Quantifying Regional Error in Surrogates by Modeling its Relationship with Sample Density

Ali Mehmani[*], Souma Chowdhury[†],
Jie Zhang[‡],
Weiyang Tong[*], and Achille Messac[§]
Syracuse University, Syracuse, NY, 13244

Approximation models (or surrogate models) provide an efficient substitute to expensive physical simulations and an efficient solution to the lack of physical models of system behavior. However, it is challenging to quantify the accuracy and reliability of such approximation models in a region of interest or the overall domain without additional system evaluations. Standard error measures, such as the mean squared error, the cross-validation error, and the Akaikes information criterion, provide limited (often inadequate) information regarding the accuracy of the final surrogate. This paper introduces a novel and model independent concept to quantify the level of errors in the function value estimated by the final surrogate in any given region of the design domain. This method is called the Regional Error Estimation of Surrogate (REES). Assuming the full set of available sample points to be fixed, intermediate surrogates are iteratively constructed over a sample set comprising all samples outside the region of interest and heuristic subsets of samples inside the region of interest (i.e., intermediate training points). The intermediate surrogate is tested over the remaining sample points inside the region of interest (i.e., intermediate test points). The fraction of sample points inside region of interest, which are used as intermediate training points, is fixed at each iteration, with the total number of iterations being pre-specified. The estimated median and maximum relative errors within the region of interest for the heuristic subsets at each iteration are used to fit a distribution of the median and maximum error, respectively. The estimated statistical mode of the median and the maximum error, and the absolute maximum error are then represented as functions of the density of intermediate training points, using regression models. The regression models are then used to predict the expected median and maximum regional errors when all the sample points are used as training points. Standard test functions and a wind farm power generation problem are used to illustrate the effectiveness and the utility of such a regional error quantification method.

Keywords: Surrogate model; Error quantification; Model selection; Sampling; Regional accuracy;

[*]Doctoral Student, Multidisciplinary Design and Optimization Laboratory, Department of Mechanical and Aerospace. AIAA Student Member

[†]Research Assistant Professor, Multidisciplinary Design and Optimization Laboratory, Department of Mechanical and Aerospace. AIAA Member

[‡]Postdoctoral Research Associate, Multidisciplinary Design and Optimization Laboratory, Department of Mechanical and Aerospace. AIAA Senior Member. Currently at the National Renewable Energy Laboratory (NREL), Golden, CO, 80401

[§]Distinguished Professor and Department Chair. Department of Mechanical and Aerospace Engineering. AIAA Lifetime Fellow. Corresponding author. Email: messac@syr.edu

I. Introduction

A. Approximation Models

Mathematical approximation models are commonly used for providing a tractable and inexpensive approximation of the actual system behavior in many routine engineering analysis and design activities, e.g., domain exploration, sensitivity analysis, development of empirical models, and optimization. One of the most popular classes of approximation models are surrogate models or metamodels,[1] which are purely mathematical models, i.e., they are not based on the physics of the system. Major surrogate modeling methods include Polynomial Response Surfaces,[2] Kriging,[3,4] Moving Least Square,[5,6] Radial Basis Functions (RBF),[7] Neural Networks,[8] and hybrid surrogate modeling.[9,10] These methods have been applied to a wide range of disciplines, such as aerospace design, automotive design, chemistry, and material science.[11] The four main steps typically involved in constructing a surrogate model are:

1. Choosing an appropriate method for performing the design of experiments (DoE);
2. Evaluating the system behavior from high fidelity simulations or physical experiments;
3. Identifying the appropriate surrogate model and training it using the system evaluations obtained in the previous step; and
4. Validating the performance of the surrogate to ensure reasonable accuracy (prior to actual application).

In the fourth step, specialized error measures can be used to assess the accuracy of the surrogate estimation. The utility of the knowledge of the local and global accuracy of a surrogate goes beyond validation of the surrogate for application. Such knowledge can be crucial (i) for domain exploration, (ii) for further improvement of the surrogate using direct or sequential sampling (adaptive sampling[12] or active learning[13]), (iii) for assessing the reliability (and updating) of the optimal design obtained through surrogate based optimization,[14,15] and (iv) for quantifying the uncertainty (and user confidence) associated with the surrogate. Other possible applications include surrogate model selection, and the construction of weighted surrogate model and conservative surrogate model. These applications generally demand a reliable measure of the *local and regional accuracy* of surrogates. However, the current state of the art in surrogate modeling and application either provides limited information regarding the local and regional accuracy of surrogates or requires additional (expensive) system evaluations, and is often model-dependent. In this paper, we introduce a novel concept towards quantifying the level of surrogate errors in any given region of the design domain, without requiring any additional system evaluations. This concept is also *model independent* and seeks to be universal in application. In the following subsection, the popular methods for assessing the accuracy of surrogates are briefly reviewed.

B. Modeling Errors: Quantification and Impact

Error quantification methods can be broadly classified, based on their computational expense, into: (i) methods that require additional data, and (ii) methods that use existing data.[16] The former can be significantly expensive and might not be a practical option in a majority of applications. Error quantification methods can also be classified into global and local error estimation methods.[17] The performance of the surrogate over the entire domain is evaluated by global error measures, while local or point-wise error measures provide the surrogate accuracy in different locations of the design domain.

Popular approaches of global error measures include:[9] (i) split sample, (ii) cross-validation, (iii) bootstrapping, and (iv) Akaike's information criterion (AIC). It should be noted that these techniques are model independent. In split sample strategy, the sample data is divided into training and test data. The former is used to construct a surrogate; and the latter is used to test the performance of the surrogate. The cross-validation is a popular technique to estimate the error of a developed surrogate. In q-fold cross-validation approach, the data set is split randomly into q (approximately) equal subsets. The surrogate is constructed q times, each time leaving out one of the subsets from training points. The omitted subset, at each iteration, is used to evaluate the cross-validation error.[18] A k-fold cross-validation approach is a variation of q-fold in which all possible subset of size k are used to evaluate cross-validation error. In *leave-one-out cross-validation* approach ($k = 1$), at each iteration, the training set is created by taking all sample points except one, and the left out point is used for estimating the error between the surrogate prediction and the actual value.

It should be noted that the cross-validation approach might provide information about the local accuracy of the intermediate surrogate (which is constructed using subset of training points). The bootstrapping approach generates m sub-samples from the sample points. Each sub-sample is a combination of all samples with replacement. Different variants of the bootstrapping approach can be used for (i) model identification, and (ii) identifying confidence intervals for surrogates.[9] In Akaike's original AIC, the performance of the surrogate is predicted based on a penalized likelihood which is a negative log likelihood plus a penalty term, which is given by

$$AIC = -2\log L(\hat{\theta}) + 2k \tag{1}$$

In this equation, $L(\hat{\theta})$ is the maximized likelihood function, and k is the number of free parameters in the model ,which is a measure of complexity or the compensation for the bias in the lack of fit when the maximum likelihood estimators are used.[19]

Standard local error measures include: (i) the mean squared errors for Kriging,[14] and (ii) the linear reference model (LRM).[20] In stochastic surrogate models like Kriging, the errors at two different points of the design domain are not independent; and the correlation between the points is related to the distance between them. When the distance between the two points is small, the correlation is close to one, and when the distance is large, the correlation tends to zero. According to this correlation strategy, if the point x^* is close to sample points, the prediction confidence at that point is higher than when it is far away from all the sample points. This concept is reflected in the local error measurement method for Kriging predictor at the special point x^*. This error is equal to zero at the sample points and is equal to $\sqrt{\sigma^2}$ at a point far away from sample points, where σ^2 is the approximation error variance in the stochastic process. The LRM is a model independent method for quantifying the local performance of a surrogate. The LRM considers the region with oscillations (complex behavior) as a high-error location. This method categorizes errors of a surrogate in the design domain based on the deviation of the surrogate from the local linear interpolation.[20]

The mean squared error (MSE)(or root mean square error (RMSE)) which provides a global error measure over the entire design domain. The RMSE at a large number of test points (N_{Test}) is defined by:

$$RMSE = \sqrt{\frac{1}{N_{Test}} \sum_{i=1}^{N_{Test}} (y_i - \hat{y}_i)^2} \tag{2}$$

where y_i and $\hat{y}_i$ are the actual and predicted values on i^{th} test point, respectively. The RMSE provides information about the actual accuracy of the surrogate, which requires additional system evaluations on test points in the case of interpolating surrogates. The maximum absolute error (MAE) and relative absolute error (RAE) are indicative of local deviations:

$$MAE = \max_{i=1,\ldots,N_{Test}} |y_i - \hat{y}_i| \tag{3}$$

$$RAE_i = |\frac{y_i - \hat{y}_i}{y_i}| \tag{4}$$

The prediction sum of square (PRESS) is based on the *leave-one-out cross-validation* error:

$$PRESS = \frac{1}{N} \sum_{i=1}^{N} (\hat{y}_i - \hat{y}_i^{(-i)}) \tag{5}$$

where $\hat{y}_i$ and $\hat{y}_i^{(-i)}$ are the surrogate estimations at the i^{th} training point, respectively predicted by the surrogate constructed using all sample points and the surrogate constructed using all sample points except the i^{th} point.[9]

Meckesheimer et al.[16] used the root mean square error of k-fold cross-validation ($\mathrm{RMSE_{CV}}$) or root mean square of PRESS ($\mathrm{PRESS_{RMS}}$) to measure the global accuracy of the surrogate over the entire design domain:

$$RMSE_{CV} = \sqrt{\frac{1}{k} \sum_{i=1}^{k} (\hat{y}_i - \hat{y}_i^{(-i)})^2} \tag{6}$$

where k is the number of omitted points. They studied the variations of k between 1 to 10; and for each k values, the average of error measured on all of the combinations of selecting k points from all sample points

are used to measure the global accuracy of surrogate. They also compared the error estimated with the actual error approximated on additional test points to show the practicality of the $\mathrm{RMSE_{CV}}$ as a fidelity characterizing method without using additional system evaluations.

Viana et al.[21] applied the $\mathrm{RMSE_{CV}}$ approach as an criterion in surrogate model selection, and constructing a weighted average surrogate. They also showed that better results can be achieved using the *leave-one-out* approach. Zhang et al.[22] applied relative absolute error of cross-validation ($\mathrm{RAE_{CV}}$) for characterizing the uncertainty in the prediction of surrogate models. In this study, the normalized $\mathrm{RAE_{CV}}$ is defined based on the *leave-one-out* approach as given by:

$$RAE_{CV_i} = |\frac{\hat{y}_i - \hat{y}_i^{(-i)}}{\hat{y}_i}| \tag{7}$$

II. Regional Error Estimation of Surrogates (REES)

A. Regional Error Estimation of Surrogates

The objective of the proposed approach is to provide a measure of the level of regional errors in an estimated function without investing additional (expensive) system evaluations; this method is called the Regional Error Estimation of Surrogate (REES). The REES method predicts the error level in any given region of interest (based on the necessity of the user) by modeling the variation of the error with increasing density of training points inside the region of interest. The size of the intermediate training data set is increased in each step while the remaining sample points are used as test points to evaluate the error of the intermediate surrogate. The overall flowchart of the REES method is illustrated in Fig.1, and the major steps are described below:

A.1 *Generation of sample data*
In this step, a set of experimental designs are selected based on a specified distribution. Then, the system is evaluated over the selected test data points. The entire set of sample points is represented by $\{X\}$.

A.2 *Identification of sample points inside/outside region of interest*
The entire set of sample points are divided into inside and outside sets based on the user-defined region of interest boundaries, and represented by $\{X_{in}\}$ and $\{X_{out}\}$, respectively.

A.3 *Estimation of the variation of the error with sample density*
This step consists of an iterative process. The number of iteration (N^{it}) is defined based on the dimension of a problem, number of inside sample points, and the preference of the user. At each iteration, sample points are divided into an intermediate training data set $\{\mathrm{X_{TR}}\}$ and an intermediate test data set $\{\mathrm{X_{TE}}\}$

$$\begin{aligned} &\{X_{TR}\} = \{X_{out}\} + \{ß^k\} \\ &\{X_{TE}\} = \{X\} - \{X_{TR}\} \\ &\text{where, } \{ß_{\mathrm{i}}^{\mathrm{t}}\} \subset \{\mathrm{X_{in}}\} \\ &t = 1, 2, 3, ..., N^{it} \end{aligned} \tag{8}$$

where $\{ß^k\}$ represents a k^{th} subset of inside-region sample points. At each iteration t, the size of $\{ß^k\}$ is defined by n^t and $n^t = s(t)$. The function $s(t)$ returns an element of an ascending series, where $s(t) \leq s(t+1),\ \forall\ t$, and $s(t=1) \geq 1$.

At each iteration, the total number of sample combinations is defined by K^t where $K^t \leq \binom{\#\{X_{in}\}}{n^t}$. The term $\#\{X_{in}\}$ represents the number of inside-region sample points. In low dimensional problems, all possible subsets of size n^t are used, while in high dimensional problems a fraction of subsets are used due to the computational cost. The intermediate surrogates $f^k,\ k = 1, 2, ..., K^t$ are constructed for all combinations using the intermediate training points, and are tested over the intermediate test points. The median and the maximum errors are then estimated for each combination as follows

$$E_{med}^k = Median(e_1, e_2, ..., e_{m^t}) \tag{9}$$
$$E_{max}^k = Max(e_1, e_2, ..., e_{m^t}) \tag{10}$$

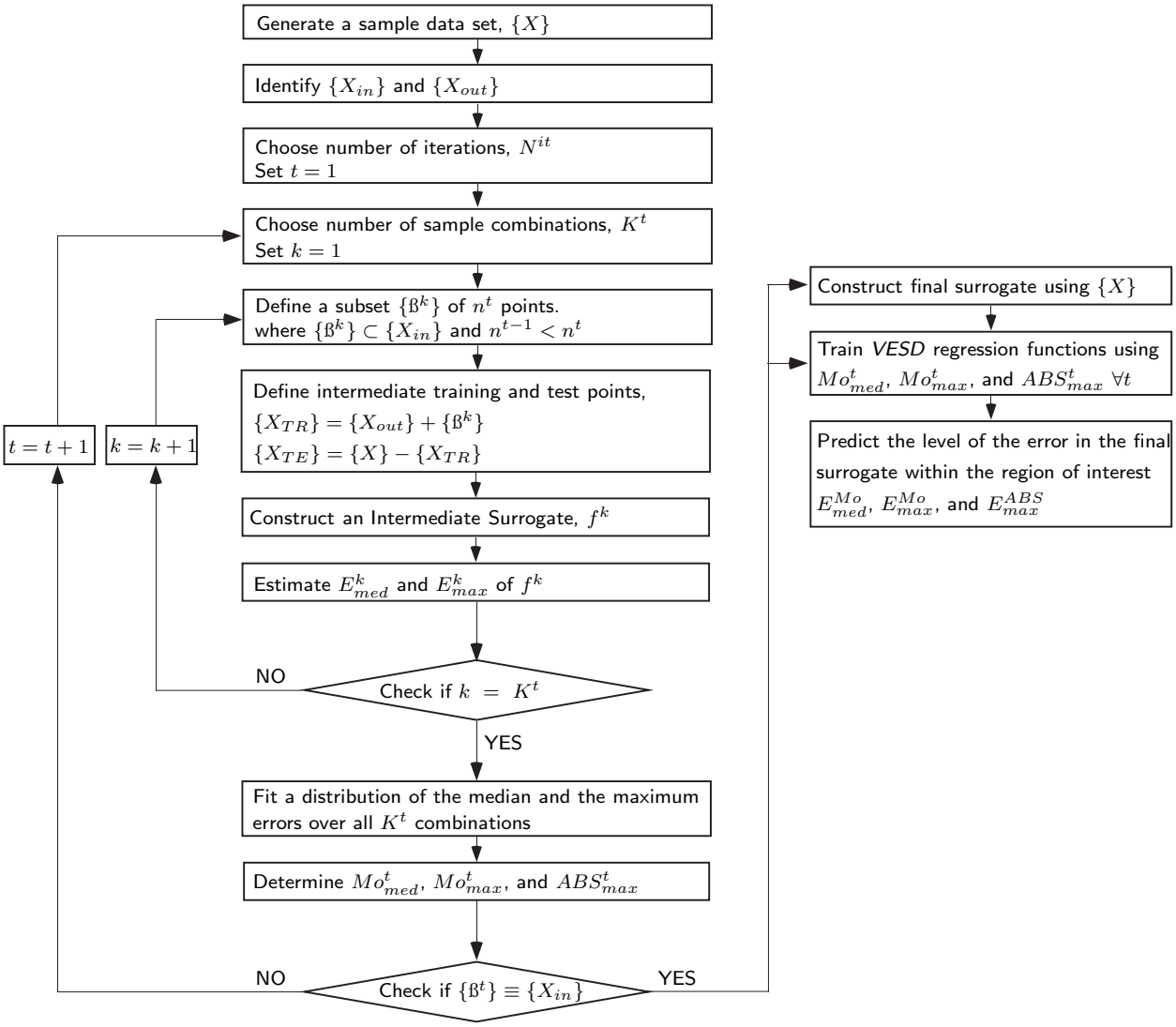


Figure 1. The framework for the Regional Error Estimation of Surrogate (REES) method

where m^t represents the number of test points in t^{th} iteration; and e represents the RAE value estimated on intermediate test points. Probabilistic models are developed using a lognormal distribution to represent median and maximum errors estimated over all K^T combinations at each iterations. It should be noted that outliers are removed from median errors estimated at each iteration. The lognormal distribution is selected because there is often orders of magnitude variation in the median error. It should be noted that the lognormal distribution is more suitable where the data range is high, while in a narrow data range it approximates to a normal distribution. The density function of a lognormal distribution is given by

$$f(x;\mu,\sigma) = \frac{1}{x\sqrt{2\pi\sigma^2}} e^{-\frac{(\ln x-\mu)^2}{2\sigma^2}} \tag{11}$$

where the σ^2 and μ represents the shape and log-scale parameters of the distribution, respectively. The mode of the median and maximum error distribution at each iteration ($\mathrm{Mo}^{\mathrm{t}}_{\mathrm{med}}$ and $\mathrm{Mo}^{\mathrm{t}}_{\mathrm{max}}$) are evaluated as a center of tendency, and are used to relate the variation of the surrogate error with sample density.

To illustrate the relation of the error with sample density, a two design variable benchmark problem (Branin-Hoo function) is used. The size of sample data set, $\#\{X\}$, is defined to be 550, and all of them are selected as inside sample points (the region of interest is entire design space). The numerical settings of this problem are defined as

$$\begin{aligned} &\#\{X_{in}\} = \#\{X\},\ n^t = 5t,\ t = 1,2,3,...,100, \\ &\textit{and}\ K^t = 400,\ \forall t \end{aligned} \tag{12}$$

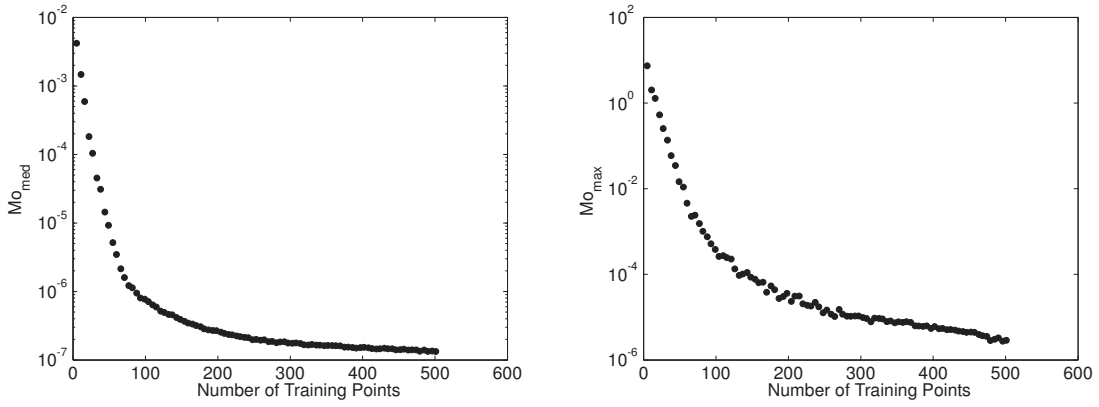
Based on this definition at the first iteration the intermediate surrogates are constructed using 5 sample points, and tested on 545 points;

$$\#\{X_{TR}\} = 5,\ \textit{and}\ \#\{X_{TE}\} = 545 \tag{13}$$

In the same way at the last iteration;

$$\#\{X_{TR}\} = 500,\ \textit{and}\ \#\{X_{TE}\} = 50 \tag{14}$$

The relation of $\mathrm{Mo}^{\mathrm{t}}_{\mathrm{med}}$ and $\mathrm{Mo}^{\mathrm{t}}_{\mathrm{max}}$ with increasing sample density are illustrated in Figs. 2(a) and 2(b), respectively. These figures illustrate that the errors decrease in a practically monotonic relation with increasing training points, thereby illustrating the effectiveness of the proposed error measure.



(a) Estimated mode of median error

(b) Estimated mode of maximum error

Figure 2. Relation of Error Estimated with Sample Density (Branin-Hoo Function with 2 design variables)

A.4 Prediction of regional error in the final surrogate

The final surrogate model is constructed using the full set of training data (sample data). Regression models are applied to relate the Statistical Mode of the median error distribution($\mathrm{Mo}_{\mathrm{med}}$), the statistical mode of the maximum error distributions($\mathrm{Mo}_{\mathrm{max}}$), and the absolute maximum error ($\mathrm{ABS}^{\mathrm{t}}_{\mathrm{max}}$) at each iteration to

the size of the inside-region training points, n^t. These regression models, called the *variation of error with sample density* (VESD), are expressed as

$$E_{med}^{Mo} = F_{med}\left(n^t\right) \tag{15}$$

$$E_{max}^{Mo} = F_{max}\left(n^t\right) \tag{16}$$

$$E_{max}^{ABS} = F_{ABS_{max}}\left(n^t\right) \tag{17}$$

The regression models are then used to predict the level of the error in the final surrogate within the region of interest.

B. Modeling the Variation of Regional Error with Training Point Density

The selection of the type of regression function is critical to the modeling of the variation of regional error with training point density. In this study, three types of the regression functions are used to represent the variation of regional error with respect to the inside-region training points and they are

Type 1 Exponential regression model

$$F(x) = a_0\ e^{a_1 x} \tag{18}$$

Type 2 Multiplicative regression model

$$F(x) = a_0\ x^{a_1} \tag{19}$$

Type 3 Linear regression model

$$F(x) = a_0 + a_1\ x \tag{20}$$

where a_0 and a_1 are unknown coefficients to be determined. The choice of these functions assume *a smooth monotonic decrease of the regional error with the training point density within that region.* In this paper, the root mean squared error metric is used to select the best-fit regression model.

III. Application of the REES Method

The effectiveness of the REES method is explored for applications with Kriging, Radial Basis Functions (RBF), Extended Radial Basis Functions (E-RBF), and Quadratic Response Surface (QRS). To evaluate practical and numerical efficiencies of the REES method, four benchmark problems and an engineering design problem are tested. The error evaluated using REES is compared with the actual error evaluated using relative absolute error on additional test points ($\mathrm{RAE}_{\mathrm{actual}}$). At the same time, the relative absolute error given by *leave-one-out cross-validation* ($\mathrm{RAE}_{\mathrm{CV}}$) is also compared with $\mathrm{RAE}_{\mathrm{actual}}$ to illustrate the potential greater effectiveness of the REES error metric over cross-validation error metrics.

A. Benchmark Problems

The performance of the proposed REES method is evaluated using the following analytical test problems:

Branin-Hoo function (2 variables)

$$f(x) = \left(x_2 - \frac{5.1x_1^2}{4\pi^2} + \frac{5x_1}{\pi} - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right) cos(x_1) + 10 \tag{21}$$

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

Hartmann functions (3 and 6 variables)

$$f(x) = -\sum_{i=1}^{4} c_i\ exp\left\{-\sum_{j=1}^{n} A_{ij}\left(x_j - P_{ij}\right)^2\right\} \tag{22}$$

$$where \quad x = (x_1\ \ x_2\ \ \ldots\ \ x_n) \quad x_i \in [0\ \ 1]$$

In Hartmann-3, the number of variables, $n = 3$; the constants c, A, and P, are a 1×4 vector, a 4×3 matrix, and a 4×3 matrix, respectively.[10] In Hartmann-6, $n = 6$; the constants c, A, and P, are a 1×4 vector, a 4×6 matrix, and a 4×6 matrix, respectively.[10]

B. Wind Farm Power Generation

The effectiveness of the REES method for practical problems is illustrated using a wind farm power generation model, adopted from the Unrestricted Wind Farm Layout Optimization (UWFLO) framework.[23,24] Surrogates are developed using Kriging, RBF, E-RBF, and QRS to represent the *power generation of an array-like wind farm*. It is assumed that turbines are arranged in a row-column pattern over the farm site. Hence, the wind farm power generation can be represented as a function of the *streamwise spacing* and *spanwise spacing* between turbines, with respect to the south direction (as shown in Fig. 3).

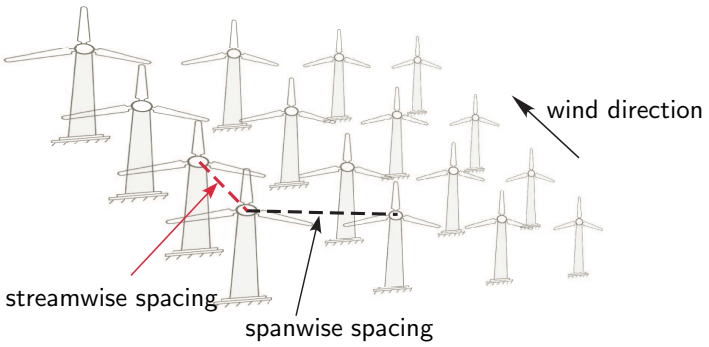


Figure 3. Wind farm array schematic

The annual average power generation of a wind farm is a complex and expensive function of the turbine features, the turbine arrangement (or farm layout), and the local wind resource variations. A surrogate model offers a more tractable (and inexpensive) representation of the farm power generation in terms of key design parameters. To train the surrogate model, the actual annual-average wind farm power generation is estimated using an advanced power generation model developed by Chowdhury et al.[23,24] The reliability of surrogate applications in wind farm design and analysis is subject to the accuracy of the surrogate in the concerned parameter ranges.

The new REES method enables a unique quantification of the fidelity of the surrogate, thereby allowing more informed decision-making in wind farm design and analysis (than possible with conventional surrogate application). This case study is expected to illustrate the effectiveness of the REES method for a complex practical engineering problem. In this case study, the surrogates are constructed to represent the power generation of an array-like *100-turbine wind farm* as a function of the *streamwise spacing* (x_h) and the *spanwise spacing* (x_l) between turbines. The turbines are arranged in a 10×10 patterns in this case. The bivariate normal distributions of wind data obtained for a site in North Dakota[25] is used for this case study. The lower and upper bounds of x_h and x_l, based on the wind turbine rotor diameter (D), are specified as

$$5D < x_h < 30D$$
$$1.1D < x_l < 10D$$

C. Numerical Settings

The numerical settings for the application of REES are provided in Table 1, which lists (i) the number of input variables, (ii) the number of training points, (iii) the number of iterations, and (iv) the number of inside-region training points based on the predefined region's boundary. In these problems, the size of inside-region training points at each iteration, t, is defined as

$$n^t = \lfloor \frac{\#\{X_{in}\}}{N^{itr}+1} t \rfloor \tag{23}$$

where the function $\lfloor x \rfloor$ returns the largest integer that is less than or equal to x. The Optimal Latin Hypercube based on Translational Propagation algorithm[26] is adopted to determine the locations of the full set of sample points $\{X\}$, and additional test points for the benchmark problems. This design of experiments method is accompanied by the modified maximum distance criterion.

Table 1. Numerical setup for test problems

Function	No. of variables	Total No. of sample points, $\{X\}$	No. of iteration, N^{it}	No. of inside points, $\{X_{in}\}$
Branin-Hoo function	2	30	4	11
Hartmann-3 function	3	45	4	14
Hartmann-6 function	6	100	4	22
Wind farm power generation	2	30	4	30

To predict the level of the surrogates error in the entire domain in the wind farm power generation problem, the region of interest is defined as equal to the design space ($\{X\} = \{X_{in}\}$). In this problem, the size of the inside-region training points at each iteration is defined by the following ascending series; $n^t = 21, 23, 25, 27$.

D. Selection of Parameters

To implement the Kriging method, the DACE (design and analysis of computer experiments) package developed by Lophaven et al.[27]is used. The bounds on the correlation parameters in the nonlinear optimization in Kriging, θ_{l} and θ_{u} are specified to be 0.1 and 20, respectively. Under the Kriging approach, the zero-order polynomial function is used as a regression model. To implement the RBF, the multiquadric radial basis function[7] is used where the shape parameter is set to c = 0.9. In implementation of E-RBF,[28] the shape parameter is set to c = 0.9; the λ parameter is set to 4.75; and the order of monomial in non-radial basis functions is fixed at 2.

E. Performance Criteria

The relative absolute error (RAE) evaluated on additional test points are used to evaluate the actual accuracy of the final surrogate within the region of interest. This criterion is used to compare the performance of the newly developed error estimation method, *REES*, and the relative absolute error of cross-validation, $\mathrm{RAE_{CV}}$, within the region of interest. The fractions of the errors evaluated using *REES* and $\mathrm{RAE_{CV}}$ with respect to the actual error ($\mathrm{RAE_{CV}}$) are represented by $\mathrm{R_{REES}}$ and $\mathrm{R_{RAE_{CV}}}$, respectively.

$$R_{REES} = \frac{E^{REES}_{Mo_{med}}}{E^{RAE}_{med}} \tag{24}$$

$$R_{RAE_{CV}} = \frac{E^{RAE_{CV}}_{med}}{E^{RAE}_{med}} \tag{25}$$

where $\mathrm{E^{REES}_{Mo_{med}}}$ represents the predicted mode of median errors in a surrogate; $\mathrm{E^{RAE}_{med}}$ represents the median of RAEs evaluated on test points; and $\mathrm{E^{RAE_{CV}}_{med}}$ represents the median of relative absolute errors of cross-validation, i.e.,

$$E^{RAE}_{med} = median(RAE_i), \quad i = 1, 2, ..., \#\{X_{test}\} \tag{26}$$

$$E^{RAE_{CV}}_{med} = median(RAE_{CV_j}), \quad j = 1, 2, ..., \#\{X_{in}\} \tag{27}$$

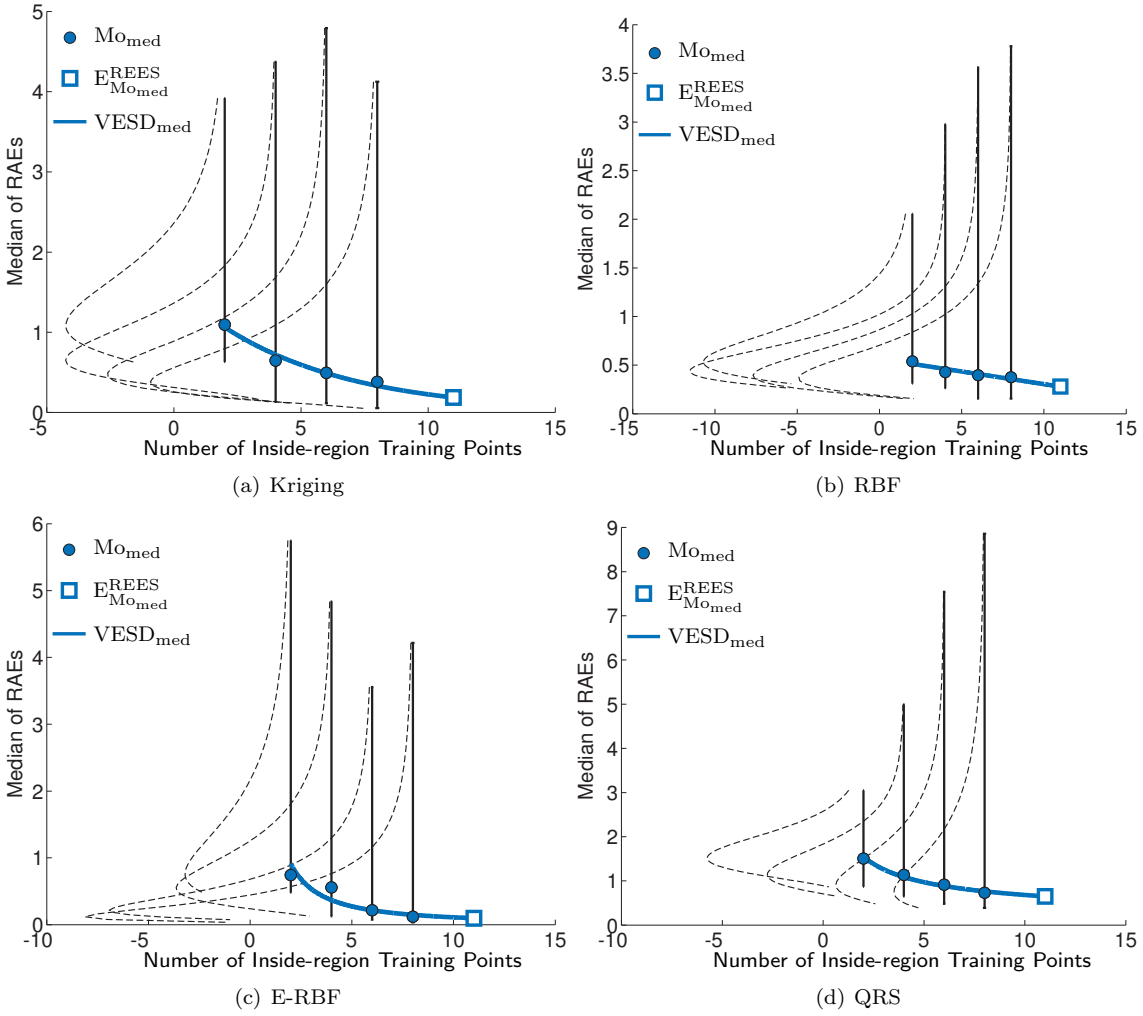
where the parameter $\#\{\mathrm{X_{test}}\}$ represents the number of test points. In this study, the total number of test points used to calculate RAE is equal to 50 times of the number of inside-region sample points ($50 \times \#\{\mathrm{X_{in}}\}$). The closer the fractions $\mathrm{R_{REES}}$ and $\mathrm{R_{RAE_{CV}}}$ are to one, the better the corresponding error measure.

IV. Results and Discussion

A. Benchmark Problems

Figures 4-6 represent VESD regression models within the region of interest of surrogate models (Kriging, RBF, E-RBF, and QRS) constructed for the three benchmark problems. The distributions of median errors, and the mode of the median error distributions ($\mathrm{Mo_{med}}$) are illustrated at each iteration. In Figs. 4-6 the solid block circles represent the quantified mode of median errors at each iteration; the square represents the predicted mode of median error in the final surrogate. It is observed that the modes of the error distributions relatively decrease with increasing density of inside-region training points, as hypothesized. Based on these observations, the variation of error with training point density can be modeled by applying the regression models (exponential, multiplicative, and linear) described in Section I.B.

The type and coefficients of the VESD regression models fitted to the modes of median errors ($\mathrm{Mo_{med}}$) of surrogates in the benchmark problems are provided in Tables 2-4. The predicted median of errors ($\mathrm{E^{REES}_{Mo_{med}}}$) in the final surrogates can also be observed in these figures.Further numerical details of the results are provided in Table 5.

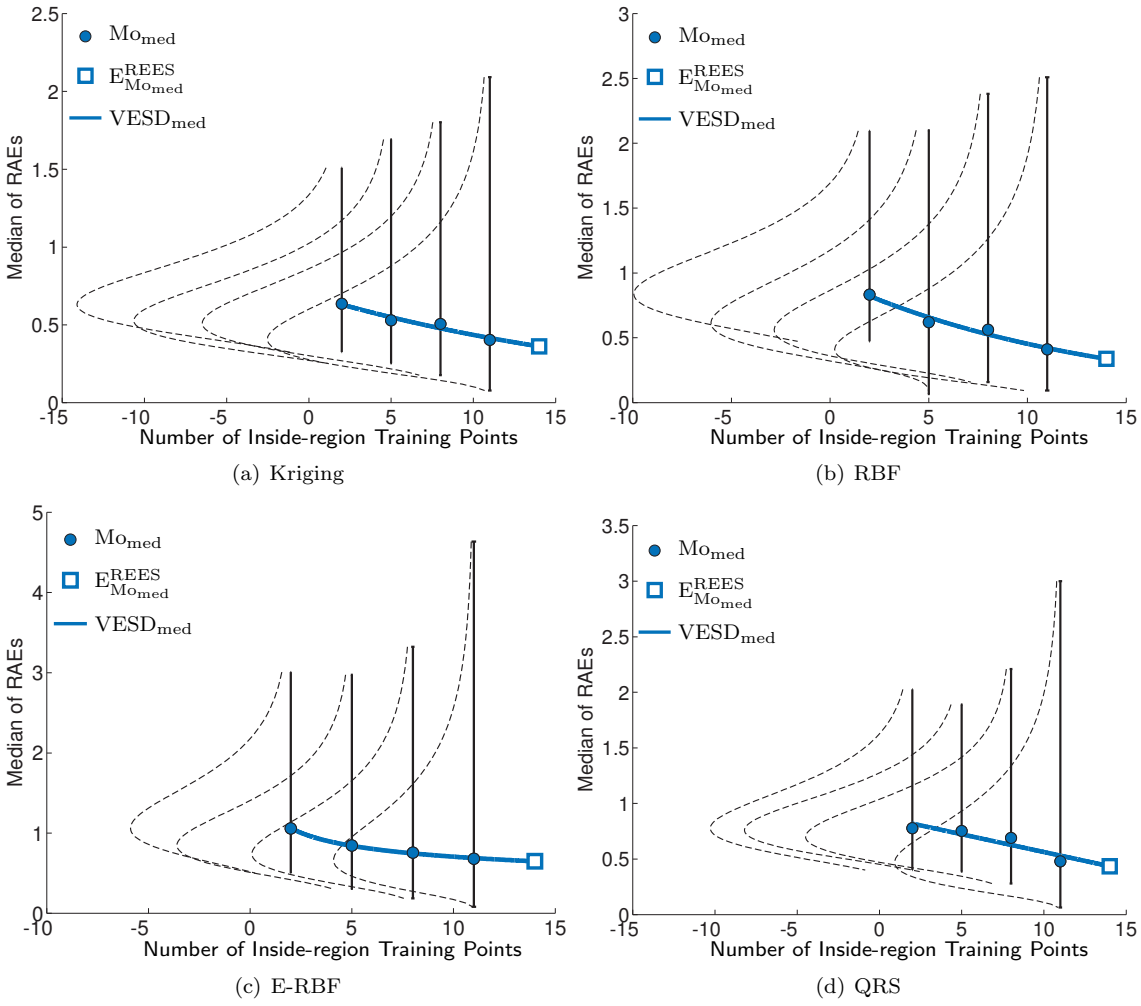


(a) Kriging (b) RBF

(c) E-RBF (d) QRS

Figure 4. VESD model to predict $\mathrm{E^{REES}_{Mo_{med}}}$ in Branin-Hoo Function (with 2 design variables)

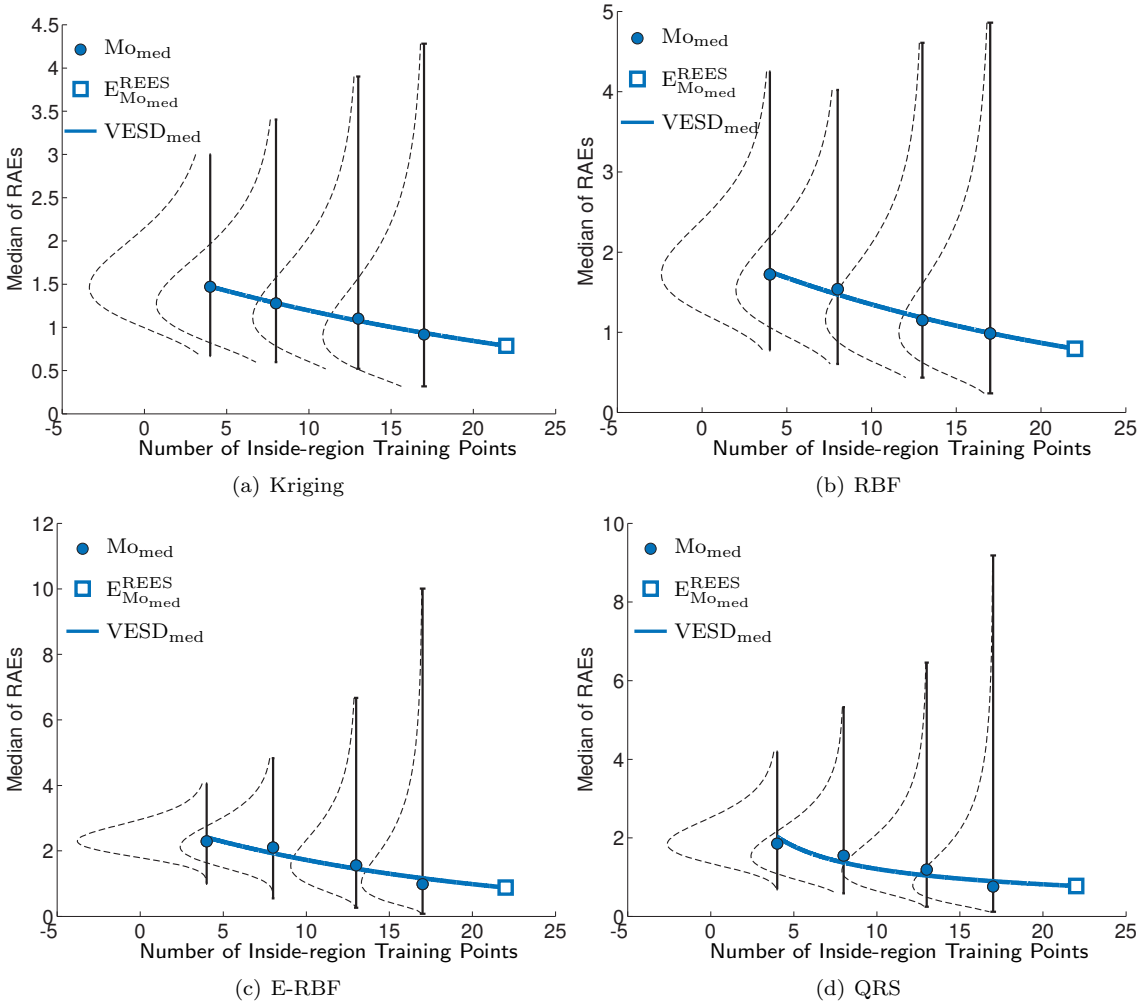
The VESD models used to predict the mode of maximum ($\mathrm{E^{REES}_{Mo_{max}}}$) and the absolute maximum ($\mathrm{E^{REES}_{ABS_{max}}}$) error of surrogates for the three benchmark problems are illustrated in Figs. 7-9. The $\mathrm{VESD_{max}}$ and $\mathrm{VESD_{ABS}}$ regression models are trained using mode of the maximum error distributions ($\mathrm{Mo_{max}}$) and absolute maximum error ($\mathrm{ABS_{max}}$), respectively. It is observed that the modes of the maximum error distributions



(a) Kriging (b) RBF

(c) E-RBF (d) QRS

Figure 5. VESD model to predict $E^{REES}_{Mo_{med}}$ in Hartmann Function (with 3 design variables)



(a) Kriging

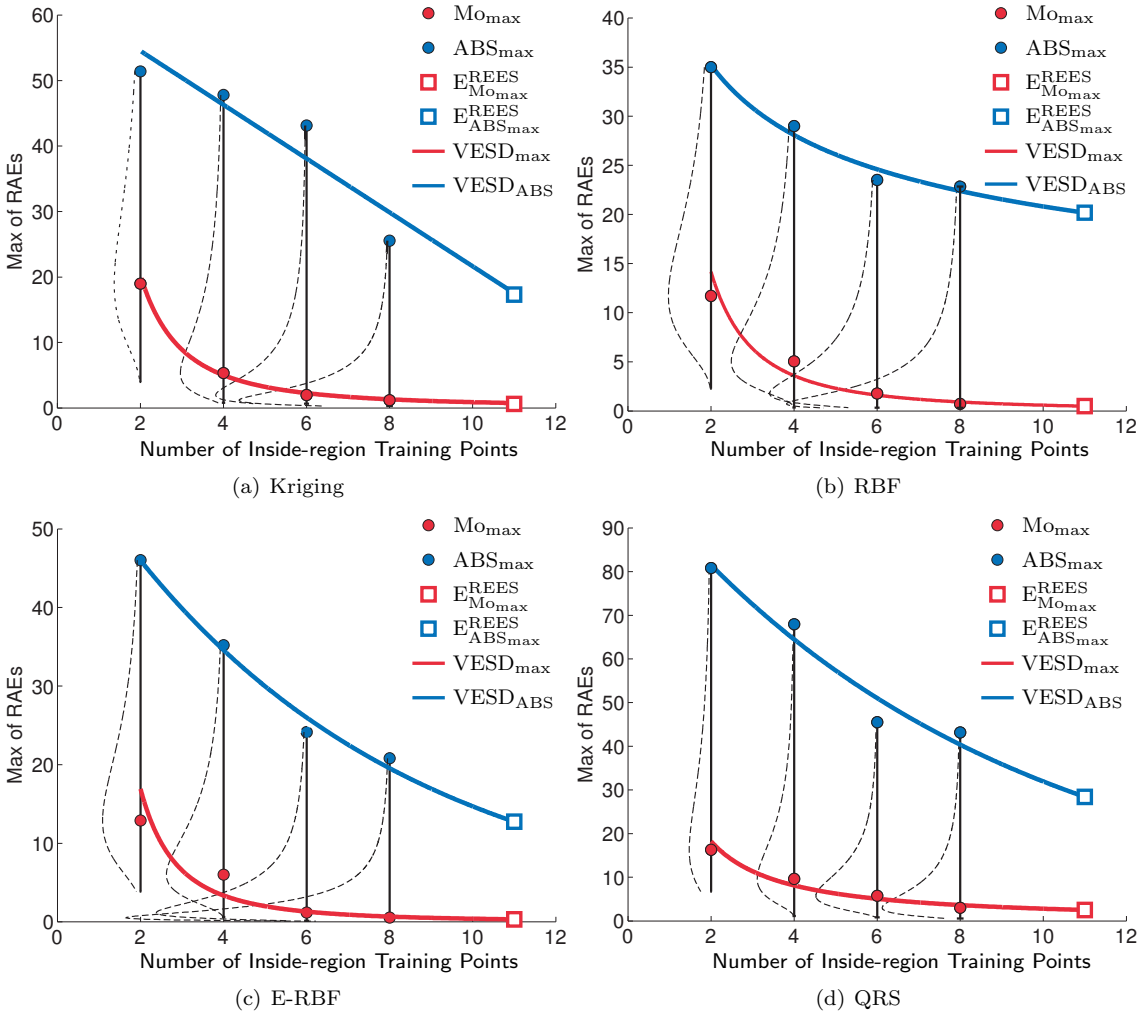
(b) RBF

(c) E-RBF

(d) QRS

Figure 6. VESD model to predict $E^{REES}_{Mo_{med}}$ in Hartmann Function (with 6 design variables)

relatively decrease with increasing density of inside-region training points, as expected. While, in certain iterations the level of the absolute maximum error is higher than that in preceding iteration. This observation can be primarily attributed to the outliers which are not eliminated in the modeling of the maximum error distributions and evaluation of the absolute maximum error. The type and coefficients of the $\mathrm{VESD_{max}}$ and $\mathrm{VESD_{ABS}}$ regression models for the three test problems are given in Tables 2-4.



(a) Kriging (b) RBF

(c) E-RBF (d) QRS

Figure 7. VESD model to predict $\mathrm{E^{REES}_{Mo_{max}}}$ and $\mathrm{E^{REES}_{ABS_{max}}}$ in Branin-Hoo Function (with 2 design variables)

The comparison of the performance of the REES error measure and the RAE given by *leave-one-out cross-validation* ($\mathrm{RAE_{CV}}$) with actual error evaluated on additional test point ($\mathrm{RAE_{actual}}$) are illustrated through bar diagrams in Figs. 10-12. In these figures, the y-axis represents the normalized surrogate error in the log scale. These comparisons are given for surrogates (Kriging, RBF, E-RBF, and QRS) constructed for the three benchmark test problems.

In these figures the modes of the median and the maximum errors, and the absolute maximum errors predicted using REES method are illustrated. Likewise, the median, the 75^{th} percentile, the 95^{th} percentile, and the absolute maximum error estimated using $\mathrm{RAE_{CV}}$ and $\mathrm{RAE_{actual}}$ are also presented. The mode of the median error predicted using RESS method ($\mathrm{E_{Mo^{REES}_{med}}}$), the median of the error estimated using $\mathrm{RAE_{CV}}$ method ($\mathrm{E^{RAE_{CV}}_{med}}$), and the median of the actual error evaluated using RAE method ($\mathrm{E^{RAE}_{med}}$) are listed in Table 5.

The performance criteria defined in Sec. III (R-values in Eqs. 25 and 25) is applied to compare the performance of the REES method and $\mathrm{RA_{CV}}$ with actual error on predicting the median error of the final surrogate. The comparison results are provided in Table 6. It is helpful to note that values closer to one are desirable for these metrics. From Figs. 10-12 and Table 6, it is observed that the overall performance of the

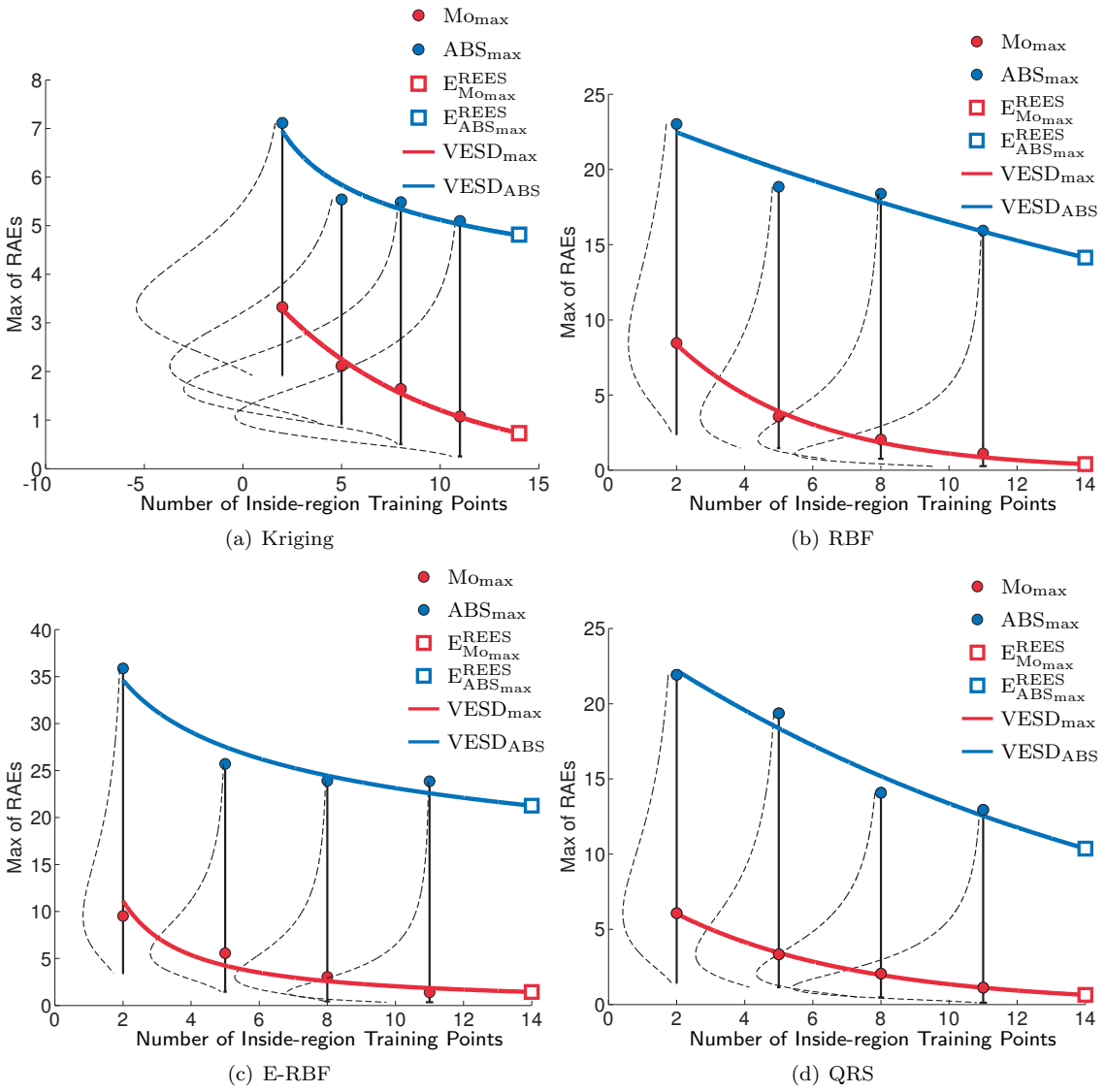
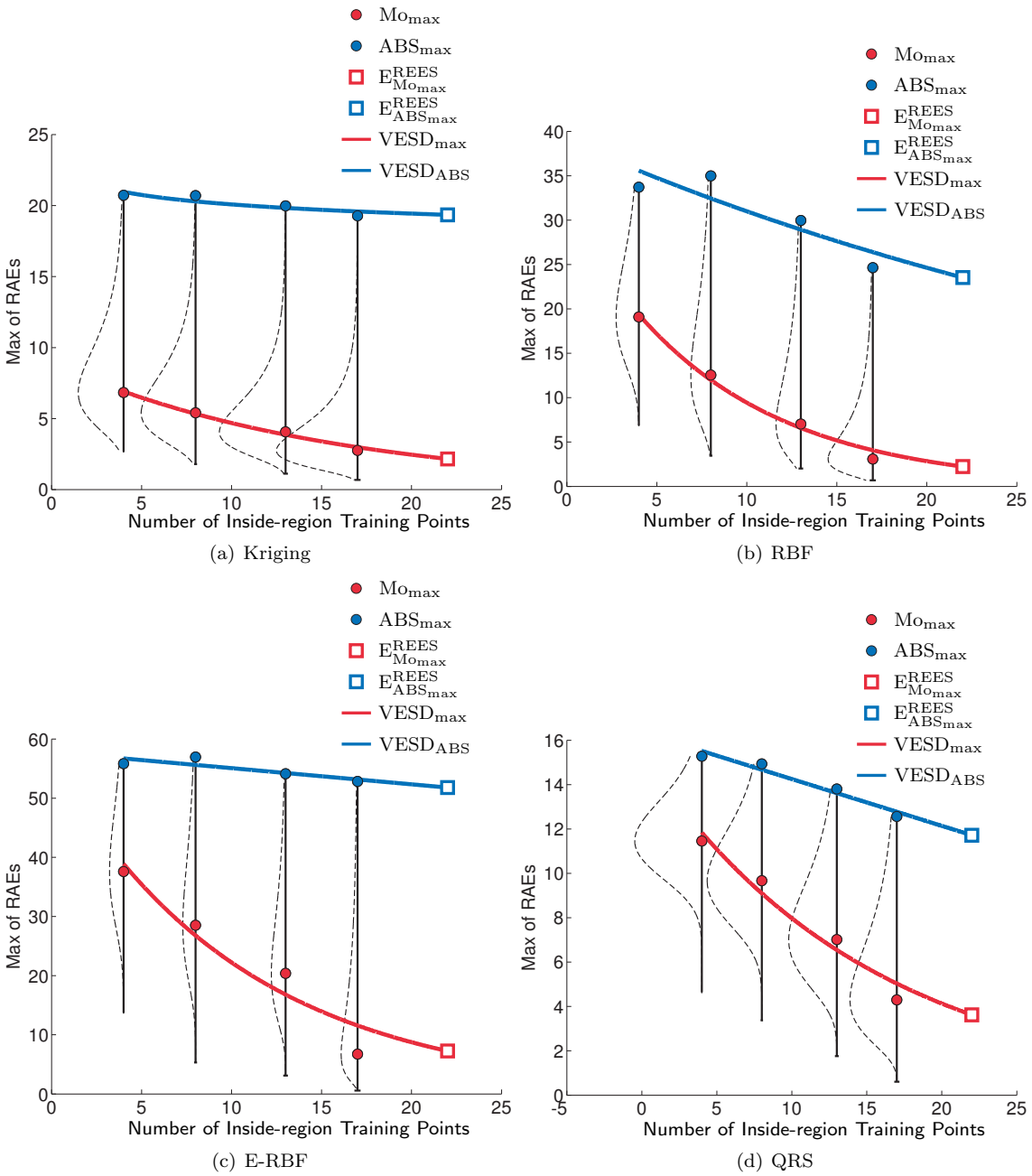


Figure 8. VESD model to predict $E^{REES}_{Mo_{max}}$ and $E^{REES}_{ABS_{max}}$ in Hartmann Function (with 3 design variables)

Table 2. VESD Coefficients for Branin-Hoo Function (with 2 design variables)

	$VESD_{med}$			$VESD_{max}$			$VESD_{ABS}$		
Surrogate	VESD model	a_0	a_1	VESD model	a_0	a_1	VESD model	a_0	a_1
Kriging	Exponential	1.56	-0.19	Multiplicative	80.24	-2.02	Linear	62.52	-4.11
RBF	Linear	0.56	-0.02	Multiplicative	55.82	-1.97	Multiplicative	44.18	-0.32
E-RBF	Multiplicative	2.27	-1.31	Multiplicative	85.37	-2.33	Linear	61.17	-0.14
QRS	Multiplicative	2.20	-0.50	Multiplicative	40.82	-1.16	Linear	103.03	-0.11



(a) Kriging (b) RBF

(c) E-RBF (d) QRS

Figure 9. VESD model to predict $E^{REES}_{Mo_{max}}$ and $E^{REES}_{ABS_{max}}$ in Hartmann Function (with 6 design variables)

Table 3. VESD Coefficients for Hartmann Function (with 3 design variables)

	$VESD_{med}$			$VESD_{max}$			$VESD_{ABS}$		
Surrogate	VESD model	a_0	a_1	VESD model	a_0	a_1	VESD model	a_0	a_1
Kriging	Exponential	0.69	-0.04	Exponential	4.02	-0.12	Multiplicative	7.94	-0.18
RBF	Exponential	0.95	-0.07	Exponential	13.85	-0.25	Exponential	24.28	-0.03
E-RBF	Multiplicative	1.26	-0.25	Multiplicative	23.08	-1.05	Multiplicative	41.18	-0.25
QRS	Linear	0.88	-0.03	Exponential	8.78	-0.18	Exponential	25.23	-0.06

Table 4. VESD Coefficients for Hartmann Function (with 6 design variables)

	$VESD_{med}$			$VESD_{max}$			$VESD_{ABS}$		
Surrogate	VESD model	a_0	a_1	VESD model	a_0	a_1	VESD model	a_0	a_1
Kriging	Exponential	1.69	-0.03	Exponential	8.94	-0.06	Multiplicative	22.38	-0.04
RBF	Exponential	2.09	-0.04	Exponential	31.28	-0.11	Exponential	38.99	-0.02
E-RBF	Exponential	3.01	-0.05	Exponential	56.47	-0.09	Linear	57.80	-0.27
QRS	Multiplicative	4.42	-0.56	Exponential	15.40	-0.06	Linear	16.36	-0.21

REES method for surrogate error quantification is significantly better than that of the RAE estimated by *leave-one-out cross-validation*.

Table 5. Median Error Estimated using $REES$, RAE_{CV}, and RAE_{actual}

	Kriging			RBF			E-RBF			QRS		
Function	REES	RAE_{cv}	RAE_{actual}	REES	RAE_{cv}	RAE_{actual}	REES	RAE_{cv}	RAE_{actual}	REES	RAE_{cv}	RAE_{actual}
Branin-Hoo	0.185	0.846	0.141	0.279	0.428	0.152	0.097	0.145	0.125	0.650	0.011	0.856
Hartmann-3	0.361	0.454	0.255	0.337	0.547	0.203	0.647	1.512	0.792	0.435	0.018	0.491
Hartmann-6	0.786	1.078	0.504	0.794	1.765	0.595	0.881	1.517	4.333	0.774	0.078	3.147
Wind farm problem	0.0016	0.0050	0.0032	.0054	0.0090	.0059	0.0049	0.0084	0.0050	0.0020	1.55e-6	0.0044

Table 6. Fraction of error evaluated using $REES$ and RAE_{CV} over RAE

	Kriging		RBF		E-RBF		QRS	
Function	R_{REES}	$R_{RAE_{CV}}$	R_{REES}	$R_{RAE_{CV}}$	R_{REES}	$R_{RAE_{CV}}$	R_{REES}	$R_{RAE_{CV}}$
Branin-Hoo	1.31	2.56	1.83	2.81	0.77	1.16	0.75	0.01
Hartmann-3	1.41	1.78	1.66	2.69	0.81	1.90	0.88	0.03
Hartmann-6	1.55	2.13	1.33	2.96	0.20	0.35	0.24	0.02
Wind farm problem	0.50	1.56	0.91	1.52	0.98	1.68	0.45	3.5e-4

B. Wind Farm Power Generation Problem

In this problem, the REES method is applied to explore the performance of different surrogates for the wind farm power generation problem. The variation of error with sample density (VESD) within the entire design space is illustrated in Fig. 13. The VESD regression models are trained using the statistical mode of the median error distributions (Mo_{med}) at each step. Expectedly, it is observed that in the four surrogates, the level of the statistical mode (Mo_{med}) decreases when increasing density of training points as desired. It is also observed that, at each step the relative accuracy of the intermediate surrogates constructed using Kriging and QRS is relatively higher than those given by RBF and E-RBF.

The predicted levels of errors in surrogates by REES and RAE_{CV} are shown in Figs. 14(a) and 14(b), respectively. Figure 14(c) shows the actual error (RAE_{actual}) in the wind farm power generation problem. The comparison results based on the performance criteria (R-values), are also provided in Table 6. It is observed that the REES method has relatively higher accuracy than RAE_{CV} in predicting levels of errors. From Table 5, it is also observed that, Kriging has relatively higher accuracy in this problem based on the REESs' predicted errors and actual errors evaluated on test points. This observation shows that REES can be effectively used as a model selection method to select the best surrogate for applications such as optimization.

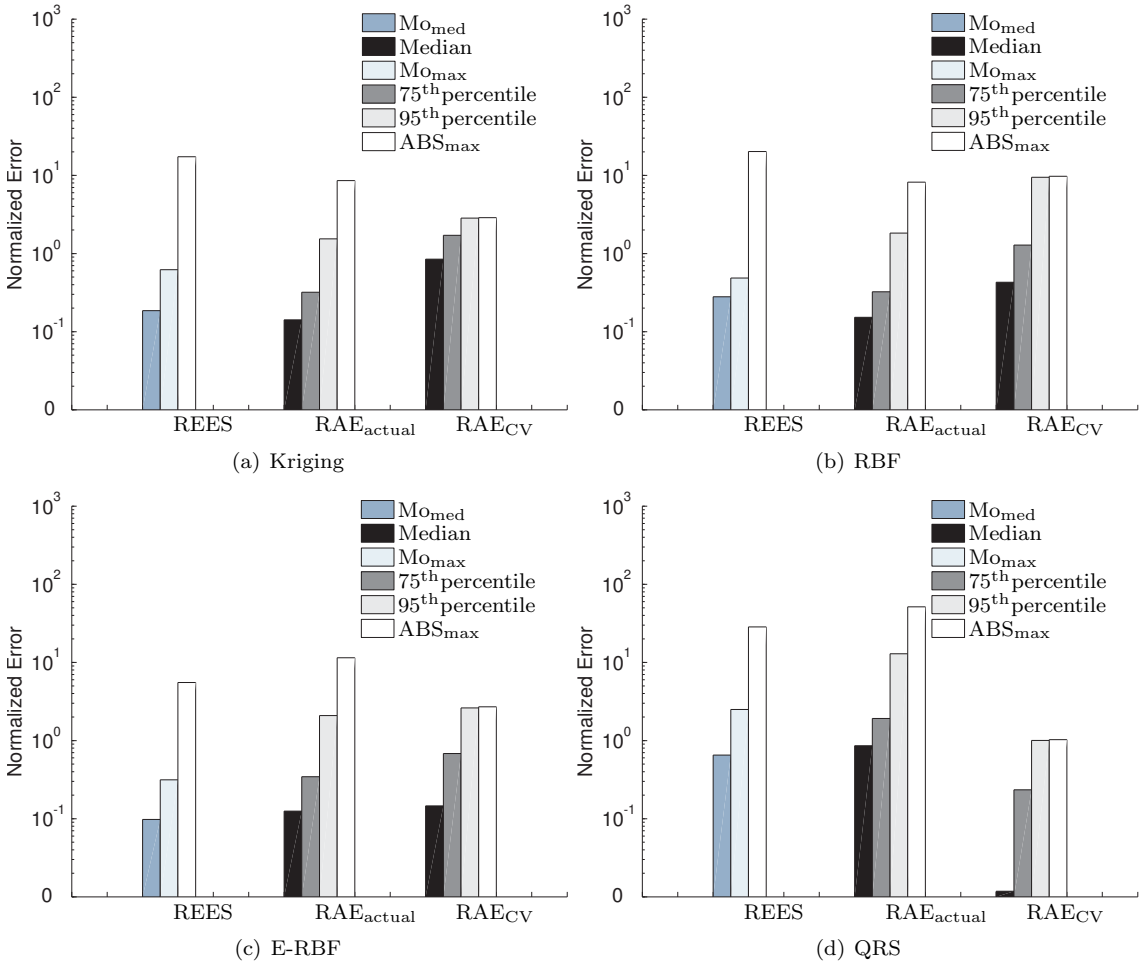


Figure 10. Comparison of the performance of the REES and $\mathrm{RAE}_{\mathrm{CV}}$ with $\mathrm{RAE}_{\mathrm{actual}}$ in Branin-Hoo Function (with 2 design variables)

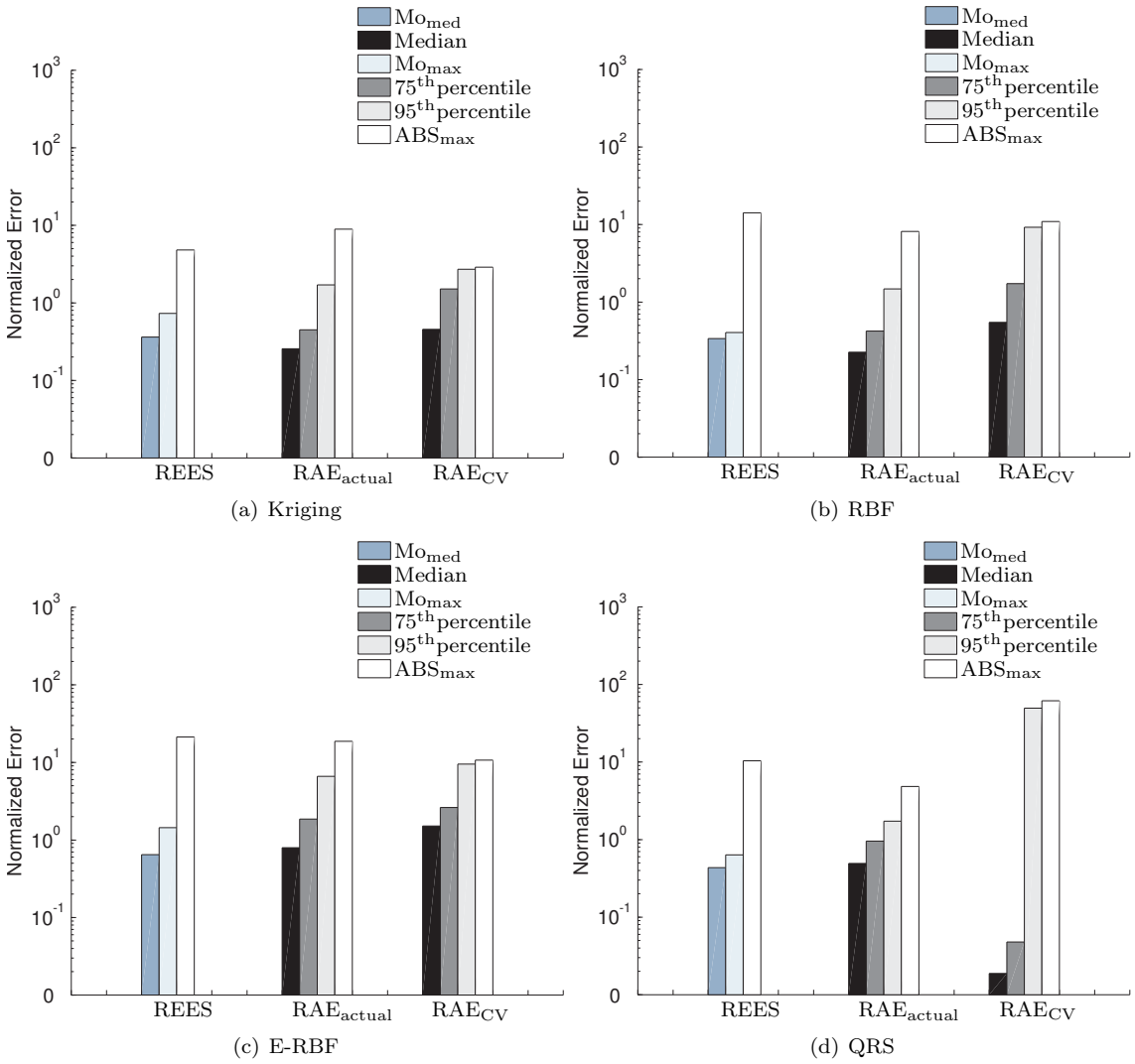


Figure 11. Comparison of the performance of the REES and $\mathrm{RAE_{CV}}$ with $\mathrm{RAE_{actual}}$ in Hartmann Function (with 3 design variables)

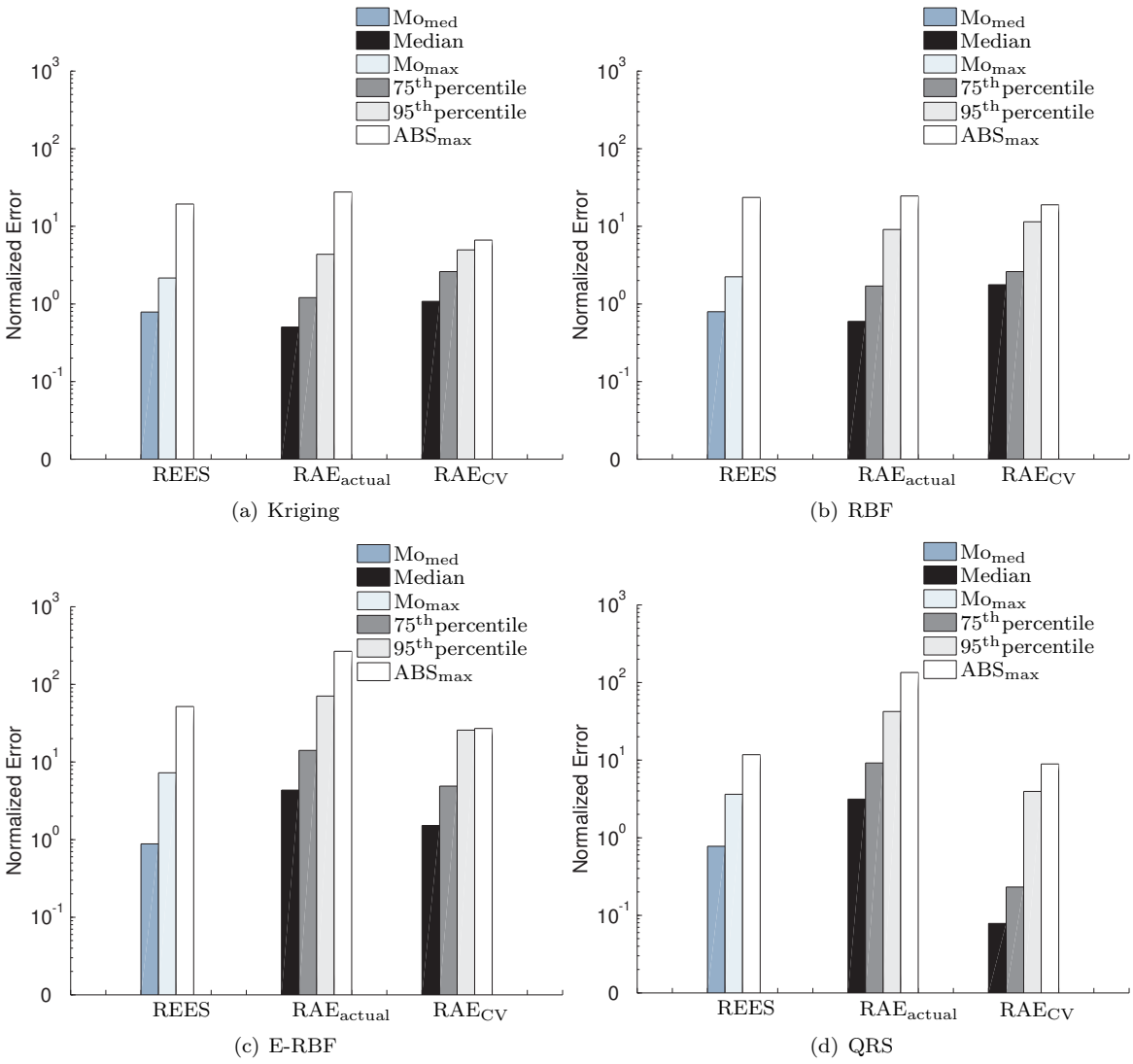


Figure 12. Comparison of the performance of the REES and $\mathrm{RAE_{CV}}$ with $\mathrm{RAE_{actual}}$ in Hartmann Function (with 6 design variables)

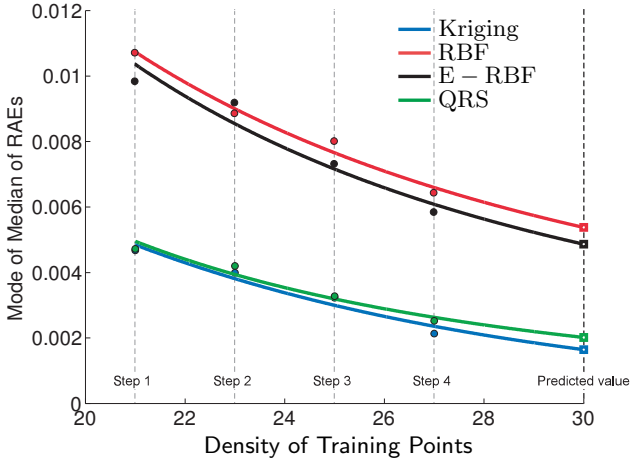
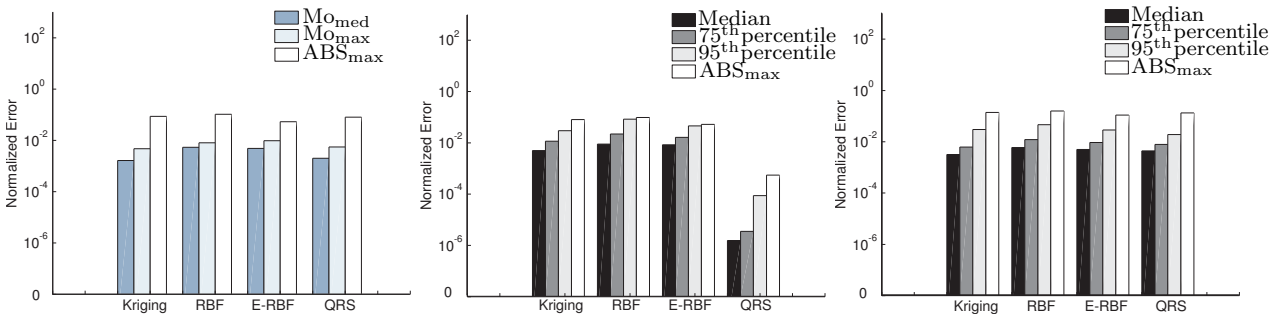


Figure 13. VESD regression models in different surrogates for the wind farm power generation problem



(a) REES (b) $\mathrm{RAE_{CV}}$ (c) $\mathrm{RAE_{actual}}$

Figure 14. Comparison of the performance of the REES and $\mathrm{RAE_{CV}}$ with $\mathrm{RAE_{actual}}$ in the wind farm power generation problem

V. Conclusion

This paper develops a new approach to quantify the surrogate accuracy in any given region of the design domain. Such an approach can be useful for informed decision-making when using surrogates for analysis and optimization. This approach is based on the Regional Error Estimation in Surrogates method, which is a model independent method for quantifying regional errors. Intermediate surrogates are iteratively created with heuristic subsets of the available sample points and the remaining sample points are used to evaluate the error in the estimated function. Regression models are then developed to represent the regional error in the surrogate as functions of the training point densities inside and outside the region of interest. These regression models are then used to predict the regional error levels for the final surrogate that is trained using all available sample points. The effectiveness of the new regional error estimation method for surrogates are illustrated using standard test problems and a wind farm power generation problem. The application of *REES* to a standard test problems and practical engineering problem show that the proposed method provides significantly better measure of surrogate error compared to relative absolute error estimated by *leave-one-out cross-validation*. The preliminary results show that, a model selection based on *REES* criterion can be more reliable than that using other error measures such as *RAE* given by cross-validation. The future application of *REES* for model selection criterion will further establish the potentials of this error quantification method.

References

[1]Kleijnen, J., *Statistical techniques in simulation*, Vol. II, New York, 1975, Russian translation, Publishing House Statistics, Moscow, 1978.

[2]Jin, R., Chen, W., and Simpson, T. W., "Comparative Studies of Metamodeling Techniques Under Multiple Modeling Criteria," *AIAA*, , No. 4801, 2000.

[3]Simpson, T., Korte, J., Mauery, T., and Mistree, F., "Kriging Models for Global Approximation in Simulation-Based Multidisciplinary Design Optimization," *AIAA Journal*, Vol. 39, No. 12, 2001, pp. 2233–2241.

[4]Forrester, A. and Keane, A., "Recent Advances in Surrogate-based Optimization," *Progress in Aerospace Sciences*, Vol. 45, No. 1-3, 2009, pp. 50–79.

[5]Choi, K., Young, B., and Yang, R., "Moving Least Square Method for Reliability-based Design Optimization," 4th World Congress of Structural and Multidisciplinary Optimization, Dalian, China, June 2001, pp. 4–8.

[6]Toropov, V. V., Schramm, U., Sahai, A., Jones, R. D., and Zeguer, T., "Design Optimization and Stochastic Analysis based on the Moving Least Squares Method," 6th World Congresses of Structural and Multidisciplinary Optimization, Rio de Janeiro, 30 May - 03 June 2005.

[7]Hardy, R. L., "Multiquadric Equations of Topography and Other Irregular Surfaces," *Journal of Geophysical Research*, Vol. 76, 1971, pp. 1905–1915.

[8]Yegnanarayana, B., *Artificial Neural Networks*, PHI Learning Pvt. Ltd., 2004.

[9]Queipo, N., Haftka, R., Shyy, W., Goel, T., Vaidyanathan, R., and Tucker, P., "Surrogate-based Analysis and Optimization," *Progress in Aerospace Sciences*, Vol. 41, No. 1, 2005, pp. 1–28.

[10]Zhang, J., Chowdhury, S., and Messac, A., "An Adaptive Hybrid Surrogate Model," *Structural and Multidisciplinary Optimization*, Vol. 46, No. 2, 2012, pp. 223–238.

[11]Wang, G. and Shan, S., "Review of Metamodeling Techniques in Support of Engineering Design Optimization," *Journal of Mechanical Design*, Vol. 129, No. 4, 2007, pp. 370–381.

[12]Lehmensiek, R., Meyer, P., and Muller, M., "Adaptive sampling applied to multivariate, multiple output rational interpolation models with application to microwave circuits," *International Journal of RF and Microwave Computer-Aided Engineering*, Vol. 12, No. 4, 2002, pp. 332–340.

[13]Sugiyama, M., "Active Learning in Approximately Linear Regression Based on Conditional Expectation of Generalization Error," *Journal of Machine Learning Research*, Vol. 7, 2006, pp. 141–166.

[14]Jones, D., Schonlau, M., and Welch, W., "Efficient Global Optimization of Expensive Black-box Functions," *Journal of Global Optimization*, Vol. 13, No. 4, 1998, pp. 455–492.

[15]Mehmani, A., Zhang, J., Chowdhury, S., and Messac, A., "Surrogate-based Design Optimization with Adaptive Sequential Sampling," *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, Hawaii, USA, April 2012.

[16]Meckesheimer, M., Barton, R., Simpson, T., and Booker, A., "Computationally inexpensive Metamodel Assessment Strategies," *AIAA*, Vol. 40, No. 10, 2002, pp. 2053–2060.

[17]Goel, T., Haftka, R., and Shyy, W., "Ensemble of Surrogates," *Structural and Multidisciplinary Optimization*, Vol. 38, 2009, pp. 429442.

[18]Viana, F., Haftka, R., and Steffen, V., "Multiple Surrogates: How Cross-validation Errors Can Help Us to Obtain the Best Predictor," *Structural and Multidisciplinary Optimization*, Vol. 39, No. 4, 2009, pp. 439–457.

[19]Bozdogan, H., "Akaike's information criterion and recent developments in information complexity," *Journal of Mathematical Psychology*, Vol. 44, 2000, pp. 62–91.

[20]Nguyen, H. M., Couckuyt, I., Knockaert, L., Dhaene, T., Gorissen, D., and Saeys, Y., "An Alternative Approach to Avoid Overfitting for Surrogate Models," *Proceedings of the 2011 Winter Simulation Conference*, 2011, pp. 2765–2776.

[21]Viana, F. A. C., Haftka, R. T., and Jr., V. S., "Multiple surrogates: how cross-validation errors can help us to obtain the best predictor," *Structural and Multidisciplinary Optimization*, Vol. 39, 2009, pp. 439457.

[22]Zhang, J., Chowdhury, S., Mehmani, A., and Messac, A., "Uncertainty Quantification in Surrogate Models Based on Pattern Classification of Cross-validation Errors," *14th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Indiana, USA, September 2012.

[23]Chowdhury, S., Zhang, J., Messac, A., and Castillo, L., "Unrestricted Wind Farm Layout Optimization (UWFLO): Investigating Key Factors Influencing the Maximum Power Generation," *Renewable Energy*, Vol. 38, No. 1, 2012, pp. 16–30.

[24]Chowdhury, S., Zhang, J., Messac, A., and Castillo, L., "Optimizing the Arrangement and the Selection of Turbines for a Wind Farm Subject to Varying Wind Conditions," *Renewable Energy*, Vol. 52, 2013, pp. 273–282.

[25]NDAWN, "The North Dakota Agricultural Weather Network," 2010, `http://ndawn.ndsu.nodak.edu/`.

[26]Viana, F. A. C., Venter, G., and Balabanov, V., "An algorithm for fast optimal latin hypercube design of experiments," *International Journal for Numerical Methods in Engineering*, Vol. 82, No. 2, 2010, pp. 135156.

[27]Lophaven, S. N., Nielsen, H. B., and Sondergaard, J., "DACE - A Matlab Kriging Toolbox, Version 2.0," Tech. Rep. IMM-REP-2002-12, Informatics and Mathematical Modelling Report, Technical University of Denmark, 2002.

[28]Mullur, A. and Messac, A., "Extended Radial Basis Functions: More Flexible and Effective Metamodeling," *AIAA Journal*, Vol. 43, No. 6, 2005, pp. 1306–1315.