PIER C

(Received 14 May 2025, Accepted 29 July 2025, Scheduled 8 August 2025)

Research on the Improvement Strategy of Initial Sampling Point Selection in Bayesian Optimization-Based Uncertainty Analysis Method

Jinjun Bai[1,*], Xiangrui Ji[1], Qing Liu[1], Yujia Song[2], and Zhongjiu Zheng[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: In recent years, uncertainty analysis methods have become a research hotspot in the field of Electromagnetic Compatibility (EMC), and non-intrusive uncertainty analysis methods are widely used in the field of EMC due to their advantages such as easy solver generalization and easy programming. The proposal of Bayesian optimization-based uncertainty analysis method further enhances the competitiveness of non-intrusive uncertainty analysis methods in solving complex EMC simulation problems. However, in traditional Bayesian optimization-based uncertainty analysis methods, Latin hypercube sampling strategy is used to construct the initial Gaussian process model, which lacks adaptive adjustment capability, and the quality of the initial Gaussian process model has a significant impact on the efficiency of subsequent calculations and the accuracy of the final results. This defect limits the computational efficiency and accuracy of Bayesian optimization methods in uncertainty analysis applications. In response to this issue, this paper proposes an active sampling strategy based on the Stochastic Reduced Order Model (SROM) method. This strategy improves the fitness function used by the SROM method in clustering to enhance the representativeness of the training set to the sampling space. By using this active sampling strategy instead of Latin hypercube sampling strategy, a higher quality initial Gaussian process model can be constructed, and the accuracy of Bayesian optimization method uncertainty analysis calculation is improved in the example, verifying the effectiveness of the proposed initial sampling point selection improvement strategy.

1. INTRODUCTION

In recent years, uncertainty analysis has become a hot research topic in the field of EMC. This method significantly improves the accuracy and reliability of EMC prediction by considering factors such as parameter variability, model error, and measurement uncertainty [1].

The uncertainty analysis methods can be divided into two categories based on whether the original solver needs to be modified: intrusive and non-intrusive. Typical non-intrusive uncertainty analysis methods include Perturbation Method (PM) [2], Stochastic Testing Method (STM) [3], and Stochastic Galerkin Method (SGM) [4]. In practical engineering applications, to achieve high reliability EMC simulation, it is often necessary to rely on commercial electromagnetic simulation software to simulate the actual electromagnetic environment. However, due to the lack of open-source deterministic simulation methods in commercial electromagnetic simulation software, intrusive uncertainty analysis methods lose their competitiveness. Therefore, non-intrusive uncertainty analysis methods are more practical for EMC field, as they only require a stable deterministic solver.

Typical non-intrusive uncertainty analysis methods include Monte Carlo Method (MCM) [5], Stochastic Collocation Method (SCM) [6], Stochastic Reduced Order Model (SROM) method [7], Surrogate model method [8], etc. MCM is based on the Wiener-Khinchin law of large numbers and has extremely high computational accuracy, but its computational efficiency is extremely low. It is commonly used as a standard result to compare with other methods [6–9]. SCM has the dual advantages of high computational accuracy and high computational efficiency. However, as the number of random variables increases, the computational efficiency of SCM will also exponentially decrease, indicating a curse of dimensionality [10]. SROM does not have a curse of dimensionality, but can only provide mean and variance predictions in uncertainty analysis results [7]. Recently, Surrogate model method has received widespread attention due to its controllable accuracy and resistance to curse of dimensionality. However, Latin hypercube sampling strategy is commonly used when training Surrogate models. This training set selection strategy is relatively mechanical and passive, and cannot actively adjust the sampling points according to the specific characteristics of different simulation situations, which seriously affects the practical application effect [8]. To solve the above problems, Bayesian optimization-based uncertainty analysis method is proposed. This method actively selects sampling points by maximizing the acquisition function, iteratively updates the Gaussian process (GP) model, and ultimately obtains uncertain

* Corresponding author: Jinjun Bai (baijinjun@dlmu.edu.cn).

analysis results. This method can achieve both computational efficiency and accuracy, and is adept at solving complex EMC simulation uncertainty analysis problems [11].

However in Bayesian optimization-based uncertainty analysis methods, the quality of the initial model has a significant impact on the efficiency of subsequent calculations and the accuracy of the final results. Traditional methods typically use Latin hypercube sampling to construct the initial GP model. Although this sampling method is more efficient than simple random sampling, it is essentially a passive space filling strategy that lacks adaptive adjustment capabilities, resulting in a lack of initiative in the sampling process, which limits the computational efficiency and accuracy of Bayesian optimization methods in uncertainty analysis applications. This paper proposes an active sampling strategy based on SROM method to address this issue. The fitness function used in SROM method during clustering is improved to enhance the representativeness of the training set to the sampling space, replacing Latin hypercube sampling strategy to improve the quality of the initial GP model, thereby improving the computational performance of Bayesian optimization-based uncertainty analysis method.

The structure of this paper is as follows. Section 2 introduces the basic principle of the traditional Bayesian optimization-based uncertainty analysis method. Section 3 provides a detailed introduction to the improved Bayesian optimization-based uncertainty analysis method. Section 4 verifies the effectiveness of the algorithm using crosstalk prediction of parallel cables example and electromagnetic interference of lightning electromagnetic pulse example. Section 5 summarizes the entire text.

2. BASIC PRINCIPLES OF THE TRADITIONAL BAYESIAN OPTIMIZATION-BASED UNCERTAINTY ANALYSIS METHOD

The Bayesian optimization-based uncertainty analysis method mainly consists of two parts, namely probabilistic surrogate model and acquisition function [12]. Below is a brief introduction to the traditional Bayesian optimization-based uncertainty analysis method.

Assuming that the uncertainty analysis problem can be represented by the following formula:

$$y_{\text{unc}}(\xi) = f_{\text{EMC}}\left[x_C(\xi_1), x_D(\xi_2)\right] \tag{1}$$

where $f_{\text{EMC}}()$ is a stable EMC simulation solver; $x_C(\xi_1)$ and $x_D(\xi_2)$ are random variables; and $\xi_1 \sim \text{pdf}(\xi_1)$, $\xi_2 \sim \text{pdf}(\xi_2)$. The uncertainty analysis result $y_{\text{unc}}(\xi)$ is also a random variable.

In Bayesian optimization, the most commonly used probabilistic surrogate model is GP model [13], where each finite subset follows a multivariate normal distribution consisting of a mean function m and a covariance function k. Assuming that the sampling space is $\mathbf{S}_{1,\xi} = [\mathbf{x}_{1,\xi}(1), \mathbf{x}_{1,\xi}(2), \cdots, \mathbf{x}_{1,\xi}(n_1)]$, n_2 small sampling points, $\mathbf{S}_{2,\xi} = [\mathbf{x}_{2,\xi}(1), \mathbf{x}_{2,\xi}(2), \cdots, \mathbf{x}_{2,\xi}(n_2)]$ can be obtained using Latin hypercube sampling in $\mathbf{S}_{1,\xi}$, which can be input into $f_{\text{EMC}}()$ to obtain $\mathbf{y}_{\text{unc},2} = \{f_{\text{EMC}}[\mathbf{x}_{2,\xi}(1)], f_{\text{EMC}}[\mathbf{x}_{2,\xi}(2)], \cdots, f_{\text{EMC}}[\mathbf{x}_{2,\xi}(n_2)]$. For input sample $\mathbf{S}_{2,\xi}$ and their corresponding output values $\mathbf{y}_{\text{unc},2}$, the predicted point $\{x(*), y(*)\}$ satisfies the following Gaussian distribution:

$$p\left[y(*)|x(*), \mathbf{S}_{2,\xi}, \mathbf{y}_{\text{unc},2}\right] = \text{N}\left[y(*)|\mu_*, \Sigma_*\right] \tag{2}$$

where μ_* is the mean of the predicted points, and Σ_* is the posterior covariance.

The initial GP model can be constructed using training set $\mathbf{R} = (\mathbf{S}_{2,\xi}, \mathbf{y}_{\text{unc},2})$:

$$y_{\text{GP},U} = \text{GP}_{\text{mod},U}(x_C, x_D) \sim f(x_C, x_D) \tag{3}$$

Input a large number of sampling points $\mathbf{S}_{1,\xi} = [\mathbf{x}_{1,\xi}(1), \mathbf{x}_{1,\xi}(2), \cdots, \mathbf{x}_{1,\xi}(n_1)]$ into $y_{\text{GP},U}$ to obtain $\mathbf{y}_{\text{unc},1}$, where $\mathbf{y}_{\text{unc},1}(i) = y_{\text{GP},U}[x_C(i), x_D(i)]$ is a random variable that follows a Gaussian distribution. The predicted points of the GP model also belong to the posterior probability density and follow a Gaussian distribution $p[y(*)|x(*), \mathbf{S}_{2,\xi}, \mathbf{y}_{\text{unc},2}] = \text{N}[y(*)|\mu_*, \Sigma_*]$.

The acquisition function is constructed based on the posterior probability distribution $\text{N}[y(*)|\mu_*, \Sigma_*]$ of the observed sample, and by maximizing the acquisition function, the next most "promising" evaluation point can be selected [14]. The purpose of using the acquisition function is to implement an exploration exploitation trade-off strategy to guide the iterative sampling process of the algorithm. Common acquisition functions include Probability of Improvement (PI), Expected Improvement (EI), Upper Confidence Bound (UCB), Expected Variance (EV), etc. [12]. Traditional Bayesian optimization-based uncertainty analysis methods combine EI and EV acquisition functions to fully utilize their advantages. The process framework of Bayesian optimization-based uncertainty analysis method is shown in Figure 1, and the specific steps are as follows:

(1) Select the next most "promising" evaluation point $\mathbf{x}_{2,\xi}(n_2+1)$ by maximizing the acquisition function, and input the selected evaluation point $\mathbf{x}_{2,\xi}(n_2+1)$ into $f_{\text{EMC}}()$ to obtain $f_{\text{EMC}}[\mathbf{x}_{2,\xi}(n_2+1)]$.

(2) Add the newly obtained input observation pair $\mathbf{R}(n_2+1) = [\mathbf{x}_{2,\xi}(n_2+1), \mathbf{y}_{\text{unc},2}(n_2+1)]$ to the historical observation set $\mathbf{R} = (\mathbf{S}_{2,\xi}, \mathbf{y}_{\text{unc},2})$. At this point, the training set becomes $\mathbf{R}_{\text{new}} = (\mathbf{S}^{\text{new}}_{2,\xi}, \mathbf{y}^{\text{new}}_{\text{unc},2})$, and the updated probability surrogate model becomes $y^{\text{new}}_{\text{GP},U} = \text{GP}^{\text{new}}_{\text{mod},U}(x_C, x_D)$. Continuously iterate until the number of sampling points in the training set reaches a certain set value Q, and establish the final GP model $y^{\text{Final}}_{\text{GP},U}$.

(3) Input $\mathbf{S}_{1,\xi}$ into $y^{\text{Final}}_{\text{GP},U}$, calculate the mean of the posterior result of the obtained Gaussian distribution to obtain $y_{S_i} = m[x_C(i), x_D(i)]$, and obtain the uncertainty analysis result $\text{pdf}\{f_{\text{EMC}}[x_C(\xi_1), x_D(\xi_2)]\}$ in the form of a probability density curve through statistical calculation.

In traditional Bayesian optimization algorithms, Latin hypercube sampling is used to obtain the training set of the initial GP model. The core idea is to divide the range of each input variable into a finite number of equally probable intervals and randomly select a sample point within each interval [15]. Although Latin hypercube sampling has higher sampling efficiency than simple random sampling, this sampling method also uses random sampling to sample each layer during the layered sampling

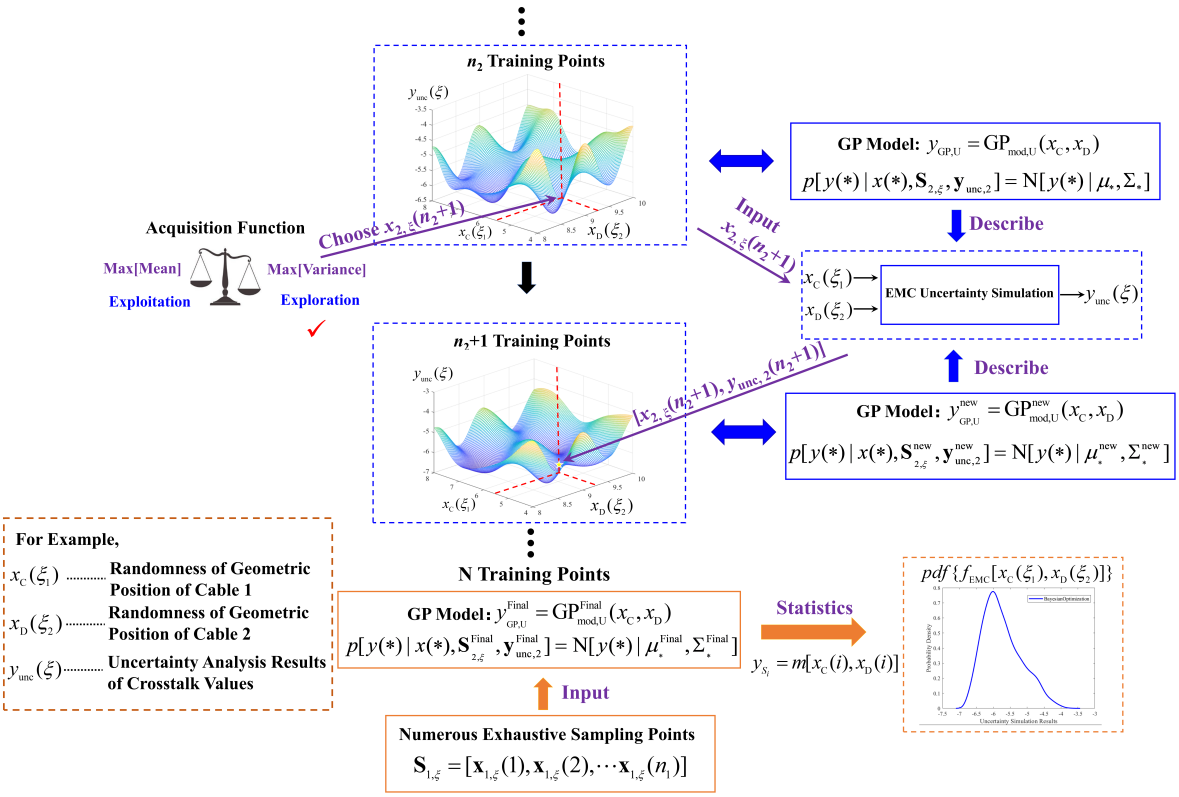

FIGURE 1. The process framework of Bayesian optimization-based uncertainty analysis method.

process, and the final sampling results are still affected by randomness. Essentially, Latin hypercube sampling is a mechanical and passive sampling method that lacks initiative. When the input dimension is high, or the nonlinearity of the example is large, this sampling method cannot accurately select the training set that contributes more to the model. When the single simulation time is long, it will also waste a lot of computing resources.

In Bayesian optimization-based uncertainty analysis methods, the quality of the initial sampling points directly determines the quality of the initial model, and the quality of the initial model has a significant impact on the efficiency of subsequent calculations and the accuracy of the final results. Therefore, in order to further improve the computational efficiency of Bayesian optimization-based uncertainty analysis methods in dealing with EMC simulation uncertainty analysis problems, a more proactive sampling strategy needs to be adopted to replace Latin hypercube sampling.

3. IMPROVED BAYESIAN OPTIMIZATION-BASED UNCERTAINTY ANALYSIS METHOD

SROM method is a non-intrusive uncertainty analysis method that aims to approximate the random variables used in the original MCM using a finite number of representative samples and their corresponding probability weights [7]. The flowchart of the SROM method is shown in Figure 2, and the specific implementation process is as follows:

(1) M large number of discrete sampling points $P = \{P_1, P_2, ..., P_M\}$ are randomly generated, and m samples are randomly selected from them, denoted as $p = \{p_1, p_2, ..., p_m\}$. The value of m is much smaller than M.

(2) The sampling space composed of $P = \{P_1, P_2, ..., P_M\}$ is divided into m regions, and the ith region is denoted as Γ_i, with the center of the region being p_i. The points in this area are those in the sampling space that are closer to the center p_i of the sample than to the centers of other samples. Calculate the Euclidean distances between sample center p_i and all other points in region Γ_i and sum them up. The result can be denoted as d_i. Repeat this operation to calculate the sum of Euclidean distances in other regions, and finally obtain $\{d_1, d_2, ..., d_m\}$. Sum $\{d_1, d_2, ..., d_m\}$ again to obtain D, which is calculated using the following formula:

$$D = \sum_{i=1}^{m} d_i \tag{4}$$

(3) By repeating the previous step, multiple sets of D values can be obtained. Genetic algorithm is used to select the small sample $p^r = \{p_1^r, p_2^r, ..., p_m^r\}$ that minimizes the D value as the optimal representative sampling point. The weight value of each representative sampling point in $p^r = \{p_1^r, p_2^r, ..., p_m^r\}$ is calculated. Taking the representative sampling point p_1^r in the first region Γ_1 as an example, assuming that there are n sampling points in this region and M sampling points in the sampling space, the weight value K_1 of p_1^r representing the sampling space $P = \{P_1, P_2, ..., P_M\}$ can be calculated by the

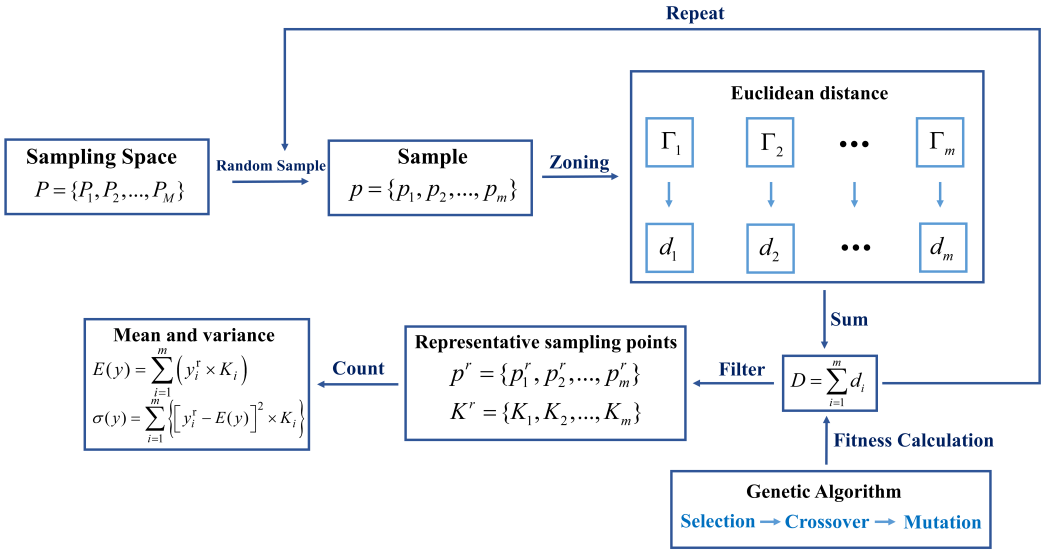

FIGURE 2. Schematic diagram of the SROM method.

following formula:

$$K_1=\frac{n}{M} \tag{5}$$

By inputting $p^r=\{p_1^r,p_2^r,...,p_m^r\}$ into the simulation model, m output results can be obtained, denoted as $y^{\text{SROM}}=\{y_1^r,y_2^r,...,y_m^r\}$. Through statistical calculations, the mean $E(y)$ and variance $\sigma(y)$ can be calculated as follows:

$$E(y)=\sum_{i=1}^{m}\left(y_i^r\times K_i\right) \tag{6}$$

$$\sigma(y)=\sum_{i=1}^{m}\left\{\left[y_i^r-E(y)\right]^2\times K_i\right\} \tag{7}$$

The fitness function mentioned above is improved in this section to ensure that the weight values of each representative sampling point are close, so that the selected representative sampling points can directly represent a large number of samples in the overall sampling space. This is the core idea of the SROM based active sampling strategy. The initial GP model constructed using this strategy can more accurately describe the overall characteristics of the system than the GP model constructed using Latin hypercube sampling. The flowchart of improved Bayesian optimization-based uncertainty analysis method is shown in Figure 3, and the specific implementation process is as follows:

(1) Assuming that there are H_ξ points in the sampling space, represented by $W=\{W_1,W_2,...,W_{H_\xi}\}$, where a single point takes the form of $W_i=\{W_{\xi_1}^i,W_{\xi_2}^i,...,W_{\xi N}^i\}$, K_ξ representative sampling points are randomly selected from them. In genetic algorithm, $\{Num_1,Num_2,...,Num_{K_\xi}\}$ represents the number of sampling points; the Euclidean distance between two points is $L(W_i,W_j)$; the specific calculation formula is $L(W_i,W_j)=\sqrt{\sum_{k=1}^{N}\left(W_{\xi_k}^i-W_{\xi_k}^j\right)^2}$.

(2) The Euclidean distance between each remaining point in the sampling space and the K_ξ representative sampling points is calculated, and its minimum value is defined as $L_{\min}(W_i)$. The minimum Euclidean distance is used as the standard to calculate the weight $\{\omega_1,\omega_2,...,\omega_{K_\xi}\}$ of each of the K_ξ representative sampling points in the sampling space. The fitness function of chromosome $\{Num_1,Num_2,...,Num_{K_\xi}\}$ can be calculated as:

$$Fin\,(Num,\omega)=\sum_{i=1}^{H_\xi}L_{\min}(W_i)+k_{weight}\times\sum_{j=1}^{K_\xi}\left|\omega_j-\frac{1}{K_\xi}\right| \tag{8}$$

where $\sum_{i=1}^{H_\xi}L_{\min}(W_i)$ represents the minimum Euclidean distance of all points, and the smaller the value of $\sum_{j=1}^{K_\xi}\left|\omega_j-\frac{1}{K_\xi}\right|$ is, the more evenly weighted the selected representative sampling points are. In the clustering process, in order to make the weights of each sampling point as close as possible and better cover the entire sampling space, a penalty term k_{weight} needs to be added to this fitness function. The method to determine the value of the penalty term k_{weight} is to run the genetic algorithm once, calculate the fitness function Fin_1 at this time, and multiply it by 10 to 50 times to obtain the value of the penalty term. When the number of sampling points is large (more than 100), the value of the penalty term is 10 times of Fin_1, and when the number of sampling points is small (less than 10), the value of the penalty term is 50 times of Fin_1. That is, $k_{weight}=\frac{Fin_1}{Q}$, where Q ranges from 0.02 to 0.1. When the number of sampling points is large, Q is taken as 0.1, and when the number of sampling points is small, Q is taken as 0.02, taking values proportionally.

(3) After the conventional selection, crossover, and mutation operations of genetic algorithm, the optimal representative sampling point $\{Num_1^{\text{Rep}},Num_2^{\text{Rep}},...,Num_{K_\xi}^{\text{Rep}}\}$ can be obtained, which corresponds to the final representative sampling point $W^{\text{Rep}}=\{W_1^{\text{Rep}},W_2^{\text{Rep}},...,W_{K_\xi}^{\text{Rep}}\}$ and its weight $\omega^{\text{best}}=\left\{\omega_1^{\text{best}},\omega_2^{\text{best}},...,\omega_{K_\xi}^{\text{best}}\right\}$. W^{Rep} is input into the EMC simulation model to obtain the training set $\mathbf{R}'$. Using this train-

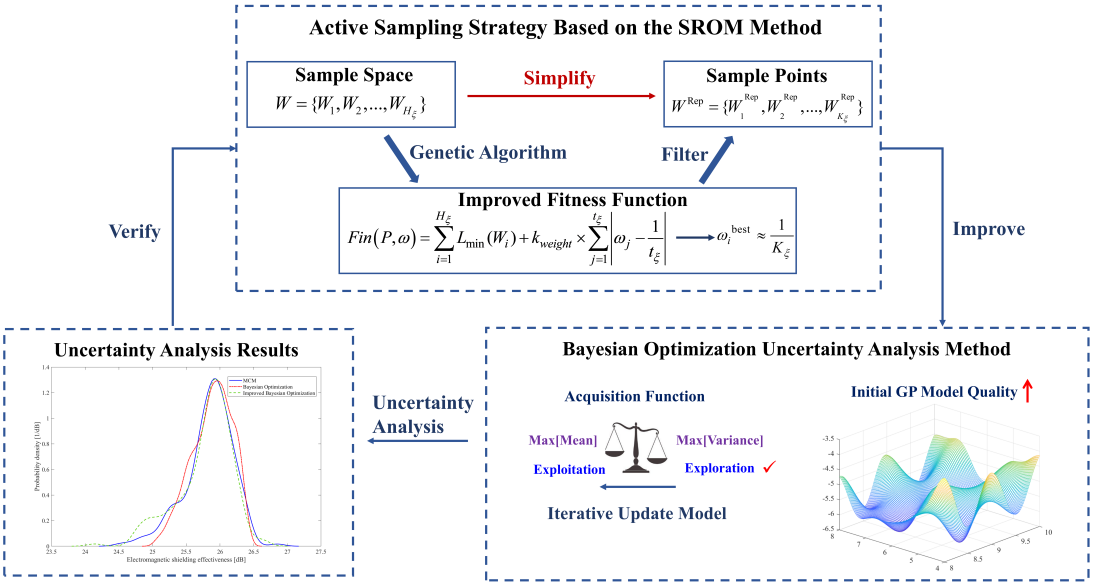

FIGURE 3. Schematic diagram of improved Bayesian optimization-based uncertainty analysis method.

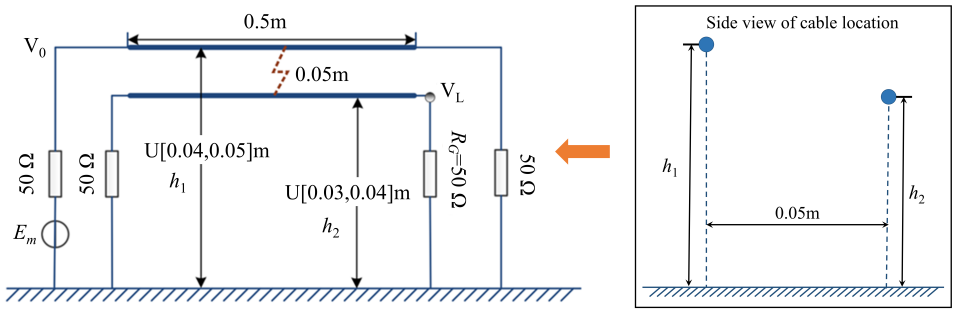

FIGURE 4. Parallel cable crosstalk prediction example in Reference [11].

ing set instead of the training set obtained by the Latin hypercube sampling method in traditional Bayesian optimization methods, the initial GP model is trained, and then the operations in traditional Bayesian optimization algorithm are continued to obtain uncertainty analysis results.

Through the analysis of the above steps, it can be concluded that firstly, in terms of spatial coverage, under the same sample size, the improved SROM method proposed in this paper outperforms Latin hypercube sampling in terms of spatial coverage. The core lies in improving SROM method by introducing a penalty mechanism to actively optimize the spatial distribution of sample points, thereby achieving more sufficient and uniform coverage of the space. The spatial coverage of Latin hypercube sampling depends on uniform spatial partitioning, and its coverage effect is limited by the use of random sampling.

Secondly, in terms of accuracy, the improved SROM method has a higher accuracy in describing the entire probability space than Latin hypercube sampling. Improving the SROM method by aggregating a large number of raw sampling points and clustering them based on probability density distribution enables the sample set to accurately represent the statistical properties of the original random variables. The spatial coverage of Latin hypercube sampling cannot be dynamically adjusted with the distribution pattern. For example, when dealing with Gaussian distributions, and the data are dense in certain areas and sparse in others, the Latin hypercube sampling method cannot automatically adjust the density of sample points to match this non-uniformity due to its fixed, uniform space based partitioning method. Improving SROM can adaptively focus by placing more sample points in high probability areas and fewer sample points in low probability areas.

Thirdly, the applicability of improving SROM method is also stronger. The sample generation process of the improved SROM method is strongly correlated with distribution types, making it universal for various complex nonuniform distributions. However, the uniform spatial partitioning mechanism of Latin hypercube sampling has insufficient applicability in dealing with nonuniform distributions.

4. ALGORITHM VERIFICATION

The crosstalk prediction of parallel cables example shown in Figure 4 is first adopted in this section to verify the performance of the improved Bayesian optimization-based uncertainty anal-

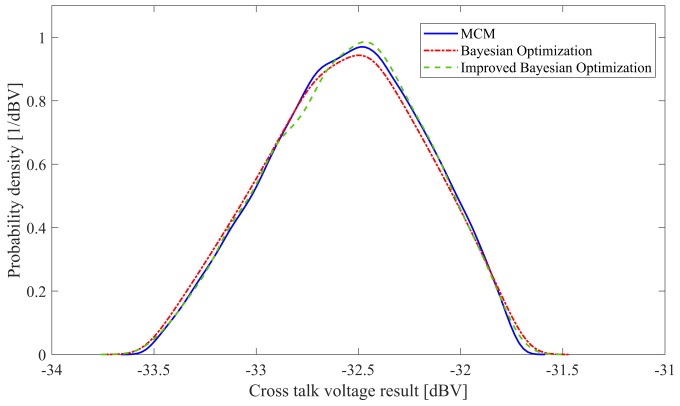

FIGURE 5. Probability density curve results of V_{dB} at 75 MHz.

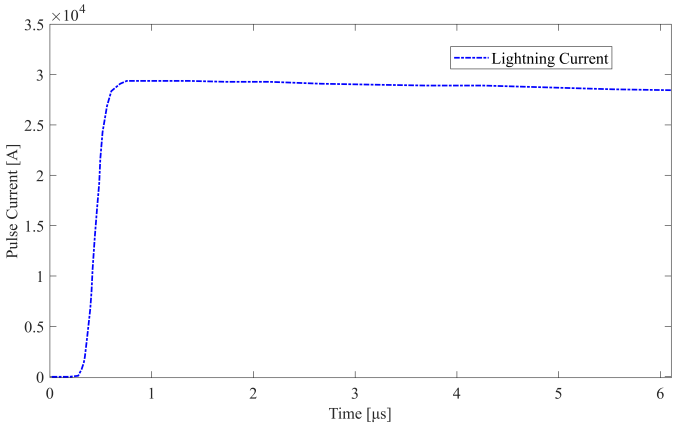

FIGURE 6. Lightning current transient current curve within 0–6 μs.

ysis method. This example is a standard example from [11], where the parallel cable height is assumed to be an uncertain input parameter, described by the following random variable model:

$$\begin{cases} h_1(\xi_1) = 0.045 + 0.005 \times \xi_1 \,[\text{m}] \\ h_2(\xi_2) = 0.035 + 0.005 \times \xi_2 \,[\text{m}] \end{cases} \tag{9}$$

where h_1 and h_2 are cable heights, and ξ_1 and ξ_2 are uniformly distributed random variables within the interval $[-1, 1]$. The horizontal distance between two cables is 0.05 m, and the output result is the far end crosstalk voltage V_{dB} in decibels. The calculation formula for V_{dB} is as follows. V is the near end voltage of interference source, and V_L is the voltage at the far end of the disturbed source

$$V_{dB} = 20 \log_{10} \frac{|V_L|}{|V_0|} \tag{10}$$

Figure 5 shows the probability density curve results obtained by each method at 75 MHz in the parallel cable crosstalk prediction example. The MCM conducts 10000 deterministic EMC simulations, and its uncertainty analysis results are used as standard data to quantitatively evaluate the effectiveness of each uncertainty analysis result using Mean Equivalent Area Method (MEAM) [16]. Since this study is based on the model in [11] and makes improvements, both the improved Bayesian optimization method and traditional Bayesian optimization method undergo 16 deterministic simulations. According to [16], due to the simplicity of this example, the MEAM result of 0.9881 from the traditional Bayesian optimization method is already an "excellent" result, leaving little room for improvement. Therefore, the accuracy of the results obtained by the improved Bayesian optimization method shows minimal improvement compared to the original method, with the equivalent area result increasing from 0.9881 to 0.9893.

Then, the performance of the improved Bayesian optimization-based uncertainty analysis method is verified using the electromagnetic interference of lightning electromagnetic pulse example. This example is also a standard example in [11], and its deterministic electromagnetic

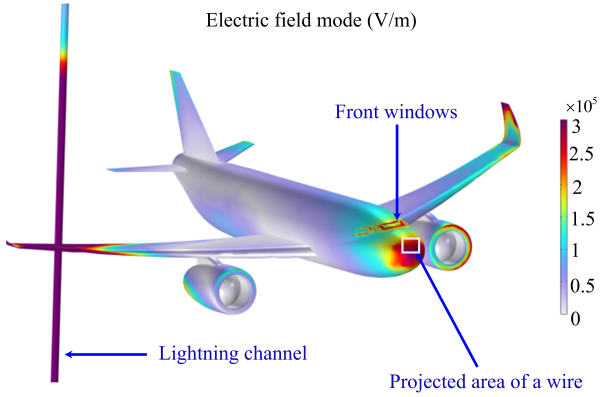

FIGURE 7. Transient simulation results of aircraft surface electric field at 2.5 μs.

simulation model is derived from the official example of COMSOL [17].

Because lightning only causes significant electromagnetic interference to aircraft in a short period of time after its occurrence, only the lightning current within 0–6 μs is modeled. The transient current curve is shown in Figure 6, and the transient simulation results of the aircraft surface electric field at 2.5 μs are shown in Figure 7. The initial deterministic simulation results are shown in Figure 8, and the simulation output is the induced voltage values of sensitive components inside the aircraft (replaced with wires in official cases) within 0–6 s under the conditions of only the front window being unshielded and completely unshielded. V_1 is the absolute value of the maximum induced voltage of sensitive components inside the aircraft when only the front window is unshielded, and V_2 is the absolute value of the maximum induced voltage of sensitive components inside the aircraft when it is completely unshielded. Therefore, the electromagnetic shielding effectiveness SE of the aircraft can be calculated as shown in Equation (11).

$$SE = 20 \lg \frac{|V_2|}{|V_1|} \tag{11}$$

The x and y coordinate values of lightning location information are assumed to be uncertain input parameters, described by

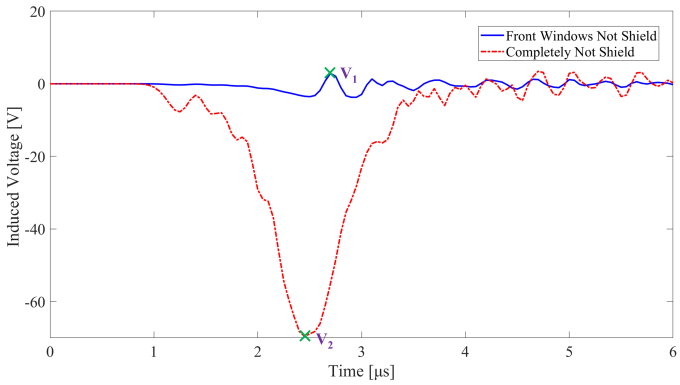

FIGURE 8. Initial deterministic simulation results in official example.

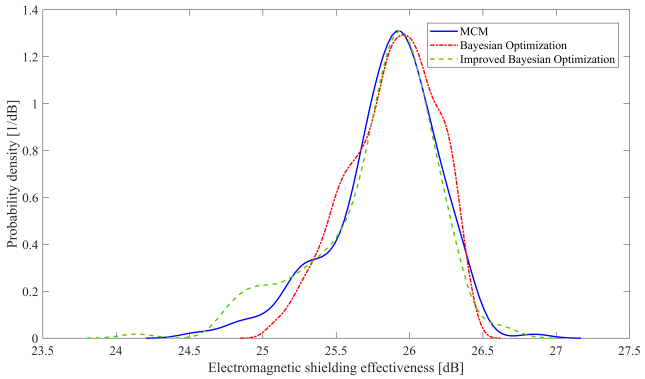

FIGURE 9. Uncertainty simulation results of SE_{EMC}.

TABLE 1. Computational time of two Bayesian optimization methods in case studies.

Methodology/Examples	Sampling preprocessing time	Single deterministic simulation time	Bayesian optimization-based uncertainty analysis time	Total time
Traditional Bayesian optimization method in example 1	0.13 s	1.32 s ×16 times	15.41 s	0.61 min
Improved Bayesian optimization method in example 1	3.82 min	1.31 s ×16 times	15.17 s	4.42 min
Traditional Bayesian optimization method in example 2	0.14 s	6.67 min ×15 times	14.82 s	100.30 min
Improved Bayesian optimization method in example 2	3.53 min	6.65 min ×15 times	14.76 s	103.53 min

the following random variable model:

$$\begin{cases} x(\xi_1) = -30 - 2 \times \xi_1 \,[\mathrm{m}] \\ y(\xi_2) = 8 + 2 \times \xi_2 \,[\mathrm{m}] \end{cases} \tag{12}$$

where ξ_1 and ξ_2 are uniformly distributed random variables in the interval $[-1, 1]$, and the value to be solved is the aircraft electromagnetic shielding effectiveness value SE_{EMC}.

Figure 9 shows the probability density curve results calculated by various methods in the electromagnetic interference of lightning electromagnetic pulse example. The MCM conducts 2000 deterministic EMC simulations, and the uncertainty analysis results are also used as standard data. Since this study is an improvement on the model based on [11], both the improved Bayesian optimization method and traditional Bayesian optimization method were subjected to 15 deterministic simulations. At this point, the MEAM result of the Bayesian optimization method has increased from 0.9288 (very good) to 0.9528 (excellent), which is a significant improvement and proves the effectiveness of the improved Bayesian optimization method proposed in this study.

The computational times of the two Bayesian optimization methods in the case study are shown in Table 1.

Based on the above calculation results and the data in Table 1, it can be concluded that although the total computation time required by the improved Bayesian optimization method is indeed slightly increased compared to the traditional Bayesian optimization method in both examples, the required sampling preprocessing time remains fixed. As the single deterministic simulation time and the number of single deterministic simulations increase, the impact of the additional sampling preprocessing time on the computational burden becomes smaller and smaller, and the computational efficiency becomes higher and higher. Therefore, it will not affect the overall conclusion of the study. For example, in example 2 of the improved Bayesian optimization method, the proportion of sampling preprocessing time to the total time is 3.53 min/103.53 min = 3.41%, which is very small.

The improved Bayesian optimization-based uncertainty analysis method achieves an improvement in the accuracy of the calculation results compared to the traditional method in both examples, proving the effectiveness of the improved

strategy for selecting initial sampling points proposed in this paper.

5. CONCLUSION

The non-intrusive uncertainty analysis method has the advantage of obtaining uncertainty analysis results without modifying the original solver and is currently widely used. The Bayesian optimization-based uncertainty analysis method has strong nonlinear processing ability and high computational efficiency and accuracy. The active sampling strategy based on SROM method is proposed in this paper, replacing the traditional Latin hypercube sampling strategy in order to solve the problem of low training efficiency of the initial GP model and its impact on algorithm computational performance. This paper improves the fitness function used by the SROM method in clustering to increase the representativeness of the training set, so that a higher quality initial GP model can be trained, and the computational accuracy of Bayesian optimization method is further improved. In the crosstalk prediction of parallel cables example and the electromagnetic interference of lightning electromagnetic pulse example, the accuracy of Bayesian optimization method uncertainty analysis calculation is improved. Especially in the electromagnetic interference of lightning electromagnetic pulse example, the improved Bayesian optimization-based uncertainty analysis method increased the equivalent area quantification result of traditional Bayesian optimization method uncertainty analysis from 0.9288 to 0.9528 in the same 15 deterministic simulations. This proves that the initial sampling point selection improvement strategy proposed in this paper has a significant improvement on the computational performance of Bayesian optimization-based uncertainty analysis method.

ACKNOWLEDGEMENT

This paper is supported by "the Open Fund of National Center for International Research of Subsea Engineering Technology and Equipment" (Project No. HG20240201).

REFERENCES

[1] Manfredi, P., D. V. Ginste, I. S. Stievano, D. D. Zutter, and F. G. Canavero, "Stochastic transmission line analysis via polynomial chaos methods: An overview," *IEEE Electromagnetic Compatibility Magazine*, Vol. 6, No. 3, 77–84, Nov. 2017.

[2] 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, Apr. 2020.

[3] Zhang, Z., T. A. El-Moselhy, I. M. Elfadel, and L. Daniel, "Stochastic testing method for transistor-level uncertainty quantification based on generalized polynomial chaos," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, Vol. 32, No. 10, 1533–1545, Oct. 2013.

[4] 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, Nov. 2013.

[5] Pignari, S. A., G. Spadacini, and F. Grassi, "Modeling field-to-wire coupling in random bundles of wires," *IEEE Electromagnetic Compatibility Magazine*, Vol. 6, No. 3, 85–90, Nov. 2017.

[6] Wang, T., Y. Gao, L. Gao, C.-Y. Liu, J. Wang, and Z. An, "Statistical analysis of crosstalk for automotive wiring harness via polynomial chaos method," *Journal of the Balkan Tribological Association*, Vol. 22, No. 2, 1503–1517, 2016.

[7] 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.

[8] 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.

[9] Xie, H., J. F. Dawson, J. Yan, A. C. Marvin, and M. P. Robinson, "Numerical and analytical analysis of stochastic electromagnetic fields coupling to a printed circuit board trace," *IEEE Transactions on Electromagnetic Compatibility*, Vol. 62, No. 4, 1128–1135, Aug. 2020.

[10] Teckentrup, A. L., P. Jantsch, C. G. Webster, and M. Gunzburger, "A multilevel stochastic collocation method for partial differential equations with random input data," *SIAM/ASA Journal on Uncertainty Quantification*, Vol. 3, No. 1, 1046–1074, 2015.

[11] 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, Apr. 2025.

[12] Greenhill, S., S. Rana, S. Gupta, P. Vellanki, and S. Venkatesh, "Bayesian optimization for adaptive experimental design: A review," *IEEE Access*, Vol. 8, 13 937–13 948, Jan. 2020.

[13] Lindauer, M., K. Eggensperger, M. Feurer, A. Biedenkapp, D. Deng, C. Benjamins, T. Ruhkopf, R. Sass, and F. Hutter, "SMAC3: A versatile Bayesian optimization package for hyperparameter optimization," *Journal of Machine Learning Research*, Vol. 23, No. 54, 1–9, 2022.

[14] Candelieri, A., A. Ponti, and F. Archetti, "Fair and green hyperparameter optimization via multi-objective and multiple information source Bayesian optimization," *Machine Learning*, Vol. 113, No. 5, 2701–2731, 2024.

[15] Olsson, A., G. Sandberg, and O. Dahlblom, "On Latin hypercube sampling for structural reliability analysis," *Structural Safety*, Vol. 25, No. 1, 47–68, Jan. 2003.

[16] Bai, J., J. Sun, and N. Wang, "Convergence determination of EMC uncertainty simulation based on the improved mean equivalent area method," *Applied Computational Electromagnetics Society Journal (ACES)*, Vol. 36, No. 11, 1446–1452, Dec. 2021.

[17] 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 (in Chinese).