Hindawi
Shock and Vibration
Volume 2019, Article ID 6357104, 13 pages
https://doi.org/10.1155/2019/6357104

Research Article

Reliability Analysis of Structures by Iterative Improved Ensemble of Surrogate Method

Bolin Liu [1,2] and Liyang Xie [1,2]

[1] *School of Mechanical Engineering & Automation, Northeastern University, Shenyang 110819, China*
[2] *Key Laboratory of Vibration and Control of Aero-Propulsion System Ministry of Education, Northeastern University, Shenyang, 110819, China*

Correspondence should be addressed to Bolin Liu; strive_1980@163.com

Received 23 April 2019; Accepted 3 October 2019; Published 24 October 2019

Academic Editor: Laurent Mevel

Surrogate models have been widely adopted for reliability analysis. The common approach is to construct a series of surrogates based on a training set and then pick out the best one with the highest accuracy as an approximation of the time-consuming limit state function. However, the traditional method increases the risk of adopting an inappropriate model and does not take full advantage of the data devoted to constructing different surrogates. Furthermore, obtaining more samples is very expensive and sometimes even impossible. Therefore, to save the cost of constructing the surrogate and improve the prediction accuracy, an ensemble strategy is proposed in this paper for efficiently analyzing the structural reliability. The values of the weights are obtained by a recursive process and the leave-one-out technique, in which the values are updated in each iteration until a given prediction accuracy is achieved. Besides, a learning function is used to guide the selection of the next sampling candidate. Because the learning function utilizes the uncertainty estimator of the surrogate to guide the design of experiments (DoE), to accurately calculate the uncertainty estimator of the ensemble of surrogates, the concept of weighted mean square error is proposed. After the high-quality ensemble of surrogates of the limit state function is available, the Monte Carlo method is employed to calculate the failure probabilities. The proposed method is evaluated by three analytic problems and one engineering problem. The results show that the proposed ensemble of surrogates has better prediction accuracy and robustness than the stand-alone surrogates and the existing ensemble techniques.

1. Introduction

Nowadays, computer simulations are a major tool to design engineering structures for accurate analysis of their performance. Although the computer processing power along with memory and storage capacities has been drastically increased, Goel et al. [1] pointed out that analysis models of acceptable accuracy have required at least six to eight hours of CPU time. At the same time, uncertainty in the parameters characterizing the mechanical behavior of a structure and loads acting on it calls for reliability analysis. When computer simulations are combined with the reliability assessments, the computational cost tends to increase, especially when complex nonlinear limit state functions are involved [2]. Thus, assessing the reliability of a complex structure requires a transaction between the reliability algorithms and numerical simulation methods used to analyze the mechanical behavior of the structure.

In a reliability assessment problem, the safety domain of the structure under a given failure mode is described by the limit state function $g(x)$, which is often determined by the FEM G and a given threshold value z, and then the response function $g(x)$ is defined by

$$g(x) = G(x) - z, \tag{1}$$

where $x = [x_1, \ldots, x_n]^T$ represents the vector of random variables and $g(x) = 0$ is the limit state function of the structure. The failure domain is defined by $g(x) > 0$, and the safe domain is defined by $g(x) < 0$.

The failure probability of the structure is computed by a multidimensional integral of the joint probability density function f_x:

$$P_x = \int \ldots \int_{g(x)>0} f(x)\mathrm{d}x. \quad (2)$$

It is typically not feasible to calculate the integral of equation (2), particularly for problems involving implicit limit state functions.

Recently, surrogate models have been introduced by several scholars to replace the limit state functions in the reliability field, including quadratic response surfaces [3], neural networks [4], support vector machines [5], and Kriging [6–8]. These methods obtain a compromise between the complexity and nonlinearity of surrogates.

Different from other surrogate models, Kriging not only predicts the mean value of the structural response but also provides the local uncertainty measure (the so-called Kriging variance). In addition to surrogate formulation, the design of experiments (DoE) also has a considerable influence on the efficiency and accuracy of reliability analysis [9, 10]. Therefore, Kriging-based sequential strategies of the design of experiments (DoE) have drawn more and more attention because it is an active learning process and can update itself by adding new training point based on the statistical information provided by the Kriging model. So far, several Kriging-based reliability methods with an adaptive DoE have been proposed utilizing the Kriging variance [11–13].

The above surrogate-based methods improve the accuracy of structural reliability analysis and reduce the number of calls to the real performance function to some extent. However, because there is not enough information to describe the relationship between the output response and the input variables, it is difficult for engineers to know which surrogate is the best for a specific limit state response. In addition, according to Goel et al. [1], due to the influence of the selected DoE type, the number of design points in the training data set, and the form (e.g., linear, nonlinear) of the limit state function, there is also uncertainty in surrogate predictions.

At present, for the surrogate-based reliability evaluation problem, scholars mainly focus on the selection of different surrogates and rarely pay attention to the application of an ensemble of surrogates. But some researchers observed that no single surrogate model was found to be most effective for all problems.

The idea of combination can be traced to the development of committees of neural networks by Perrone and Cooper [14] with further refinement by Bishop [21]. Zerpa et al. [22] and Goel et al. [1] extended this idea to the ensemble of metamodels and found that multiple metamodels can identify the possible regions with high errors where predictions of metamodels vary widely. Thereby, it can guide the researchers to obtain more sample points in this uncertain region and reduce prediction errors. At the same time, they pointed out that combining metamodels can provide researchers with a more robust prediction and effectively eliminate the negative impact brought by inappropriate stand-alone metamodels. In 2009, Acar and Rais-Rohani [15] proposed a combining technique with optimized weight coefficients. They get the weights by minimizing GMSE or RMSE using a formal optimization algorithm. Later, other ensemble techniques are proposed by scholars, such as BestPRESS [1] and OWS [17]. Essentially, they are the same; the difference between them is that the method of calculating the weight coefficient is different. Those techniques could achieve a more satisfactory result than individual surrogates in some cases. However, those methods are not only time-consuming but also have the following two disadvantages. One is that they could not ensure obtaining a globally optimal solution, easily fall into a local optimum, and sometimes even have no locally optimal solution. The other is that the range of weight coefficients (w_i) is not constrained and any value can be taken. For the actual problem, $w_i < 0$ is unreasonable.

In view of the shortcomings of the above methods, Zhou proposed an ensemble technique with recursive arithmetic average [16], in which the weights are obtained using a recursive process. These weights are updated in each iteration until the last ensemble reaches a desirable prediction accuracy. Unlike the previous method of arithmetically averaging the responses of the stand-alone metamodels just once, this technique builds an ensemble of metamodels by recursive arithmetic average several times. It provides a balance between model prediction accuracy and modeling time. But this method uses the randomly selected samples to construct the ensemble of metamodels. This results in a low precision if too few samples are used, or wastes cost creating models that are accurate in areas where they need not be. Moreover, Zhou's method has the following shortcoming: if the initial weights are incorrectly selected, this will lead to more iterations.

As mentioned above, the design of experiments (DoE) also has huge influence on the convergence rate and the accuracy of reliability analysis, and the surrogate uncertainty estimator can guide the design of experiments (DoE). However, there is relatively little research on the uncertainty of the ensemble of surrogates. Goel et al. [1] proposed a method to identify the region of ensemble of surrogates with large uncertainty. He used the standard deviation (SD) of multiple individual surrogates' predictions to identify areas where the prediction error of ensemble of surrogates is high. The SD value of the predictions will be large in regions where the surrogates differ greatly. A high SD may indicate that there is a region with high uncertainty in the predictions of any of the surrogates, and adding sampling points can reduce the uncertainty in such region. However, this method does not take into account the influence of the prediction precision of an individual surrogate. It is unreasonable that the value of the standard deviation is completely determined by the single surrogate when the prediction of a single surrogate differs greatly from that of an ensemble of surrogates.

In this paper, we propose an ensemble technique with recursive arithmetic average for structural reliability analysis. In calculating the uncertainty of the ensemble of surrogates, we consider the influence of the precision of the prediction of each stand-alone surrogate and then utilize an active learning function to add new sample point in the vicinity of the limit state function based on the statistical information provided by the ensemble of surrogates model.

The remainder of this paper is organized as follows. Section 2 gives an overview of the ensemble of surrogates method briefly. Section 3 introduces our method for running the reliability analysis algorithm with ensemble of surrogates. Section 4 presents the academic validation and compares the performance of our proposed ensemble technique to that of other available methods. Finally, Section 5 presents the conclusion.

2. Ensembles of Surrogates

In recent years, in order to improve the accuracy of the surrogate and make up for the lack of a stand-alone surrogate, the ensemble technology has been developed. The ensemble of surrogates is composed of several stand-alone surrogates which are multiplied by different weight coefficients. Using the weight-sum formulation, the ensemble of surrogates for approximation of response can be expressed as

$$\widehat{y}_{\mathrm{EN}}(x) = \sum_{i=1}^{N} w_i(x)\widehat{y}_i(x), \tag{3}$$

where x is the input variable, $\widehat{y}_{\mathrm{EN}}(x)$ denotes the predicted response by the ensemble of surrogates, N is the number of surrogates in the ensembles, $w_i(x)$ is the weight coefficient for the ith surrogate, and $\widehat{y}_i(x)$ is the predicted response of the ith surrogate.

The weight coefficients in equation (3) are usually satisfied:

$$\sum_{i=1}^{N} w_i(x) = 1. \tag{4}$$

Generally, the weight coefficients of individual surrogates can be obtained based upon global and/or local measures [14, 15]. The surrogates with high accuracy have large weight factor and vice versa. Considering the calculation cost of the actual engineering problem, global error metric with a generalized mean square cross-validation error (GMSE) is proposed by Acar and Rais-Rohani [15]. The GMSE can be written as follows:

$$\mathrm{GMSE} = \frac{1}{n}\sum_{j=1}^{n}\left(y^j - \widehat{y}_{\mathrm{EN}}^j\right)^2, \tag{5}$$

where y^j and $\widehat{y}_{\mathrm{EN}}^j$ denote the true response at x_j and corresponding predicted value from the ensemble of surrogates constructed by using all but the jth design point (i.e., leave-one-out cross-validation strategy), respectively, and n is the number of sampling points.

2.1. Weight Coefficients Selection. At present, a variety of techniques for constructing more accurate ensemble of surrogates are proposed. There are mainly the following methods.

2.1.1. Weight Factors Selection Based on Cross-Validation Errors

(1) Heuristic computation of the weight coefficient: Goel et al. [1] proposed a heuristic weight scheme, namely, the prediction-sum-of-squares-based weighted average surrogate (PWS). The weight coefficients are computed as follows:

$$\omega_i = \frac{\omega_i^*}{\sum_{j=1}^{N}\omega_j^*},$$
$$\omega_i^* = \left(E_i + \alpha E_{\mathrm{average}}\right)^{\beta}, \tag{6}$$
$$E_{\mathrm{average}} = \frac{1}{N}\sum_{i=1}^{N} E_i,$$

where E_i is the generalized mean square error (GMSE) of the ith surrogate with $\alpha = 0.05$ and $\beta = -1$ suggested by Goel et al.

(2) The approach based on minimizing GMSE: this optimal weighted surrogate approach was proposed by Acar and Rais-Rohani [15], which is achieved by minimizing some error metric, such as GMSE error. The optimization problem is presented as follows.

Find $\omega_i, i = 1, 2, \ldots, N$

$$\min\ \mathrm{GMSE}_e = \frac{1}{N}\sum_{i=1}^{N}\left[y(x_i) - \widehat{y}_{\mathrm{EN}}\left(w_i, \widehat{y}^{(-i)}(x_i)\right)\right]^2 \tag{7}$$
$$\text{s.t.}\quad w_i \geq 0, \sum_{i=1}^{N} w_i = 1,$$

where $y(x_i) - \widehat{y}_{\mathrm{EN}}(w_i, \widehat{y}^{(-i)}(x_i))$ is the predicted value of the ensemble of surrogates for all training sample points except point $(x_i, y(x_i))$.

2.1.2. The Ensemble Technique with Recursive Arithmetic Average. Zhou proposed an ensemble technique with recursive arithmetic average [16], in which the weights are obtained using a recursive process. It is composed of the following steps:

(1) initial weight coefficients

(2) Generate initial samples using sampling method, call the performance function to calculate the structural response at those initial samples, use these training samples, and construct N candidate surrogates

(3) Calculate the prediction mean square errors (MSE) of N candidate surrogates separately

(4) Find out the individual surrogate with the largest MSE and the individual surrogate with the smallest MSE, while $(\mathrm{MSE}_{\mathrm{largest}} - \mathrm{MSE}_{\mathrm{smallset}}) > \mathrm{tolerance}$

(5) Obtain the arithmetic average of the candidate N surrogates

(6) Use the simple average surrogate made in Step 5 to replace the surrogate which has the largest prediction MSE, obtain N new surrogates, of which $N-1$ surrogates are not changed, and then the weights of the initial individual surrogates are calculated and updated

(7) Perform the same work as Step 4; if the condition in while $(\cdot)$ is satisfied, go back to Step 5; otherwise, break out of the loop

End While

(8) Output the optimal weight coefficients

2.2. Prediction Metrics. The coefficient of determination (R^2) and the root mean square error (RMSE) are often used to compare the prediction capabilities of different surrogate models. The coefficient of determination (R^2) is given as

$$R^2 = 1 - \frac{\sum_{k=1}^{n}(y_k - \tilde{y}_k)^2}{\sum_{k=1}^{n}(y_k - \overline{y}_k)^2}, \quad (8)$$

and the root mean square error (RMSE) is given as

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{k=1}^{n}(y_k - \tilde{y}_k)^2}, \quad (9)$$

where n is the number of test samples, $\overline{y}_k$ is the mean of actual response, $\tilde{y}_k$ is the mean of predicted response, and y_k denotes the true response. R^2 indicates how well the surrogate model fitted the actual response model; a higher value represents a better fit. RMSE is an indicator of the precision of a surrogate model, where smaller value means a more accurate surrogate model.

Since the experiments are repeated many times, the coefficient of variation of R^2 and RMSE are used to evaluate the prediction accuracy of the surrogate. The definition of coefficient of variation is as follows:

$$\psi = \frac{\delta}{\mu}, \quad (10)$$

where μ and δ denote the mean and the standard deviation of R^2 and RMSE, respectively.

3. Ensemble of Surrogates for Structural Reliability Analysis

3.1. The Proposed Ensemble of Surrogates with Iterative Weight Coefficients. As mentioned above, the ensemble of surrogates can identify the possible regions with high errors and can provide researchers with a more robust prediction. So in this paper, we propose the ensemble technique for structural reliability analysis.

In practical applications, the ensemble of surrogates can be established by the following two ways:

(i) A set of surrogates created based on different techniques—such as polynomial response surface (PRS), radial basis function (RBF), and support vector regression (SVR)

(ii) Different instances of one surrogate; for example, several surrogates can be obtained by changing the types of regression functions of Kriging (KRG)

In this article, in order to use the statistical information provided by the Kriging to guide DoE, we use the second approach to build the ensemble of surrogates.

Kriging assumes that the response of interest $F(x)$ includes the linear regression part and the nonparametric part:

$$F(x) = h(x)^T\beta + Z(x), \quad (11)$$

where $h(x)$ is a scalar or multivariable polynomial, β is the coefficient vector of $h(x)$ estimated with generalized least squares, $Z(x)$ follows a Gaussian process with zero-mean and constant variance, and the covariance between $Z(x_i)$ and $Z(x_j)$ is defined as

$$\operatorname{cov}(Z(x_i), Z(x_j)) = \sigma^2 R(x_i, x_j, \theta), \quad (12)$$

where σ^2 and $R(x_i, x_j, \theta)$ denote the variance and correlation function of $Z(x)$, respectively. θ is a parameter vector.

The predictions of the KRG at a point are

$$\widehat{G}(x) = F\widehat{\beta} + r^T R^{-1}(G(x) - F\widehat{\beta}), \quad (13)$$

where F is the matrix of linear equations constructed by the regression function and the experimental design, $\widehat{\beta} = (F^T R^{-1} F)^{-1} F^T R^{-1} G$ is the generalized least square estimate of β, R is an $n \times n$ matrix correlation between Z at design sites, r is the vector of correlations between point x and the points of the training design, and $G = [G_1 \ldots G_n]^T$ is a vector of the training point observations. The Kriging prediction variance can be calculated as

$$s^2(x) = \sigma^2 - \begin{bmatrix} F & r^T \end{bmatrix} \begin{bmatrix} 0 & F^T \\ F & R \end{bmatrix} \begin{bmatrix} F \\ r \end{bmatrix}^T. \quad (14)$$

The prediction precision of KRG is determined by regression functions and correlation functions. For a certain problem, Kriging models with different accuracy can be obtained by different combinations of correlation functions and regression models. However, the choice of correlation functions and regression models depends on human experience. Therefore, ensemble technique is an effective way to make up for the shortfalls of the above strategy.

For the most current combining techniques, they build an ensemble of surrogates by arithmetically averaging the responses of the stand-alone metamodel just once. But they could not ensure obtaining the optimal weight coefficients or take a lot of time to build the surrogate. Therefore, we propose an iterative method to obtain the weight coefficients, in which the values of these weight coefficients are updated in each iteration until a desirable prediction accuracy is achieved. In order to improve the modeling efficiency, we fist use a cross-validation strategy and the prediction-sum-of-squares-based weighted average surrogate (PWS) to calculate the initial weight coefficients and then employ recursive process to obtain the best weight coefficients.

3.2. Adaptive Sequential Sampling. According to [18] and [12], besides surrogate formulation, the design of experiments (DoE) also has great influence on the convergence rate and the accuracy of reliability analysis. A good strategy of DoE enables the reliability calculation process to converge quickly and provides higher computational accuracy at the same time.

Within recent years, several Kriging-based reliability methods with an adaptive DoE have been proposed utilizing the KRG variance [11, 12]. Examples of such methods include the efficient global reliability analysis (EGRA) proposed by Bichon et al. [11] and the active learning reliability method combining Kriging and MCS (AK-MCS) proposed by Echard et al. [12]. These methods use both the prediction and uncertainty estimates offered by the Kriging model to select the new sampling point for building an accurate surrogate. Such adaptive methods not only improve the accuracy of structural reliability analysis but also reduce the number of calls to the real performance function.

In our proposed method, to speed up the construction of the ensemble of surrogates, we use EGRA method to add sequentially training set and then update the surrogate model during each iteration until the predefined criterion is satisfied.

The EGRA algorithm uses the expected feasibility function (EFF) to search for points in the vicinity of the limit state function. The maximum point of EFF is added into the DoE step by step. EFF is defined as

$$\mathrm{EFF}[x] = \int_{z-\varepsilon}^{z+\varepsilon} [\varepsilon - |z - G(x)] f_{\widehat{G}} \mathrm{d}x, \tag{15}$$

where G is a realization of the distribution $\widehat{G}$. This integral can be calculated as

$$\begin{aligned}\mathrm{EFF}[x] = &\left(\mu_{\widehat{G}}(x) - z\right)\left[2\Phi\left(\frac{z-\mu_{\widehat{G}}(x)}{\sigma_{\widehat{G}}(x)}\right) - \Phi\left(\frac{(z-\varepsilon)-\mu_{\widehat{G}}(x)}{\sigma_{\widehat{G}}(x)}\right)\right.\\ &\left.- \Phi\left(\frac{(z+\varepsilon)-\mu_{\widehat{G}}(x)}{\sigma_{\widehat{G}}(x)}\right)\right] - \sigma_{\widehat{G}}(x)\left[2\phi\left(\frac{z-\mu_{\widehat{G}}(x)}{\sigma_{\widehat{G}}(x)}\right)\right.\\ &\left.- \phi\left(\frac{(z-\varepsilon)-\mu_{\widehat{G}}(x)}{\sigma_{\widehat{G}}(x)}\right) - \phi\left(\frac{(z+\varepsilon)-\mu_{\widehat{G}}(x)}{\sigma_{\widehat{G}}(x)}\right)\right]\\ &+ \varepsilon\left[\Phi\left(\frac{(z+\varepsilon)-\mu_{\widehat{G}}(x)}{\sigma_{\widehat{G}}(x)}\right) - \Phi\left(\frac{(z-\varepsilon)-\mu_{\widehat{G}}(x)}{\sigma_{\widehat{G}}(x)}\right)\right],\end{aligned} \tag{16}$$

where $\varepsilon = 2\sigma_{\widehat{G}}(x)$ and z is a constant ($z = 0$). The new sample can be obtained by maximizing the EFF and adding this new sample to the previous set.

$$x^* = \arg(\max(\mathrm{EFF}(x))). \tag{17}$$

EGRA iteratively adds training points to the data set by maximizing the EFF until the stopping criterion is met.

3.3. The Weighted Mean Square Error. Because EGRA uses the surrogate uncertainty estimator to guide the selection of the next sampling candidate, it requires that the surrogate can provide uncertainty estimates. As mentioned above, there is relatively little research on the uncertainty of the ensemble of surrogates. Although Goel et al. [1] proposed a method to identify the region of ensemble of surrogates with standard deviation (SD), this method does not take into account the influence of the prediction precision of the stand-alone surrogate. SD is defined as

$$\begin{aligned}\mathrm{SD}_{\mathrm{EN}}(\widehat{y}(x)) &= \sqrt{\frac{\sum_{i=1}^{N}\left(\widehat{y}_j(x) - \overline{y}(x)\right)^2}{N-1}},\\ \overline{y} &= \frac{\sum_{i=1}^{N} y_i(x)}{N},\end{aligned} \tag{18}$$

where N is the number of surrogates, $\widehat{y}_i(x)$ denotes the predicted response of the ith surrogate, and $\overline{y}$ is the mean predicted value of all surrogates.

In this research, we propose the concept of weighted mean square error to calculate the uncertainty for ensemble of surrogates, which is constructed as

$$\mathrm{MSE}_{\mathrm{EN}}(\widehat{y}(x)) = \sum_{j=1}^{N} \omega_i \left[\widehat{y}_j(x) - \widehat{y}_{\mathrm{EN}}(x)\right]^2, \tag{19}$$

where x is the input variable, $\widehat{y}_{\mathrm{EN}}(x)$ denotes the predicted response by the ensemble of surrogates, N is the number of surrogates in the ensembles, $w_i(x)$ is the weight coefficient for the ith surrogate, and $\widehat{y}_i(x)$ is the predicted response of the ith surrogate.

We demonstrate the application of the weighted mean square error to identify the region of high uncertainty of the ensemble of surrogates. The result for a single instance of a DoE for a two-dimensional limit state function [11] is presented in detail. This function is given as follows:

$$\begin{aligned}g(x) &= \frac{(x_1^2+4)(x_2-1)}{20} - \sin\frac{5x_1}{2} - 2,\\ x_1 &\in [-4, 7],\\ x_2 &\in [-3, 8].\end{aligned} \tag{20}$$

Figure 1 shows the contour plots of absolute error ($|y(x) - \widehat{y}(x)|$), standard deviation (SD), and the square root of the weighted mean square error in prediction. From Figure 1(a), it can be seen that the ensemble of surrogates has a high prediction accuracy in most areas (the absolute error is relatively small), but the absolute error is larger in the top middle boundary and bottom middle boundary. The contour plot of the square root of the weighted mean square error (Figure 1(b)) also shows the region of high uncertainty near the above areas, indicating that the weighted mean square error can accurately identify the region of high uncertainty.

From Figure 1(c), we can note that the standard deviation (SD) has high values near the bottom right corner and the top left corner, but the absolute error is not high. This means that the standard deviation (SD) cannot accurately identify regions of high uncertainty of the ensemble of surrogates. The reason for this may be due to the

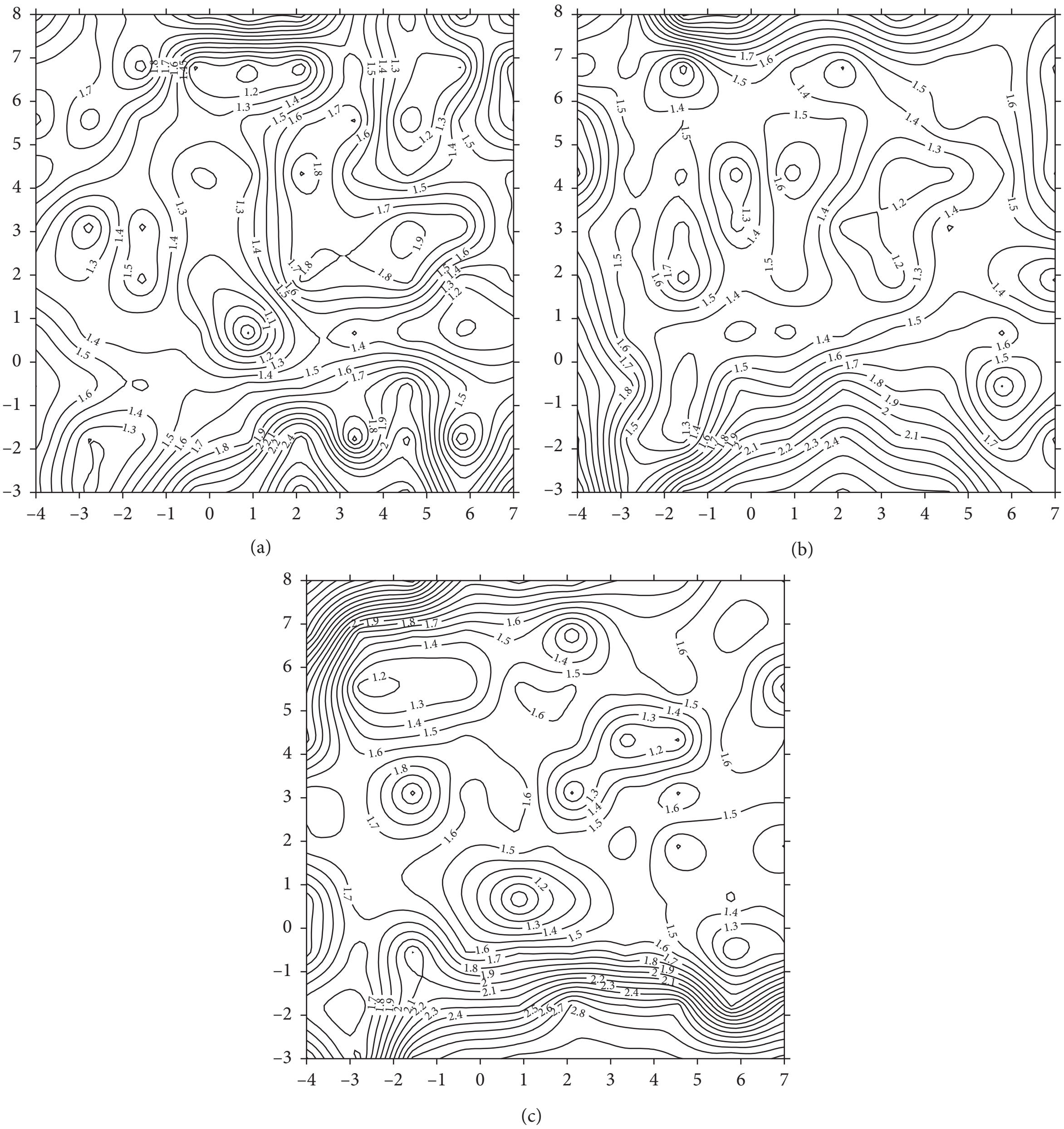

Figure 1: Contour plots of absolute error, the square root of the weighted mean square error, and standard deviation (SD) for a two-dimensional function. (a) Absolute error. (b) The root weighted mean square error. (c) Standard deviation (SD).

poor prediction accuracy of one individual surrogate. Compared with the standard deviation (SD), the weighted mean square error can make a more reasonable assessment of the uncertainty of the ensemble of surrogates.

In order to demonstrate the independence of the result with respect to the design of experiments, 1,000 DoEs are carried out for the function expressed in formulas (21), and then the maximum square root of weighted mean square error, the minimum square root of weighted mean square error, and corresponding locations are obtained. At the same time, the actual errors in the predictions of different surrogates at those locations are also calculated. Figure 2(a) shows the magnitude of maximum square root of the weighted mean square error and actual errors for different surrogates. Figure 2(b) shows the magnitude of minimum square root of the weighted mean square error and actual errors for different surrogates. In Figure 2, s-WMS denotes the square root of the weighted mean square error, and e_EN, e_KRG-con, e_KRG-lin, and e_KRG-qua represent the actual errors of EN, KRG-con, KRG-lin, and KRG-qua, respectively. By comparing the boxplots for the errors in predictions, we can see that the large weighted mean square errors correspond to areas of great uncertainty, while the small weighted mean square errors correspond to areas of small uncertainty. We also find that the actual errors at the locations of maximum weighted mean square errors are high, and the actual errors at the locations of minimum weighted mean square errors are low. This result shows that

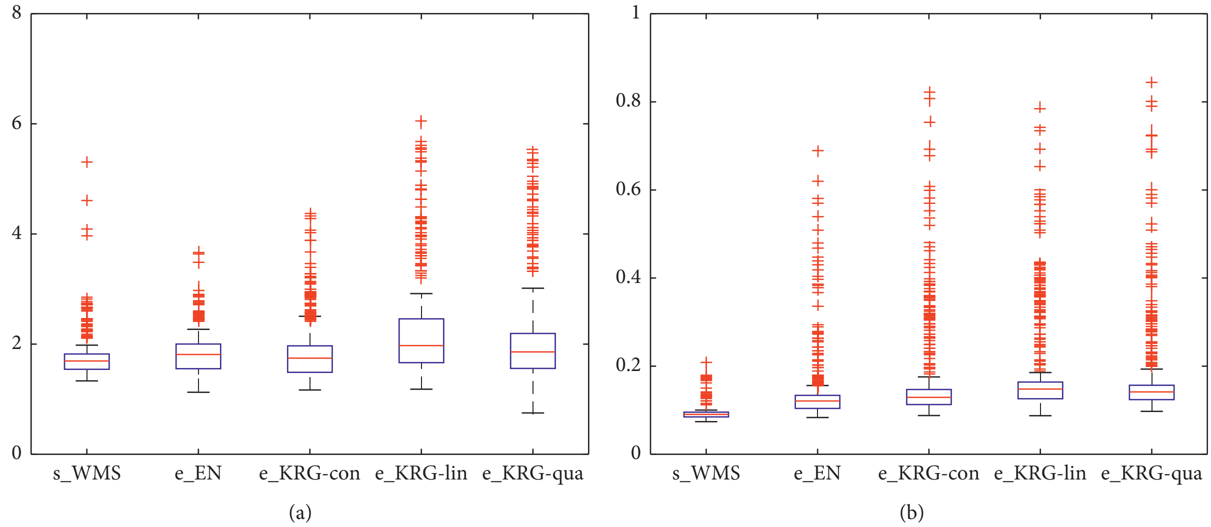

FIGURE 2: Maximum/minimum square root of the weighted mean square error and actual errors in prediction of different surrogates at corresponding locations. (a) Maximum square root of the weighted mean square error and corresponding actual errors. (b) Minimum square root of the weighted mean square error and corresponding actual errors.

there is a high correlation between the weighted mean square error and the uncertainty of the ensemble model. In the sequential sampling process, we pay more attention to find regions with greater uncertainty in prediction and add sampling points to these locations instead of quantifying the magnitude of actual errors. The weighted mean square error can qualitatively identify regions with large uncertainties in prediction; therefore, is well suitable for sequential updating of ensemble models.

3.4. Leave-One-Out Technique. As mentioned above, for an ensemble technique that uses a recursive process to obtain the weights, if the weight coefficients are incorrectly selected, this will lead to more iterations. To improve modeling efficiency, we employ the leave-one-out technique with respect to GMSE to obtain the weight coefficients in each iteration; namely,

$$\text{GMSE} = \frac{1}{n}\sum_{k=1}^{n}\left(y_k - \tilde{y}_{(k)}\right)^2, \tag{21}$$

where n denotes the number of test samples, y_k is the actual response at x_k, and $\tilde{y}_{(k)}$ is the corresponding predicted response from the metamodel constructed using all except the kth design point.

The basic frame of our proposed algorithm is composed of the following steps:

(1) Generate a small number of samples from the true response function. Here, Latin hypercube sampling (LHS) is used to generate the initial samples.

(2) Use these samples to construct multiple surrogates. Those candidate surrogates can be obtained by changing the types of regression functions.

(3) Calculate the generalized mean square cross-validation error (GMSE) of each surrogate and use the heuristic weight scheme to calculate weight coefficients of ensemble of surrogates.

(4) Calculate the uncertainty of ensemble of surrogates by the proposed weighted mean square error method (equation (19)).

(5) Find the point with the maximum EFF (equation (17)) of ensemble of surrogates and determine whether max(EFF(x)) is smaller than the given e_given. If so, go directly to Step 9.

(6) Evaluate the true response function at the point obtained from Step 5.

(7) Add this new sample to the previous set and update all surrogates.

(8) If max(EFF(x)) is smaller than the given e_givens, end the iteration, and go to Step 9. Otherwise, go to step 2 and continue the iterative process.

(9) Use this surrogate model to calculate the probability of failure.

4. Academic Validation

4.1. Analytical Problems. In order to test the performance of our proposed ensemble technique, we choose the following analytic functions which are used in the literature, and then one engineering example with nonlinear behavior and high dimension is used to demonstrate the advantage of the proposed method. The details of each test problem are given as follows:

(1) Numerical Example 1: this example is a two-dimensional nonlinear function already analyzed in [11]:

$$G(x) = x_1^2 + x_2^2 - 18, \tag{22}$$

where x_1 and x_2 are subject to normal distribution, respectively, and the variables are uncorrelated. The distribution of x_1 is normal ($\mu = 10, \sigma = 5$) and x_2 is normal ($\mu = 9.9, \sigma = 5$). In this example, the response level is $\overline{z} = 0$, and the probability of failure P_{f} is then

$$P_{\text{f}} = P(g(x) > 0). \quad (23)$$

(2) Numerical Example 2: this example consists of a modified Rastrigin function adopted in [18]:

$$G(x) = 10 - \sum_{n=1}^{2}\left(x_n^2 - 5\cos\left(2\pi x_n\right)\right), \quad (24)$$

where x_1 and x_2 are subject to the standard normal distribution, and the variables are uncorrelated. In this example, the response level is $\overline{z} = 0$, and the probability of failure P_{f} is then

$$P_{\text{f}} = P(g(x) > 0). \quad (25)$$

The limit state of equation (24) is much more complicated than the one in Example 1, due to the complex failure domain composed of multiple failure regions.

(3) Numerical Example 3: this example is a multidimensional problem which was proposed in Rackwitz [19]:

$$G\left(x_1, \ldots, x_m\right) = m + 3\sigma\sqrt{m} - \sum_{j=1}^{m} x_j, \quad (26)$$

where x_j, $j = 1, \ldots, m$, are subject to lognormal distribution, and the variables are uncorrelated with unit means and standard deviations ($\sigma = 0.2$). In this example, the response level is $\overline{z} = 0$, and the probability of failure P_{f} is then

$$P_{\text{f}} = P(g(x) > 0). \quad (27)$$

The limit state of equation (28) has equal curvature and its concavity is pointed toward the origin.

(4) Stiffened plate element problem: an implicit limit state function is involved which has already been studied in [20].

When comparing the accuracy of one surrogate, considering only analytical problem functions can be very misleading, because they often do not represent the various characteristics of true engineering design problems. Therefore, we also consider the following engineering design problems from the literature. The limit state function is given by

$$g(x) = \sigma_u(x) - \left(\sigma_{\text{sw}}(x) + \sigma_{\text{wi}}(x)\right), \quad (28)$$

where $\sigma_u(x)$ is the ultimate compressive stress of the stiffened plate elements, and $\sigma_{\text{sw}}(x)$ and $\sigma_{\text{wi}}(x)$ are the still water component and the wave-induced component of the uniaxial compressive stress, which are induced by the ship hull girder bending moments.

The ultimate compressive stress is an implicit function of the vector of basic random variables, which is defined as the maximum stress value of the average stress-strain curve of the stiffened plate elements under uniaxial compression and computed through nonlinear FEA:

$$\sigma_u(x) = \max_{\varepsilon x}\left\{\sigma_x\left(x; \varepsilon_x\right)\right\}. \quad (29)$$

For the distributions of the basic random variables, see [20]. Due to the nonlinear behavior of the material, the finite element analysis process is very time-consuming. To increase the analytical efficiency, symmetry boundary conditions are imposed at the plate longitudinal edges and the stiffener mid-span transverse sections. Figure 3 shows the nonlinear FEA structural model.

4.2. Design Experiments. For the above problems, the Latin hypercube sampling (LHS) method is used to generate the training sample. In order to reduce the effect of random sampling, 1000 different training sets are used for all the analytical problems. For stiffened plate element problem, taking into account the computational cost of each simulation, the surrogates are constructed using only one single training set with 120 training points. For all the problems, the samples are selected in the space over the bounds $\pm 5\sigma$. Additional information about the training and test data sets is provided in 1.

4.3. Ensemble of Surrogates Techniques. According to the theoretical researches of some scholars [16], when the number of metamodels is maintained at 3–5, the prediction accuracy of ensemble models is relatively high. Therefore, three types of regression functions are considered here, including constant regression, linear regression, and quadratic regression. The three Kriging models constructed from these three regression functions serve as the three members of the ensemble of surrogates, and the ensemble of surrogates is developed based on our proposed techniques. These three Kriging models are also used as the three members of the ensemble that is developed based on the three previously described techniques.

5. Results and Discussion

5.1. Comparison and Analysis of Surrogate Models. In order to facilitate the comparison and analysis of the results, each stand-alone and all ensembles of surrogates are marked as shown in Table 2.

To compare the performance of different metamodeling models, the mean and the coefficient of variation (COV) of R^2 and RMSE are used in all cases to illustrate the accuracy and robustness of the stand-alone and ensemble of surrogates, respectively.

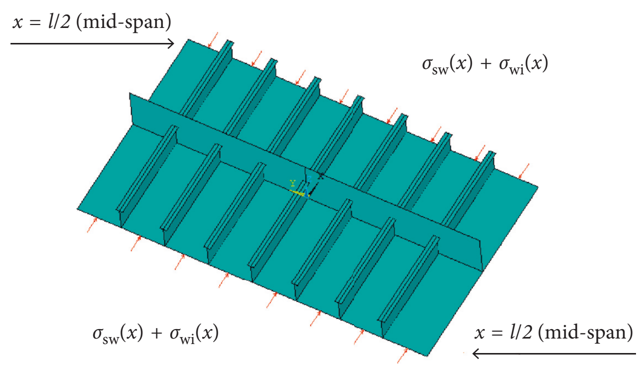

Figure 3: Stiffened plate under uniaxial compression in the longitudinal direction.

Table 1: Summary of training data and test data used in each problem.

Problem	Design variables	Training sets	Design points	Test points
Numerical Example 1	2	1000	12	405
Numerical Example 2	2	1000	12	405
Numerical Example 3	6	200	56	512
Numerical Example 4	8	1	120	40

Table 2: Summary of surrogate modeling strategies.

Acronym	Description
KRG-con	The Kriging model with constant regression
KRG-lin	The Kriging model with linear regression
KRG-qua	The Kriging model with quadratic regression
EH	The ensemble of surrogates which is constructed by the heuristic method of Goel et al. [1]
EM	The ensemble of surrogates which is based on minimizing GMSE method in Acar and Rais-Rohani [15]
ER	The ensemble of surrogates which is based on recursive arithmetic average technique of Zhou et al. [16]
ERR	The ensemble of surrogates which is constructed based on our proposed method

5.1.1. The Coefficient of Determination (R^2). The boxplots can provide us with a graphical depiction of how the value of the metric varies over the range of training sets used. The bottom of the box represents the lower quartile, the top of the box represents the upper quartile, and the inner line of the box represents the median value. The broken line, which is extended from the end of the box, represents the range of the remaining data relative to the upper and lower quartiles. Figure 4 shows the boxplots of the coefficient of determination (R^2) errors for all cases. Based on the comparison and analysis in Figure 4, we can make the following findings. (1) No single surrogate performs the best on all problems. Compared with them, the ensemble models have better prediction ability than the stand-alone metamodel and are less influenced by the experimental design. This suggests that using an ensemble of surrogate models potentially yields robust approximation. (2) For all examples, the EM, ER, and ERR have higher prediction accuracy, but the long tail of EM indicates that EM is less robust than ER and ERR. (3) For most problems, ERR has the best median value, indicating that ERR performs best in all ensemble models. (4) Of all the ensemble models, EH has the worst performance in Example 1, Example 3, and Example 4 and has the second worst performance in Example 2, which reveals that EH cannot perfectly capture the true prediction errors.

For all metamodeling models, the mean and ψ value of R^2 of each test example are shown in Table 3. From the table, it can be seen the following: except for Example 3, the average coefficient of determination for ERR is the best, and the performances of the other ensembles and individual surrogates have changed significantly due to different test problems. Even the average coefficient of determination of ERR in Example 3 is larger than that of ER, but the value of ψ is minimal, indicating that ERR is more robust. We also note that although the performances of other ensembles are worse than that of ERR, they are better than the worst individual surrogates. This suggests that using an ensemble of surrogate models, we can protect against poor choice of a surrogate. Additionally, for all test examples, each stand-alone metamodel does not perform perfectly.

5.1.2. The Root Mean Square Error (RMSE). Next, the RMSEs of different surrogates for all test examples are compared. As shown in Figure 5, we can note that no single surrogate performs the best for all problems, and the root mean square error for individual surrogates varied with the experimental design. In addition, all ensemble models work better than the worst individual surrogate, and RMSE for ensemble models does not vary with DoE significantly, which indicates the necessity of adopting the ensemble techniques. At the same time, we can also find that our proposed method has better performance than other ensemble models in RMSE.

Numerical quantification of the RMSE is given in Table 4. The results in Table 4 show that for the test

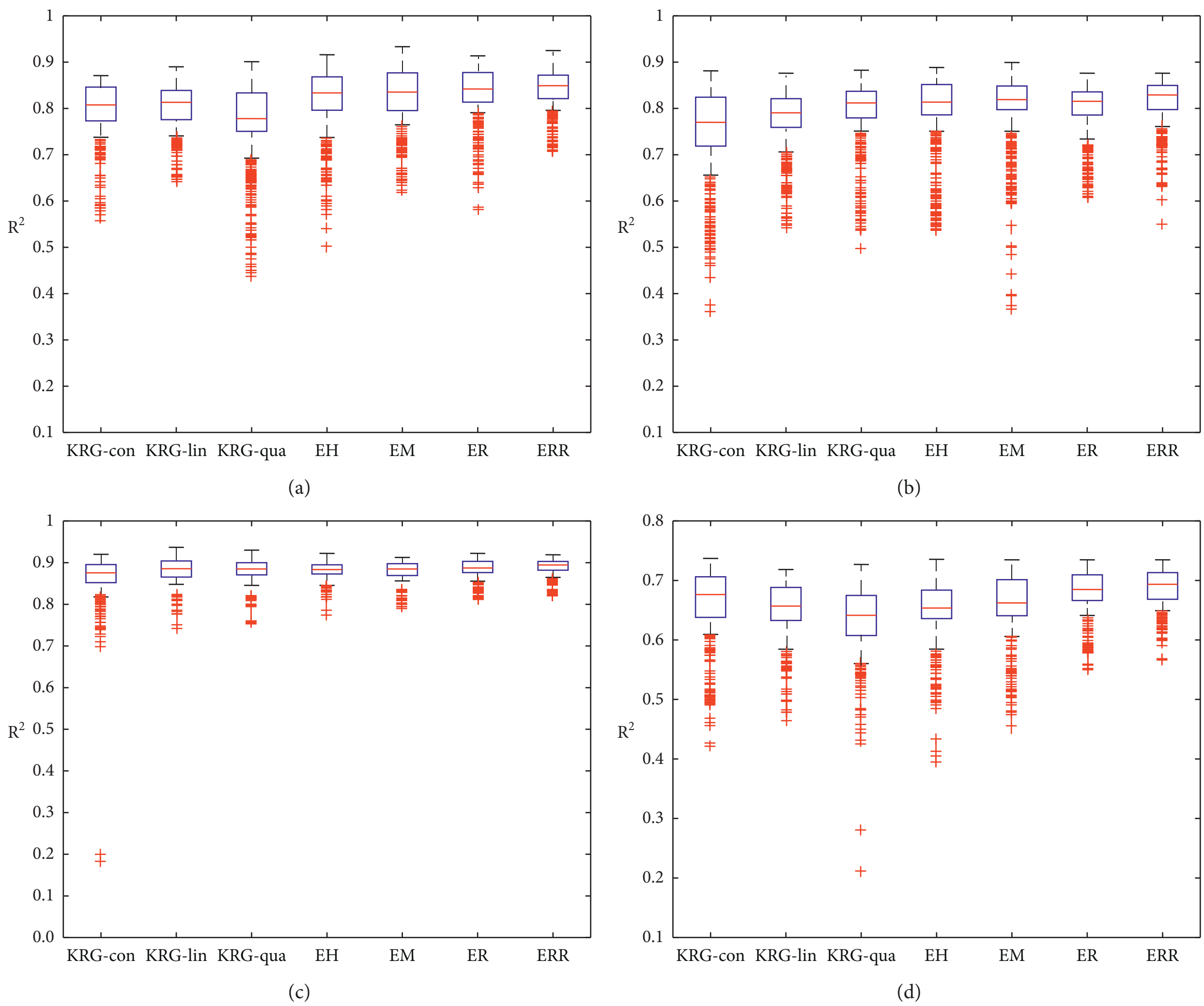

Figure 4: Coefficients of determination for different surrogate models. (a) Numerical Example 1. (b) Numerical Example 2. (c) Numerical Example 3. (d) Numerical Example 4.

Table 3: Mean and ψ value of R^2 using different metamodeling models for all test examples.

	Numerical Example 1	Numerical Example 2	Numerical Example 3	Stiffened plate element problems
krg-con	0.8111 (0.0580)	0.7725 (0.0849)	0.8756 (0.0308)	0.6660 (0.0737)
kg-lin	0.8238 (0.0882)	0.7917 (0.0601)	0.8846 (0.0269)	0.6645 (0.0990)
krg-qua	0.7936 (0.0777)	0.8101 (0.0495)	0.8848 (0.0258)	0.6404 (0.0687)
EH	0.8269 (0.0629)	0.8157 (0.0562)	0.8820 (0.0243)	0.6563 (0.0543)
EM	0.8298 (0.0515)	0.8104 (0.0495)	**0.8865 (0.0222)**	0.6807 (0.0683)
ER	0.8476 (0.0296)	0.8128 (0.0443)	0.8864 (0.0226)	0.6861 (0.0462)
ERR	**0.8602 (0.0303)**	**0.8194 (0.0426)**	0.8930 (0.0191)	**0.6926 (0.0454)**

problems 1, 2, and 4, ERR has the lowest RMSE errors compared to other surrogates.

Although the mean of RMSE for ERR in test problem 3 is gently larger than EM, ERR has a lower ψ, which indicates that ERR is more robust than EM. We can also observe that EH, EM, and ER can significantly reduce the errors compared to the worst individual surrogate, which suggests that using an ensemble of surrogate models can prevent us from making wrong choices of a surrogate.

5.2. Reliability Analysis Results. In order to compare the performance of different surrogates, after high-quality approximate models of the limit state equations for the four cases are obtained by using the above-mentioned metamodeling techniques, the Monte Carlo (MCS) method is employed to perform the reliability analysis. The results are provided in Table 5 including the time of surrogate construction (T_{con}), the estimation of failure probability (P_{f}), and the relative error (ΔP_{f}) compared with MCS.

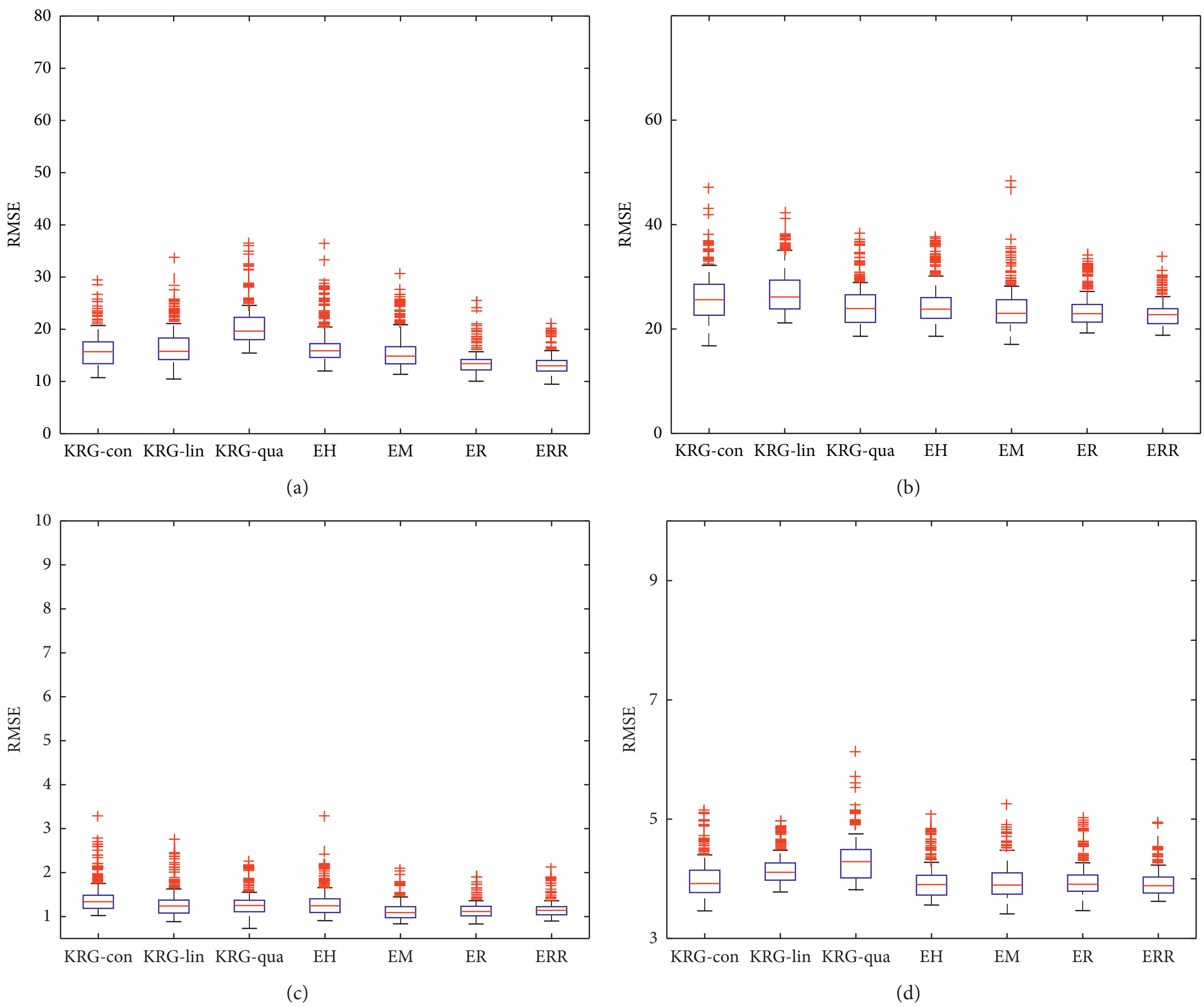

FIGURE 5: RMSE for different surrogate models. (a) Numerical Example 1. (b) Numerical Example 2. (c) Numerical Example 3. (d) Numerical Example 4.

TABLE 4: The mean and the coefficient of variation (ψ) of RMSE errors using different metamodels for different problems.

	Numerical Example 1	Numerical Example 2	Numerical Example 3	Stiffened plate element problems
Krg-con	16.0254 (0.1712)	26.0915 (0.1553)	1.3606 (0.1816)	3.9573 (0.0913)
Krg-lin	16.0601 (0.2042)	27.0418 (0.1246)	1.2540 (0.1715)	4.1229 (0.0514)
krg-qua	19.9411 (0.1450)	23.7087 (0.1366)	1.2622 (0.1802)	4.3224 (0.0830)
EH	15.8835 (0.1489)	23.6902 (0.1237)	1.2511 (0.1451)	3.9136 (0.0508)
EM	15.0233 (0.1563)	23.4337 (0.1209)	**1.1114 (0.1715)**	3.9190 (0.0804)
ER	13.5372 (0.1513)	23.1461 (0.1175)	1.1496 (0.1542)	3.9224 (0.0559)
ERR	**12.9343 (0.1177)**	**22.4847 (0.0878)**	1.1159 (0.12590)	**3.8767 (0.0546)**

For all the examples, from Table 5, we can see the following: an individual surrogate overestimates the failure probabilities (P_f), an individual surrogate underestimates the failure probabilities (P_f), or the relative error (ΔP_f) is the largest. Combined with Tables 3 and 4, we observe that no single surrogate model performs well in terms of either prediction accuracy or reliability evaluation.

In contrast, the ensembles of surrogates have considerably better performance than the individual surrogates from both the prediction accuracy and the failure probability obtained. For the same surrogate, the ER and ERR converge to a more accurate result than the EH and EM. However, ERR appears to be more effective in assessing the failure probability with good accuracy. Except for Example 3, the failure probability obtained by ERR is the closest to that of MCS and the relative error is minimal.

Table 5 also shows the time consumption of all the ensembles and individual surrogates. From the table, we can see that the construction time of individual surrogate is shorter than that of all ensembles of surrogates. Of all

Table 5: Results of the reliability analysis of all numerical examples.

Method	Example 1			Example 2			Example 3			Example 4		
	T_{con} (s)	P_f (10^{-3})	ΔP_f (%)	T_{con} (s)	P_f (10^{-2})	ΔP_f (%)	T_{con} (s)	P_f (10^{-3})	ΔP_f (%)	T_{con} (s)	P_f (10^{-5})	ΔP_f (%)
MCS	—	5.681	—	—	*7.33*	—	—	*1.581*	—	—	*1.910*	—
krg-con	35.261	5.576	1.85	1087.3	*7.711*	*5.20*	245.6	*1.660*	*4.99*	3215.6	*1.837*	*3.82*
kg-lin	39.432	5.983	5.32	1106.5	*7.657*	*4.46*	201.3	*1.614*	*2.08*	3456.8	*2.043*	*6.96*
krg-qua	45.583	5.902	3.89	1035.2	*7.485*	*2.11*	230.5	*1.643*	*3.91*	3279.4	*2.022*	*5.86*
EH	52.146	5.873	3.38	1288.55	*7.591*	*3.56*	289.1	*1.615*	*2.14*	3726.3	*2.028*	*6.18*
EM	61.531	5.842	2.83	1312.3	*7.614*	*3.87*	353.6	*1.620*	*2.46*	4533.2	*2.016*	*5.55*
ER	80.152	5.824	2.52	1401.8	*7.563*	*3.18*	396.7	*1.630*	*3.09*	5284.5	*2.01*	*5.24*
ERR	75.573	5.791	1.94	1369.4	*7.502*	*2.35*	369.2	*1.621*	*2.59*	4216.1	*1.832*	*4.08*

ensembles of surrogates, EH and EM are constructed in a shorter time, because they are both built by the method of obtaining weight factors at a time. Due to the concept of weighted mean square error proposed and the use of the heuristic weight scheme to determine the initial weight coefficients, ERR takes less time to build than ER.

6. Conclusion

Surrogate models have been widely adopted for reliability analysis. Traditionally, the researchers tend to select the so-called most accurate surrogate model as an approximation of the time-consuming limit state function by assessing its error metrics. However, the choice of the surrogate model relies on specific problems and generally there is no prior information for identifying a suitable surrogate and it is always man-made. Therefore, in order to save the cost of constructing the surrogate and improve the prediction accuracy, an ensemble strategy is proposed in this paper. The weight coefficients are obtained by a recursive process and the leave-one-out technique. In each iteration, weight factors are constantly updated until a given prediction accuracy is achieved. Besides, a learning function is used to guide the next sampling candidate selection by using the uncertain estimate of the ensemble of surrogates. In order to accurately evaluate the uncertainty for ensemble of surrogates, the concept of weighted mean square error is proposed. After a high-quality ensemble of surrogates of the limit state is available, the Monte Carlo method is used for reliability analysis. The effectiveness of the proposed method is validated by three analytical functions and a stiffened plate element problem, which requires high fidelity simulation of complex models with nonlinear response. Meanwhile, the proposed method is compared with the previous surrogate modeling strategy in prediction precision and reliability analysis result.

The results show that no single surrogate model performs well for all problems, in terms of both prediction accuracy and reliability assessment. In addition, due to the lack of sufficient information describing the relationship between response and input variables, it is difficult for researchers to know which metamodel is the best for a specific problem using traditional metamodeling strategies. However, our comparative study demonstrates that using an ensemble of surrogate models can provide more ideal prediction accuracy and higher robustness, which can effectively eliminate the negative impact brought by inappropriate stand-alone metamodel. Of all ensembles of surrogates, although the construction time of EH and EM is short, the prediction accuracy is low. Compared with ER, our proposed method has higher accuracy and efficiency due to the active learning function based on weighted mean square error and the heuristic weight scheme to determine the initial weight coefficients.

Data Availability

All data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the Natural Science Foundation of China "Research on reliability theory and method of total fatigue life for large complex mechanical structures" (Grant no. U1708255).

References

[1] T. Goel, R. T. Haftka, W. Shyy, and N. V. Queipo, "Ensemble of surrogates," *Structural and Multidisciplinary Optimization*, vol. 33, no. 3, pp. 199–216, 2007.

[2] B. Richard, C. Cremona, and L. Adelaide, "A response surface method based on support vector machines trained with an adaptive experimental design," *Structural Safety*, vol. 39, pp. 14–21, 2012.

[3] P. K. Das and Y. Zheng, "Cumulative formation of response surface and its use in reliability analysis," *Probabilistic Engineering Mechanics*, vol. 15, no. 4, pp. 309–315, 2000.

[4] M. Papadrakakis and N. Lagaros, "Reliability-based structural optimization using neural networks and Monte Carlo simulation," *Computer Methods in Applied Mechanics and Engineering*, vol. 191, no. 32, pp. 3491–3507, 2001.

[5] J.-M. Bourinet, F. Deheeger, and M. Lemaire, "Assessing small failure probabilities by combined subset simulation and support vector machines," *Structural Safety*, vol. 33, no. 6, pp. 343–353, 2011.

[6] I. Kaymaz, "Application of kriging method to structural reliability problems," *Structural Safety*, vol. 27, no. 2, pp. 133–151, 2005.

[7] P. Wang, Z. Lu, and Z. Tang, "An application of the Kriging method in global sensitivity analysis with parameter uncertainty," *Applied Mathematical Modelling*, vol. 37, no. 9, pp. 6543–6555, 2013.

[8] S. Sakata, F. Ashida, and M. Zako, "Structural optimization using Kriging approximation," *Computer Methods in Applied Mechanics and Engineering*, vol. 192, no. 7-8, pp. 923–939, 2003.

[9] D. Busby, "Hierarchical adaptive experimental design for Gaussian process emulators," *Reliability Engineering and System Safety*, vol. 94, no. 7, pp. 1183–1193, 2009.

[10] V. Picheny, D. Ginsbourger, O. Roustant et al., "Adaptive Designs of experiments for accurate approximation of a target region," *Journal of Mechanical Design*, vol. 132, no. 7, pp. 461–471, 2010.

[11] B. J. Bichon, M. S. Eldred, L. P. Swiler, S. Mahadevan, and J. M. McFarland, "Efficient global reliability analysis for nonlinear implicit performance functions," *AIAA Journal*, vol. 46, no. 10, pp. 2459–2468, 2008.

[12] B. Echard, N. Gayton, and M. Lemaire, "AK-MCS: an active learning reliability method combining Kriging and Monte Carlo Simulation," *Structural Safety*, vol. 33, no. 2, pp. 145–154, 2011.

[13] Z. Lv, Z. Lu, and P. Wang, "A new learning function for Kriging and its applications to solve reliability problems in engineering," *Computers and Mathematics with Applications*, vol. 70, no. 5, pp. 1182–1197, 2015.

[14] M. P. Perrone and L. N. Cooper, "When networks disagree: ensemble methods for hybrid neural networks," in *Artificial Neural Networks for Speech and Vision*, R. J. Mammone, Ed., pp. 126–142, Chapman and Hall, London, UK, 1993.

[15] E. Acar and M. Rais-Rohani, "Ensemble of metamodels with optimized weight factors," *Structural and Multidisciplinary Optimization*, vol. 37, no. 3, pp. 279–294, 2009.

[16] X. Zhou, Y. Ma, and X. LI, "Ensemble of surrogates with recursive arithmetic average," *Structural and Multidisciplinary Optimization*, vol. 44, no. 5, pp. 651–671, 2011.

[17] F. A. C. Viana, R. T. Haftka, and V. Steffen, "Multiple surrogates: how cross-validation errors can help us to obtain the best predictor," *Structural and Multidisciplinary Optimization*, vol. 39, no. 4, pp. 439–457, 2009.

[18] M. Balesdent, J. Morio, and J. Marzat, "Kriging-based adaptive importance sampling algorithms for rare event estimation," *Structural Safety*, vol. 44, pp. 1–10, 2013.

[19] R. Rackwitz, "Reliability analysis—A review and some perspectives," *Structural Safety*, vol. 23, no. 4, pp. 365–395, 2001.

[20] B. Gaspar, A. Naess, B. J. Leira, and C. Guedes Soares, "System reliability analysis by Monte Carlo based method and finite element structural models," *Journal of Offshore Mechanics and Arctic Engineering*, vol. 136, no. 3, 2014.

[21] C. Bishop, *Neural Networks for Pattern Recognition*, Oxford University Press, New York, NY, USA, 1995.

[22] L. E. Zerpa, N. V. Queipo, S. Pintos, and J.-L. Salager, "An optimization methodology of alkaline-surfactant-polymer flooding processes using field scale numerical simulation and multiple surrogates," *Journal of Petroleum Science and Engineering*, vol. 47, no. 3-4, pp. 197–208, 2005.