(Received 27 March 2026, Accepted 13 May 2026, Scheduled 3 June 2026)

# A Preprocessing Dimensionality Reduction Framework for Improved Polynomial Chaos Expansion in EMC Uncertainty Quantification

Yitong Lu, Zhengyu Xue*, and Shenghang Huo

*College of Marine Electrical Engineering, Dalian Maritime University, Dalian 116026, China*

**ABSTRACT:** Polynomial Chaos Expansion (PCE) is widely utilized in uncertainty quantification (UQ) for electromagnetic compatibility (EMC) due to its robust global predictive capabilities. However, its computational overhead increases exponentially with stochastic dimensionality, leading to the notorious *curse of dimensionality*. To address this bottleneck, this paper proposes a generalized preprocessing dimensionality-reduction framework designed to enhance the performance of PCE. By decoupling dimensional screening from predictive modeling, the proposed framework first employs low-cost estimators to identify significant random variables. Subsequently, an improved PCE model is constructed within the reduced feature space. Given the prohibitively high computational cost of acquiring EMC simulation samples, this study instantiates a screening module within the framework that integrates Least Squares Support Vector Regression (LSSVR) with Sobol indices. Finally, the proposed framework-based method is applied to a cable crosstalk case study to validate its effectiveness and engineering applicability.

## 1. INTRODUCTION

Uncontrolled variations in system parameters — stemming from manufacturing tolerances, material aging, and operating conditions — can severely degrade EMC performance [1, 2]. Consequently, UQ has garnered increasing attention within the EMC community [3]. Safety design margins are typically established using extensive Monte Carlo methods (MCM) [4]; however, this approach significantly impedes the efficiency of the design phase.

PCE [5] has emerged as a prominent surrogate tool for UQ, owing to its computational efficiency and global predictive fidelity. The cornerstone of PCE-based UQ lies in accurately determining expansion coefficients. Methods for coefficient estimation are generally categorized into two paradigms: intrusive and non-intrusive approaches [6]. The stochastic Galerkin method [7] is a representative intrusive approach; it involves reformulating the stochastic governing equations to derive a set of augmented deterministic coupled equations for the unknown coefficients. Despite their theoretical high-fidelity advantages, intrusive methods suffer from high implementation complexity, which precludes their widespread adoption in large-scale engineering problems. In contrast, non-intrusive methods treat the original system as a black-box paradigm, utilizing input-output data pairs obtained through sampling to estimate coefficients via projection or regression techniques [8, 9], thereby preserving the integrity of the original deterministic solver. To address high-dimensional challenges in complex systems, non-intrusive methods typically leverage sparsity-promoting schemes [10, 11]. Bayesian sparse PCE methods have effectively addressed high-dimensional problems by incorporating joint priors or adaptive mechanisms [12, 13]. Similarly, adaptive least angle regression (LAR) significantly enhances computational efficiency in high-dimensional spaces by dynamically updating the basis function library during the iterative process [14]. Nonetheless, even with sparse schemes, the number of quadrature nodes scales superlinearly with the cardinality of the basis set. Consequently, this inevitably creates an accuracy bottleneck in high-dimensional scenarios.

In recent years, the rapid advancement of machine learning (ML) has paved the way for its integration into the realm of UQ [15, 16]. Learning-based paradigms, such as Support Vector Machine (SVM) [17] and Gaussian Process Regression (GPR) [18, 19], have been successfully deployed to address UQ challenges. Gaussian process regression (GP) demonstrates exceptional scalability when handling large-scale datasets [20]. As proposed by Trinchero et al. [21], LSSVR exhibits remarkable robustness against high input dimensionality, contingent upon the availability of a sufficient training dataset. However, in practical engineering scenarios, the prohibitive computational overhead of high-fidelity EMC simulations often restricts the sample size, thereby imposing a bottleneck on the predictive accuracy of LSSVR models.

As a pivotal application of UQ, sensitivity analysis quantifies the individual contributions of stochastic input variables to the system output, facilitating the identification of high-impact parameters. Sobol-index-based sensitivity analysis has emerged as one of the most prevalent and widely adopted frameworks [22, 23]. Its theoretical underpinnings rely on the functional decomposition of the model into a summation of or-

* Corresponding author: Zhengyu Xue (xuezy@dlmu.edu.cn).

thogonal terms. Models, such as Bayesian Adaptive Splines (BMARS) [24] and Bayesian Regression Trees (BART) [25], are capable of efficiently computing the Sobol index using closed-form solutions. However, calculating the Sobol index also requires a large sample size.

To further enhance predictive performance under extremely small sample sizes, this paper proposes a generalized preprocessing dimensionality-reduction framework. By decoupling dimensionality screening from predictive modeling, the framework employs low-cost estimators to identify critical random variables. This facilitates the construction of an improved PCE model within a reduced subspace. Specifically, in the context of EMC engineering applications involving limited data, this study instantiates the framework by integrating LSSVR with Sobol indices, leveraging their nonlinear robustness under severely restricted samples to achieve precise dimensionality reduction. Finally, a case study involving parallel cable crosstalk validates the significant advantages of this framework in improving predictive accuracy and overcoming dimensionality bottlenecks.

The remainder of this paper is organized as follows. Section 2 delineates the fundamental principles of PCE-based UQ. Section 3 describes the proposed general preprocessing dimensionality reduction framework. Section 4 provides a detailed description of the specific implementation process of this framework under LSSVR-Sobol. In Section 5, the proposed method is implemented in a cable crosstalk case study to validate its practical engineering utility. Section 6 summarizes the paper.

## 2. PCE-BASED UNCERTAINTY QUANTIFICATION METHOD

Suppose that a certain EMC stochastic process can be represented as:

$$Y = f(\boldsymbol{\xi}) \tag{1}$$

Here, $\boldsymbol{\xi} = [\xi_1, \xi_2, \ldots, \xi_d]^T$ denotes a vector of $d$-dimensional independent stochastic input variables.

The PCE of this process is formulated as:

$$Y = \sum_{i=0}^{P_{full}-1} \alpha_i \varphi_i(\boldsymbol{\xi}) \tag{2}$$

where $\varphi_i(\boldsymbol{\xi})$ is an orthogonal polynomial that depends on the distribution of the random variable, and $\alpha_i$ is the deterministic expansion coefficient to be determined. The number of terms in the expansion is:

$$P_{full} = \frac{(d+q)!}{d!q!} \tag{3}$$

where $q$ is the order of the expansion.

The orthonormal basis functions $\varphi_i(\boldsymbol{\xi})$ satisfy the orthogonality condition:

$$\langle \varphi_i, \varphi_j \rangle = \langle \varphi_i^2 \rangle \, \delta_{ij} \tag{4}$$

where $\delta_{ij}$ is the Kronecker function, defined by:

$$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases} \tag{5}$$

The unknown deterministic coefficients $\alpha_i$ are determined by projecting the model response onto the orthogonal polynomial bases:

$$\alpha_i = \langle Y(\boldsymbol{\xi}), \varphi_i(\boldsymbol{\xi}) \rangle \, / \, \langle \varphi_i(\boldsymbol{\xi})^2 \rangle \tag{6}$$

The mean and variance of the model's response are quantified as follows:

$$E_{\mathrm{PCE}} = E(Y(\boldsymbol{\xi})) = \alpha_0 \tag{7}$$

$$\mathrm{Var}_{\mathrm{PCE}} = \mathrm{Var}(Y(\boldsymbol{\xi})) = \sum_{i=1}^{P_{full}-1} \left[ \alpha_i^2 \langle \varphi_i^2 \rangle \right] \tag{8}$$

As the input dimension $d$ increases, the number of basis functions grows exponentially, manifesting the classic curse of dimensionality. Although advanced methods, such as sparse PCE, partially alleviate this issue, guaranteeing the accuracy of a surrogate model constructed directly in the full-dimensional space remains challenging under severely restricted sample conditions in engineering applications.

## 3. PREPROCESSING DIMENSION REDUCTION FRAMEWORK

To overcome the modeling bottleneck under the conditions of high dimensionality and small sample sizes, this paper proposes a generalized preprocessing dimensionality-reduction framework. As illustrated in Figure 1, the core concept of this framework is to decouple dimensionality screening of the stochastic space from model construction. Before proceeding to high-precision PCE prediction, low-cost estimators are employed to identify and extract key features.

Let $S$ define a generalized dimensionality screening estimator. By calculating the sensitivity indices of random variables and comparing them against a preset threshold $\epsilon$, the index set $\mathcal{I}$ for high-sensitivity variables is determined as follows:

$$\mathcal{I}(\epsilon) = \mathcal{S}(\xi, Y, \epsilon) = \{ i \in \{1, ..., d\} \mid S_i^T > \epsilon \} \tag{9}$$

The effective dimension is denoted by $k = |\mathcal{I}|$, where $k \ll d$. The original random space is mapped to the reduced-dimension feature space via the projection operator $P_{\mathcal{I}}$:

$$\xi_{\mathcal{I}} = P_{\mathcal{I}}(\xi) = \{ \xi_i \mid i \in \mathcal{I} \} \tag{10}$$

For the non-significant variable $\xi_j (j \notin \mathcal{I})$, it is held constant at its mean.

The improved PCE model is constructed in the reduced-dimension feature space $\xi_{\mathcal{I}}$, and its expression is:

$$Y(\xi_{\mathcal{I}}) = \sum_{i=0}^{P_{red}-1} \alpha_i \varphi_i(\xi_{\mathcal{I}}) \tag{11}$$

Here, the number of basis functions $P_{red}$ after dimensionality reduction is reduced to:

$$P_{red} = \frac{(k+q)!}{k!q!} \tag{12}$$

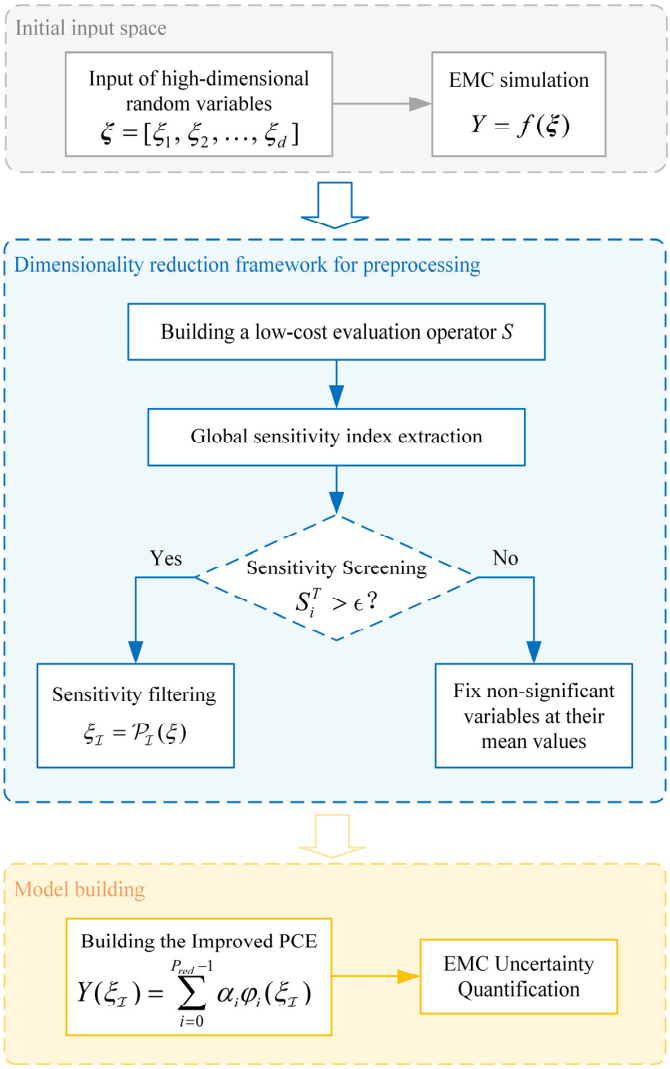


**FIGURE 1**. An improved PCE based on the preprocessing dimension reduction framework.

By reducing the number of dimensions from $d$ to $k$, the computational complexity has been drastically reduced from exponential in $d$ to exponential in $k$. It is worth noting that if a physical system exhibits strong coupling among all variables (i.e., all inputs are highly sensitive), this framework naturally degenerates into the full-dimensional space, which represents a common physical limit for all methods based on front-end filtering.

## 4. SENSITIVITY EVALUATION OPERATOR BASED ON LSSVR-SOBOL

Under the aforementioned generalized framework, given the practical challenges of prohibitively high EMC simulation costs and severely restricted sample sizes, this study selects the integration of LSSVR and Sobol indices as the specific implementation of the screening estimator $S$ as detailed in Algorithm 1.

Leveraging its robust generalization capability under extremely small sample sizes, LSSVR emerges as an ideal choice for constructing low-cost evaluation estimators. Its optimization objective function is defined as follows:

$$\min_{w,b,e} J(w,e) \tag{13}$$

where $J(w,e) = \frac{1}{2}w^T w + \frac{\gamma}{2}\sum_{i=1}^{N} e_i^2$, subject to the equality constraints:

$$y_i = w^T \phi(x_i) + b + e_i, \quad i = 1, \ldots, N \tag{14}$$

where $e_i$ represents the error variable, $\gamma$ the regularization parameter, $w$ the weight vector, and $b$ the bias term.

By introducing the Lagrange multipliers $\alpha_i \in \mathbb{R} (i = 1, \ldots, N)$, the corresponding dual problem is obtained as:

$$\begin{bmatrix} 0 & \mathbf{1}_N^T \\ \mathbf{1}_N & \mathbf{K} + \gamma^{-1}\mathbf{I}_N \end{bmatrix} \begin{bmatrix} b \\ \boldsymbol{\alpha} \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix} \tag{15}$$

where $\mathbf{1}_N = [1, \cdots, 1]^T \in \mathbb{R}^N$, $\mathbf{K}_{ij} = \kappa(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^T \phi(\mathbf{x}_j), i, j = 1, \ldots, N$, and $\kappa(\mathbf{x}_i, \mathbf{x}_j)$ is kernel function. $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_N]^T$ represents the Lagrange multiplier vector.

The LSSVR model serves as:

$$M_{\text{LSSVR}}(\mathbf{x}) = \sum_{i=1}^{N} \alpha_i \kappa(\mathbf{x}_i, \mathbf{x}) + b \tag{16}$$

The Sobol method, anchored in the Analysis of Variance decomposition, partitions the total output variance into first-order effects and higher-order interactions. The original model can be expressed as the sum of incremental terms, as follows:

$$\begin{aligned} Y(\boldsymbol{\xi}) &= y_0 + \sum_{i=1}^{d} y_i(\xi_i) + \sum_{1 \le i < j \le d} y_{ij}(\xi_i, \xi_j) + \cdots \\ &+ y_{1,2,\ldots,d}(\xi_1, \xi_2, \ldots, \xi_d) \end{aligned} \tag{17}$$

The variance of the model response can be expanded as

$$\text{Var}[Y] = \sum_{i=1}^{d} D_i + \sum_{1 \le i < j \le d} D_{ij} + \cdots + D_{1,2,\ldots,d} \tag{18}$$

The Sobol index is defined as

$$S_{i_1,\ldots,i_s} = \frac{D_{i_1,\ldots,i_s}}{D}, \ 1 \le i_1 < \ldots < i_s \le d, \ s = 1, \ldots, d \tag{19}$$

where $D$ represents the total variance.

The total Sobol index $S_i^T$ of the input variable is the sum of all Sobol indices associated with that variable:

$$S_i^T = \sum_{\{i_1,\ldots,i_s\} \supset i} S_{i_1,\ldots,i_s} \tag{20}$$

Exhaustive sampling predictions are executed on the constructed LSSVR model to efficiently estimate the Sobol indices.

**Algorithm 1**

**Input:** PCE order $p$, random variable dimension $d$, sample size $N$, sensitivity threshold ε

**Output:** Improved PCE model

**1. Initial Sampling:**

Generate $N$ samples using Latin Hypercube Sampling [26]: $\boldsymbol{S}=\left[\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_N\right]$

Perform EMC simulations to obtain the response vector: $\boldsymbol{Y}=\left[y_1,y_2,\cdots,y_N\right]^{\mathrm{T}}$

**2. LSSVR Modeling:**

Solve the optimization problem: $\min_{w,b,e} J(w,e)$

Construct the surrogate model: $\mathrm{M}_{\mathrm{LSSVR}}(\boldsymbol{x})=\sum_{i=1}^{N}\alpha_i\kappa(\boldsymbol{x}_i,\boldsymbol{x})+b$

**3. Sensitivity Evaluation Operator Based on LSSVR-Sobol:**

Calculate the mean ($\hat{y}_0$), total variance ($\hat{D}$), and partial variances ($\hat{D}_i$) using the LSSVR model through Sobol indices:

$$\begin{cases}\hat{y}_0=\dfrac{1}{N_s}\sum_{g=1}^{N_s}\mathrm{M}_{\mathrm{LSSVR}}(\boldsymbol{x}^{(g)})\\ \hat{D}=\dfrac{1}{N_s}\sum_{g=1}^{N_s}\mathrm{M}_{\mathrm{LSSVR}}{}^2(\boldsymbol{x}^{(g)})-\hat{y}_0^2\\ \hat{D}_i=\dfrac{1}{N_s}\sum_{g=1}^{N_s}\mathrm{M}_{\mathrm{LSSVR}}(x_i^{(g)},\boldsymbol{x}_{\sim i}^{(g)})\mathrm{M}_{\mathrm{LSSVR}}(x_i^{(g)},\tilde{\boldsymbol{x}}_{\sim i}^{(g)})-\hat{y}_0^2\end{cases}$$

**4. Improved PCE Model Construction:**

Utilizing sensitivity operators to identify highly sensitive variables:

$\mathcal{I}(\epsilon)=\mathcal{S}(\xi,Y,\epsilon)=\{i\in\{1,\ldots,d\}\mid S_i^T>\epsilon\}$

Mapping the original random space to the reduced-dimension feature space:

$\xi_{\mathcal{I}}=\mathcal{P}_{\mathcal{I}}(\xi)=\{\xi_i\mid i\in\mathcal{I}\}$

**5. Output:**

$Y(\xi_{\mathcal{I}})=\sum_{i=0}^{P_{\mathrm{red}}-1}\alpha_i\varphi_i(\xi_{\mathcal{I}})$

Let $N_s$ denote the number of sampling points, with $x$ and $\tilde{x}$ representing two independent sample matrices. The Sobol indices based on the LSSVR model are then derived as follows:

$$\begin{cases}\hat{y}_0=\frac{1}{N_s}\sum_{g=1}^{N_s}M_{\mathrm{LSSVR}}(\mathbf{x}^{(g)})\\ \hat{D}=\frac{1}{N_s}\sum_{g=1}^{N_s}M_{\mathrm{LSSVR}}^2(\mathbf{x}^{(g)})-\hat{y}_0^2\\ \hat{D}_i=\frac{1}{N_s}\sum_{g=1}^{N_s}M_{\mathrm{LSSVR}}(x_i^{(g)},\mathbf{x}_{\sim i}^{(g)})M_{\mathrm{LSSVR}}(x_i^{(g)},\tilde{\mathbf{x}}_{\sim i}^{(g)})-\hat{y}_0^2\end{cases}\tag{21}$$

The total sensitivity index $S_i^T$ obtained for each variable will serve as the basis for dimensionality reduction.

## 5. APPLICATION OF THE PROPOSED METHOD IN CABLE CROSSTALK CASES

To validate the proposed algorithm, a parallel cable crosstalk model is established, as illustrated in Figure 2 [27, 28]. Both the signal generator line and the victim line have a length of 0.5 m, with a near-end termination impedance of 50 Ω. The stochastic modeling of the geometric and electrical parameters — including height above ground, cable radius, insulation thickness, far-end impedance, and lateral separation — is summarized in Table 1. The cable conductors are made of copper, while the insulation layer consists of polyvinyl chloride, with a relative permittivity of $\varepsilon_r = 3.19$.

**TABLE 1**. Uncertainty modeling for input variables.

| Uniform random variables | | U [min, max] | Unit |
|---|---|---|---|
| Height of signal line | $h_1$ | U [2.3 2.7] | cm |
| Height of victim line | $h_2$ | U [2.2 2.6] | cm |
| Radius of signal line | $r_1$ | U [0.36 0.44] | mm |
| Radius of victim line | $r_2$ | U [0.36 0.44] | mm |
| Insulation thickness of signal line | $t_1$ | U [0.27 0.33] | mm |
| Insulation thickness of victim line | $t_2$ | U [0.27 0.33] | mm |
| Far-end impedance of signal line | $R_1$ | U [45, 55] | Ω |
| Far-end impedance of victim line | $R_2$ | U [45, 55] | Ω |
| Lateral spacing between two lines | $d$ | U [4.5 5.5] | mm |

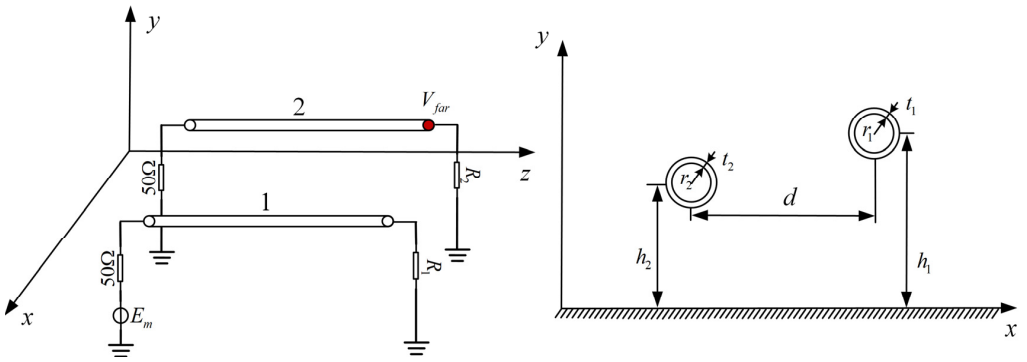


**FIGURE 2**. Parallel cable crosstalk model.

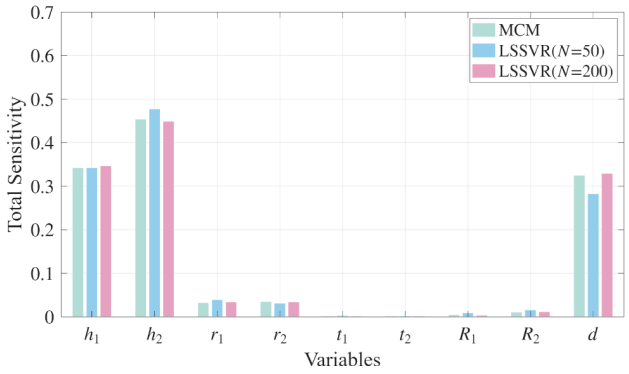


**FIGURE 3**. Total sensitivity index.

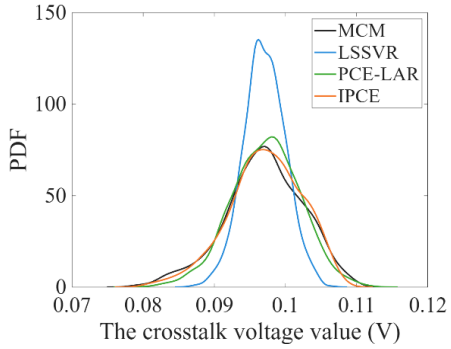


**FIGURE 4**. PDF curve of far-end crosstalk voltage ($N = 50$).

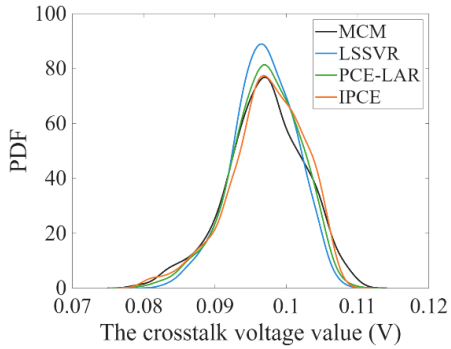


**FIGURE 5**. PDF curve of far-end crosstalk voltage ($N = 200$).

The far-end crosstalk voltage at 75 MHz is defined as the output of interest. Assuming the sensitivity threshold $\epsilon = 0.1$. An LSSVR approximation model is constructed using a limited sample set ($N = 50$) to perform the LSSVR-Sobol sensitivity analysis; the resulting total sensitivity indices are presented in Figure 3. To validate the accuracy of the LSSVR approximate model constructed with a small sample size in LSSVR-Sobol sensitivity analysis, an LSSVR model trained with a large sample size ($N = 200$) and MCM are established as control groups. It is observed that the sensitivity indices of each variable remain largely consistent across all three scenarios. The sensitivity indices for the signal line height $h_1$, victim line height $h_2$, and lateral separation $d$ are prominent, reaching or exceeding 0.3. In contrast, the indices for other input variables remain below 0.05, which is significantly lower than the prescribed sensitivity threshold $\epsilon$.

In accordance with the proposed general preprocessing dimension reduction framework, low-sensitivity variables are fixed at their mean values: $r_1 = r_2 = 0.4\,\text{mm}$, $t_1 = t_2 = 0.3\,\text{mm}$, and $R_1 = R_2 = 50\,\Omega$. Consequently, the initial nine-dimensional stochastic space is reduced to three key variables. The improved PCE (IPCE) model is subsequently applied to the parallel cable crosstalk case; the resulting probability density function (PDF) curves for $N = 50$ and $N = 200$ are depicted in Figures 4 and 5, respectively. The MCM based on 2,000 simulation runs serves as the benchmark data, whereas the LSSVR before dimensionality reduction and the PCE-LAR are established as control groups. At $N = 50$, a significant discrepancy exists between the LSSVR and the MCM benchmark, indicating insufficient predictive accuracy. Conversely, both the PCE-LAR and the IPCE demonstrate a high goodness-of-fit with the MCM, with the IPCE slightly outperforming the PCE-LAR. Notably, although the global accuracy of the LSSVR is limited, it remains sufficient to meet the requirements for the sensitivity analysis shown in Figure 3. When the sample size is increased to $N = 200$, the accuracy of the LSSVR improves significantly, but it still falls short of that of the IPCE and the PCE-LAR. At this juncture, the accuracies of the PCE-LAR and IPCE are highly comparable, making it difficult to distinguish between them only by observing the PDF curves.

To quantify the discrepancies between the PDFs generated by various UQ methods and the benchmark data, the Mean Equivalent Area Method (MEAM) [29] is employed. The MEAM evaluation result is a constant between 0 and 1. The closer the value is to 1, the smaller the discrepancy and the more accurate

the UQ method [30]. The MEAM-based assessment results are summarized in Table 2. Consistent with the findings in Figures 4 and 5, the LSSVR model exhibits limited precision at $N = 50$. When the sample size is increased to $N = 200$, the LSSVR achieves a high level of accuracy; yet it still underperforms compared to the IPCE and the PCE-LAR at $N = 50$. Specifically, at $N = 50$, the PCE-LAR yields an MEAM value of 0.8761, significantly outperforming the LSSVR. The implemented sparse strategy enables it to achieve considerable accuracy under small sample conditions. However, this sparse strategy suffers from an inherent accuracy bottleneck. Even as the sample size increases to $N = 200$, the MEAM value of the PCE-LAR exhibits only marginal improvement, failing to reach 0.9. In stark contrast, the IPCE yields an MEAM value of 0.9435 at $N = 50$, achieving an exceptionally high degree of accuracy.

**TABLE 2**. MEAM index.

| | $N = 50$ | $N = 200$ |
|---|---|---|
| **LSSVR** | 0.5337 | 0.8029 |
| **PCE-LAR** | 0.8761 | 0.8900 |
| **IPCE** | 0.9435 | 0.9553 |

Table 3 presents the root mean square errors (RMSEs) of the respective methods to provide a more comprehensive performance comparison. A vertical comparison clearly reveals that the predictive accuracy of all methods improves as the sample size increases. Conversely, a horizontal comparison demonstrates that the proposed IPCE exhibits the lowest predictive error under both sample sizes, highlighting its exceptional global fitting capability.

**TABLE 3**. RMSE.

| | $N = 50$ | $N = 200$ |
|---|---|---|
| **LSSVR** | 0.0034 | 0.0029 |
| **PCE-LAR** | 0.0027 | 0.0021 |
| **IPCE** | 0.0018 | 0.0014 |

Notably, the PCE-LAR, leveraging its advantages in screening high-dimensional orthogonal basis functions, achieves a predictive accuracy significantly superior to that of the baseline LSSVR model. However, under small sample conditions, the PCE-LAR is prone to falling into local optima or suffering from inevitable truncation errors. In contrast, the proposed preprocessing dimensionality-reduction framework exhibits exceptional robustness. Specifically, at $N = 50$, the RMSE of the IPCE is a mere 0.0018, representing a further reduction of approximately 33.3% compared to the PCE-LAR under identical conditions. This result compellingly substantiates the theoretical advantage of the framework's decoupling strategy: by employing low-cost estimators to preemptively execute rigid dimensionality reduction on the high-dimensional stochastic space (compressing it from 9 dimensions to a 3-dimensional critical feature subspace), the IPCE fundamentally eliminates the underdetermined nature of the subsequent PCE modeling phase.

Excessive cable crosstalk significantly degrades the electromagnetic compatibility of a system; thus, an accurate estimation of worst-case crosstalk scenarios is critical for practical engineering applications. In this study, the proposed IPCE method is employed to estimate the worst-case cable crosstalk within the 1–100 MHz frequency band, and the results are compared with those of the PCE-LAR, as illustrated in Figure 6. In worst-case cable crosstalk estimation studies, the $\mu + 3\sigma$ curve is typically utilized to characterize the worst-case estimates [31]. If the worst-case estimation curve is positioned slightly above the upper bound derived from all MCM simulations, it indicates that the method under evaluation provides a *safe envelope while avoiding overly pessimistic estimations* in worst-case estimation applications. The results demonstrate that, given a sample size of $N = 50$, the IPCE effectively envelopes the upper bound derived from 2,000 MCM simulations. Conversely, the worst-case envelope generated by the PCE-LAR falls significantly below the actual upper boundary of the 2,000 MCM runs. In practical engineering applications, such an *underestimation* poses an extremely high risk of EMC failure.

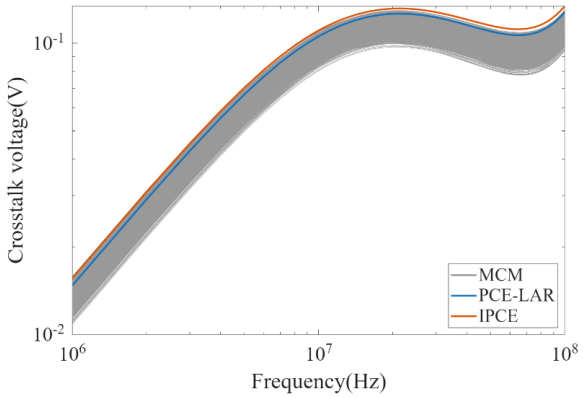


**FIGURE 6**. Worst-case estimate of far-end crosstalk voltage.

## 6. CONCLUSION

This paper proposes a generalized preprocessing dimensionality-reduction framework to enhance the performance of PCE, effectively addressing the curse of dimensionality in high-dimensional EMC uncertainty quantification. By transforming the complex high-dimensional modeling process into a modular workflow — comprising a generalized screening estimator, dimensionality projection, and PCE prediction — the framework exhibits significant flexibility and scalability. The results indicate that instantiating the LSSVR-Sobol module under small sample size conditions enables the precise extraction of critical system features at a minimal computational cost, thereby facilitating a substantial performance enhancement for subsequent PCE modeling within the reduced-dimensional space. Validation results from a cable crosstalk case study demonstrate that the framework successfully compresses the 9-dimensional parameter space into 3 key dimensions. The resulting improved PCE model outperforms traditional sparse methods in terms of predictive accuracy, providing a generalized and efficient preprocessing

strategy for addressing high-dimensional EMC problems with limited samples.

## REFERENCES

[1] Manfredi, P. and R. Trinchero, "Nonparametric formulation of polynomial chaos expansion based on least-square support-vector machines," *Engineering Applications of Artificial Intelligence*, Vol. 133, 108182, Jul. 2024.

[2] Hu, R., V. Monebhurrun, R. Himeno, H. Yokota, and F. Costen, "An uncertainty analysis on finite difference time-domain computations with artificial neural networks: Improving accuracy while maintaining low computational costs," *IEEE Antennas and Propagation Magazine*, Vol. 65, No. 1, 60–70, Feb. 2023.

[3] Huo, S., Y. Song, A. Duffy, and J. Bai, "Uncertainty quantification of human electromagnetic exposure from mobile phone antenna based on Migration Co-Kriging," *IEEE Transactions on Antennas and Propagation*, Vol. 73, No. 11, 9203–9212, Nov. 2025.

[4] Spence, R. and R. S. Soin, *Tolerance Design of Electronic Circuits*, World Scientific, 1997.

[5] 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.

[6] Bai, J., G. Zhang, D. Wang, A. P. Duffy, and L. Wang, "Performance comparison of the SGM and the SCM in EMC simulation," *IEEE Transactions on Electromagnetic Compatibility*, Vol. 58, No. 6, 1739–1746, Dec. 2016.

[7] Ghanem, R. G. and P. D. Spanos, *Stochastic Finite Elements: A Spectral Approach*, Courier Corporation, 2003.

[8] Lüthen, N., S. Marelli, and B. Sudret, "Sparse polynomial chaos expansions: Literature survey and benchmark," *SIAM/ASA Journal on Uncertainty Quantification*, Vol. 9, No. 2, 593–649, 2021.

[9] Hadigol, M. and A. Doostan, "Least squares polynomial chaos expansion: A review of sampling strategies," *Computer Methods in Applied Mechanics and Engineering*, Vol. 332, 382–407, Apr. 2018.

[10] Manfredi, P., "A hybrid polynomial chaos expansion — Gaussian process regression method for Bayesian uncertainty quantification and sensitivity analysis," *Computer Methods in Applied Mechanics and Engineering*, Vol. 436, 117693, Mar. 2025.

[11] Bürkner, P., I. Kröker, S. Oladyshkin, and W. Nowak, "The sparse polynomial chaos expansion: A fully Bayesian approach with joint priors on the coefficients and global selection of terms," [Online]. Available: https://api.semanticscholar.org/CorpusID:248157210, 2022.

[12] Shao, Q., A. Younes, M. Fahs, and T. A. Mara, "Bayesian sparse polynomial chaos expansion for global sensitivity analysis," *Computer Methods in Applied Mechanics and Engineering*, Vol. 318, 474–496, May 2017.

[13] Rumsey, K. N., D. Francom, G. Gibson, J. D. Tucker, and G. Huerta, "Bayesian adaptive polynomial chaos expansions," *Stat*, Vol. 15, No. 1, e70151, Mar. 2026.

[14] Hu, R., V. Monebhurrun, R. Himeno, H. Yokota, and F. Costen, "An adaptive least angle regression method for uncertainty quantification in FDTD computation," *IEEE Transactions on Antennas and Propagation*, Vol. 66, No. 12, 7188–7197, Dec. 2018.

[15] Psaros, A. F., X. Meng, Z. Zou, L. Guo, and G. E. Karniadakis, "Uncertainty quantification in scientific machine learning: Methods, metrics, and comparisons," *Journal of Computational Physics*, Vol. 477, 111902, Mar. 2023.

[16] Rumsey, K. N., G. C. Gibson, D. Francom, and R. Morris, "All emulators are wrong, many are useful, and some are more useful than others: A reproducible comparison of computer model surrogates," [Online]. Available: https://api.semanticscholar.org/CorpusID:283721610, 2025.

[17] He, J., S. A. Mattis, T. D. Butler, and C. N. Dawson, "Data-driven uncertainty quantification for predictive flow and transport modeling using support vector machines," *Computational Geosciences*, Vol. 23, No. 4, 631–645, 2019.

[18] Bhattacharyya, B., "Uncertainty quantification of dynamical systems by a POD-Kriging surrogate model," *Journal of Computational Science*, Vol. 60, 101602, Apr. 2022.

[19] Rasmussen, C. E., "Gaussian processes in machine learning," in *Advanced Lectures on Machine Learning*, Vol. 3176, 63–71, O. Bousquet, U. Von Luxburg, and G. Rätsch (eds.), in Lecture Notes in Computer Science, Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[20] Liu, H., Y.-S. Ong, X. Shen, and J. Cai, "When Gaussian process meets big data: A review of scalable GPs," *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 31, No. 11, 4405–4423, Nov. 2020.

[21] Trinchero, R., M. Larbi, H. M. Torun, F. G. Canavero, and M. Swaminathan, "Machine learning and uncertainty quantification for surrogate models of integrated devices with a large number of parameters," *IEEE Access*, Vol. 7, 4056–4066, 2019.

[22] Sudret, B., "Global sensitivity analysis using polynomial chaos expansions," *Reliability Engineering & System Safety*, Vol. 93, No. 7, 964–979, Jul. 2008.

[23] Ehre, M., I. Papaioannou, and D. Straub, "Global sensitivity analysis in high dimensions with PLS-PCE," *Reliability Engineering & System Safety*, Vol. 198, 106861, Jun. 2020.

[24] Francom, D., B. Sansó, A. Kupresanin, and G. Johannesson, "Sensitivity analysis and emulation for functional data using Bayesian adaptive splines," *Statistica Sinica*, Vol. 28, No. 2, 791–816, 2018.

[25] Horiguchi, A., M. T. Pratola, and T. J. Santner, "Assessing variable activity for Bayesian regression trees," *Reliability Engineering & System Safety*, Vol. 207, 107391, 2021.

[26] McKay, M. D., R. J. Beckman, and W. J. Conover, "A comparison of three methods for selecting values of input variables in the analysis of output from a computer code," *Technometrics*, Vol. 42, No. 1, 55–61, 2000.

[27] 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.

[28] 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.

[29] 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.

[30] 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, Jun. 2017.

[31] Wang, Y., G. Li, Q. Yu, H. Jiang, S. Pi, T. Wang, and Y. Chi, "Mixed uncertainty quantification and imprecise sensitivity analysis for EMC assessment based on probability box theory and polynomial chaos," *IEEE Transactions on Electromagnetic Compatibility*, Vol. 66, No. 2, 493–501, Apr. 2024.