Xueguan Song[1]
School of Mechanical Engineering,
Dalian University of Technology,
No. 2 Linggong Road,
Ganjingzi District,
Dalian 116024, China
e-mail: sxg@dlut.edu.cn

Liye Lv
School of Mechanical Engineering,
Dalian University of Technology,
No. 2 Linggong Road,
Ganjingzi District,
Dalian 116024, China
e-mail: lvhexiaoye@mail.dlut.edu.cn

Jieling Li
School of Mechanical Engineering,
Dalian University of Technology,
No. 2 Linggong Road,
Ganjingzi District,
Dalian 116024, China
e-mail: 349783872@mail.dlut.edu.cn

Wei Sun
School of Mechanical Engineering,
Dalian University of Technology,
No. 2 Linggong Road,
Ganjingzi District,
Dalian 116024, China
e-mail: sunwei@dlut.edu.cn

Jie Zhang
Department of Mechanical Engineering,
The University of Texas at Dallas,
Richardson, TX 75080
e-mail: jiezhang@utdallas.edu

An Advanced and Robust Ensemble Surrogate Model: Extended Adaptive Hybrid Functions

Hybrid or ensemble surrogate models developed in recent years have shown a better accuracy compared to individual surrogate models. However, it is still challenging for hybrid surrogate models to always meet the accuracy, robustness, and efficiency requirements for many specific problems. In this paper, an advanced hybrid surrogate model, namely, extended adaptive hybrid functions (E-AHF), is developed, which consists of two major components. The first part automatically filters out the poorly performing individual models and remains the appropriate ones based on the leave-one-out (LOO) cross-validation (CV) error. The second part calculates the adaptive weight factors for each individual surrogate model based on the baseline model and the estimated mean square error in a Gaussian process prediction. A large set of numerical experiments consisting of up to 40 test problems from one dimension to 16 dimensions are used to verify the accuracy and robustness of the proposed model. The results show that both the accuracy and the robustness of E-AHF have been remarkably improved compared with the individual surrogate models and multiple benchmark hybrid surrogate models. The computational time of E-AHF has also been considerately reduced compared with other hybrid models. [DOI: 10.1115/1.4039128]

Keywords: hybrid surrogate model, adaptive weight factor, model selection, Gaussian-process error, robustness

1 Introduction

Computational simulation techniques have been making a notable progress in presenting the true physics of phenomena, and thus are playing an important role in the design and optimization of complex engineering systems [1–4]. Recently, the needs for high-fidelity computational simulations have been growing dramatically in various engineering applications due to their high level of accuracy [5]. On the other hand, the expensive computational cost of high-fidelity simulations is still prohibitive in today's competitive and demanding market [6]. To address this challenge, surrogate models (also called metamodels) have been widely used as substitutes for computationally expensive simulation models [7]. Surrogate models are used to represent computationally expensive simulation models or experiments by fitting the relationship between the system inputs and outputs of interest based on the limited sampling data. Surrogate models have been significantly improved over the past two decades, and many kinds of surrogate models such as polynomial response surface (PRS) [8], Kriging (KRG) [9–11], radial basis function (RBF) [12], and support vector regression (SVR) [13,14], have been developed and successfully applied in many structure and/or multidisciplinary design optimization problems.

However, the rapid development of various surrogate models does not only provide researchers flexibility in surrogate models selection for different problems, but also brings a challenge in model selection for specific applications [15]. It has been proved that no single surrogate model always performs the best for all engineering practice [16]. This is because that the actual engineering application often presents different linear or nonlinear characteristics, and each surrogate model has its own advantages and disadvantages [17–19]. For example, the relationship between the inputs and outputs in a high-dimensional problem may be linear, so a simple first-order PRS model might be sufficient and better for the fitting [20]. The relationship in a low-dimensional problem could be highly nonlinear, so that an RBF model may be suitable for capturing the trend [21]. Owing to the complexity of practical engineering applications and the lack of sufficient information, it is extremely challenging to choose the most appropriate surrogate model prior to optimization. To this end, an ensemble of surrogate models (or hybrid surrogate model) has been developed, which seeks to make use of advantages of each individual surrogate, as well as to eliminate the effort of selecting the appropriate individual [16,22].

A number of hybrid surrogate models have been developed in the literature and most of them are based on error correlation or prediction variance. For example, Acar and Rais-Rohani [23] developed a strategy of building a combination of five different individual surrogate models by optimizing the weight factors to minimize a selected error metric. Ferreira and Serpa [24] proposed a hybrid surrogate model based on the augmented least squares that can calculate the ensemble weights by minimizing the effects

[1]Corresponding author.
Contributed by the Design Automation Committee of ASME for publication in the Journal of Mechanical Design. Manuscript received February 19, 2017; final manuscript received January 5, 2018; published online February 27, 2018. Assoc. Editor: Christina Bloebaum.

of nonlinearity inherent. Viana et al. [25] utilized the cross-validation (CV) error for weights calculation. In these studies, hybrid surrogate models have been shown able to provide a better accuracy than individual surrogate models. However, it is also observed that the ensemble of surrogate models does not always guarantee a better solution than the best selected individual, and the potential gains diminish especially for high-dimensional problems [24,25]. In addition, the weight factors in these hybrid models are constant over the whole design space. The calculation of weight factors is thus relatively straightforward and computationally inexpensive. However, these globally fixed coefficients may not reflect the local accuracy of each individual surrogate, thereby resulting in inaccurate predictions in specific local regions. An adaptive hybrid surrogate model is expected to overcome this shortcoming.

Recently, many researchers have attempted to develop hybrid surrogate models that can capture both the global and local predicting accuracy [16]. For example, Acar [26] adopted the pointwise CV error as a local error measure for constructing a hybrid model. Zhang et al. [27] developed an adaptive hybrid functions (AHF) model that used a self-defined distance-based trust region to define the weight factors of each individual. Liu et al. [28] developed an optimal weighted pointwise ensemble to combine the locally accurate predictions of RBF models with different basis functions together. In these studies, adaptive hybrid surrogate models generally perform better than those hybrid models with constant weight factors. However, adaptive hybrid models also have their shortcomings. First, most of adaptive hybrid models demand auxiliary optimization to search for the weight factors, which requires additional computational cost as well, no matter whether the optimization is a one-dimensional (1D) or n-dimensional search [26,27]. In most cases, the cost of constructing or running a surrogate model is negligible compared to the cost of computationally expensive simulations [29]. However, there still exist many practical problems that require a large number of surrogate model evaluations. For example, a global sensitivity analysis may require more than 100,000 calculations to get converged. If each surrogate evaluation takes more than 10 s without parallel computing, the computational time of each global sensitivity analysis may exceed 10 days. Second, most of the above discussed hybrid models mainly focus on low-dimensional problems rather than high-dimensional problems. Therefore, an adaptive hybrid model that is suitable for both low-dimensional and high-dimensional problems, with both high accuracy and low computational time, is still desired for engineering systems design.

Motivated by this analysis, this paper develops an advanced hybrid surrogate model, namely, extended adaptive hybrid functions (E-AHF), which is constructed with adaptive weights based on the Gaussian process estimated prediction error and a redundant model eliminating strategy. The E-AHF surrogate aims to take advantage of the diversity of well-performing individual surrogate models to guarantee the prediction accuracy and robustness for various problems from low to high dimensions.

The remainder of the paper is organized as follows: Section 2 presents the development of the E-AHF surrogate model, followed by the settings of numerical experiments in Sec. 3. The validation of the proposed E-AHF hybrid model on a large set of test problems is given in Sec. 4. Concluding remarks and perspectives are summarized in Sec. 5.

2 Extended Adaptive Hybrid Functions

A hybrid surrogate model intends to approximate a response of interest by taking full advantage of the prediction ability of each individual surrogate model. Members of the hybrid model are multiplied by weight factors in the form of a weighted-sum combination

$$\hat{y}(\mathbf{x}) = \sum_{i=1}^{m} \omega_i \hat{y}_i(\mathbf{x}) \tag{1}$$

where $\hat{y}(\mathbf{x})$ is the predicted response by the hybrid model, m is the number of surrogate models in the combination, ω_i is the weight factor associated with the ith individual surrogate model, and $\hat{y}_i(\mathbf{x})$ is the response estimated by the ith individual surrogate model at the input vector $\mathbf{x}$. Furthermore, the sum of weight factors is constant and equals one, so that if all the individual surrogates produce the same output at certain points, then the weighted surrogate will also reflect this output [30]

$$\sum_{i=1}^{m} \omega_i = 1 \tag{2}$$

As mentioned earlier, the accuracy of weights selection plays an important role in the accuracy of the hybrid surrogate model. One of the traditional methods is to define the weight factors in terms of the global performance, so the weights remain constant throughout the design space [16,23,25]. However, the accuracy of each surrogate model may vary significantly in the design space. Recently, researchers have been developing hybrid surrogate models with adaptive weight factors that depend on both their global and local performance, aiming to further improve the overall accuracy of the hybrid surrogate model [26,27,31].

Zhang et al. [27] recently developed the AHF surrogate model. The AHF formulates a trust region based on the density of available sample points, and adaptively combines characteristically differing surrogate models. The weight of each contributing surrogate model is represented as a function of the input domain based on a local measure of accuracy for that surrogate model. Such an approach exploits the advantages of each component surrogate, thereby, capturing both the global and the local trends of complex functional relationships. Figure 1 [27] shows the overall framework of the AHF surrogate model.

By extending the AHF model [27], this paper proposes a novel strategy to develop a new hybrid surrogate model, namely, E-AHF, by considering both the global and local accuracies of individual surrogate models. The construction of the E-AHF model can be outlined by the following two parts:

- Part (I)—Selection of individual surrogate models: A group of widely used individual surrogate models are compared based on the same sampling data set and a global accuracy metric, so that a subgroup of models with relatively high accuracies can be selected.
- Part (II)—Weights calculation and model ensemble: An adaptive weights calculation strategy is developed to obtain the weights for each individual surrogate model, and E-AHF is constructed based on these adaptive weight factors.

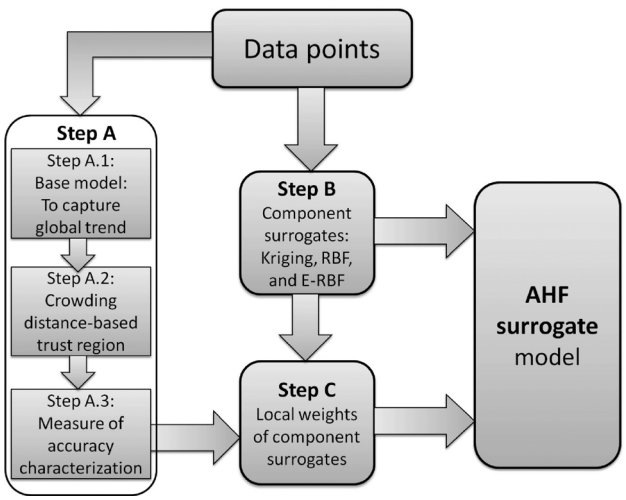


Fig. 1 The framework of the AHF surrogate model [27]

2.1 Selection of Individual Surrogate Models. It has been proved that introducing a poorly performing individual surrogate into the ensemble may significantly reduce the average prediction accuracy [15]. Thus, instead of randomly determining a group of individual component surrogates, a filtering process could be performed first to exclude the poorly performing individual surrogates. To avoid adding extra sampling data, the leave-one-out (LOO) CV is adopted as the global accuracy measure to assess the performance of each individual. The LOO CV error of each individual surrogate model on the training points is calculated as

$$\text{CVerror}_i = \frac{1}{n-1}\sum_{j=1}^{n-1}(y_j - \hat{y}_{ij}^{-j})^2, \quad i = 1, 2, \ldots, m \tag{3}$$

where CVerror_i is the CV error of the ith individual surrogate model, y_j is the true response at the jth sample point, and $\hat{y}_{ij}^{-j}$ stands for the prediction of the ith surrogate model at the jth sample point, which is calculated using the $n-1$ sample points except the jth sample point. m denotes the number of candidate surrogate models, and n represents the number of sample points. To compare the performance and filter out the individual surrogates with large errors, a normalized CV error (NCVerror) is calculated for each individual surrogate model, which is given as

$$\text{NCVerror}_i = \frac{\text{CVerror}_i - \text{CVerror}_{\min}}{\text{CVerror}_{\max} - \text{CVerror}_{\min}} \tag{4}$$

where $\text{CVerror}_{\min}$ and $\text{CVerror}_{\max}$ are the minimum CV error and the maximum CV error, respectively. A smaller NCVerror indicates a better surrogate model. Then, a threshold value β ranging from 0 to 1 is defined. The individual surrogate models with NCVerrors smaller than β are selected from the initial set to form a new subset for the hybrid model construction. In addition, the surrogate model with the smallest NCVerror is selected as the baseline model for the follow-on adaptive weight factors calculation, which will be explained in Sec. 2.2.

2.2 Calculation of Adaptive Weight Factors. An adaptive hybrid surrogate model is expected to capture the local performance of each surrogate model. Therefore, in addition to selecting the baseline model, we also need to characterize the local performance of each individual surrogates. To this end, the Gaussian process based models are adopted to characterize the local variance, thereby calculating prediction errors for the component surrogate models. The process of calculating the adaptive weight factors is described in the following three steps:

Step 1. Local measure estimation

Calculate the estimated mean squared error by a Gaussian-process based prediction:

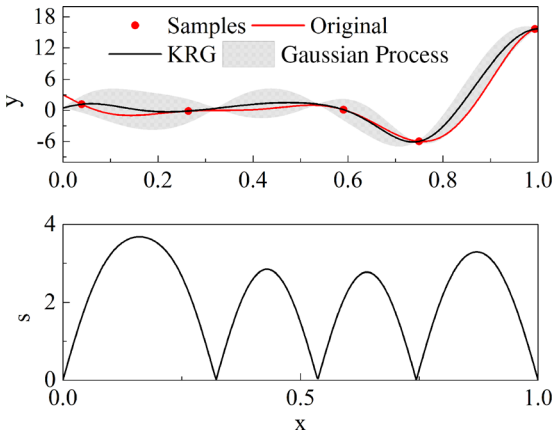


Fig. 3 Error region and values from the Gaussian process

$$\hat{s}^2(x) = \sigma^2\left[1 - \boldsymbol{\psi}^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\boldsymbol{\psi} + \frac{\mathbf{1} - \mathbf{1}^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\boldsymbol{\psi}}{\mathbf{1}^{\mathrm{T}}\boldsymbol{\Psi}^{-1}\mathbf{1}}\right] \tag{5}$$

where σ denotes the constant process variance of a Gaussian field. $\boldsymbol{\Psi}$ is the correlation matrix with the elements, which can be expressed as

$$\Psi_{jk} = \text{cor}[y_j, y_k], \quad j = 1, \ldots, n \text{ and } k = 1, \ldots, n \tag{6}$$

And $\boldsymbol{\psi}$ is a vector with correlations between the observed data and the new prediction, given by

$$\psi = \{\text{cor}[y_1, y(\mathbf{x})], \ldots, \text{cor}[y_n, y(\mathbf{x})]\}^{\mathrm{T}} \tag{7}$$

where $y(\mathbf{x})$ is the prediction at the test point.

Step 2. Probability estimation

The baseline model can represent the global trend of the hybrid surrogate model, due to its high accuracy across the entire design space. Therefore, this prediction by the baseline model can be deemed as an expected value of the hybrid model. With the local prediction estimation s^2 obtained from the previous local measure estimation, the probability coefficient of each individual surrogate model can be calculated by

$$P_i = \exp\left\{-\frac{[y_{ij} - y_{\text{base}j}]^2}{2s_j^2}\right\} \tag{8}$$

where $y_{\text{base}j}$ is the prediction value of the baseline model at the jth sample point, P_i is the probability coefficient associated with the

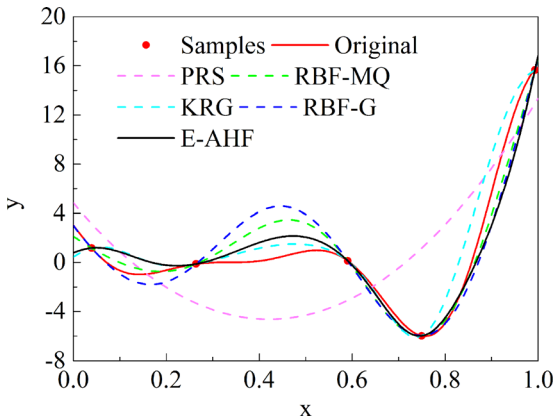


Fig. 2 Different surrogate models based on five training points

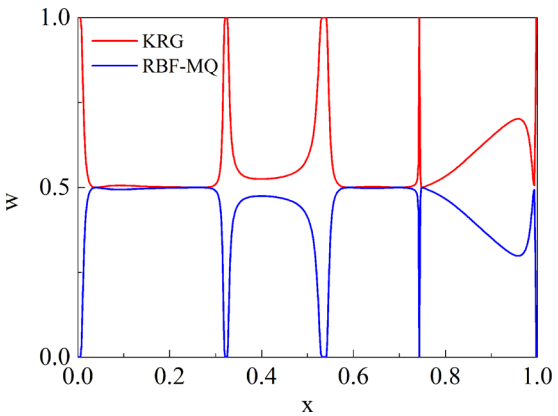


Fig. 4 Adaptive weight factors for different individual models

Table 1 Test functions

No.	D	Test function	S
1	1	$y = (6x-2)^2 \sin[2(6x-2)]$	$[0, 1]^D$
2	1	$y = \frac{\sin(10\pi x)}{2x} + (x-1)^4$	$[0.5, 2.5]^D$
3	1	$y = 6 - \frac{1}{(x-0.3)^2 + 0.01} - \frac{1}{(x-9)^2 + 0.04}$	$[0,1]^D$
4	2	$y = \left(x_2 - \frac{5.1x_1^2}{4\pi^2} + \frac{5x_1}{\pi} - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos(x_1) + 10$	$[-5, 0;10, 15]$
5	2	$y = 10^5 x_1^2 + x_2^2 - (x_1^2 + x_2^2)^2 + 10^5(x_1^2 + x_2^2)^4$	$[-20, 20]^D$
6	2	$y = [1 - 2x_2 + 0.05\sin(4\pi x_2 - x_1)]^2 + [x_2 - 0.5\sin(2\pi x_1)]^2$	$[-5,10]^D$
7	2	$y = [1 - 2x_2 + 0.05\sin(4\pi x_2 - x_1)]^2 + [x_2 - 0.5\sin(2\pi x_1)]^2$	$[-10,10]^D$
8	2	$y = (x_2 - 1.275x_1/\pi)^2 + 5(x_1/\pi - 6)^2 + 10(1 - 0.125/\pi)\cos(x_1)$	$[-5, 0;10, 15]$
9	2	$y = [10 + x_1\sin(x_1)][4 + \exp((-x_2)^2)]$	$[-2, 2]^D$
10	2	$y = x_1 e^{(-x_1^2 - x_2^2)}$	$[-2, 2]^D$
11	2	$y = 2 + 0.01(x_2 - x_1^2)^2 + (1 - x_1)^2 + 2(2 - x_2)^2 + 7\sin(0.5x_1)\sin(0.7x_1x_2)$	$[0, 5]^D$
12	2	$y = 2x_1^2 - 1.05x_1^4 + x_1^6/6 + x_1x_2 + x_2^2$	$[-5,5]^D$
13	2	$y = \left(4 - 2.1x_1^2 + \frac{1}{3}x_1^4\right)x_1^2 + x_1x_2 + (-4 + 4x_2^2)x_2^2$	$[-3\ -2;3\ 2]$
14	2	$y = \left(4 - 2.1x_1^2 + \frac{1}{3}x_1^4\right)x_1^2 + x_1x_2 + (-4 + 4x_2^2)x_2^2$	$[-2\ -1;2\ 1]$
15	2	$y = [1 + (x_1 + x_2 + 1)^2(19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2)] \times [30 + (2x_1 - 3x_2)^2 \times (18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2)]$	$[-2, 2]^D$
16	2	$y = 3(1 - x_1)^2\exp(-x_1^2 - (x_2 + 1)^2) - 10\left(\frac{x_1}{5} - x_1^3 - x_2^5\right)\exp((-x_1^2 - x_2^2)) - \frac{1}{3}\exp((-(x_1 + 1)^2 - x_2^2))$	$[-4,4]^D$
17	2	$y = \sin(x_1 + x_2) + (x_1 - x_2)^2 - 1.5x_1 + 2.5x_2 + 1$	$[0,1]^D$
18	3	$y = -\sum_{i=1}^{4} \alpha_i \exp\left(-\sum_{j=1}^{3} A_{ij}(x_j - P_{ij})^2\right)$	$[0, 1]^D$
19	3	$y = -x_1x_2x_3$	[0 0 0;20 11 42]
20	4	$y = 100(x_1^2 - x_2)^2 + (x_1 - 1)^2 + 90(x_3^2 - x_4)^2 + 10.1[(x_2 - 1)^2 + (x_4 - 1)^2] + 19.8(x_2 - 1)(x_4 - 1)$	$[-10, 10]^D$
21	4	$y = 100(x_1 - x_2)^2 + (x_1 - 1)^2 + (x_3 - 1)^2 + 90(x_3^2 - x_4) + 10.1[(x_2 - 1)^2 + (x_4 - 1)^2] + 19.8(x_2 - 1)(x_4 - 1)$	$[-10, 10]^D$
22	4	$y = 10\sin[2(x_1 - 0.6\pi)] + x_2 + x_3 + x_4 + x_1x_2 + x_3x_4 + x_1^3 + x_4^3$	$[0, 1]^D$
23	4	$y = \frac{1}{0.839}\left[1.1 - \sum_{i=1}^{4} \alpha_i \exp\left(-\sum_{j=1}^{4} A_{ij}(x_j - P_{ij})^2\right)\right]$	$[0, 1]^D$
24	4	$y = \sum_{i=1}^{d}\left[\left(\sum_{j=1}^{d} x_j^i\right) - b_i\right]^2$	$[0, 4]^D$
25	4	$y = \sum_{i=1}^{d/4}[(x_{4i-3} + 10x_{4i-1})^2 + 5(x_{4i-1} - x_{4i})^2 + (x_{4i-2} - 2x_{4i-1})^4 + 10(x_{4i-3} + x_{4i})^4]$	$[-4, 5]^D$
26	4	$y = 1 + \exp\{-2[(x_1 - 1)^2 + x_2^2] - 0.5(x_3^2 + x_4^2)\} + \exp\{-2[x_1^2 + (x_2 - 1)^2] - 0.5(x_3^2 + x_4^2)\}$	$[0, 1]^D$
27	6	$y = -\sum_{i=1}^{4} \alpha_i \exp\left(-\sum_{j=1}^{6} A_{ij}(x_j - P_{ij})^2\right)$	$[0, 1]^D$
28	6	$y = 0.0204x_1^2x_4 + x_2x_3 + 0.01870.0204x_1x_2x_3 + 1.57x_2x_4 + 0.0607x_1x_4x_5^2(x_1 + x_2 + x_3) + 0.0437x_2x_3x_6^2(x_1 + 1.57x_2 + x_4)$	$[0, 10^5]^D$
29	8	Refer to HS105 in Ref. [33]	
30	8	$y = \sum_{i}^{8}\left[\frac{3}{10} + \sin\left(\frac{16}{15}x_i - 1\right) + \sin^2\left(\frac{16}{15}x_i - 1\right)\right]$	$[-1, 1]^D$
31	10	$y = \sum_{i}^{10}[(x_{i+1}^2 - x_i)^2 + (x_i - 1)^2]$	$[-3, 3]^D$
32	10	$y = \sum_{i}^{10} e^{x_i}\left[c_i + x_i - \log\left(\sum_{j}^{10} x_j\right)\right]$	$[-5, 5]^D$
33	10	$y = \prod_{i=1}^{10} x_i + \sum_{j=1}^{10}[\ln(x_j - 2)^2 + \ln(10 - x_j)^2]$	$[2.1,9.9]^D$
34	10	$y = \sum_{i}^{10} x_i^2 + \left(\sum_{j}^{10} 0.5jx_j\right)^2 + \left(\sum_{j}^{10} 0.5jx_j\right)^4$	$[-5, 10]^D$
35	10	$y = \sum_{i}^{10}\left[\frac{3}{10} + \sin\left(\frac{16}{15}x_i - 1\right) + \sin^2\left(\frac{16}{15}x_i - 1\right)\right]$	$[-1, 1]^D$
36	10	$y = x_1^2 + x_2^2 + x_1x_2 - 14x_1 - 16x_2 + (x_3 - 10)^2(x_4 - 5)^2 + (x_5 - 3)^2 + 2(x_6 - 1)^2 + 5x_7^2 + 2(x_9 - 10)^2 + 2(x_{10} - 7)^2 + 45$	$[0, 1]^D$
37	15	$y = (x_1 - 1)^2 + \sum_{i=2}^{15} i(2x_i^2 - x_{i-1})^2$	$[-10, 10]^D$
38	15	$y = \sum_{i=1}^{14} 100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2$	$[-5, 10]^D$
39	16	$y = (x_1 - 1)^2 + \sum_{i=2}^{16} i(2x_i^2 - x_{i-1})^2$	$[-5, 5]^D$
40	16	$y = \sum_{i}^{16}\sum_{j}^{16} a_{ij}(x_i^2 + x_i + 1)(x_j^2 + x_j + 1)$	$[-1, 1]^D$

ith individual surrogate model, and s_j^2 is the local prediction estimation at the jth sample point.

Step 3. Local weight determination

With the probability coefficient assigned to each surrogate model, the weight factors can be computed by

$$\omega_i = \frac{P_i}{\sum_{j=1}^{m} P_j} \tag{9}$$

As P is a function of the input vector $\mathbf{x}$, the weight factors obtained for each individual surrogate model is adaptive to the input vector as well.

2.3 Example of Extended Adaptive Hybrid Functions Modeling. To further clarify the process of building of the E-AHF model, a 1D benchmark problem as shown in Eq. (10) is examined below:

$$y = (6x-2)^2 \sin[2(6x-2)], \quad 0 \le x \le 1 \tag{10}$$

First, a group of surrogate models including PRS, KRG, RBF-multiquadric (RBF-MQ), and RBF-Gaussian (RBF-G) are considered in this case. PRS is a typical second-order polynomial model; KRG uses a zeroth-order polynomial function and Gaussian correlation function; both RBF-MQ and RBF-G use $c = 0.34$ in the basis function. Five training points are located at $x = [0.039, 0.260, 0.590, 0.750, 0.990]$ as shown in Fig. 2. Then, the process of building the hybrid model starts from Eq. (3), where the CV errors of PRS, KRG, RBF-MQ, and RBF-G are calculated to be 1177, 740, 1085, and 1121, respectively. The NCVerrors of the four surrogate models are thus 1, 0, 0.79, and 0.87 based on Eq. (4). Here, we set the threshold value β to be 0.8, so PRS and RBF-G are filtered out from the candidate pool due to their larger NCVerrors, KRG and RBF-MQ are kept for the ensemble, and KRG is selected as the baseline model due to its smallest NCVerror. Then, the estimated root-mean-square error (RMSE) from the Gaussian process is obtained from Eqs. (5)–(7). Figure 3 illustrates the trust region based on the baseline and the estimated RMSE. It is seen from Fig. 3 that the estimated (RMSE) (s value in the figure) is zero at the five training points, and increases with the increase of distance between the test point and the training points. According to Eq. (8), the probability coefficients P_i can be obtained for each surrogate model. Then, according to Eq. (9), the adaptive weight factors are obtained as shown in Fig. 4. Finally, the E-AHF surrogate model is constructed as shown in Fig. 2. The performance of the E-AHF model as well as the KRG and RBF-MQ models are compared by using 500 test points evenly distributed in the design space. It is found that the global accuracy defined by the coefficient of determination (R-square or R^2) for E-AHF, KRG, and RBF-MQ are 0.94, 0.91, and 0.91, respectively. Overall, the proposed E-AHF performs better than the two individual surrogate models.

3 Settings of Numerical Experiments

3.1 Test Problems. In order to thoroughly test the performance of the proposed hybrid surrogate model, up to 40 mathematical problems are used in this work. As shown in Table 1, these test problems are chosen from the literature [20,32–35], consisting of 30 low-dimensional (from 1D to 8D) and ten high-dimensional (from 10D to 16D) problems.

3.2 Design of Experiments. Design of experiments (DoE) is the strategy to generate sampling or training points for computer simulations and surrogate modeling. Among many available DoE methods, the latin hypercube sampling has been proved capable of balancing the trade-off between accuracy and robustness by generating a near-random set of sample points [36]. In this work, the MATLAB function *lhsdesign* is used to generate the sample points.

Both the number and distribution of the sample points from DoE have significant impacts on the performance of surrogate models. To thoroughly compare the performance of the proposed hybrid model and benchmark models, different numbers of sample points classified as small, medium, and large sets are tested in this work. Specifically, each set is defined as k times of n design variables: the small sets contain three sets of ($3n$, $5n$, $8n$), the medium sets contain three sets of ($10n$, $15n$, $20n$), and the large sets contain three sets of ($25n$, $30n$, $40n$). The size in each set is guaranteed to be larger than $(n+1)(n+2)/2$, which is the minimum number required by the quadratic PRS model.

To eliminate the effects of the randomness of sample points, a preliminary analysis on the effects of the sampling points distribution has been performed for test problems 19, 26, and 40. The E-AHF and KRG model are used, as KRG has been found to be the most sensitive model to sample points [20]. Figure 5 shows the accuracies of the E-AHF and KRG models under different sets of sample points, which are generated by *lhsdesign* (in MATLAB) with different criteria and iterations. For KRG, the mean R-square converges until the number of sampling sets reaches 90, 71, and 70 for test problems 19, 26, and 40, respectively, when the convergence criterion is set to be 0.0001 (i.e., the change of the mean R-square is less than 0.0001). For the E-AHF, the mean R-square converges until the number of sampling sets reaches 55, 64, and 50 for test problems 19, 26, and 40, respectively. As there are 40 test problems and ten types of surrogate models, it is computationally expensive to repeat 90 times of surrogate model construction for each test problem. Therefore, the number of sampling set of 20 is used. In terms of the accuracy, the 20 sampling sets may generate at most 0.25% deviation compared to the converged accuracy at 64 and 71 sampling sets for E-AHF and KRG models, respectively, and will not significantly affect the results of numerical experiments.

3.3 Performance Criteria. Two kinds of performance criteria are used for the comparison of surrogate models, namely, the global performance metrics such as RMSE and R-square, and the local performance metrics such as the relative maximum absolute error. As RMSE is highly correlated with R-square, and relative maximum absolute error cannot show the overall performance in the design space [20], R-square is selected as the only criterion for the following comparison:

$$R^2 = 1 - \frac{\sum_j (y_j - \hat{y}_j)^2}{\sum_j (y_j - \bar{y})^2} = 1 - \frac{\text{MSE}}{\text{variance}} \tag{11}$$

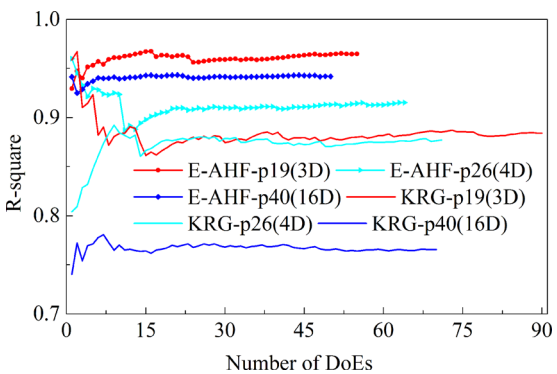
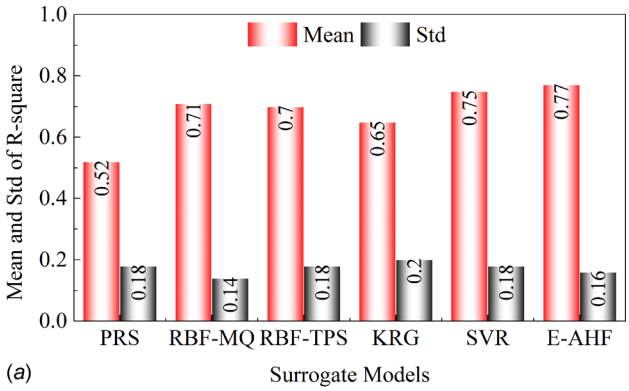
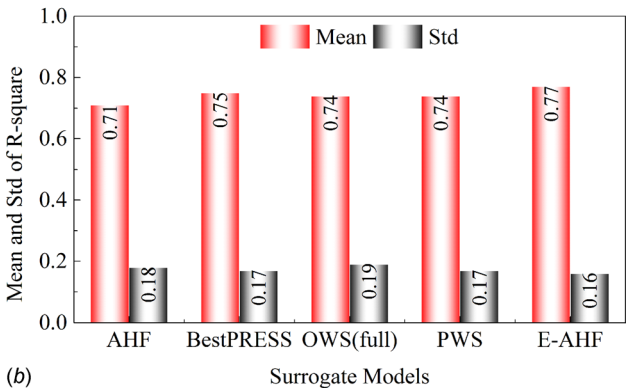


Fig. 5 Effect of DoE sets on the accuracy of E-AHF and Kriging models

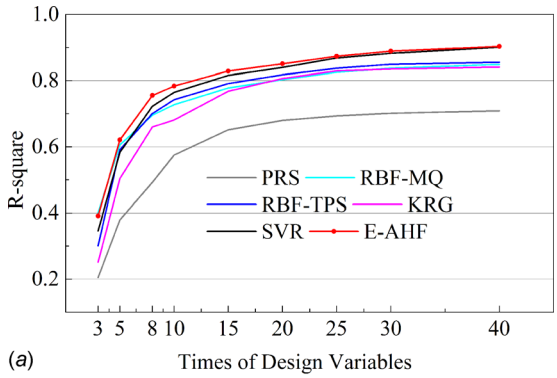


(*a*)

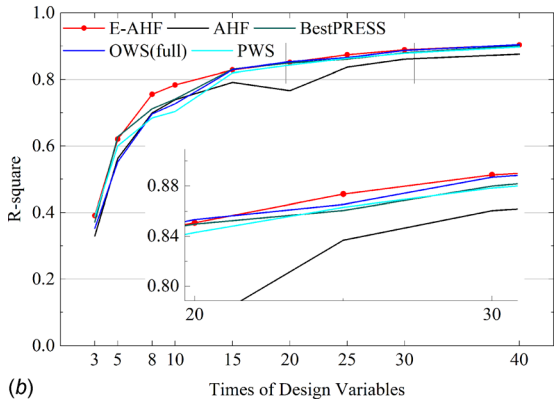


(*b*)

Fig. 6 R-square comparison of different surrogate models for 40 test problems: (*a*) compared with individual models and (*b*) compared with hybrid models



(*a*)



(*b*)

Fig. 7 R-square comparison under different sample sets: (*a*) compared with individual models and (*b*) compared with hybrid models

where y_j denotes the observed response for x_j, $\hat{y}_j$ denotes the corresponding prediction, t is the number of evaluation points, and $\bar{y}$ is the mean of the observed responses. R-square ranges from 0 to 1, and a larger value indicates a better accuracy over the whole design space. A total of 500n test points (n is the number of design variables) are used for R-square calculation for each surrogate model.

3.4 Surrogate Models. To take advantage of the diversity of different individual surrogate models, five widely used models (i.e., PRS, RBF-MQ, RBF-thin plate spline, KRG, and SVR) are initially selected for the hybrid model construction. The threshold value β is set to be 0.5 for the numerical experiments. To evaluate the performance of E-AHF, the five individual surrogate models and another four hybrid surrogate models including BestPRESS, optimal weighted surrogate using full C-matrix (OWS(full)), PRESS weighted surrogate (PWS), and AHF, are used as benchmark models. The PRS is a quadratic model that can generally capture the global trend of training points. The RBF model is constructed based on the multiquadric and thin plate spline basis functions due to their high performance [21]. The KRG model uses a zero-order polynomial function and Gaussian correlation function. The SVR model is constructed by using the Gaussian kernel function. Individual surrogate models and three hybrid models including BestPRESS, OWS(full), and PWS are implemented in the MATLAB toolbox developed by Viana [37], where SVR model is replaced by the LSSVR Toolbox developed by Brabanter et al. [38] and Suykens and Vandewalle [39].

4 Results and Discussion

4.1 Overall Performance. Figure 6(*a*) shows the R-square results of different surrogate models estimated by the five individual surrogate models and the proposed E-AHF model, in which the mean R-square indicates the average accuracy of a surrogate model over all 40 test problems, and the standard deviation (Std) represents the robustness of the surrogates. It is found that the E-AHF model has the largest mean R-square 0.77, which is better than 0.75 calculated by the second best one generated by SVR model. Given that the small DoE sets (3n, 5n, 8n) account for one-third of the total DoE sets, this 0.77 accuracy of E-AHF is considerably high. Meanwhile, the Std of R-square of E-AHF is 0.16, which is the second best compared with the individual component models. Overall, the E-AHF model presents high accuracy and robustness compared with individual surrogate models for these 40 test problems.

Figure 6(*b*) compares the performance between E-AHF and another four benchmark hybrid models (i.e., AHF, BestPRESS, OWS(full), and PWS). Though the improvement of E-AHF is not as significant as that when compared with the individual surrogate models, the E-AHF model still outperforms the four benchmark hybrid models in terms of both accuracy and robustness. Meanwhile, BestPRESS and PWS have the second best value of 0.17, implying that hybrid models are generally capable of providing a more robust prediction than individual surrogate models. This is because hybrid models are able to take full advantages of the diversity of different surrogate models, which could potentially reduce the sensitivity of one individual surrogate model.

4.2 Performance Under Different Sample Sets. Figure 7(*a*) compares the performance of E-AHF with the individual models under different sample sets. As expected, it is seen that small DoE sets present a relatively low accuracy and large DoE sets present a high accuracy. Since the accuracy rank of the five individual surrogate models is varying among the sample sets, it is hard to determine the best individual model. For the small sample sets of (3n, 5n, 8n), the RBF-MQ model is the best among the individual models. For the medium and large sets, the SVR model is the

best. This also shows that no single individual surrogate model is always better than others [16]. However, by integrating these individual surrogate models, the proposed E-AHF model is always better than all the individual models under all sample sets. It shows that the E-AHF model performs slightly better than the best individual model (i.e., SVR) with large sample sets, and this is due to the fact that individual models such as SVR and RBF-thin plate spline are already considerably accurate with a large number of training points, combining them cannot lead to distinct improvement for the accuracy. However, when the medium and small sample sets are used, the E-AHF model has improved the accuracy by approximately 2–10% compared to the best individual surrogate model.

Figure 7(*b*) further compares the performance of E-AHF with the hybrid models under different sample sets. It is seen that the E-AHF model still performs the best, though several benchmark hybrid models have shown better performance than the individual surrogate models.

4.3 Performance for Different Problems. Figure 8 illustrates the R-square results of different-dimensional problems. It is observed that the E-AHF model is significantly better than all individual and hybrid surrogate models for low-dimensional test problems. For high-dimensional problems, the E-AHF model is slightly better than the SVR model and OWS (full) hybrid model. This indicates E-AHF's superior predictive capability for both low-dimensional and high-dimensional problems. It is also noted that the prediction accuracies for 8D and 10D problems are relatively low, and this is because all the 8D problems used in this work are highly nonlinear, as well as part of the 10D problems. It implies that surrogate modeling still needs further improvement, especially for high-dimensional and high-nonlinear problems.

4.4 Computational Cost. Regarding the computational cost, it is noticed that hybrid models generally take much more computational time than individual surrogate models, especially for high-dimensional problems. This is due to the fact that most hybrid models rely on the LOO CV calculation, and high-dimensional problems usually require more sample points. To compare the computational cost in detail, the 16D problems are used again based on a computer with a 3.5 GHz processor and 8G RAM. It is noted from Fig. 9 that BestPRESS, OWS(full), PWS, AHF, and E-AHF take 1.52, 1.61, 1.51, 2.77, and 0.31 s for 1000 predictions with 160 (10 n) sample points, respectively. The E-AHF model is the most computationally efficient model among these hybrid surrogate models. This is because that E-AHF uses a straightforward way to select the baseline model and calculate the weight factors for the predictions. In another word, except the LOO CV process, E-AHF neither requires additional calculation to identify the baseline model nor additional searching algorithm to calculate the adaptive weight factors. This computational time reduction becomes more critical when the surrogate model is used for a large number of calculations such as global sensitivity analysis. In a case of global sensitivity analysis with 100,000 repeated calculations, E-AHF may take approximately 30 min, and the second most efficient model (OWS(full)) may take 150 min for the same analysis.

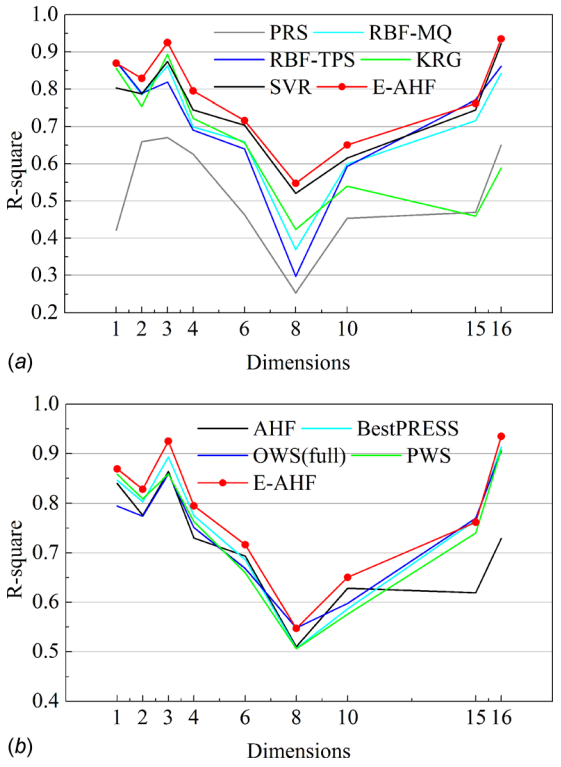


Fig. 8 R-square comparison for different-dimensional test problems: (*a*) compared with individual models and (*b*) compared with hybrid models

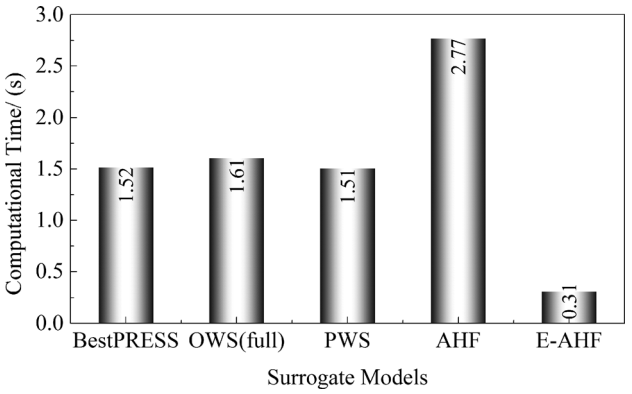


Fig. 9 Comparison of computational time of hybrid models

4.5 Impact of Threshold Value β. In the numerical experiments discussed earlier, the E-AHF model uses five individual surrogate models, and a threshold value, $\beta = 0.5$, is used to filter out the poorly performing individual models. Different threshold values (i.e., 0.0, 0.2, 0.5, 0.8, and 1.0) are compared here to investigate the effect of the threshold value and also help determine an appropriate β value for practical applications. A threshold value of $\beta = 0$ means only the best individual surrogate model is selected, and the hybrid model is essentially the same to the BestPRESS model. A threshold value of $\beta = 1.0$ means all the five

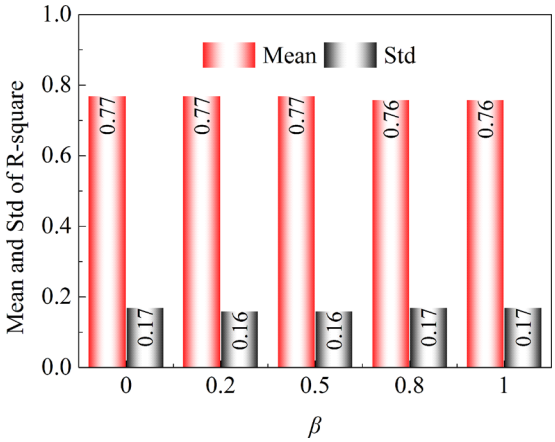


Fig. 10 Effect of threshold value on the E-AHF model

individual surrogate models are used for the E-AHF construction. Figure 10 shows the effect of the threshold value on the E-AHF model. It is seen that the threshold values of $\beta = 0.2$ and $\beta = 0.5$ yield the best performance in terms of both accuracy and robustness, and other threshold values yield worse performance in accuracy/robustness or both of them.

By analyzing the individual surrogate models integrated in the E-AHF model, it is found that the cases of $\beta = 0.2$ and $\beta = 0.5$ filter out three and two poorly performing individual models, respectively. The best two or three individual surrogate models are selected to form the E-AHF model when the threshold value is set to be 0.2 or 0.5, respectively. We also find that the model selected by the BestPRESS is not always the best individual surrogate. This is due to the fact that the CV process is not always able to identify the best one from a set of individual surrogate models [25]. Therefore, using multiple well-performing models is a good choice to guarantee that the best individual model is integrated in the hybrid model. In addition, the diversity of different individual models could potentially improve the accuracy and robustness of surrogate modeling. However, the potential benefit from the diversity diminishes as the number of individual models increase, since including the poorly performing models would worsen the overall performance of the E-AHF model. Therefore, the threshold value is recommended to be 0.2 or 0.5 when these five individual surrogate models are combined.

5 Conclusions

This paper developed an advanced hybrid surrogate model, namely, the E-AHF, which takes into account both individual surrogate models selection and adaptive weight factors determination. The LOO CV was utilized as the measure for filtering out poorly performing individual surrogate models and determining the baseline model for the ensemble of surrogate models. The adaptive weight factors were calculated based on the Gaussian-process estimated errors and the baseline prediction at any new points.

To evaluate the performance of the E-AHF model, a total of 40 test problems including low-dimensional and high-dimensional problems were used. The mean R-square was used to evaluate the accuracy of the surrogate models, and the standard deviation of R-square was used to evaluate the robustness of the surrogate models for different problems. The results showed that the proposed E-AHF model performed better than both the individual and benchmark hybrid surrogate models for these test problems in terms of both accuracy and robustness. The computational time of hybrid surrogate models was also compared, showing that E-AHF model was the most efficient model among the hybrid models.

In addition, we found that the LOO CV was not capable of identifying the best individual surrogate model for many test problems. In the future, we will focus on finding or developing a global measure to identify the best individual surrogate model. The impact of threshold value was studied, and an appropriate value was recommended, but it must be noted that the recommended value is only valid for the ensemble of the five individual surrogate models used in this study.

Funding Data

- National Natural Science Foundation of China, Dalian University of Technology (Grant Nos. 51505061 and U1608256).

References

[1] Queipo, N. V., Haftka, R. T., Shyy, W., Goel, T., Vaidyanathan R., and Tucker, P. K., 2005, "Surrogate-Based Analysis and Optimization," Prog. Aerosp. Sci., **41**(1), pp. 1–28.

[2] Song, X., Sun, G., Li, G., Gao, W., and Li, Q., 2013, "Crashworthiness Optimization of Foam-Filled Tapered Thin-Walled Structure Using Multiple Surrogate Models," Struct. Multidiscipli. Optim., **47**(2), pp. 221–231.

[3] Gorissen, D., Couckuyt, I., Demeester, P., Dhaene, T., and Crombecq, K., 2010, "A Surrogate Modeling and Adaptive Sampling Toolbox for Computer Based Design," J. Mach. Learn. Res., **11**(1), pp. 2051–2055.

[4] Forrester, A. I. J., Sóbester, A., and Keane, A. J., 2007, "Multi-Fidelity Optimization Via Surrogate Modeling," Proc. R. Soc. London A, **463**(2088), pp. 3251–3269.

[5] Peri, D., and Campana, E. F., 2005, "High-Fidelity Models and Multiobjective Global Optimization Algorithms in Simulation-Based Design," J. Ship Res., **49**(3), pp. 159–175.

[6] Yang, J., Zhan, Z., Zheng, K., Hu, J., and Zheng, L., 2016, "Enhanced Similarity-Based Metamodel Updating Strategy for Reliability-Based Design Optimization," Eng. Optim., **48**(12), pp. 2026–2045.

[7] Ong, Y. S., Nair, P. B., Keane, A. J., and Wong, K. W., 2005, "Surrogate-Assisted Evolutionary Optimization Frameworks for High-Fidelity Engineering Design Problems," *Knowledge Incorporation in Evolutionary Computation*, Springer, Berlin, pp. 307–331.

[8] Draper, N. R., and Smith, H., 2014, *Applied Regression Analysis*, 3rd ed., Wiley, New York.

[9] Matheron, G., 1963, "Principles of Geostatistics," Econ. Geol., **58**(8), pp. 1246–1266.

[10] Sacks, J., Schiller, S. B., and Welch, W. J., 1989, "Designs for Computer Experiments," Technometrics, **31**(1), pp. 41–47.

[11] Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P., 1989, "Design and Analysis of Computer Experiments," Stat. Sci., **4**(4), pp. 409–423.

[12] Fang, H., and Horstemeyer, M. F., 2006, "Global Response Approximation With Radial Basis Functions," Eng. Optim., **38**(4), pp. 407–424.

[13] Vapnik, V. N., and Vapnik, V., 1998, *Statistical Learning Theory*, Wiley, New York.

[14] Girosi, F., 1998, "An Equivalence Between Sparse Approximation and Support Vector Machines," Neural Comput., **10**(6), pp. 1455–1480.

[15] Zhou, X. J., and Jiang, T., 2016, "Metamodel Selection Based on Stepwise Regression," Struct. Multidiscip. Optim., **54**(3), pp. 641–657.

[16] Goel, T., Haftka, R. T., Shyy, W., and Queipo, N. V., 2007, "Ensemble of Surrogates," Struct. Multidiscip. Optim., **33**(3), pp. 199–216.

[17] Wang, G. G., and Shan, S., 2007, "Review of Metamodeling Techniques in Support of Engineering Design Optimization," ASME J. Mech. Des., **129**(4), pp. 370–380.

[18] Simpson, T. W., Mauery, T. M., Korte, J. J., and Mistree, F., 2001, "Kriging Models for Global Approximation in Simulation-Based Multidisciplinary Design Optimization," AIAA J., **39**(12), pp. 2233–2241.

[19] Clarke, S. M., Griebsch, J. H., and Simpson, T. W., 2005, "Analysis of Support Vector Regression for Approximation of Complex Engineering Analyses," ASME J. Mech. Des., **127**(6), pp. 1077–1087.

[20] Jin, R., Chen, W., and Simpson, T. W., 2001, "Comparative Studies of Metamodelling Techniques Under Multiple Modelling Criteria," Struct. Multidiscip. Optim., **23**(1), pp. 1–13.

[21] Fang, H., Rais-Rohani, M., Liu, Z., and Horstemeyer, M. F., 2005, "A Comparative Study of Metamodeling Methods for Multiobjective Crashworthiness Optimization," Comput. Struct., **83**(25–26), pp. 2121–2136.

[22] Zerpa, L. E., Queipo, N. V., Pintos, S., and Salager, J. L., 2005, "An Optimization Methodology of Alkaline–Surfactant–Polymer Flooding Processes Using Field Scale Numerical Simulation and Multiple Surrogates," J. Pet. Sci. Eng., **47**(3–4), pp. 197–208.

[23] Acar, E., and Rais-Rohani, M., 2009, "Ensemble of Metamodels With Optimized Weight Factors," Struct. Multidiscip. Optim., **37**(3), pp. 279–294.

[24] Ferreira, W. G., and Serpa, A. L., 2016, "Ensemble of Metamodels: The Augmented Least Squares Approach," Struct. Multidiscip. Optim., **53**(5), pp. 1019–1046.

[25] Viana, F. A. C., Haftka, R. T., and Steffen, V., Jr., 2009, "Multiple Surrogates: How Cross-Validation Errors Can Help Us to Obtain the Best Predictor," Struct. Multidiscip. Optim., **39**(4), pp. 439–457.

[26] Acar, E., 2010, "Various Approaches for Constructing an Ensemble of Metamodels Using Local Measures," Struct. Multidiscip. Optim., **42**(6), pp. 879–896.

[27] Zhang, J., Chowdhury, S., and Messac, A., 2012, "An Adaptive Hybrid Surrogate Model," Struct. Multidiscip. Optim., **46**(2), pp. 223–238.

[28] Liu, H., Xu, S., Wang, X., Meng, J., and Yang, S., 2016, "Optimal Weighted Pointwise Ensemble of Radial Basis Functions With Different Basis Functions," AIAA J., **54**(10), pp. 3117–3133.

[29] Forrester, A., Sobester, A., and Keane, A., 2008, *Engineering Design Via Surrogate Modelling: A Practical Guide*, Wiley, Chichester, UK.

[30] Lee, Y., and Choi, D., 2014, "Pointwise Ensemble of Meta-Models Using v Nearest Points Cross-Validation," Struct. Multidiscip. Optim., **50**(3), pp. 383–394.

[31] Glaz, B., Goel, T., Liu, L., and Haftka, R. T., 2007, "Application of a Weighted Average Surrogate Approach to Helicopter Rotor Blade Vibration Reduction," AIAA Paper No. 2007-1898.

[32] Gramacy, R. B., and Lee, H. K. H., 2012, "Cases for the Nugget in Modeling Computer Experiments," Stat. Comput., **22**(3), pp. 713–722.

[33] Talgorn, B., Kokkolaras, M., and Digabel, S. L., 2015, "Statistical Surrogate Formulations for Simulation-Based Design Optimization," ASME J. Mech. Des., **137**(2), p. 021405.

[34] Mullur, A. A., and Messac, A., 2006, "Metamodeling Using Extended Radial Basis Functions: A Comparative Approach," Eng. Comput., **21**(3), pp. 203–217.
[35] Cai, X., Qiu, H., Gao, L., and Shao, X., 2016, "An Enhanced RBF-HDMR Integrated With an Adaptive Sampling Method for Approximating High Dimensional Problems in Engineering Design," Struct. Multidiscip. Optim., **53**(6), pp. 1209–1229.
[36] McKay, M. D., Beckman, R. J., and Conover, W. J., 2000, "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output From a Computer Code," Technometrics, **42**(1), pp. 55–61.
[37] Viana, F. A. C., 2010, "SURROGATES Toolbox User's Guide," Gainesville, FL.
[38] Brabanter, K. D., Karsmakers, P., Ojeda, F., Alzate, C., Brabanter, J. D., Pelckmans, K., Moor, B. D., Vandewalle, J., and Suykens, J. A. K., 2011, "LS-SVMlab Toolbox User's Guide. Version 1.8," ESAT-SISTA, Katholieke Universiteit Leuven, Belgium, Technical Report No. 10-146.
[39] Suykens, J. A. K., and Vandewalle, J., 1999, "Least Squares Support Vector Machine Classifiers," Neural Process. Lett., **9**(3), pp. 293–300.