Structural and Multidisciplinary Optimization (2022) 65:113
https://doi.org/10.1007/s00158-022-03202-3

RESEARCH PAPER

A pointwise ensemble of surrogates with adaptive function and heuristic formulation

Hao Chen[1,2,3] · Weikun Li[2,3] · Weicheng Cui[2,3] · Qimeng Liu[1,2,3]

Received: 26 June 2021 / Revised: 18 January 2022 / Accepted: 18 February 2022 / Published online: 10 March 2022

Abstract

Due to the advantages of easy implementation and high efficiency, surrogate models have been widely used in designing complex engineering systems. In general, a stand-alone surrogate model cannot perform well for all engineering design problems, and the performance of a surrogate model is not known in advance. Ensembles of surrogates that combine various stand-alone surrogates have been developed to improve the robustness of stand-alone surrogate models. Inspired by the previous research on using the heuristic formulation to calculate weights of stand-alone surrogate models, we propose a pointwise ensemble of surrogates with adaptive function and heuristic formulation (PEAH) in this paper. The adaptive function presented in this paper contains the local accuracy and prediction uncertainty information around a prediction point. Thus, the adaptive function can adapt to the local characteristics of the prediction point. Various analytical test functions and two engineering design problems have been selected to test PEAH, and existing well-known ensembles of surrogates are employed to compare with the proposed pointwise ensemble model. The test results indicate that PEAH performs better in those problems with a better balance between accuracy and robustness.

1 Introduction

The application of high-fidelity numerical simulations, including computational fluid dynamics (CFD) and finite element analysis (FEA), in the design of complex engineering systems leads to extensive computation. When high-fidelity simulation techniques are integrated into the design optimization of complex engineering systems, the resulting computational cost is expensive (Jin et al. 2001). Hence, surrogate models are widely used in place of costly high-fidelity computer simulations.

Surrogate models approximate the relationship between the system response and input variables using interpolation fitting methods or regression methods at a set of training points. Many stand-alone surrogate models have been developed, including Radial Basis Function (RBF) (Hardy 1971), Polynomial Response Surface (PRS) (Box et al. 1978) , Kriging (Sacks et al. 1989), Multivariate Adaptive Regression Splines (MARS) (Friedman 1991), Artificial Neural Network (ANN), (Hassoun et al. 1995), Gaussian Process (GP) (MacKay 1998) and Support Vector Regression (SVR) (Cristianini et al. 2000). Many researchers have focused on comparing the performance of various stand-alone surrogate models (Carpenter and Barthelemy 1993; Giunta and Watson 1998; Simpson et al. 1998; Clarke et al. 2004; Krishnamurthy 2005; Yang et al. 2005).

In general, one surrogate model cannot perform well for all engineering design problems (Forrester and Keane 2009). For a given design problem, the performance of a surrogate model is not known in advance, so it is challenging to select an appropriate surrogate model for a given problem. In addition, for the same design problem, the performance of the surrogate model may vary from one design of experiments (DOE) to another DOE (Viana et al. 2009).

Over the past few decades, ensembles of surrogates that combine various surrogate models have been investigated by

Responsible Editor: Mehmet Polat Saka

✉ Weicheng Cui
cuiweicheng@westlake.edu.cn

[1] Zhejiang University-Westlake University Joint Training, Zhejiang University, Hangzhou 310024, Zhejiang Province, China

[2] Key Laboratory of Coastal Environment and Resources of Zhejiang Province (KLaCER), School of Engineering, Westlake University, Hangzho 310024, Zhejiang Province, China

[3] Institute of Advanced Technology, Westlake Institute for Advanced Study, Hangzhou 310024, Zhejiang Province, China

researchers to improve the accuracy and robustness of the prediction (Goel et al. 2007; Acar and Rais-Rohani 2009; Viana et al. 2009; Acar 2010; Lee and Choi 2014; Liu et al. 2016). Each surrogate is assigned a weight factor, which reflects the importance of the surrogate in the ensemble. The larger the weight, the more important the corresponding surrogate model is in the ensemble. According to the error measures used to determine weight factors, existing ensembles of surrogates can be generally classified as the average ensemble and the pointwise ensemble.

Using global error measures, the weight factors of the average ensemble keep constant over the entire design space. Many researchers employed the generalized mean square cross-validation error (*GMSE*) as the global error measure to calculate the weight factors. Goel et al. (2007) proposed a heuristic formulation to evaluate the weight factors based on *GMSE* . Acar and Rais-Rohani (2009) converted the selection of weight factors to an optimization problem, and the objective was to minimize *GMSE*. Viana et al. (2009) calculated weight factors using an approach based on the optimization problem with the aim to reduce the mean square error (*MSE*). Zhou et al. (2011) proposed a new approach to calculate weight factors using a recursive process. Liu et al. (2017) employed an evolutionary multi-agent system to construct the ensemble of surrogates, and the *GMSE* was adopted as the global error metric. Yin et al. (2018) proposed a new ensemble of surrogates. The design space was divided into multiple regions. One set of weight factors was obtained by minimizing the *GMSE* on the training points in each area.

Based on local error measures, weight factors of the pointwise ensemble change with the prediction point. As an alternative to using a global error measure, Sanchez et al. (2008) adopted the prediction variance as the local error measure to evaluate weight factors of the ensemble of surrogates, and the prediction variance for each surrogate model in the ensemble was assessed using the k nearest neighbors of the prediction point x . Acar (2010) proposed an ensemble of surrogates, in which the pointwise cross-validation error is employed as the local error metric. Zhang et al. (2012) presented an adaptive hybrid surrogate model, and the weight factors were determined based on a local measure of accuracy. Lee and Choi (2014) developed a new ensemble of surrogates using the v nearest points cross-validation error (vCV) rather than the *GMSE*. Liu et al. (2016) proposed an optimal weighted pointwise ensemble that combined the locally accurate predictions of radial basis function models constructed by different basis functions. Chen et al. (2018) developed an ensemble of surrogates with local and global measures. Song et al. (2018) employed the Gaussian process estimated prediction error as the local error metric to determine weight factors of the ensemble of surrogates. Ye et al. (2020) developed an optimal weighted pointwise ensemble. The weight factors were selected based on minimizing the local mean square error constructed with the global-local error (*GLE*). Zhang et al. (2021) proposed a unified ensemble of surrogates using the combination of global and local error metrics.

Due to constant weights over the design space, the average ensemble of surrogates cannot adapt to the local characteristics of the prediction point. Thus, the performance of average ensemble models is limited. In contrast, the pointwise ensemble of surrogates employs local error measures to evaluate the weights of stand-alone surrogates. The weight factor of each surrogate changes with the prediction point. Compared with the average ensemble, the pointwise ensemble did not show obvious advantages because there is no accurate local error metric (Ye et al. 2020). Therefore, the local error measure that can accurately adapt to the local features around the prediction point is the key to the pointwise ensemble of surrogates.

Inspired by the heuristic formulation proposed by Goel et al. (2007), a pointwise ensemble with adaptive function and heuristic formulation (PEAH) is proposed in this work. The adaptive function combines the pointwise cross-validation error and the distance information between the training set and the prediction point. In addition, the uncertainty information of local accuracy for all stand-alone surrogates at the prediction point is also included in the adaptive function. Therefore, the adaptive function can adapt to the local characteristics of a prediction point.

In the following of this paper, existing well-known ensembles of surrogates are reviewed in Sect. 2. And the proposed pointwise ensemble of surrogates is presented in Sect. 3. Section 4 gives the comparison results on numerical examples. Finally, concluding remarks are given in Sect. 5.

2 Existing ensembles of surrogates

Conventional multiple surrogate technique assesses the performance of various stand-alone surrogate models and selects the best surrogate model (Glaz et al. 2009). This conventional approach has two disadvantages. Firstly, most of the time spent on building stand-alone surrogate models is wasted. Secondly, for the same problem, the performance of various surrogate models depends on the selection of the training points, the chosen surrogate model may not perform the best when the training set changes (Acar 2010). Those two problems can be solved by constructing a weighted ensemble of surrogates.

2.1 Weighted average ensemble model

Using a weighted average of various surrogate models, an ensemble of surrogates can be constructed. The built ensemble can be formulated as:

$$\hat{y}_E(x) = \sum_{i=1}^{M} w_i \hat{y}_i(x) \tag{1}$$

$$\sum_{i=1}^{M} w_i = 1 \tag{2}$$

where M refers to the number of stand-alone surrogate models in the ensemble, $\hat{y}_i$ represents the predicted value of the *ith* surrogate model at the prediction point x, and $\hat{y}_E$ refers to the predicted value of the ensemble of surrogates at the prediction point. w_i indicates the weight factor of the *ith* surrogate model. The sum of weight factors in the ensemble equals to 1, so when all surrogates have the same predicted value at the point x, the ensemble of surrogates also keeps the same prediction.

2.2 Average ensemble model proposed by Goel et al. (2007)

Using the global error metric, the weight factors of average ensemble keep constant over the entire design space. Goel et al. (2007) proposed a prediction sum of squares (PRESS) based weighting model (PBW) using the following heuristic formulation.

$$w_i = \frac{w_i^*}{\sum_{j=1}^{M} w_j^*} \tag{3}$$

$$w_i^* = (E_i + \alpha \bar{E})^{\beta} \quad \alpha < 1, \beta < 0 \tag{4}$$

$$E_i = \sqrt{\frac{1}{N}\sum_{k=1}^{N} (y_{(k)} - \hat{y}_{i(k)})^2} = \sqrt{PRESS} \tag{5}$$

$$\bar{E} = \frac{1}{M}\sum_{i=1}^{M} E_i \tag{6}$$

where E_i indicates the $\sqrt{PRESS}$ (also known as $\sqrt{GMSE}$) of the *ith* surrogate model at the training set. α and β are two parameters used to control the importance of averaging and the importance of each surrogate in the ensemble, respectively. N refers to the number of training points in the training set. $y_{(k)}$ represents the actual response at the *kth* training point, and $\hat{y}_{i(k)}$ represents the predicted value at the *kth* training point from the *ith* surrogate model constructed with all the training points except the *kth* one. The average ensemble model proposed by Goel et al. (2007) is denoted as PBW.

2.3 Average ensemble model proposed by Acar and Rais-Rohani (2009)

Acar and Rais-Rohani (2009) converted the selection of weight factors to an optimization problem, and the objective was to minimize the *GMSE* of the proposed ensemble model. The optimization problem is formulated as:

$$\begin{aligned} &Min \quad GMSE(\hat{y}_E(w_i)) = \frac{1}{N}\sum_{k=1}^{N} (y_{(k)} - \hat{y}_{E(k)})^2 \\ &w.r.t \quad w_i \\ &s.t \quad \sum_{i=1}^{M} w_i = 1 \end{aligned} \tag{7}$$

where $\hat{y}_{E(k)}$ represents the predicted value at the *kth* training point from the ensemble of surrogates constructed with all except the *kth* training point. The average ensemble model proposed by Acar and Rais-Rohani (2009) is denoted as EP in this paper.

2.4 Average ensemble model proposed by Viana et al. (2009)

Inspired by Bishop's approach to calculate the mean square error (*MSE*) (Bishop et al. 1995), Viana et al. (2009) calculated weight factors using an approach based on the optimization problem with the objective to minimize *MSE*:

$$MSE_E = \frac{1}{V}\int_V e_E^2(x)dx = w^T C w \tag{8}$$

$$e_E(x) = y(x) - \hat{y}_E(x) \tag{9}$$

where $e_E(x)$ represents the prediction error of the ensemble of surrogates. C refers to the covariance matrix in Bishop's formulation.

$$c_{ij} = \frac{1}{V}\int_V e_i(x) e_j(x) dx \tag{10}$$

c_{ij} represents the element of C, Viana et al. (2009) approximated C by the following equation:

$$c_{ij} \simeq \frac{1}{N} e_i^T e_j \tag{11}$$

where N represents the number of training points and $e_i = [e_i(x_1), e_i(x_2), ..., e_i(x_N)]^T$ is the pointwise cross-validation error vector of the *ith* surrogate model at the training set. Therefore, the optimization problem proposed by Viana et al. (2009) can be formulated as:

$$\begin{aligned} &Min \quad w^T C w \\ &w.r.t \quad w_i \\ &s.t \quad \sum_{i=1}^{M} w_i = 1 \end{aligned} \tag{12}$$

The optimal results can be obtained using Lagrange multipliers:

$$w = \frac{C^{-1}\mathbf{1}}{\mathbf{1}^T C^{-1}\mathbf{1}} \tag{13}$$

The weight factors in the above equation may be larger than one or less than zero, Viana et al. (2009) enforced the weight factors positive by solving Eq.(13) using only the diagonal elements of C. Then the weights of the ensemble of surrogates can be obtained using Eq.(13). For convenience, the average ensemble model proposed by Viana et al. (2009) is denoted by OWSd in this paper.

2.5 Pointwise ensemble model proposed by Acar (2010)

Acar (2010) proposed an ensemble of surrogates, in which the pointwise cross-validation error was adopted as the local error metric. The weight factors are calculated by the following equations:

$$w_i(x) = \frac{W_i^*(x)}{\sum_{j=1}^{M} W_j^*(x)} \tag{14}$$

$$W_i^*(x) = \sum_{j=1}^{N} W_{ji} I_j(x) \tag{15}$$

$$I_j(x) = \frac{1}{d_j^2(x)} \tag{16}$$

$$d_j(x) = \left\| x - x_j \right\| \tag{17}$$

where $d_j(x)$ represents the Euclidean distance between the prediction point x and the training point x_j. W_{ji} is equal to one for the *ith* surrogate model with the smallest cross-validation error at the *jth* training point and equal to zero for all other surrogate models at the *jth* training point. The pointwise ensemble model proposed by Acar (2010) is denoted by SP1 in this paper.

2.6 Pointwise ensemble model proposed by Lee and Choi (2014)

Lee and Choi (2014) developed a new ensemble of surrogates, in which the v nearest pointwise cross-validation error (vCV) was employed as the local error metric. The weight factors are evaluated by the following equations:

$$w_i(x) = \frac{\frac{1}{vCV_i^*(x)}}{\sum_{j=1}^{M} \frac{1}{vCV_j^*(x)}} \tag{18}$$

$$vCV_i^*(x) = \begin{cases} \alpha(x) \cdot \sqrt{\frac{1}{v}\sum_{k=1}^{v} (y_i(x) - \hat{y}_{i(k)}(x))^2} \\ for \ \ interpolation \ \ models \\ \sqrt{\frac{1}{v}\sum_{k=1}^{v} (y_i(x) - \hat{y}_{i(k)}(x))^2} \\ for \ \ regression \ \ models \end{cases} \tag{19}$$

$$\alpha(x) = 3\left(\frac{D_1(x)}{D_2(x)}\right)^2 - 2\left(\frac{D_1(x)}{D_2(x)}\right)^3 \tag{20}$$

where $D_1(x)$ represents the distance between the prediction point x and the closest training point to x, and $D_2(x)$ indicates the distance between the prediction point x and the second closest training point to x. The pointwise ensemble model proposed by Lee and Choi (2014) is denoted by PEM-vCV.

2.7 Pointwise ensemble model proposed by Liu et al. (2016)

Liu et al. (2016) proposed a pointwise ensemble of surrogates, in which the weight functions are calculated as follows:

$$w_j(x) = \begin{cases} if \quad x \neq x_i : \sum_{i=1}^{N} \frac{d_i^{-B_i\theta}\bar{W}_{ij}}{\sum d_i^{-B_i\theta}} \\ if \quad x = x_i : \bar{W}_{ij} \end{cases} \tag{21}$$

where $\bar{W}_{ij}$ refers to the observed weights of the *jth* surrogate at the *ith* training point, the 0-1 strategy is used to determine $\bar{W}_{ij}$. d_i indicates the Euclidean distance between the *ith* training point and the prediction point. And $B_i = (1/GMSE_u)/\max_{k=1,2,\ldots,M}(1/GMSE_k)$ is the normalized global accuracy of the *uth* surrogate model that has an observed weight with a value of 1 at the *ith* training point. θ represents a constant attenuation coefficient, which is obtained by minimizing the $\sqrt{GMSE}$ of the ensemble of surrogates as:

$$\begin{aligned} &Min \quad \sqrt{GMSE_e} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \left(y(x_i) - \tilde{y}_e^{(-i)}(x_i, \theta)\right)^2} \\ &w.r.t \quad \theta \end{aligned} \tag{22}$$

$$\tilde{y}_e^{(-i)}(x_i,\theta)=\sum_{j=1}^{M} w_j^{(-i)}(x_i,\theta)\tilde{y}_j^{(-i)}(x_i) \tag{23}$$

where $\tilde{y}_e^{(-i)}(x_i,\theta)$ represents the predicted value at x_i obtained by the ensemble model constructed with all the training points except the *ith* one. $w_j^{(-i)}(x_i,\theta)$ represents the predicted weight of the *jth* surrogate model at x_i, it can be obtained by deleting the *ith* row of the observed weights matrix $\bar{W}$. The pointwise ensemble model proposed by Liu et al. (2016) is denoted by OWPE.

3 Proposed pointwise ensemble of surrogates

Generally, the pointwise ensemble can be more accurate than the average ensemble as the former can adapt to local features of the prediction point. The local error measure which can accurately adapt to the local characteristics around the prediction point is the key to the pointwise ensemble of surrogates. This paper proposed a pointwise ensemble of surrogates with adaptive function and heuristic formulation (PEAH).

3.1 Weight function of the PEAH model

The adaptive function presented in this paper can adapt to the local characteristics around the prediction point. In the proposed PEAH, the weight function is calculated as follows:

$$w_i(x)=\frac{w_i^*(x)}{\sum_{j=1}^{M} w_j^*(x)} \tag{24}$$

$$w_i^*(x)=(E_i(x)+\alpha\bar{E}(x))^{\beta}\quad \alpha<1,\quad \beta<0 \tag{25}$$

$$E_i(x)=E_i\cdot adapt_i(x) \tag{26}$$

$$\bar{E}(x)=\frac{1}{M}\sum_{i=1}^{M} E_i(x) \tag{27}$$

where E_i indicates the $\sqrt{PRESS}$ of the *ith* surrogate model at the training set and $adapt_i(x)$ represents the adaptive function of the *ith* surrogate model at the prediction point x, the adaptive function is formulated as:

$$adapt_i(x)=g_i(x)^{[1+e^{(-1/u(x))}]} \tag{28}$$

where $g_i(x)$ represents the local feature function of the *ith* surrogate model at the prediction point x, calculated as:

$$g_i(x)=d(x)^T\cdot ac_i \tag{29}$$

where $d(x)=[d(x,x_1),d(x,x_2),...,d(x,x_N)]^T$ represents the distance vector, $d(x,x_N)$ refers to the Euclidean distance between the prediction point and the *Nth* training point. $ac_i=[\bar{e}_i(x_1),\bar{e}_i(x_2),...,\bar{e}_i(x_N)]^T$ is the normalized pointwise cross-validation error vector of the *ith* surrogate at the training set, and $\bar{e}_i(x_N)$ is determined by:

$$\bar{e}_i(x_N)=\frac{|e_i(x_N)|}{\sum_{k=1}^{M}|e_k(x_N)|} \tag{30}$$

From the above equations, we can observe that the local feature function $g_i(x)$ combines the normalized pointwise cross-validation vector ac_i and the distance information between the training point set and the prediction point x, thus $g_i(x)$ can reflect the local accuracy of the *ith* surrogate. The smaller value of the $g_i(x)$, the higher local accuracy of the *ith* surrogate model. And $u(x)$ represents the uncertainty of local accuracy for all the surrogates at the prediction point x, calculated as:

$$u(x)=\sqrt{\frac{\sum_{i=1}^{M}(g_i(x)-\bar{g}(x))^2}{M-1}} \tag{31}$$

$$\bar{g}(x)=\frac{\sum_{i=1}^{M} g_i(x)}{M} \tag{32}$$

A large value of $u(x)$ indicates a region near the prediction point x with high uncertainty of local accuracy. Combining the local feature function $g_i(x)$ and the uncertainty of local accuracy at the prediction point, the adaptive function $adapt_i(x)$ can reflect the local characteristics of the prediction point x. The larger the value of the $adapt_i(x)$ function, the larger the corresponding local error. Therefore, the value of $E_i(x)$ will increase with the increase of $adapt_i(x)$, and the weight of the *ith* stand-alone surrogate model $w_i^*(x)$ will be smaller (Since $\beta<0$). The proposed PEAH model involves the effective properties of the PBW model. And the weights of the PEAH model change with the prediction point, which can reflect local features of the region near the prediction point.

In order to better explain the characteristics of the adaptive function, we use an example that includes one-dimensional input space and two local feature functions $g_i(x)(i=1,2)$ to illustrate the trends of the adaptive function. We set a training point x_{tr} at the origin in the one-dimensional input space, and two stand-alone surrogate models are used to construct the ensemble. The cross-validation error of the first stand-alone surrogate model at the x_{tr} $(e_1(x_{tr}))$ is assumed to be 1. To illustrate the influence of the uncertainty of local accuracy u in the adaptive function, the cross-validation error of the second stand-alone surrogate model at the x_{tr} $(e_2(x_{tr}))$ is assumed to be 2, 4 and 6, respectively. Then the distance vector in the one-dimensional input space is $d(x)^T=|x|$. In the case of three different $e_2(x_{tr})$ values,

the first stand-alone surrogate's ac_1 value is $\frac{1}{3}$, $\frac{1}{5}$, and $\frac{1}{7}$, respectively. The corresponding ac_2 value of the second stand-alone surrogate is $\frac{2}{3}$, $\frac{4}{5}$, and $\frac{6}{7}$, respectively. Using Eqs. (28-32), we can obtain the adaptive functions of the two stand-alone surrogate models:

$$adapt_1(x) = (ac_1|x|)^{[1+e^{-(\frac{\sqrt{2}}{|ac_1-ac_2||x|})}]} \tag{33}$$

$$adapt_2(x) = (ac_2|x|)^{[1+e^{-(\frac{\sqrt{2}}{|ac_1-ac_2||x|})}]} \tag{34}$$

We set the value range of x to (0,2) to show the trends of the adaptive function. In the case of three different $e_2(x_{tr})$ values, we plotted the function curves of $adapt_1(x)$ and $adapt_2(x)$. The results are shown in Fig. 1.

As shown in Fig. 1, the solid red line $adapt_2(x)^{(1)}$, the solid black line $adapt_2(x)^{(2)}$, and the solid green line $adapt_2(x)^{(3)}$ are the function curves of the $adapt_2(x)$ when $e_2(x_{tr})$ is 2, 4, and 6, respectively. The dashed red line $adapt_1(x)^{(1)}$, the dashed black line $adapt_1(x)^{(2)}$, and the dashed green line $adapt_1(x)^{(3)}$ are the function curves of the $adapt_1(x)$ when $e_2(x_{tr})$ is 2, 4, and 6, respectively. When the value of x increases, it indicates that the distance between the prediction point and the training point gradually increases, and the values of all the adaptive functions increase. With the increase of $e_2(x_{tr})$, the value of $adapt_2(x)$ increases (See solid lines in Fig. 1), this shows that an increase in the local error value leads to an increase in the value of $adapt(x)$. In addition, when the error value $e_2(x_{tr})$ increases and the $e_1(x_{tr})$ remains the same, it means that the uncertainty of local accuracy u increases. Then from Fig. 1 we can observe

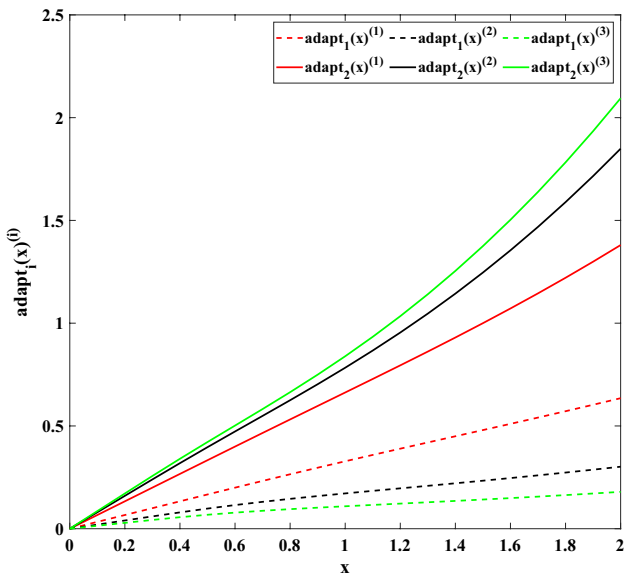

Fig. 1 The trends of the adaptive functions

that with the increase of the $e_2(x_{tr})$, the difference between the value of $adapt_1(x)$ and the value of $adapt_2(x)$ gradually increases. This is because when the uncertainty of local accuracy increases, the stand-alone surrogate model with higher local accuracy is given more importance in the proposed PEAH ensemble model.

3.2 Selection of parameters α and β

Parameters α and β are employed to control the importance of averaging and the importance of each surrogate in the ensemble, respectively (Goel et al. 2007). In the proposed PEAH, α and β are obtained by minimizing the *GMSE* value on the training set. The optimization problem is formulated as:

$$\begin{aligned} &Min \quad GMSE_P = \frac{1}{N}\sum_{i=1}^{N}(y_{(i)} - \hat{y}_{P(i)})^2 \\ &w.r.t \quad 0 < \alpha < 1, -2 < \beta < 0 \end{aligned} \tag{35}$$

where $\hat{y}_{P(i)}$ represents the response at the *ith* training point predicted by the PBW model constructed with all training points except the *ith* one. The parameter α controls the importance of averaging. It is a positive parameter, and the upper bound of the α is 1 (presented by (Goel et al. 2007)), so the lower limit and upper limit of α are 0 and 1, respectively. As for β, it is a negative parameter. The sensitivity analysis of the β in (Goel et al. 2007) indicates that if the value of β is too small, the accuracy of the ensemble may be reduced, and the recommended value of β is -1, so the interval (-2, 0) of β is selected as the bound for β. $\hat{y}_{P(i)}$ is calculated as:

$$\hat{y}_{P(i)} = \sum_{j=1}^{M} w_j \hat{y}_j(x_i) \tag{36}$$

where $\hat{y}_j(x_i)$ represents the predicted value at the *ith* training point x_i from the *jth* surrogate constructed with all training points except x_i.

3.3 Flowchart of the proposed PEAH model

The flowchart of the proposed PEAH model is shown in Fig. 2.

In general, the construction of the PEAH model consists of five steps:

Step 1: Generate the training set and the prediction points.
Step 2: Construct each stand-alone surrogate model using the training set.
Step 3: Optimize the parameters α and β using Eq.(35)
Step 4: Construct the PEAH model using Eqs.(24–32)

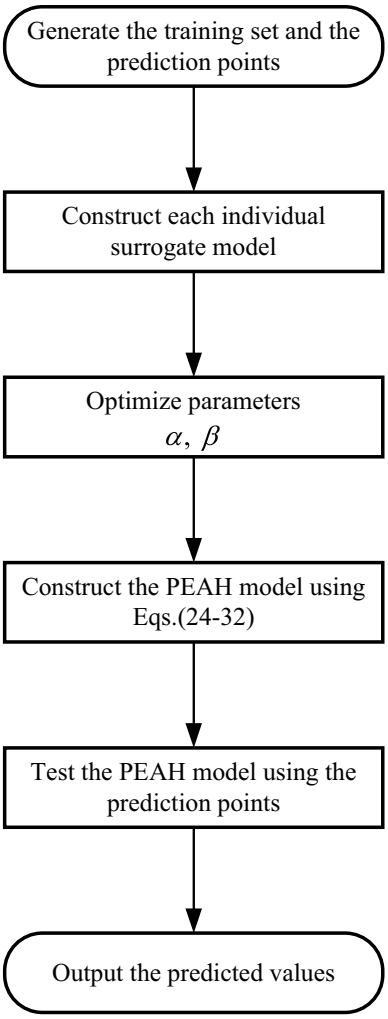

Fig. 2 Flowchart of the proposed PEAH model

Step 5: Test the PEAH model using the prediction points and output the predicted values.

4 Experiments and analysis

Six well-known ensembles of surrogates introduced in Sect. 2, including PBW, EP, OWSd, SP1, PEM-vCV and OWPE, are employed to compare with the proposed PEAH. PBW, EP and OWSd belong to average ensemble models. SP1, PEM-vCV and OWPE are pointwise ensemble models. Section 4.1 presents the introduction of the parameter setting and test examples (including twenty analytical test functions and two engineering design problems), the CFD modeling process of the shape design problem is introduced in Sect. 4.2. The introduction of the I-beam design problem is given in Sect. 4.3. The performance metrics used to evaluate the prediction accuracy of surrogate models are given in Sect. 4.4. Section 4.5 shows the simulation results of various surrogate models on the analytical test functions. The simulation results of different surrogate models on the shape design problem of the underwater robot are shown in Sect. 4.6, and the comparison results of surrogate models on the I-beam design problem are presented in Sect. 4.7. The comparison of computational costs of various ensembles of surrogates is given in Sect. 4.8.

4.1 Parameter setting and test examples

For fair comparison, the parameters used in those ensembles of surrogates are selected in accordance with the original references. Three typical stand-alone surrogate models, including RBF, Kriging and PRS, are used to construct the above ensembles of surrogates. These stand-alone surrogates include interpolation models and regression models. The RBF surrogate model is based on the multiquadric basis function with the constant $c = 1$. A Gaussian correlation function and a zero-order polynomial model are used in the Kriging surrogate model. A fully quadratic polynomial model is used to represent the PRS surrogate model. The parameter settings of the three stand-alone surrogate models are the same in all the ensembles of surrogates.

In order to fully verify the effectiveness of the proposed PEAH in this paper, we utilize twenty analytical functions that are widely applied for optimization from the literature (Acar 2010; Surjanovic and Bingham 2013; Lee and Choi 2014). Both low-dimensional and high-dimensional problems are included in these test functions. The mathematical formulations of these analytical functions are listed in Table 1. Parameters in these test functions are listed in Table 2.

As for the design of experiments, the Latin hypercube sampling technique is adopted. The routine *lhsdesign* set with *maximin* criterion with 20 iterations in MATLAB is utilized to generate a set of training points for each test function. In order to eliminate the influence of random sampling, 1000 different training sets are generated for comparison. 1000 prediction points are generated by the random sampling to test surrogate models constructed with the training set. The three ensembles of surrogates, including PEAH, OWPE and EP, are based on optimization problems. The *fmicon* function in MATLAB are utilized and sequential quadratic programming (SQP) is adopted as the optimizer for the above three ensembles of surrogates. The simulation process is performed on the computer with a AMD 16-Core processor clocking at 2.40 GHz, and the RAM of the desktop computer is 1 TB.

In addition to these analytical test functions, the CFD-based shape design problem of an underwater robot from the literature (Chen et al. 2021) and an I-beam design problem with the explicit function (Chen et al. 2019) are

Table 1 Analytical test functions

ID	Function Name	Mathematical formulation	Range
F1	Bukin N.6	$f(x) = 100\sqrt{\left\|x_2 - 0.01x_1^2\right\|} + 0.01\left\|x_1 + 10\right\|$	$x_1 \in [-15, -5]$, $x_2 \in [-3, 3]$
F2	Branin-Hoo	$f(x) = (x_2 - \frac{5.1}{4\pi^2}x_1^2 + \frac{5}{\pi}x_1 - 6)^2 + 10(1 - \frac{1}{8\pi})\cos(x_1) + 10$	$x_1 \in [-5, 10]$, $x_2 \in [0, 15]$
F3	Cross-in-Tray	$f(x) = -0.0001(\left\|\sin(x_1)\sin(x_2)\exp(\left\|100 - \frac{\sqrt{x_1^2+x_2^2}}{\pi}\right\|)\right\| + 1)^{0.1}$	$x_1 \in [-10, 10]$, $x_2 \in [-10, 10]$
F4	De Jong N.5	$f(x) = (0.002 + \sum_{i=1}^{25} \frac{1}{i+(x_1-a_{1i})^6+(x_2-a_{2i})^6})^{-1}$	
		where $a = \begin{pmatrix} -32 & -16 & 0 & 16 & 32 & -32 & \dots & 0 & 16 & 32 \\ -32 & -32 & -32 & -32 & -32 & -16 & \dots & 32 & 32 & 32 \end{pmatrix}$	$x_1 \in [-65.536, 65.536]$, $x_2 \in [-65.536, 65.536]$
F5	Drop-Wave	$f(x) = -\frac{1+\cos(12\sqrt{x_1^2+x_2^2})}{0.5(x_1^2+x_2^2)+2}$	$x_1 \in [-5.12, 5.12]$, $x_2 \in [-5.12, 5.12]$
F6	Holder Table	$f(x) = -\left\|\sin(x_1)\cos(x_2)\exp(\left\|1 - \frac{\sqrt{x_1^2+x_2^2}}{\pi}\right\|)\right\|$	$x_1 \in [-10, 10]$, $x_2 \in [-10, 10]$
F7	Levy N.13	$f(x) = \sin^2(3\pi x_1) + (x_1 - 1)^2[1 + \sin^2(3\pi x_2)] + (x_2 - 1)^2[1 + \sin^2(2\pi x_2)]$	$x_1 \in [-10, 10]$, $x_2 \in [-10, 10]$
F8	Shubert	$f(x) = (\sum_{i=1}^{5} i\cos((i+1)x_1 + i))(\sum_{i=1}^{5} i\cos((i+1)x_2 + i))$	$x_1 \in [-10, 10]$, $x_2 \in [-10, 10]$
F9	Six-Hump Camel	$f(x) = (4 - 2.1x_1^2 + \frac{x_1^4}{3})x_1^2 + x_1x_2 + (-4 + 4x_2^2)x_2^2$	$x_1 \in [-3, 3]$, $x_2 \in [-2, 2]$
F10	Hartmann 4-D	$f(x) = \frac{1}{0.839}[1.1 - \sum_{i=1}^{4} \alpha_i \exp(-\sum_{j=1}^{4} A_{ij}(x_j - P_{ij})^2)]$	$x_j \in [0, 1]$, $j = 1, 2, ..., 4$
F11	Shekel	$f(x) = -\sum_{i=1}^{p} (\sum_{j=1}^{4} (x_j - C_{ji})^2 + \beta_i)^{-1}$	$x_j \in [0, 10]$, $j = 1, 2, ..., 4$
F12	Hartmann 6-D	$f(x) = -\sum_{i=1}^{4} \alpha_i \exp(-\sum_{j=1}^{6} A_{ij}(x_j - P_{ij})^2)$	$x_j \in [0, 1]$, $j = 1, 2, ..., 6$
F13	Langermann	$f(x) = \sum_{i=1}^{q} c_i \exp(-\frac{1}{\pi}\sum_{j=1}^{6} (x_j - B_{ij})^2)\cos(\pi \sum_{j=1}^{6} (x_j - B_{ij}))^2$	$x_j \in [0, 10]$, $j = 1, 2, ..., 6$
F14	Levy	$f(x) = \sin^2(\pi w_1) + \sum_{j=1}^{5} (w_j - 1)^2[1 + 10\sin^2(\pi w_j + 1)] + (w_6 - 1)^2[1 + \sin^2(2\pi w_6)]$	
		where $w_j = 1 + \frac{x_j - 1}{4}$	$x_j \in [-10, 10]$, $j = 1, 2, ..., 6$
F15	Rastrigin	$f(x) = 60 + \sum_{j=1}^{6} [x_j^2 - 10\cos(2\pi x_j)]$	$x_j \in [-5.12, 5.12]$, $j = 1, 2, ..., 6$
F16	Styblinski-Tang	$f(x) = \frac{1}{2}\sum_{j=1}^{6} (x_j^4 - 16x_j^2 + 5x_j)$	$x_j \in [-5, 5]$, $j = 1, 2, ..., 6$
F17	Extended Rosenbrock	$f(x) = \sum_{j=1}^{8} [100(x_{j+1} - x_j^2)^2 + (x_j - 1)^2]$	$x_j \in [-5, 10]$, $j = 1, 2, ..., 9$
F18	Dixon-Price	$f(x) = (x_1 - 1)^2 + \sum_{j=2}^{12} j(2x_j^2 - x_{j-1})^2$	$x_j \in [-5, 10]$ $j = 1, 2, ..., 12$
F19	Power Sum	$f(x) = \sum_{i=1}^{12} [(\sum_{j=1}^{12} x_j^i) - b_i]^2$	$x_j \in [0, 12]$, $j = 1, 2, ..., 12$
F20	Sum of Different Power	$f(x) = \sum_{j=1}^{12} \left\|x_j\right\|^{j+1}$	$x_j \in [-1, 1]$, $j = 1, 2, ..., 12$

Table 2 Parameters used in some test functions

Hartmann 4-D & Hartmann 6-D functions	Shekel function	Langermann function	Power Sum function
$\alpha = \begin{pmatrix} 1.0, & 1.2, & 3.0, & 3.2 \end{pmatrix}^T$	$p = 10$	$q = 5$	$b = \begin{pmatrix} 8 \\ 18 \\ 44 \\ 114 \\ ... \\ 8 \\ 18 \\ 44 \\ 114 \end{pmatrix}^T$
$A = \begin{pmatrix} 10 & 3 & 17 & 3.5 & 1.7 & 8 \\ 0.05 & 10 & 17 & 0.1 & 8 & 14 \\ 3 & 3.5 & 1.7 & 10 & 17 & 8 \\ 17 & 8 & 0.05 & 10 & 0.1 & 14 \end{pmatrix}$	$\beta = \frac{1}{10}\begin{pmatrix} 1, & 2, & 2, & 4, & 4, & 6, & 3, & 7, & 5, & 5 \end{pmatrix}^T$	$c = \begin{pmatrix} 1, & 2, & 5, & 2, & 3 \end{pmatrix}$	
$P = 10^{-4}\begin{pmatrix} 1312 & 1696 & 5569 & 124 & 8283 & 5886 \\ 2329 & 4135 & 8307 & 3736 & 1004 & 9991 \\ 2348 & 1451 & 3522 & 2883 & 3047 & 6650 \\ 4047 & 8828 & 8732 & 5743 & 1091 & 381 \end{pmatrix}$	$C = \begin{pmatrix} 4 & 1 & 8 & 6 & 3 & 2 & 5 & 8 & 6 & 7 \\ 4 & 1 & 8 & 6 & 7 & 9 & 3 & 1 & 2 & 3.6 \\ 4 & 1 & 8 & 6 & 3 & 2 & 5 & 8 & 6 & 7 \\ 4 & 1 & 8 & 6 & 7 & 9 & 3 & 1 & 2 & 3.6 \end{pmatrix}$	$B = \begin{pmatrix} 3 & 5 \\ 5 & 2 \\ 2 & 1 \\ 1 & 4 \\ 7 & 9 \end{pmatrix}$	

employed to compare the performance of various ensembles of surrogates.

4.2 CFD modeling of the shape design problem

The shape has an important influence on the hydrodynamic performance of the underwater robot; it is necessary for the engineer to design a shape of the underwater robot with the least fluid resistance. In this engineering example, various surrogate models are constructed to estimate the fluid resistance of the underwater robot. The shape of the underwater robot shown in Fig. 3 is determined by 5 input variables, including the fore-body shape parameters $qf1$ and $qf2$, the after-body shape parameters $qa1$ and $qa2$, and the length of parallel middle body Lp. The design variables of this engineering problem is listed in Table 3.

The numerical process to obtain the fluid resistance of the underwater robot is shown in Fig. 4, which consists of three steps:

Step 1: UG 10.0 is employed to import the input variables and generate the geometric model, and the Parasolid file of the geometric model is then exported to the ICEM 20.0 for meshing.

Step 2: In this step, ICEM 20.0 is used to generate meshes for the flow field. The computational domain shown in Fig. 5 consists of a semi-sphere and a cylinder. The geometric model of the underwater robot is set as the wall. The bottom of the cylinder defines the pressure outlet (with reference pressure 0). The hemispherical surface and the side of the cylinder are selected as the Velocity-inlet (with 5m/s in the x-axis direction). The radius of the hemisphere is the same as the length of the WALL, and the length of the cylinder is

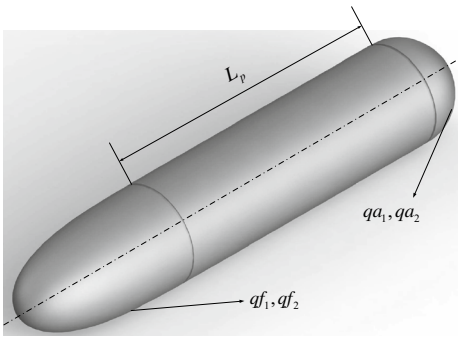

Fig. 3 The shape of the underwater robot

Table 3 Design variables of the shape design problem

Design variable	Range
x_1 (qf_1)	[2, 4]
x_2 (qf_2)	[0, 4]
x_3 (qa_1)	[2, 4]
x_4 (qa_2)	[10, 20]
x_5 ($L_p(cm)$)	[24, 30]

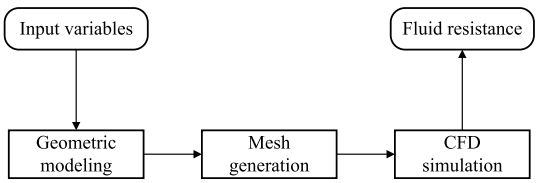

Fig. 4 The CFD numerical modeling process of the shape design problem

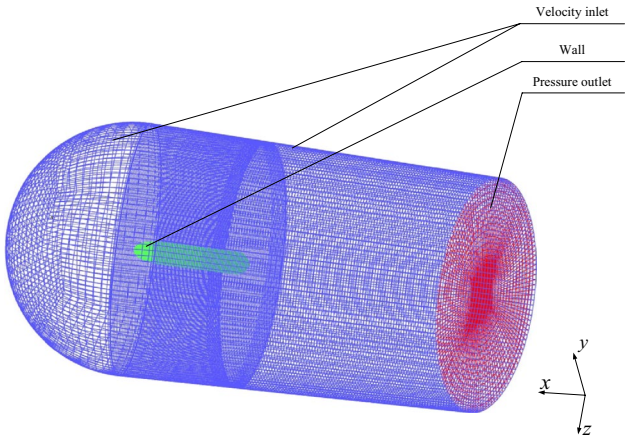

Fig. 5 Computational domain of the underwater robot

three times the length of the WALL. Structured meshes are utilized to generate nodes and elements.

Step 3: Then Fluent 20.0 is adopted as the CFD solver. RANS equations are employed as the control equations for the CFD simulation, and control equations are discretized by the finite volume method (FVM). As for the turbulence model, we choose the Standard $k-\varepsilon$ model, and the standard wall function is used to handle the area close to the wall. The wall in the flow field is defined as a no-slip wall. The SIMPLEC algorithm is employed as the solution method. The turbulent viscosity ratio for the Velocity-inlet and the pressure outlet is set to 2, and the intensity for the Velocity-inlet and the pressure outlet is 2%. The standard discretization scheme is applied for pressure. As for the turbulence kinetic energy, turbulence dissipation rate, and momentum, the second order upwind scheme is adopted. Under-relaxation factors are set to default values.

Through the above CFD modeling process, the fluid resistance of the underwater robot can be obtained. When the input variables change, this CFD simulation process will output corresponding fluid resistance of the underwater robot. A more detailed description of the numerical process of the design problem can be found in Chen et al. (2021). For the shape design problem, the Latin hypercube sampling technique is employed to generate the training set, and random sampling is used to generate prediction points. Due to the extensive computational cost of the CFD numerical simulation, only 100 prediction points are utilized to test the performance of various ensembles of surrogates.

4.3 The four variable I-beam design problem

The four variable I-beam design problem shown in Fig. 6 is chosen from Chen et al. (2019). And the output response for the I-beam design problem is the vertical deflection of the four variable I-beam, which is calculated from:

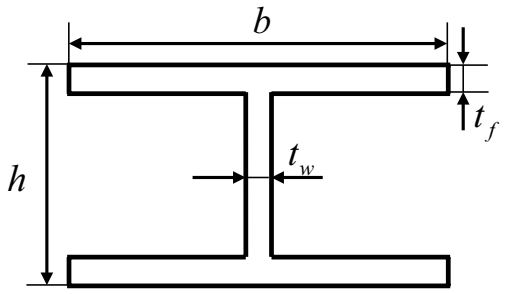

Fig. 6 The four variable I-beam design problem

$$f(b,h,t_w,t_f)=\frac{5000}{\frac{t_w(h-2t_f)^3}{12}+\frac{bt_f^3}{6}+2bt_f\left(\frac{h-t_f}{2}\right)^2} \tag{37}$$

where b and h represent the width and the height of the I-beam, and t_w, t_f indicate two thicknesses of this problem. And f is the vertical deflection of the four variable I-beam. The range of the four input variables of the I-beam problem is shown in Table 4. The surrogate models are employed to approximate the relationship between the output response and the input variables of the I-beam design problem, so the constraint of the I-beam design problem is not considered in this work.

4.4 Performance metrics

In this section, root mean square error (RMSE) and mean absolute percentage error (MAPE) are employed as two accuracy indexes to evaluate the prediction accuracy of different ensembles of surrogates. The mathematical formulations of RMSE and MAPE are shown in Eq.(38) and Eq.(39), respectively.

$$RMSE=\sqrt{\frac{1}{Pt}\sum_{i=1}^{Pt}(\hat{Y}_i-Y_i)^2} \tag{38}$$

$$MAPE=\frac{100\%}{Pt}\sum_{i=1}^{Pt}\left|\frac{\hat{Y}_i-Y_i}{Y_i}\right| \tag{39}$$

where $\hat{Y}_i$ and Y_i represent the predicted response and the true response at the *ith* prediction point, respectively. *Pt* indicates

Table 4 Design variables of the four variable I-beam design problem

Design variable	Range (in)
x_1 (b)	[10, 50]
x_2 (h)	[10, 99]
x_3 (t_w)	[0.9, 5]
x_4 (t_f)	[0.9, 5]

the number of prediction points. The smaller the value of RMSE and MAPE, the higher the prediction accuracy of the surrogate model.

4.5 Performance comparison based on analytical test functions

The number of training points for constructing each surrogate model depends on the number of input variables. To investigate the effect of the sample size on the prediction accuracy of different ensembles of surrogates, the number of training points for each analytical test function is set to $5n$, $10n$, $20n$ and $30n$, respectively. n represents the dimension of each test function. To give a fair comparison, the Wilcoxon rank-sum test is employed to compare the RMSE results obtained by PEAH and other surrogate models at a significance level of 0.05. Based on the surrogate model that provides the most accurate prediction accuracy on each test function, the average RMSE value and the average MAPE value of different surrogate models on the analytical test functions are normalized. The closer the value of the accuracy metric is to 1, the higher the accuracy of the surrogate model. The normalized average RMSE results for those test functions are listed in Tables 5, 6, 7, and 8, in which the "+/=/−" indicates that the proposed PEAH is better than, similar to or worse than its corresponding competitor respectively according to the Wilcoxon rank-sum test result on the test function. When the Wilcoxon rank-sum test results show that the two groups of data are significantly different, the group with the lower median is better. The Wilcoxon rank-sum test results on the twenty test functions are denoted by "*w*/*t*/*l*", which means that compared to the competitor, PEAH wins on w test functions, ties on t test functions and loses on l test functions, it can reflect the robustness of PEAH. To further compare the impact of the dimension of test functions on the prediction accuracy of different ensembles of surrogates, these test functions are divided into two groups, low-dimensional (From F1 to F11) and high-dimensional (From F12 to F20). In the cases of selecting different numbers of training points, the normalized average MAPE values of each ensemble of surrogates on the low-dimensional test functions and high-dimensional test functions are shown in Fig. 7.

As for the case with a small number of training points (5n), the statistical results listed in Table 5 indicate that PEAH shows a similar prediction accuracy to OWPE on low-dimensional test functions. Compared to the other six ensembles of surrogates, the Wilcoxon rank-sum test results on functions F12 to F20 (see Table 5) suggest that PEAH has superiority in the prediction accuracy on high-dimensional test functions. In addition, from Fig. 7 we can observe that PEAH and OWPE provide better prediction results than other five ensembles of surrogates on low-dimensional test functions in terms of the MAPE index, and EP shows the poorest prediction accuracy on high-dimensional test functions.

In the case with $5n$ training points, we can also observe that the RBF has the most accurate prediction ability on most of the test functions. Compared to the RBF, PEAH shows similar prediction accuracy on low-dimensional functions. The statistical results listed in Table 5 suggest that in cases with a small number of training points, ensembles of surrogates are able to filter out the stand-alone surrogate models that provide poor prediction accuracy.

As shown in Table 6, the normalized average RMSE values of PEAH on the twenty test functions suggest that PEAH shows better prediction accuracy than other six ensembles of surrogates on both low-dimensional and high-dimensional test functions in the case with $10n$ training points. Using the adaptive function, PEAH can better characterize the local characteristics of the prediction point. Compared to the PBW, PEAH performs better than PBW on 16 test functions in terms of the Wilcoxon rank-sum test results on the twenty test functions (see the last row of Table 6), and the two ensembles of surrogates show similar prediction accuracy in the other four test functions. The statistical results listed in Table 6 indicate that PEAH shows better performance in prediction accuracy than SP1 on 19 test functions. From the normalized average MAPE results shown in Fig. 7, we can observe that PBW provides the poorest prediction results on the low-dimensional test problems, and the prediction accuracy of EP on the high-dimensional test functions in the case with $10n$ training points is poor. Compared with stand-alone surrogate models, although PEAH only performs the best on two test functions (F11 and F12), the Wilcoxon rank-sum test results on the twenty test functions suggest that the robustness of PEAH is better than the three stand-alone surrogate models in the case with $10n$ training points.

As for the case with $20n$ training points, the statistical results listed in Table 7 suggest that PEAH and EP show better performance in prediction accuracy on high-dimensional test functions than other five ensembles of surrogates, and PEAH provides better prediction accuracy than EP on low-dimensional test functions. Compared to PBW, PEAH exhibits better prediction accuracy than PBW in 19 of the 20 test functions, which verifies the effectiveness of PEAH. From Fig. 7, we can observe that PBW provides the poorest prediction results on the twenty test functions in the case with $20n$ training points. In contrast, PEAH provides better prediction results on high-dimensional test functions than other six ensembles of surrogates (see Fig. 7). The Wilcoxon rank-sum test results shown in Table 7 indicate that the robustness of PEAH is better than the three stand-alone surrogate models. Although PRS shows the best prediction accuracy on some low-dimensional test functions (F3-F8), the performance of PRS on other test functions is not stable. It means that

Table 5 Normalized average RMSE values of different surrogate models on the test functions with 5*n* training points

Function	PEAH	PBW	OWPE	SP1	EP	PEM-vCV	OWSd	RBF	Kriging	PRS
F1	1.0323	1.0027	1.0516	1.0464	1.2002	1.0007	**1**	1.1750	1.0702	1.3367
		−	=	=	+	−	−	+	=	+
F2	1.0887	1.0746	1.0900	1.0870	1.2665	1.0679	1.0712	1.2293	**1**	1.4518
		=	=	=	+	=	=	+	−	+
F3	1.0450	1.0473	1.0542	1.0609	1.1237	1.0630	1.0454	**1**	1.2695	1.0761
		=	=	=	+	+	=	−	+	=
F4	1.0169	1.0746	1.0269	1.0598	1.0432	1.0918	1.0559	**1**	1.4013	1.0925
		+	=	+	=	+	+	=	+	=
F5	1.0301	1.0609	1.0427	1.0621	1.1045	1.0829	1.0488	**1**	1.3917	1.0961
		+	=	+	+	+	+	−	+	=
F6	1.0167	1.0093	1.0207	1.0171	1.1077	1.0153	1.0122	**1**	1.0868	1.0373
		=	=	=	+	=	=	−	+	=
F7	1.0547	1.0397	1.0552	1.0612	1.0958	1.0732	1.0378	1.1695	1.2138	**1**
		=	=	=	+	+	=	+	+	−
F8	1.0367	1.0681	1.0470	1.0644	1.0832	1.0823	1.0594	**1**	1.3187	1.0912
		+	=	=	+	+	=	−	+	=
F9	1.0405	1.0324	1.0584	1.0425	1.1833	1.0479	1.0344	**1**	1.2539	1.0316
		=	+	=	+	+	=	−	+	−
F10	1.0458	1.0995	1.0547	1.1067	1.0724	1.1019	1.0565	1.2381	**1**	1.8224
		+	=	+	+	+	+	+	−	+
F11	**1**	1.0268	1.0117	1.0483	1.0304	1.0327	1.0062	**1**	1.0006	1.6001
		+	=	+	+	+	=	=	=	+
F12	1.0288	1.1559	1.0380	1.1663	1.0612	1.1820	1.0447	1.2134	**1**	4.0887
		+	+	+	+	+	+	+	+	+
F13	1.0360	1.0438	1.0457	1.0469	1.1174	1.0522	1.0437	**1**	1.1813	1.0869
		=	=	=	+	+	=	−	+	=
F14	1.1184	1.1341	1.1254	1.2339	1.1399	1.1437	1.0945	**1**	1.1132	3.3136
		+	=	+	+	+	+	−	+	+
F15	1.0723	1.1215	1.0872	1.2179	1.1065	1.1411	1.0654	**1**	1.0870	3.7356
		+	+	+	+	+	=	−	+	+
F16	1.0787	1.1280	1.0953	1.2145	1.1347	1.1339	1.0668	**1**	1.0564	3.7768
		+	+	+	+	+	+	−	+	+
F17	1.0446	1.3879	1.0766	1.3516	1.0060	1.3629	1.1633	**1**	1.7545	6.5810
		+	+	+	−	+	+	−	+	+
F18	1.0316	1.2445	1.0534	1.2069	**1**	1.2338	1.1315	1.0162	1.7977	3.0264
		+	+	+	−	+	+	−	+	+
F19	1.0119	1.2144	1.0253	1.1466	1.7732	1.1787	1.0752	**1**	1.2692	4.1852
		+	+	+	+	+	+	−	+	+
F20	1.0238	1.2421	1.0352	1.2284	1.0167	1.2276	1.1058	**1**	1.4565	3.6957
		+	+	+	−	+	+	−	+	+
*w**t**l*		13\6\1	8\12\0	12\8\0	16\1\3	17\2\1	10\9\1	5\2\13	16\2\2	12\6\2

the robustness of PRS on the twenty test functions in the case with 20*n* training points is unsatisfactory. In addition, among the nine high-dimensional test functions, the six best RMSE values are obtained by ensembles of surrogates (see Table 7), indicating that ensembles of surrogates are able to show better prediction accuracy than the three stand-alone surrogate models on high-dimensional test functions.

The statistical results of the surrogate models for the normalized average RMSE in Table 8 suggest that PEAH shows better prediction accuracy than other six ensembles of surrogates on low-dimensional test functions in the case with 30*n* training points, and OWPE provides

Table 6 Normalized average RMSE values of different surrogate models on the test functions with $10n$ training points

Function	PEAH	PBW	OWPE	SP1	EP	PEM-vCV	OWSd	RBF	Kriging	PRS
F1	1.0245	1.0442	1.0552	1.1148	1.0877	1.0760	1.0181	**1**	2.6141	1.3836
		+	+	+	+	+	=	−	+	+
F2	1.0614	1.0401	1.0701	1.0486	1.1686	**1**	1.0483	1.0440	1.3337	1.6299
		=	=	=	+	−	=	=	+	+
F3	1.0280	1.1481	1.0508	1.2792	1.1009	1.2960	1.0417	1.1182	4.5538	**1**
		+	+	+	+	+	+	+	+	−
F4	1.0612	1.1231	1.0396	1.1905	1.0658	1.2428	1.0306	1.0934	5.0248	**1**
		+	−	+	−	+	−	+	+	−
F5	1.0009	1.1131	1.0190	1.2149	1.0651	1.2537	**1**	1.1013	5.5675	1.0016
		+	+	+	+	+	=	+	+	=
F6	1.0377	1.0704	1.0584	1.0965	1.1314	1.1145	1.0388	1.0480	1.9503	**1**
		+	+	+	+	+	+	+	+	−
F7	1.0606	1.1206	1.0893	1.1636	1.1121	1.1981	1.0729	1.1701	2.2247	**1**
		+	+	+	+	+	+	+	+	−
F8	1.0453	1.1424	1.0536	1.2111	1.0887	1.2619	1.0601	1.1292	3.5975	**1**
		+	=	+	+	+	+	+	+	−
F9	1.0521	1.0395	1.1052	1.0885	1.1923	1.0537	1.0386	**1**	1.4631	1.1920
		=	+	+	+	=	=	−	+	+
F10	1.0146	1.0669	1.0241	1.0671	1.0424	1.0633	1.0434	1.3761	**1**	1.6032
		+	=	+	+	+	+	+	−	+
F11	**1**	1.0304	1.0093	1.0218	1.0014	1.0269	1.0214	1.0102	1.0867	1.2388
		+	+	+	=	+	+	+	+	+
F12	**1**	1.0634	1.0152	1.0348	1.0210	1.0586	1.0438	1.2674	1.0107	1.3972
		+	+	+	+	+	+	+	+	+
F13	1.0593	1.1617	1.0793	1.2029	1.1385	1.2732	1.0788	1.1490	3.0088	**1**
		+	=	+	+	+	+	+	+	−
F14	1.0146	1.0163	1.0211	1.0236	1.0266	1.0189	1.0164	**1**	1.1420	1.1593
		=	+	+	+	+	=	−	+	+
F15	1.0138	1.0166	1.0219	1.0259	1.0293	1.0185	1.0161	**1**	1.0934	1.1637
		=	+	+	+	+	=	−	+	+
F16	1.0126	1.0178	1.0206	1.0257	1.0328	1.0184	1.0162	**1**	1.0672	1.1768
		+	+	+	+	+	+	−	+	+
F17	1.0158	1.1047	1.0279	1.0855	**1**	1.1038	1.0565	1.0394	2.0402	1.2327
		+	+	+	−	+	+	+	+	+
F18	1.0090	1.0986	1.0191	1.0680	**1**	1.0912	1.0459	1.0332	2.1281	1.4310
		+	=	+	−	+	+	+	+	+
F19	1.0102	1.1153	1.0292	1.0852	1.1593	1.1018	1.0810	**1**	1.3677	1.7710
		+	+	+	+	+	+	−	+	+
F20	1.0139	1.1189	1.0326	1.1104	1.0125	1.1132	1.0814	**1**	1.6903	1.6332
		+	+	+	=	+	+	−	+	+
$w\backslash t\backslash l$		16\4\0	14\5\1	19\1\0	15\2\3	18\1\1	13\6\1	12\1\7	19\0\1	13\1\6

similar prediction accuracy to PEAH on high-dimensional test functions. The normalized average MAPE on the twenty test functions shown in Fig. 7 indicate that PEAH and OWPE show better prediction results than other five ensembles of surrogates. From Tables 5, 6, 7, and 8 we can conclude that the PEAH can provide better performances in these test functions with a better balance between accuracy and robustness. In addition, it is worth noting that the prediction accuracy of the stand-alone surrogate model on the test functions changes with the number of training points. As for the case with $5n$ training points, RBF performs better than Kriging and PRS. However, the prediction accuracy of PRS improved with the number of training points, and PRS provides better prediction results than RBF and Kriging in the cases with $10n$, $20n$ and $30n$ training points.

Table 7 Normalized average RMSE values of different surrogate models on the test functions with $20n$ training points

Function	PEAH	PBW	OWPE	SP1	EP	PEM-vCV	OWSd	RBF	Kriging	PRS
F1	**1**	1.3782	1.0231	1.9016	1.0484	1.7077	1.0546	1.0023	17.2037	1.8665
		+	=	+	+	+	+	=	+	+
F2	1.0873	1.0992	1.0764	1.0701	1.0261	**1**	1.0577	1.0730	1.7993	2.4348
		+	=	−	−	−	=	=	+	+
F3	1.0097	1.4817	1.0493	2.6700	1.0645	1.9718	1.0209	1.1974	44.4172	**1**
		+	+	+	+	+	+	+	+	−
F4	1.0465	1.4707	1.0430	2.0318	1.0330	1.6054	1.0137	1.1175	55.2022	**1**
		+	=	+	−	+	−	+	+	−
F5	1.0200	1.5515	1.0506	2.6019	1.0423	1.7370	1.0171	1.2605	58.6370	**1**
		+	+	+	+	+	+	+	+	−
F6	1.0161	1.2359	1.0454	1.6871	1.1024	1.6180	1.0171	1.1232	11.8035	**1**
		+	+	+	+	+	=	+	+	−
F7	1.0386	1.2964	1.0709	1.8301	1.0934	1.7165	1.0488	1.2082	14.3674	**1**
		+	+	+	+	+	+	+	+	−
F8	1.0250	1.4444	1.0398	2.2284	1.0482	1.9445	1.0523	1.2393	35.4572	**1**
		+	+	+	+	+	+	+	+	−
F9	**1**	1.1234	1.0315	1.0617	1.0928	1.0552	1.0564	1.3280	1.0145	2.4236
		+	+	+	+	+	+	+	=	+
F10	1.0310	1.0919	1.0090	1.0792	**1**	1.0817	1.0587	1.4053	1.0002	2.1119
		+	−	+	−	+	+	+	−	+
F11	1.0035	1.0462	1.0073	1.0228	**1**	1.0393	1.0287	1.0073	1.3919	1.2882
		+	=	+	−	+	+	=	+	+
F12	**1**	1.0481	1.0115	1.0283	1.0117	1.0413	1.0349	1.2710	1.0068	1.3924
		+	+	+	+	+	+	+	=	+
F13	1.0461	1.3962	1.0784	1.9307	1.0918	1.9472	1.0622	1.3488	22.1010	**1**
		+	+	+	+	+	+	+	+	−
F14	1.0024	1.0095	1.0017	1.0057	**1**	1.0147	1.0067	1.0133	1.1552	1.0034
		+	=	+	−	+	+	+	+	=
F15	1.0090	1.0145	1.0113	1.0082	1.0050	1.0197	1.0121	1.0273	1.1512	**1**
		+	=	=	−	+	+	+	+	−
F16	**1**	1.0022	1.0062	1.0019	1.0068	1.0051	1.0013	1.0043	1.1072	1.0239
		=	+	=	+	+	=	+	+	+
F17	1.0119	1.1252	1.0238	1.1171	**1**	1.1314	1.0534	1.0662	2.2731	1.0393
		+	+	+	−	+	+	+	+	+
F18	1.0081	1.1201	1.0106	1.0967	**1**	1.1164	1.0399	1.1210	2.4780	1.0255
		+	=	+	−	+	+	+	+	+
F19	1.0151	1.0615	1.0251	1.0542	1.0737	1.0600	1.0508	**1**	1.3926	1.1810
		+	+	+	+	+	+	−	+	+
F20	**1**	1.0335	1.0018	1.0356	1.0042	1.0302	1.0094	1.0142	1.7408	1.0595
		+	=	+	+	+	+	+	+	+
$w\backslash t\backslash l$		19\1\0	11\8\1	17\2\1	12\0\8	19\0\1	16\3\1	16\3\1	17\2\1	11\1\8

4.6 Performance comparison based on the shape design problem

To investigate the effect of the number of training points on the prediction accuracy of different surrogate models. The number of training points for the shape design problem in this section is set to $5n$, $10n$, $20n$, $30n$, respectively. The simulation results of different surrogate models on the shape design problem of the underwater robot are shown in Table 9, Table 10 and Fig. 8. The data in bold in Tables 9, 10 represent the smallest RMSE and MAPE values in each group, respectively.

From Fig. 8(a) we can observe that the prediction accuracy of different ensembles of surrogates on the engineering problem increases with the number of training points in terms of the RMSE values. As listed in Table 9, PEAH

Table 8 Normalized average RMSE values of different surrogate models on the test functions with 30n training points

Function	PEAH	PBW	OWPE	SP1	EP	PEM-vCV	OWSd	RBF	Kriging	PRS
F1	**1**	1.7837	1.0221	4.2996	1.0394	2.2998	1.0255	1.0015	48.1104	2.1367
		+	=	+	+	+	+	=	+	+
F2	1.1785	1.2005	1.0778	1.0623	1.0796	**1**	1.0669	1.1493	1.3307	4.1996
		+	−	−	=	−	=	+	+	+
F3	1.0088	1.8453	1.0763	5.8468	1.0446	2.4861	1.0215	1.2720	100.2829	**1**
		+	+	+	+	+	+	+	+	−
F4	1.0583	2.1283	1.0678	4.6013	1.0283	1.6739	1.0156	1.1456	142.8551	**1**
		+	+	+	−	+	−	+	+	−
F5	1.0232	2.3857	1.0640	6.2160	1.0351	2.1265	1.0275	1.4039	160.4234	**1**
		+	+	+	+	+	+	+	+	−
F6	1.0050	1.4016	1.0505	2.9192	1.0901	2.0648	**1**	1.1414	28.6835	1.0019
		+	+	+	+	+	=	+	+	=
F7	1.0404	1.5256	1.0863	3.5653	1.0838	2.5608	1.0487	1.2756	41.5158	**1**
		+	+	+	+	+	+	+	+	−
F8	1.0169	1.8036	1.0454	4.3747	1.0288	2.5454	1.0463	1.2764	88.7846	**1**
		+	+	+	+	+	+	+	+	−
F9	1.0371	2.4632	1.0246	1.5918	1.0890	1.9490	1.1359	7.4456	**1**	20.1925
		+	=	+	+	+	+	+	=	+
F10	1.1221	1.1264	1.0685	1.1009	**1**	1.1033	1.1080	1.4219	1.0695	2.7520
		=	−	−	−	−	−	+	−	+
F11	1.0147	1.1040	1.0151	1.0463	**1**	1.0907	1.0515	1.0179	2.1243	1.4443
		+	=	+	−	+	+	=	+	+
F12	**1**	1.0517	1.0053	1.0326	1.0107	1.0414	1.0355	1.2687	1.0025	1.5211
		+	=	+	+	+	+	+	=	+
F13	1.0350	1.7615	1.0829	3.9742	1.0728	2.7495	1.0489	1.4379	63.2121	**1**
		+	+	+	+	+	+	+	+	−
F14	1.0167	1.0323	1.0089	1.0230	1.0055	1.0384	1.0265	1.0434	1.2035	**1**
		+	−	+	−	+	+	+	+	−
F15	1.0267	1.0448	1.0157	1.0306	1.0089	1.0521	1.0376	1.0679	1.2417	**1**
		+	−	+	−	+	+	+	+	−
F16	**1**	1.0040	1.0044	1.0021	1.0007	1.0077	1.0021	1.0108	1.1285	1.0149
		+	=	+	=	+	+	+	+	+
F17	1.0096	1.1071	1.0210	1.1083	**1**	1.1156	1.0454	1.0464	2.1456	1.0369
		+	+	+	−	+	+	+	+	+
F18	1.0054	1.1102	1.0049	1.0942	**1**	1.1099	1.0298	1.1064	2.6547	1.0136
		+	=	+	−	+	+	+	+	+
F19	1.0129	1.0665	1.0276	1.0656	1.0862	1.0677	1.0505	**1**	1.4525	1.0836
		+	+	+	+	+	+	−	+	+
F20	1.0004	1.0454	**1**	1.0465	1.0008	1.0428	1.0124	1.0461	1.8815	1.0199
		+	=	+	=	+	+	+	+	+
$w\backslash t\backslash l$		19\1\0	9\7\4	18\0\2	10\3\7	18\0\2	16\2\2	17\2\1	17\2\1	11\1\8

shows the best prediction accuracy on the shape design problem in the cases with different numbers of training points (5n, 10n, 20n and 30n). And in the case with 10n training points, PEAH, OWPE and EP show the same prediction accuracy. As for the MAPE values shown in Fig. 8(b), with the increase of training point numbers from 5n to 10n, the MAPE values of ensembles of surrogates on the shape design problem drops significantly. When the number of training points increases from 10n to 20n, the MAPE values of these surrogate models has a small drop except EP. From the MAPE values of surrogate models on the shape design problem listed in Table 10, we can observe that OWPE provides the smallest value of MAPE in the case with 5n training points and EP shows better prediction results than other surrogate models in the case with 10n training points. PEAH provides the best MAPE values in the cases with

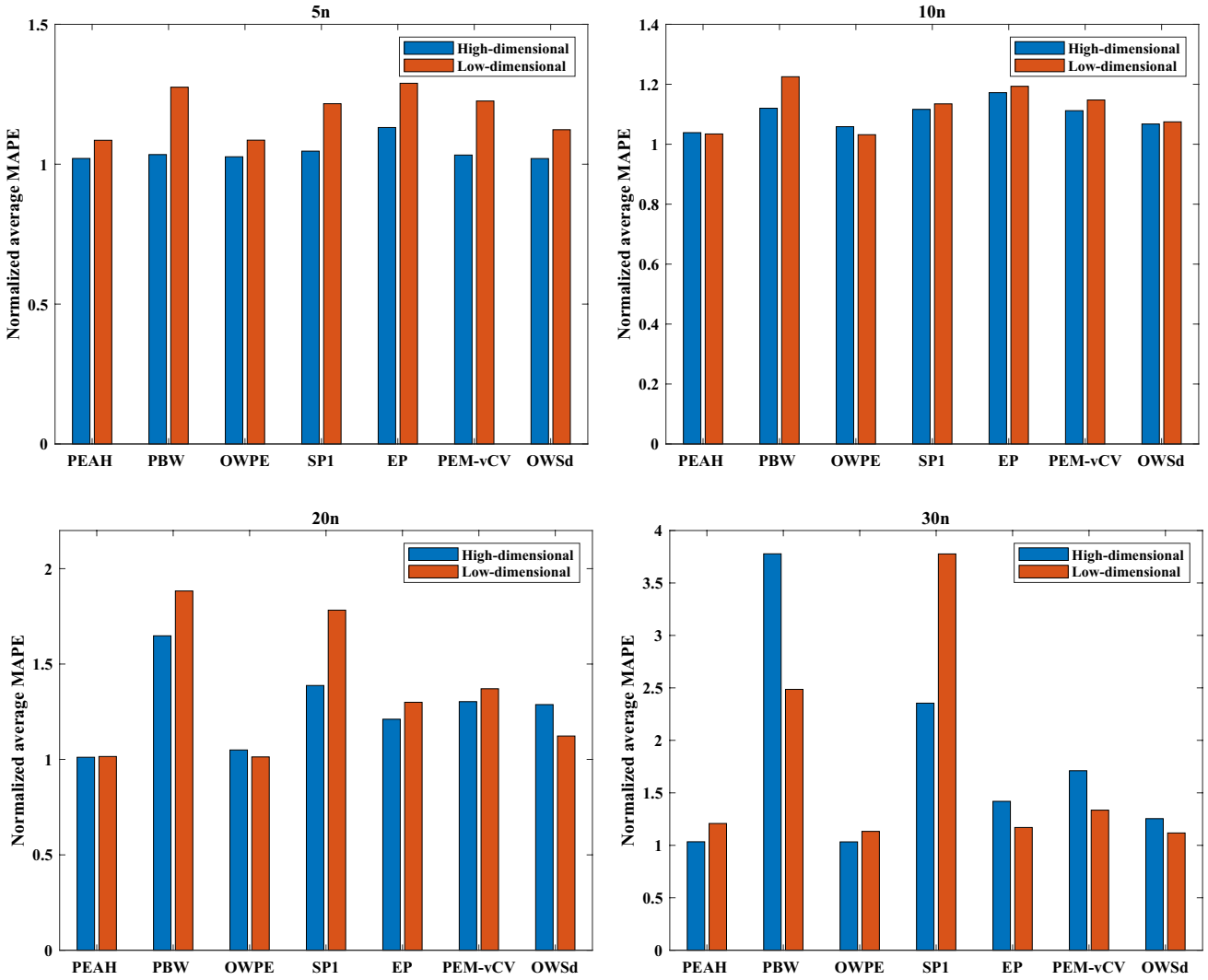

Fig. 7 Normalized average MAPE values of the ensembles of surrogate on test functions with different numbers of training points

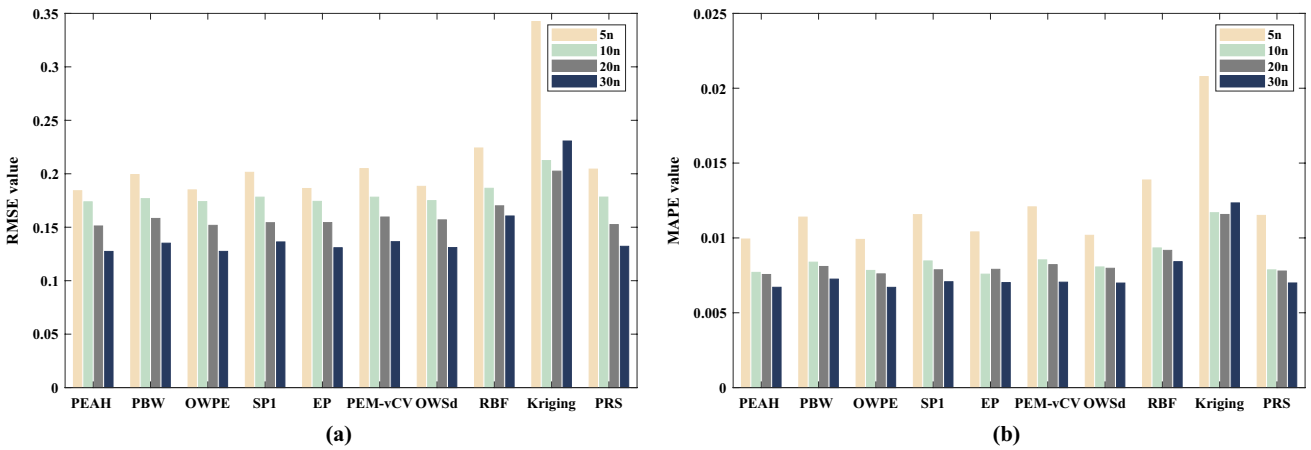

Fig. 8 Prediction accuracy of surrogate models for the shape design problem with different numbers of training points using accuracy metrics (a) RMSE (b) MAPE

$20n$ and $30n$ training points. In general, PEAH, OWPE and EP exhibit better prediction accuracy on the shape design problem than other ensembles of surrogates. In addition, it is worth noting that the prediction accuracy of PEAH on the shape design problem is better than its stand-alone surrogate models in all the cases.

4.7 Performance comparison based on the I-beam design problem

The number of training points for the I-beam design problem with the explicit function is set to $5n$, $10n$, $20n$, $30n$, respectively. In order to eliminate the influence of random sampling, 1000 different training sets are generated for comparison. 1000 prediction points are generated by random sampling to test surrogate models constructed with the training set. The simulation results of different surrogate models on the I-beam design problem are shown in Table 11, Table 12, Fig. 9 and Fig. 10. The data in bold in Tables 11, 12 represent the smallest average RMSE and MAPE values in each group, respectively.

The statistical results listed in Table 11 indicate that the prediction accuracy of different surrogate models on the I-beam design problem increases with the number of training points in terms of the RMSE values. As shown in Table 11 and Fig. 9, PEAH and RBF show the best prediction accuracy on the I-beam design problem in the cases with $10n$, $20n$ and $30n$ training points. The boxplot of different ensembles of surrogates on the I-beam design problem shown in Fig. 10 suggests that that PEAH has higher

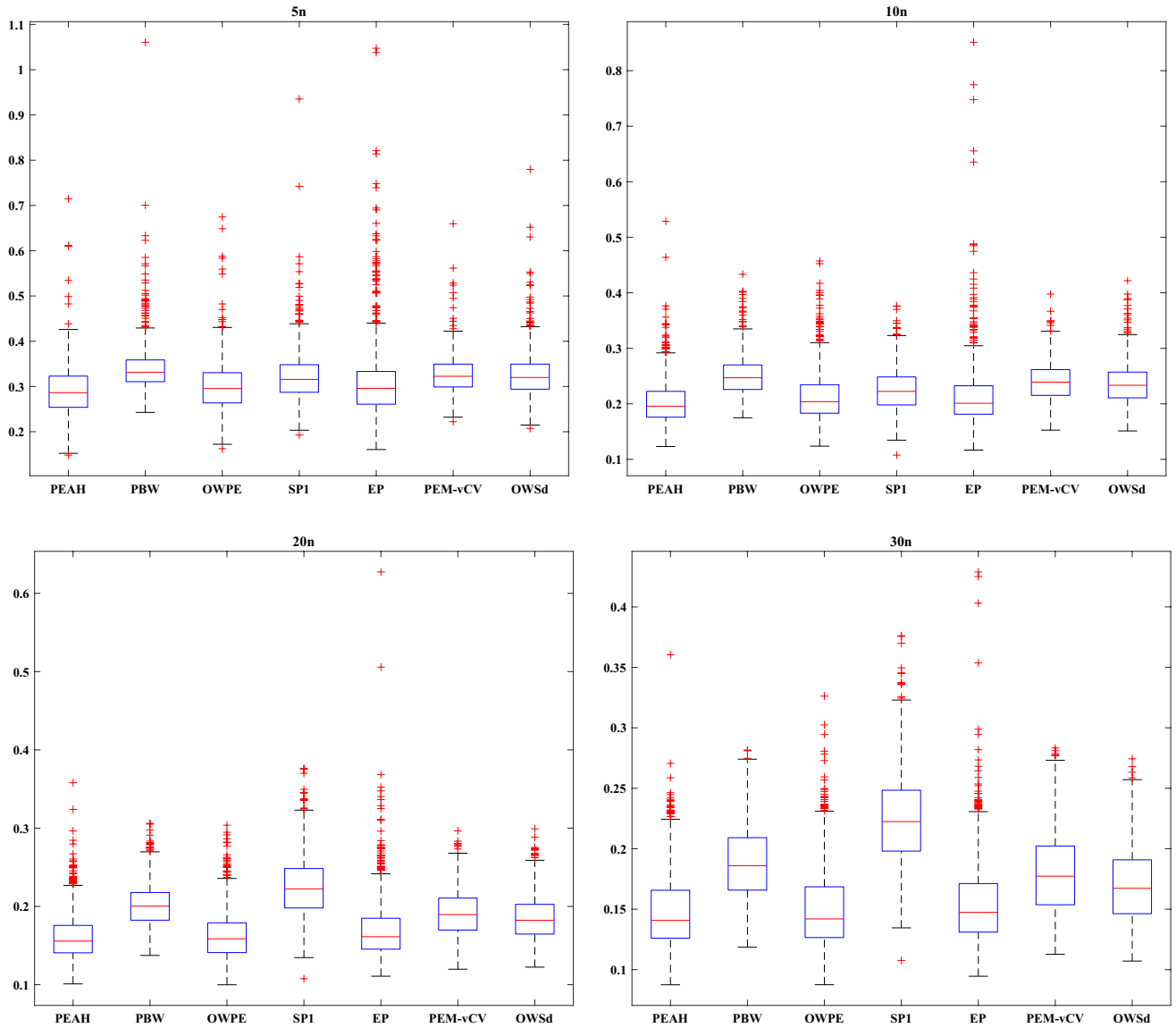

Fig. 9 Boxplot of the statistical results for RMSE values obtained by different ensembles of surrogates on the I-beam design problem

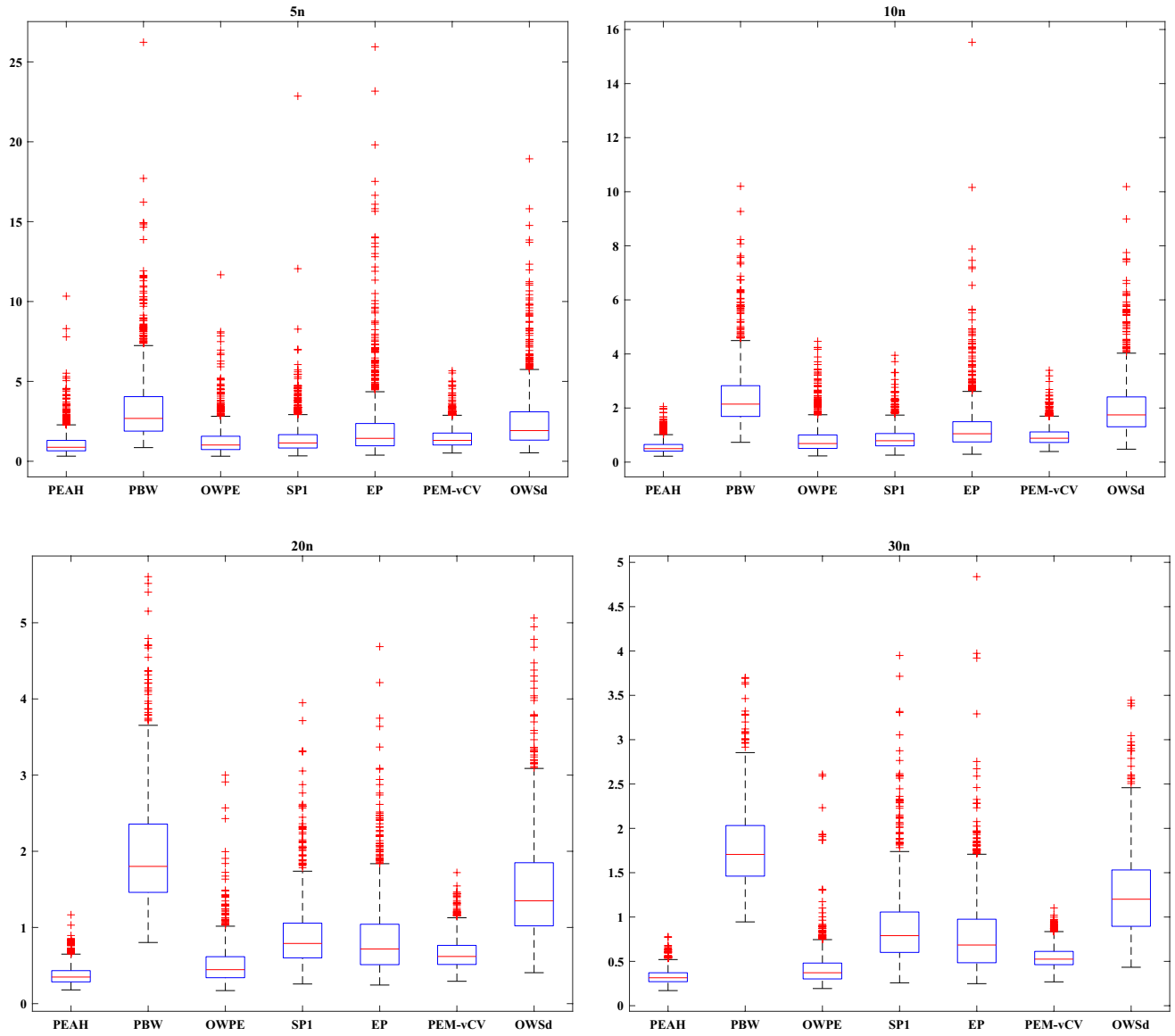

Fig. 10 Boxplot of the statistical results for MAPE values obtained by different ensembles of surrogates on the I-beam design problem

prediction accuracy and stability of prediction than the other six ensembles of surrogates. As listed in Table 12, RBF and PEAH provide better MAPE results than other surrogate models. Compared to other ensembles of surrogates, the simulation results on the I-beam design problem with the explicit function indicate that the proposed PEAH is able to accurately track the stand-alone surrogate model that performs the best and filter out the stand-alone surrogate models that show poor prediction accuracy.

4.8 Computational cost

The total running time of each ensemble of surrogates on the twenty test functions is shown in Fig. 11 and Table 13. The data in bold in Table 13 represent the smallest computational time in each group. The computational time of each ensemble of surrogates is calculated using the MATLAB routine *tic* and *toc*. From Fig. 11 we can observe that with the number of training points, the computational costs of ensembles of surrogates increase. PBW and OWSd have smaller computational costs than the other five ensembles of surrogates. Since PEAH, OWPE and EP include optimization

Table 9 RMSE values of surrogate models on the shape design problem with different numbers of training points

Number of **DOE**	PEAH	PBW	OWPE	SP1	EP	PEM-vCV	OWSd	RBF	Kriging	PRS
5n	**1.84E-01**	1.99E-01	1.85E-01	2.01E-01	1.86E-01	2.05E-01	1.88E-01	2.24E-01	3.42E-01	2.04E-01
10n	**1.74E-01**	1.77E-01	**1.74E-01**	1.78E-01	**1.74E-01**	1.78E-01	1.75E-01	1.86E-01	2.12E-01	1.78E-01
20n	**1.51E-01**	1.58E-01	1.52E-01	1.54E-01	1.54E-01	1.60E-01	1.57E-01	1.70E-01	2.02E-01	1.52E-01
30n	**1.27E-01**	1.35E-01	**1.27E-01**	1.36E-01	1.31E-01	1.37E-01	1.31E-01	1.61E-01	2.31E-01	1.32E-01

Table 10 MAPE values of surrogate models on the shape design problem with different numbers of training points

Number of DOE	PEAH	PBW	OWPE	SP1	EP	PEM-vCV	OWSd	RBF	Kriging	PRS
5n	9.93E-03	1.14E-02	**9.90E-03**	1.16E-02	1.04E-02	1.21E-02	1.02E-02	1.39E-02	2.08E-02	1.15E-02
10n	7.70E-03	8.38E-03	7.83E-03	8.46E-03	**7.58E-03**	8.53E-03	8.06E-03	9.33E-03	1.17E-02	7.87E-03
20n	**7.56E-03**	8.09E-03	7.60E-03	7.87E-03	7.90E-03	8.22E-03	7.97E-03	9.17E-03	1.16E-02	7.79E-03
30n	**6.70E-03**	7.20E-03	**6.70E-03**	7.10E-03	7.00E-03	7.00E-03	7.00E-03	8.40E-03	1.23E-02	7.00E-03

Table 11 Average RMSE values of surrogate models on the I-beam design problem with different numbers of training points

Number of DOE	PEAH	PBW	OWPE	SP1	EP	PEM-vCV	OWSd	RBF	Kriging	PRS
5n	2.88E-01	3.39E-01	2.98E-01	3.21E-01	3.08E-01	3.26E-01	3.24E-01	**2.87E-01**	4.11E-01	5.20E-01
10n	**2.03E-01**	2.50E-01	2.12E-01	2.25E-01	2.14E-01	2.40E-01	2.36E-01	**2.03E-01**	3.59E-01	3.37E-01
20n	**1.61E-01**	2.02E-01	1.63E-01	1.75E-01	1.70E-01	1.91E-01	1.85E-01	**1.61E-01**	3.20E-01	2.98E-01
30n	**1.49E-01**	1.89E-01	1.50E-01	1.63E-01	1.55E-01	1.80E-01	1.71E-01	**1.49E-01**	3.03E-01	2.98E-01

Table 12 Average MAPE values of surrogate models on the I-beam design problem with different numbers of training points

Number of DOE	PEAH	PBW	OWPE	SP1	EP	PEM-vCV	OWSd	RBF	Kriging	PRS
5n	1.14E+00	3.38E+00	1.34E+00	1.44E+00	2.16E+00	1.51E+00	2.60E+00	**8.93E-01**	3.25E+00	1.18E+01
10n	5.68E-01	2.43E+00	8.70E-01	8.95E-01	1.32E+00	9.67E-01	2.03E+00	**5.44E-01**	2.91E+00	6.81E+00
20n	3.77E-01	1.98E+00	5.18E-01	5.91E-01	8.60E-01	6.62E-01	1.52E+00	**3.75E-01**	2.85E+00	5.97E+00
30n	**3.31E-01**	1.78E+00	4.22E-01	4.81E-01	7.92E-01	5.51E-01	1.27E+00	**3.31E-01**	2.99E+00	5.62E+00

Table 13 Computational time (s) of ensembles of surrogates on the test functions with different numbers of training points

Ensemble	5n	10n	20n	30n
PEAH	4.565E+03	7.295E+03	2.661E+04	5.493E+04
PBW	**4.121E+03**	**6.839E+03**	**2.613E+04**	**5.446E+04**
OWPE	5.390E+03	8.981E+03	3.129E+04	6.337E+04
SP1	4.306E+03	7.037E+03	2.636E+04	5.469E+04
EP	4.382E+03	7.129E+03	2.656E+04	5.487E+04
PEM-vCV	4.276E+03	7.002E+03	2.632E+04	5.466E+04
OWSd	4.126E+03	6.844E+03	2.614E+04	5.447E+04

problems in the construction process, these three ensembles of surrogates are more computationally expensive than other ensemble models. As listed in Table 13, OWPE has the largest amount of calculations in all the cases with different numbers of training points. From Fig. 11 and Table 13 we can observe that as the training point number increases, the computational cost of PEAH is similar to EP.

5 Conclusion

Inspired by the heuristic formulation for the average ensemble model, a pointwise ensemble of surrogates with adaptive function and heuristic formulation (PEAH) is proposed in this paper. The adaptive function proposed in this paper is the key to PEAH; it includes the local accuracy and uncertainty prediction information of a prediction point. Therefore, the adaptive function can adapt to local characteristics of the prediction point. The effectiveness of PEAH is validated by twenty analytical test functions, a shape design problem that requires high-fidelity CFD simulations, and an I-beam design problem with the explicit function. We choose three typical stand-alone surrogate models and six well-known ensembles of surrogates for comparison. In order to intuitively compare the merits and demerits of different ensembles of surrogates, a comparison table (see Table 14) is made based on the simulation results in this work. The test problems used in this paper are divided into three categories, including low-dimensional analytical test functions (From F1 to F11), high-dimensional analytical test functions (From F12 to F20) and engineering design problems (The CFD-based shape design problem and the I-beam design problem).

From the numerical results, the following findings can be drawn:

1. In the case with the small number of training points (5n), compared to other ensembles of surrogates, PEAH and OWPE show superior prediction accuracy on low-dimensional test functions, and PEAH has superiority in the prediction accuracy on high-dimensional test functions. As for stand-alone surrogate models, RBF provides the best performance on most of the analytical test functions, and the compared ensembles couldn't outperform the RBF model. The reason is that for a given problem, the performance of some stand-alone surrogate models may be poor and differs significantly from the best model. This means that compared to the best surrogate model, other single surrogate models may show poor prediction accuracy in the whole design space. In these cases, although the single surrogate model that performs the best has the largest weight in the ensemble, the performance of the constructed ensembles of surrogates will also be affected by other surrogate models that perform poorly. However, ensembles of surrogates can effectively filter out the stand-alone surrogate models that show poor prediction results.
2. With the increase of the number of training points (from 5n to 30n), the Wilcoxon rank-sum test results on the test functions suggest that the robustness of PEAH is better than the three stand-alone surrogate models. This means

Table 14 Comparative analysis of various ensembles of surrogates on different types of test problems

Surrogates	Low-dimensional functions	High-dimensional functions	Engineering design problems
PEAH PBW OWPE SP1 EP PEM-vCV OWSd	PEAH and OWPE show superior prediction accuracy on low-dimensional analytical test functions, and EP's prediction accuracy on low-dimensional test functions increases with the number of training points. Since PEAH, OWPE and EP include optimization problems in the construction process, these three models are more likely to obtain the optimal weights for prediction. Compared to the ensembles of surrogates that include the optimization process, PBW, OWSd, PEM-vCV and SP1 are less computationally expensive.	Compared to other six ensembles of surrogates, PEAH has high prediction accuracy on high-dimensional analytical test functions with a small number of training points. This is because when the number of training points is small, the uncertainty in the design space is large, and the adaptive function in PEAH can adapt to the local characteristics in the design space. And the prediction accuracy of EP on these test functions increases with the number of training points. The prediction accuracy of PBW and SP1 in the high-dimensional test problems is not high. Compared to PEM-vCV and OWSd, PEAH has superiority in the robustness of prediction accuracy.	OWPE and PEAH provide better prediction accuracy on the CFD-based shape design problem with the implicit function than the other five ensembles of surrogates. Compared to the other six ensembles of surrogates, PEAH has superiority in the prediction accuracy on the I-beam design problem with the explicit function, which validates the effectiveness of the adaptive function in the PEAH model.

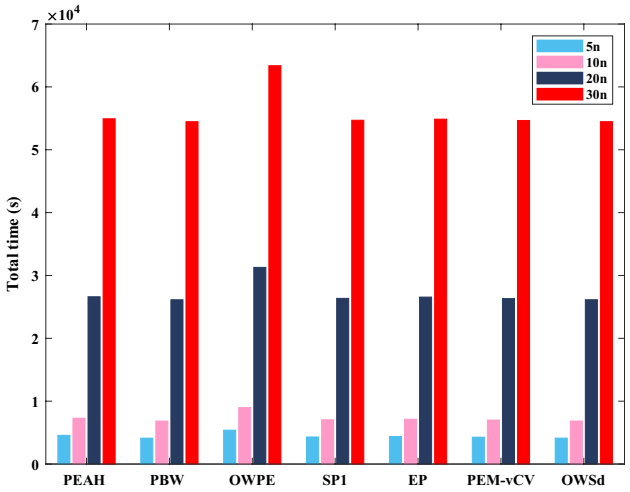

Fig. 11 Total time (s) consumed by the ensembles of surrogates on the test functions with different numbers of training points

that the proposed PEAH model is able to provide stable prediction on different types of test problems.

3. Among the three stand-alone surrogate models, the surrogate model with the highest prediction accuracy in the test functions varies with the increase of the number of training points. In the case with $5n$ training points, RBF performs better than Kriging and PRS. In contrast, the prediction accuracy of PRS improved significantly with the increase of the number of training points, and PRS exhibits better prediction accuracy than RBF and Kriging in the cases with $10n$, $20n$ and $30n$ training points. With the increase of the number of training points, the proposed PEAH model can also maintain stable high-precision predictions on the same test problem. In addition, compared with stand-alone surrogate models, the proposed PEAH model has higher robustness in prediction.
4. The construction processes of PEAH, OWPE and EP include optimization problems, so the computational costs of these three ensemble models are higher than the other four ensembles of surrogates. The computational cost of OWPE is the highest. With the increase of the number of training points, the computational cost of PEAH is similar to EP. In general, to obtain the output response for a practical engineering design problem, we need to conduct a high-fidelity numerical simulation procedure (such as the CFD method) to get the output response. This process is usually time-consuming, which means that the number of training points significantly impacts the amount of calculation in the construction of the surrogate models. However, the construction of ensembles of surrogates is based on the same training set and does not need extra training points. Thus compared with the impact of the number of training points on the computational cost, the computational cost of building an ensemble of surrogates will be smaller.
5. As for the shape design problem, PEAH, OWPE and EP show better prediction accuracy than the other four ensembles of surrogates. In addition, it is worth noting that the prediction accuracy of PEAH on the shape design problem is always better than its stand-alone surrogate models in all the cases. Compared to other ensembles of surrogates, PEAH shows superiority in the prediction accuracy on the I-beam design problem with the explicit function.

The numerical experimental results indicate that PEAH performs better in these problems with a better balance between accuracy and robustness. In the future, to further investigate the performance of PEAH, more stand-alone surrogate models should be employed to construct PEAH. The proposed PEAH is expected to have a wide application in engineering design problems.

Appendix: Explicit functions of of PRS, RBF and Kriging

PRS

PRS model (Box et al. 1978) has been widely applied for estimating the output response in engineering systems. A second-order PRS model can be expressed as:

$$\hat{y} = \beta_0 + \sum_{i=1}^{d} \beta_d x_i + \sum_{i=1}^{d} \beta_{ii} x_i^2 + \sum_{i} \sum_{j} \beta_{ij} x_i x_j \tag{40}$$

where $\hat{y}$ and x represent the predicted response and the input variable vector, respectively. β indicates the coefficient vector of the PRS model and d refers to the dimension of the input variable vector.

RBF

RBF (Hardy 1971) has been developed for approximating the relationship between input variables and the output response. The RBF model utilizes combinations of a radially symmetric function based on Euclidean distance or other such metrics to approximate the response function (Jin et al. 2001). The radial basis function can be expressed as:

$$\hat{y} = \sum_{k=1}^{N} \beta_k \phi(\left\| x - x^{(k)} \right\|) \tag{41}$$

where N refers to the number of training points in the training set. ϕ is the the basis function and β is the coefficient

vector of the basis function. $\left\|x - x^{(k)}\right\|$ represents the Euclidean distance between the prediction point x and the *kth* training point $x^{(k)}$.

Kriging

The Kriging model (Sacks et al. 1989) postulates that the output response at a prediction point can be expressed as a linear combination of the output responses of the training points in the neighborhood of the prediction point. The Kriging model can be expressed as:

$$\hat{y} = \sum_{k=1}^{N} \beta_k f_k(x) + Z(x) \tag{42}$$

where $f_k(x)$ is the output response at the *kth* training point. $Z(x)$ is assumed to be the realization of a stochastic process with mean zero and spatial correlation function (Jin et al. 2001), given by:

$$\text{cov}[Z(x_i), Z(x_j)] = \delta^2 R(x_i, x_j) \tag{43}$$

where δ^2 indicates the process variance, and $R(x_i, x_j)$ refers to the correlation function.

Acknowledgements This research was supported by Guangdong Key R&D Program of 2021 Ocean Six Industrial project No.2021-45, the Construction of a Leading Innovation Team project by the Hangzhou Municipal government, the Startup funding of New-joined PI of Westlake University with Grant Number (041030150118), the Priority Postdoctoral Projects in Zhejiang Province, China (Grant No. ZJ2021046)

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Replication of results The authors state that they have the willingness to share the source codes of the proposed PEAH model and input data used to replicate figures and tables in this paper. The source code of the PEAH model (developed in MATLAB) can be obtained by contacting the corresponding author via email.

References

Acar E (2010) Various approaches for constructing an ensemble of metamodels using local measures. Struct Multidisc Optim 42(6):879–896

Acar E, Rais-Rohani M (2009) Ensemble of metamodels with optimized weight factors. Struct Multidisc Optim 37(3):279–294

Bishop CM et al (1995) Neural networks for pattern recognition. Oxford University Press, Oxford

Box GE, Hunter WH, Hunter S et al (1978) Statistics for experimenters: an introduction to design, data analysis and model building, vol 664. Wiley, New York

Carpenter W, Barthelemy JF (1993) A comparison of polynomial approximations and artificial neural nets as response surfaces. Struct Optim 5(3):166–174

Chen H, Xu Y, Wang M, Zhao X (2019) A balanced whale optimization algorithm for constrained engineering design problems. Appl Math Model 71:45–59

Chen H, Li W, Cui W, Yang P, Chen L (2021) Multi-objective multidisciplinary design optimization of a robotic fish system. J Mar Sci Eng 9(5):478

Chen L, Qiu H, Jiang C, Cai X, Gao L (2018) Ensemble of surrogates with hybrid method using global and local measures for engineering design. Struct Multidisc Optim 57(4):1711–1729

Clarke SM, Griebsch JH, Simpson TW (2004) Analysis of support vector regression for approximation of complex engineering analyses. J Mech Des 127(6):1077–1087

Cristianini N, Shawe-Taylor J et al (2000) An introduction to support vector machines and other kernel-based learning methods. Cambridge University Press, Cambridge

Forrester AI, Keane AJ (2009) Recent advances in surrogate-based optimization. Prog Aerospace Sci 45(1–3):50–79

Friedman JH (1991) Multivariate adaptive regression splines. Ann Stat 19(1):1–67

Giunta A, Watson L (1998) A comparison of approximation modeling techniques-polynomial versus interpolating models. In: 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, p 4758

Glaz B, Goel T, Liu L, Friedmann PP, Haftka RT (2009) Multiple-surrogate approach to helicopter rotor blade vibration reduction. AIAA J 47(1):271–282

Goel T, Haftka RT, Shyy W, Queipo NV (2007) Ensemble of surrogates. Struct Multidisc Optim 33(3):199–216

Hardy RL (1971) Multiquadric equations of topography and other irregular surfaces. J Geophys Res 76(8):1905–1915

Hassoun MH et al (1995) Fundamentals of artificial neural networks. MIT Press, Cambridge

Jin R, Chen W, Simpson TW (2001) Comparative studies of metamodelling techniques under multiple modelling criteria. Struct Multidisc Optim 23(1):1–13

Krishnamurthy T (2005) Comparison of response surface construction methods for derivative estimation using moving least squares, kriging and radial basis functions. In: 46th AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference, p 1821

Lee Y, Choi DH (2014) Pointwise ensemble of meta-models using v nearest points cross-validation. Struct Multidisc Optim 50(3):383–394

Liu H, Xu S, Wang X, Meng J, Yang S (2016) Optimal weighted pointwise ensemble of radial basis functions with different basis functions. AIAA J 54(10):3117–3133

Liu Y, Chen W, Hu J, Zheng X, Shi Y (2017) Ensemble of surrogates with an evolutionary multi-agent system. In: 2017 IEEE 21st International Conference on Computer Supported Cooperative Work in Design (CSCWD), IEEE, pp 521–525

MacKay DJ (1998) Introduction to gaussian processes. NATO ASI series F computer and systems sciences 168:133–166

Sacks J, Welch WJ, Mitchell TJ, Wynn HP (1989) Design and analysis of computer experiments. Stat Sci 4(4):409–423

Sanchez E, Pintos S, Queipo NV (2008) Toward an optimal ensemble of kernel-based approximations with engineering applications. Struct Multidisc Optim 36(3):247–261

Simpson T, Mistree F, Korte J, Mauery T (1998) Comparison of response surface and kriging models for multidisciplinary design optimization. In: 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, p 4755

Song X, Lv L, Li J, Sun W, Zhang J (2018) An advanced and robust ensemble surrogate model: extended adaptive hybrid functions. J Mech Des 140(4):1402

Surjanovic S, Bingham D (2013) Virtual library of simulation experiments: Test functions and datasets. http://www.sfu.ca/~ssurjano. Accessed 3 June 2021

Viana FA, Haftka RT, Steffen V (2009) Multiple surrogates: how cross-validation errors can help us to obtain the best predictor. Struct Multidisc Optim 39(4):439–457

Yang RJ, Wang N, Tho CH, Bobineau JP, Wang BP (2005) Metamodeling development for vehicle frontal impact simulation. J Mech Des 127(5):1014–1020

Ye Y, Wang Z, Zhang X (2020) An optimal pointwise weighted ensemble of surrogates based on minimization of local mean square error. Struct Multidisc Optim 62(2):529–542

Yin H, Fang H, Wen G, Gutowski M, Xiao Y (2018) On the ensemble of metamodels with multiple regional optimized weight factors. Struct Multidisc Optim 58(1):245–263

Zhang J, Chowdhury S, Messac A (2012) An adaptive hybrid surrogate model. Struct Multidisc Optim 46(2):223–238

Zhang J, Yue X, Qiu J, Zhang M, Wang X (2021) A unified ensemble of surrogates with global and local measures for global metamodelling. Eng Optim 53(3):474–495

Zhou XJ, Ma YZ, Li XF (2011) Ensemble of surrogates with recursive arithmetic average. Struct Multidisc Optim 44(5):651–671