

DETC2014-35358

**DRAFT: A FRAMEWORK FOR COLLABORATIVE SURROGATE MODEL
SELECTION (COSMOS)**

Souma Chowdhury*

Mississippi State University
Starkville, NY, 39759

Email: chowdhury@bagley.msstate.edu

Ali Mehmani†

Syracuse University
Syracuse, NY, 13244

Email: amehmani@syr.edu

Achille Messac‡

Mississippi State University
Syracuse, NY, 39762

Email: messac@bagley.msstate.edu

ABSTRACT

One of the primary drawbacks plaguing wider acceptance of surrogate models is their low fidelity (in general), which can be in a large part attributed to the lack of quantitative guidelines regarding the suitability of different models for diverse classes of problems. In this context, model selection techniques are immensely helpful in ensuring the selection and use of an optimal model for a particular design problem. A novel model selection technique was recently developed to perform optimal model search at three levels: (i) optimal model type (e.g., RBF), (ii) optimal kernel type (e.g., multiquadric), and (iii) optimal values of hyper-parameters (e.g., shape parameter) that are conventionally kept constant. The maximum and the median error measures to be minimized in this optimal model selection process are given by the REES error metrics, which have been shown to be significantly more accurate than typical cross-validation-based error metrics. Motivated by the promising results given by REES-based model selection, in this paper we develop a framework called Collaborative Surrogate Model Selection (COSMOS). The primary goal of COSMOS is to allow the selection and usage of globally competitive surrogate models. More specifically, this framework will offer an open online platform where users from within and beyond the Engineering Design (and MDO) community can submit

training data to identify best surrogates for their problem, as well as contribute new and advanced surrogate models to the pool of models in this framework. This first-of-its-kind global platform will facilitate sharing of ideas in the area of surrogate modeling, benchmarking of existing surrogates, validation of new surrogates, and identification of the right surrogate for the right problem. In developing this framework, this paper makes three important fundamental advancements to the original REES-based model selection - (i) The optimization approach is modified through binary coding to allow surrogates with differing numbers of candidate kernels and kernels with differing numbers of hyper-parameters. (ii) A robustness criterion, based on the variance of errors, is added to the existing criteria for model selection. (iii) Users are allowed to perform model selection for a specified region of the input space and not only for the entire input domain, subject to empirical constraints that are related to the relative sample strength of the region. The effectiveness of the COSMOS framework is demonstrated using a comprehensive pool of five major surrogate model-types (with up to five constitutive kernel types), which are tested on two standard test problems and an airfoil design problem.

Keywords: Error Estimation, Mixed-Integer Non-Linear Programming (MINLP), Model Selection, Optimization, Response Surface, Surrogate.

*Assistant Research Professor, Department of Mechanical Engineering, Center for Advanced Vehicular Systems, ASME Professional Member. Corresponding Author.

†Doctoral Candidate, Department of Mechanical and Aerospace Engineering, ASME Student Member.

‡Dean of Engineering, and Professor and Earnest W. and Mary Ann Deavenport Jr. Chair of Aerospace Engineering, Bagley College of Engineering, Mississippi State University. ASME Lifetime Fellow.

INTRODUCTION
Surrogate Model Selection

In the era of complex engineering systems, computational models play a central role in systems design and analysis. A

popular class of stochastic computational models are *surrogate models*, which are also known as *metamodels* [1] or *response surfaces* (depending on the research community). Surrogate models are purely mathematical models (i.e., not derived from the system physics) that are commonly used to provide a tractable and inexpensive approximation of the actual system behavior, as an alternative to expensive computational simulations (e.g., CFD) or to the lack of a physical model in the case of experiment-derived data (e.g., creation and testing of new metallic alloys). Primary applications of surrogate models include application in domain exploration, sensitivity analysis, uncertainty quantification, and optimization. Owing to the typical low-fidelity of surrogate models, the selection of a suitable model for a given experimental or simulation data set is critical to effective and reliable usage of surrogates in any of the above applications. While a suitable surrogate model can be selected intuitively (experience-based selection) based on an understanding of the data characteristics and/or the application constraints [2], the development of general guidelines might not be practical due to the diversity of system behavior among design applications. However, applicability of intuitive model selection will remain limited to scenarios where only a few candidate surrogates (which the user is well acquainted with) are available, thereby failing to exploit the large fast-growing pool of surrogates models. *Automated model selection* approaches could fill this crucial gap in the surrogate modeling paradigm.

Existing *automated model selection* techniques generally operate at one of the following levels:

1. Selecting a **model type**, e.g., Quadratic Response Surfaces (QRS) [3], Kriging [4,5], Radial Basis Functions (RBF) [6], Support Vector Regression(SVR) [7], Artificial Neural Networks [8], and Moving Least Square [9,10].
2. selecting a **kernel function** or a basis function, e.g., Linear, Gaussian, and Multiquadric functions.
3. optimizing the hyper-parameters (e.g., shape parameter in RBF); these parameters are the ones that are conventionally kept constant or prescribed by the user.

In the literature, error measures have been used to separately select model type and kernel functions. Popular error measures used for model type selection include [11]: (i) split sample, (ii) cross-validation, (iii) bootstrapping, (iv) Schwarz’s Bayesian information criterion (BIC) [12], and (v) Akaike’s information criterion (AIC) [13,12]. In addition to the model type and basis (or correlation) function selection, error measures can also be applied in hyper-parameter optimization to select the parameter that minimizes surrogate errors. This hyper-parameter optimization is highly sensitive to the basis functions and the data distribution [14]. Viana et al. [15] applied the cross-validation method to select the best predictor function and weights for different surrogates to construct a hybrid weighted surrogate. Martin and Simpson [16] used maximum likelihood estimation (MLE) and cross-validation methods to find the optimum hyper-parameter value for the Gaussian correlation function in Kriging. The likelihood

function in that case defines probability of observing the training data for a particular set of parameters. Gorissen et al. [17] provided the leave-one-out cross-validation and AIC error measures in the SURrogate MODELing (SUMO) Toolbox to automatically select the best model type for a given problem.

However, unified approaches to perform *automated surrogate model selection* simultaneously at all three levels (surrogate type, kernel type, and hyper-parameter value) is rare in the literature. Mehmani et al. [18] recently presented a new *automated surrogate model selection* technique called *REES-based model selection*, which operates at all the three levels, i.e., selecting model type, kernel type, and hyper-parameter value. This approach significantly mitigates the likelihood of sub-optimal selection, being only limited by the accuracy of the optimization algorithm (which is state of the art) and the underlying error measure (which is reportedly among the most reliable). The following section provides brief description of this model selection technique [18], and how this paper aims to advance both the fundamental and the practical scope of this technique.

REES-based Model Selection

The REES-based model selection uses the error measures given by the Regional Error Estimation in Surrogates (RESS) method, which has been shown to be up to two orders of magnitude more accurate than standard leave-one-out cross validation [19]. The REES method is derived from the hypothesis that the accuracy of approximation models is related to the amount of data resources leveraged to train the model. In REES, intermediate surrogates are iteratively constructed over heuristic subsets of training points (intermediate training points). The median and maximum errors estimated over the unused sample points (intermediate test points) at each iteration are used to fit distributions of the median and maximum error, respectively. The estimated statistical mode of the median and the maximum error are then represented as functions of the density of intermediate training points, with the aim to predict the median and maximum errors in the final surrogate. The REES-based model selection operates at all the three levels of selection as shown in Fig. 1 by using one of following techniques:

- (I) *Cascaded Technique*: In this technique, REES error measures are used to perform selection in a nested loop starting with “hyper-parameter optimization” at the lowest level and ending with “model-type selection” at the topmost level. Although easy to implement, this approach could become computationally expensive as the number of model-types and kernel/basis functions considered increases.
- (II) *One-Step Technique*: In this technique, a *mixed integer non-linear programming* (MINLP) problem is formulated and solved to minimize the surrogate error (derived from REES) in order to simultaneously select the best surrogate, the best constitutive kernel, and the optimal hyper-parameter value for that kernel. This unique approach, although more complex than the former, is expected to be significantly less ex-

pensive (computationally).

In this 3-level model selection, the selection criteria could depend on the class of applications. For example, in standard surrogate-based analysis the lower median error could be treated as the selection criteria, whereas in structural optimization applications that demands conservative surrogates, the lower maximum error could be treated as the selection criteria. Depending on the problem and the available data set, the median and maximum errors might be mutually conflicting or mutually promoting, which would then respectively lead to a set of Pareto models or a single optimum model. In the original *REES-based model selection* [19], the predicted modal values of the median and maximum error in the surrogate are simultaneously minimized, using a multi-objective optimization algorithm (NSGA-II [20]). Surrogate types and constitutive kernel types were treated as two integer variables, and hyper-parameters were treated as continuous variables. In the original implementation, kernels with the same number of hyper-parameters and surrogates with the same number of constitutive kernels were only considered, so as to avoid candidate solutions with differing numbers of variables.

The vision behind the *REES-based model selection* was to enable the identification of globally best surrogate models for any given application – a technique that is neither limited to any particular class of models nor makes a limiting assumption regarding the surrogate/kernel type and the value of hyper-parameters. In doing so, this model selection technique is also expected to create unique opportunities for benchmarking of existing surrogates, validation of new surrogates, and identification of the right surrogate for the right problem. With this vision, this paper seeks to develop an open framework called *Collaborative Surrogate Model Selection* (COSMOS). This framework will offer an open online platform where users from within and beyond the Engineering Design (and MDO) community can submit training data to identify best surrogates for their application, as well as contribute new and advanced surrogate models to the pool of models in COSMOS - thereby offering an opportunity to the design community to collaboratively advance the state of the art in surrogate model development and usage.

The specific objectives of this paper are:

1. To make fundamental modifications to the original *REES-based model selection* method to allow consideration of additional selection criteria such as (i) the variance of the surrogate error and (ii) the predicted error at a greater number of sample points than yet available.
2. To modify the optimization approach in order to allow the competition of surrogates with differing numbers of candidate kernels and kernels with differing numbers of hyper-parameters, without having to implement a full cascaded approach (which could be prohibitively expensive).
3. To test the COSMOS framework with a sufficiently comprehensive set of surrogate types and constitutive kernel types (a total of 16 surrogate-kernel combinations with 0 to 2 hyper-parameters).

The development of the COSMOS framework is further described in the following section. In Section , numerical experiments, including standard test problems and an engineering problem, are provided to illustrate the potential of COSMOS.

A corollary objective of this paper is to develop the *open online platform* (with a visual interface) for COSMOS, which will allow users to both use COSMOS for their own data set and/or contribute new surrogates and kernels. The construction of this online platform is however ongoing, and is expected to be completed within the next month (well ahead of the actual ASME IDETC 2014 conference). Further details of the online platform (although not a central component of this paper), will be included in the final draft of this paper.

COLLABORATIVE SURROGATE MODEL SELECTION (COSMOS)

The COSMOS Framework

The major components and the flow of information in the COSMOS framework is illustrated in Fig. 2

COSMOS provides a framework for automated selection of surrogate models from a candidate pool of surrogate models, which involves simultaneous determination of (i) optimal surrogate type(s), (ii) optimal kernel type(s), and (iii) optimal hyper-parameter value(s). The available pool of surrogates is the leftmost block in Fig. 2. Each type of surrogate model offers one or more kernel/basis/correlation function choices, e.g., under the RBF surrogate, linear, cubic, Gaussian, or multiquadric basis functions could be selected. Some of these kernel or basis functions include hyper-parameters which are conventionally kept fixed at recommended values (e.g., the shape parameter in multiquadric RBF). In COSMOS, these hyper-parameters are however determined through an optimization process, and hence the candidate pool provides a guideline range for each one these hyper-parameters. The candidate pool used in the current version of COSMOS is summarized in Table 1.

COSMOS offers five different criteria for selection of optimal surrogates, as represented by the leftmost block in Fig. 2. Any combination of these criteria can be chosen by a user. In the current implementation, three of the most practically-pertinent pairwise combinations (1-2, 1-3, and 1-5) are allowed to be selected – using more than two criteria could prove to be taxing on the complex *optimal model selection process*. Further description of how the selection criteria are estimated using the REES method (with certain modifications) is provided in the following section.

The central component of the COSMOS framework is the “Optimal Model Selector”. This process accepts the user-specified selection criteria and the training data (also given by the user) to search for the optimal surrogate models from the available candidate pool. Owing to the differing numbers of hyper-parameters in different kernels, optimal model selection is performed through multiple parallel optimizations (different from cascaded technique), which is further described in the following

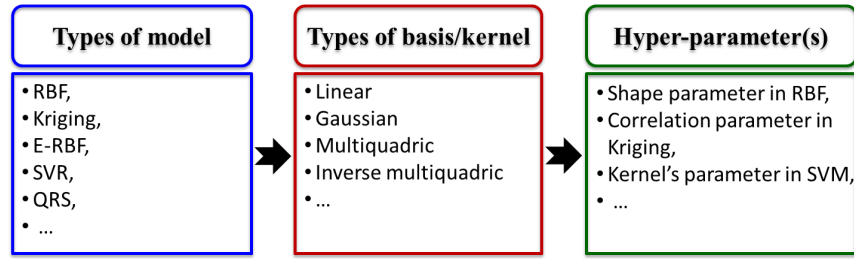


FIGURE 1. Elements of 3-level surrogate model selection

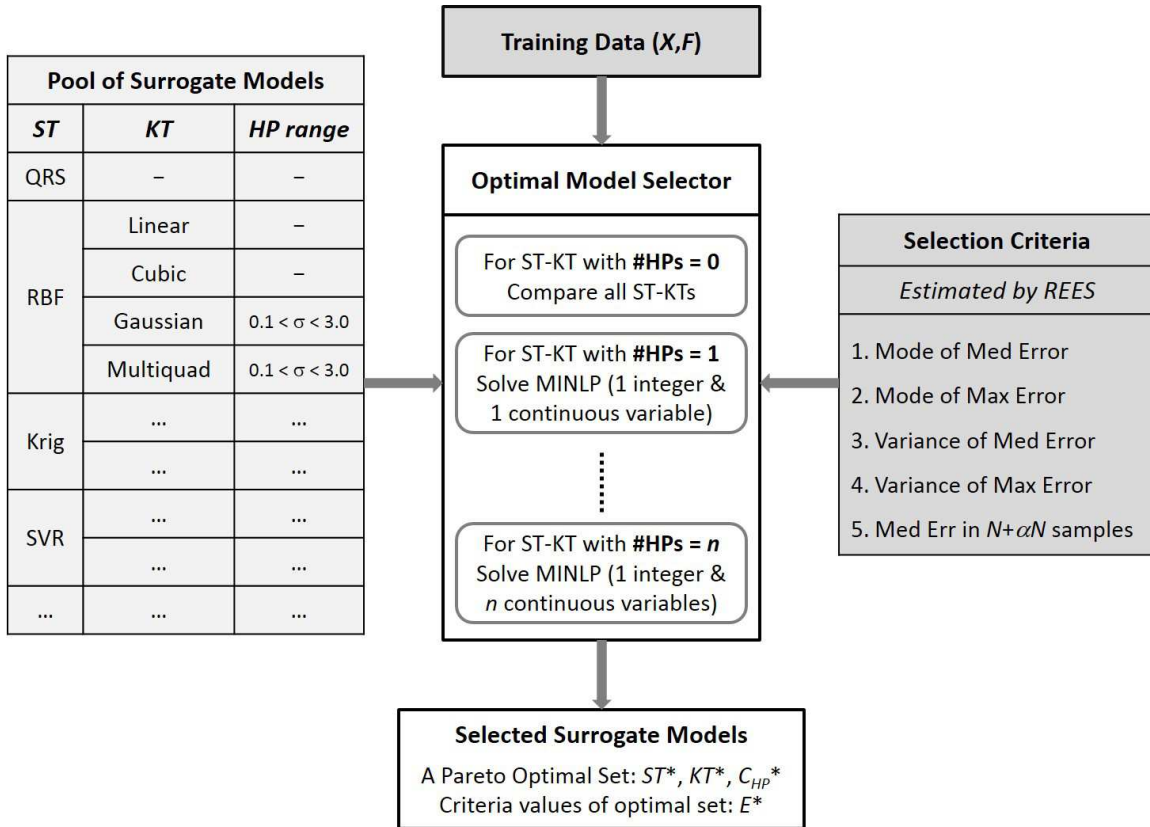


FIGURE 2. The COSMOS framework (ST: Surrogate Type; KT: Kernel Type; HP: Hyper-parameter; C_{HP^*} : vector of HP values; $\alpha \in (0, 1)$)

Section. The *optimal model selector* produces a set of Pareto optimal surrogate models (with optimal surrogate-kernel types and optimal hyper-parameter values), and also provides the error distribution of the optimal surrogates.

REES-based Criteria Functions for Model Selection

As evident from the COSMOS framework in Fig. 2, a comprehensive set of five different selection criteria can be specified (in any combination) to drive the surrogate model selection. The first two criteria respectively represent the (predicted) most probable values of the median error and the maximum error in the surrogate – which stochastically accounts for the sensitivity of the surrogate accuracy to choice (number and location) of train-

ing points. Advancements are made to REES in this paper to formulate and use three new selection criteria. Criteria 3 and 4 in Fig. 2, which are two new selection criteria, respectively represent the variance in the median and the maximum error of the surrogate model – these two measures reflect the robustness of the surrogate. Together, criteria 1-4 provide an understanding of the uncertainty in the surrogate. Another new selection criteria formulated in this paper is the (expected) most probable value of the median error on a hypothetically-larger sample set, assuming both the actual and the hypothetically-larger sets of sample points follow the same overall distribution.

The REES method previously developed by the same authors [19] is applied to evaluate all the five selection criteria for any given candidate surrogate. The inputs and outputs of the

REES method can be expressed as:

$$[E_{med}^{mo}, E_{max}^{mo}, E_{med}^{\sigma}, E_{max}^{\sigma}, E_{med,\alpha}^{mo}] = f_{REES}(T_S, T_K, C_{HP}, X, F) \quad (1)$$

where T_S , T_K , and C_{HP} respectively represent the surrogate type, the kernel type, and the vector of hyper-parameter values for that surrogate-kernel combination; the vectors X and F represent the sample data (input and output) used for training the surrogate. In Eq. 1, the terms E_{med}^{mo} and E_{max}^{mo} respectively represent the predicted modal values of the median and maximum errors in the surrogate; the terms E_{med}^{σ} and E_{max}^{σ} respectively represent the standard deviation values of the median and maximum errors in the surrogate; and the term $E_{med,\alpha}^{mo}$ represents the expected modal value of the median error if the same surrogate (i.e., same T_S , T_K , and C_{HP}) was constructed over α -times greater number of sample points (assuming the overall distribution is similar to that of X). This error measure, $E_{med,\alpha}^{mo}$, indicates the expected capability of candidate surrogates if a larger set of sample data could be made available, thereby helping users understand the cost-to-benefit ratio of performing additional high-fidelity simulations or experiments in this context.

The REES method predicts the surrogate error by modeling the variation of the error with increasing density of training points, as described in the following pseudocode.

In this method, for a set of N sample points intermediate surrogates are constructed at each iteration, t , using M^t heuristic subsets of n^t training points (intermediate training points). These intermediate surrogates are then tested over the corresponding remaining $N - n^t$ points (intermediate test points). The median error is then estimated for each of the M^t intermediate surrogates at that iteration, and a parametric probability distribution is fitted to yield the modal value ($E_{med}^{mo,t}$) and the standard deviation ($E_{med}^{\sigma,t}$) of the median error at that iteration. The smart use of the modal value of median error promotes a monotonic variation of error with sample point density [19], unlike mean or root mean squared error which are highly susceptible to outliers. This approach gives REES an important advantage over conventional cross-validation-based error measures as illustrated by Mehmani et al. [19]. A similar approach is used to estimate the modal value ($E_{max}^{mo,t}$) and the standard deviation ($E_{max}^{\sigma,t}$) of the maximum error, at any t^{th} iteration. It is important to note that all error quantifications are done in terms of the relative absolute error which is given by:

$$E_{RAE} = \left| \frac{F_A(X_i) - F_S(X_i)}{F_A(X_i)} \right| \quad (2)$$

where F_A and F_S are respectively the actual function value at X_i (given by high fidelity simulation of data) and the function value estimated by the surrogate; in this definition, a denominator of 1 is used instead of F_A , if $F_A = 0$.

In the original REES method, distribution functions to be fitted over the median and maximum errors at each iteration

Algorithm 1 Regional Error Estimation of Surrogate (REES)

INPUT Number of sample points N , sample set (X, F) , and α
Set Number of iterations N^u ; indexed by t
Set Size of intermediate training points at each iteration, n^t , where $n^t < n^{t+1}$
Set Number of heuristic subsets of training points at each iteration equal to M^t , where $M^t \leq \binom{N}{n^t}$; indexed by k

```

X=Experimental Design(N)
F | X = Evaluate System (X)
{X} = {(Xi, Fi)i=1N}
for t = 1, ..., Nu do
  for k = 1, ..., Mt do
    Choose {β} ⊂ {X}, where # {β} = nt
    Define intermediate training points, {XTR} = {β}
    Define intermediate test points, {XTE} = {X} - {β}
    Construct intermediate surrogate Sk using {XTR}
    Estimate median and maximum errors;
    Emedk,t = median(em)m=1, ..., # {XTE}
    Emaxk,t = max(em)m=1, ..., # {XTE}
  end for
  Fit distributions of the median error over all Mt heuristic subsets
  Determine the mode of the median and maximum error distributions: Emedmo,t and Emaxmo,t
  Determine the standard deviation of the median and maximum error distributions: Emedσ,t and Emaxσ,t
end for
Construct the final surrogate using all N sample points
Use the estimated Emedmo,t, Emaxmo,t, Emedσ,t and Emaxσ,t ∀t < Nu, to quantify their variation with # training points (nt), using regression functions.
return The modal values of the median and maximum errors in the final surrogate: Emedmo and Emaxmo
return The standard deviation of the median and maximum errors in the final surrogate: Emedσ and Emaxσ
return The modal value of the median errors in the hypothetical surrogate with (N + αN) sample points: Emed,αmo

```

were selected using the chi-square goodness-of-fit criterion [21] among the following candidates: lognormal, Gamma, Weibull, logistic, log logistic, t location scale, inverse gaussian, and generalized extreme value distribution. However, in this paper, to reduce the computational expense of REES (which is a function evaluator in the MINLP process), we only use the lognormal distribution, which has been observed to be the most effective from earlier numerical experiments. The pdf's of the median and the maximum errors, p_{med} and p_{max} , can thus be expressed as

$$p_{med} = \frac{1}{E_{med} \sigma_{med} \sqrt{2\pi}} \exp\left(-\frac{(\ln(E_{med}) - \mu_{med})^2}{2\sigma_{med}^2}\right) \quad (3)$$

$$p_{max} = \frac{1}{E_{max}\sigma_{max}\sqrt{2\pi}} \exp\left(-\frac{(\ln(E_{max}) - \mu_{max})^2}{2\sigma_{max}^2}\right) \quad (4)$$

In the above equations, E_{med} and E_{max} respectively represent the median and the maximum errors estimated over a heuristic subset of training points at any given iteration in REES. The parameters, $(\mu_{med}, \sigma_{med})$ and $(\mu_{max}, \sigma_{max})$ represent the generic parameters of the lognormal distribution (which are respectively the mean and the standard deviation of the corresponding variables' natural logarithm).

The modal values of the median and the maximum error at any iteration can then be expressed as

$$E_{med}^{mo} = \exp(\mu_{med} - \sigma_{med}^2) \quad (5)$$

$$E_{max}^{mo} = \exp(\mu_{max} - \sigma_{max}^2) \quad (6)$$

The standard deviation of the median and the maximum error at any iteration can be expressed as

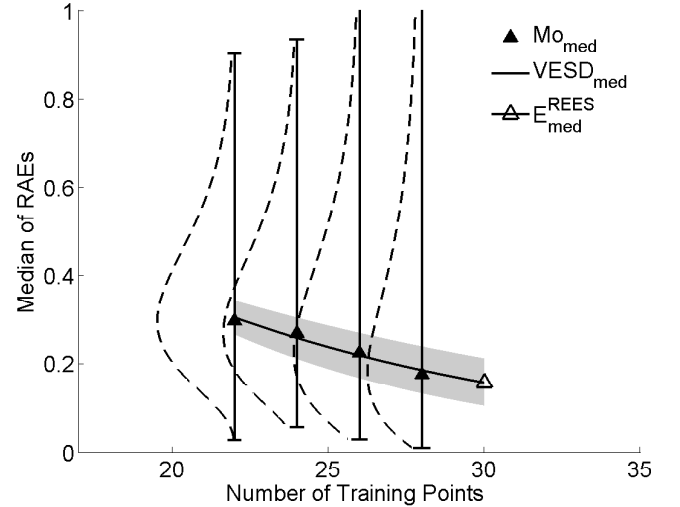
$$E_{med}^{\sigma} = \sqrt{(\exp(\sigma_{med}^2) - 1) \exp(2\mu_{med} + \sigma_{med}^2)} \quad (7)$$

$$E_{max}^{\sigma} = \sqrt{(\exp(\sigma_{max}^2) - 1) \exp(2\mu_{max} + \sigma_{max}^2)} \quad (8)$$

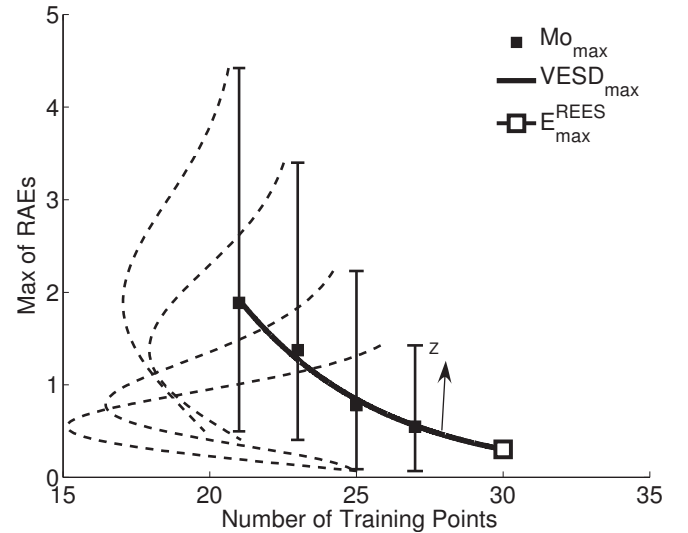
The above four equations represent the modal values and the standard deviation of the median and maximum errors at any iteration t ; the subscript, t , is however not included in these equations for ease of representation.

The variation of the modal values of the median and maximum errors (Eqs. 5 and 6) with sample density is modeled using the multiplicative ($E = a_0 n^{a_1}$) or the exponential ($E = a_0 e^{a_1 n}$) regression functions. Alternatively, if the type of median/maximum error distribution is fixed (e.g., lognormal distribution with parameters μ and σ), standard regression functions could be used to model the variation of the parameters of the distribution as functions of the number of training points. The choice of these regression functions assume a *smooth monotonic decrease of the error values with the training point density*. Since the standard deviation in the errors could both increase or decrease with increasing number of sample points (for practical sample sizes), other regression functions (e.g., linear and polynomial functions) are allowed to be used to fit the variation of E_{med}^{σ} and E_{max}^{σ} . The root mean squared error metric is used to select the best-fit regression model in each of these cases. These regression functions are then used to predict the modal values and the standard deviation of the median and maximum errors in the final surrogate (that is trained using all N sample points).

Illustration of how the REES method works is provided in Figs. 3 to 5, which are generated by applying Kriging to the 2-variable Branin Hoo function. The solid symbols in these figures represent the estimated median and maximum errors at the sets of intermediate training points (in each iteration), and the hollow symbols at the 30th iteration represent the predicted error in the final surrogate which uses all the available training points. These predicted error values are used as criteria functions in COSMOS. The REES error measures promote a monotonic decrease of the



(a) Variation of the mode of median errors with increasing sample points



(b) Variation of the mode of maximum errors with increasing sample points

FIGURE 3. Application of REES using Kriging to model the Branin Hoo function

median and maximum errors with sample density (as observed in Figs. 3 to 5). Since the rate of error decrease (given by the

regression functions) is likely different for different surrogates, it is possible that the order of dominance among candidate surrogates may change if a higher number of sample points were made available – this scenario is uniquely accounted for by the 5th selection criteria in COSMOS. Since the modal value of the median error at $N + \alpha N$ sample points (the 5th selection criteria) is predicted by further extrapolation of the regression function, the specified value of α should be small (lower than 0.3 or 30%).

The Optimal Model Selector

In the original REES-based model selection method, optimal model selection was performed using two approaches, the Cascaded technique and the One-step technique [18]. The Cascaded technique, although easy to implement, becomes computationally expensive when a large pool of surrogate candidates is considered (which is indeed the objective of REES). The One-step technique which performs selection by solving a MINLP problem, is computationally more efficient and novel in implementation. However, the One-step technique cannot be readily applied if the pool of candidate surrogate models include surrogates with different number of available kernel choices, and kernels comprising differing numbers of hyper-parameters. To address these limitations and allow a flexible global selection of surrogate models, a *hybrid optimal model selection* technique is developed in COSMOS.

Since most surrogate kernel functions involve three or fewer hyper-parameters (e.g., shape parameter in multiquadric RBF), the global variation in the number of hyper-parameters can be assumed to be small. Therefore, the global pool of surrogate-kernel candidates is divided into multiple smaller pool of surrogate-kernel candidates based on the number of hyper-parameters in them. In other words, the candidate pool- I , represented by $HP - I$, comprises all those surrogate-kernel combinations which include I hyper-parameters (that can be tuned). Subsequently, optimal model selection is performed separately (in parallel) for each candidate pool, $HP - I$. Surrogate-kernel combinations with 0-2 hyper-parameters are considered in this paper. Each surrogate-kernel combination/candidate (associated with a particular $HP - I$) is then assigned a single integer code, as opposed to two separate integer codes given in the original REES-based model selection. The candidate surrogate-kernels considered in this paper are listed in Table 1, where the integer code assigned to each candidate is shown under their respective $HP - I$ class.

For the $HP - 0$ class of surrogate-kernel combinations, REES is applied to all the candidates, followed by the application of a Pareto filter to determine the set of non-dominated/optimal surrogate-kernel combinations. For all other $HP - I$ classes of surrogate-kernel combinations (with $i > 0$), a mixed integer non-linear programming (MINLP) problem is formulated to simultaneously perform model selection at all three levels. This MINLP

for the I^{th} pool can be expressed as:

$$\text{Min } f(T_{SK}, C_{HP}) = [E_{med}^{mo}, E_{max}^{mo}, E_{med}^{\sigma}, E_{max}^{\sigma}, E_{med,\alpha}^{mo}]$$

where

$$T_{SK} = [1, 2, \dots, N_{SK}^I]$$

$$0 \leq C_{HP} \leq 1, C_{HP} \in R^I$$

(9)

In Eq. 9, T_{SK} is the integer variable that denotes the combined surrogate-kernel type; C_{HP} is the vector of continuous variables that represent the hyper-parameter values; and N_{SK}^I represents the total number of candidate surrogates-kernel types available under the class $HP - I$. It should be noted that a range of (0,1) is used for each hyper-parameter variable in the optimization process to avoid disparity. In order to conform to this uniform range across all hyper-parameters, each hyper-parameter is scaled using a user-specified upper and lower bound (listed in Table 2).

Although, all five criteria functions (to be minimized) are simultaneously included in this optimization problem definition, in practice it is recommended to pursue bi-objective optimization – i.e., at a time considering any two objectives/criteria of interest. Depending on the data set, the considered criteria functions could be mutually promoting or conflicting. The former yields a single optimum surrogate model and the latter yields a set of Pareto optimal surrogate models. Once the optimum surrogate models for each $HP - I$ have been obtained, a Pareto filter is applied to determine the globally optimal set of surrogate models. The criteria functions in this MINLP problem could often be multimodal, since the integer codes of the surrogate-kernel types do not necessarily follow the order of error in these candidate types. Hence, *heuristic multi-objective optimization* algorithms that are capable of dealing with mixed-integer variables are considered to be the most appropriate for solving these MINLP problems. In this paper, the binary Non-Dominated Sorting Genetic Algorithm II (NSGA-II) [20] is used to solve the multi-objective MINLP problems.

NUMERICAL EXPERIMENTS WITH COSMOS

Pool of Surrogate Candidates

In this paper, we have considered the participation of five major types of surrogate models: (i) Quadratic Response Surface (QRS), (ii) Radial Basis Functions (RBF), (iii) Kriging, (iv) Support Vector Regression (SVR), and (iii) Extended-Radial Basis Functions (E-RBF). The different forms of the kernel/basis/correlation functions currently considered in COSMOS are given in Table 4 in Appendix B. It could be said that COSMOS is one of the comprehensive model selection framework (in the literature) both in methodology and implementation (i.e., the pool of candidate surrogate-kernels considered). *Once the open online interface is created for COSMOS, it is expected to*

TABLE 1. Candidate surrogate-kernel combinations and their integer-codes

Surrogate	Kernel	HP-0	HP-1	HP-2	Hyper-parameter
Quadratic Response Surface	-	1	-	-	-
	Linear	2	-	-	-
	Cubic	3	-	-	-
Radial Basis Functions	TPS	4	-	-	-
	Gaussian	-	1	-	shape parameter, σ
	Multiquadric	-	2	-	shape parameter
	Linear	-	3	-	correlation parameter, θ
Kriging	Exponential	-	4	-	correlation parameter, θ
	Gaussian	-	5	-	correlation parameter, θ
	Cubic	-	6	-	correlation parameter, θ
	Spherical	-	7	-	correlation parameter, θ
	Linear	-	X	-	-
Support Vector Regression	RBF	-	8	-	parameter γ
	Polynomial	-	-	X	-
	Sigmoid	-	-	1	parameters γ and r
	Multiquadric	-	-	2	parameters λ and σ

TABLE 2. Range of hyper-parameters

Surrogate candidate	Hyper-parameter	Lower bound	Upper bound
RBF	shape parameter, σ	0.1	3
Kriging	correlation parameter, θ	0.1	20
SVR	kernel parameter, γ	0.1	8
SVR (Sigmoid)	kernel parameter, r	X	X
E-RBF	non-radial parameter, λ	X	X

provide a first-of-its-kind one-stop platform for surrogate model construction, selection, and benchmarking. The candidate surrogate models in COSMOS are briefly described below.

Quadratic Response Surfaces: QRS is one of the oldest and most popular type of surrogate model. The generic expression for a n -variable QRS is given by:

$$f(x) = c_0 + \sum_i^n c_i x_i + \sum_i^n c_{ii} x_i^2 + \sum_{\forall i \neq j} c_{ij} x_i x_j \quad (10)$$

where c_0 , c_i , c_{ii} , and c_{ij} are arbitrary coefficients that can be determined by the least squares approach.

Radial Basis Functions: The idea of using Radial Basis Functions (RBF) as interpolating functions was introduced by Hardy [6]. The RBF approximation is a linear combination of basis functions (Ψ) across all sample points, given by

$$\tilde{f}(x) = W^T \Psi = \sum_{i=1}^{n_p} w_i \psi(\|x - x^i\|) \quad (11)$$

where n_p denotes the number of selected sample points; w_i are the weights evaluated using the pseudo inverse method on sample points' responses; ψ is basis function expressed in terms of the Euclidean distance, $r = \|x - x^i\|$, of a point x from a given sample point, x^i . The 5 different types of basis functions for RBF

considered in the paper are listed in Table 1, where σ represents the shape parameter of the basis function. The shape parameter, $\sigma = 0$, corresponds to a constant basis function [22].

Kriging: Kriging [23, 24, 25] is a stochastic modeling approach to approximate irregular data. To implement Kriging in COSMOS, the DACE (design and analysis of computer experiments) package developed by Lophaven et al. [26] is used with a modification (related to the estimation of the correlation function parameter). The Kriging approximation function consists of two components: (i) a global trend function, and (ii) a deviation function representing the departure from the trend function. The trend function is generally a polynomial (e.g., constant, linear, or quadratic). The general form of the kriging surrogate model is given by [27]:

$$\begin{aligned} \tilde{f}(x) &= F(x, \varphi) + Z(x) \\ \text{where } F &= f(x)^T \varphi \end{aligned} \quad (12)$$

and where $\tilde{f}(x)$ is the unknown function of interest; $Z(x)$ is the realization of a stochastic process with the mean equal to zero and a nonzero covariance; F is the known approximation (usually polynomial) function; and φ is the regression parameter matrix. The (i, j) th element of the covariance matrix of $Z(x)$ is expressed as

$$\text{COV}[Z(x^i), Z(x^j)] = \sigma_z^2 R_{ij} \quad (13)$$

where R_{ij} is the correlation function between the i^{th} and the j^{th} data points, and σ_z^2 is the process variance, which scales the spatial correlation function. The popular forms of the correlation functions are listed in Table 1. The correlation function controls the smoothness of the Kriging predicted values based on the influence of other nearby points on a test point.

In Kriging, the regression function coefficients, the process variance, and the correlation function parameters ($\varphi, \sigma_z^2, \theta$) could be predefined or estimated using parameter estimation methods. The parameter, θ , in the correlation function is however treated as a hyper-parameter in COSMOS, and estimated through the solution of the MINLP problem (9). The regression function coefficients (φ) and the process variance (σ_z^2) are still estimated using MLE. In the current implementation, only the zeroth-order regression function is considered (for ease of computation); however first-order (linear) and second-order (quadratic) regression functions could be readily enabled without requiring any change in the formulation – it will basically increase the total number of candidates (for the MINLP) under the Kriging surrogate model.

Support Vector Regression: SVR is a relatively newer type of surrogate model. For a given training set of instance-label pairs (x_i, y_i) , $i = 1, \dots, n_p$, where $x_i \in R^n$ and $y \in 1, -1^m$, a linear SVR is defined by $f(x) = \langle w, x \rangle + b$, where b is a bias. To train the SVR, the error, $|\xi| = |y - f(x)|$, is minimized by solving

the following convex optimization problem:

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 + C \sum_{i=1}^{n_p} \xi_i + \tilde{\xi}_i \\ \text{s.t.} \quad & (w^T x_i + b) - y_i \leq \varepsilon + \xi_i \\ & y_i - (w^T x_i + b) \leq \varepsilon + \tilde{\xi}_i \\ & \xi_i, \tilde{\xi}_i \geq 0, \quad i = 1, 2, \dots, n_p \end{aligned} \quad (14)$$

In Eq.14, $\varepsilon \geq 0$ represents the difference between the actual and the predicted values; ξ_i and $\tilde{\xi}_i$ are the slack variables; C represents the flatness of the function; n_p represents number of training points. By applying kernel functions, $K(\alpha, \beta) = \langle \phi(\alpha), \phi(\beta) \rangle$, under KKT conditions, the original problem is mapped into higher dimension space. The dual form of SVR on nonlinear regression could be represented as

$$\begin{aligned} \max \quad & \sum_{i=1}^{n_p} \alpha_i y_i - \varepsilon \sum_{i=1}^{n_p} |\alpha_i| - \frac{1}{2} \sum_{i,j=1}^{n_p} \alpha_i \cdot \alpha_j \langle \phi(x_i), \phi(x_j) \rangle \\ \text{s.t.} \quad & \sum_{i=1}^{n_p} \alpha_i = 0, \quad -C \leq \alpha_i \leq C \text{ for } i = 1, \dots, n_p \end{aligned} \quad (15)$$

The standard Kernel functions used in SVR are listed in Table 4. The performance of SVR depends on its penalty parameter C and kernel parameters γ, r , and d . Using hyper parameter optimization, correlation parameters can be estimated such that it minimizes the model error. To implement SVR in this paper, the LIBSVM (A Library for Support Vector Machines) package developed [28] is used.

Applications

Benchmark Problems: The effectiveness of COSMOS framework is illustrated by performing model selection for the following two analytical test functions: (i) the 2-variable Branin Hoo function, and (ii) the 6-variable Hartmann function. These two test functions are described in Appendix A. Due to time constraints, only the results of these two test functions are provided in the draft paper. The full paper and subsequent presentation will likely include a more comprehensive set of test functions, e.g., the Dixon and Price function, the Shubert function, Goldstein and Price function, and the Michalewicz function.

Airfoil Design Problem: In this problem, the lift-to-drag ratio (C_L/C_D) of a Wortmann FX60.126 series airfoil is maximized [29]. The design variables include the angle of incidence (ranging from 0 to 10) and three shape variables (each ranging from -0.01 to 0.01). The three shape variables represent the distances between the middle of suction side and horizontal axis, between the middle of pressure side and horizontal axis, and between the trailing edge and horizontal axis. These three shape variables allow a modification of the un-deformed airfoil profile.

With respect to the initial airfoil design, two splines (e.g., cubic spline interpolation) are added to the suction and pressure sides. Each of these splines is characterized by 3 points defined on the leading edge, the middle span, and the trailing edge. To develop the high fidelity aerodynamics model of this airfoil, the commercial Finite Volume Method tool, FLUENT, is used, which solves the Reynolds-averaged Navier-Stokes (RANS) equations. The chord length of the airfoil is specified to be 1.0m, and incoming wind velocity is specified to be 25.0m/s. The structured CFD mesh is constructed using 9,838 quadrangular cells and 10,322 node points.

Results and Discussion

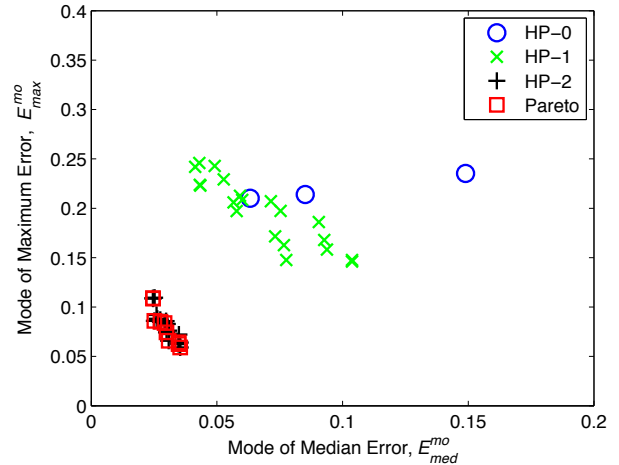
Three different model selections tests are performed using COSMOS, each involving a unique pair of selection criteria, as listed below:

- Test 1** Apply COSMOS to (i) minimize the modal value of the median error E_{med}^{mo} , and (ii) minimize the modal value of the maximum error E_{max}^{mo} ;
- Test 2** Apply COSMOS to (i) minimize the modal value of the median error E_{med}^{mo} , and (ii) minimize the standard deviation of the median error E_{max}^{σ}
- Test 3** Apply COSMOS to (i) minimize the modal value of the median error E_{med}^{mo} , and (ii) minimize the expected modal value of the median error at ($\alpha = 0.2$) more number of sample points ($E_{med,\alpha}^{mo}$)

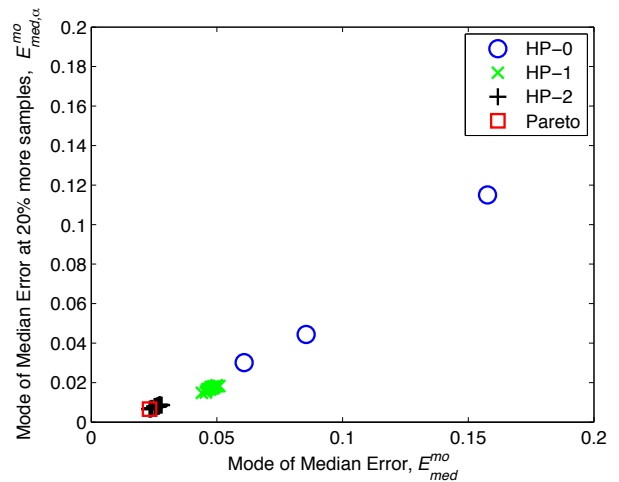
In each one of these tests, a population of 20 candidate solutions is used for the Brannin Hoo and airfoil design problems, and a population of 40 candidates is used for the relatively higher dimensional Hartman problem. For the $HP - 1$ and $HP - 2$ classes of surrogates, the optimal model selection process (using NSGA-II) is allowed to run for a maximum number of 50 generations. The test results are discussed below.

The surrogate models in the Pareto Optimal set for each test are listed in Table 3. The criteria functions for the Pareto solutions as well as the solutions in the final population (in $HP - 1$ and $HP - 2$ classes) for the three problems are shown in Figs. 4 to 6. Each figure shows the results from a particular test for a particular problem. In each of these figures the blue 'O', the green 'X', and the black '+' symbols respectively represent the final results of model selection under $HP - 0$, $HP - 1$, and $HP - 2$ classes (which represent the number of hyper-parameters involved). The Pareto solutions in these figures are encircled by red square symbols.

The test-2 for Branin Hoo function crashed, and hence has not been included in the draft paper (will be duely included in the final paper). For this problem, E-RBF Multiquadric (under class $HP - 2$) was selected as the best surrogate model by COSMOS in both Tests 1 and 3 (Table 3 and Fig. 4). The set of Pareto models for Test 1 in this case corresponded to different values of the hyper-parameters (HP) (λ and σ) in the E-RBF Multiquadric model (Fig.). Test 3 in this case yielded a single optimal E-RBF Multiquadric model, with $\lambda = 0.0232$ and $\sigma = .0066$ (Fig.).



(a) Trade-offs between modal values of median and maximum error



(b) Trade-offs between modal values of median error at N and $1.2N$ samples

FIGURE 4. Branin Hoo function: Pareto models (red sq.) and final population of models from all HP classes

With respect to the median error criteria, in this case surrogate-kernels with no HPs performed the worst and those with two HPs performed the best (Figs. and).

For the Hartmann function, SVR-RBF (with different HP values) constituted the set of Pareto models in both Tests 1 and 3 (Table 3, Figs. and); and Kriging-Spherical, SVR-RBF, and SVR-Sigmoid constituted the set of Pareto models in Test 2 (Table 3 and Fig.). In this test problem, surrogate-kernels with one one HP performed the best and those with no HP performed the worst.

For the airfoil design problem, RBF-TPS and ERBFMultiquadric constituted the set of Pareto models in Test 1 (Table 3 and Fig.); RBF-Cubic, RBF-TPS, and ERBF-Multiquadric constituted the set of Pareto models in Test 2 (Table 3 and Fig.); and E-RBF Multiquadric was selected as the single optimum surrogate model in Test 3 (Table 3 and Fig.). The relative accuracy

TABLE 3. The set of Pareto Surrogate Models given by COSMOS

Problem	Test 1, $\min[E_{med}^{mo}, E_{max}^{mo}]$	Test 2, $\min[E_{med}^{mo}, E_{med}^{\sigma}]$	Test 3, $\min[E_{med}^{mo}, E_{med,\alpha}^{mo}]$
Branin Hoo	ERBF-Multiquadric	-	ERBF-Multiquadric
Hartmann-6	SVR-RBF	Kriging-Spherical, SVR-RBF, SVR-Sigmoid	SVR-RBF
Airfoil Design	RBF-TPS, ERBF-Multiquadric	RBF-Cubic, RBF-TPS, ERBF-Multiquadric	ERBF-Multiquadric

of the optimal surrogates models in the airfoil design problem is observed to be significantly higher than that in the other to analytical test problems.

Ideally, for any given problem, the mode of median error estimated for a particular $HP - 0$ class surrogate should remain the same across the three tests, since they do have HPs that can be optimized. However, minor variations in their mode of median error are observed among different tests in all three problems (Figs. 4 to 6). This deviation can be attributed to the sensitivity of the REES error measures themselves to the surrogate robustness, unlike leave-one-out cross validation which is deterministic for a given set of sample points. This is because each time REES is applied, different heuristic subsets of intermediate training points could be chosen given the same surrogate and the same overall set of sample points. Future research in surrogate error measures should pursue the mitigation of this undesirable sensitivity.

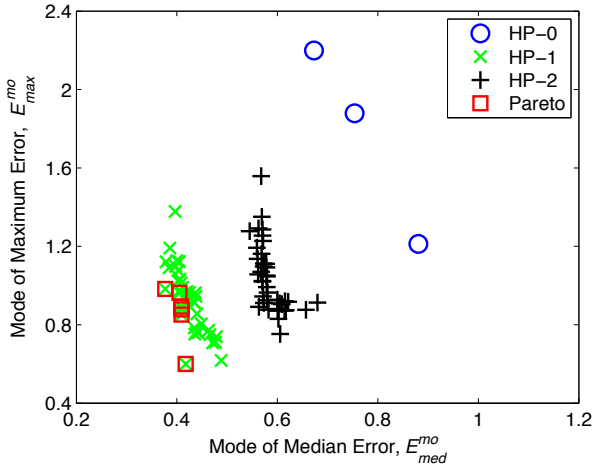
Overall, the COSMOS output, as illustrated in Table 3 and Figs. 4 to 6 not only yields sets of best trade-off surrogate models, but also provides important insights into the performance of the surrogate. For example, Fig. shows that the (modal value of the) median error in some of the best surrogate models could be reduced by approximately 50% if 20% more sample points are made available. It is also interesting to observe that the same surrogate-kernel combination (e.g., SVR-RBF in Fig.) could provide a wide spread of best trade-offs between median and maximum errors, depending on the hyper-parameter values.

CONCLUSION

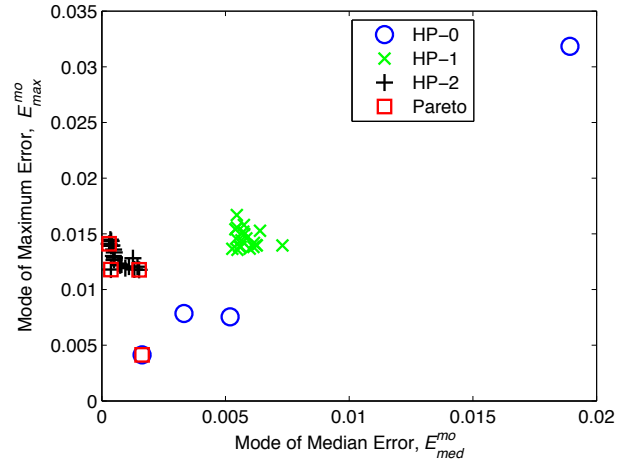
This paper advanced the surrogate model selection framework that is based on the error measure given by the Regional Error Estimation of Surrogates (REES) method. This framework is called the Collaborative Surrogate Model Selection (COSMOS) framework, and is aimed to allow the selection of globally-optimal surrogates, where the pool of candidate surrogate models can be expanded through further collaboration and contribution from the design engineering research community. To this end, COSMOS performs surrogate model selection simultaneously at all three levels: (i) selecting optimal surrogate type, (ii) selecting optimal kernel/basis function type, and (iii) finding optimum values of the hyper-parameters in the kernel. The two primary components of the COSMOS framework are (i) the error esti-

mation method (REES) which provides the criteria functions to guide surrogate model selection, and (ii) the optimal model selector that searches for the Pareto optimal set of surrogate models. The REES method provides a model-independent approach to quantify the modal values of the median and maximum errors in a surrogate, and has been shown to be significantly more accurate than typical leave-one-out cross validation. In this paper we advance the REES method to also yield the variance (or the standard deviation) of the median and maximum errors (indicating the surrogate robustness), which are estimated by modeling their relationship with increasing number of sample points. In addition, we also estimate the expected median error if the surrogate was constructed on a hypothetically larger set of sample points (with similar distribution).

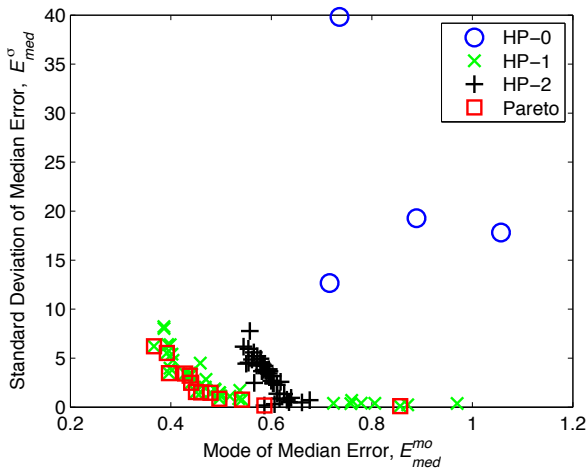
The optimal model selector solves a small set of parallel optimizations, each corresponding to candidate surrogate-kernel combinations involving a particular number of hyper-parameters. Each optimization solves a multi-objective MINLP problem; subsequently a Pareto filter is applied to determine the final Pareto set of surrogate models. In this paper, COSMOS was applied to two analytical test functions and an airfoil design problem. The comprehensive candidate pool included five major surrogate model types, with up to five different kernel types available per surrogate model. Widely different sets of surrogates models were selected (as the optimum set) in the three problems. In some cases, the same surrogate-kernel combination was observed to yield a noticeable spread of best trade-offs (between median and maximum error), accomplished through different hyper-parameter values. These numerical experiments showed that in addition to model selection, COSMOS can provide uniquely important insights into the relative capability of different surrogate models, e.g., their capacity for accuracy improvement if a larger set of samples is made available. With further advancement, we believe that COSMOS would lay the foundation for not only selection of globally-competitive surrogates for any given application, but also for benchmarking of existing and new surrogate models.



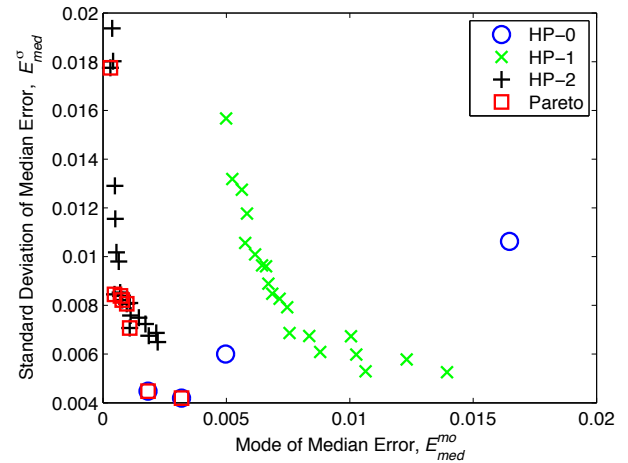
(a) Trade-offs between modal values of median and maximum error



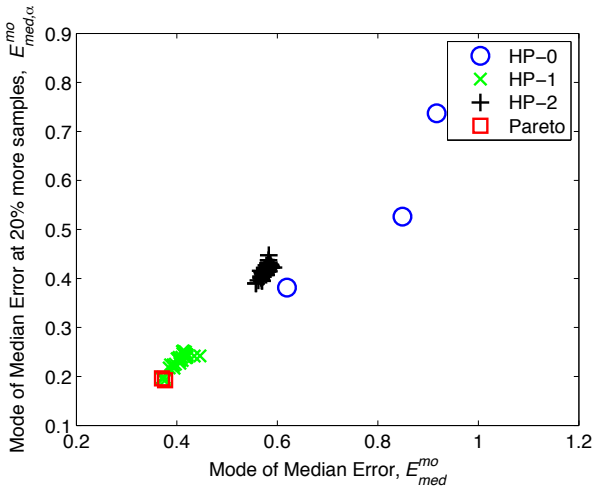
(a) Trade-offs between modal values of median and maximum error



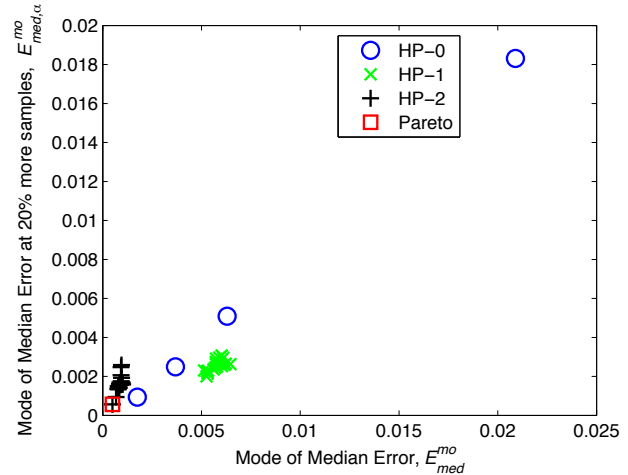
(b) Trade-offs between modal values and standard deviation of median error



(b) Trade-offs between modal values and standard deviation of median error



(c) Trade-offs between modal values of median error at N and $1.2N$ samples



(c) Trade-offs between modal values of median error at N and $1.2N$ samples

FIGURE 5. Hartmann function: Pareto models (red sq.) and final population of models from all *HP* classes

FIGURE 6. Airfoil Design: Pareto models (red sq.) and final population of models from all *HP* classes

ACKNOWLEDGEMENT

Support from the National Science Foundation Award CMMI 1100948 is gratefully acknowledged. Any opinions, findings, and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the NSF.

REFERENCES

- [1] Kleijnen, J., 1975. *Statistical techniques in simulation*, Vol. II. New York. Russian translation, Publishing House Statistics, Moscow, 1978.
- [2] Gorissen, D., Dhaene, T., and Turck, F. D., 2009. “Evolutionary model type selection for global surrogate modeling”. *Journal of Machine Learning Research*, **10**, pp. 2039–2078.
- [3] Jin, R., Chen, W., and Simpson, T. W., 2000. “Comparative studies of metamodeling techniques under multiple modeling criteria”. *AIAA*(4801).
- [4] Simpson, T., Korte, J., Mauery, T., and Mistree, F., 2001. “Kriging models for global approximation in simulation-based multidisciplinary design optimization”. *AIAA Journal*, **39**(12), pp. 2233–2241.
- [5] Forrester, A., and Keane, A., 2009. “Recent advances in surrogate-based optimization”. *Progress in Aerospace Sciences*, **45**(1-3), pp. 50–79.
- [6] Hardy, R. L., 1971. “Multiquadric equations of topography and other irregular surfaces”. *Journal of Geophysical Research*, **76**, pp. 1905–1915.
- [7] Clarke, S. M., Griebisch, J. H., and Simpson, T. W., 2005. “Analysis of support vector regression for approximation of complex engineering analyses”. *Journal of mechanical design*.
- [8] Yegnanarayana, B., 2004. *Artificial Neural Networks*. PHI Learning Pvt. Ltd.
- [9] Choi, K., Young, B., and Yang, R., 2001. “Moving least square method for reliability-based design optimization”. 4th World Congress of Structural and Multidisciplinary Optimization, pp. 4–8.
- [10] Toropov, V. V., Schramm, U., Sahai, A., Jones, R. D., and Zeguer, T., 2005. “Design optimization and stochastic analysis based on the moving least squares method”. 6th World Congresses of Structural and Multidisciplinary Optimization.
- [11] Queipo, N., Haftka, R., Shyy, W., Goel, T., Vaidyanathan, R., and Tucker, P., 2005. “Surrogate-based analysis and optimization”. *Progress in Aerospace Sciences*, **41**(1), pp. 1–28.
- [12] Claeskens, G., and Hjort, N. L. *Model selection and model averaging*. Cambridge Books.
- [13] Bozdogan, H., 2000. “Akaike’s information criterion and recent developments in information complexity”. *Journal of Mathematical Psychology*, **44**, pp. 62–91.
- [14] Gorissen, D., Tommasi, L. D., Hendrickx, W., Croon, J., and Dhaene, T., 2008. “Rf circuit block modeling via kriging surrogates”. In *Proceedings of the 17th International Conference on Microwaves, Radar and Wireless Communications (MIKON)*.
- [15] Viana, F. A. C., Haftka, R. T., and Jr., V. S., 2000. “Multiple surrogates: how cross-validation errors can help us to obtain the best predictor”. *Structural and Multidisciplinary Optimization*, **39**, p. 439457.
- [16] Martin, J. D., and Simpson, T. W., 2005. “Use of kriging models to approximate deterministic computer models”. *AIAA journal*, **43**(4), pp. 853–863.
- [17] Gorissen, D., Couckuyt, I., Demeester, P., Dhaene, T., and Crombecq, K., 2010. “Surrogate modeling and adaptive sampling toolbox for computer based design”. *The Journal of Machine Learning Research*, **99**, pp. 2051–2055.
- [18] Mehmani, A., Chowdhury, S., and Messac, A., 2014. “A novel approach to simultaneous selection of surrogate models, constitutive kernels, and hyper-parameter values”. In *AIAA SciTech 2014*, AIAA.
- [19] Mehmani, A., Chowdhury, S., Zhang, J., Tong, W., and Messac, A., 2013. “Quantifying regional error in surrogates by modeling its relationship with sample density”. In *54th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, AIAA.
- [20] Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T., 2002. “A fast and elitist multi-objective genetic algorithm: Nsgaii”. *IEEE Transactions on Evolutionary Computation*, **6**(2), April, pp. 182–197.
- [21] Haldar, A., and Mahadevan, S., 2000. *Probability, reliability, and statistical methods in engineering design*. Wiley.
- [22] Mongillo, M., 2011. “Choosing basis functions and shape parameters for radial basis function methods”. In *SIAM Undergraduate Research Online*.
- [23] Giunta, A., and Watson, L., 1998. A comparison of approximation modeling techniques: Polynomial versus interpolating models. Tech. Rep. AIAA-98-4758.
- [24] Sakata, S., Ashida, F., and Zako, M., 2003. “Structural optimization using kriging approximation”. *Computer Methods in Applied Mechanics and Engineering*, **192**(7-8), pp. 923–939.
- [25] Simpson, T., 1998. “A concept exploration method for product family design”. PhD thesis, Georgia Institute of Technology.
- [26] Lophaven, S. N., Nielsen, H. B., and Sondergaard, J., 2002. Dace - a matlab kriging toolbox, version 2.0. Tech. Rep. IMM-REP-2002-12, Informatics and Mathematical Modelling Report, Technical University of Denmark.
- [27] Cressie, N., 1993. *Statistics for Spatial Data*. Wiley, New York.
- [28] Chang, Chih-Chung, and Lin, C.-J., 2011. “LIBSVM: A library for support vector machines”. *ACM Transactions on Intelligent Systems and Technology*, **2**(3), pp. 27:1–27:27.
- [29] Coelho, F., Breitkopf, P., and Knopf-Lenoir, C., 2008. “Model reduction for multidisciplinary optimization: Ap-

plication to a 2d wing”. *Structural and Multidisciplinary Optimization*, **37**(1), pp. 29–48.

- [30] Zhang, J., Chowdhury, S., and Messac, A., 2012. “An adaptive hybrid surrogate model”. *Structural and Multidisciplinary Optimization*, **46**(2), pp. 223–238.

Appendix A: Analytical Test Functions

Branin-Hoo function (2 variables):

$$f(x) = \left(x_2 - \frac{5.1x_1^2}{4\pi^2} + \frac{5x_1}{\pi} - 6 \right)^2 + 10 \left(1 - \frac{1}{8\pi} \right) \cos(x_1) + 10 \quad (16)$$

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

Hartmann function (6 variables):

$$f(x) = - \sum_{i=1}^4 c_i \exp \left\{ - \sum_{j=1}^n A_{ij} (x_j - P_{ij})^2 \right\} \quad (17)$$

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

In Hartmann-6, $n = 6$, and the constants c , A , and P , are a 1×4 vector, a 4×6 matrix, and a 4×6 matrix, respectively [30].

Appendix B: Expressions for Candidate Surrogates-Kernels in COSMOS

TABLE 4. Basis or Kernel functions and their hyper-parameters in the candidate surrogate models

Type of model	Type of basis/correlation/kernel function	Hyper parameter	
RBF	Linear:	r	
	Cubic:	r^3	
	Thin plate spline:	$r^2 \ln(r)$	shape parameter, σ
	Gaussian:	$e^{(-r^2/2\sigma)}$	
	Multiquadric:	$(r^2 + \sigma^2)^{1/2}$	
Kriging	Linear:	$\max(1 - \theta r, 1)$	
	Exponential:	$e^{(-\theta r)}$	
	Gaussian:	$e^{(-\theta r)^2}$	
	Cubic:	$1 - 0.5\xi + 0.5\xi^2$; $\xi = \max(1 - \theta r, 1)$	
	Spherical:	$1 - 3\xi^2 + 2\xi^3$; $\xi = \max(1 - \theta r, 1)$	
SVR	Linear:	$(x_i^T x_j)$	kernel parameters; γ , r , and d
	Polynomial:	$(\gamma x_i^T x_j + r)^d$, $\gamma > 0$	
	Radial Basis Function:	$e^{-\gamma \ x_i - x_j\ ^2}$, $\gamma > 0$	
	Sigmoid:	$\tanh(\gamma x_i^T x_j + r)$	