(Received 24 June 2025, Accepted 10 September 2025, Scheduled 20 September 2025)

Active Sampling Strategies for Non-Embedded EMC Uncertainty Simulation

Jinjun Bai[1], Jiasheng Wang[1], Xiangrui Ji[1], Yujia Song[2,*], and Haichuan Cao[1]

[1] College of Marine Electrical Engineering, Dalian Maritime University, Dalian 116026, China
[2] School of Electrical Engineering, Dalian University of Technology, Dalian 116024, China

ABSTRACT: Non-embedded uncertainty analysis methods are widely used in the field of electromagnetic compatibility (EMC). Their essence is to construct a surrogate model to simulate the actual electromagnetic simulation process and obtain the desired uncertainty simulation results through exhaustive sampling. However, when performing complex electromagnetic compatibility simulations, non-embedded uncertainty analysis methods face an inherent problem. This problem arises from the excessive number of deterministic simulations, which leads to computational inefficiency. In this paper, an active sampling strategy based on Bayesian optimization is proposed. By selecting the locations of deterministic simulation sampling points in a more reasonable manner, the overall number of sampling points required for the uncertainty simulation can be minimized, thereby improving the computational efficiency. Finally, the effectiveness of the sampling strategy proposed in this paper was verified using a typical parallel cable crosstalk example and a lightning electromagnetic pulse electromagnetic interference simulation example.

1. INTRODUCTION

Uncertainty analysis is a means of providing a scientific description of complex environments. In electromagnetic compatibility (EMC) simulation, the actual electromagnetic environment is complex and difficult to describe. Therefore, it is more reasonable to model some simulation inputs using uncertainty variables (such as random variables and random fuzzy variables) rather than using constant models [1].

Uncertainty analysis methods are divided into three categories, namely Monte Carlo method (MCM), embedded uncertainty simulation methods, and non-embedded uncertainty simulation methods [2]. MCM is uncompetitive due to its slow convergence rate, which can be unattainable due to time cost when a single deterministic simulation takes a long time [3, 4]. Because of the development of finite element simulation technology, commercial electromagnetic simulation software is widely used in EMC simulation. The closed-source nature of its solver means that embedded uncertainty simulation methods lack competitiveness because they cannot be implemented [5]. Non-embedded uncertainty simulation methods have become the mainstream research direction due to their applicability, fast convergence, and high accuracy. They are also the first choice for practical applications.

The essence of the non-embedded uncertainty analysis method is to construct the EMC simulation process as a surrogate model, i.e., a black box. Then, exhaustive sampling is performed on the uncertain input variables, which are substituted into the surrogate model point by point for calculation. Finally, the uncertainty analysis results are obtained. Traditional surrogate modelling methods include stochastic collocation method (SCM) [6, 7], stochastic reduced order models (SROMs) [8], etc. However, they all suffer from dimensionality catastrophe problem, i.e., they are computationally inefficient when the dimensionality of uncertain input variables is high. In recent years, some novel surrogate modelling approaches have gradually received attention in EMC field, such as Kriging Model [9], Radial Basis Function (RBF) Neural Network [10], and Support Vector machine (SVM) Model [11]. Their use can initially alleviate dimensionality catastrophe problem, but has poor accuracy in dealing with higher nonlinearity (meaning that the difference in the simulation outputs is large as the inputs to the EMC simulation change) of uncertainty analysis problems with poor accuracy. Therefore, when uncertainty simulation is relatively complex, the uncertainty analysis problem itself is more difficult. At this point, the number of sampling points required by the non-embedded uncertainty analysis method increases significantly, with the inherent problem of computational inefficiency.

How sampling points are selected is a key to determining the computational performance of non-embedded uncertainty analysis methods. SCM selects sampling points based on generalized polynomial chaos theory [12]. SROM selects sampling points based on the central clustering algorithm. Kriging model [13], RBF Neural Network model [14, 15], and SVM model [16] determine sampling points based on Latin Hypercubic Sampling method. In addition, some new sampling strategies have gradually gained attention. Bai et al. proposed a Kriging Model sampling point selection strategy based on genetic algorithms and centroid clustering algorithms [17]. Huo et al. proposed a sampling strategy based on Kriging-LSSVR [18]. However, these existing sampling point selection strategies are relatively passive. (The EMC deterministic simulation results

* Corresponding author: Yujia Song (songyujia@dlut.edu.cn).

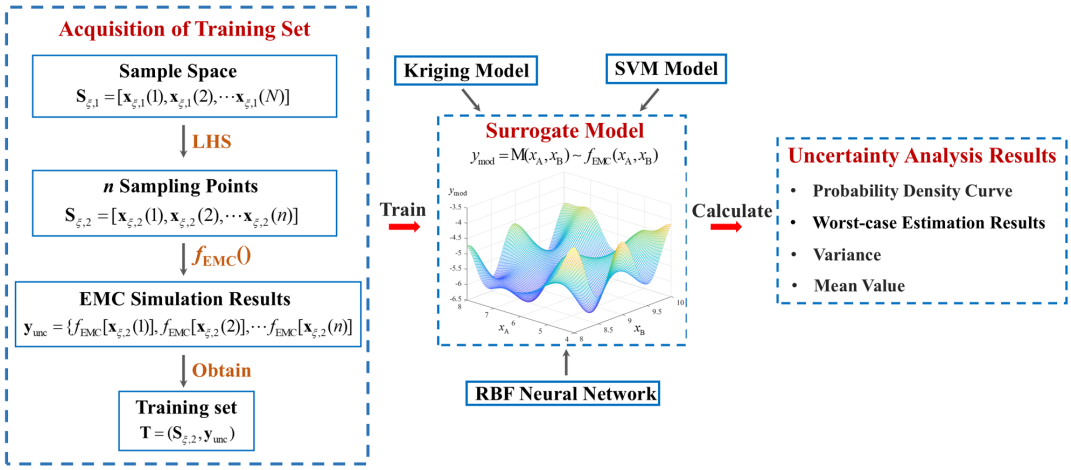

FIGURE 1. Simulation approach for non-embedded EMC uncertainty based on a surrogate model.

at the previous sampling point do not affect the selection of the next sampling point, and still follow fixed principles such as generalized polynomial chaos theory and Latin hypercube sampling.) Therefore, sampling points cannot be adjusted in a timely manner based on the characteristics of EMC deterministic simulation. This passive property is the root cause of the computational inefficiency under complex electromagnetic compatibility simulation conditions.

Bayesian optimization originated from the field of artificial intelligence, which is a reliable algorithm for performing optimal tuning of hyperparameters of the machine learning models [19, 20]. It selects the optimal training sampling points for machine learning models by constructing Gaussian process surrogate models and choosing reasonable sampling functions, and is effective in dealing with complex electromagnetic compatibility simulation conditions. Bayesian optimization is essentially an optimization algorithm, although it incorporates the solution idea of surrogate model construction. In this paper, we innovatively apply Bayesian optimization to EMC uncertainty analysis method and propose an active sampling strategy for the surrogate model to improve the overall computational efficiency of non-embedded uncertainty analysis method.

The structure of this paper is as follows. A non-embedded EMC uncertainty analysis method is introduced in Section 2. In Section 3, an active sampling strategy for surrogate models is proposed. The effectiveness of the algorithm is verified in Section 4 using a classical parallel cable crosstalk example and the simulation example of electromagnetic interference of lightning electromagnetic pulse. Section 5 summarizes this paper.

2. INTRODUCTION TO NON-EMBEDDED EMC UNCERTAINTY ANALYSIS METHODS

The process of non-embedded EMC uncertainty simulation is shown in Equation (1), where ξ represents a set of random events, which can be represented by random variables or random fuzzy variables. In this example, it is assumed to be a combination of events ξ_1 and ξ_2. $x_{\mathrm{A}}(\xi_1)$ and $x_{\mathrm{B}}(\xi_2)$ are EMC uncertainty input parameters affected by random events. $f_{\mathrm{EMC}}()$ is a stable EMC simulation solver, and non-embedded means that no internal modifications are required. Finally, $y(\xi)$ represents the EMC uncertainty simulation output. The process of obtaining it is called uncertainty analysis, which can be represented by the mean, standard deviation, worst-case estimate, probability density curve, etc.

$$y(\xi)=f_{\mathrm{EMC}}[x_{\mathrm{A}}(\xi_1),x_{\mathrm{B}}(\xi_2)] \tag{1}$$

Assuming that random events can be described by random variable models, i.e., $\xi_1 \sim \mathrm{pdf}(\xi_1)$ and $\xi_2 \sim \mathrm{pdf}(\xi_2)$, large-scale sampling can be performed based on probability density functions, with sampling points $\mathbf{S}_{\xi,1}=[\mathbf{x}_{\xi,1}(1),\mathbf{x}_{\xi,1}(2),\ldots,\mathbf{x}_{\xi,1}(N)]$ and each point taking the form $\mathbf{x}_{\xi,1}(i)=[x_{\mathrm{A},1}(i),x_{\mathrm{B},1}(i)]$. N is a large number, so it is clear that the computational cost of performing a deterministic EMC simulation $f_{\mathrm{EMC}}[\mathbf{x}_{\xi,1}(i)]$ for each point to obtain the uncertainty output result $y(\xi)$ is high.

The research approach for non-embedded EMC uncertainty simulation based on the surrogate model is shown in Figure 1. It follows the above-mentioned large number of sampling points $\mathbf{S}_{\xi,1}=[\mathbf{x}_{\xi,1}(1),\mathbf{x}_{\xi,1}(2),\ldots,\mathbf{x}_{\xi,1}(N)]$ to obtain n small sampling points $\mathbf{S}_{\xi,2}=[\mathbf{x}_{\xi,2}(1),\mathbf{x}_{\xi,2}(2),\ldots,\mathbf{x}_{\xi,2}(n)]$ using Latin Hypercube Sampling, where $\mathbf{x}_{\xi,2}(i)=[x_{\mathrm{A},2}(i),x_{\mathrm{B},2}(i)]$. These n sampling points are input into EMC simulation solver $f_{\mathrm{EMC}}()$ to obtain a combination of point-by-point deterministic EMC simulation results $\mathbf{y}_{\mathrm{unc}}=\{f_{\mathrm{EMC}}[\mathbf{x}_{\xi,2}(1)],f_{\mathrm{EMC}}[\mathbf{x}_{\xi,2}(2)],\ldots,f_{\mathrm{EMC}}[\mathbf{x}_{\xi,2}(n)]$. This simulation result is based on sampling point $\mathbf{S}_{\xi,2}$ being included in the training set $\mathbf{T}=(\mathbf{S}_{\xi,2},\mathbf{y}_{\mathrm{unc}})$. Each training point is in the form of $\mathbf{T}(i)=\{[x_{\mathrm{A},2}(i),x_{\mathrm{B},2}(i)],\mathbf{y}_{\mathrm{unc}}(i)\}$. A combination of n training sets $\mathbf{T}(i)=\{[x_{\mathrm{A},2}(i),x_{\mathrm{B},2}(i)],\mathbf{y}_{\mathrm{unc}}(i)\}$ is used as input to train the surrogate model. The process is as follows:

$$y_{\mathrm{mod}}=M(x_{\mathrm{A}},x_{\mathrm{B}})\sim f_{\mathrm{EMC}}(x_{\mathrm{A}},x_{\mathrm{B}}) \tag{2}$$

Among them, y_{mod} is a surrogate model, which can be regarded as a black box that can replace the deterministic EMC simulation process. Different methods construct surrogate models in various ways. Kriging Model is based on a geostatistical

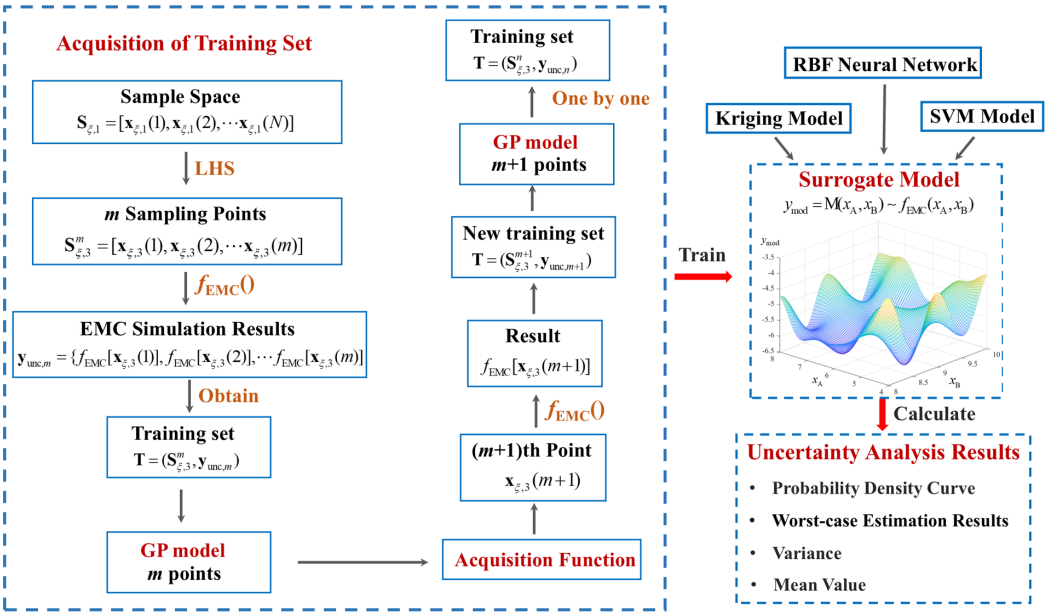

FIGURE 2. Research approach for active sampling strategies based on Bayesian optimization.

interpolation method, and RBF Neural Network is based on a feedforward neural network and a radial basis function activation function, and SVM Model is based on statistical learning theory classification and regression methods.

After constructing the surrogate model, many sampling points $\mathbf{S}_{\xi,1} = [\mathbf{x}_{\xi,1}(1), \mathbf{x}_{\xi,1}(2), \ldots, \mathbf{x}_{\xi,1}(N)]$ can be entered into y_{mod} to estimate the results $y_{\text{mod}}(\mathbf{S}_{\xi,1})$ for all points. The purpose is to estimate $f_{\text{EMC}}(\mathbf{S}_{\xi,1})$ approximately. Finally, after a simple statistical analysis of $y_{\text{mod}}(\mathbf{S}_{\xi,1})$, the results of uncertainty analysis, such as mean, standard deviation, worst-case estimate, and probability density curve, can be obtained.

Notably, uncertainty analysis based on the surrogate model only requires n deterministic EMC simulations, which is far less than the N simulations required by MCM. In other words, when the cost of a single EMC simulation is high, the surrogate model has a significant advantage over MCM in terms of computational efficiency.

3. ACTIVE SAMPLING STRATEGY BASED ON BAYESIAN OPTIMIZATION

For uncertainty analysis methods based on surrogate models, the scientific selection of n sampling points $\mathbf{S}_{\xi,2} = [\mathbf{x}_{\xi,2}(1), \mathbf{x}_{\xi,2}(2), \ldots, \mathbf{x}_{\xi,2}(n)]$ is critical for determining the accuracy of the surrogate model. There are two key issues here: First, how to determine whether the surrogate model has converged at a given number of sampling points. This issue has been addressed by the method proposed in [21]. Second, it is critical to determine the rules for selecting sampling points that can maximize the coverage of the random variable space.

The Latin hypercube sampling method performs well in terms of computational performance when the random variable dimension is low, and the deterministic EMC simulation is nonlinear. This is because uncertainty analysis problem is relatively simple at this point, and the entire random variable space can be traversed with a small number of sampling points. However, when the random variable dimension is high, and the deterministic EMC simulation is highly nonlinear (complex electromagnetic compatibility simulation), the required number of sampling points n will increase accordingly, until the time cost becomes unfeasible.

Sensitivity analysis methods can be used for ranking and dimension reduction in simulation problems involving high-dimensional random variables. Bai et al. provided an excellent solution [22]. Therefore, the focus of this study is on how to deal with complex electromagnetic compatibility simulation, and an active sampling strategy based on Bayesian optimization is proposed, as shown in Figure 2.

Bayesian optimization framework mainly consists of two parts, namely probabilistic surrogate model (note: this is not the same concept as the surrogate model in the surrogate model uncertainty analysis method mentioned in this paper) and acquisition function. The most used probabilistic surrogate model is Gaussian Process (GP) model. The modelling process of this model is recursive, involving the continuous modification of model parameters to approximate the current objective function. GP model is a non-parametric model. Formally, each finite subset of the model follows a multivariate normal distribution, consisting of a mean function m and a covariance function k. Given input samples $\mathbf{S}_2 = [\mathbf{x}_2(1), \mathbf{x}_2(2), \ldots, \mathbf{x}_2(n_2)]$ and their corresponding output values $\mathbf{y}_{\text{op},2} = \{f_{\text{EMC}}[\mathbf{x}_2(1)], f_{\text{EMC}}[\mathbf{x}_2(2)], \ldots, f_{\text{EMC}}[\mathbf{x}_2(n_2)]$, the specific multivariate Gaussian distribution is as follows:

$$\mathbf{y}_{\text{op},2} \sim \text{N}\left[m(\mathbf{S}_2), k(\mathbf{S}_2, \mathbf{S}_2')\right] \tag{3}$$

Among them, $m(\mathbf{S}_2)$ is the mean function; $k(\mathbf{S}_2, \mathbf{S}_2')$ represents a semi-positive covariance matrix; k is the corresponding covariance function, which is used to characterize the differ-

ences between paired input points. Covariance function is also known as kernel function. Commonly used kernel functions include Gaussian kernel function, exponential kernel function, and polynomial kernel function.

The next prediction point $\{x(*), y(*)\}$ has the following joint distribution with the training set:

$$\begin{bmatrix} \mathbf{y}_{\text{op},2} \\ y(*) \end{bmatrix} \sim \text{N}\left(\begin{bmatrix} m(\mathbf{S}_2) \\ m[x(*)] \end{bmatrix}, \begin{bmatrix} K & K_* \\ K_*^T & K_{**} \end{bmatrix} \right) \quad (4)$$

Among them, K is the covariance matrix formed by $k(\mathbf{S}_2, \mathbf{S}_2')$, $K_* = k[\mathbf{S}_2, x(*)]$, and $K_{**} = k[x(*), x(*)]$. The conditional distribution of the predicted value $y(*)$ also satisfies the multivariate Gaussian distribution:

$$p[y(*)|x(*), \mathbf{S}_2, \mathbf{y}_{\text{op},2}] = \text{N}[y(*)|\mu_*, \Sigma_*] \quad (5)$$

From conditional distribution properties of the multivariate Gaussian distribution, Equation (6) can be obtained:

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix} \sim \text{N}\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} A & C \\ C^T & B \end{bmatrix} \right)$$

$$\Rightarrow \mathbf{x}|\mathbf{y} \sim \text{N}(\mu_x + CB^{-1}(\mathbf{y} - \mu_y), A - CB^{-1}C^T) \quad (6)$$

Based on Equation (4), the estimated values of mean μ_* and posterior covariance $\sum_*$ can be obtained.

$$\mu_* = m[x(*)] + K_*^T K^{-1}(\mathbf{y}_{\text{op},2} - m(\mathbf{S}_2)) \quad (7)$$

$$\sum_* = K_{**} - K_*^T K^{-1} K_* \quad (8)$$

Based on the above principle, a GP model can be constructed according to the training set $\mathbf{T}$:

$$y_{\text{GP}} = \text{GP}_{\text{mod}}(x_\text{A}, x_\text{B}) \sim f(x_\text{A}, x_\text{B}) \quad (9)$$

As can be clearly seen from Figure 2, the GP model can be continuously adjusted, and adjustments are made point by point to maximize the convergence of the surrogate model.

Another key component of Bayesian optimization framework is acquisition function, which is constructed based on the posterior probability distribution $\text{N}[y(*)|\mu_*, \Sigma_*]$ of the observed samples. By maximizing the acquisition function, the next most "promising" evaluation point can be selected. The collection function needs to consider two aspects: First, harvesting function utilizes the currently developed area (Exploitation), i.e., it continues to search near the current extreme value. Second, harvesting function explores the undeveloped area (Exploration), i.e., it searches in areas that have not yet been searched.

It is worth noting that in optimization problems, the purpose of Bayesian optimization algorithms is to find the optimal solution. Therefore, the algorithm's ability to explore unknown areas only plays a supporting role, with the aim of preventing the omission of the global optimal solution. However, when being used for uncertainty simulation, exploring undeveloped areas is clearly more important, which maximizes the descriptive power of the surrogate model for deterministic EMC simulations. Common collection functions include Expected Improvement (EI) function, the Probability of Improvement (PI) function, and collection functions based on confidence boundaries, such as Upper Confidence Bound (UCB) function. Among them, EI focuses on improving the optimal solution $y_{\text{best}} = \min_{m \le k} f(x_m)$ of the current k sample points, and the improvement value is recorded as $\max(0, y_{\text{best}} - y)$. Therefore, when the next experimental point is selected, the goal is to maximize the expected result of this value:

$$\text{EI} = \max\{\text{E}[\mathbf{y}_{\text{op},1}(i)]\} \quad (10)$$

The difference is that expected variance (EV) uses the maximum variance value at existing grid points as the criterion for judgement [23], with the aim of exploring unknown areas to establish a more accurate Gaussian process model:

$$\text{EV} = \max\{\sigma[\mathbf{y}_{\text{op},1}(i)]\} \quad (11)$$

Clearly, in uncertainty analysis methods based on surrogate models, the key issue is how to quickly construct surrogate models that accurately describe deterministic EMC simulation processes. Therefore, EV should be selected as the sampling function [24], which is also the key point in the innovative application of Bayesian optimization to uncertainty analysis in this paper.

Returning to the overall algorithm research approach in Figure 2, m sampling points $\mathbf{S}_{\xi,3}^m = [\mathbf{x}_{\xi,3}(1), \mathbf{x}_{\xi,3}(2), \ldots, \mathbf{x}_{\xi,3}(m)]$ are initially selected based on Latin hypercube sampling, with the number of points being less than the actual number of convergent sampling points n. It is worth noting that the usual meaning of "hyperparameter" corresponds to "ξ" in Figure 2. Similarly, m sampling points are subjected to a point-by-point deterministic EMC simulation process to obtain the corresponding result $\mathbf{y}_{\text{unc},m} = \{f_{\text{EMC}}[\mathbf{x}_{\xi,3}(1)], f_{\text{EMC}}[\mathbf{x}_{\xi,3}(2)], \ldots, f_{\text{EMC}}[\mathbf{x}_{\xi,3}(m)]$, forming training set combination $\mathbf{T} = (\mathbf{S}_{\xi,3}^m, \mathbf{y}_{\text{unc},m})$. First, an initial GP model is constructed for m sampling points. The sampling function EV is used to select and increase the number of sampling points point by point. Assume that the $m+1$th sampling point is $\mathbf{x}_{\xi,3}(m+1)$. A single deterministic EMC simulation is performed on the $m+1$th sampling point to obtain $f_{\text{EMC}}[\mathbf{x}_{\xi,3}(m+1)]$, forming a new training point combination $\mathbf{T} = (\mathbf{S}_{\xi,3}^{m+1}, \mathbf{y}_{\text{unc},m+1})$. By using Equations (4) to (9), a new GP model can be obtained. Based on the sampling function, the process of "selecting the next sampling point → performing a single deterministic EMC simulation → applying Equations (4) to (9)" is repeated iteratively to obtain the final training set $\mathbf{T} = (\mathbf{S}_{\xi,3}^n, \mathbf{y}_{\text{unc},n})$ for the surrogate model. It is worth noting that the convergence criterion in [18] can be used to determine whether to stop increasing the number of sampling points. Finally, apply Equation (2) to construct the final surrogate model, i.e., Kriging Model, RBF Neural Network, or SVM Model.

In summary, the active sampling strategy based on Bayesian optimization can effectively utilize deterministic EMC simulation results. By increasing the number of sampling points point by point, it can maximize the accuracy of the surrogate model's description of the EMC simulation process, thereby improving the overall computational performance of EMC uncertainty simulation.

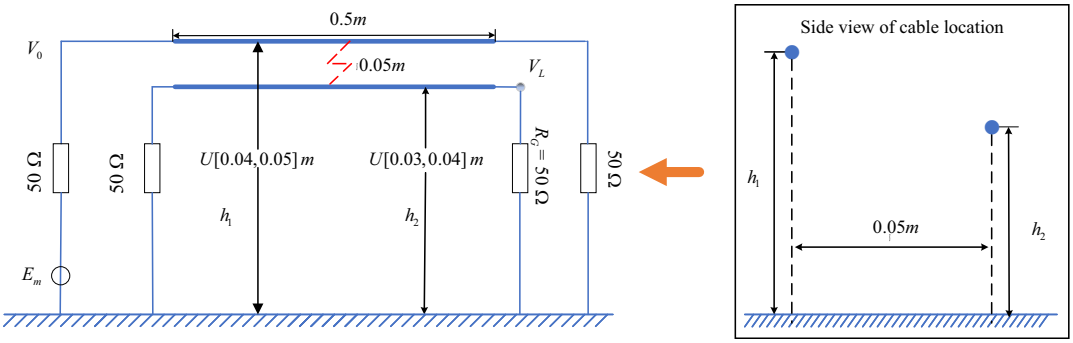

FIGURE 3. Example of parallel cable crosstalk.

TABLE 1. Number of sampling points required under the Latin hypercube sampling method.

Surrogate model	Number of simulations	MEAM values at 75 MHz	MEAM values for the 1 MHz ∼ 200 MHz frequency band
SVM model	36	0.97	0.98
Kriging model	46	0.97	0.98
RBF neural network model	64	0.96	0.98

TABLE 2. Number of sampling points and MEAM values for two sampling methods.

Surrogate model	Number of simulations	MEAM values at 75 MHz	MEAM values for the 1 MHz ∼ 200 MHz frequency band
SVM model	$10 + 12 = 22$	0.96	0.96
Kriging model	$10 + 24 = 34$	0.97	0.99
RBF neural network model	$10 + 37 = 47$	0.97	0.98

4. VALIDATION OF THE EFFECTIVENESS OF ACTIVE SAMPLING STRATEGIES

4.1. The Example of Parallel Cable Crosstalk

A classic parallel cable crosstalk example is used to validate the active sampling strategy proposed in this paper, as shown in Figure 3. To clearly present the sampling results, the parameters used in this paper are consistent with those in [18]. The amplitude E_m of the excitation source is 1 V. Both conductors have a radius of 0.1 mm and a length of 0.5 m. The horizontal distance between the two conductors is 0.05 m. All loads are $50\,\Omega$. Both conductors are surrounded by vacuum, and the relative dielectric constant and relative magnetic permeability of the vacuum are both 1. The height of the cable is an uncertain input parameter, which can be described by the following random variable model:

$$\begin{cases} h_1(\xi_1) = 0.045 + 0.005 \times \xi_1 \text{ [m]} \\ h_2(\xi_2) = 0.035 + 0.005 \times \xi_2 \text{ [m]} \end{cases} \tag{12}$$

Among them, ξ_1 and ξ_2 are uniformly distributed random variables in the interval $[-1, 1]$.

The overall simulation uncertainty output result is the remote crosstalk voltage in decibels at each frequency point, as shown below:

$$\mathrm{V_{dB}}(\xi) = 20 \log_{10} \frac{|\mathrm{V_L}(\xi_1, \xi_2)|}{|\mathrm{V_0}|} \tag{13}$$

Figure 4 shows the simulation results of three types of efficient non-embedded uncertainty analysis methods: Kriging Model, RBF Neural Network, and SVM Model. MCM is also provided as standard data. Under the traditional Latin hypercube sampling method, the number of sampling points required for the three surrogate models is determined based on the convergence criteria in [18], as shown in Table 1. Among them, mean equivalent area method (MEAM) values represent the difference between the uncertainty analysis results and standard data. The closer the value is to 1, the smaller the difference is, indicating higher accuracy of the uncertainty analysis results [22]. Detailed descriptions of all the above results can be found in [18].

The uncertainty analysis results obtained using the active sampling strategy described in this paper are shown in Figure 5, and the corresponding number of sampling points required is shown in Table 2. Among them, the number of sampling points $m = 10$ for constructing the initial GP model is $\mathbf{S}_{\xi,3}^{10} = [\mathbf{x}_{\xi,3}(1), \mathbf{x}_{\xi,3}(2), \ldots, \mathbf{x}_{\xi,3}(10)]$. Following the strategy outlined in Figure 3, the process of "selecting the next sampling

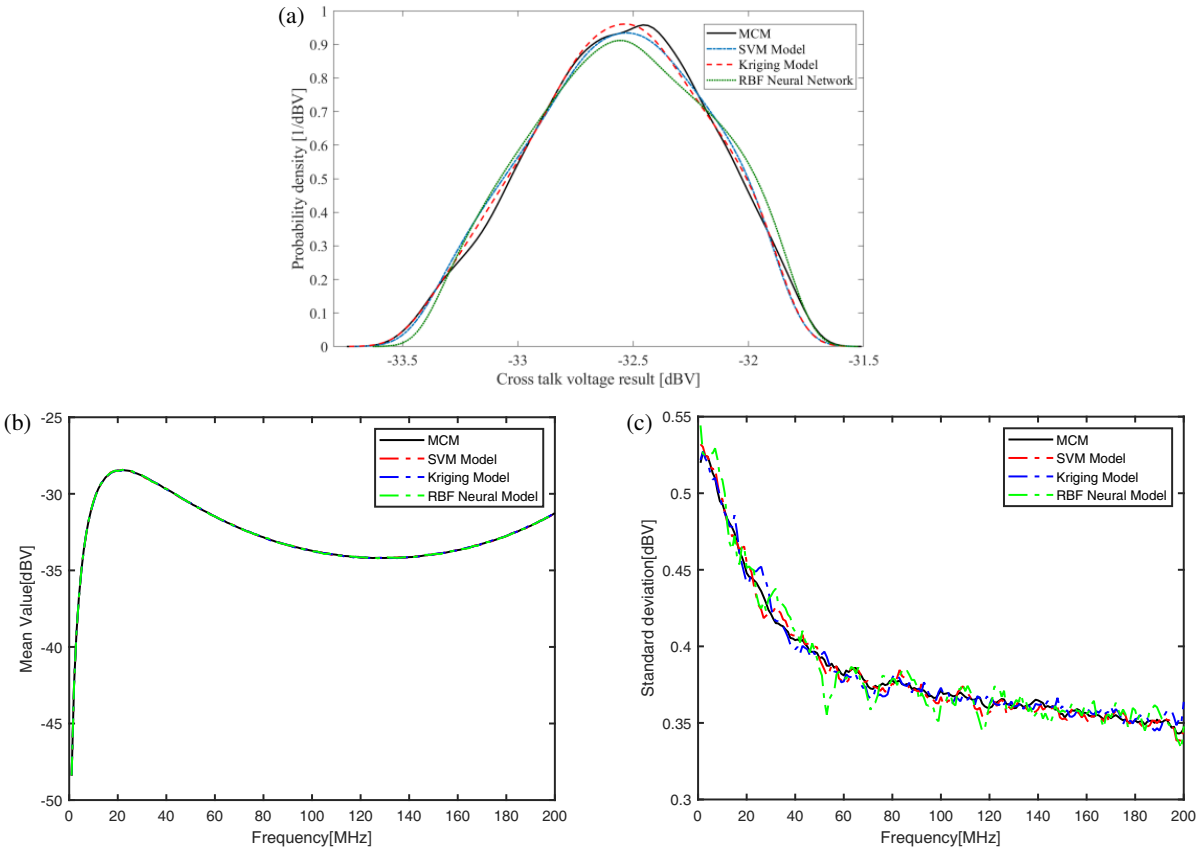

FIGURE 4. Uncertainty analysis results under the Latin hypercube sampling method. (a) Probability density curve results of V_{dB} at 75 MHz. (b) Mean results for the 1 MHz $\sim$ 200 MHz frequency band. (c) Standard deviation results for the 1 MHz $\sim$ 200 MHz frequency band.

TABLE 3. The time cost of the active sampling strategy compared with the traditional sampling strategy in the parallel cable crosstalk example.

Surrogate model	Active sampling strategies simulation				Traditional simulation			
	Active sampling time	Uncertainty analysis method time	Deterministic simulation time	Active sampling total time	Latin hypercube sampling time	Uncertainty analysis method time	Deterministic simulation time	Latin hypercube sampling total time
SVM model	3.4 s	1.50 s	0.46 s × 22 = 10.12 s	15.02s	0.01s	1.50 s	0.46 s × 36 = 16.56 s	18.07 s
Kriging model	5.3 s	1.54 s	0.46 s × 34 = 15.64 s	22.48 s		1.54 s	0.46 s × 46 = 21.16 s	22.71 s
RBF neural network model	71 s	1.57 s	0.46 s × 47 = 21.62 s	30.29 s		1.57 s	0.46 s × 64 = 29.44 s	31.02 s

point $\rightarrow$ single deterministic EMC simulation $\rightarrow$ Equations (4) to (9)" is repeated in a loop, with sampling points added one by one. The convergence determination strategy follows the rules outlined in [18] to achieve uncertainty analysis under three surrogate models. The SVM model added 12 additional sampling points, requiring a total of 22 sampling points, which is $22 \div 36 = 61\%$ in the cost calculation under the Latin hypercube sampling method. The Kriging model added 24 additional sampling points, requiring a total of 34 sampling points, which is $34 \div 46 = 74\%$ in the calculation of costs under the Latin hypercube sampling method. The RBF neural network model added 37 additional sampling points, requiring a total of 47 sampling points, which is $47 \div 64 = 73\%$ in the calculation cost under the Latin hypercube sampling method. Therefore, this example verifies that the active sampling strategy proposed in this paper can effectively improve the computational performance of the surrogate model method in performing EMC uncertainty analysis.

It is worth noting that three MEAM values greater than 0.95 can ensure the accuracy of uncertainty analysis results [22].

As shown in Table 2, the active sampling strategy proposed in this paper significantly reduces the number of deterministic simulations. However, as shown in Table 3, due to the short simulation time per run in this example, advantages of

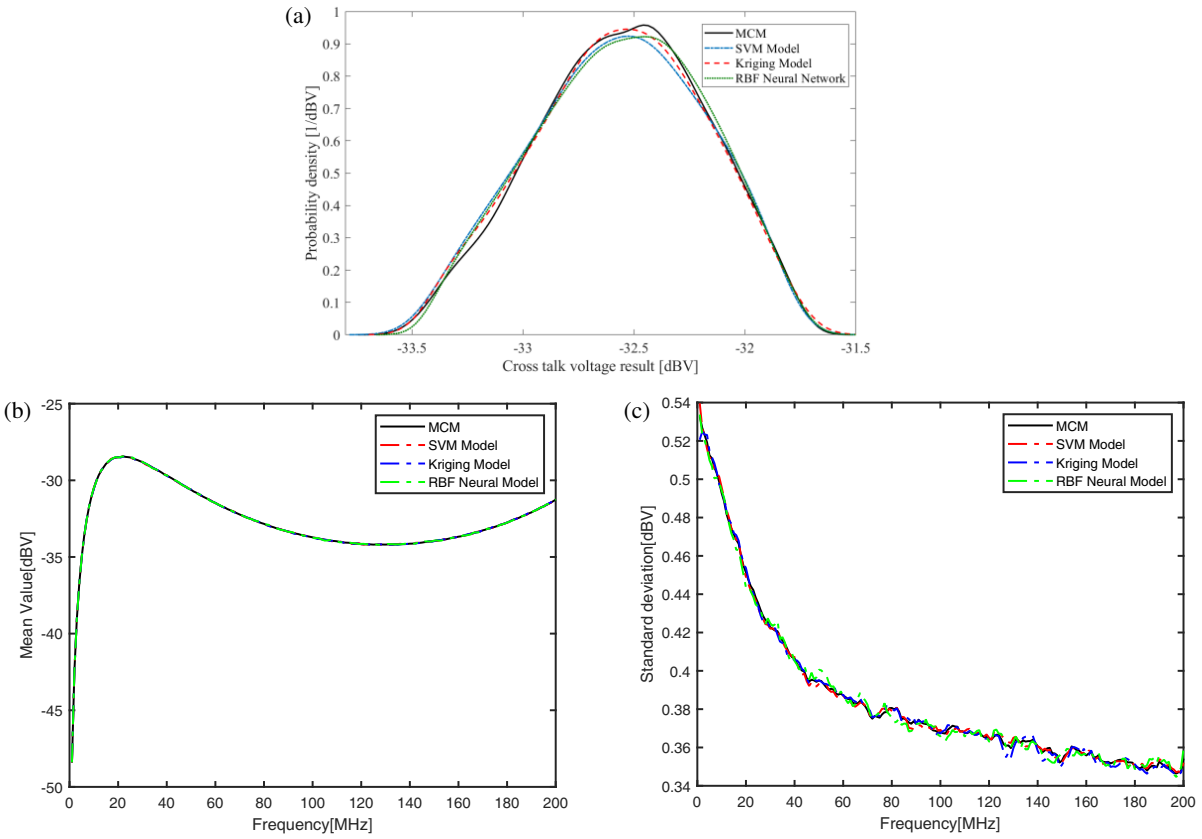

FIGURE 5. Uncertainty analysis results under the active sampling strategy proposed in this paper. (a) Probability density curve results of V_{dB} at 75 MHz. (b) Mean results for the 1 MHz $\sim$ 200 MHz frequency band. (c) Standard deviation results for the 1 MHz $\sim$ 200 MHz frequency band.

the active sampling strategy proposed in this paper are not fully demonstrated. In order to demonstrate advantages of the active sampling strategy, the following will introduce a simulation example of lightning electromagnetic pulse electromagnetic interference. It will discuss the computational cost, compared to the traditional Latin hypercube sampling method, when a single deterministic simulation takes a long time and needs to be repeated many times. This will serve to verify the superiority of the active sampling strategy proposed in this paper.

4.2. The Simulation Example of Electromagnetic Interference of Lightning Electromagnetic Pulse

An example of electromagnetic interference of lightning electromagnetic pulse is used to validate the active sampling strategy proposed in this paper. To clearly present sampling results, the parameters used in this paper are consistent with those in [24]. This calculation example is the prediction of the electromagnetic radiation interference of a lightning electromagnetic pulse on the sensitive devices inside the aircraft, which is derived from the example simulation case of COMSOL Multiphysics software [25]. Because lightning only causes significant electromagnetic interference to the aircraft for a short period of time after its occurrence [26], the lightning current is only modeled within 0 to 6 µs, and the transient current curve is shown in Figure 6. Simulated output is the electromagnetic shielding effectiveness of the aircraft shielding effectiveness

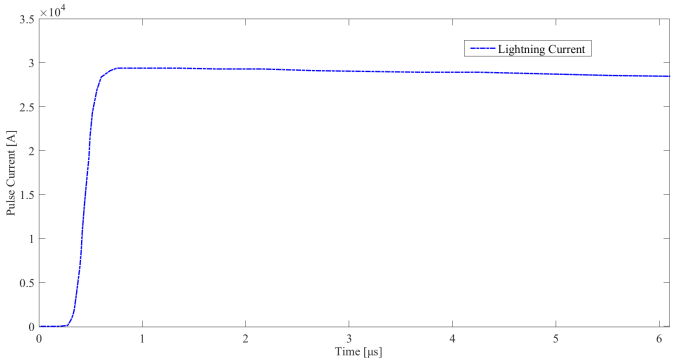

FIGURE 6. Lightning current transient current curve within 0–-6 µs [24].

(SE), as shown in

$$\mathrm{SE} = 0 \lg \frac{|V_2|}{|V_1|} \tag{14}$$

The transient simulation results of the aircraft surface electric field at 2.5 µs are shown in Figure 7.

Assuming that L_x and L_y coordinate values of lightning location information are uncertain input parameters, described by the following random variables model:

$$\begin{cases} L_x(\xi_1) = -30 - 2 \times \xi_1 \text{ [m]} \\ L_y(\xi_2) = 8 + 2 \times \xi_2 \text{ [m]} \end{cases} \tag{15}$$

TABLE 4. Number of sampling points and MEAM values for two sampling methods.

Surrogate model	Number of simulations (Traditional simulation)	MEAM values (Traditional simulation)	Number of simulations (active sampling strategies simulation)	MEAM values (active sampling strategies simulation)
SVM model	81	0.96	$20 + 34 = 54$	0.97
Kriging model	95	0.97	$20 + 42 = 62$	0.98
RBF neural network model	125	0.98	$20 + 59 = 79$	0.97

TABLE 5. The time cost of the active sampling strategy compared with the traditional sampling strategy in the electromagnetic interference of lightning electromagnetic pulse example.

Surrogate model	Active sampling strategies simulation				Traditional simulation			
	Active sampling time	Uncertainty analysis method time	Deterministic simulation time	Total time	Latin hypercube sampling time	Uncertainty analysis method time	Deterministic simulation time	Total time
SVM model	8.3 s	2.23 s	4.26 min × 54 = 230.04 min	230.2 min		2.23 s	4.26 min × 81 = 345.06 min	345.1 min
Kriging model	9.7 s	2.31 s	4.26 min × 62 = 264.12 min	264.3	0.5s	2.31 s	4.26 min × 95 = 404.7 min	404.7 min
RBF neural network model	13.4 s	2.37 s	4.26 min × 79 = 336.54 min	336.8 min		2.37 s	4.26 min × 125 = 532.5 min	532.5 min

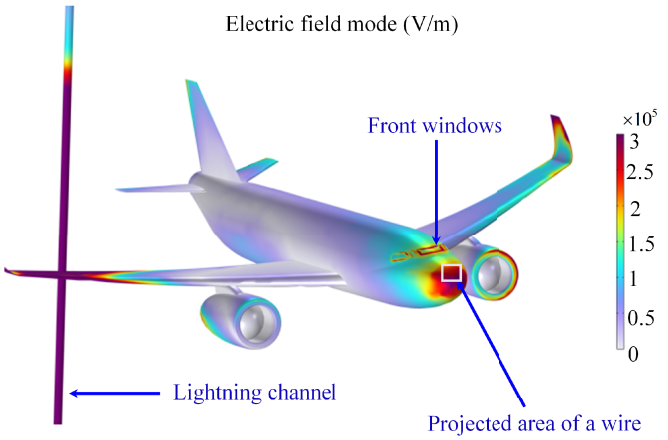

FIGURE 7. Transient simulation results of aircraft surface electric field at 2.5 μs [24].

where ξ_1 and ξ_2 are random variables uniformly distributed in the interval $[-1, 1]$, the value to be solved is the electromagnetic SE of the aircraft.

The uncertainty simulation results of two sampling methods are shown in Figure 8, and the corresponding number of sampling points required is shown in Table 4. Among them, the SVM model added 34 additional sampling points, requiring a total of 54 sampling points, which corresponds to the computational cost $54 \div 81 = 67\%$ under the Latin hypercube sampling method; the Kriging model added 42 additional sampling points, requiring a total of 62 sampling points, which corresponds to the computational cost $62 \div 95 = 65\%$ under the Latin hypercube sampling method; the RBF neural network model added 68 additional sampling points, requiring a total of 88 sampling points, corresponding to computational cost $79 \div 125 = 63\%$ under the Latin hypercube sampling method.

From the perspective of time cost, as shown in Table 5, the total computation time for the SVM model decreased from 345.1 minutes to 230.04 minutes; the total computation time for the Kriging model decreased from 404.7 minutes to 264.3 minutes; the total computation time decreased from 532.5 minutes to 336.8 minutes. Therefore, the active sampling strategy proposed in this paper significantly improves the computational efficiency of this example. The effect will be even more pronounced when single deterministic simulations are performed with longer durations and a greater number of iterations.

Therefore, this example verifies that the active sampling strategy proposed in this paper can effectively improve the computational performance of the surrogate model method when performing EMC uncertainty analysis. It is especially suitable for scenarios where the single deterministic simulation time is protracted, and the number of simulations is elevated. Similarly, three "MEAM values" greater than 0.95 ensure the accuracy of the uncertainty analysis results [22].

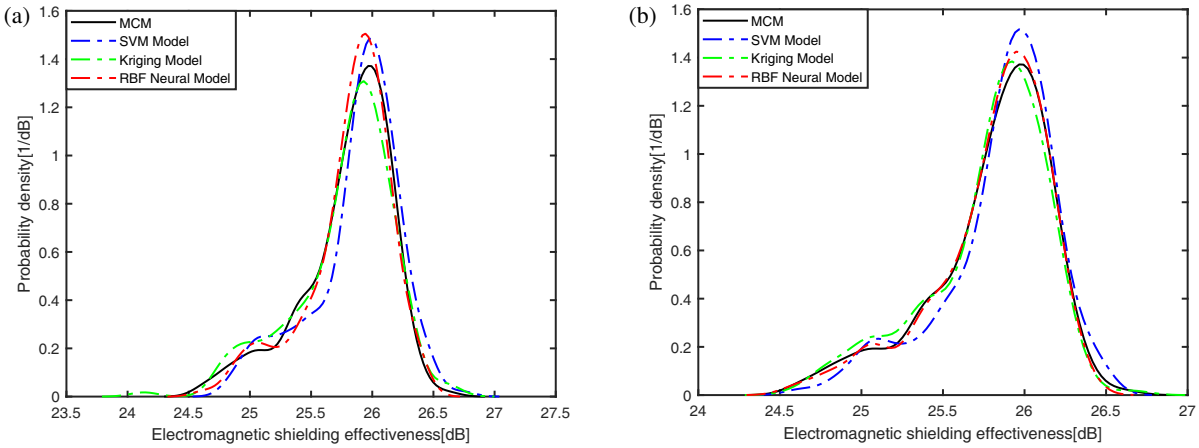

FIGURE 8. Uncertainty simulation results of two sampling methods. (a) Latin hypercube sampling. (b) Active sampling strategy proposed in this pap.

5. CONCLUSION

For non-embedded EMC uncertainty analysis methods, this paper proposes an active sampling strategy based on Bayesian optimization, which improves computational accuracy by constructing surrogate models with fewer sampling points. This strategy makes the reasonable use of all deterministic EMC simulation results, effectively exploring the random variable input space and selecting the next most "promising" sampling point. Compared with traditional Latin hypercube sampling methods, the active sampling strategy proposed in this paper can theoretically handle complex electromagnetic compatibility simulation more effectively. In the typical parallel cable crosstalk example, quantitative verification showed that the numbers of deterministic simulations of the Kriging Model, RBF Neural Network, and SVM Model were 74%, 73%, and 61% of those under traditional sampling strategies. In the simulation example of electromagnetic interference of lightning electromagnetic pulse, quantitative verification showed that uncertainty analysis time costs of the Kriging Model, RBF Neural Network, and SVM Model were 67%, 65%, and 63% of those under traditional sampling strategies, respectively, fully proving the effectiveness of the active sampling strategy proposed in this paper.

ACKNOWLEDGEMENT

This paper is supported by "the Fundamental Research Funds for the Central Universities" (Project No. 3132025101).

REFERENCES

[1] Lalléchère, S., C. F. M. Carobbi, and L. R. Arnaut, "Review of uncertainty quantification of measurement and computational modeling in EMC Part II: Computational uncertainty," *IEEE Transactions on Electromagnetic Compatibility*, Vol. 61, No. 6, 1699–1706, 2019.

[2] Bai, J., S. Huo, A. Duffy, and B. Hu, "Improvement of nonembedded EMC uncertainty analysis methods based on data fusion technique," *IEEE Transactions on Electromagnetic Compatibility*, Vol. 66, No. 6, 1999–2009, Dec. 2024.

[3] Zhu, Z., J. Yang, Y. Shi, X. Nie, W. Wu, and W. Chen, "Monte Carlo simulation based uncertainty analysis of coupling electric fields in HEMP field tests," *IEEE Transactions on Electromagnetic Compatibility*, Vol. 63, No. 6, 1951–1961, 2021.

[4] Zhang, Y., C. Liao, R. Huan, Y. Shang, and H. Zhou, "Analysis of nonuniform transmission lines with a perturbation technique in time domain," *IEEE Transactions on Electromagnetic Compatibility*, Vol. 62, No. 2, 542–548, 2020.

[5] Bauer, S., W. Renhart, and O. Bíró, "FEM-based computation of circuit parameters for testing fast transients for EMC problems," *IEEE Transactions on Magnetics*, Vol. 53, No. 6, 1–4, 2017.

[6] Bai, J., G. Zhang, A. P. Duffy, and L. Wang, "Dimension-reduced sparse grid strategy for a stochastic collocation method in EMC software," *IEEE Transactions on Electromagnetic Compatibility*, Vol. 60, No. 1, 218–224, 2018.

[7] Manfredi, P., D. V. Ginste, D. D. Zutter, and F. G. Canavero, "On the passivity of polynomial chaos-based augmented models for stochastic circuits," *IEEE Transactions on Circuits and Systems I: Regular Papers*, Vol. 60, No. 11, 2998–3007, 2013.

[8] Fei, Z., Y. Huang, J. Zhou, and Q. Xu, "Uncertainty quantification of crosstalk using stochastic reduced order models," *IEEE Transactions on Electromagnetic Compatibility*, Vol. 59, No. 1, 228–239, 2017.

[9] Ren, Z., J. Ma, Y. Qi, D. Zhang, and C.-S. Koh, "Managing uncertainties of permanent magnet synchronous machine by adaptive Kriging assisted weight index Monte Carlo simulation method," *IEEE Transactions on Energy Conversion*, Vol. 35, No. 4, 2162–2169, 2020.

[10] Hu, B., Y. Wang, S. Huo, and J. Bai, "Application of improved SROM based on RBF neural network model in EMC worst case estimation," *Progress In Electromagnetics Research Letters*, Vol. 119, 51–57, 2024.

[11] Memon, Z. A., R. Trinchero, P. Manfredi, F. Canavero, I. S. Stievano, and Y. Xie, "Machine learning for the uncertainty quantification of power networks," *IEEE Letters on Electromagnetic Compatibility Practice and Applications*, Vol. 2, No. 4, 138–141, 2020.

[12] Xiu, D. and G. E. Karniadakis, "The Wiener-Askey polynomial chaos for stochastic differential equations," *SIAM Journal on Scientific Computing*, Vol. 24, No. 2, 619–644, 2002.

[13] Ren, Z., J. Ma, Y. Qi, D. Zhang, and C.-S. Koh, "Managing uncertainties of permanent magnet synchronous machine by adaptive Kriging assisted weight index Monte Carlo simulation

method," *IEEE Transactions on Energy Conversion*, Vol. 35, No. 4, 2162–2169, 2020.

[14] Yu, Z., Z. Qing, and M. Yan, "Application of chaos immune optimization RBF network in dynamic deformation prediction," *Geodesy and Geodynamics*, Vol. 32, No. 5, 53–57, 2012.

[15] Xia, L., P. Hu, K. Ma, and L. Yang, "Research on measurement modeling of spherical joint rotation angle based on RBF-ELM network," *IEEE Sensors Journal*, Vol. 21, No. 20, 23 118–23 124, 2021.

[16] Memon, Z. A., R. Trinchero, P. Manfredi, F. Canavero, I. S. Stievano, and Y. Xie, "Machine learning for the uncertainty quantification of power networks," *IEEE Letters on Electromagnetic Compatibility Practice and Applications*, Vol. 2, No. 4, 138–141, 2020.

[17] Bai, J., B. Hu, and Z. Xue, "EMC uncertainty simulation method based on improved Kriging model," *IEEE Letters on Electromagnetic Compatibility Practice and Applications*, Vol. 5, No. 4, 127–130, 2023.

[18] Huo, S., Y. Song, Q. Liu, and J. Bai, "Improving kriging surrogate model for EMC uncertainty analysis using LSSVR," *Applied Computational Electromagnetics Society Journal (ACES)*, Vol. 39, No. 7, 614–622, Jul. 2024.

[19] Tang, S., Y. Zhu, and S. Yuan, "Intelligent fault diagnosis of hydraulic piston pump based on deep learning and Bayesian optimization," *ISA Transactions*, Vol. 129, 555–563, 2022.

[20] Sun, D., J. Xu, H. Wen, and D. Wang, "Assessment of landslide susceptibility mapping based on Bayesian hyperparameter optimization: A comparison between logistic regression and random forest," *Engineering Geology*, Vol. 281, 105972, 2021.

[21] Bai, J., Y. Wan, M. Li, G. Zhang, and X. He, "Reduction of random variables in EMC uncertainty simulation model," *Applied Computational Electromagnetics Society Journal (ACES)*, Vol. 37, No. 9, 941–947, 2022.

[22] Bai, J., L. Wang, D. Wang, A. P. Duffy, and G. Zhang, "Validity evaluation of the uncertain EMC simulation results," *IEEE Transactions on Electromagnetic Compatibility*, Vol. 59, No. 3, 797–804, 2017.

[23] Shahriari, B., K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas, "Taking the human out of the loop: A review of Bayesian optimization," *Proceedings of the IEEE*, Vol. 104, No. 1, 148–175, 2016.

[24] Bai, J., B. Hu, and A. Duffy, "Uncertainty analysis for EMC simulation based on bayesian optimization," *IEEE Transactions on Electromagnetic Compatibility*, Vol. 67, No. 2, 587–597, 2025.

[25] COMSOL Multiphysics, "Lightning-induced voltage of a wire in an airplane." [Online]. Available: https://cn.comsol.com/model/lightning-induced-voltage-of-a-wire-in-an-airplane-110121.

[26] Fuchs, F., E. U. Landers, R. Schmid, and J. Wiesinger, "Lightning current and magnetic field parameters caused by lightning strikes to tall structures relating to interference of electronic systems," *IEEE Transactions on Electromagnetic Compatibility*, Vol. 40, No. 4, 444–451, 1998.