

Article

Selecting the Best Quantity and Variety of Surrogates for an Ensemble Model

Pengcheng Ye ^{1,2,*} and Guang Pan ^{1,2}

¹ School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China; panguang@nwpu.edu.cn

² Key Laboratory for Unmanned Underwater Vehicle, Northwestern Polytechnical University, Xi'an 710072, China

* Correspondence: yepengcheng@nwpu.edu.cn

Received: 27 August 2020; Accepted: 30 September 2020; Published: 7 October 2020



Abstract: Surrogate modeling techniques are widely used to replace the computationally expensive black-box functions in engineering. As a combination of individual surrogate models, an ensemble of surrogates is preferred due to its strong robustness. However, how to select the best quantity and variety of surrogates for an ensemble has always been a challenging task. In this work, five popular surrogate modeling techniques including polynomial response surface (PRS), radial basis functions (RBF), kriging (KRG), Gaussian process (GP) and linear shepard (SHEP) are considered as the basic surrogate models, resulting in twenty-six ensemble models by using a previously presented weights selection method. The best ensemble model is expected to be found by comparative studies on prediction accuracy and robustness. By testing eight mathematical problems and two engineering examples, we found that: (1) in general, using as many accurate surrogates as possible to construct ensemble models will improve the prediction performance and (2) ensemble models can be used as an insurance rather than offering significant improvements. Moreover, the ensemble of three surrogates PRS, RBF and KRG is preferred based on the prediction performance. The results provide engineering practitioners with guidance on the superior choice of the quantity and variety of surrogates for an ensemble.

Keywords: ensemble of surrogates; surrogate models; prediction accuracy; robustness

1. Introduction

Surrogate models, also called metamodels, utilize interpolation and regression methods to approximate the computation-intensive black-box functions. They have attracted more attention in recent decades. Researchers are expecting to develop a best surrogate modeling technique for all engineering applications. Therefore, many comparative studies on surrogate modeling techniques under multiple modeling criteria are presented. Jin et al. [1] provide the first systematic comparative report on the performances of four surrogates polynomial response surface (PRS), radial basis functions (RBF), kriging (KRG) and multivariate adaptive regression splines (MARS) based on multiple performance criteria and multiple test problems. The results show that these four surrogate modeling techniques have both advantages and disadvantages. Mullur [2] proposes an extended radial basis function (E-RBF), which offers more flexibility than the typical radial basis functions (RBF). To further understand the advantages and limitations of this new surrogate modeling technique, Mullur and Messac [3] compared the performances of E-RBF to that of PRS, RBF, KRG. The E-RBF is found to outperform the other surrogate models. Zhao and Xue [4] observed the relationships between the sample quality merits and the performance measures of the PRS, RBF, KRG and Bayesian neural network (BNN) models. They provide simple guidelines to select the candidate surrogate models

according to the sample quality merits and the performance measures. Gelder et al. [5] carry out a comparative study of five surrogate modeling techniques containing PRS, RBF, KRG, MARS and neural networks (NN). This work contributes to guiding users to select a reliable and time-efficient surrogate modeling technique for building energy simulations and they recommend KRG and NN models. Salem and Tomaso [6] proposed an automatic selection method to determine the most suitable surrogate models for the specific problems. They measured the quality of general surrogate models based on the internal accuracy, predictive performance and a roughness penalty. Keane and Voutchkov [7] compared and contrasted a wide range of surrogate models for aerodynamic section performance modeling. They found that the NN model outperforms many existing surrogate modeling techniques when large quantities of data are available. Kianifar and Campean [8] evaluated the performances of surrogate modeling techniques PRS, RBF, KRG using eighteen test problems and four engineering examples. The results show that KRG performs consistently well across different problems although it can be very time-consuming for large samples. It has been proved that different surrogate modeling techniques with diverse characteristics are suitable for different engineering problems. Most researchers face the difficulties of selecting the appropriate surrogate models for their own engineering applications. Thus, there is a great interest in an intuitive way to combine multiple surrogate models.

Research focusing on the ensembles of surrogates in the literature can be divided into two parts: (1) the measures for evaluating the weight factors, (2) the engineering applications using ensembles of surrogates. According to the measures for evaluating the weight factors, existing ensemble modeling techniques can be generally classified as global measures [9–13], local measures [14–18] and combining global and local measures [19,20]. The weight factors evaluated by global measures remain constant over the entire design space. Goel et al. [9] allocate weight factors for the chosen individual surrogates PRS, RBF, and KRG by using the generalized mean square cross-validation error (GMSE). They state that the ensemble of surrogates can improve the robustness of the predictions by reducing the impact of a poor surrogate model. Acar and Rais-Rohani [10] consider that the selection of weight factors can be directly determined based on the minimization of the certain global error metric. The results show that the optimized ensemble model provides more accurate predictions than the stand-alone surrogates. In contrast, the weight factors evaluated by local measures are pointwise changed with the variation of prediction points. Lee and Choi [15] presented a new pointwise ensemble of surrogates whose weights are calculated by using a v nearest points cross-validation error. The regression models are also suggested as basic surrogates for ensemble construction, it can interpolate the function values at training points by the proposed control function. To take advantages of both global and local measures, Chen et al. [19] divided the design space into two parts, whereas the weight factors are evaluated by using the different strategies, respectively. A large proportion of research indicates that the ensemble of surrogates can provide more accurate and robust results. Therefore, successful engineering applications using ensembles of surrogates are continuously reported [21–26]. Hamza and Saitou [21] presented a multi-scenario co-evolutionary genetic algorithm (MSCGA) for vehicle structural crashworthiness via an ensemble of surrogates. Gu et al. [22] employed an ensemble of surrogates to deal with the reliability-based design optimization for a vehicle occupant protection system. Dhamotharan et al. [23] proposed an ensemble of surrogates based optimization framework for Savonius wind turbine design. Chen and Lu [24] developed a new adaptive approach for reliability analysis by the ensemble learning of multiple competitive surrogate models including KRG, support vector regression (SVR) and polynomial chaos expansion. In all instances, the quantity and variety of individual surrogates in the ensembles are chosen in advance.

This work aims to carry out a systematic and comprehensive study on selecting the best quantity and variety of surrogates for an ensemble. Five classical surrogates including PRS, RBF, KRG, GP and SHEP are used as the basic surrogate models for building twenty-six ensembles whose weight factors are determined by the authors' previous work [27]. The performances of twenty-six ensembles along with five individual surrogates are measured using prediction accuracy and robustness. Thus, guidelines for selecting the appropriate quantity and variety of surrogates for an ensemble are proposed.

2. Research Objectives

The question of how to determine the weight factors for constructing an ensemble of surrogates, attracts most researchers’ attention. Thus, many weights selection methods have been presented [9,10,14,15,27,28]. When forming an ensemble of surrogates, we generally select the basic surrogate models according to past experience and personal preference. However, it’s possible that selecting inappropriate surrogates will result in a loss of accuracy. Moreover, selecting excessive surrogates will lead to a loss of modeling efficiency. The research objective presented in this work, is to carry out a systematic and comprehensive study to answer the following two questions:

- (1) How to select the appropriate variety of surrogates for an ensemble? Do we need to use surrogates that are as accurate as possible?
- (2) How to select the appropriate quantity of surrogates for an ensemble? Do we need to use as many surrogates as possible?

We do not discuss the measures for evaluating the weight factors in this paper. The existing weights selection method presented in the authors’ previous work is used [27]. It will be briefly introduced in the next section.

3. Weights Selection Method

Intelligent selection of weights is important for building a superior ensemble of surrogates. The weights are carefully allocated to improve the overall prediction accuracy of the ensemble. Goel et al. [9] consider that the weights should not only reflect the confidence in the surrogates, but also filter out the adverse effects of surrogates which perform poorly in sampling sparse regions. A weights selection method addressing these two issues is proposed and formulated as:

$$\begin{aligned}
 w_i &= w_i^* \Big/ \sum_{i=1}^m w_i^*, & w_i^* &= (E_i + \alpha \bar{E})^\beta, \\
 \bar{E} &= \frac{1}{m} \sum_{i=1}^m E_i, & \alpha < 1, & \quad \beta < 0
 \end{aligned}
 \tag{1}$$

where w_i is the weight associated with the i th basic surrogate, E_i is the given error measure of the i th basic surrogate, \bar{E} indicates the average value of all surrogates’ error measures. This weighting scheme needs to specify two parameters α and β which control the importance of averaging and individual surrogates, respectively. Goel et al. [9] found that $\alpha = 0.05$ and $\beta = -1$ results in a better ensemble model in their study.

Acar and Rais-Rohani [10] considered that the fixed values of two key parameters α and β will lead to a loss of flexibility. Inspired by the work from Goel et al. [9], Acar and Rais-Rohani [10] selected the weights by solving an optimization problem of the form:

$$\begin{aligned}
 & \text{Find} && w_i \\
 & \min && \varepsilon_e = \text{Err}\{\hat{y}_e(w_i, \hat{y}_i(x_k)), y(x_k)\}, \quad k = 1, 2, \dots, N \\
 & \text{s.t.} && \sum_{i=1}^m w_i(x) = 1
 \end{aligned}
 \tag{2}$$

where $\text{Err}\{\}$ indicates the chosen error metric that measures accuracy of the ensemble model \hat{y}_e . N is the number of training points, $y(x_k)$ denotes the true response value at the k th sample point x_k , $\hat{y}_i(x_k)$ indicates the predicted response of the i th basic surrogate at the k th sample point x_k . It is noticed that the number of optimal parameters in this weights selection method equals to the number of basic surrogates. That is, the more surrogates that are used to construct the ensemble of surrogates, the more computational cost is required. Here, we advise the parameters α and β can be optimized to minimize

the error measure of the ensemble model. We propose a weights selection method through solving the optimization problem formulated as:

$$\begin{aligned}
 & \text{Find} \quad \alpha, \beta \\
 \min \quad & E_e = \sqrt{\frac{1}{N} \sum_{k=1}^N \left(y(x_k) - \sum_{i=1}^m w_i \hat{y}_i^{(-k)}(x_k) \right)^2} \\
 & w_i = w_i^* \sqrt{\sum_{i=1}^m w_i^*}, \quad w_i^* = (E_i + \alpha \bar{E})^\beta, \\
 & \bar{E} = \frac{1}{m} \sum_{i=1}^m E_i, \quad E_i = \sqrt{\frac{1}{N} \sum_{k=1}^N \left(y(x_k) - \hat{y}_i^{(-k)}(x_k) \right)^2} \\
 \text{s.t.} \quad & \alpha < 1, \quad \beta < 0, \quad E_i + \alpha \bar{E} > 0
 \end{aligned} \tag{3}$$

where E_e and E_i represent the error measure, i.e., Prediction sum of squares of root mean square ($PRESS_{RMS}$) of the ensemble model and i th basic surrogate, respectively. $\hat{y}_i^{(-k)}(x_k)$ indicates the predicted response of the i th basic surrogate generated using $N - 1$ training points without the k th sample point x_k (i.e., leave-one-out cross-validation strategy). Here, a more restrictive constraint $E_i + \alpha \bar{E} > 0$ is added to improve the quality of the proposed weights selection method in contrast to the work [9,10]. The MATLAB® function “fmincon” is employed to produce the sequential quadratic programming (SQP) algorithm for solving this optimization problem.

4. Numerical Experiments

4.1. Basic Surrogate Models and Derived Ensemble Models

Five basic surrogate models including PRS, RBF, KRG, GP and SHEP are selected to form twenty-six different ensembles of surrogates. For the sake of simplicity, the selected individual surrogates and formed ensemble models are named as IS1~IS5 and ES1~ES26, respectively. Table 1 provides details about twenty-six ensemble models and five basic surrogates used during the investigation. The SURROGATES toolbox developed by Viana [29] was employed to build five basic surrogate models. The SURROGATES toolbox fits the radial basis function model using the RBF toolbox of Jakabsons [30], fits the KRG model using the DACE toolbox of Lophaven et al. [31], and fits the GP model using the GP toolbox of Forrester [32]. PRS is a typical regression method that has benefits in handling convex problems. The second-order fully quadratic polynomial is used in this work. And the unknown regression coefficients can be determined by the least squares method. The RBF model is more appropriate for predicting the scattered multivariate data. It is composed of multiple radial basis functions. The RBF model is based on the multi-quadric formulation with the constant, $c = 1$. The KRG model is good at dealing with nonlinear problems. It estimates the value of a function as a combination of a known function and unknown departures. A Gaussian correlation function and a constant trend model are used in the KRG model. The GP model provides a flexible model and model-based estimate of the prediction error even if the simulation itself is deterministic. It can also be used when the simulation is stochastic, although this requires an extension of the model. The covariance function in the GP model is selected as the squared exponential function with an automatic relevance determination distance measure. The SHEP model is well suited for interpolating the material density field by using nodal density values. The linear Shepard interpolation scheme is employed to build the SHEP model.

Table 1. Information on the selected individual surrogates and forming ensemble models.

Abbreviation	Surrogates Selected	Abbreviation	Surrogates Selected
IS1	PRS	ES11	PRS-RBF-KRG
IS2	RBF	ES12	PRS-RBF-GP
IS3	KRG	ES13	PRS-RBF-SHEP
IS4	GP	ES14	PRS-KRG-GP
IS5	SHEP	ES15	PRS-KRG-SHEP
ES1	PRS-RBF	ES16	PRS-GP-SHEP
ES2	PRS-KRG	ES17	RBF-KRG-GP
ES3	PRS-GP	ES18	RBF-KRG-SHEP
ES4	PRS-SHEP	ES19	RBF-GP-SHEP
ES5	RBF-KRG	ES20	KRG-GP-SHEP
ES6	RBF-GP	ES21	PRS-RBF-KRG-GP
ES7	RBF-SHEP	ES22	PRS-RBF-KRG-SHEP
ES8	KRG-GP	ES23	PRS-RBF-GP-SHEP
ES9	KRG-SHEP	ES24	PRS-KRG-GP-SHEP
ES10	GP-SHEP	ES25	RBF-KRG-GP-SHEP
/	/	ES26	PRS-RBF-KRG-GP-SHEP

4.2. Test Problems

The prediction capability of surrogate models is tested by eight mathematical problems and two engineering examples from the publications [10,27]. The specific mathematical formulas are summarized as follows

- (1) Branin-Hoo (BH) function with $n = 2$

$$f(x) = [x_2 - 5.1(x_1/2\pi)^2 + (5/\pi)x_1 - 6]^2 + 10[1 - (1/8\pi)] \cos x_1 + 10, \quad x_1 \in [-5, 10], \quad x_2 \in [0, 15] \quad (4)$$

- (2) Camelback function (CB) with $n = 2$

$$f(x) = 4x_1^2 - 2.1x_1^4 + x_1^6/3 + x_1x_2 - 4x_2^2 + 4x_2^4, \quad x_1 \in [-3, 3], \quad x_2 \in [-2, 2] \quad (5)$$

- (3) Goldstein and Price function (GF) with $n = 2$

$$f(x) = [1 + (x_1 + x_2 + 1)^2(19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2)] * [30 + (2x_1 - 3x_2)^2(18 - 32x_1 + 12x_1^2 + 48x_2 - 36x_1x_2 + 27x_2^2)], \quad x \in [-2, 2]^n \quad (6)$$

- (4) Hartman function (HN3) with $n = 3$

$$f(x) = -\sum_{i=1}^4 c_i \exp\left[-\sum_{j=1}^n \alpha_{ij}(x_j - p_{ij})^2\right], \quad x \in [0, 1]^n \quad (7)$$

$$\alpha_{ij} = \begin{bmatrix} 3 & 10 & 30 \\ 0.1 & 10 & 35 \\ 3 & 10 & 30 \\ 0.1 & 10 & 35 \end{bmatrix}, \quad p_{ij} = \begin{bmatrix} 0.3689 & 0.1170 & 0.2673 \\ 0.4699 & 0.4387 & 0.7470 \\ 0.1091 & 0.8732 & 0.5547 \\ 0.03815 & 0.5743 & 0.8828 \end{bmatrix}, \quad c_i = \begin{bmatrix} 1 \\ 1.2 \\ 3 \\ 3.2 \end{bmatrix}$$

- (5) Shekel function (SK) with $n = 4$

$$f(x) = -\sum_{i=1}^5 \left(\sum_{j=1}^n (x_j - a_{ij})^2 + c_i \right)^{-1}, \quad x \in [0, 10]^n \quad (8)$$

$$[a_{ij}] = \begin{bmatrix} 4 & 4 & 4 & 4 \\ 1 & 1 & 1 & 1 \\ 8 & 8 & 8 & 8 \\ 6 & 6 & 6 & 6 \\ 3 & 7 & 3 & 7 \end{bmatrix}, \quad c_i = [0.1 \quad 0.2 \quad 0.2 \quad 0.4 \quad 0.4]$$

(6) Hartman function (HN6) with $n = 6$

$$f(x) = - \sum_{i=1}^4 c_i \exp \left[- \sum_{j=1}^n \alpha_{ij} (x_j - p_{ij})^2 \right], \quad i = 1, 2, \dots, n. \quad c_i = [1 \quad 1.2 \quad 3 \quad 3.2] \quad , \quad x \in [0, 1]^n \quad (9)$$

$$[a_{ij}] = \begin{bmatrix} 10 & 3 & 17 & 3.5 & 1.7 & 8 \\ 0.05 & 10 & 17 & 0.1 & 8 & 14 \\ 3 & 3.5 & 1.7 & 10 & 17 & 8 \\ 17 & 8 & 0.05 & 10 & 0.1 & 14 \end{bmatrix}, \quad [p_{ij}] = \begin{bmatrix} 1312 & 1696 & 5569 & 124 & 8283 & 5886 \\ 2329 & 4135 & 8307 & 3736 & 1004 & 9991 \\ 2348 & 1451 & 3522 & 2883 & 3047 & 6650 \\ 4047 & 8828 & 8732 & 5743 & 1091 & 381 \end{bmatrix} * 10^{-4}$$

(7) Extended Rosenbrock function (ER) with $n = 9$

$$f(x) = \sum_{i=1}^{n-1} \left[(1 - x_i)^2 + 100(x_{i+1} - x_i^2)^2 \right], \quad x \in [-5, 10]^n \quad (10)$$

(8) Dixon-Price function (DP) with $n = 12$

$$f(x) = (x_1 - 1)^2 + \sum_{i=2}^n i(2x_i^2 - x_{i-1})^2, \quad x \in [-10, 10]^n \quad (11)$$

(9) The tension spring design (TSD) problem is taken from Arora [33]. The schematic of this design is illustrated in Figure 1. The objective is to minimize the weight of a tension spring, whose response function is written as

$$f(x) = (2 + x_3)x_1^2x_2, \quad (12)$$

$$x_1 \in [0.05, 2], \quad x_2 \in [0.25, 1.3], \quad x_3 \in [2, 15]$$

Three variables are identified: diameter x_1 , mean coil diameter x_2 , and number of active coils x_3 .

(10) The I-beam design (IBD) problem was firstly presented by Messac and Mullur [34]. The schematic of this design is illustrated in Figure 2. It aims to minimize the vertical deflection of an I-beam, whose objective function is formulated as

$$f(x) = 5000 \left/ \left(\frac{1}{12}x_3(x_1 - 2x_4)^3 + \frac{1}{6}x_2x_4^3 + 2x_2x_4 \left(\frac{x_1 - x_4}{2} \right)^2 \right) \right. \quad (13)$$

$$x_1 \in [10, 80], \quad x_2 \in [10, 80], \quad x_3 \in [0.9, 5], \quad x_4 \in [0.9, 5]$$

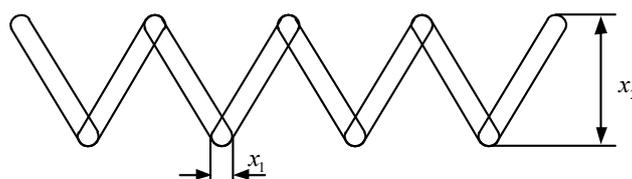


Figure 1. Schematic of tension spring design (TSD).

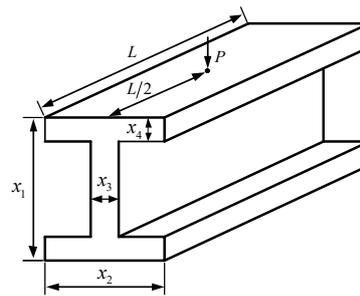


Figure 2. Schematic of I-beam design (IBD).

The known parameters for the IBD problem include: Young’s modulus of elasticity $E = 20000$ kN/cm², maximal bending forces $P = 600$ kN and length of the beam $L = 200$ cm.

The Latin hypercube sampling (LHS) method is used to determine the locations of the points for all test problems. The MATLAB® function “lhsdesign” and “maximin” criterion with a maximum of 100 and 10 iterations are employed to generate the training points and test points, separately. The details about the training and test data for each test problem are given in Table 2. As shown in Table 2, the training set for each test problem is composed of 12 to 182 points depending on the number of input variables. In order to reduce the influence of random sampling, 1000 different training sets are used for all test problems.

Table 2. Summary of training and test data used for each test problem.

Fun.	No. of Design Variables	No. of Training Points	No. of Test Points	No. of Training Sets
BH	2	12	441	1000
CB	2	12	441	1000
GF	2	12	441	1000
HN3	3	36	512	1000
TSD	3	36	512	1000
SK	4	36	512	1000
IBD	4	36	512	1000
HN6	6	56	512	1000
ER	9	110	1000	1000
DP	12	182	1000	1000

4.3. Performance Measures

The prediction accuracy of five individual surrogate models and twenty-six ensemble models is evaluated by three classical performance measures including the coefficient of determination R^2 , root mean square error (RMSE) and maximum absolute error (MAE) [35–37]. They are expressed as

$$R^2(y, \hat{y}) = \frac{\frac{1}{V} \int_V (y - \bar{y})(\hat{y} - \bar{\hat{y}}) dV}{\delta(y)\delta(\hat{y})} \tag{14}$$

$$RMSE = \sqrt{\sum_{i=1}^K (y_i - \hat{y}_i)^2 / K} \tag{15}$$

$$MAE = \text{Max}_{1 \leq i \leq K} |y_i - \hat{y}_i| \tag{16}$$

where K is the number of test points, y_i and \hat{y}_i are the actual and predicted response values at the i th test point, separately. \bar{y}_i and $\bar{\hat{y}}$ are the average values of actual and predicted response values at all test points, respectively. In this work, the MATLAB® function “corrcoef” is employed to calculate R^2 . A larger R^2 and smaller RMSE and MAE would indicate a good prediction capability.

5. Results and Discussions

The prediction accuracy and robustness of all twenty-six ensemble models derived from five basic surrogate models PRS, RBF, KRG, GP, SHEP were tested and compared and the results are presented in this section. The average values of the performance measures R^2 , RMSE and MAE for all test problems are given in Tables 3–5. In order to facilitate the comparison, the average values of RMSE and MAE are normalized with respect to the most accurate stand-alone surrogates while the R^2 metric is not (since it is already a normalized value). The best values of each performance measure are shown in boldface. Thereinto, the numbers in brackets indicate the ranks of the prediction accuracy and the summation of ranks for all test problems are also provided. As shown in Tables 3–5, there is no one surrogate model that performs the best for all test problems. The individual surrogate models IS1 (PRS) and IS2 (RBF) perform the best for test problem CB and GF, respectively, while IS3 (KRG) performs the best for test problem BH. That is, the ensemble of surrogates may be less accurate than individual surrogates for some certain problems. Moreover, the stand-alone surrogate models PRS, RBF and KRG generally perform better than the GP and SHEP models. It is unclear how to accurately and efficiently select the appropriate surrogate models.

Firstly, we will discuss the first question from Section 2. Do we need to use surrogates that are as accurate as possible? As shown in Table 3, for individual surrogate models, IS3 (KRG) performs best, IS2 (RBF) takes second place, and IS1 (PRS) comes in third. For ensembles of two surrogates ES1~ES10, the performance measures of R^2 for ES5 (RBF-KRG), ES2 (PRS-KRG), ES1 (PRS-RBF) are larger than others. It can be found that ES5, ES2 and ES1 select more accurate stand-alone surrogates KRG, RBF and PRS. Furthermore, two more accurate surrogates KRG and RBF lead to the most accurate ensemble model ES5 (RBF-KRG). Similarly, for ensembles of three surrogates ES11~ES20, the total rank of ES11 (PRS-RBF-KRG) is significantly smaller than other ensembles of three surrogates. With regard to ensembles of four surrogates, ES21~ES25, ES21 (PRS-RBF-KRG-SHEP) and ES22 (PRS-RBF-KRG-GP) rank as the top two. It is noted that the performance measure of R^2 for ES11 is larger than other surrogate models used. When considering the results for RMSE and MAE in Tables 4 and 5, the same phenomenon appears as in Table 3. For the performance measure of RMSE, the top three individual surrogate models RBF, KRG, PRS result in the most accurate ensemble models: ES11 (PRS-RBF-KRG), ES5 (RBF-KRG) and ES22 (PRS-RBF-KRG-SHEP). As is the case for performance measure R^2 , the smaller the value of the performance measure RMSE indicates that the ensemble model ES11 (PRS-RBF-KRG) has the best prediction performance. As for the performance measure MAE, the top three individual surrogate models PRS, KRG, RBF result in the most accurate ensemble models ES11 (PRS-RBF-KRG), ES2 (PRS-KRG) and ES22 (PRS-RBF-KRG-SHEP). In accordance with the performance measures R^2 and RMSE, the performance measure of MAE for ES11 (PRS-RBF-KRG) is smallest among the selected thirty-one surrogate models. In summary, we can conclude that selecting surrogates that are as accurate as possible will lead to more accurate ensemble models when using the same quantity of stand-alone surrogates.

Then, we will discuss the second question from Section 2. Do we need to use as many surrogates as possible? The comparison results shown in Table 3, indicate that using poor surrogate models will result in a loss of accuracy of ensemble models. For example, the performance measure of R^2 for ensemble model ES5 (RBF-KRG) is larger than the ensemble models ES17 (RBF-KRG-GP), ES18 (RBF-KRG-SHEP) and ES25 (RBF-KRG-GP-SHEP). On the contrary, satisfactory surrogate models will lead to the advancement of accuracy of ensemble models. For instance, the performance measure of R^2 for ensemble model ES11 (PRS-RBF-KRG) is better than the ensemble models ES1 (PRS-RBF), ES2 (PRS-KRG) and ES5 (RBF-KRG). When considering the results on RMSE and MAE, in Tables 4 and 5, the same conclusion is reached as in Table 3. That is, whether or not to use more surrogate models depends on their prediction performance.

Table 3. Comparison of performance measure coefficient of determination (R^2) for all test problems.

No.	Surr.	BH	CB	GF	HN3	TSD	SK	IBD	HN6	ER	DP	Total
1	IS1	0.813 (29)	0.768 (1)	0.802 (24)	0.742 (28)	0.986 (17)	0.584 (28)	0.724 (22)	0.507 (30)	0.888 (19)	0.911 (11)	209
2	IS2	0.885 (17)	0.754 (3)	0.850 (1)	0.695 (29)	0.961 (25)	0.687 (16)	0.846 (1)	0.541 (28)	0.904 (9)	0.861 (17)	146
3	IS3	0.914 (1)	0.611 (25)	0.830 (10)	0.881 (2)	0.989 (4)	0.690 (12)	0.618 (30)	0.678 (1)	0.600 (28)	0.429 (25)	138
4	IS4	0.414 (31)	0.454 (31)	0.733 (31)	0.665 (31)	0.872 (31)	0.255 (31)	0.128 (31)	0.557 (19)	0.023 (31)	0.026 (31)	298
5	IS5	0.862 (25)	0.527 (29)	0.790 (29)	0.823 (23)	0.895 (30)	0.469 (30)	0.717 (23)	0.484 (31)	0.580 (29)	0.214 (30)	279
6	ES1	0.873 (21)	0.757 (2)	0.843 (2)	0.808 (25)	0.986 (19)	0.682 (22)	0.842 (3)	0.546 (27)	0.907 (1)	0.914 (1)	123
7	ES2	0.904 (9)	0.724 (10)	0.833 (9)	0.882 (1)	0.991 (1)	0.692 (9)	0.711 (25)	0.670 (6)	0.888 (18)	0.911 (12)	100
8	ES3	0.807 (30)	0.712 (15)	0.790 (27)	0.797 (26)	0.986 (18)	0.592 (26)	0.725 (21)	0.557 (20)	0.888 (20)	0.911 (13)	216
9	ES4	0.858 (26)	0.747 (5)	0.817 (14)	0.824 (21)	0.985 (24)	0.591 (27)	0.756 (17)	0.536 (29)	0.887 (24)	0.911 (14)	201
10	ES5	0.910 (3)	0.722 (11)	0.842 (3)	0.878 (6)	0.988 (10)	0.697 (2)	0.838 (6)	0.671 (3)	0.904 (11)	0.861 (21)	76
11	ES6	0.884 (18)	0.697 (19)	0.811 (18)	0.686 (30)	0.959 (26)	0.688 (14)	0.846 (2)	0.554 (23)	0.904 (10)	0.861 (18)	178
12	ES7	0.883 (19)	0.745 (6)	0.840 (4)	0.828 (20)	0.959 (27)	0.684 (20)	0.828 (9)	0.550 (25)	0.904 (13)	0.861 (20)	163
13	ES8	0.913 (2)	0.536 (27)	0.790 (28)	0.879 (5)	0.989 (5)	0.692 (8)	0.618 (29)	0.671 (2)	0.600 (27)	0.429 (26)	159
14	ES9	0.909 (4)	0.595 (26)	0.822 (12)	0.881 (3)	0.988 (8)	0.687 (17)	0.707 (26)	0.666 (15)	0.634 (26)	0.349 (27)	164
15	ES10	0.852 (28)	0.471 (30)	0.749 (30)	0.823 (22)	0.897 (29)	0.490 (29)	0.705 (27)	0.558 (18)	0.580 (30)	0.214 (29)	272
16	ES11	0.902 (11)	0.732 (7)	0.837 (6)	0.873 (13)	0.990 (2)	0.693 (6)	0.836 (7)	0.670 (5)	0.907 (2)	0.913 (5)	64
17	ES12	0.870 (24)	0.727 (9)	0.818 (13)	0.793 (27)	0.985 (21)	0.684 (19)	0.841 (4)	0.550 (26)	0.906 (8)	0.911 (8)	159
18	ES13	0.872 (22)	0.749 (4)	0.837 (5)	0.842 (17)	0.985 (22)	0.680 (23)	0.826 (11)	0.553 (24)	0.906 (5)	0.913 (7)	140
19	ES14	0.902 (10)	0.692 (21)	0.810 (22)	0.872 (14)	0.989 (6)	0.694 (4)	0.713 (24)	0.670 (7)	0.888 (17)	0.911 (10)	135
20	ES15	0.900 (14)	0.716 (14)	0.829 (11)	0.880 (4)	0.989 (3)	0.687 (15)	0.740 (19)	0.666 (16)	0.888 (22)	0.911 (9)	127
21	ES16	0.857 (27)	0.700 (18)	0.795 (25)	0.836 (18)	0.986 (20)	0.599 (25)	0.754 (18)	0.560 (17)	0.888 (23)	0.911 (15)	206
22	ES17	0.909 (5)	0.673 (23)	0.810 (20)	0.878 (7)	0.988 (12)	0.697 (1)	0.839 (5)	0.671 (4)	0.904 (12)	0.861 (24)	113
23	ES18	0.906 (7)	0.717 (13)	0.835 (7)	0.875 (9)	0.988 (13)	0.693 (5)	0.823 (14)	0.667 (11)	0.904 (14)	0.861 (22)	115
24	ES19	0.879 (20)	0.694 (20)	0.810 (19)	0.820 (24)	0.958 (28)	0.683 (21)	0.826 (10)	0.555 (21)	0.904 (16)	0.861 (19)	198
25	ES20	0.908 (6)	0.536 (28)	0.791 (26)	0.875 (8)	0.989 (7)	0.686 (18)	0.704 (28)	0.668 (10)	0.634 (25)	0.330 (28)	184
26	ES21	0.901 (12)	0.707 (16)	0.817 (17)	0.867 (16)	0.988 (11)	0.695 (3)	0.836 (8)	0.669 (8)	0.906 (6)	0.913 (3)	100
27	ES22	0.900 (15)	0.728 (8)	0.834 (8)	0.875 (10)	0.988 (9)	0.690 (11)	0.821 (15)	0.667 (14)	0.906 (4)	0.913 (4)	98
28	ES23	0.871 (23)	0.722 (12)	0.817 (15)	0.831 (19)	0.985 (23)	0.680 (24)	0.825 (12)	0.555 (22)	0.906 (3)	0.913 (2)	155
29	ES24	0.900 (13)	0.685 (22)	0.809 (23)	0.874 (11)	0.987 (15)	0.689 (13)	0.740 (20)	0.667 (13)	0.888 (21)	0.911 (16)	167
30	ES25	0.906 (8)	0.672 (24)	0.810 (21)	0.874 (12)	0.988 (14)	0.692 (7)	0.823 (13)	0.668 (9)	0.904 (15)	0.861 (23)	146
31	ES26	0.899 (16)	0.705 (17)	0.817 (16)	0.868 (15)	0.987 (16)	0.691 (10)	0.821 (16)	0.667 (12)	0.906 (7)	0.913 (6)	131

Table 4. Comparison of performance measure root mean square error (RMSE) for all test problems.

No.	Surr.	BH	CB	GF	HN3	TSD	SK	IBD	HN6	ER	DP	Total
1	IS1	1.534 (29)	1.000 (1)	1.183 (29)	1.364 (28)	1.113 (17)	1.158 (27)	1.318 (24)	1.385 (24)	1.083 (20)	1.000 (12)	211
2	IS2	1.151 (17)	1.005 (2)	1.000 (1)	2.065 (29)	1.804 (25)	1.000 (14)	1.000 (2)	1.550 (30)	1.000 (9)	1.263 (17)	146
3	IS3	1.000 (1)	1.252 (25)	1.029 (5)	1.000 (7)	1.000 (4)	1.036 (23)	1.417 (29)	1.000 (1)	2.058 (27)	2.320 (27)	149
4	IS4	2.977 (31)	1.403 (29)	1.187 (31)	2.262 (31)	3.195 (31)	2.016 (31)	1.901 (31)	1.490 (28)	4.513 (31)	5.856 (31)	305
5	IS5	1.326 (25)	1.501 (31)	1.182 (28)	1.345 (25)	3.021 (30)	1.637 (30)	1.410 (28)	1.619 (31)	2.217 (29)	3.105 (29)	286
6	ES1	1.234 (21)	1.009 (3)	1.035 (6)	1.203 (21)	1.128 (18)	1.004 (17)	1.013 (4)	1.357 (23)	0.973 (1)	0.967 (1)	115
7	ES2	1.067 (9)	1.068 (13)	1.059 (11)	0.951 (1)	0.926 (1)	1.001 (15)	1.290 (23)	1.003 (4)	1.083 (22)	1.000 (11)	110
8	ES3	1.669 (30)	1.074 (14)	1.135 (25)	1.239 (22)	1.134 (20)	1.099 (26)	1.280 (21)	1.333 (18)	1.083 (21)	1.000 (13)	210
9	ES4	1.337 (26)	1.034 (6)	1.136 (26)	1.173 (20)	1.138 (23)	1.162 (28)	1.261 (18)	1.339 (19)	1.084 (24)	1.000 (14)	204
10	ES5	1.017 (2)	1.057 (9)	1.013 (2)	1.014 (11)	1.045 (10)	0.993 (12)	1.019 (6)	1.001 (3)	1.000 (12)	1.263 (21)	88
11	ES6	1.166 (18)	1.080 (15)	1.063 (12)	2.124 (30)	1.840 (26)	0.966 (1)	1.000 (1)	1.500 (29)	1.000 (10)	1.263 (18)	160
12	ES7	1.171 (19)	1.020 (4)	1.020 (3)	1.318 (24)	1.849 (27)	1.012 (20)	1.073 (9)	1.457 (27)	1.000 (11)	1.263 (20)	164
13	ES8	1.027 (3)	1.323 (27)	1.100 (23)	1.009 (8)	1.002 (5)	0.984 (6)	1.424 (30)	1.001 (2)	2.058 (28)	2.320 (28)	160
14	ES9	1.029 (4)	1.289 (26)	1.055 (10)	1.012 (9)	1.032 (8)	1.048 (24)	1.325 (25)	1.010 (9)	1.865 (25)	2.307 (26)	166
15	ES10	1.450 (28)	1.404 (30)	1.185 (30)	1.348 (26)	2.931 (29)	1.310 (29)	1.382 (27)	1.437 (25)	2.217 (30)	3.105 (30)	284
16	ES11	1.074 (10)	1.052 (7)	1.043 (7)	0.994 (3)	0.970 (2)	0.996 (13)	1.023 (7)	1.005 (5)	0.979 (2)	0.977 (6)	62
17	ES12	1.268 (24)	1.052 (8)	1.073 (17)	1.259 (23)	1.139 (24)	0.973 (5)	1.013 (3)	1.356 (22)	0.983 (8)	0.991 (8)	142
18	ES13	1.249 (22)	1.024 (5)	1.048 (8)	1.117 (17)	1.136 (21)	1.016 (21)	1.079 (10)	1.340 (20)	0.979 (4)	0.977 (7)	135
19	ES14	1.091 (14)	1.111 (21)	1.096 (21)	0.996 (4)	1.019 (7)	0.968 (3)	1.285 (22)	1.006 (6)	1.082 (18)	1.000 (10)	126
20	ES15	1.086 (12)	1.082 (16)	1.070 (15)	0.972 (2)	0.989 (3)	1.018 (22)	1.272 (19)	1.009 (8)	1.083 (19)	1.000 (9)	125
21	ES16	1.390 (27)	1.095 (20)	1.156 (27)	1.140 (18)	1.130 (19)	1.096 (25)	1.255 (17)	1.327 (17)	1.083 (23)	1.000 (15)	208
22	ES17	1.032 (5)	1.120 (22)	1.066 (13)	1.014 (10)	1.047 (11)	0.967 (2)	1.016 (5)	1.007 (7)	1.000 (13)	1.263 (24)	112
23	ES18	1.039 (6)	1.066 (12)	1.029 (4)	1.043 (15)	1.051 (14)	1.004 (18)	1.087 (14)	1.010 (10)	1.000 (15)	1.263 (22)	130
24	ES19	1.214 (20)	1.083 (17)	1.070 (14)	1.363 (27)	1.865 (28)	0.986 (7)	1.079 (11)	1.451 (26)	1.000 (14)	1.263 (19)	183
25	ES20	1.049 (8)	1.329 (28)	1.109 (24)	1.038 (14)	1.018 (6)	1.001 (16)	1.338 (26)	1.011 (11)	1.907 (26)	2.300 (25)	184
26	ES21	1.086 (13)	1.085 (18)	1.074 (18)	1.028 (12)	1.051 (13)	0.969 (4)	1.024 (8)	1.012 (12)	0.980 (7)	0.976 (5)	110
27	ES22	1.086 (11)	1.058 (10)	1.051 (9)	0.999 (6)	1.035 (9)	1.007 (19)	1.091 (15)	1.012 (13)	0.979 (5)	0.976 (3)	100
28	ES23	1.268 (23)	1.061 (11)	1.075 (20)	1.164 (19)	1.136 (22)	0.992 (11)	1.083 (12)	1.346 (21)	0.979 (3)	0.974 (2)	144
29	ES24	1.095 (16)	1.123 (24)	1.100 (22)	0.999 (5)	1.071 (15)	0.986 (9)	1.273 (20)	1.013 (14)	1.081 (17)	1.000 (16)	158
30	ES25	1.047 (7)	1.122 (23)	1.071 (16)	1.046 (16)	1.050 (12)	0.987 (10)	1.087 (13)	1.014 (15)	1.000 (16)	1.263 (23)	151
31	ES26	1.092 (15)	1.089 (19)	1.075 (19)	1.030 (13)	1.077 (16)	0.986 (8)	1.093 (16)	1.016 (16)	0.980 (6)	0.976 (4)	132

Table 5. Comparison of performance measure maximum absolute error (MAE) for all test problems.

No.	Surr.	BH	CB	GF	HN3	TSD	SK	IBD	HN6	ER	DP	Total
1	IS1	1.253 (25)	1.000 (1)	1.035 (11)	1.217 (18)	1.000 (1)	1.033 (25)	1.062 (17)	1.013 (20)	1.000 (9)	1.000 (2)	129
2	IS2	1.240 (22)	1.178 (13)	1.000 (1)	2.886 (29)	1.911 (25)	1.013 (19)	1.000 (1)	1.107 (27)	1.006 (13)	1.425 (17)	167
3	IS3	1.000 (1)	1.491 (25)	1.007 (3)	1.000 (4)	1.489 (17)	1.000 (1)	1.261 (29)	1.000 (16)	2.035 (27)	2.385 (25)	148
4	IS4	2.114 (31)	1.662 (30)	1.250 (31)	3.032 (31)	3.531 (31)	1.232 (31)	1.523 (31)	1.063 (25)	3.023 (31)	3.801 (29)	301
5	IS5	1.225 (19)	1.795 (31)	1.136 (28)	1.748 (27)	2.977 (29)	1.190 (30)	1.131 (22)	1.426 (31)	2.589 (30)	3.901 (31)	278
6	ES1	1.240 (21)	1.089 (3)	1.012 (4)	1.204 (17)	1.086 (8)	1.014 (20)	1.003 (3)	1.011 (19)	0.940 (1)	1.007 (14)	110
7	ES2	1.042 (4)	1.132 (5)	1.021 (6)	0.938 (1)	1.211 (9)	1.002 (3)	1.138 (23)	0.989 (14)	1.007 (22)	1.000 (6)	93
8	ES3	1.401 (30)	1.169 (11)	1.134 (26)	1.229 (20)	1.055 (5)	1.049 (28)	1.156 (24)	1.001 (17)	1.000 (10)	1.000 (3)	174
9	ES4	1.210 (17)	1.075 (2)	1.075 (14)	1.279 (23)	1.050 (3)	1.040 (26)	1.072 (18)	1.057 (24)	1.007 (21)	1.000 (4)	152
10	ES5	1.046 (5)	1.232 (19)	1.007 (2)	1.026 (8)	1.533 (21)	1.003 (5)	1.017 (8)	0.976 (2)	1.006 (16)	1.425 (21)	107
11	ES6	1.254 (26)	1.269 (21)	1.104 (24)	2.947 (30)	1.951 (26)	1.013 (18)	1.002 (2)	1.072 (26)	1.006 (14)	1.425 (18)	205
12	ES7	1.211 (18)	1.197 (16)	1.023 (7)	1.684 (25)	1.955 (27)	1.019 (23)	1.033 (10)	1.164 (30)	1.006 (15)	1.425 (20)	191
13	ES8	1.025 (2)	1.556 (27)	1.135 (27)	1.015 (5)	1.492 (18)	1.003 (4)	1.283 (30)	0.975 (1)	2.035 (28)	2.385 (26)	168
14	ES9	1.025 (3)	1.527 (26)	1.036 (12)	1.087 (12)	1.518 (20)	1.005 (9)	1.157 (26)	0.992 (15)	1.971 (25)	2.409 (28)	176
15	ES10	1.326 (29)	1.662 (29)	1.236 (30)	1.718 (26)	2.993 (30)	1.095 (29)	1.183 (28)	1.134 (28)	2.589 (29)	3.901 (30)	288
16	ES11	1.070 (12)	1.133 (6)	1.017 (5)	1.000 (3)	1.254 (10)	1.005 (8)	1.014 (5)	0.982 (8)	0.951 (3)	1.006 (11)	71
17	ES12	1.269 (27)	1.152 (9)	1.088 (18)	1.273 (22)	1.075 (7)	1.012 (17)	1.007 (4)	1.010 (18)	0.960 (8)	1.007 (12)	142
18	ES13	1.230 (20)	1.103 (4)	1.031 (10)	1.225 (19)	1.067 (6)	1.020 (24)	1.033 (9)	1.044 (23)	0.952 (4)	1.007 (13)	132
19	ES14	1.069 (11)	1.204 (17)	1.093 (19)	0.995 (2)	1.280 (12)	1.005 (7)	1.157 (25)	0.979 (3)	1.006 (12)	1.000 (7)	115
20	ES15	1.060 (9)	1.147 (8)	1.040 (13)	1.016 (6)	1.267 (11)	1.007 (14)	1.114 (20)	0.988 (13)	1.009 (24)	1.000 (8)	126
21	ES16	1.271 (28)	1.174 (12)	1.140 (29)	1.253 (21)	1.034 (2)	1.046 (27)	1.111 (19)	1.033 (21)	1.004 (11)	1.000 (5)	175
22	ES17	1.056 (7)	1.306 (23)	1.099 (22)	1.025 (7)	1.534 (22)	1.005 (10)	1.014 (6)	0.981 (5)	1.006 (17)	1.425 (24)	143
23	ES18	1.056 (8)	1.245 (20)	1.024 (8)	1.124 (16)	1.536 (24)	1.008 (15)	1.044 (14)	0.982 (9)	1.006 (19)	1.425 (22)	155
24	ES19	1.248 (23)	1.270 (22)	1.102 (23)	1.753 (28)	1.973 (28)	1.015 (21)	1.039 (12)	1.139 (29)	1.006 (18)	1.425 (19)	223
25	ES20	1.048 (6)	1.561 (28)	1.133 (25)	1.113 (14)	1.507 (19)	1.001 (2)	1.179 (27)	0.981 (4)	2.005 (26)	2.407 (27)	178
26	ES21	1.080 (15)	1.182 (15)	1.083 (16)	1.039 (9)	1.315 (15)	1.006 (12)	1.015 (7)	0.982 (10)	0.955 (7)	1.011 (16)	122
27	ES22	1.078 (14)	1.140 (7)	1.031 (9)	1.054 (11)	1.302 (13)	1.011 (16)	1.045 (16)	0.983 (12)	0.952 (5)	1.002 (10)	113
28	ES23	1.250 (24)	1.156 (10)	1.085 (17)	1.292 (24)	1.054 (4)	1.016 (22)	1.037 (11)	1.041 (22)	0.951 (2)	1.010 (15)	151
29	ES24	1.073 (13)	1.210 (18)	1.094 (20)	1.051 (10)	1.314 (14)	1.004 (6)	1.128 (21)	0.981 (6)	1.008 (23)	1.000 (9)	140
30	ES25	1.063 (10)	1.307 (24)	1.097 (21)	1.123 (15)	1.535 (23)	1.005 (11)	1.043 (13)	0.982 (7)	1.006 (20)	1.425 (23)	167
31	ES26	1.084 (16)	1.182 (14)	1.082 (15)	1.090 (13)	1.323 (16)	1.007 (13)	1.044 (15)	0.983 (11)	0.954 (6)	0.998 (1)	120

In order to facilitate the comparison of the prediction performances of the ensemble models with different quantities, Table 6 provides their average values of total ranks of performance measures for all test problems. It can be seen that the overall prediction performances of stand-alone surrogate models are inferior to the ensemble models. Furthermore, the more surrogate models we use for constructing ensemble models, the more accurate prediction performances we obtain. The results also show that the average prediction performance of the ensemble of four surrogates ES21–ES25 is compared to that of the ensemble of five surrogates ES26. We can conclude that using as many as possible surrogates may not improve the prediction performances of ensemble models due to the inaccurate surrogates. However, using as many as possible accurate surrogates will contribute to enhance the prediction performance. In fact, the ensemble model ES11 (PRS-RBF-KRG) shows the best prediction performance among twenty-six ensemble models and five individual surrogate models.

Table 6. Comparison of performance measures on ensemble models with different quantities.

Performance Measures	Individual Surrogate	Ensemble of Two Surrogates	Ensemble of Three Surrogates	Ensemble of Four Surrogates	Ensemble of Five Surrogates
R^2	214.0	165.2	144.1	133.2	131.0
RMSE	219.2	166.2	140.7	132.6	132.0
MAE	204.6	166.4	146.0	138.6	120.0

In order to reveal the robustness of the different surrogate models, statistical graphics i.e., boxplots are used to show the deviations of the performance measures results. A smaller/shorter box implies a small standard deviation and the symbol (+) denotes an abnormal value. Figures 3–5 show the boxplot of performance measures for R^2 , RMSE and MAE, respectively. The comparative results show that the sizes of boxes of all surrogate models for different test problems are obviously variable. Moreover, the abnormal values from the used surrogate models mainly appear for all test problems due to the substantive experiments. Meanwhile, the standard deviation of individual surrogate models PRS, RBF and KRG is smaller than many ensemble models. Certainly, the standard deviation of all ensemble models is smaller than the worst stand-alone surrogate model. The relatively smaller boxes of ensemble models for most test problems demonstrate that it has a higher robustness, compared with the individual surrogate models and other ensemble models. In conclusion, it can be seen that: (1) The prediction performances of the five individual surrogate models and twenty-six ensemble models vary, apparently with different test problems; (2) The individual surrogate models PRS, RBF and KRG reveal the satisfactory prediction performances which are not worse than most of the ensemble models; (3) The twenty-six ensemble models perform better than the worst individual surrogate model, which demonstrates the necessity for adopting the ensemble techniques; (4) The ensemble of four surrogates ES21–ES25 and ensemble of five surrogates ES26, on the whole provide better accuracy and robustness under three performance measures for ten test problems in this work. In general, we suggest that ensemble models can be used as an insurance, rather than for offering significant improvement. From the accuracy and robustness perspectives, the ensemble model ES11 (PRS-RBF-KRG) is suggested in this work.

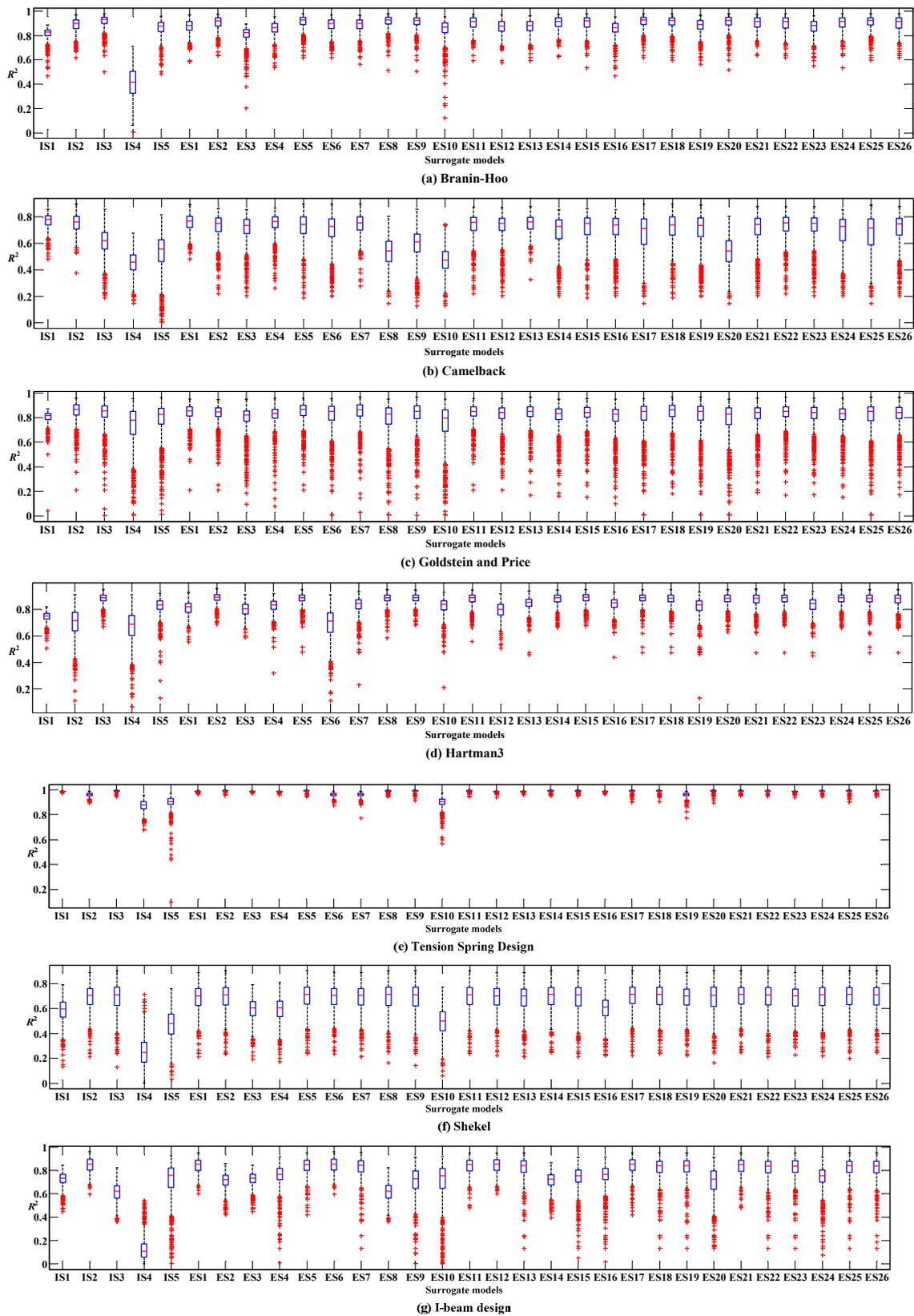


Figure 3. Cont.

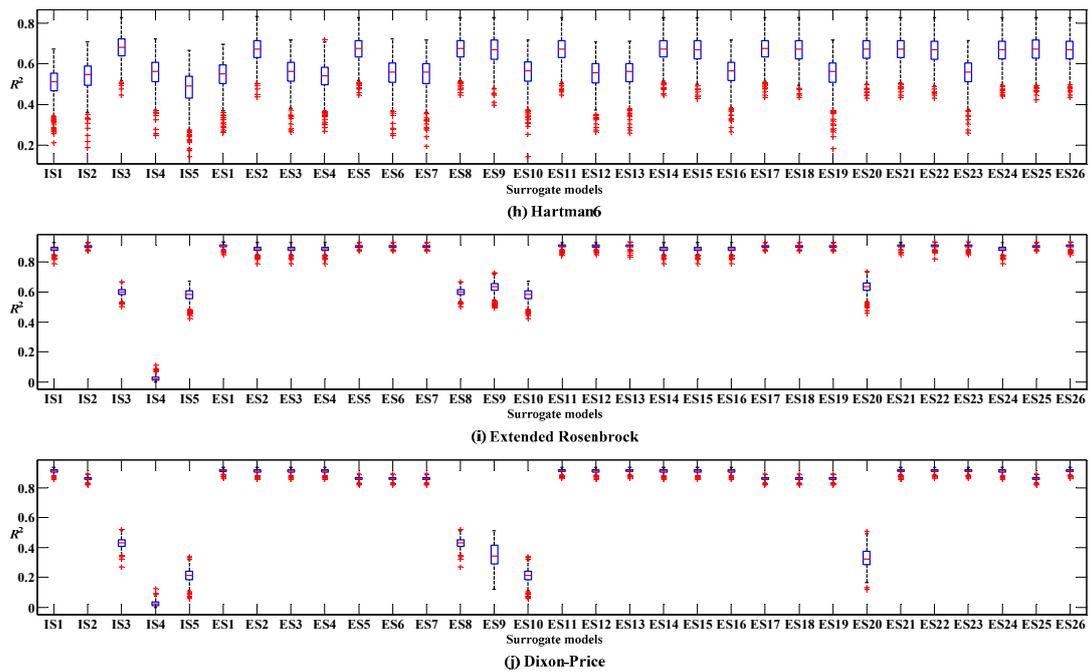


Figure 3. Boxplots of R^2 for all test problems.

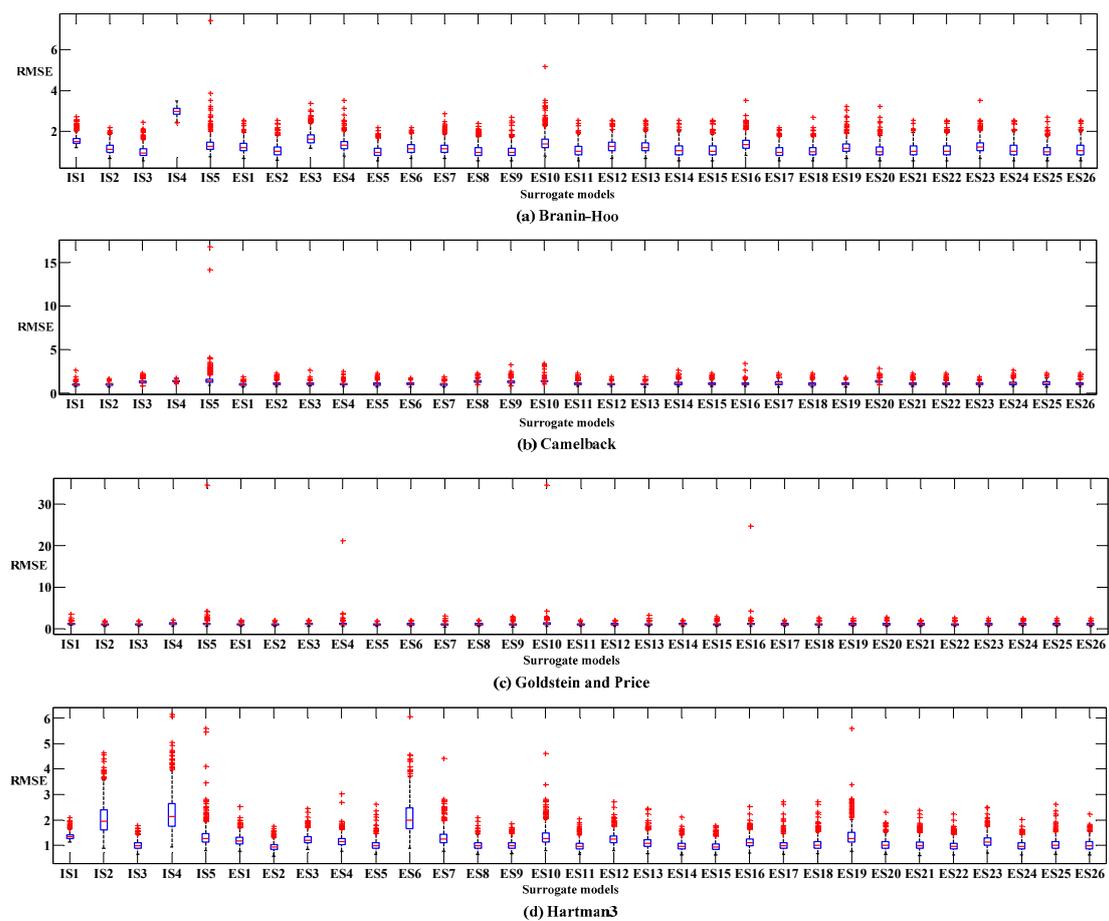


Figure 4. Cont.

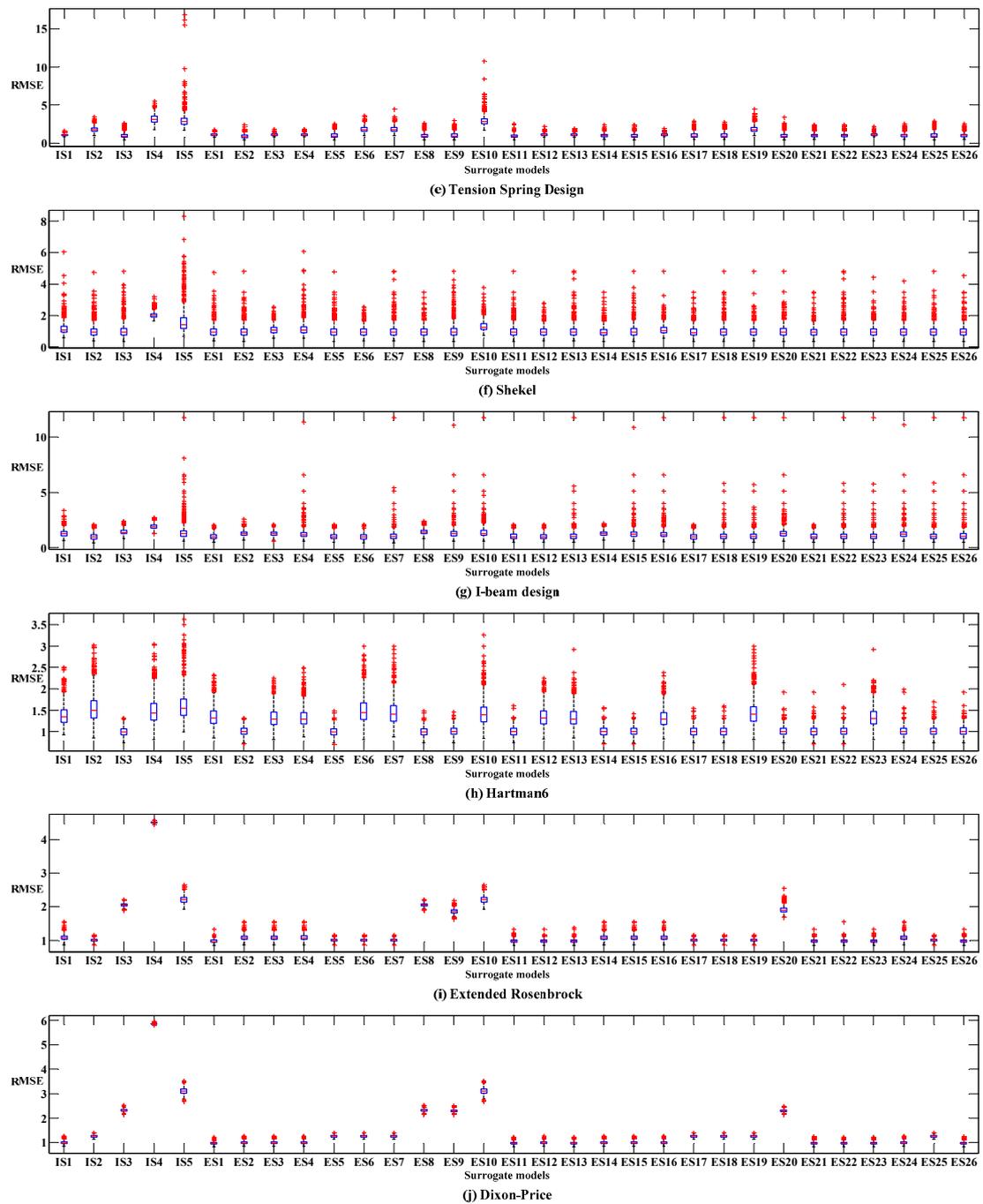


Figure 4. Boxplots of RMSE for all test problems.

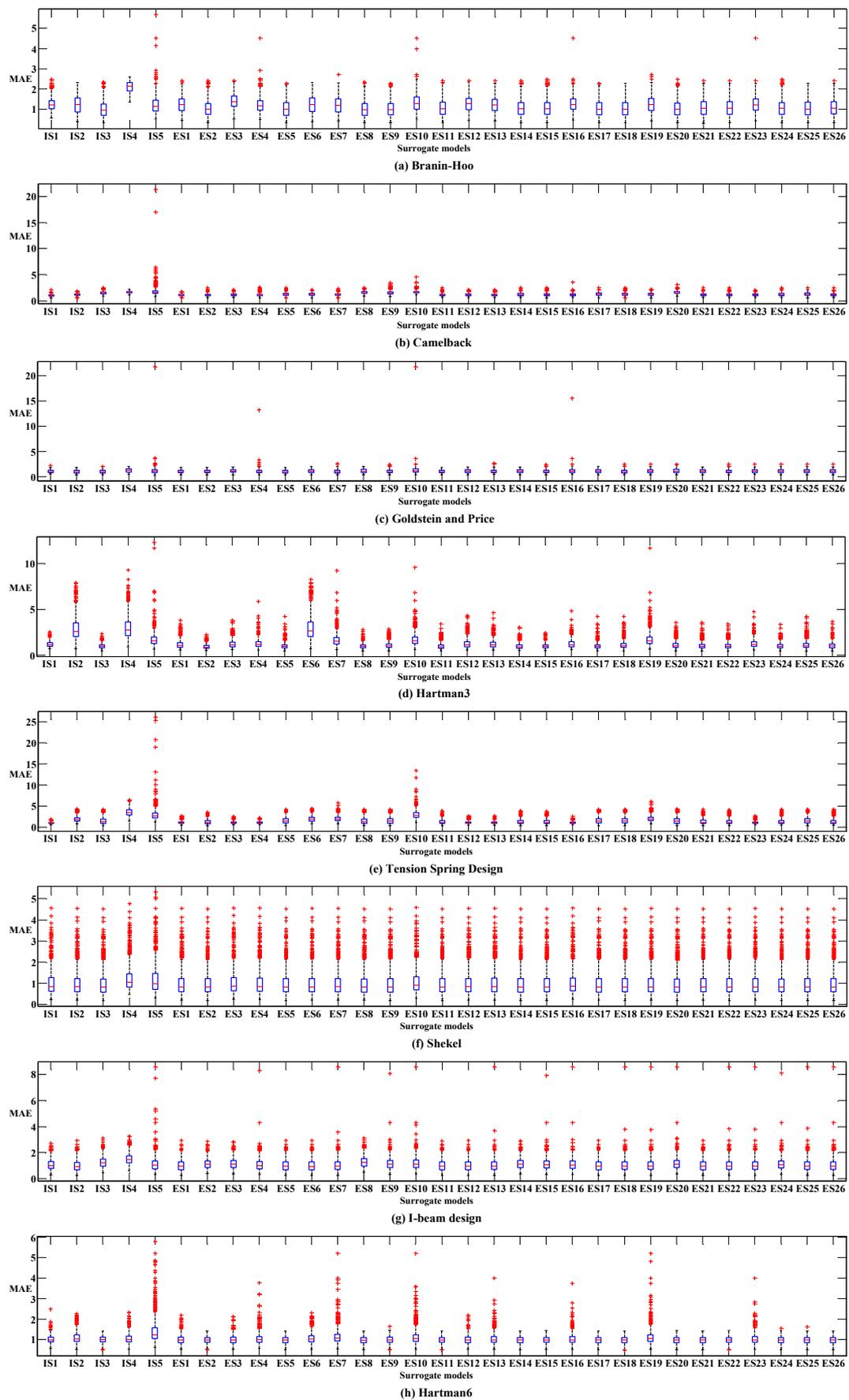


Figure 5. Cont.

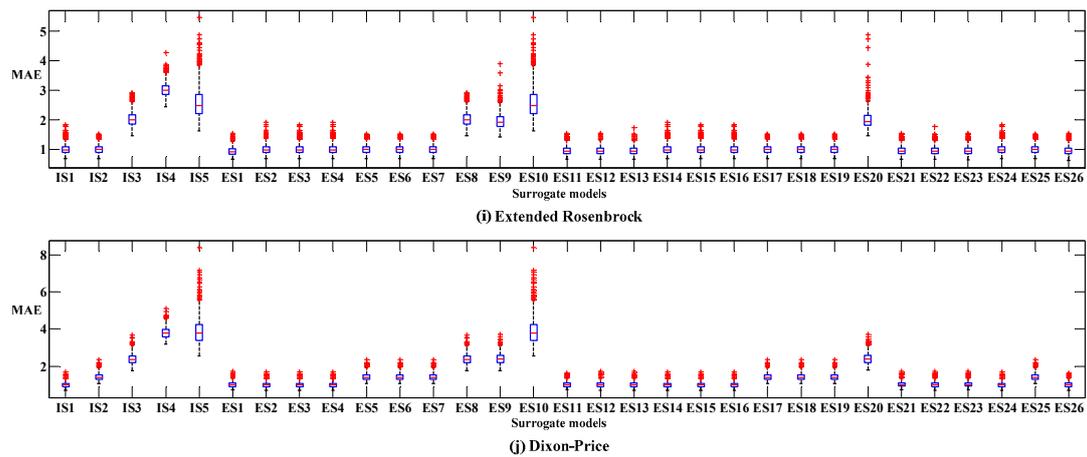


Figure 5. Boxplots of MAE for all test problems.

6. Conclusions

In this work, five basic surrogate models PRS, RBF, KRG, GP and SHEP are selected to form twenty-six ensemble models, through using a previously presented weights selection method. The prediction performances of the total thirty-one surrogate models are tested using eight mathematical problems and two engineering examples. The performance measures R^2 , RMSE and MAE are calculated to reflect the prediction accuracy and robustness. The comparative studies answer two questions raised in this paper. We suggest that more accurate surrogates should be used to construct the ensemble models on the premise of sufficient computational resources. Moreover, the ensemble of three surrogates ES11 (PRS-RBF-KRG) is preferred in view of prediction accuracy and robustness. The objective of this work is to guide researchers in selecting the appropriate quantity and variety of surrogate modeling techniques for building ensemble models, rather than concentrating on developing novel ensemble modeling methods. However, there are still some limitations in this study. Firstly, the use of only five basic surrogate models is limited. Secondly, the selection of weight factors is not investigated and discussed. More individual surrogate models will be selected as the basic surrogate models for the comparative study in future work. Additionally, the investigation on the choice of weights in different ensembles will also be studied.

Author Contributions: Conceptualization, P.Y. and G.P.; methodology, P.Y.; software, P.Y.; validation, P.Y.; formal analysis, P.Y. and G.P.; investigation, G.P.; writing—original draft preparation, P.Y. and G.P.; writing—review and editing, P.Y. and G.P. All authors have read and agreed to the published version of the manuscript.

Funding: Supports from National Natural Science Foundation of China (No. 61803306, 51709229, 51879220), China Postdoctoral Science Foundation (No. 2019M660264), National Key Research and Development Project of China (No. 2016YFC0301300), and the Fundamental Research Funds for the Central Universities (No. 3102019HHZY03009) are greatly acknowledged.

Conflicts of Interest: The authors declare no conflict of interest.

Acronyms

PRS	Polynomial response surface
RBF	Radial basis functions
KRG	Kriging
GP	Gaussian process
SHEP	Linear shepard
MARS	Multivariate adaptive regression splines
E-RBF	Extended radial basis function
BNN	Bayesian neural network
NN	Neural networks

GMSE	Generalized mean square cross-validation error
SVR	Support vector regression
$PRESS_{RMS}$	Prediction sum of squares of root mean square
SQP	Sequential quadratic programming
LHS	Latin hypercube sampling
R^2	Coefficient of determination
RMSE	Root mean square error
MAE	Maximum absolute error

References

- Jin, R.; Chen, W.; Simpson, T.W. Comparative studies of metamodeling techniques under multiple modelling criteria. *Struct. Multidiscip. Optim.* **2011**, *23*, 1–13. [[CrossRef](#)]
- Mullur, A.A.; Messac, A. Extended radial basis functions: More flexible and effective metamodeling. *AIAA J.* **2005**, *43*, 1306–1315. [[CrossRef](#)]
- Mullur, A.A.; Messac, A. Metamodeling using extended radial basis functions: A comparative approach. *Eng. Comput.* **2006**, *21*, 203–217. [[CrossRef](#)]
- Zhao, D.; Xue, D. A comparative study of metamodeling methods considering sample quality merits. *Struct. Multidiscip. Optim.* **2010**, *42*, 923–938. [[CrossRef](#)]
- Gelder, L.V.; Das, P.; Janssen, H.; Roels, S. Comparative study of metamodeling techniques in building energy simulation: Guidelines for practitioners. *Simul. Model. Pr. Theory* **2014**, *49*, 245–257. [[CrossRef](#)]
- Salem, M.B.; Tomaso, L. Automatic selection for general surrogate models. *Struct. Multidiscip. Optim.* **2018**, *58*, 719–734. [[CrossRef](#)]
- Keane, A.J.; Voutchkov, I.I. Surrogate approaches for aerodynamic section performance modeling. *AIAA J.* **2020**, *58*, 16–24. [[CrossRef](#)]
- Kianifar, M.R.; Campean, F. Performance evaluation of metamodeling methods for engineering problems: Towards a practitioner guide. *Struct. Multidiscip. Optim.* **2020**, *61*, 159–186. [[CrossRef](#)]
- Goel, T.; Haftka, R.T.; Shyy, W. Ensemble of surrogates. *Struct. Multidiscip. Optim.* **2007**, *33*, 199–216. [[CrossRef](#)]
- Acar, E.; Rais-Rohani, M. Ensemble of metamodels with optimized weight factors. *Struct. Multidiscip. Optim.* **2009**, *37*, 279–294. [[CrossRef](#)]
- Zhou, X.; Ma, Y.; Li, X. Ensemble of surrogates with recursive arithmetic average. *Struct. Multidiscip. Optim.* **2011**, *44*, 651–671. [[CrossRef](#)]
- Ferreira, W.G.; Serpa, A.L. Ensemble of metamodels: The augmented least squares approach. *Struct. Multidiscip. Optim.* **2016**, *53*, 1019–1046. [[CrossRef](#)]
- Alizadeh, R.; Jia, L.; Balu, A. Ensemble of surrogates and cross-validation for rapid and accurate predictions using small data sets. *Artif. Intell. Eng. Des. Anal. Manuf.* **2019**, *33*, 484–501. [[CrossRef](#)]
- Acar, E. Various approaches for constructing an ensemble of metamodels using local measures. *Struct. Multidiscip. Optim.* **2010**, *42*, 879–896. [[CrossRef](#)]
- Lee, Y.; Choi, D.H. Pointwise ensemble of meta-models using v nearest points cross-validation. *Struct. Multidiscip. Optim.* **2014**, *50*, 383–394. [[CrossRef](#)]
- Liu, H.; Xu, S.; Wang, X. Optimal weighted pointwise ensemble of radial basis functions with different basis functions. *AIAA J.* **2016**, *54*, 3117–3133. [[CrossRef](#)]
- Yin, H.; Fang, H.; Wen, G. On the ensemble of metamodels with multiple regional optimized weight factors. *Struct. Multidiscip. Optim.* **2018**, *58*, 245–263. [[CrossRef](#)]
- Ye, Y.; Wang, Z.; Zhang, X. An optimal pointwise weighted ensemble of surrogates based on minimization of local mean square error. *Struct. Multidiscip. Optim.* **2020**, *62*, 529–542. [[CrossRef](#)]
- Chen, L.; Qiu, H.; Jiang, C.; Cai, X.; Gao, L. Ensemble of surrogates with hybrid method using global and local measures for engineering design. *Struct. Multidiscip. Optim.* **2018**, *57*, 1711–1729. [[CrossRef](#)]
- Zhang, J.; Yue, X.; Qiu, J.; Zhang, M.; Wang, X. A unified ensemble of surrogates with global and local measures for global metamodeling. *Eng. Optim.* **2020**. [[CrossRef](#)]
- Hamza, K.; Saitou, K. A co-evolutionary approach for design optimization via ensembles of surrogates with application to vehicle crashworthiness. *J. Mech. Design* **2012**, *134*, 011001. [[CrossRef](#)]

22. Gu, X.; Lu, J.; Wang, H. Reliability-based design optimization for vehicle occupant protection system based on ensemble of metamodels. *Struct. Multidiscip. Optim.* **2015**, *51*, 533–546. [[CrossRef](#)]
23. Dhamotharan, V.; Jadhav, P.D.; Ramu, P. Optimal design of savonius wind turbines using ensemble of surrogates and CFD analysis. *Struct. Multidiscip. Optim.* **2018**, *58*, 2711–2726. [[CrossRef](#)]
24. Cheng, K.; Lu, Z. Structural reliability analysis based on ensemble learning of surrogate models. *Struct. Saf.* **2020**, *83*, 101905. [[CrossRef](#)]
25. Kajbaf, A.A.; Bensi, M. Application of surrogate models in estimation of storm surge: A comparative assessment. *Appl. Soft. Comput.* **2020**, *91*, 106184. [[CrossRef](#)]
26. Fan, C.; Hou, B.; Zheng, J.; Xiao, L.; Yi, L. A surrogate-assisted particle swarm optimization using ensemble learning for expensive problems with small sample datasets. *Appl. Soft. Comput.* **2020**, *91*, 106242. [[CrossRef](#)]
27. Ye, P.; Pan, G.; Dong, Z. Ensemble of surrogate based global optimization methods using hierarchical design space reduction. *Struct. Multidiscip. Optim.* **2018**, *58*, 537–554. [[CrossRef](#)]
28. Zerpa, L.; Queipo, N.V.; Pintos, S.; Salager, J.L. An optimization methodology of alkaline-surfactant- polymer flooding processes using field scale numerical simulation and multiple surrogates. *J. Petrol. Sci. Eng.* **2005**, *47*, 197–208. [[CrossRef](#)]
29. Viana, F.A.C. SURROGATES Toolbox User’s Guide Version 3.0. Available online: <http://sites.google.com/site/felipeacviana/surrogate-stoolbox> (accessed on 6 October 2020).
30. Jekabsons, G. RBF: Radial Basis Function Interpolation for MATLAB/OCTAVE. Available online: <http://www.cs.rtu.lv/jekabsons/regression.html> (accessed on 6 October 2020).
31. Lophaven, S.N.; Nielsen, H.B.; Søndergaard, J. DACE—A MATLAB Kriging Toolbox. Available online: <http://www2.imm.dtu.dk/~hbn/dace> (accessed on 6 October 2020).
32. Forrester, A.; Sobester, A.; Keane, A. *Engineering Design via Surrogate Modelling: A Practical Guide*; John Wiley & Sons: Hoboken, NJ, USA, 2008.
33. Arora, J.S. *Introduction to Optimum Design*; McGraw-Hill: New York, NY, USA, 1989.
34. Messac, A.; Mullur, A.A. A computationally efficient metamodeling approach for expensive multiobjective optimization. *Optim. Eng.* **2008**, *9*, 37–67. [[CrossRef](#)]
35. Ly, H.B.; Pham, B.T.; Le, T.T.; Le, V.M.; Asteris, P.G. Estimation of axial load-carrying capacity of concrete-filled steel tubes using surrogate models. *Neural. Comput. Appl.* **2020**. [[CrossRef](#)]
36. Li, X.; Lee, Y.; Mikael, S.; Simonelli, J.; Tsao, T.C.; Wu, H.H. Respiratory motion prediction using fusion-based multi-rate kalman filtering and real-time golden-angle radial MRI. *IEEE Trans. Biomed. Eng.* **2020**, *67*, 1727–1738. [[CrossRef](#)] [[PubMed](#)]
37. Xu, G.; Kareem, A.; Shen, L. Surrogate modeling with sequential updating: Applications to bridge deck–wave and bridge deck–wind interactions. *J. Comput. Civ. Eng.* **2020**, *34*, 04020023. [[CrossRef](#)]



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).