

# Ensemble of surrogates with hybrid method using global and local measures for engineering design

Liming Chen<sup>1</sup> · Haobo Qiu<sup>1</sup> · Chen Jiang<sup>1</sup> · Xiwen Cai<sup>1</sup> · Liang Gao<sup>1</sup>

Received: 27 April 2017 / Revised: 5 September 2017 / Accepted: 19 October 2017 / Published online: 6 November 2017  
© Springer-Verlag GmbH Germany 2017

**Abstract** Surrogate models are usually used as a time-saving approach to reduce the computational burden of expensive computer simulations for engineering design. However, it is difficult to choose an appropriate model for an unknown design space. To tackle this problem, an effective method is forming an ensemble model that combines several surrogate models. Many efforts were made to determine the weight factors of ensemble, which include global and local measures. This article investigates the characteristics of global and local measures, and presents a new ensemble model which combines the advantages of these two measures. In the proposed method, the design space is divided into two parts, and different strategies are introduced to evaluate the weight factors in these two parts respectively. The results from numerical and engineering design cases show that the proposed ensemble model has satisfactory robustness and accuracy (it performs best for most cases tested in this article), while spending almost the equivalent modeling time (the additional cost is not more than 6.7% for any case tested in this article) compared with the combined global and local ensemble models.

**Keywords** Ensemble model · Global measure · Local measure · Surrogate models

✉ Haobo Qiu  
hobbyqiu@163.com

<sup>1</sup> State Key Laboratory of Digital Manufacturing Equipment and Technology, School of Mechanical Science and Engineering, Huazhong University of Science & Technology, Wuhan 430074, People's Republic of China

## Nomenclature

$d$	Number of design variables.
$E_i$	Root generalized mean square cross-validation error of the $i^{\text{th}}$ surrogate.
$e_{ik}$	Cross-validation error of the $i^{\text{th}}$ surrogate at the $k^{\text{th}}$ sample point.
$\hat{f}^{\text{ens}}$	Predictor of the ensemble.
$\hat{f}_i$	Predictor of the $i^{\text{th}}$ surrogate.
$N$	Number of test points.
$N_s$	Number of surrogates used in the ensemble.
$n$	Number of sample points.
$P_k$	Ratio of the global cross-validation error to the local cross-validation error at the $k^{\text{th}}$ sample point.
$R^o$	Outer region.
$R^i$	Inner region.
$r_k$	Radius of the $k^{\text{th}}$ point's inner region.
$r_k^{\text{max}}$	Euclidean distance between the $k^{\text{th}}$ sample point and the closest sample point.
$S$	Sample points set.
$WCVE$	Weighted cross-validation error.
$w_i$	Normalized weight of the $i^{\text{th}}$ surrogate.
$w_i^*$	Unnormalized weight of the $i^{\text{th}}$ surrogate.
$w_{ik}$	Pointwise weight of the $i^{\text{th}}$ surrogate at the $k^{\text{th}}$ sample point.
$x^{\text{nearest}}$	Sample point which is nearest to the prediction point.
$\hat{y}_{ik}$	Response predicted by the $i^{\text{th}}$ surrogate at the $k^{\text{th}}$ point, the surrogate is constructed by using leave-one-out cross-validation.
$y_k$	True response at the $k^{\text{th}}$ sample/test point.
$\hat{y}_k$	Prediction response at the $k^{\text{th}}$ sample/test point.
$\rho$	Impact metric of local measure.

## 1 Introduction

Computer simulations are widely used in structural optimization of mechanical systems. However, for the design of complicated structures, computer simulation techniques such as finite element analysis may need unbearable computing time, which hinders the successful optimization of system performance. As a cheap alternative for computationally expensive simulations, surrogate modeling method can solve this problem very well and has been developed rapidly over the last few decades.

Surrogate models such as Polynomial Response Surface (PRS; Box and Draper 1987), Kriging (KRG; Sacks et al. 1989), Radial Basis Functions (RBF; Hardy 1971) and Support Vector Regression (SVR; Smola and Schölkopf 2004) are widely used in the practice of engineering design. And some reviews of various surrogates can be found in Queipo et al. (2005), Forrester and Keane (2009) and references therein. However, there is no clear consensus on which one is most suitable for an unknown problem. More recently, inspired by Bishop's work (1995) in neural network, a promising modeling technique which is known as ensemble of surrogate models was developed (Zerpa et al. 2005; Goel et al. 2007). The ensemble method combines different surrogate models through a weighted form and the weight factor of each surrogate model is determined by the model accuracy. It is reasonable to view the ensemble approach as an alternative to model selection in statistics, and there are many researches in this area, including selection methods of models based on Akaike's Information Criterion (AIC), Bayes Information Criterion (BIC), cross-validation, and the structural risk minimization methods (Madigan and Raftery 1994; Kass and Raftery 1995; Buckland et al. 1997; Cherkassky et al. 1999; Hoeting et al. 1999). Existing ensemble modeling methods can be summarized as global measures and local measures. For simplicity, weight factors evaluated from global measures and local measures are called global weight factors and local weight factors respectively in this article.

Global measures evaluate the weight factors over the entire design space. For each surrogate model, the weight factor keeps constant at every sampling point. Goel et al. (2007) proposed a heuristic algorithm in which the weight factor is calculated from generalized mean square cross-validation error (GMSE). Acar and Rais-Rohani (2009) treated the weight factors as design variables in an optimization problem, GMSE and root mean square error (RMSE) are selected as the objective function respectively. The optimization problem is solved by a numerical optimization procedure in their work. Viana et al. (2009) also obtained the optimum weight factors through minimizing RMSE, but they solved the optimization problem analytically by using Lagrange multipliers. Zhou et al. (2011) introduced a recursive algorithm in which the final averaged ensemble model is obtained by iterative modeling.

Compared with global measures, local measures evaluate the weight factors point by point, so the weight factors of each surrogate model are different at every sampling point. Sanchez et al. (2008) used prediction variance of the  $k$ -nearest sampling points around the prediction point to evaluate the weight factors. Based on the pointwise weight factors at sample points and the distances between the sample points and the prediction point, Acar (2010) proposed a spatial ensemble model.

Considering that global measures and local measures both have their pros and cons, an ensemble method (ES-HGL) which hybrids a global measure and a local measure is proposed in this article. In this method, design space is divided into two regions: the region far from the sample points (the outer region) and the region near to the sample points (the inner region). Then two strategies are introduced to evaluate the weight factors for different regions respectively: (1) in the outer region, a new weight factor named Hybrid Weight Factor is introduced; (2) in the inner region, the Hybrid Weight Factor and the local weight factor are combined in a certain way based on the location of the prediction point.

The remainder of this article is organized as follows: Some representative ensemble methods are briefly overviewed in Section 2. The development of the proposed ES-HGL model is described in Section 3. Several numerical and engineering examples are tested in Section 4. And several conclusions are presented in Section 5.

## 2 Background of ensemble methods

The common way of using surrogate modeling methods includes the following steps: constructing several candidate surrogate models, selecting the most accurate one based on some criteria and discarding the rest. However, this scenario has two major shortcomings. First, it is a waste of resource used on the construction of those so-called "inaccurate" models. Second, the performances of different surrogate models are influenced by the sample points, which means one surrogate model may be accurate on one data set but may be inaccurate on another one. To overcome these shortcomings, ensemble methods are proposed.

An ensemble model is a weighted combination of several individual surrogate models. The basic form of an ensemble model is defined as

$$\begin{aligned} \hat{f}^{ens} &= \sum_{i=1}^{N_s} w_i \hat{f}_i \\ \sum_{i=1}^{N_s} w_i &= 1 \end{aligned} \quad (1)$$

where  $\hat{f}^{ens}$  is the response value of the ensemble model,  $N_s$  is the number of used surrogate models and  $w_i$  is the weight

factor of the  $i^{th}$  surrogate model  $\hat{f}_i$ . Apparently, if one surrogate model is more accurate than another, it will occupy a larger proportion in the ensemble model, and vice versa. In this section, some representative global measures and local measures for ensemble modeling are briefly introduced.

**2.1 Global measures**

Goel et al. (2007) proposed an ensemble model which is based on GMSE. The weight factors are evaluated from a heuristic algorithm

$$w_i = \frac{w_i^*}{\sum_{j=1}^{N_s} w_j^*}, w_i^* = (E_i + \alpha \bar{E})^\beta$$

$$\bar{E} = \frac{1}{N_s} \sum_{i=1}^{N_s} E_i, E_i = \sqrt{\frac{1}{n} \sum_{k=1}^n (y_k - \hat{y}_{ik})^2}$$
(2)

where  $y_k$  is the actual response at the  $k^{th}$  sampling point,  $\hat{y}_{ik}$  is the  $i^{th}$  surrogate model’s corresponding prediction value using cross-validation and  $n$  is the number of sample points. Two unknown parameters  $\alpha(\alpha < 1)$  and  $\beta(\beta < 0)$  are determined based on the relationship between  $E_i$  and  $\bar{E}$ . Goel et al. (2007) suggested  $\alpha = 0.05$  and  $\beta = -1$  in their study.

Acar and Rais-Rohani (2009) tried both GMSE and RMSE as the global error metrics. The weight factors of different surrogate models in the ensemble model are determined by solving the following optimization problem

$$\begin{aligned} & \text{Find } w_i \\ \min & \text{ GMSE or RMSE} \\ & \text{s.t. } \sum_{i=1}^{N_s} w_i = 1 \end{aligned}$$
(3)

**2.2 Local measures**

Sanchez et al. (2008) used prediction variance as the local error metric to construct the ensemble model. The ensemble method is based on the k-nearest prediction variance, the weight factors are evaluated from

$$w_i = \frac{1}{\sum_{j=1}^{N_s} \frac{1}{V_j^{near}}} \quad V_i^{near} = \frac{1}{k-1} \sum_{i=1}^k (y_i - \hat{y}_i)^2$$
(4)

where  $V_i^{near}$  is the prediction variance of the  $i^{th}$  surrogate model. Here, Sanchez et al. (2008) suggested that  $k = 3$  is a reasonable choice.

Acar (2010) proposed a spatial model, in which the calculation of weight factors depends on the pointwise weight factors and the distances between sample points and the prediction point. Hence, the weight factors are evaluated from

$$w_i = \frac{w_{ik}^*}{\sum_{j=1}^{N_s} w_{jk}^*}, w_{ik}^* = \sum_{k=1}^n w_{ik} I_k(x)$$

$$I_k(x) = \frac{1}{d_k^2(x)}, d_k(x) = \|x - x_k\|$$
(5)

where  $w_{ik}$  is the pointwise weight factor of the  $i^{th}$  surrogate model at the  $k^{th}$  sample point.  $w_{ik}$  equals one for the surrogate model with the lowest cross-validation error at the  $k^{th}$  sample point, and equals zero for all other surrogate models at this sample point.  $I_k(x)$  is a distance metric, especially  $w_{ik}^* = w_{ik}$  when  $d_k(x) = 0$ . Three other approaches for determining  $w_{ik}$  and  $I_k(x)$  were also proposed by Acar (2010).

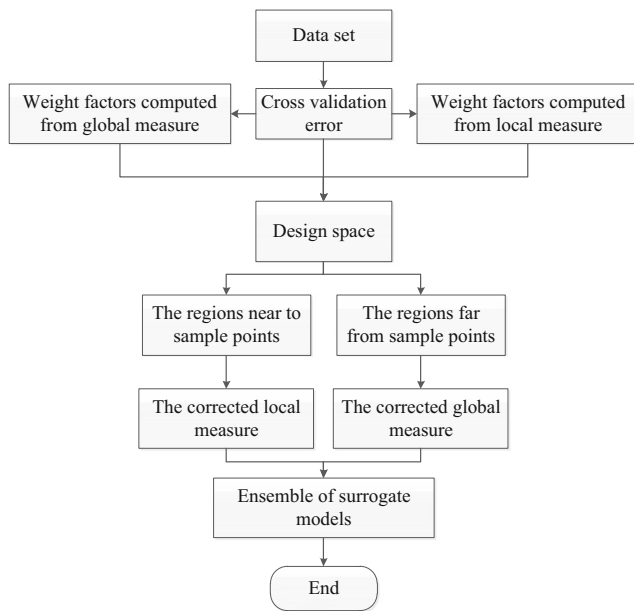
**3 The proposed hybrid ensemble method**

Constructing an ensemble model merely with global measures or local measures both have pros and cons. Global measures can guarantee the modeling accuracy in a global perspective but ignore the diversity of the combined surrogate models. Local measures can make the ensemble model more flexible but less robust, because inaccurate local error metrics may influence the model accuracy severely. This article integrates a global measure with a local measure in ensemble modeling, attempting to make the ensemble model more robust and accurate. Thus we call this approach ensemble of surrogates with hybrid method using global and local measures. In this method, the same error matrix, which is the most time-consuming part in the modeling process, is used for both global and local measures. It means that the modeling accuracy can be enhanced with the amount of modeling time remaining nearly the same.

The flowchart for ES-HGL model is shown in Fig. 1. Three key steps of this proposed method are: calculation of the weight factors using global measure and local measure based on the same error matrix, division of the design space and construction of ES-HGL model in the divided design space with different weight calculation strategies. These will be discussed in detail in Section 3.1 to 3.3.

**3.1 Calculation of weight factors using global and local measures**

Modeling error is an important criterion to evaluate the accuracy of the surrogate model. Commonly used measures include prediction variance and cross-validation error. With no more additional test points needed for error calculation, the cross-validation error is used as the modeling error to construct ES-HGL model in this article. Cross-validation error is the prediction error at each sample point when the surrogate model is constructed by using the other  $(n - 1)$  points (it is also called leave-one-out cross-validation



**Fig. 1** Development and evaluation steps for ES-HGL model

error). Cross-validation error of the  $i^{th}$  surrogate at the  $k^{th}$  sample point is evaluated as

$$e_{ik} = y_k - \hat{y}_{ik} \tag{6}$$

In order to save the repeated computational time and make a fair comparison, the same error matrix is used for constructing all ensemble models. Two ensemble modeling methods are selected to construct the proposed ES-HGL: the heuristic algorithm proposed by Goel et al. (2007) as the global measure, and the spatial model presented by Acar (2010) as the local measure. These two ensemble methods both use cross-validation error as the error matrix and can provide excellent modeling performances. The formulation of each method can be found in Section 2 and the measures can be described as follows

$$\begin{cases} w^G \Leftarrow f^G(e^{CV}) \\ w^L \Leftarrow f^L(e^{CV}) \end{cases} \tag{7}$$

where  $e^{CV}$  represents the error matrix which is computed from cross-validation error,  $f(\cdot)$  represents the strategy of weight factor calculation, and the superscript G and L denote the weight factor is obtained by using global measure or local measure.

### 3.2 Division of the design space

It is crucial to choose a proper method to calculate weight factors of the combined surrogate models. Researchers have summarized the existing ensemble methods into two classes: the global measures and the local measures. However there is no consensus that which measure is the best one

when it comes to an unknown problem. So in this article, the weight factors of individual surrogate models are calculated using different measures according to the location of the prediction point.

Due to the weight factors at a prediction point are evaluated based on the error at the sample points, the calculation strategies of weight factors need to be different for areas near to and far from the sample points. Among many of the existing methods, different measures have their own characteristics in calculating weight factors: (1) Local measures consider that the weight factors should indicate the diversity of combined surrogate models at different locations in the design space, so the weight factors are evaluated point by point and the weight factors at a prediction point are heavily influenced by the nearest sample point; (2) Global measures regard weight factors as the representative of overall accuracy, so that the weight factors are evaluated from the entire error matrix. Hence the weight factors at a prediction point are not affected greatly by the modeling error of any one sample point. Therefore, local measures are more suitable for regions near to the sample points while global measures are more appropriate for regions far from the sample points. Thus it is reasonable to define the weight factors as:

$$w_i^* = \begin{cases} w_i^G & \text{if } x \in R^o \\ w_i^L & \text{if } x \in R^i \end{cases} \tag{8}$$

$$w_i = \frac{w_i^*}{\sum_{j=1}^{N_s} w_j^*}$$

where  $R^o$  and  $R^i$  are the outer and inner regions. Notice that the actual landscape of interest is unknown, every inner part  $R^i$  should have an n-sphere shape whose center is the according sample point. Thus the design space can be divided from:

$$x \in \begin{cases} R^i & \text{if } \|x - x_k^{nearest}\| \leq r_k \\ R^o & \text{if } \|x - x_k^{nearest}\| > r_k \end{cases} \tag{9}$$

where  $R^o$  and  $R^i$  denote the regions far from and near to the sample points respectively,  $\|\cdot\|$  denotes the Euclidean distance between the prediction point and the closest sample point and  $r_k$  is the radius of the  $k^{th}$  sample point's inner region.

Once the division of design space is implemented, the determination of the region radius of each sample point is crucial. To solve this problem, the errors of ensemble models at sample points are calculated first, and then the region radii are computed from those errors. Obviously, the modeling errors evaluated from sample points are important criteria in the determination of region radii. In this article, the weighted cross-validation error (WCVE for short), which is a weighted sum of cross-validation errors from combined surrogate models, is utilized to evaluate the

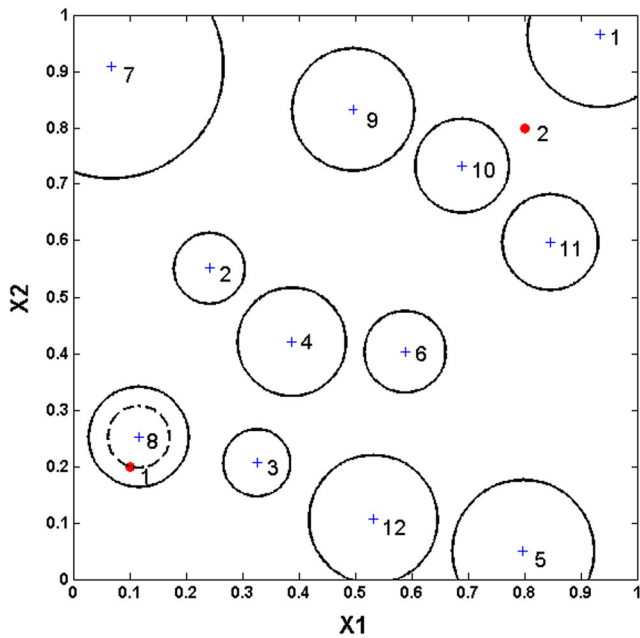


Fig. 2 A 2-D problem for illustrating the division of the design space

modeling error of ES-HGL. Apparently, with the strategy of weight factor calculation changing, the  $WCVEs$  of sample points are also different. For simplicity,  $WCVEs$  computed from the global measure and the local measure are denoted as  $WCVE_k^G$  and  $WCVE_k^L$  respectively in this article.

In ES-HGL, to evaluate the region radius of each sample point, two criteria should be followed:

(1) The region radius of the  $k^{th}$  sample point is proportional to the distance ( $r_k^{max}$ ) between the current sample point and the closest sample point;

(2) The region radius depends on the ratio of  $WCVE_k^G$  to  $WCVE_k^L$  at the current sample point. That is to say, the region radius  $r_k(\cdot)$  is the function of the error ratio  $P_k$ . Moreover, the region radius  $r_k(P_k)$  monotonically increases as the error ratio  $P_k$  increasing from 1 to positive infinity. At the same time, two boundary conditions should be satisfied:

$$\lim_{P_k \rightarrow 1} r(P_k) = 0 \tag{10}$$

$$\lim_{P_k \rightarrow +\infty} r(P_k) = r_k^{max} \tag{11}$$

That is to say, when the global weighted cross-validation error  $WCVE_k^G$  at the current sample point is smaller than the local one (the error ratio  $P_k$  is less than one), the region radius should be equal to zero. That is because the global measure is deemed to be more accurate than the local measure in this case, so we just adopt the global measure only. And when the global cross-validation error at the current sample point is far less than the local one (the error ratio  $P_k$  is tend to be positive infinity), the region radius should be equal to the distance between the current sample point and the closest sample point  $r_k^{max}$ . And the local measure is deemed to be more accurate than the local measure in this case, so we just adopt the local measure only.

We consider the following two feasible formulas of the region radius  $r_k(P_k)$  in the range of the elementary function:

$$r_k^1(P_k) = r_k^{max} \left(1 - \frac{1}{P_k}\right) \tag{12}$$

$$r_k^2(P_k) = r_k^{max} (1 - e^{1-P_k}) \tag{13}$$

After testing these two formulas respectively, we decide to adopt the first form  $r_k^1(P_k)$  according to the test results (see in Appendix C.1). Hence, the region radius in (9) is evaluated from

$$r_k = \begin{cases} r_k^{max} \cdot \left(1 - \frac{1}{P_k}\right) & \text{if } P_k > 1 \\ 0 & \text{else} \end{cases} \tag{14}$$

where

$$r_k^{max} = \frac{1}{2} \min_{i \in S, i \neq k} \|x_k - x_i\| \tag{15}$$

$$P_k = \frac{WCVE_k^G}{WCVE_k^L} \tag{16}$$

$$WCVE_k^G = \sum_{i=1}^{N_s} |w_{ik}^G e_{ik}|, WCVE_k^L = \sum_{i=1}^{N_s} |w_{ik}^L e_{ik}| \tag{17}$$

where  $S$  denotes the sample points set,  $r_k^{max}$  is half of the minimum distance between the  $k^{th}$  sample point and the nearest sample point to the  $k^{th}$  sample point and  $P_k$  denotes the  $k^{th}$  ratio of the weighted cross-validation error computed from different measures.

Table 1 Summary of the training and test point sets used in each problem

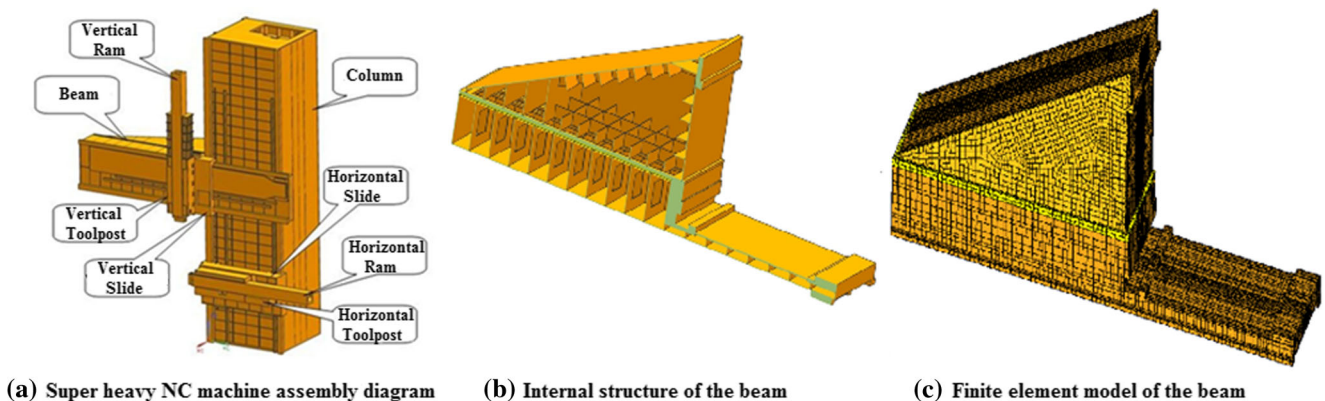
Function	Design variables	Training sets	Sample points	Test points
Branin-Hoo	2	1000	12	1000
Camelback	2	1000	12	1000
Hartman-3	3	500	20	1000
Hartman-6	6	100	56	1000
Extended-Rosenbrock	9	20	110	1000
Dixon-Price	12	10	182	1000



**Table 2** Mean and COV of RMSE, AAE and MAE for different surrogate models

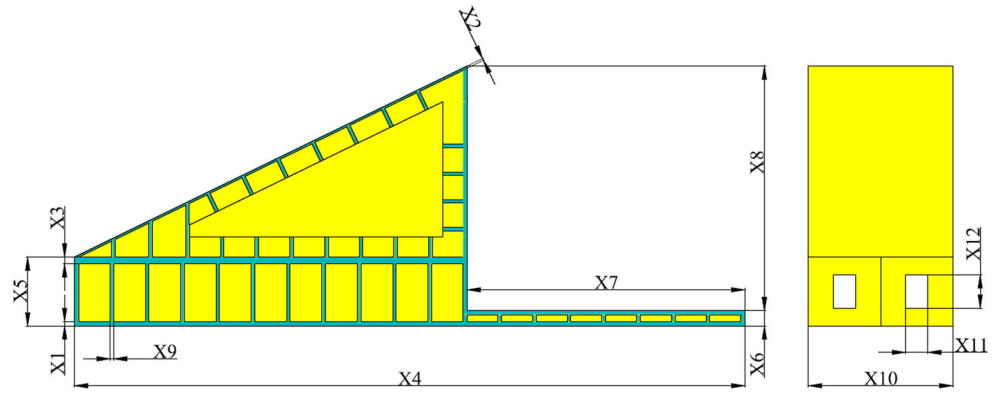
	PRS	RBF	KRG	OM	EG	SP	ES-HGL
Branin-Hoo							
RMSE	32.08/ <b>0.14</b>	25.51/0.21	<b>24.48</b> /0.34	26.44/0.27	24.17/ <b>0.23</b>	24.08/0.24	<b>23.97</b> /0.24
AAE	23.07/ <b>0.14</b>	<b>15.81</b> /0.17	15.87/0.35	17.43/0.28	15.61/ <b>0.20</b>	15.48/0.21	<b>15.40</b> /0.21
MAE	146.86 /0.26	148.43 /0.33	<b>136.61</b> /0.39	141.14 /0.34	135.40 /0.34	134.79 /0.34	<b>132.27</b> /0.35
Camelback							
RMSE	17.99/0.13	<b>17.68</b> / <b>0.10</b>	22.97/0.20	19.02/0.20	17.52/ <b>0.13</b>	17.82/0.14	<b>17.51</b> / <b>0.13</b>
AAE	12.80/0.16	<b>11.41</b> / <b>0.11</b>	15.40/0.18	12.82/0.20	11.41/0.13	11.63/0.14	<b>11.34</b> / <b>0.11</b>
MAE	<b>74.86</b> /0.20	85.16/ <b>0.16</b>	97.22/0.21	83.73/0.22	82.12/ <b>0.19</b>	82.85/ <b>0.19</b>	<b>80.90</b> / <b>0.19</b>
Hartman-3							
RMSE	0.79/ <b>0.18</b>	0.74/0.22	<b>0.45</b> /0.27	0.50/0.34	0.51/0.23	<b>0.49</b> /0.25	<b>0.49</b> /0.22
AAE	0.58/ <b>0.16</b>	0.53/0.19	<b>0.29</b> /0.26	0.40/0.33	0.35/ <b>0.19</b>	0.34/0.21	<b>0.33</b> /0.20
MAE	3.26/0.33	3.15/0.36	<b>1.82</b> / <b>0.27</b>	2.40/0.44	2.12/0.34	2.08/0.35	<b>2.03</b> / <b>0.29</b>
Hartman-6							
RMSE	0.40/0.19	0.35/ <b>0.13</b>	<b>0.34</b> /0.18	0.34/0.16	0.31/0.12	0.31/0.11	<b>0.30</b> / <b>0.10</b>
AAE	0.29/0.22	0.25/ <b>0.17</b>	<b>0.21</b> /0.18	0.22/0.20	0.21/0.15	0.20/0.14	<b>0.19</b> / <b>0.13</b>
MAE	2.03/ <b>0.17</b>	<b>1.90</b> /0.18	2.00/0.19	1.94/0.18	<b>1.85</b> / <b>0.17</b>	1.86/ <b>0.17</b>	<b>1.85</b> / <b>0.17</b>
Extended-Rosenbrock							
RMSE	7.68*10 <sup>4</sup> /0.76	<b>7.64</b> *10 <sup>4</sup> /0.76	7.67*10 <sup>4</sup> /0.76	7.65*10 <sup>4</sup> /0.76	7.65*10 <sup>4</sup> /0.76	7.65*10 <sup>4</sup> /0.76	<b>7.64</b> *10 <sup>4</sup> /0.62
AAE	7.64*10 <sup>4</sup> /0.76	<b>7.60</b> *10 <sup>4</sup> /0.77	7.63*10 <sup>4</sup> /0.76	7.61*10 <sup>4</sup> /0.77	7.62*10 <sup>4</sup> /0.76	7.62*10 <sup>4</sup> /0.76	<b>7.59</b> *10 <sup>4</sup> /0.63
MAE	1.38*10 <sup>5</sup> /0.37	<b>1.33</b> *10 <sup>5</sup> /0.35	1.33*10 <sup>5</sup> /0.36	1.33*10 <sup>5</sup> /0.36	1.33*10 <sup>5</sup> /0.36	1.32*10 <sup>5</sup> /0.36	1.33*10 <sup>5</sup> /0.36
Dixon-Price							
RMSE	<b>3.70</b> *10 <sup>5</sup> /0.69	3.76*10 <sup>5</sup> /0.67	3.74*10 <sup>5</sup> /0.67	3.72*10 <sup>5</sup> /0.68	3.70*10 <sup>5</sup> /0.69	3.70*10 <sup>5</sup> /0.69	<b>3.35</b> *10 <sup>5</sup> /0.60
AAE	<b>3.62</b> *10 <sup>5</sup> /0.73	3.66*10 <sup>5</sup> /0.71	<b>3.62</b> *10 <sup>5</sup> /0.72	3.62*10 <sup>5</sup> /0.73	3.62*10 <sup>5</sup> /0.73	3.61*10 <sup>5</sup> /0.73	<b>3.28</b> *10 <sup>5</sup> /0.63
MAE	<b>6.56</b> *10 <sup>5</sup> /0.33	6.85*10 <sup>5</sup> /0.33	7.47*10 <sup>5</sup> /0.33	6.97*10 <sup>5</sup> /0.28	6.63*10 <sup>5</sup> /0.33	6.79*10 <sup>5</sup> /0.32	<b>5.99</b> *10 <sup>5</sup> /0.37

For ease of comparison, the best results of ensemble models and individual surrogate models are shown in bold respectively (the lowest value for RMSE, AAE, MAE and COV).



**Fig. 3** The structure of a super-heavy machine

**Fig. 4** Diagram of the simplified beam with twelve variables



### 3.3 Construction of ES-HGL model

Though using a global measure and a local measure to construct the ensemble model in different regions is a good strategy, the defects are also distinct. The global measure ignores the diversity of combined surrogate models in different areas, and the local measure may be inaccurate when the error matrix cannot represent the actual modeling error very well. To balance the global measure and the local measure, a new weight factor named the Hybrid Weight Factor (HWF) is introduced. The value of HWF should be evaluated between the global weight and the local weight, i.e.,  $w^G \leq w^H \leq w^L$  or  $w^L \leq w^H \leq w^G$ . Then we have tried the following four types of HWFs:

$$w_1^H = \sqrt{\frac{(w^G)^2 + (w^L)^2}{2}} \tag{18}$$

$$w_2^H = \sqrt{w^G \times w^L} \tag{19}$$

$$w_3^H = \frac{w^G + w^L}{2} \tag{20}$$

$$w_4^H = \frac{WCVE^L}{WCVE^G + WCVE^L} w^G + \frac{WCVE^G}{WCVE^G + WCVE^L} w^L \tag{21}$$

We decide to adopt the first form  $w_1^H$  according to the test results (see in Appendix C.2).

The Hybrid Weight Factor can possess the following advantages:

- (1) The Hybrid Weight Factor makes the model more robust and accurate, because the modeling error in some local areas would be eliminated by the overall accuracy;
- (2) The Hybrid Weight Factor makes the model more flexible, because the diversities of combined surrogate models are sufficiently considered in the whole design space;
- (3) The Hybrid Weight Factor is succinct and easy to construct, for almost no extra computational burden is brought in the computing process.

Because the weight factors are based on the errors evaluated from the sample points, they have higher probability to be more accurate in regions near to the sample points. Then local weight factor, which can make an ensemble model more flexible, is used to amend the Hybrid Weight Factor in regions near to the sample points. In addition, considering that the local weight factors may be less accurate as the distance between the prediction point and the nearest sample point increasing, the effect of modification should be weakened. Thus in the new weight factor calculation strategy, the weight factor evaluated from (8) is modified based on the following two criteria:

- (1) in the outer region  $R^o$ , the weight factor is equal to the Hybrid Weight Factor;
- (2) in the inner region  $R^i$ , the weight factor consists of two parts: the local weight factor and the Hybrid Weight Factor. The proportion of the local weight factor decreases with the distance between the prediction point and the nearest sample point increasing.

Therefore, (8) is replaced by

$$w_i^* = \begin{cases} w_i^H & \text{if } x \in R^o \\ w_i^L \rho + w_i^H (1-\rho) & \text{if } x \in R^i \end{cases} \tag{22}$$

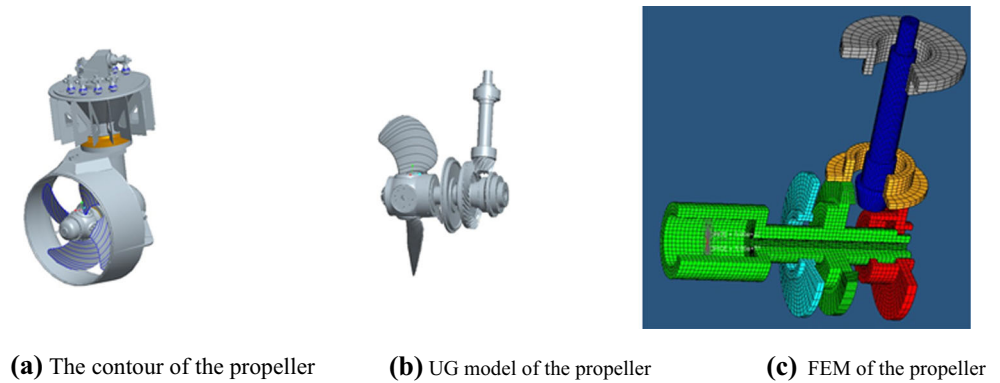
$$w_i = \frac{w_i^*}{\sum_{j=1}^{N_s} w_j^*}$$

**Table 3** Result comparison for the design of NC machine beam

	PRS	RBF	KRG	OM	EG	SP	ES-HGL
RMSE	<b>0.2368</b>	0.3730	0.4054	0.2578	0.2832	0.2587	<b>0.2328</b>
AAE	<b>0.1590</b>	0.2401	0.2392	0.1583	0.1670	0.1579	<b>0.1571</b>
MAE	<b>0.9322</b>	1.4571	1.9178	1.1479	1.2454	1.0567	<b>1.0465</b>

For ease of comparison, the best results of ensemble models and individual surrogate models are shown in bold respectively (the lowest value for RMSE, AAE and MAE).

**Fig. 5** Different models of the propeller



**(a)** The contour of the propeller

**(b)** UG model of the propeller

**(c)** FEM of the propeller

Considering that the volume of n-sphere (Rennie 2005) is

$$V_n(r) = \frac{\pi^{\frac{n}{2}}}{\Gamma\left(\frac{n}{2} + 1\right)} r^n \quad (23)$$

where the volume  $V_n$  is proportional to  $r^n$  and  $n$  denotes the dimension. So the impact metric of local measure  $\rho$  is calculated from

$$\rho = 1 - \left( \frac{\|x - x_k^{nearest}\|}{r_k} \right)^n \quad (24)$$

The results of ES-HGL generated by using (8) and (22) respectively are given in the Appendix C.3. From the result comparison we can see that it is reasonable to use (22) in place of (8).

For better understanding how the design space is divided and how the various measures are applied, a 2-D problem is used for illustration in Fig. 2.

In Fig. 2, the design space is normalized to  $[0,1]^2$  and 12 sample points are denoted as “+”. The circular regions with 12 sample points as the center points are inner regions defined in ES-HGL. The radii of these regions are evaluated from (14). The rest region in the design space is the outer region. Two random selected prediction points (denoted as “•”) are used to give a demonstration. The prediction point No. 1 is located in the circular region whose center point is the sample point No. 8, so it is in the inner region. Then the weights of surrogate models are determined by (22), (23) and (24). More exactly, the prediction point  $x$  and the nearest sample point  $x_k^{nearest}$  in (24) are the prediction point No. 1 and the sample point No. 8 respectively. The inner region radius  $r_k$  in (24) is the radius of the circular region whose center point is the sample point No. 8. For the prediction point No. 2, the weights of surrogate models in ES-HGL are obtained

from (22). Because this prediction point is not located in any circular region, that is to say, it is in the outer region.

#### 4 Case studies

The approximation performance of ES-HGL model is compared with three existing ensemble models: the heuristic algorithm EG proposed by Goel et al. (2007) as the global measure, the spatial model SP introduced by Acar (2010) as the local measure and the optimization-based method OM (minimizing GMSE through a numerical optimization procedure to obtain the optimum weight factors) proposed by Acar and Rais-Rohani (2009). Three typical surrogate models: PRS, RBF and KRG (the detailed construction and tuning processes of these three surrogate models can be seen in Appendix A) are all used as the components of each ensemble model in this article and are also compared with the ensemble models. We use these surrogate models because they are commonly used by practitioners and they can represent different parametric and nonparametric approaches (Queipo et al. 2005).

Acar has studied the effect of several error metrics on ensemble of surrogate models (2015). Inspired by his work, four kinds of error metrics are used to evaluate the performances of different models: root mean squared error (RMSE) which evaluates the degree of deviation between the prediction value and the true response value over the entire design space, average absolute error (AAE) which ensures the positive and negative errors will not offset, maximum absolute error (MAE) which shows the maximum error within the whole design domain and coefficient of variation (COV) values which measure the dispersion of RMSE, AAE and MAE.

**Table 4** Result comparison of all-direction propeller bearings design in power flow

	Initial design	PRS	RBF	KRG	OM	EG	SP	ES-HGL
Power flow ( $P^*$ )	29.6499	<b>22.6279</b>	23.4511	25.5421	22.5322	22.4472	22.4614	<b>22.2021</b>

For ease of comparison, the best results of ensemble models and individual surrogate models are shown in bold respectively (the lowest value for RMSE, AAE and MAE).



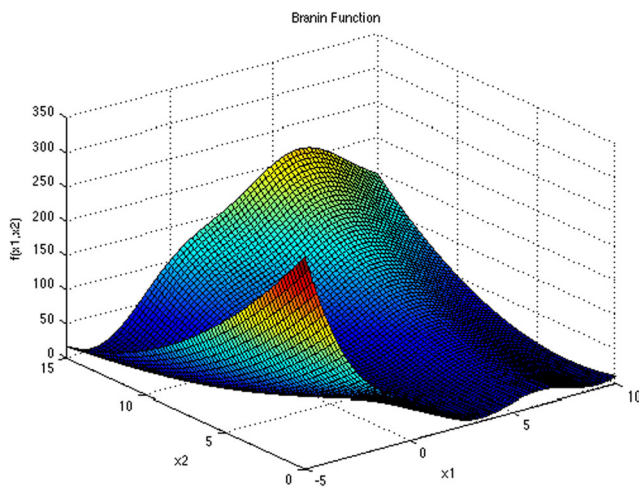


Fig. 6 The landscape of Branin-Hoo function

$$\begin{aligned}
 RMSE &= \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \\
 AAE &= \frac{\sum_{i=1}^N |y_i - \hat{y}_i|}{N} \\
 MAE &= \max |y_i - \hat{y}_i| \\
 COV &= \frac{std}{mean}
 \end{aligned}
 \tag{25}$$

In these four definitions above,  $N$  is the number of test points,  $std$  denotes the sample standard deviation and  $mean$  denotes the mean value.

### 4.1 Numerical examples

Six well-known numerical examples varying from 2-D to 12-D are chosen from previous works (Dixon and Szegö 1978; Goel et al. 2007; Acar 2010) to test the performance of ES-HGL

model: (1) Branin-Hoo function (2-D); (2) Camelback function (2-D); (3) and (4) are Hartman functions (3-D and 6-D); (5) Extended-Rosenbrock function (9-D); (6) Dixon-Price function (12-D). Description of these test functions can be seen in Appendix B.

In order to guarantee the accuracies of the constructed ensemble models and individual surrogate models, the number of sample points is set as twice the number of the coefficients in a full quadratic PRS (Acar 2010). Latin Hypercube Sampling (LHS; McKay et al. 1979) with good space-filling quality is used to generate the sample and test sets by the MATLAB® routine “lhsdesign” and the “maximin” criterion with a maximum of 100 iterations. To take the cost for constructing surrogate models into account and to make a full comparison, we randomly select an appropriate quantity of test sets to eliminate the effect brought by certain distribution of sample points and test points. The summary of the training and test point sets used in each problem is provided in Table 1.

For ease of comparison, the best results of ensemble models and individual surrogate models are shown in bold respectively (the lowest value for  $RMSE$ ,  $AAE$ ,  $MAE$  and  $COV$ ).

From Table 2 we can see that no individual surrogate model is always accurate for different test functions while RBF is relatively better than KRG, and PRS is the worst one. On the contrary, ensemble models perform better than most of the individual surrogate models in most cases. ES-HGL outperforms in most of the error metrics for all six numerical examples, and is still among the top three models in situations when the results of ES-HGL are not the best. Because the advantages of both global and local measures are combined. The COV values of error metrics ( $RMSE$ ,  $AAE$  and  $MAE$ ) and the error distributions in boxplots (see Fig. 8, 9 and 10) indicate that the ES-HGL is robust, because it has small COV values and the error distributions are stable. On the other hand, the performances of the other three ensemble models and the three individual surrogate models vary apparently

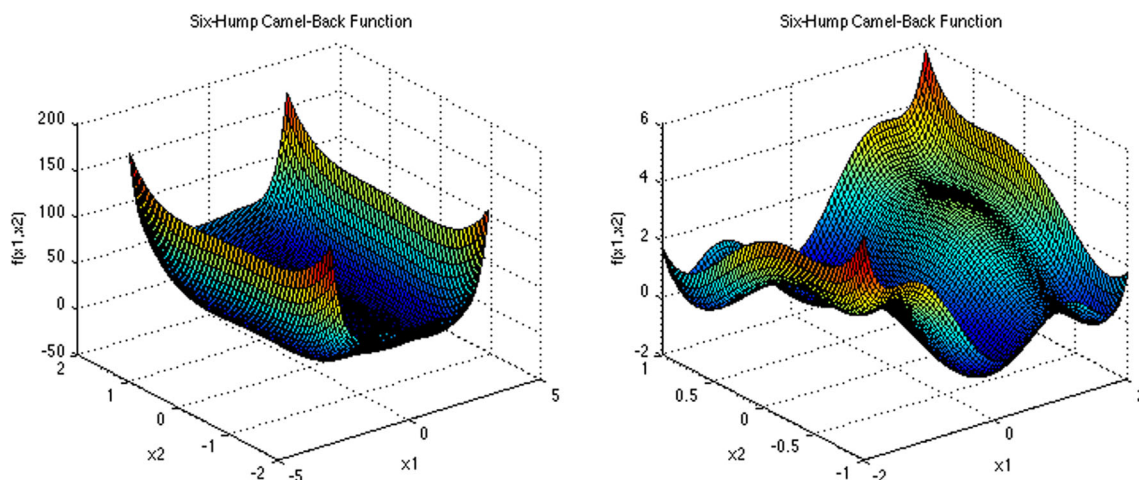


Fig. 7 The landscape of Camelback function

with different test numerical examples. It means that without losing accuracy, ES-HGL can offer more reliable approximations for problems with varying degrees of complexity and dimension. In general, the comparison results show that it is worth implementing the ensemble methods, and the proposed ES-HGL model provides more satisfactory robustness and accuracy under the four metrics for the six test functions in this article. But we should still suggest that ensemble method be used as an “insurance” rather than offering significant improvement.

#### 4.2 NC machine beam design problem

A super heavy NC vertical lathe machine weighs over one thousand tons. It can process the workpieces with a maximum diameter of 28 m and maximum height of 13 m. The major part of this NC machine consists of beam, columns, slides, toolposts and other parts which are shown in Fig. 3(a). Among these parts, the beam is the main moving part that has a welded irregular triangular structure. It can conduct vertical feeding along the guide rail of the column, as well as provide support to the vertical toolpost and vertical slide. Two horizontal guide rails are located on the upper and lower side of the beam that can guide the vertical slide moving along the horizontal direction. Due to the great mass of the vertical slide and vertical toolpost, large weight is imposed on the guide rails of the beam. In addition, the beam is also affected by its large self weight and the cutting force of the vertical toolpost. Under these complex machining situations, the beam is easily deformed along the negative Z-direction, which has a great impact on the machining accuracy.

The simplified structure of the beam is mainly controlled by twelve variables which are shown in Fig. 4: the thicknesses of the plate ( $x_1, x_2$  and  $x_3$ ), the total length of the beam ( $x_4$ ), the widths of the front and rear end face ( $x_5$  and  $x_6$ ), the sizes of the tail ( $x_7$  and  $x_8$ ), the thickness of the rib plates ( $x_9$ ), the width of the beam ( $x_{10}$ ) and the geometry parameters of the rectangular holes ( $x_{11}$  and  $x_{12}$ ).

To reveal the relationship between the deformation along the negative Z-direction and the relative design variables, finite element analysis (FEA) simulations are implemented. Considering that one simulation for this beam needs great computing time, the ES-HGL model is constructed to evaluate the deflection of the negative Z-direction, which is the most representative indicator of load effect imposed on the beam. Three existing ensemble methods and three individual surrogate models are used to make a comparison. 150 sample points are selected for model construction, and 50 test points are adopted to evaluate the performances of each model.

The results of the test are given in Table 3. ES-HGL is the best in RMSE and AAE, meanwhile the second in MAE. The performance of SP is better than RBF and KRG, but worse than the best individual surrogate model PRS. And EG

performs not well than the other three ensemble methods in this engineering case. It reveals that when confronted with a black-box problem with high dimensions, using individual surrogate model to approximate the unknown design space has a risk to obtain inaccurate results. However, these great inaccuracies of individual surrogates do not affect the approximate ability of ES-HGL. Obviously, ES-HGL is a promising ensemble modeling method dealing with engineering problem when the sampling cost is high and a reliable model is needed.

#### 4.3 Optimal design of bearings for an all-direction propeller

In this section, an optimal design of bearings for an all-direction propeller is used to demonstrate the superiority of ES-HGL from another aspect, exploring the ability to find the optimal solution when confronted with a complex engineering problem.

Because of the tough environment of ocean work, marine equipment is required to have high positioning accuracy, good mobility and high stability. The all-direction propeller, as a core dynamic positioning system, is widely used in drilling platform and large ships. Different models of the propeller studied in this article can be seen in Fig. 5. In the design of a propeller, the vibration resist property is important. Because large vibration will reduce the life of the propeller and deteriorate the service performance. Power flow which combines the effect of response speed and power, can give an absolute measure of vibration transmission. So this physical quantity is adopted here to evaluate the vibration characteristic. The power flow with ten structural parameters of shafting system is utilized as the optimization objective to obtain a better dynamic performance of the propeller. FEA simulations are used to build the relationship of several significant structural parameters and the power flow of the propeller. Considering that these simulations are time-consuming (21 mins for one simulation with Intel i3–2120 3.30GHz CPU and 4 GB RAM), we regard this simulation as a black-box problem.

The optimization problem can be summarized as:

$$\begin{aligned} \text{Find } K &= [k_1, k_2, k_3, k_4, k_5, c_1, c_2, c_3, c_4, c_5]^T \\ \min P^* &= \sqrt{\sum_{i=1}^n P_i^2} \quad n = 5 \\ P_i &= \frac{\omega}{2\pi} \int_0^{2\pi} \text{Re}(f_i) \text{Re}(v_i) dt \\ \text{s.t. } k_i^L &\leq k_i \leq k_i^U, c_i^L \leq c_i \leq c_i^U \quad i = 1, 2, 3, 4, 5 \end{aligned} \quad (26)$$

where  $k_i$  and  $c_i$  are the stiffness and damping coefficients of the  $i^{\text{th}}$  bearing,  $k_i^L$  and  $k_i^U$  are the lower and upper boundaries of  $k_i$ ,  $c_i^L$  and  $c_i^U$  are the lower and upper boundaries of  $c_i$  and  $P_i$  is the corresponding power flow of the  $i^{\text{th}}$  bearing. The

optimum stiffness and damping coefficients of each bearing need to be found at the minimum value of power flow.

Three other existing ensemble models (EG, SP and OM) and three individual surrogate models (PRS, RBF and KRG) are also adopted to make a comparison. For its good global search ability, the GA (Genetic Algorithm) toolbox of MATLAB is used to find the optimal solution of each model. After implementing one hundred FEA simulations and constructing the above mentioned four ensemble methods and three individual surrogate models, the optimized parameters are obtained by running the GA procedure.

Then we run FEA simulations with specified initial parameters (the mid-range of each variable) and these optimized parameters respectively for comparison. The result can be seen in Table 4. From the result we can see that all of the solutions found by the ensemble models and individual surrogate models are proved to be feasible. ES-HGL presents the best result which reduces 25.12% compared to the initial objective value 29.6499. EG and SP perform nearly the same as the best individual surrogate model PRS with a decrease of around 24%. KRG is the most inaccurate individual model which only reduces 13.85% on the basis of the initial objective value. The result shows that ensemble models perform better than the combined individual surrogate models. It explains the necessity of ensemble methods again. Above all, the successful application of ES-HGL model indicates that the proposed new ensemble model has a relative better accuracy and robustness than other individual surrogate models and ensemble models compared in this article.

## 5 Conclusions

In this article, a new method which combines the advantages of both global and local measures is proposed to construct a better ensemble model in the cases that only a small number of sample points are available. In this method, design space is divided into two parts: the region far from the sample points and the region near to the sample points. Two strategies are introduced to evaluate weight factors in these two different regions respectively: (1) in the outer region, the Hybrid Weight Factor is adopted, which is composed of global weight factor and local weight factor; (2) in the inner region, for its good approximating ability around sample points, the local weight factor is used to amend the Hybrid Weight Factor according to the distance between the prediction point and the nearest sample point. Six numerical functions, a 12-D NC machine beam design problem and a design optimization problem for the bearings of an all-direction propeller are used to test the proposed ES-HGL method. Three other ensemble models and three individual surrogate models are adopted to make comparisons with ES-HGL. The results show that ES-HGL model can provide more robust and accurate

approximations within limited sample points, while spending almost the equivalent modeling time compared with the combined ensemble models in this article.

**Acknowledgments** Financial support from the National Natural Science Foundation of China under Grant No. 51675198, 973 National Basic Research Program of China under Grant No. 2014CB046705 and National Natural Science Foundation of China under Grant No. 51421062 are gratefully acknowledged.

## Appendix A: Description of the selected surrogate models

In this appendix, a brief overview of the mathematical formulations of PRS, RBF and KRG surrogate models is provided.

### A.1 Polynomial response surface (PRS)

The PRS approximation is one of the most well-established surrogate models. The most commonly used PRS model is the following second-order form

$$\hat{f}(x) = \beta_0 + \sum_{i=1}^d \beta_i x_i + \sum_{i=1}^d \sum_{j \geq i}^d \beta_{ij} x_i x_j \quad (27)$$

where  $d$  is the number of design variables,  $\beta_0$ ,  $\beta_i$  and  $\beta_{ij}$  are the unknown coefficients to be determined by the least squares technique. Here we run the MATLAB® routine “pinv” to obtain the Moore-Penrose generalized inverse matrix of the unknown coefficients, which was proved to be the optimal least square solution by Penrose (1955).

### A.2 Radial basis function (RBF)

RBF models were originally developed to approximate multivariate functions. The general form of the RBF approximation can be expressed as

$$\hat{f}(x) = \sum_{i=1}^n w_i \varphi(\|x - x_i\|) \quad (28)$$

where  $n$  denotes the number of sample points,  $w_i$  are the unknown coefficients to be determined,  $\|\cdot\|$  represents the Euclidean norm and  $\varphi(\cdot)$  is the so-called basis function. Powell (1987) suggested several forms of the basis function  $\varphi(\cdot)$ :

- Gaussian  $\varphi(r) = e^{-\frac{r^2}{2\sigma^2}}$
- ultriquadric  $\varphi(r) = \sqrt{r^2 + \sigma^2}$
- nverse Multiquadric  $\varphi(r) = 1/\sqrt{r^2 + \sigma^2}$
- Thin-Plate Spline  $\varphi(r) = r^2 \ln(r)$

where  $\sigma \geq 0$ . In this article we use the multiquadric basis function with  $\sigma = 1$  (suggested by Acar and Rais-Rohani (2009)),

for its prediction accuracy and convergence ability with increased sample points. In order to obtain the unknown coefficients  $w_i$ , we substitute the  $n$  sample points into the (28) to form an equation as

$$y = \Phi \cdot w \tag{29}$$

where  $y$  is the vector of sample responses and  $\Phi$  is an  $n \times n$  matrix of basis functions. The coefficients vector  $w$  is obtained by solving (29).

### A.3 Kriging (KRG)

The basic assumption of KRG is the estimation of the response in the form

$$Y(x) = \mu(x) + Z(x) \tag{30}$$

where the response  $Y$  consists of a known polynomial  $\mu(x)$  which globally approximates the trend of the function and a stochastic component  $Z(x)$  which generates deviations, so that the Kriging model interpolates the sample points. The correlation between the random variables  $Y(x^{(i)})$  and  $Y(x^{(j)})$  is given by

$$Corr[Y(x^{(i)}), Y(x^{(j)})] = \exp\left(-\sum_l^d \theta_l |x_l^{(i)} - x_l^{(j)}|^{P_l}\right) \tag{31}$$

where  $d$  is the number of design variables,  $\theta_l$  and  $P_l(l=1, \dots, d)$  are unknown parameters to be estimated. Here we only consider a constant term to represent the mean of the overall surface (the Ordinary Kriging) and fix the parameters  $P_l=2(l=1, \dots, d)$  (the stationary Gaussian correlation function case). Then we search the optimal  $\theta_l$  in the range of  $[10^{-3}, 10^2]$  (suggested by Forrester et al. (2008)) with the GA (Genetic Algorithm) toolbox of MATLAB®.

Once the correlation function has been selected, the response is predicted as

$$\hat{y} = \hat{\mu} + r^T R^{-1} (y - \mathbf{1}\hat{\mu}) \tag{32}$$

where the matrix  $R^{-1}$  is the inverse of the correlation matrix  $R$  whose element  $R_{ij}$  is equal to the (31),  $y$  is the vector of sample responses and  $\mathbf{1}$  represents an  $n \times 1$  vector of ones. The estimated value of  $\hat{\mu}$  and the expressions of  $r^T$  are

$$\hat{\mu} = \frac{\mathbf{1}^T R^{-1} y}{\mathbf{1}^T R^{-1} \mathbf{1}} \tag{33}$$

$$r^T = \left( Corr[Y(x^{(1)}), Y(x)] \cdots Corr[Y(x^{(n)}), Y(x)] \right) \tag{34}$$

Detailed derivation of Kriging can be found in Jones (2001) and Forrester et al. (2008)

## Appendix B: Description of the numerical test functions

In this appendix, the description of six numerical test functions is provided. The landscapes of two-variable functions are depicted in Figs. 6 and 7.

### B.1 Branin-Hoo function

$$f(x) = \left(x_2 - \frac{5.1x_1^2}{4\pi^2} + \frac{5x_1}{\pi} - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos(x_1) + 10 \tag{35}$$

where  $x_1 \in [-5, 10]$  and  $x_2 \in [0, 15]$ .

### B.2 Camelback function

$$f(x) = \left(4 - 2.1x_1^2 + \frac{x_1^4}{3}\right)x_1^2 + x_1x_2 + (-4 + 4x_2^2)x_2^2 \tag{36}$$

where  $x_1 \in [-2, 2]$  and  $x_2 \in [-2, 2]$ .

### B.3 and B.4 Hartman functions

$$f(x) = -\sum_{i=1}^4 c_i \exp\left[-\sum_{j=1}^n a_{ij} (x_j - p_{ij})^2\right] \tag{37}$$

where  $x_i \in [0, 1]$ . Two types of Hartman functions are given based on different number of input variables: (1) Hartman-3 with three input variables (test function 3), and (2) Hartman-6 with six input variables (test function 4). While the parameter  $c$  in each function is the same vector  $[1 \ 1.2 \ 3 \ 3.2]^T$ , the other two parameters  $a$  and  $p$  are shown in Table 5 and 6.

### B.5 Extended-Rosenbrock function

$$f(x) = \sum_{i=1}^{m-1} \left[ (1-x_i)^2 + 100(x_{i+1}-x_i^2)^2 \right] \tag{38}$$

where  $x_i \in [-5, 10], i = 1, 2, \dots, m = 9$ .

**Table 5** Parameters used in Hartman-3 function,  $j = 1, 2, 3$

$a_{ij}$	$p_{ij}$				
3.0	10	30	0.3689	0.1170	0.2673
0.1	10	35	0.4699	0.4387	0.7470
3.0	10	30	0.1091	0.8732	0.5547
0.1	10	35	0.03815	0.5743	0.8828



**Table 6** Parameters used in Hartman-6 function,  $j = 1, 2, \dots, 6$

$a_{ij}$		$P_{ij}$										
10	3.0	17.0	3.5	1.7	8.0	0.1312	0.1696	0.5569	0.0124	0.8283	0.5886	
0.05	10.0	17.0	0.1	8.0	14.0	0.2329	0.4135	0.8307	0.3736	0.1004	0.9991	
3.0	3.5	1.7	10.0	17.0	8.0	0.2348	0.1451	0.3522	0.2883	0.3047	0.6650	
17.0	8.0	0.05	10.0	0.1	14.0	0.4047	0.8828	0.8732	0.5743	0.1091	0.0381	

**B.6 Dixon-Price function**

$$f(x) = (x_1 - 1)^2 + \sum_{i=2}^m i [2x_i^2 - x_{i-1}]^2 \tag{39}$$

$x_i \in [-10, 10], i = 1, 2, \dots, m = 12.$  (where)

**Appendix C: Test results for determining the form of ES-HGL**

In this appendix, the test results referenced in Section 3.3 are provided in Table 7, 8 and 9.

**Table 7** Test results of different region radius forms

	$r_k^1(P_k)$	$r_k^2(P_k)$
<b>Branin-Hoo</b>		
RMSE/COV	23.97/0.24	23.98/0.24
AAE/COV	15.40/0.21	15.41/0.21
MAE/COV	132.27/0.35	134.17/0.35
<b>Camelback</b>		
RMSE/COV	17.51/0.13	17.61/0.14
AAE/COV	11.34/0.11	11.48/0.13
MAE/COV	80.90/0.19	82.11/0.19
<b>Hartman-3</b>		
RMSE/COV	0.49/0.22	0.49/0.22
AAE/COV	0.33/0.20	0.34/0.21
MAE/COV	2.03/0.29	2.06/0.31
<b>Hartman-6</b>		
RMSE/COV	0.30/0.10	0.31/0.12
AAE/COV	0.19/0.13	0.21/0.15
MAE/COV	1.85/0.17	1.85/0.17
<b>Extended-Rosenbrock-9</b>		
RMSE/COV	$7.64 \cdot 10^4 / 0.62$	$7.69 \cdot 10^4 / 0.77$
AAE/COV	$7.59 \cdot 10^4 / 0.63$	$7.66 \cdot 10^4 / 0.78$
MAE/COV	$1.33 \cdot 10^5 / 0.36$	$1.33 \cdot 10^5 / 0.36$
<b>Dixon-Price-12</b>		
RMSE/COV	$3.35 \cdot 10^5 / 0.60$	$3.73 \cdot 10^5 / 0.70$
AAE/COV	$3.28 \cdot 10^5 / 0.63$	$3.64 \cdot 10^5 / 0.74$
MAE/COV	$5.99 \cdot 10^5 / 0.37$	$6.74 \cdot 10^5 / 0.33$
<b>NC machine beam design</b>		
RMSE	0.2328	0.2355
AAE	0.1571	0.1572
MAE	1.0465	1.0477
<b>All-direction propeller bearings design</b>		
Power flow ( $P^*$ )	22.2021	22.3469



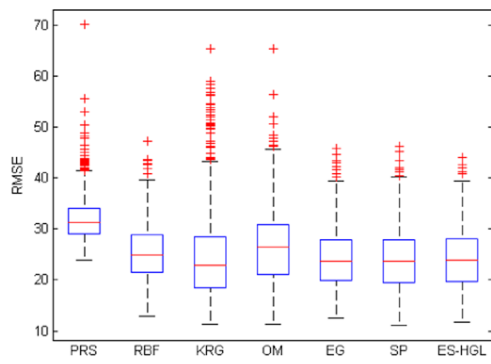
**Table 8** Test results of different HWF forms

	$w_1^H$	$w_2^H$	$w_3^H$	$w_4^H$
Branin-Hoo				
RMSE/COV	23.97/0.24	25.46/0.24	24.87/0.22	24.20/0.23
AAE/COV	15.40/0.21	16.26/0.19	16.08/0.19	15.45/0.21
MAE/COV	132.27/0.35	146.37/0.33	138.72/0.32	136.06/0.34
Camelback				
RMSE/COV	17.51/0.13	17.96/0.13	18.43/0.12	17.56/0.14
AAE/COV	11.34/0.11	11.35/0.12	11.43/0.13	11.44/0.13
MAE/COV	80.90/0.19	86.36/0.17	90.31/0.15	81.24/0.19
Hartman-3				
RMSE/COV	0.49/0.22	0.50/0.24	0.52/0.19	0.49/0.24
AAE/COV	0.33/0.20	0.34/0.20	0.34/0.19	0.34/0.20
MAE/COV	2.03/0.29	2.09/0.34	2.08/0.23	2.08/0.34
Hartman-6				
RMSE/COV	0.30/0.10	0.31/0.12	0.31/0.11	0.31/0.12
AAE/COV	0.19/0.13	0.21/0.15	0.21/0.15	0.21/0.15
MAE/COV	1.85/0.17	1.93/0.18	1.86/0.17	1.85/0.17
Extended-Rosenbrock-9				
RMSE/COV	$7.64 \cdot 10^4/0.62$	$7.69 \cdot 10^4/0.77$	$7.69 \cdot 10^4/0.77$	$7.69 \cdot 10^4/0.77$
AAE/COV	$7.59 \cdot 10^4/0.63$	$7.66 \cdot 10^4/0.78$	$7.66 \cdot 10^4/0.78$	$7.66 \cdot 10^4/0.78$
MAE/COV	$1.33 \cdot 10^5/0.36$	$1.36 \cdot 10^5/0.32$	$1.33 \cdot 10^5/0.36$	$1.33 \cdot 10^5/0.36$
Dixon-Price-12				
RMSE/COV	$3.35 \cdot 10^5/0.60$	$3.73 \cdot 10^5/0.70$	$3.73 \cdot 10^5/0.70$	$3.73 \cdot 10^5/0.70$
AAE/COV	$3.28 \cdot 10^5/0.63$	$3.64 \cdot 10^5/0.74$	$3.64 \cdot 10^5/0.74$	$3.64 \cdot 10^5/0.74$
MAE/COV	$5.99 \cdot 10^5/0.37$	$6.74 \cdot 10^5/0.33$	$6.74 \cdot 10^5/0.33$	$6.74 \cdot 10^5/0.33$
NC machine beam design				
RMSE	0.2328	0.2335	0.2338	0.2335
AAE	0.1571	0.1577	0.1583	0.1577
MAE	1.0465	1.0512	1.0509	1.0499
All-direction propeller bearings design				
Power flow (P*)	22.2021	22.2130	22.2128	22.2133

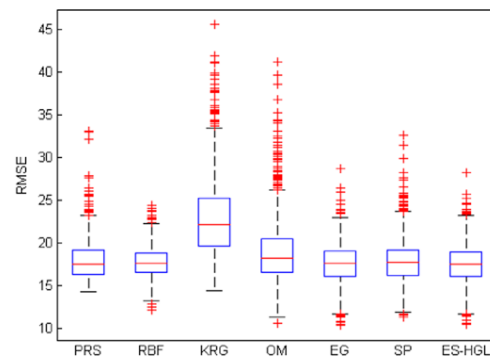
**Table 9** Test result for evaluating the effect of the hybrid method

	Eq. (8)	Eq. (22)
Branin-Hoo		
RMSE/COV	24.98/0.26	23.97/0.24
AAE/COV	16.41/0.24	15.40/0.21
MAE/COV	134.17/0.37	132.27/0.35
Camelback		
RMSE/COV	17.61/0.14	17.51/0.13
AAE/COV	11.48/0.13	11.34/0.11
MAE/COV	82.11/0.19	80.90/0.19
Hartman-3		
RMSE/COV	0.51/0.22	0.49/0.22
AAE/COV	0.34/0.21	0.33/0.20
MAE/COV	2.11/0.31	2.03/0.29
Hartman-6		
RMSE/COV	0.32/0.12	0.30/0.10
AAE/COV	0.21/0.15	0.19/0.13
MAE/COV	1.85/0.17	1.85/0.17
Extended-Rosenbrock-9		
RMSE/COV	$7.69 \times 10^4/0.77$	$7.64 \times 10^4/0.62$
AAE/COV	$7.66 \times 10^4/0.78$	$7.59 \times 10^4/0.63$
MAE/COV	$1.33 \times 10^5/0.36$	$1.33 \times 10^5/0.36$
Dixon-Price-12		
RMSE/COV	$3.73 \times 10^5/0.70$	$3.35 \times 10^5/0.60$
AAE/COV	$3.64 \times 10^5/0.74$	$3.28 \times 10^5/0.63$
MAE/COV	$6.74 \times 10^5/0.33$	$5.99 \times 10^5/0.37$
NC machine beam design		
RMSE	0.2379	0.2328
AAE	0.1590	0.1571
MAE	1.0477	1.0465
All-direction propeller bearings design		
Power flow ( $P^*$ )	22.3346	22.2021

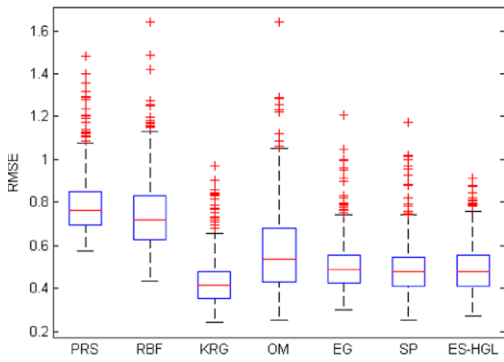
**C.1 Test result for determining the form of the region radius****C.2 Test result for determining the form of the Hybrid Weight Factor (HWF)****C.3 Test result for evaluating the effect of the hybrid method****Appendix D: Boxplots for six numerical examples (Figures 8, 9, and 10)**



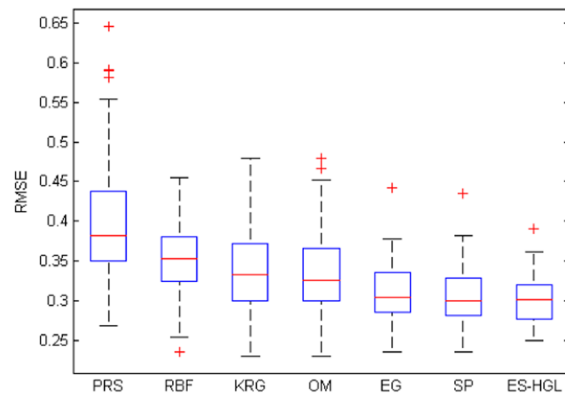
(a) Branin-Hoo



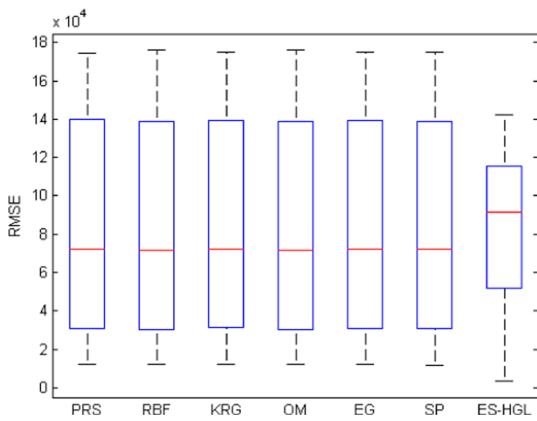
(b) Camelback



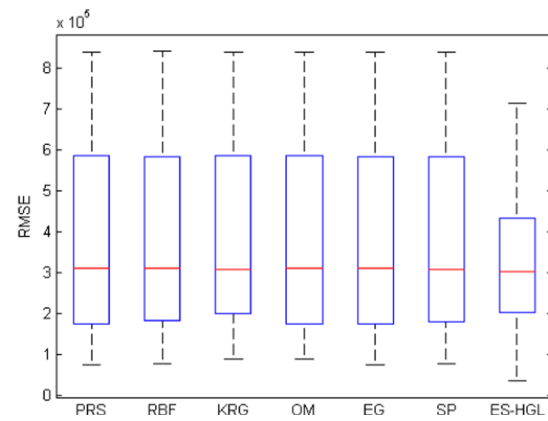
(c) Hartman-3



(d) Hartman-6

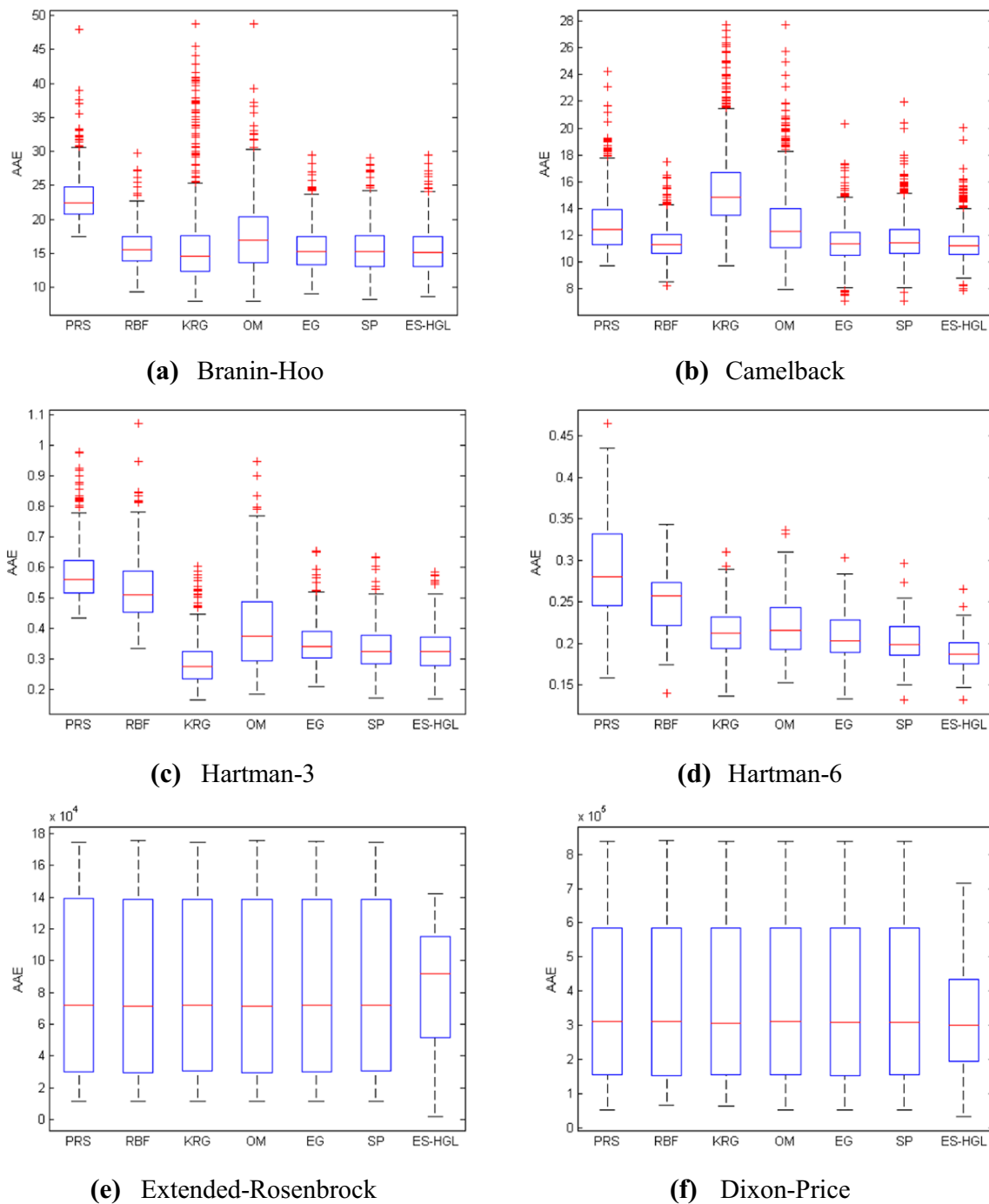


(e) Extended-Rosenbrock

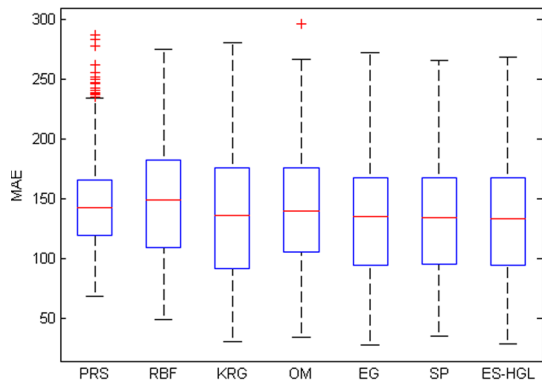


(f) Dixon-Price

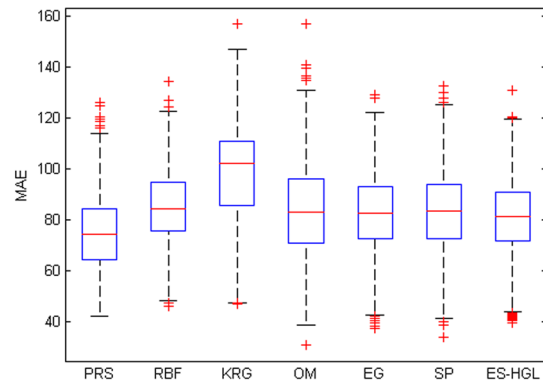
Fig. 8 Boxplots of RMSE for six numerical examples



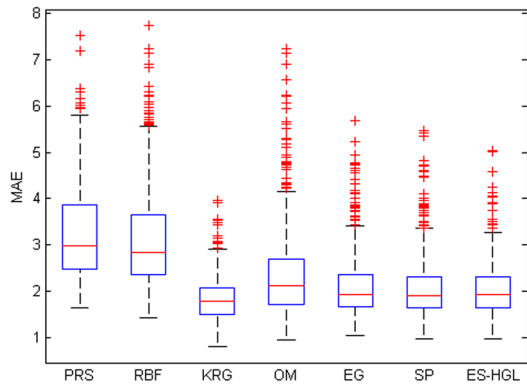
**Fig. 9** Boxplots of AAE for six numerical examples



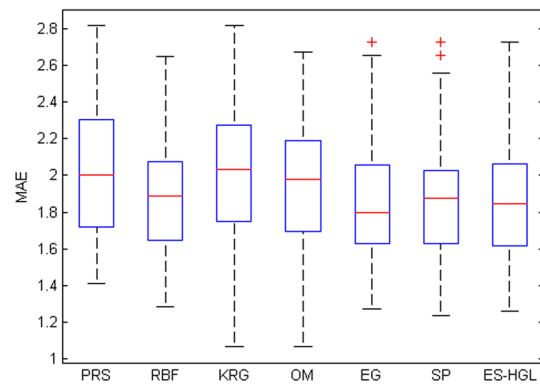
(a) Branin-Hoo



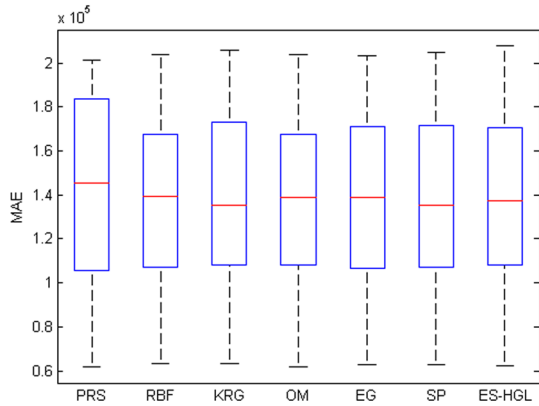
(b) Camelback



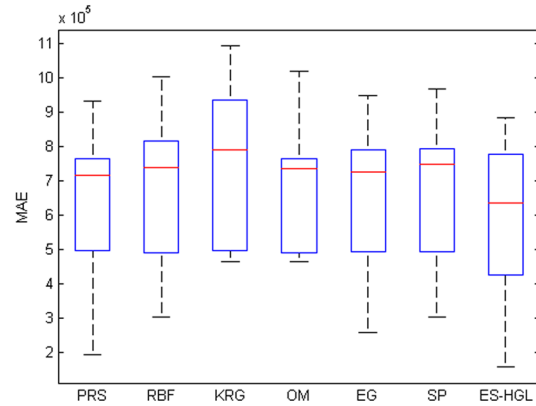
(c) Hartman-3



(d) Hartman-6



(e) Extended-Rosenbrock



(f) Dixon-Price

Fig. 10 Boxplots of MAE for six numerical examples



## References

- Acar E, Rais-Rohani M (2009) Ensemble of metamodels with optimized weight factors. *Struct Multidiscip Optim* 37(3):279–294
- Acar E (2010) Various approaches for constructing an ensemble of metamodels using local measures. *Struct Multidiscip Optim* 42(6):879–896
- Acar E (2015) Effect of error metrics on optimum weight factor selection for ensemble of metamodels. *Expert Syst Appl* 42(5):2703–2709
- Bishop CM (1995) *Neural networks for pattern recognition*. Oxford university press, Oxford
- Box GE, Draper NR (1987) *Empirical model-building and response surfaces*, vol 424. Wiley, New York
- Buckland ST, Burnham KP, Augustin NH (1997) Model selection: an integral part of inference. *Biometrics* 53:603–618
- Cherkassky V, Shao X, Mulier FM, Vapnik VN (1999) Model complexity control for regression using VC generalization bounds. *IEEE Trans Neural Netw* 10(5):1075–1089
- Dixon LCW, Szegő GP (eds) (1978) *Towards global optimisation*. North-Holland, Amsterdam
- Forrester AI, Keane AJ (2009) Recent advances in surrogate-based optimization. *Prog Aerosp Sci* 45(1):50–79
- Forrester A, Sobester A, Keane A (2008) *Engineering design via surrogate modelling: a practical guide*. John Wiley & Sons, Chichester
- Goel T, Haftka RT, Shyy W, Queipo NV (2007) Ensemble of surrogates. *Struct Multidiscip Optim* 33(3):199–216
- Hardy RL (1971) Multiquadric equations of topography and other irregular surfaces. *J Geophys Res* 76(8):1905–1915
- Hoeting JA, Madigan D, Raftery AE, Volinsky CT (1999) Bayesian model averaging: a tutorial. *Stat Sci*:382–401
- Jones DR (2001) A taxonomy of global optimization methods based on response surfaces. *J Glob Optim* 21(4):345–383
- Kass RE, Raftery AE (1995) Bayes factors. *J Am Stat Assoc* 90(430):773–795
- Madigan D, Raftery AE (1994) Model selection and accounting for model uncertainty in graphical models using Occam's window. *J Am Stat Assoc* 89(428):1535–1546
- McKay MD, Beckman RJ, Conover WJ (1979) Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* 21(2):239–245
- Penrose R (1955) A generalized inverse for matrices. *Math Proc Camb Philos Soc* 51(03):406–413 Cambridge University Press
- Powell MJD (1987) Radial basis functions for multivariable interpolation: a review. In: Mason JC, Cox MG (eds) *Proceedings of the IMA conference on algorithms for the approximation of functions and data*, Oxford University Press, London, pp 143–167
- Queipo NV, Haftka RT, Shyy W, Goel T, Vaidyanathan R, Tucker PK (2005) Surrogate-based analysis and optimization. *Prog Aerosp Sci* 41(1):1–28
- Rennie JDM (2005) Volume of the n-sphere. Retrieved in April 2017, from <http://people.csail.mit.edu/jrennie/writing/sphereVolume.pdf>
- Sacks J, Welch WJ, Mitchell TJ, Wynn HP (1989) Design and analysis of computer experiments. *Stat Sci* 4:409–423
- Sanchez E, Pintos S, Queipo NV (2008) Toward an optimal ensemble of kernel-based approximations with engineering applications. *Struct Multidiscip Optim* 36(3):247–261
- Smola AJ, Schölkopf B (2004) A tutorial on support vector regression. *Stat Comput* 14(3):199–222
- Viana FA, Haftka RT, Steffen V (2009) Multiple surrogates: how cross-validation errors can help us to obtain the best predictor. *Struct Multidiscip Optim* 39(4):439–457
- Zerpa LE, Queipo NV, Pintos S, Salager JL (2005) An optimization methodology of alkaline–surfactant–polymer flooding processes using field scale numerical simulation and multiple surrogates. *J Pet Sci Eng* 47(3):197–208
- Zhou XJ, Ma YZ, Li XF (2011) Ensemble of surrogates with recursive arithmetic average. *Struct Multidiscip Optim* 44(5):651–671