A unified ensemble of surrogates with global and local measures for global metamodeling (Supplemental Material)

Jian Zhang[a]*, Xinxin Yue[a], Jiajia Qiu[a]*, Muyu Zhang[a]* and Xiaomei Wang[b]

[a]Department of Mechanics and Engineering Science, Jiangsu University, Zhenjiang 212013, PR China; [b]Department of Ocean Engineering, Tianjin University, Tianjin 300072, PR China

A: Description of the test problems

In this appendix, a suite of 38 test problems is provided in Table S1.

B: Results for selecting the coefficient function $\boldsymbol{\lambda}(\boldsymbol{x})$

The coefficient function described in Section 3.2 of the main text can be any function that satisfy Equation (11). Herein, nine candidate functions as shown in Figure S1 were chosen:

$$
\begin{aligned}
&\textit{Quarter}:\ \lambda(x) = \left(\frac{d_1(x)}{d_2(x)}\right)^{1/4} \\
&\textit{One-half}:\ \lambda(x) = \left(\frac{d_1(x)}{d_2(x)}\right)^{1/2} \\
&\textit{Linear}:\ \lambda(x) = \frac{d_1(x)}{d_2(x)} \\
&\textit{Cubic}:\ \lambda(x) = 3\left(\frac{d_1(x)}{d_2(x)}\right)^{2} - 2\left(\frac{d_1(x)}{d_2(x)}\right)^{3} \\
&\textit{Quartic}:\ \lambda(x) = \left(\frac{d_1(x)}{d_2(x)}\right)^{4} \\
&\textit{Sine}:\ \lambda(x) = sin\left\{\frac{\pi}{2}\left(\frac{d_1(x)}{d_2(x)}\right)\right\} \\
&\textit{Exp1}:\ \lambda(x) = exp\left\{\left(\frac{d_1(x)}{d_2(x)}\right)^{2} \cdot ln2\right\} - 1 \\
&\textit{Exp2}:\ \lambda(x) = 1 - exp\left\{-15 \cdot \left(\frac{d_1(x)}{d_2(x)}\right)^{2}\right\} \\
&\textit{Exp3}:\ \lambda(x) = 1 - exp\left\{-ln1000 \cdot \left(\frac{d_1(x)}{d_2(x)}\right)^{8}\right\}
\end{aligned}
\tag{1}
$$

The proposed ensemble of surrogates was generated using each candidate function, and the mean values of R^2 were obtained in Table S2 for UES1 and UES2 by using one thousand test points for all test problems as in Table S1. It can be concluded that the *sine* function has the best accuracy for both UES1 and UES2 as compared to other candidate functions. Therefore, it is selected as the coefficient function of $\lambda(x)$ for the proposed UES models.

C: Description of selected surrogate models

This appendix gives a brief overview of the mathematical formulations of PRS, RBF and KRG surrogate models.

C.1 Polynomial response surface (PRS)

PRS model is one of the popular surrogate models, of which the second-order polynomial function is most commonly used:

$$\hat{y}(x) = \beta_0 + \sum_{i=1}^{d} \beta_i x_i + \sum_{i=1}^{d} \beta_{ii} x_i^2 + \sum_{i=1}^{d-1} \sum_{j=i+1}^{d} \beta_{ij} x_i x_j \tag{2}$$

where d is the number of input variables, β_0, β_i, β_{ii} and β_{ij} represent the unknown coefficients to be determined by the method of least squares.

C.2 Radial basis function (RBF)

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

$$\hat{y}(x) = \sum_{i=1}^{n} \omega_i \varphi(\|x - x_i\|) \tag{3}$$

where n is the number of sample points, ω_i is the unknown interpolation coefficient, $\|\cdot\|$ denotes the Euclidean norm, $\varphi(\cdot)$ is a radially symmetric basis function. Some of the frequently used RBF models include:

- Multiquadric $\varphi(r) = (r^2 + c^2)^{1/2}$
- Inverse multiquadric $\varphi(r) = (r^2 + c^2)^{-1/2}$
- Gaussian $\varphi(r) = exp(-cr^2)$
- Thin-plate spline $\varphi(r) = r^2 ln(r)$

where $c \geq 0$. For its prediction accuracy and high convergence rate with increasing sample points, the multiquadric formulation of RBF is selected with $c = 1$ (Acar 2010).

C.3 Kriging (KRG)

The basic assumption of KRG is that the response is estimated in the form

$$Y(x) = \mu(x) + Z(x) \tag{4}$$

where $\mu(x)$ is a known polynomial that globally approximates the response, and $Z(x)$ is a stochastic component that generates deviations such that the Kriging model interpolates the sample points.

The stochastic component $Z(x)$ has a mean value of zero and covariance of

$$Cov\left[Z(x^i), Z(x^j)\right] = \sigma^2 \mathbf{R}\left[R(x^i, x^j)\right] \tag{5}$$

where σ^2 is the process variance, $\mathbf{R}$ is $n \times n$ correlation matrix if $\boldsymbol{n}$ is the number of sample points, $R(x^i, x^j)$ is the correlation function between two sample points x^i and x^j. The correlation function is commonly assumed to be stationary Gaussian, that is,

$$R(x^i, x^j) = exp\left\{-\sum_{k=1}^{d} \theta_k \left(x_k^i - x_k^j\right)^2\right\} \tag{6}$$

where d is the number of variables, θ_k is the unknown parameter to be determined by using optimization algorithm.

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

$$\hat{y}(x) = \hat{\mu} + \mathbf{r}^T(x)\mathbf{R}^{-1}(\boldsymbol{y} - \hat{\mu}\mathbf{1}) \tag{7}$$

where $\mathbf{r}^T(x)$ is the correlation vector of length n between a prediction point x and the n sample points, $\boldsymbol{y}$ represents the vector of sample responses, $\mathbf{1}$ is an $n \times 1$ vector of ones. The scalar $\hat{\mu}$ and vector $\mathbf{r}$ are given by

$$\hat{\mu} = (\mathbf{1}^T \mathbf{R}^{-1} \mathbf{1})^{-1} \mathbf{1}^T \mathbf{R}^{-1} \boldsymbol{y}$$
$$\mathbf{r}^T(x) = [R(x, x^1), R(x, x^2), \cdots, R(x, x^n)]^T \quad (8)$$

In this study, the MATLAB® Kriging toolbox is used, and the lower and upper bounds of $\boldsymbol{\theta_k}$ used for all the investigated examples are 0.01 and 20 respectively.

D: Boxplots for two test problems (Test function 7 & 35 in Table S1)

Table S1. A suite of test functions.

No.	Design variable (D)	Test function	Domain
1	2	$f(x) = \left(x_2 - \frac{5.1x_1^2}{4\pi^2} + \frac{5x_1}{\pi} - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)cos(x_1) + 10$	[-5,10; 0,15]
2	2	$f(x) = (x_1 - 1)^2 + 2(2x_2^2 - x_1)^2$	[-10, 10]^D
3	2	$f(x) = [1 + (x_1 + x_2 + 1)^2(19 - 14x_1 + 3x_1^2 - 14x_2 + 6x_1x_2 + 3x_2^2)]$ $\times [30 + (2x_1 - 3x_2)^2(18 - 32x_1 + 12x_1^2 + 48x_2 - 6x_1x_2 + 27x_2^2)]$	[-2, 2]^D
4	2	$f(x) = x_1^2 + x_2^2 + 200 - (x_1^2 + x_2 - 11)^2 - (x_1 + x_2^2 - 7)^2$	[-4, 4]^D
5	2	$f(x) = 4x_1^2 - 2.1x_1^4 + x_1^6/3 + x_1x_2 - 4x_2^2 + 4x_2^4$	[-5, 5]^D
6	2	$f(x) = cos(10x_1^2) + 3.1\|x_1 - 0.7\| + 2x_1^2$ $+sin[1/(\|x_1 - 0.7)\| + 0.31)] + 2x_2^2$	[0, 1]^D
7	2	$f(x) = (C + 1)/9.32\,,\ A = 6x_1 - 2\,,\ B = 6x_2 - 2\,,$ $C = (A^2 + B^2 + 3)/2 + sin[(A^2 + B^2 + 2)/2]$	[0, 1]^D
8	2	$f(x) = e^{(x_1 - x_2)^2} + e^{(10 - x_1)^2} - x_1x_2$	[0, 10]^D
9	2	$f(x) = \|x_1\| + x_2^2$	[-8, 8]^D
10	2	$f(x) = sin(\pi x_1/12)\, cos(\pi x_2/16)$	[0, 10]^D
11	2	$f(x) = sin(x_1 + x_2) + (x_1 - x_2)^2 - 1.5x_1 + 2.5x_2 + 1$	[0, 10]^D
12	2	$f(x) = cos(\sqrt{x_1^2 + x_2^2})$	[-5, 5]^D
13	2	$f(x) = \sum_{i=1}^{n} \frac{x_i^2}{4000} - \prod_{i=1}^{n} cos(x_i/\sqrt{i}) + 1\,,\ n = 2$	[-20, 20]^D
14	2	$f(x) = 20 + x_1^2 + x_2^2 - 10[cos(2\pi x_1) + cos(2\pi x_2)]$	[-1, 1]^D
15	2	$f(x) = 0.5 + sin\left\{\sqrt{x_1^2 + x_2^2}/[1 + 0.1(x_1^2 + x_2^2)^2]\right\}$	[-2, 2]^D
16	2	$f(x) = cos[6(x_1 - 0.5)] + 3.1\|x_1 - 7\| + 2(x_1 - 0.5)^2$ $+sin[0.1/(\|x_1 - 0.5\| + 0.31)] + 0.5x_2$	[0, 1]^D
17	2	$f(x) = 2 + 0.01(x_2 - x_1^2)^2 + (1 - x_1)^2 + 2(2 - x_2)$ $+7\, sin(0.5x_1)\, sin(0.7x_1x_2)$	[0, 5]^D
18	2	$f(x) = \sin(x_1)\cos(x_2)$	[-3, 3]^D
19	2	$f(x) = (30 + x_1\, sinx_1)\left(4 + e^{-x_2^2}\right)$	[0, 10]^D
20	2	$f(x) = cos\left(\sqrt{(x_1 - 3)^2 + (x_2 - 2)^2}\right)$	[-5, 5]^D
21	2	$f(x) = (x_1 - 0.2)^2 + (x_2 - 11)^2 + [x_1 + (x_2 + 0.3)^2 - 7]^2$	[-4, 4]^D
22	3	$f(x) = -\sum_{i=1}^{4} c_i e^{-\sum_{j=1}^{3} a_{ij}(x_j - p_{ij})^2}$	[0, 1]^D
23	4	$f(x) = 100(x_1^2 - x_2)^2 + a_1^2 + a_3^2 + 90(x_3^2 - x_4)$ $+10.1(a_2^2 + a_4^2) + 19.8a_2a_4,\ \ a_i = x_i - 1$	[-10, 10]^D
24	4	$f(x) = (x_1 - 1)^2 + \sum_{i=2}^{n} i(2x_i^2 - x_{i-1})^2\,, n = 4$	[-10, 10]^D

Table S1. A suite of test functions (continued).

No.	Design variable (D)	Test function	Domain
25	4	$f(x)=\sum_{i=1}^{n}\frac{x_i^2}{4000}-\prod_{i=1}^{n}cos\left(x_i/\sqrt{i}\right)+1,\ n=4$	$[-20, 20]^D$
26	4	$f(x)=10sin[2(x_1-0.6\pi)]+x_2+x_3+x_4+x_1x_2 +x_3x_4+x_1^3+x_4^3$	$[0, 1]^D$
27	4	$f(x)=\sum_{i=1}^{d/4}[(x_{4i-3}+10x_{4i-1})^2+5(x_{4i-1}-x_{4i})^2 +(x_{4i-2}-2x_{4i-1})^4+10(x_{4i-3}+x_{4i})^4]$	$[-4, 5]^D$
28	4	$f(x)=1+\ e^{-2[(x_1-1)^2+x_2^2]-0.5(x_3^2+x_4^2)}+e^{-2[x_1^2+(x_2-1)^2]-0.5(x_3^2+x_4^2)}$	$[0, 1]^D$
29	6	$f(x)=-\sum_{i=1}^{4}c_ie^{-\sum_{j=1}^{6}a_{ij}(x_j-p_{ij})^2}$	$[0, 1]^D$
30	8	$f(x)=4(x_1-2+8x_2-8x_2^2)+(3-4x_2^2)^2 +16\sqrt{x_3+1}(2x_3-1)^2+\sum_{i=4}^{8}i\,ln\left(1+\sum_{j=3}^{i}x_j\right)$	$[0, 1]^D$
31	8	$f(x)=(x_1-1)^2+\sum_{i=2}^{n}i(2x_i^2-x_{i-1})^2,\ n=8$	$[-10, 10]^D$
32	8	$f(x)=\sum_{i=1}^{n}\frac{x_i^2}{4000}-\prod_{i=1}^{n}cos(x_i/\sqrt{i})+1, n=8$	$[-20, 20]^D$
33	10	$f(x)=\sum_{i=1}^{10}[(x_{i+1}^2-x_i)^2+(x_i-1)^2]$	$[-3, 3]^D$
34	10	$f(x)=\sum_{i=1}^{10}{x_i}^2+(\sum_{i=1}^{10}0.5ix_i)^2+(\sum_{i=1}^{10}0.5ix_i)^4$	$[-5, 10]^D$
35	10	$f(x)=\sum_{i=1}^{10}\left\{\frac{3}{10}+sin\left(\frac{16}{15}x_i-1\right)+\left[sin\left(\frac{16}{15}x_i-1\right)\right]^2\right\}$	$[-1, 1]^D$
36	12	$f(x)=(x_1-1)^2+\sum_{i=2}^{n}i(2x_i^2-x_{i-1})^2,\ n=12$	$[-10, 10]^D$
37	15	$f(x)=(x_1-1)^2+\sum_{i=2}^{n}i(2x_i^2-x_{i-1})^2,\ n=15$	$[-10, 10]^D$
38	15	$f(x)=\sum_{i=1}^{14}[100(x_{i+1}-x_i^2)^2+(x_i-1)^2]$	$[-5, 10]^D$

Table S2. Mean of R^2 for the UES model with different candidate coefficient functions.

UES	Quarter	One-half	Linear	Cubic	Quartic	Sine	Exp1	Exp2	Exp3
UES1	0.7314	0.7312	0.7301	0.7299	0.7253	**0.7322**	0.7269	0.7318	0.7253
UES2	0.7339	0.7341	0.7245	0.7339	0.7299	**0.7355**	0.7322	0.7323	0.7306

1. The best results of coefficient functions are marked in bold.

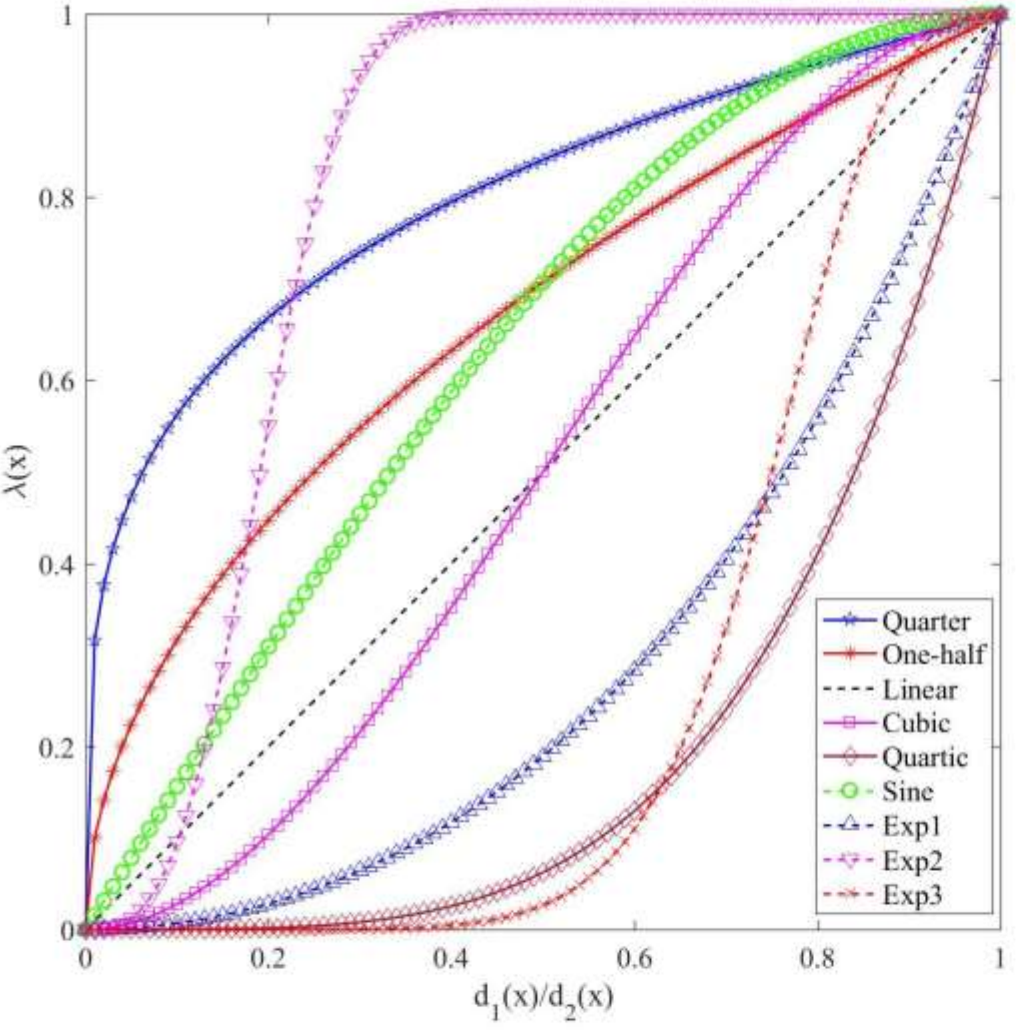


Figure S1. Candidate coefficient function of $\boldsymbol{\lambda}(\boldsymbol{x})$.

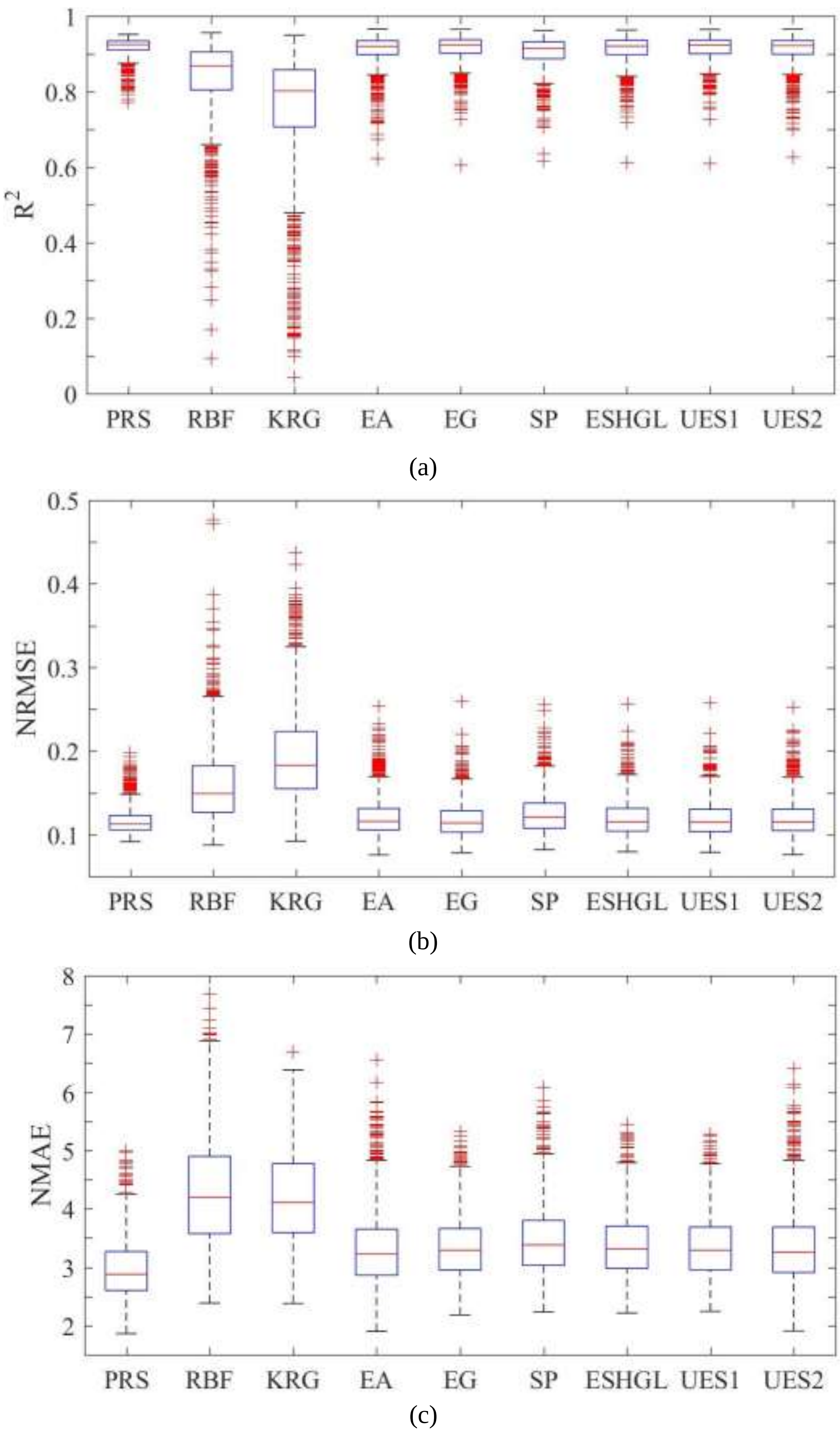
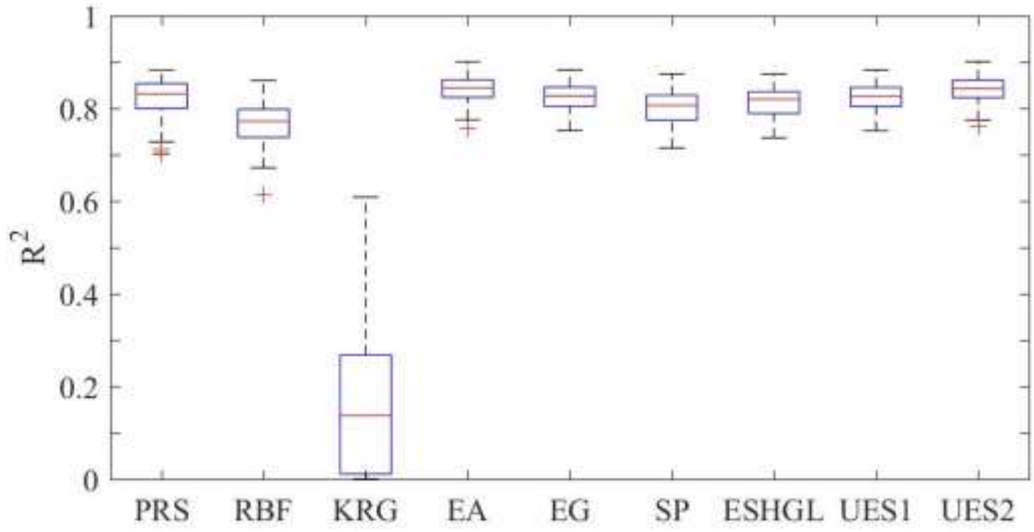
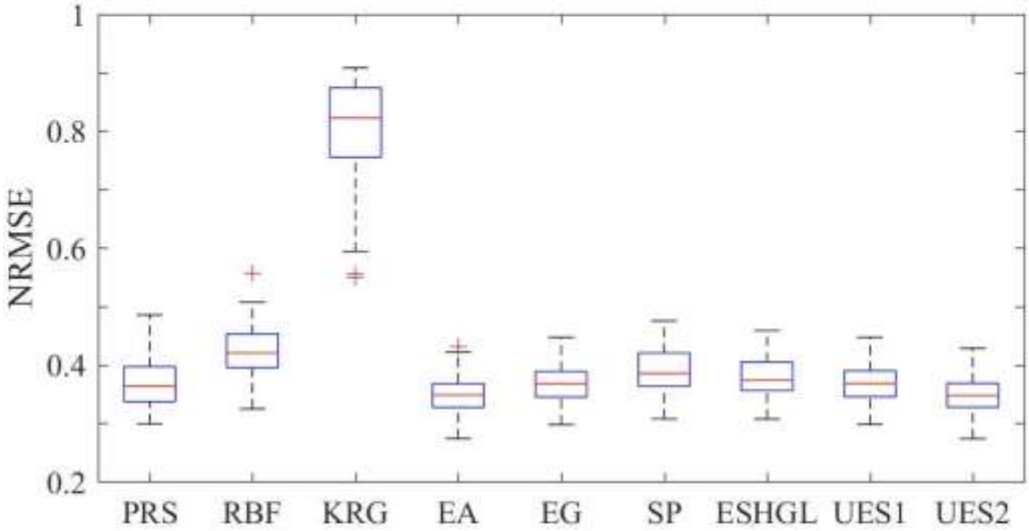


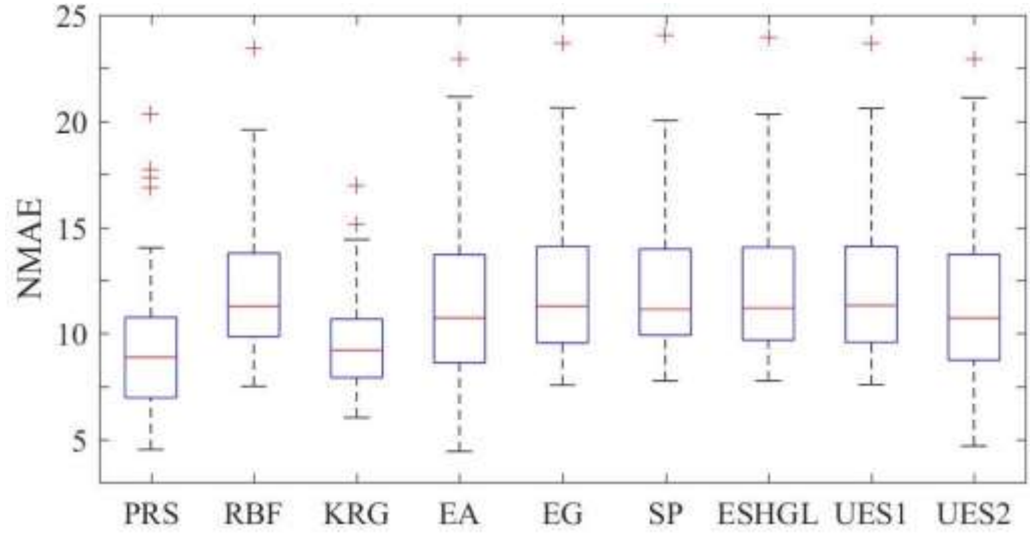
Figure S2. Boxplots for Test function 7: (a) R^2, (b) NRMSE and (c) NMAE.



(a)



(b)



(c)

Figure S3. Boxplots for Test function 35: (a) R^2, (b) NRMSE and (c) NMAE.