Regional Error Estimation of Surrogates (REES)

Ali Mehmani

Syracuse University, Syracuse, NY, 13244

Souma Chowdhury† and Jie Zhang †

Rensselaer Polytechnic Institute, Troy, New York 12180

Achille Messac‡

Syracuse University, Syracuse, NY, 13244

Surrogate-based design is an effective approach for modeling computationally expensive system behavior. In such application, it is often challenging to characterize the expected accuracy of the surrogate. In addition to global and local error measures, regional error measures can be used to understand and interpret the surrogate accuracy in the regions of interest. This paper develops the Regional Error Estimation of Surrogate (REES) method to quantify the level of the error in any given subspace (or region) of the entire domain, when all the available training points have been invested to build the surrogate. In this approach, the accuracy of the surrogate in each subspace is estimated by modeling the variations of the mean and the maximum error in that subspace with increasing number of training points (in an iterative process). A regression model is used for this purpose. At each iteration, the intermediate surrogate is constructed using a subset of the entire training data, and tested over the remaining points. The evaluated errors at the intermediate test points at each iteration are used for training the regression model that represents the error variation with sample points. The effectiveness of the proposed method is illustrated using standard test problems. To this end, the predicted regional errors of the surrogate constructed using all the training points are compared with the regional errors estimated over a large set of test points.

Keywords: Kriging; Regional error measures; Sampling; Surrogate models;

I. Introduction

A. Surrogate Modeling

Engineering design problems often involve computationally intensive simulation models (high fidelity models) or expensive experiment-based system evaluations. An accurate surrogate model is an effective tool for providing a tractable and an inexpensive approximation of the actual system evaluation. During the last two decades, the use of mathematical approximation models in design space exploration, sensitivity analysis, and optimization has become popular for reducing the computational cost and filtering the numerical noise of computationally intensive system evaluations.1 This approximation model is known as a Surrogate or Metamodel (model of the models).2 Popular surrogate modeling methods include polynomial response surfaces,3 Kriging,4,5 Moving Least Square,6,7 radial basis functions,8 neural networks,9 and hybrid surrogate modeling.10,11 These methods have been applied to a wide range of disciplines, such as aerospace design,
B. Review of Surrogate Model Error Measurement Methods

In this section, the methods for assessing errors in surrogates are reviewed. 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.17 Expensive methods are typically using additional data to compare the surrogate’s response to the actual response. When extra data is not available, inexpensive methods use existing data (the training data for constructing of the surrogate) to quantify the level of errors in surrogates. Error measure methods can also be classified into global and local error estimations.18 The overall performance of the developed surrogates can be evaluated using global error measures, while the local or point-wise error measures evaluate the surrogate in different location of the surrogates.

Popular approaches of global error measures include:10 (i) split sample, (ii) cross-validation, and (iii) bootstrapping. 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 Leave-one-out cross-validation, the training set is created by taking all sample points except one, and the left out point is used for estimating error between the surrogate prediction and the actual value. In q-fold cross-validation, 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, in each iteration, is used to evaluate the cross-validation error.19 The bootstrapping approach generates m subsamples from the sample points. Each subsample is a random sample with replacement from the full sample. Different variants of the bootstrapping approach can be used for (i) model identification, and (ii) identifying confidence intervals for surrogates.10

Local approaches of error measures are used to estimate the level of errors in specific locations of a design space. Examples of the local error measures include: (i) the mean square errors for Kriging15 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. In particular, 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 the point $x^*$ is much more than it would be if $x^*$ was 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 sample points and is equal to $\sqrt{\sigma^2}$ at a point far away from sample points, where $\sigma^2$ 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

Besides the error measures mentioned above, there are other error metrics: (i) mean squared error (MSE) or root mean square error (RMSE) which provides a global error measure over the entire design domain, and (ii) maximum absolute error (MAE) and relative absolute error (RAE), which are indicative of local deviation.

C. Research Objectives

The global measures, which can be helpful for representing the overall performance of the surrogates, may not adequately represent the regional accuracy of a complex system. The primary objective of this paper is to develop an effective methodology for quantifying the surrogate error in each region of the design domain;
this method is called the Regional Error Estimation of Surrogate (REES). The proposed regional error measure method is model independent (it is not limited to a specific kind of surrogate model). The REES uses the available training data set to predict the error in different regions without any additional system evaluations. This measure method provides useful information about the performance of the surrogate in different regions, thereby improving the usability of the surrogates in real-life engineering design problems.

The paper is organized as follows: Section II describes the developed formulation of the proposed regional error measure method. In Section III, the REES is evaluated by two two-dimensional standard test problems. The results are discussed in Section IV. Concluding remarks and future work are provided in Section V.

II. Regional Error Estimation for Surrogate

A. Regional Error Estimation of Surrogate (REES) Measure

The goal of the proposed approach is to provide a regional measure of the level of errors in an estimated function without investing additional (expensive) system evaluations. The REES method estimates the error of the surrogate in a given region by predicting the variation of the error in that region with increasing number of training points, which the total number of available training points remains fixed. The size of the training data set is increased in each step while the remaining sample points are used as test points to evaluate the error. The framework of the proposed methodology is illustrated in Fig. 1, and is described below:

Step 1. Generation of Training Data In this step, a set of experimental designs are chosen to cover the entire design space as uniformly as possible. Then, the system is evaluated over selected test data points. The entire set of sample points is represented by \( \{X\} \).

Step 2. Estimation of the Variation of the Error in Different Regions This step consists of an iterative process. In each iteration, the sample points are divided into a tentative training data set \( \{X^t_{TR}\} \) and a tentative test data set \( \{X^t_{TE}\} \), given by

\[
\{X^t_{TR}\} = \{X^{t-1}_{TR}\} + \{X^t_{TRA}\} \tag{1}
\]

where

\[
\{X^0_{TR}\} = \emptyset, \quad \{X^1_{TRA}\} \subset \{X\}, \quad \text{and} \quad \{X^t_{TRA}\}_{t>1} \subset \{X^t_{TE}\} \tag{2}
\]

The tentative set of training points \( \{X^t_{TR}\} \) at the iteration \( t \) consists of the previous set of training points \( \{X^{t-1}_{TR}\} \) and the current additional selected points \( \{X^t_{TRA}\} \). The additional selected points are a subset of the full set of the training data in the first iteration, and they are a subset of the test points in subsequent iterations.

The intermediate surrogate \( (f_t) \) at each iteration is constructed using the tentative training points and is evaluated over the tentative test points. Then, domain segmentation strategy is applied to divide the design space into subspaces. Based on domain segmentation, the whole design domain is divided into physically meaningful classes. The boundary between classes can be determined using error-based pattern classification methods or can be predefined by the designer based on the level of the knowledge regarding the design problem. The mean error \( (E^t_M) \) and the maximum error \( (E^t_{Max}) \) are then estimated at each iteration, as given by

\[
E^t_M = [E^t_{Mean}], \quad E^t_{Mean} = \text{Mean}(e_{i1}, e_{i2}, ..., e_{im}),
\]

\[
E^t_{Max} = [E^t_{Max}], \quad E^t_{Max} = \text{Max}(e_{i1}, e_{i2}, ..., e_{im}), \tag{3}
\]

where \( i = 1, ..., n \).

In Eq. 3, the \( e_{ij} \) is the error of the \( j^{th} \) test point in the \( i^{th} \) subspace at iteration \( t \). The parameter \( n \) represents the number of subspaces and \( m_i \) is the number of test points in the \( i^{th} \) subspace at the \( t^{th} \) iteration.

Step 3. Final Surrogate Construction and Error Prediction in Different Regions The final surrogate model is constructed using the full set of training data. A regression model is developed to predict the level of errors in each region of the final surrogate. This regression model, called the Variation of the Error with Sample Points (VESP), is trained using the mean and the maximum error information from all the preceding iterations.

3 of 14

American Institute of Aeronautics and Astronautics
B. Implementation of the REES Measure

For the sake of simplicity, a two-dimensional problem is used to illustrate the implementation of the proposed method. The total number of sample points \( n_X \) allowed for this function is 32. The first tentative surrogate (at \( t = 1 \)) is constructed using 8 training points, out of the 32 points; and each subsequent surrogate is constructed using additional 8 points \( \{X_{TRA}^{t+1}\} \). The sampling scheme developed by Audze-Eglais is adopted to determine the locations of the full set of sample points \( \{X\} \). It is a Latin Hypercube-based design of experiments method that is accompanied by an optimality criterion as given by

\[
\sum_{p}^{P} \sum_{q=p+1}^{P} \frac{1}{L_{pq}^2} \rightarrow \text{min}
\]  

(4)

where \( L_{pq} \) is the distance between the points \( p \) and \( q \), and \( P \) is the total number of sample points. The Latin Hypercube based on max-min criterion is used for selecting additional points \( \{X_{TRA}\} \) from the total sample set \( \{X\} \) in each subsequent iteration. This criterion seeks to maximize the minimum distance between selected points. In each iteration, the set of training points consist of selected additional points \( \{X_{TRA}^{t}\} \) and the training points from the previous iteration \( \{X_{TR}^{t-1}\} \). The training points and the test points for the first, second, third, and forth iteration (before termination of iterative process at \( n_{X_{TR}} = 32 \)) are illustrated in Figs. 2(a), 2(b), 2(c), and 2(d), respectively.

The surrogates in this paper are developed using Kriging method, which is a widely used surrogate modeling method. The standard Kriging model is given by

\[
\hat{f}(x) = G(x) + Z(x)
\]

(5)

where \( \hat{f}(x) \) is the approximation of the actual response \( f(x) \); \( G(x) \) is the known type of approximation function (often a polynomial); and \( Z(x) \) is a weak stationary stochastic process with zero mean and variance \( \sigma^2 \). The \( ij^{th} \) element of the covariance matrix of \( Z(x) \) is given by

\[
\text{COV}[Z(x^i), Z(x^j)] = \sigma^2 R_{ij}
\]

(6)
Figure 2. Tentative Training and Test Points in 2D Function (the red data points represent training points and blue data points represent test points)
where $R_{ij}$ is the correlation function between the $i^{th}$ and the $j^{th}$ data points; and $\sigma^2$ is the process variance. In this paper, a Gaussian function is used as the correlation function. The intermediate surrogate model at each iteration is tested on the set of test points at that iteration. The Relative Accuracy Error (RAE) is used to evaluate the error at each test point. The RAE for test point $x_i$ is defined as

$$\text{RAE}(x_i) = \left| \frac{\tilde{f}(x_i) - f(x_i)}{f(x_i)} \right|$$

where $f(x_i)$ is the actual function value at $x_i$, and $\tilde{f}(x_i)$ represents the function value estimation by the surrogate model. The mean and maximum values of the error in each subspace of the domain are determined from the RAE estimated at all test points allowed at the concerned iteration. In the current paper, the entire design space is manually divided into a specific number of subspaces. In the case of the two-dimensional function, the design space is divided into four equal rectangular subspaces as shown in Fig. 3. The errors ($E_{\text{Mean}}$ and $E_{\text{Max}}$) in each subspace estimated during the iterative process are used to train the VESP regression model. The selection of the type of regression function is critical to the prediction of the regional error levels in the surrogate. In this paper, we explore an exponential regression function to train the VESP model, which is expressed as

$$\tilde{F}(x, a) = a_0 e^{(a_1 n_{XTR})}$$

where $a_0$ and $a_1$ are unknown coefficients, and $n_{XTR}$ represents the number of training points in the $t^{th}$ iteration. In the next section, the proposed method is applied to two benchmark test problems.

III. Numerical Examples and Testing Procedure

The performance of the Regional Error Estimation of Surrogate (REES) method is evaluated using the following analytical test problems: (i) Dixon & Price function, and (ii) Branin function.

Test Function 1: Dixon & Price Function

$$f(x) = (x_1 - 1)^2 + 2(2x_2^2 - x_1)^2$$

where $x_1 \in [-10, 10]$, $x_2 \in [-10, 10]$

Test Function 2: Branin Function

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

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

Figure 3. Illustration of dividing domain design into four equal rectangular subspaces in the two-dimensional function.
Numerical settings of two problems are given in Table 1. The table lists (i) the number of input variables, (ii) the total number of sample points \( n_X \), and (iii) the number of training \( (n_X^t_R) \) and test points \( (n_X^t_E) \) at each iteration. For the two test problems, the first tentative surrogate model is constructed using 16 points.

\[
(n_X^t_R = n_X^t_{TRA}). \text{ The number of additional points at each iteration is defined to be 4 } (n_X^t_{TRA} = 4, \ t > 1). \text{ The remaining sample points at each iteration } (|X| - \{X^t_R\} \text{ are used as test points to evaluate the performance of the tentative surrogate in each predefined subspace. For the two problems, the design space is divided into four equal rectangular subspaces (as illustrated in Fig. 3). The Kriging surrogate model is developed for each problem and the surrogate accuracy is evaluated using the REES method. To implement the Kriging method, the DACE (design and analysis of computer experiments) package developed by Lophaven et al.}^{24}\text{ is used. The bounds on the correlation parameters in the nonlinear optimization, } \theta_l \text{ and } \theta_u, \text{ are specified to be 0.1 and 20, respectively. The order of the global polynomial trend function is specified to be zero.}
\]

### IV. Results and Discussion

Before implementing the Regional Error Estimation of Surrogate (REES) method on the test problems, the actual performance of the final surrogate in each subspace is evaluated using Relative Accuracy Error (RAE) between surrogate predictions and actual values on larger set of test points. To this end, different sets of test points ranging from 100 to 15,000 points are studied. The optimal Latin Hypercube based on the max-min criterion\(^ {25}\) is adopted to determine the locations of test points for each set. The performance of the final surrogate on 100, 500, 1000, and 10,000 test points are illustrated in Figs. 4-7 for the Dixon & Price function, and in Figs. 8-11 for the Branin function. In this paper, RAES are classified into four levels based on the estimated error. The predefined lower and upper bounds of each level for the mean and the maximum errors are defined as

\[
E_i^{Mean} \in \begin{cases} 
\text{Level 1} & \text{if } E_i^{Mean} \leq 0.5 \mu^{Mean} \\
\text{Level 2} & \text{if } 0.5 \mu^{Mean} < E_i^{Mean} \leq \mu^{Mean} \\
\text{Level 3} & \text{if } \mu^{Mean} < E_i^{Mean} \leq 1.5 \mu^{Mean} \\
\text{Level 4} & \text{if } E_i^{Mean} > 1.5 \mu^{Mean} \\
\end{cases}
\]

where \( \mu^{Mean} = \text{Mean}(E_i^{Mean}), \text{ and } i=1,\ldots,n \)  

\[
E_i^{Max} \in \begin{cases} 
\text{Level 1} & \text{if } E_i^{Max} \leq 0.5 \mu^{Max} \\
\text{Level 2} & \text{if } 0.5 \mu^{Max} < E_i^{Max} \leq \mu^{Max} \\
\text{Level 3} & \text{if } \mu^{Max} < E_i^{Max} \leq 1.5 \mu^{Max} \\
\text{Level 4} & \text{if } E_i^{Max} > 1.5 \mu^{Max} \\
\end{cases}
\]

where \( \mu^{Max} = \text{Mean}(E_i^{Max}), \text{ and } i=1,\ldots,n \)
In Eqs. 11 and 12, $E_{\text{Mean}}^{i}$ and $E_{\text{Max}}^{i}$ are normalized values of the mean and maximum errors evaluated in the $i^{th}$ subspace, respectively. Interesting observations are made from Figs. 4-11: (i) The error level classifications change when the number of test points is increased from 100 to 10,000. We observe that surrogate performance evaluated on a specific number of test points (i.e., 100 or 200 test points) does not necessarily represent the actual performance of the surrogate. In these two test problems, the error level classification does not change extremely beyond 1000 test points. It should be noted that this number may vary depending on the complexity of the problems. (ii) According to the difference among the level of errors in each subspace of the design domain, global error measures might not provide adequate knowledge of the surrogate reliability in the process of the surrogate-based design.

![Figure 4. The error levels of the surrogate in different subspaces on 100 Test Points (Dixon & Price function)](image)

(a) Mean of the RAE in each subspace - $E_{\text{Mean}}$
(b) Maximum of the RAE in each subspace - $E_{\text{Max}}$

In REES method, the VESP regression model in each subspace is used to predict the level of the error in that subspace of the final surrogate. The trends of the VESP models for the Dixon & Price and the Branin functions are illustrated in Figs. 12 and 14, respectively. The VESP regression model is trained by the mean and the maximum errors evaluated for the intermediate surrogates. These intermediate surrogates are constructed and tested using the tentative training and test points defined in Table 1. Expectedly, the level of the error decrease with increasing the number of training points. It is observed that the RAE in certain iterations is relatively higher than that in preceding iteration (i.e., second iteration in 12(b)). The coefficients of the exponential regression function (Eq. 8) used in the VESP model for the two test problems are given in Tables 2 and 3.

The regional errors (maximum and mean) predicted based on the REES method in each subspace are...
Figure 6. The error levels of the surrogate in different subspaces on 1000 Test Points (Dixon & Price function)

Figure 7. The error levels of the surrogate in different subspaces on 10000 Test Points (Dixon & Price function)

Figure 8. The error levels of the surrogate in different subspaces on 100 Test Points (Branin function)
Figure 9. The error levels of the surrogate in different subspaces on 500 Test Points (Branin function)

Figure 10. The error levels of the surrogate in different subspaces on 1000 Test Points (Branin function)

Figure 11. The error levels of the surrogate in different subspaces on 10000 Test Points (Branin function)
Table 2. VESP coefficients in each subspace for Dixon & Price function

<table>
<thead>
<tr>
<th>Region</th>
<th>VESP for $\tilde{E}_{\text{Mean}}$</th>
<th>VESP for $\tilde{E}_{\text{Max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$a_0$ $a_1$</td>
<td>$a_0$ $a_1$</td>
</tr>
<tr>
<td>Subspace 1</td>
<td>2203.4515 -0.3050</td>
<td>5991.1503 -0.3087</td>
</tr>
<tr>
<td>Subspace 2</td>
<td>7.4808 -0.0041</td>
<td>17.9141 -0.0116</td>
</tr>
<tr>
<td>Subspace 3</td>
<td>126.8177 -0.1999</td>
<td>215.1334 -0.1960</td>
</tr>
<tr>
<td>Subspace 4</td>
<td>209.6296 -0.1048</td>
<td>516.8065 -0.0978</td>
</tr>
</tbody>
</table>

Table 3. VESP coefficients in each subspace for Branin function

<table>
<thead>
<tr>
<th>Region</th>
<th>VESP for $\tilde{E}_{\text{Mean}}$</th>
<th>VESP for $\tilde{E}_{\text{Max}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$a_0$ $a_1$</td>
<td>$a_0$ $a_1$</td>
</tr>
<tr>
<td>Subspace 1</td>
<td>2.3351 -0.0634</td>
<td>10.00 -0.0923</td>
</tr>
<tr>
<td>Subspace 2</td>
<td>0.2650 -0.0413</td>
<td>0.6625 -0.0627</td>
</tr>
<tr>
<td>Subspace 3</td>
<td>0.7180 0.0021</td>
<td>1.5777 -0.0276</td>
</tr>
<tr>
<td>Subspace 4</td>
<td>3.9213 -0.0057</td>
<td>16.6621 -0.0215</td>
</tr>
</tbody>
</table>

illustrated in Figs. 13 and 15. The predicted RAEs are classified into four levels defined in Eqs. 11 and 12. For the Dixon & Price function, the combination of the predicted performance (Fig. 13) with the actual performance evaluated by 10000 test points (Fig. 7) shows that the REES method successfully predicts the level of the error (mean and max) in each subspace. In Branin function, the comparison between the predicted performance and the actual performance evaluated by 10000 test points shows that: (i) the predicted maximum error levels (Fig. 15(b)) are identical to the actual maximum error levels (Fig. 11(b)); and (ii) there is slightly difference between the predicted and the actual levels of mean error (Fig. 15(a) and Fig. 11(a)). Overall, the REES method is capable of locating the region with the high level of error.

V. Conclusion

This paper developed a new methodology to quantify the surrogate error in the different regions of the design domain, which is called the Regional Error Estimation of Surrogate (REES) method. The REES method provides a model independent error measure that does not require any additional system evaluations. In the REES method, after segregating the design space into subspaces (or regions), Variation of the Error with Sample Points (VESP) regression models are constructed to predict the accuracy of the surrogate in each subspace. These regression models are trained by the errors (the mean and the maximum error) evaluated for the intermediate surrogates in an iterative process. At each iteration, the intermediate surrogate is constructed using different subsets of training points and tested over the remaining points. In this paper, the boundaries of the subspaces are predefined manually, and an exponential regression function is used to train the VESP model. Two numerical test problems are examined to evaluate the predicted performance of the REES measure. The preliminary results indicate that the REES measure is capable of evaluating the regional performance of a surrogate with reasonable accuracy.

Future work may include applying pattern classification methods to segregate the domain space at each iteration based on the evaluated error, instead of pre-specifying subspace boundaries.

VI. Acknowledgements

Support from the National Science Foundation Awards CMMI 0946765 and CMMI 1100948 is gratefully acknowledged. Any opinions, findings, and conclusions or recommendations expressed in this paper are those
Figure 12. The trend of VESP regression model in different subspaces (Dixon & Price function)

Figure 13. The predicted performance of the surrogate in different subspaces based on the REEM measure (Dixon & Price function)
Figure 14. The trend of VESP regression model in different subspaces (Branin function)

Figure 15. The predicted performance of the surrogate in different subspaces based on the REEM measure (Branin function)
of the authors and do not necessarily reflect the views of the NSF.

References


