

Adaptive Hybrid Surrogate Modeling for Complex Systems

Jie Zhang,* Souma Chowdhury,* and Junqiang Zhang*
Rensselaer Polytechnic Institute, Troy, New York 12180

Achille Messac[†]

Syracuse University, Syracuse, New York 13244

and

Luciano Castillo[‡]

Texas Tech University, Lubbock, Texas 79409

DOI: 10.2514/1.J052008

This paper explores the effectiveness of the recently developed surrogate modeling method, the Adaptive Hybrid Functions, when applied in conjunction with different sampling techniques and different sample sizes. The Adaptive Hybrid Functions is a hybrid surrogate modeling method that seeks to exploit the advantages of each component surrogate. To explore its effectiveness, the Adaptive Hybrid Functions is applied to model four disparate complex engineering systems, namely: 1) wind farm power generation, 2) product platform planning (for universal electric motors), 3) three-pane window heat transfer, and 4) onshore wind farm cost estimation. The effectiveness of the following three distinct sampling techniques is investigated: 1) Latin hypercube sampling, 2) Sobol's quasi-random sequence, and 3) the Hammersley sequence sampling. Cross-validation is used to evaluate the accuracy of the resulting surrogate models. It was observed that the Sobol's sequence and the Latin hypercube sampling techniques provided better accuracy in the case of high-dimensional problems, whereas the Hammersley sequence sampling technique performed better in the case of low-dimensional problems. It was further observed that a monotonic increase in modeling accuracy is not necessarily accomplished by simply increasing the sample size for different problems. Overall, the Adaptive Hybrid Functions provided acceptable to high accuracy in representing complex system behavior.

Nomenclature

CD_i	=	crowding distance of the i th point on the base model
C_t	=	total annual cost of a wind farm per kilowatt installed
Δx_{10}	=	full width at one-tenth maximum
d^i	=	adaptive distance between the i th corresponding point on the boundary and the base model
\tilde{f}_{AHF}	=	function value estimated by Adaptive Hybrid Functions
\tilde{f}_{erbf}	=	function value estimated by extended radial basis functions
\tilde{f}_{kri}	=	function value estimated by Kriging
\tilde{f}_{qrs}	=	function value estimated by quadratic response surface method
\tilde{f}_{rbf}	=	function value estimated by radial basis functions
η_{farm}	=	farm efficiency
λ_j^{ik}	=	commonality constraint
n_{dim}	=	number of variables
n_{int}	=	number of integer variables
n_t	=	number of test points
P_{farm}	=	power generated by a wind farm
$P_i(x)$	=	measure of accuracy of the i th surrogate at point x
\dot{Q}_w	=	the heat flux through the inner pane

ρ	=	local density of input data
w_i	=	weight of the i th component surrogate in the hybrid surrogate

I. Introduction

COMPLEX systems, such as human bodies, aerospace systems, energy systems, and wireless networking generally tend to be highly interdisciplinary. Understanding, designing, building, and controlling these complex systems remain a central challenge in academia and industry [1]. The determination of complex underlying relationships between system parameters from a limited number of simulated and/or recorded data can be accomplished using advanced interpolating/approximating functions, also known as surrogates. Engineering design systems often present varying levels and nature of complexities over the domain under consideration. These complexities may entail high nonlinearities, function discontinuity, and a regional scarcity of data. Hybrid surrogate models can provide a powerful solution to such systems. In the hybrid surrogate modeling paradigm, characteristically differing surrogate models are intelligently aggregated to offer a robust modeling solution.

Since 1990, advances in function estimation methods and approximation-based optimization have progressed remarkably. Surrogate models are extensively used in the analysis and optimization of computationally expensive simulation-based models. Surrogate modeling techniques have been used for a variety of applications in multidisciplinary design optimization to reduce the analysis time and to improve the tractability of complex analysis codes [2,3]. Surrogate modeling has been shown to be an effective approach in the development of aerospace systems, which often involves finite element analysis, computational fluid dynamics (CFD), and other computationally intensive analyses. For instance, surrogate modeling has been applied to airfoil shape optimization [4], wing box design [5], and diffuser shape optimization [6].

The literature offers a wide variety of surrogate modeling techniques, including: 1) polynomial response surface method [7], 2) Kriging [8,9], 3) radial basis functions (RBF) [10], 4) extended radial basis functions (E-RBF) [11,12], 5) artificial neural networks

Received 4 April 2012; revision received 15 August 2012; accepted for publication 9 September 2012; published online 24 January 2013. Copyright © 2012 by Achille Messac. Published by the American Institute of Aeronautics and Astronautics, Inc., with permission. Copies of this paper may be made for personal or internal use, on condition that the copier pay the \$10.00 per-copy fee to the Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923; include the code 1533-385X/13 and \$10.00 in correspondence with the CCC.

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

[†]Distinguished Professor and Department Chair, Department of Mechanical and Aerospace Engineering; messac@syr.edu. Lifetime Fellow AIAA (Corresponding Author).

[‡]Don-Kay-Clay Cash Distinguished Engineering Chair in Wind Energy, Executive Director/President of the National Wind Resource Center (NWRC), and Professor, Mechanical Engineering Department, Senior Member AIAA.

Table 1 Techniques for response surface development

Sampling/design of experiments	Surrogate modeling	Coefficient estimation
(Fractional) factorial	Polynomial (linear, quadratic)	Least-squares regression
Central composite	Splines (linear, cubic)	Best weighted least-squares regression
Latin hypercube	Kriging	Best linear predictor
Hammersley sequence	RBF	Log likelihood
Uniform designs	Extended RBF	Multipoint approximation
Sobol's sequence	SVR	Adaptive response surface
Random selection	Neural network	Back propagation
Box–Behnken	Hybrid models	Entropy
Plackett–Burman		Linear unbiased predictor
Orthogonal arrays		

[13], 6) support vector regression (SVR) [14,15], and 7) hybrid surrogate modeling methods [16–19]. Table 1 provides a list of standard sampling techniques, surrogate modeling methods, and function-coefficient estimation methods, which extends the preexisting work in Ref. [13].

More recently, researchers have presented the development of a combination of different approximate models into a single hybrid model for developing weighted-average surrogates [16,17,20–24]. Zepa et al. [16] showed one application using an ensemble of surrogate models to construct a weighted-average surrogate for the optimization of alkaline–surfactant–polymer flooding processes. They found that the weighted average surrogate provided better performance than individual surrogates. Goel et al. [17] combined three surrogate models (polynomial response surface, Kriging, and radial basis neural network), and used the *generalized mean square cross-validation error* of individual surrogate models to select appropriate weight factors. Acar and Rais-Rohani [21] treated the selection of weight factors in the general weighted-sum formulation of an ensemble as an optimization problem with the objective to minimize an error metric. The results showed that the optimized ensemble provided more accurate predictions than the stand-alone surrogate models. Acar [23] investigated the efficiency of using various local error measures for constructing an ensemble of surrogate models, and also presented the use of the pointwise cross-validation error as a local error measure. Zhou et al. [24] used a recursive process to obtain the values of weights in which the values of surrogate weights are updated in each iteration until the last ensemble achieves a desirable prediction accuracy. Zhang et al. [18] developed a high-fidelity surrogate modeling technique, the Adaptive Hybrid Functions (AHF), by adaptively combining the favorable characteristics of different surrogate models. In this paper, we are extending the work on AHF by investigating 1) its applicability to engineering design systems and 2) its performance variations with the type of sampling technique and sample size.

Complex engineering systems, which can be highly nonlinear and computationally expensive, present significant challenges to accurate representation through surrogate modeling. At the same time, the choice of appropriate sampling techniques is generally considered critical to the performance of any surrogate modeling approach. It would be helpful to investigate the effectiveness of a surrogate model through complex system modeling in conjunction with distinct sampling techniques.

The recently developed hybrid surrogate modeling method, the AHF, combines the component surrogate models by characterizing and evaluating the local *measure of accuracy* of each model. A novel *crowding distance-based trust region* (CD-TR) was proposed to capture both the global trend and the local variations of the system behavior. The weights of the component surrogates are adaptively selected based on the *measure of accuracy* of each surrogate in the trust region. This paper explores the original AHF methodology by 1) applying the AHF to complex engineered systems design and economic system design problems; 2) implementing three representative sampling techniques: Latin hypercube sampling (LHS), Sobol's quasi-random sequence, and the Hammersley sequence sampling (HSS); and 3) investigating the effects of sample size and problem dimensionality on the accuracy of the surrogate model.

The remainder of this paper is organized as follows: Sec. II briefly introduces the AHF methodology; Secs. III and IV respectively present the formulations of the complex problems and the experimental design strategy; and the results of the AHF surrogate modeling are discussed in Sec. V.

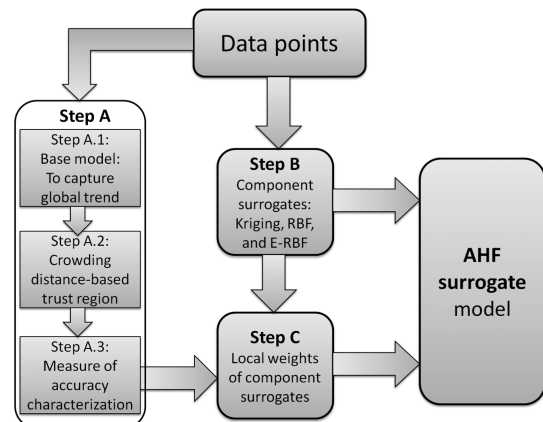
II. Adaptive Hybrid Functions

The AHF methodology was recently developed by Zhang et al. [18,25]. The AHF formulates a trust region based on the density of available sample points, and adaptively combines characteristically differing surrogate models. The weight of each contributing surrogate model is represented as a function of the input domain based on a local *measure of accuracy* for that surrogate model. Such an approach exploits the advantages of each component surrogate, thereby, capturing both the global and the local trends of complex functional relationships. In this paper, the AHF combines the three component surrogate models, 1) RBF, 2) E-RBF, and 3) Kriging, by characterizing and evaluating the local *measure of accuracy* of each model. The key components of the hybrid surrogate modeling methodology include: 1) generation of the component surrogates (RBF, E-RBF, and Kriging in this paper); 2) determination of a CD-TR: numerical bounds of the estimated parameter (output) as a function of the independent parameters (input vector) over the feasible input space (Fig. 1, step A.2); 3) characterization of the local *measures of accuracy* (using kernel functions) of the estimated function value, and the representation of the corresponding kernel function parameters as functions of the input vector (Fig. 1, step A.3); and 4) weighted aggregation of the function values estimated by the individual surrogates, based on their local measures of accuracy.

The key components of the training process of the AHF are illustrated in Fig. 1 and discussed in the following sections.

A. Generation of Component Surrogates

Different interpolating surrogate models (component surrogates) are constructed. The selected component surrogate models are intended to be locally accurate, with greater accuracy in the region close to the training points. Three component surrogates are

**Fig. 1** The framework of the AHF surrogate model [18].

constructed based on the set of training points D using Kriging, RBF, and E-RBF. We can also integrate other standard surrogates that are locally accurate. For each test point, the estimated function vector is represented as $\tilde{f} = \{f_{\text{kri}}, f_{\text{rbf}}, f_{\text{erbf}}\}$. The parameters f_{kri} , f_{rbf} , and f_{erbf} represent function values estimated by the Kriging, the RBF, and the E-RBF methods, respectively.

The three component surrogates are expected to offer different levels of numerical fidelity and tractability. In general, Kriging models are more accurate for nonlinear problems [3]. The RBF approach can accurately model scattered multivariate data. The RBF method is also generally easier to construct compared to Kriging [2,11] because the parameters in Kriging are typically determined through suboptimization (via maximum likelihood estimation or cross-validation). The E-RBF approach provides additional degrees of freedom compared to RBF, in order to ensure certain desirable properties, such as nonlinearity and convexity. However, E-RBF involves a significantly larger number of parameters compared to RBF.

1. Kriging

The general form of the Kriging surrogate model is given in [26]

$$\tilde{f}(x) = G(x) + Z(x) \tag{1}$$

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

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

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. In the present paper, a Gaussian function is used as the correlation function, defined as

$$R(x^i, x^j) = R_{ij} = \exp\{-\sum_{k=1}^{n_d} \theta_k (x_k^i - x_k^j)^2\} \tag{3}$$

where θ_k is distinct for each dimension, and these unknown parameters are generally estimated by solving a nonlinear optimization problem. In this paper, we use an efficient MATLAB implementation of the Kriging surrogate, *the design and analysis of computer experiments*, developed by Lophaven et al. [27]. The order of the global polynomial trend function is specified to be zero in this paper.

2. Radial Basis Functions

RBFs are expressed in terms of the Euclidean distance, $r = \|x - x^i\|$, of a point x from a given data point x^i . The multiquadric function [10,28] is adopted in this paper, which is defined as

$$\psi(r) = \sqrt{r^2 + c^2} \tag{4}$$

where $c > 0$ is a prescribed real valued parameter. The final approximation function is a linear combination of these basis functions across all the data points, as given by

$$\tilde{f}(x) = \sum_{i=1}^{n_p} \sigma_i \psi(\|x - x^i\|) \tag{5}$$

where σ_i s are the unknown coefficients (to be determined), and n_p denotes the number of training points.

3. Extended Radial Basis Functions

The E-RBF [11] approach uses a combination of the radial and the nonradial basis functions (N-RBFs). The N-RBFs are not functions of the Euclidean distance, r . Instead, they are functions of individual

coordinates of generic points x relative to a given data point x^i , in each dimension separately. We define the coordinate vector as $\xi^i = x - x^i$, which is a vector of n_d elements, each corresponding to a single coordinate dimension. Thus, ξ_j^i is the coordinate of any point x relative to the data point x^i along the j th dimension. The N-RBF for the i th data point and the j th dimension is denoted by ϕ_{ij} . It is composed of three distinct components, as given by

$$\phi_{ij}(\xi_j^i) = \alpha_{ij}^L \phi^L(\xi_j^i) + \alpha_{ij}^R \phi^R(\xi_j^i) + \beta_{ij} \phi^\beta(\xi_j^i) \tag{6}$$

where α_{ij}^L , α_{ij}^R , and β_{ij} are coefficients to be determined for the given problem. The E-RBF approach presents a linear combination of RBF and N-RBF. The approximation function takes the form

$$\tilde{f}(x) = \sum_{i=1}^{n_p} \sigma_i \psi\left(\|x - x^i\|\right) + \sum_{i=1}^{n_p} \sum_{j=1}^{n_d} \{\alpha_{ij}^L \phi^L(\xi_j^i) + \alpha_{ij}^R \phi^R(\xi_j^i) + \beta_{ij} \phi^\beta(\xi_j^i)\} \tag{7}$$

where ϕ^L , ϕ^R , and ϕ^β are components of the N-RBF. The vectors α^L , α^R , and β , defined above, contain $n_d n_p$ elements each, and the vector σ contains n_p coefficients.

B. Crowding Distance-Based Trust Region

The CD-TR is formulated using the following two-step procedure:

- 1) the determination of the base model using smooth functions and
- 2) the formulation of trust region boundaries.

1. Determination of Base Model

The base model is developed using smooth functions to obtain a global approximation for the given set of points, D . This base model is intended to capture the global trend of the training points, thereby, addressing the global accuracy for the overall surrogate. This research constructs the base model using the quadratic response surface method (QRSM). The coefficients of the QRSM are determined using the least-squares method. However, the base model also has the flexibility to use other smooth functions (regression functions).

2. Formulation of Trust Region Boundaries

In this step, a trust region is formulated for surrogate modeling based on the density of sample points, which we call the CD-TR. A set of points are selected on the base model, and the crowding distance is evaluated for each point. The trust region boundaries of the surrogate model are adaptively constructed according to the base model and the estimated crowding distances of the training points. The base model is relaxed along either directions of the output axis to obtain the boundaries of the surrogate.

Crowding distance is used to evaluate the density of training points surrounding any point on the base model. A larger crowding distance value at a point reflects a lower sample density (fewer points around that point), and the *measure of accuracy* of a surrogate is expected to be relatively lower around that point. Therefore, we need wider boundaries at that point. Based on the crowding distance value at each point on the base model, we construct adaptive boundaries of the CD-TR. In this paper, the crowding distance of the i th point on the base model (CD^i) is evaluated by

$$CD^i = \sum_{j=1}^{n_p} \|x^j - x^i\|^2 \tag{8}$$

where n_p is the number of training points used. A parameter ρ is defined to represent the local density of input data, which is given by

$$\rho^i = \frac{1}{CD^i} \tag{9}$$

The parameter, ρ , is then normalized to obtain α_i s, as given by

Downloaded by SYRACUSE UNIVERSITY LIBRARY on February 19, 2013 | http://arc.aiaa.org | DOI: 10.2514/1.1052008

$$\alpha_i = \frac{\max(\rho) - \rho^i}{\max(\rho) - \min(\rho)} \quad (10)$$

The adaptive distance d^i between the i th corresponding point on the boundary and the base model along either directions of the output axis is expressed as

$$d^i = (1 + \alpha_i) \times \max_{j \in D} |\tilde{f}_{\text{qrsm}}(x^j) - y^j| \quad (11)$$

where D represents the original training data set. In Eq. (11), the index, j , represents training points, and the index, i , represents a uniform set of points selected on the base model. In Eq. (11), the adaptive distance is divided into two parts:

1) $\max_{j \in D} |\tilde{f}_{\text{qrsm}}(x^j) - y^j|$ is a constant that ensures that all of the training points are located between the boundaries.

2) $\alpha_i \times \max_{j \in D} |\tilde{f}_{\text{qrsm}}(x^j) - y^j|$ is an adaptive distance based on the distance coefficients (α_i 's).

Crowding distance is evaluated with respect to each of the selected points based on which α is previously formulated. The extent of the boundary region is scaled using the maximum of the training data deviation from the base model. Here, $\tilde{f}_{\text{qrsm}}(x^j)$ is the output value of the j th training point estimated by QRSM and $|\tilde{f}_{\text{qrsm}}(x^j) - y^j|$ is the distance from the j th training point to the base model along the direction of the output axis. Subsequently, we obtain two sets of points, D^U and D^L , for constructing the two boundaries. QRSM is adopted to estimate the upper and the lower boundary surfaces, using the generated data points D^U and D^L , respectively. The probabilities associated with individual component surrogates are determined within these trust region boundaries.

The use of crowding distance for evaluating the density of training points can be modified and adopted by other surrogate modeling techniques. For other methods on ensemble of surrogates, researchers can also use the idea of trust regions based on local error measures, such as prediction variance [16] or pointwise cross-validation error [23].

C. Accuracy Measure of Surrogate Modeling

With the CD-TR, it is important to be able to determine the *measure of accuracy* of the estimated function value at any given point in the trust region. Based on the value of the *measure of accuracy*, we can adaptively combine different component surrogate models. A metric is developed, which we call the *accuracy measure of surrogate modeling*, to represent the uncertainty in the estimated function value.

Function estimation is performed between the two boundary surfaces, using a local *measure of accuracy* technique. The uncertainty in the estimated function value at a location in the input variable domain is modeled using a kernel function. This kernel function is expressed as a function of the output parameter. The corresponding coefficients of the kernel function are represented as functions of the input vector, thereby, characterizing the *measure of accuracy* of the estimated function over the entire input domain.

The kernel function used to represent the *measure of accuracy* must have the following properties:

1) The kernel function value must be a maximum of one at the actual output, $y(x^i)$.

2) The kernel function must be equal to the specified small tolerance value at the upper and the lower boundaries of the trust region.

3) The function must increase monotonically from either boundary to the actual output value.

4) The function must be continuous.

The following kernel function is adopted for use in representing the *measure of accuracy*. This kernel function, which satisfies the specified requirements, is expressed as

$$P(z) = a \exp\left[-\frac{(z - \mu)^2}{2\sigma^2}\right] \quad (12)$$

where the amplitude coefficient a is specified as one; the coefficients μ and σ represent the mean and the standard deviation of the kernel

function, respectively. Note that other kernel functions that have similar properties can also be used to represent the *measure of accuracy*.

The distance between the two boundaries is normalized. At each training data point, x^i , the output value at the lower and upper boundaries, $f_L^i(x^i)$, and $f_U^i(x^i)$, respectively, are also normalized. It is assumed that the estimated *measure of accuracy* (kernel function) is a maximum of 1 at the actual output value $y(x^i)$, and a minimum of 0.1 at the boundaries (within the trust region). The actual output value of a training point does not necessarily occur midway between the two boundaries. To ensure the continuity of the kernel function, the function is divided into two parts, with distinct standard deviations and the same mean. Then, the kernel function is represented as

$$P(x^i) = \begin{cases} a \exp\left\{-\frac{[y(x^i) - \mu(x^i)]^2}{2\sigma_1^2(x^i)}\right\} & \text{if } 0 \leq y(X^i) < \mu(X^i) \\ a \exp\left\{-\frac{[y(x^i) - \mu(x^i)]^2}{2\sigma_2^2(x^i)}\right\} & \text{if } \mu(X^i) \leq y(X^i) \leq 1 \end{cases} \quad (13)$$

where the parameters σ_1 and σ_2 , controlled by the full width at one-tenth maximum (Δz_{10}) requirement, are given by

$$\begin{aligned} \sigma_1(x^i) &= \frac{\Delta z_{10}(x^i)}{2\sqrt{2} \ln 10} = \frac{2[\mu(x^i) - f_L^i(x^i)]}{2\sqrt{2} \ln 10} \\ &= \frac{2\mu(x^i)}{2\sqrt{2} \ln 10} = \frac{\mu(x^i)}{\sqrt{2} \ln 10} \end{aligned} \quad (14)$$

and

$$\begin{aligned} \sigma_2(x^i) &= \frac{\Delta z_{10}(x^i)}{2\sqrt{2} \ln 10} = \frac{2[f_U^i(x^i) - \mu(x^i)]}{2\sqrt{2} \ln 10} \\ &= \frac{2[1 - \mu(x^i)]}{2\sqrt{2} \ln 10} = \frac{1 - \mu(x^i)}{\sqrt{2} \ln 10} \end{aligned} \quad (15)$$

where

$$P(\mu \pm 0.5\Delta z_{10}) = \frac{1}{10} \quad (16)$$

From Eq. (13), the *measure of accuracy* coefficient $\mu(x^i)$ is determined for the i th training point. The coefficient μ is expressed in terms of input variables x_j^i using a polynomial response surface.

D. Ensemble of Component Surrogates

The AHF surrogate model is formulated by the adaptive selection of weights for the three component surrogate models (RBF, E-RBF, and Kriging). The AHF is a weighted summation of function values estimated by the component surrogates, as given by

$$\tilde{f}_{\text{AHF}} = \sum_{i=1}^{n_s} w_i(x) \tilde{f}_i(x) \quad (17)$$

where n_s is the number of component surrogates combined into the AHF, and $\tilde{f}_i(x)$ represents the value estimated by the i th component surrogate. The generic weight w_i is expressed in terms of the estimated *measure of accuracy*, which is given by

$$w_i(x) = \frac{P_i(x)}{\sum_{i=1}^{n_s} P_i(x)} \quad (18)$$

where $P_i(x)$ is the *measure of accuracy* of the i th surrogate at point x .

III. Complex Engineered Systems and Economic System

The AHF is applied to complex engineering systems and an economic system, which are 1) wind farm power generation, 2) product platform planning (for universal electric motors), 3) three-pane window heat transfer, and 4) onshore wind farm cost estimation. The attributes of the four examples discussed in this paper offer similarities to other aerospace engineering problems. For instance, wind energy and aerospace (e.g., aircraft, missiles, and launch

vehicles) involve similar disciplinary systems. In addition, product platform planning concepts are also widely used in the design of aircraft.

A. Wind Farm Power Generation

The power generated by a wind farm is an intricate function of the configuration and location of the individual wind turbines. The power generation model used in this paper is adopted from Ref. [29]. The flow pattern inside a wind farm is complex, primarily due to the wake effects and the highly turbulent flow. The power generated by a wind farm (P_{farm}) comprised of N wind turbines is evaluated as a sum of the powers generated by the individual turbines, which is expressed as [29]

$$P_{\text{farm}} = \sum_{j=1}^N P_j \tag{19}$$

Accordingly, the farm efficiency can be expressed as

$$\eta_{\text{farm}} = \frac{P_{\text{farm}}}{\sum_{j=1}^N P_{0j}} \tag{20}$$

where P_{0j} is the power that turbine- j would generate if operating as a stand-alone entity for the given incoming wind velocity. Detailed formulation of the power generation model can be found in Ref. [29].

The power generation is a function of the location coordinates of each turbine. With the turbine type and operating conditions remaining fixed, the $x - y$ coordinates are the design variables for wind farm layout optimization. In this paper, a hybrid surrogate is developed using AHF to represent the net power generation as a function of the turbine location coordinates. In the case of a wind farm comprised of N turbines, the power generation model presents a $2N$ dimensional problem.

The power generation model is a complex combination of nonlinear submodels, including the wake model, the wake overlap model, and the turbine power response model. The overall farm power generation function is, thus, a highly nonlinear function that generally does not have a straightforward analytical expression. In addition, owing to the inherent nature of the layout design approach, the power generation function is also expected to be multimodal.

With respect to the wind farm layout design problem, four cases are considered: 1) a wind farm with 4 turbines (8 variables), 2) a wind farm with 9 turbines (18 variables), 3) a wind farm with 16 turbines (32 variables), and 4) a wind farm with 25 turbines (50 variables). The GE-1.5MW-xle [30] turbine is chosen as the specified turbine type for all cases.

B. Product Platform Planning (for Universal Electric Motors)

A product family is a group of related products that are derived from a common product platform to satisfy a variety of market niches [31]. The sharing of a common platform by different products is expected to result in 1) reduced overhead, 2) lower per product cost, and 3) increased profit. The recently developed comprehensive product platform planning (CP³) framework [32] formulated a generalized mathematical model to represent the complex platform planning process.

The CP³ model formulates a generic equality constraint (the *commonality constraint*) to represent the variable-based platform formation. The presence of a combination of integer variables (specifically binary variables) and continuous variables can be attributed to the combinatorial process of platform identification. The nonlinearity of this problem is attributed to the nonlinear performance functions and nonlinear constraints in the physical system of the products. The process of testing whether a product family (comprising N products and n design variables) satisfies this commonality constraint is explained in Fig. 2. In this figure, the black arrows represent the process direction, and the grey arrows represent the flow of information (on an as-needed basis). The details of the CP³ model can be found in [32,33].

The net performance of the universal motor family is estimated as

$$f_{\text{perf}} = \frac{1}{10} \sum_{k=1}^{10} \eta^k \tag{21}$$

where η^k is the efficiency of the motor k , which is a function of design variables.

In the case of a scaling product family, where each product comprises n parts, the commonality index (CI) can be simplified to

$$\text{CI} = 1 - \frac{R_\lambda - n}{n(N - 1)} \tag{22}$$

where R_λ is the rank of the commonality matrix λ . The expression of λ can be found in Ref. [32].

The first equality constraint is termed the commonality constraint [Eq. (23)], which involves the parameters λ_j^{lk} . $X^T \Lambda X = 0$, where

$$\Lambda = \begin{bmatrix} \sum_{k \neq 1} \lambda_1^{1k} & \dots & -\lambda_1^{1N} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\lambda_1^{N1} & \dots & \sum_{k \neq N} \lambda_1^{Nk} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \vdots & \vdots & \vdots & \vdots & \vdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \vdots & \sum_{k \neq 1} \lambda_j^{1k} & \dots & -\lambda_j^{1N} & \vdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \vdots & \vdots & \vdots & \vdots & \vdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \vdots & -\lambda_j^{N1} & \dots & \sum_{k \neq N} \lambda_j^{Nk} & \vdots & 0 & 0 & 0 \\ 0 & 0 & 0 & \vdots & \vdots & \vdots & \vdots & \vdots & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \sum_{k \neq 1} \lambda_n^{1k} & \dots & -\lambda_n^{1N} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\lambda_n^{N1} & \dots & \sum_{k \neq N} \lambda_n^{Nk} \end{bmatrix} \tag{23}$$

where $k = 1, 2, \dots, N$; and $j = 1, 2, \dots, n$;

$$X = [x_1^1 \ x_1^2 \ \dots \ x_1^N \ \dots \ x_j^1 \ x_j^2 \ \dots \ x_j^N \ \dots \ x_n^1 \ x_n^2 \ \dots \ x_n^N]^T$$

$$\lambda_j^{kl} = \begin{cases} 1, & \text{if } \lambda_j^{kk} = \lambda_j^{ll} = 1 \text{ and } x_j^k = x_j^l \\ 0, & \text{otherwise} \end{cases} \quad \forall k \neq l$$

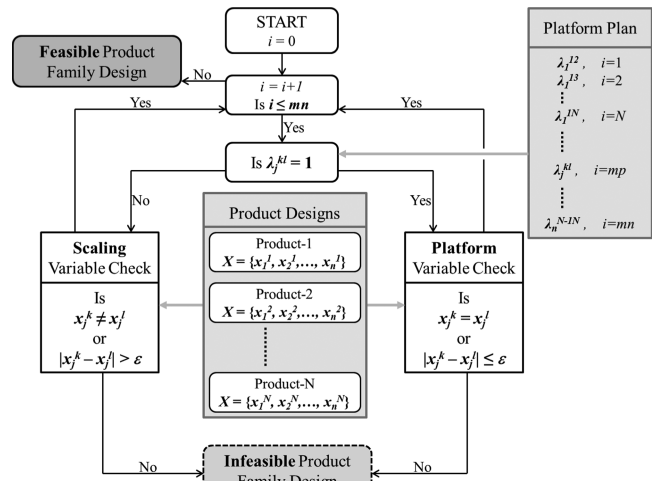


Fig. 2 The process of applying the commonality constraint.

$$\lambda_j^{kk} = \begin{cases} 1, & \text{if the } j\text{th variable is included in product } -k \\ 0, & \text{if the } j\text{th variable is NOT included in product } -k \end{cases}$$

where $l = 1, 2, \dots, N$; and $j = 1, 2, \dots, n$;

The net constraint violation $f_c(X)$ is determined by

$$f_c(X) = \sum_{j=1}^p \max(g_j, 0) + \sum_{k=1}^q \max(h_k - \epsilon, 0) \quad (24)$$

where g and h represent the inequality and equality constraints related to the physical design of the product, respectively. The functions g_j and h_k represent the normalized values of the j th inequality constraint and k th equality constraint, respectively. The parameter ϵ represents the tolerance specified to relax each equality constraint.

In this paper, surrogates are developed to model the key criteria functions involved in the design of a family of universal electric motors. Universal motors are frequently used in a variety of applications, e.g., electric drills and saws, blenders, vacuum cleaners, and sewing machines [32]. In product family design, the overall objective is to design a family of motors that have high efficiencies and low masses, as well as a high degree of commonality among them. Each motor is subjected to additional design constraints. The design of each motor involves seven design variables; the corresponding variable limits are given in Table 2.

The complexities of the performance objective and the net system constraint of the product platform planning problem depend on the system complexity of the products being designed. The case study in this paper corresponds to the design of a family of universal electric motors for which the performance function and the system constraint are fairly nonlinear. The commonality constraint is nonlinear and multimodal. Overall, the product platform planning problem, thus, presents a set of four complex nonlinear criterion functions.

The AHF method is used to represent the performance objective (f_{perf}), the CI, and the two constraints (system constraint and commonality constraint) as functions of the design variables. In the case of universal electric motors, the total number of design variables is $(N_{\text{pro}} + 1) \times 7$, where N_{pro} represents the number of product variants in the family. Three cases have been considered: 1) 2 products (21 variables), 2) 3 products (28 variables), and 3) 4 products (35 variables).

C. Three-Pane Window Heat Transfer

The performance of the three-pane window varies with climatic conditions [34]. The heat transfer through the window is a function of environmental conditions and the window material and dimensions. The constant indoor temperature and the window design are specified in this paper. A simplified schematic of the three-pane window is shown in Fig. 3.

The heat transfer simulation model of the side channels and the air gap is created using the CFD software Fluent. The model simulates the steady-state heat transfer process. In this study, the middle tinted pane is made of a generic bronze glass, and the other two panes are made of clear glass. The CFD simulation of the three-pane window is computationally expensive.

To reduce the computational expense of the Fluent model, a surrogate model is developed using the AHF. The inputs for the surrogate model are 1) the atmospheric temperature, 2) the wind speed, and 3) the solar radiation. The output of the surrogate model is the heat flux through the inner pane, \dot{Q}_w .

Table 2 Design variable limits of the electric motors

Design variable	Lower limit	Upper limit
Number of turns on the armature (N_a)	100	1500
Number of turns on each field pole (N_f)	1	500
Cross-sectional area of the armature wire (A_{wa})	0.01 mm ²	1.00 mm ²
Cross-sectional area of the field pole wire (A_{wf})	0.01 mm ²	1.00 mm ²
Radius of the motor (r_o)	10.00 mm	100.00 mm
Thickness of the stator (t)	0.50 mm	10.00 mm
Stack length of the motor (L)	1.00 mm	100.00 mm

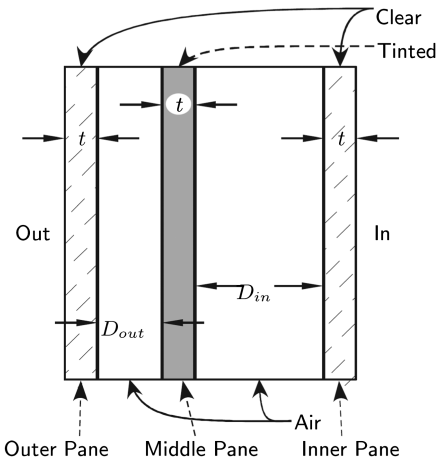


Fig. 3 Schematic of the three-pane window [34].

D. Onshore Wind Farm Cost Model

A response surface-based wind farm cost (RS-WFC) model was developed in Ref. [35]. The RS-WFC model, for onshore wind farms in the United States, was implemented using E-RBF. The RS-WFC model estimates the total annual cost of a wind farm per kilowatt installed, C_t . In the RS-WFC model, C_{LC} , C_{LT} , and C_{LM} represent the wage per hour for the construction labor, the technician labor, and the management labor, respectively; N is the number of turbines in a farm; and D is the rotor diameter. The wind turbine lifetime (in years), n , the number of years financed, n_f , the percentage financed, θ , and the interest rate, η , are constants specified as 20 years, 10 years, 80%, and 10%, respectively. Figure 4 shows the inputs and the outputs of the RS-WFC model.

In this paper, the total annual cost of a wind farm is estimated using the AHF method. The input parameters to the total annual cost model are 1) the number and 2) the rated power of wind turbines installed in a wind farm, and the output is the total annual cost of a wind farm. Data points collected for the state of North Dakota are used to develop the cost model, thereby, showing the effectiveness of the AHF method in economic application.

It is important to note that the CD-TR estimation is particularly useful for problems when the user does not have control over sampling, when the sample comprises commercial data or data from prior experiments. The wind farm cost data is not generated through the design of experiments, and hence the RS-WFC model is suitable for illustrating the effectiveness of the CD-TR concept in the AHF surrogate.

IV. Experimental Designs

Sampling techniques are used to select a specified number of data points over the design domain where the expensive system function is evaluated. The choice of an appropriate sampling technique is generally critical to the performance of a surrogate modeling approach. However, in several practical situations, the designer does not have full control over the choice of design samples (e.g., surrogate development from recorded data). In these situations, it is important to select a surrogate modeling approach whose performance is less sensitive to the sampling technique used. The following three representative sampling techniques are used to study their effects on the performance of the resulting surrogates.

A. Latin Hypercube Sampling

Latin Hypercube Sampling (LHS) is a strategy for generating random sample points. A Latin Hypercube sample set containing n_p sample points (between zero and one) over m dimensions is a matrix of n_p rows and m columns. Each row corresponds to a sample point. The values of the n_p points in each column are randomly selected, one from each of the intervals $(0, 1/n_p)$, $(1/n_p, 2/n_p)$, \dots , and $(1 - 1/n_p, 1)$ [36]. Different criteria can be applied to generate uniform representation, such as maximizing the minimum distance between points and reducing the correlation between points. It is important to

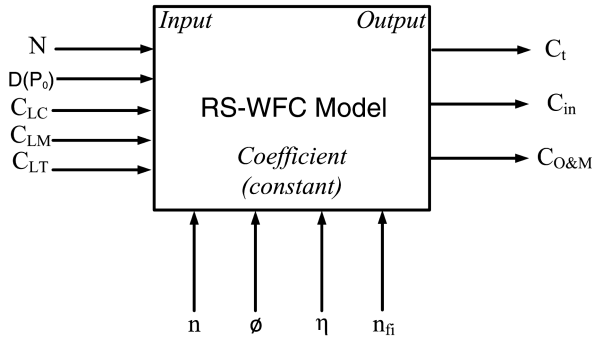


Fig. 4 The input and output of the RS-WFC model.

note that the basic LHS method without any optimality criterion is adopted in this paper for a fair comparison of sampling methods.

B. Sobol’s Quasi-Random Sequence Generator

Sobol’s sequences [37] use a base of two to form successively finer uniform partitions of the unit interval and reorder the coordinates in each dimension. The algorithm for generating Sobol’s sequences is discussed in Ref. [38].

C. Hammersley Sequence Sampling

HSS is based on the representation of a decimal number in the inverse radix format, where the radix values are chosen as the first $(m - 1)$ prime numbers, m being the number of dimensions. A detailed description of this technique can be found in [39]. The Hammersley sequence provides a highly uniform set of sample points containing n_p points in a m -dimensional unit hypercube.

LHS is useful if one desires a sample that is predominantly random but is uniform in each dimension separately. We also specifically investigate whether the uniformity properties in Sobol’s and the HSS methods indeed translate into higher accuracy.

D. Sampling Strategy for Problems

For the engineering design problems in this paper, the training data and the test data are obtained through simulation. Four sample sizes are considered: 1) 5 times, 2) 10 times, 3) 15 times, and 4) 20 times of the given problem dimension (the number of input variables) for each

sampling technique. Table 3 lists the details of the design of experiments. The variable limits are given in Table 4. For the onshore wind farm cost problem, available cost data is used [40].

LHS generates different sets of sample points everytime it is initiated. For the engineering-design problems, a set of 100 different experimental designs with the LHS method are used to account for the randomness of the LHS.

Two types of integer variables are involved in the product family problem. The first type of integer variables are physical design variables, including 1) the number of turns on the armature, N_a and 2) the number of turns on each field pole, N_f (see Table 2). For these two integer variables, a set of real numbers are generated using a sampling method and rounded to the nearest integers. The second type of integer variables are commonality design variables. The values of the second type of integer variables (z_j in Table 4) belong to a feasible set of integers, $Z = [0, 1, \dots, 2^{N_{pro}(N_{pro}-1)} - 1]$. It is seen from Table 1 that: for two products, $z_j \in \{0, 1\}$; for three products, $z_j \in \{0, 1, 2, 4, 7\}$; and for four products, $z_j \in \{0, 1, 2, 4, 7, 8, 12, 16, 18, 28, 32, 33, 42, 52, 63\}$. To sample points from the set z_j , we generate a set of real numbers between 1 and p and round the numbers to the nearest feasible integers, \hat{z}_j . Here, p is the number of elements in the feasible set Z , e.g., $p = 2$ for two products, $p = 5$ for three products, and $p = 15$ for four products. Finally, we map the value of \hat{z}_j to the value of the variable z_j , using the following mapping rules: 1) $\{1, 2\} \rightarrow \{0, 1\}$ for two products; 2) $\{1, 2, 3, 4, 5\} \rightarrow \{0, 1, 2, 4, 7\}$ for three products; and 3) $\{1, 2, \dots, 15\} \rightarrow \{0, 1, 2, 4, 7, 8, 12, 16, 18, 28, 32, 33, 42, 52, 63\}$ for four products.

V. Numerical Experiments

A. Selection of Parameters

Through numerical experiments, it was found that the following prescribed coefficient values generally produced more accurate function estimations. We set $c = 0.9$ for the RBF approach. We use $c = 0.9$ and $\lambda = 4.75$ for the E-RBF approach. The parameter t of the E-RBF approach is fixed at two (second degree monomial). For the Kriging approach, the bounds on the correlation parameters in the nonlinear optimization, θ_l and θ_u , are selected to be 0.1 and 20, respectively. During the cross-validation process, the number of subsets (q) is specified to be five. The prescribed values are summarized in Table 5.

Table 3 Experimental design for each problem

Engineering problem	N/N_{pro}^a	n_{int}^a	n_{dim}^a	Sample size (training points)				No. of test points
				$5 \times n_{dim}$	$10 \times n_{dim}$	$15 \times n_{dim}$	$20 \times n_{dim}$	
Wind farm 1	4	0	8	40	80	120	160	100
Wind farm 2	9	0	18	90	180	270	360	200
Wind farm 3	16	0	32	160	320	480	640	320
Wind farm 4	25	0	50	250	500	750	1000	500
Product family 1	2	11	21	105	210	315	420	210
Product family 2	3	13	28	140	280	420	560	280
Product family 3	4	15	35	175	350	525	700	350
Three-pane window	—	0	3	15	30	45	60	216

^a N : No. of turbines for wind farm; n_{dim} : No. of variables; N_{pro} : No. of products for product family; n_{int} : No. of integer variables.

Table 4 The limits of design variables

Problem	Variable limits
Wind farm 1	$0 < x_i < 7D_0, 0 < y_i < 3D_0$, where $i = 1, 2, 3, 4$
Wind farm 2	$0 < x_i < 2 \times 7D_0, 0 < y_i < 2 \times 3D_0$, where $i = 1, 2, \dots, 9$
Wind farm 3	$0 < x_i < 3 \times 7D_0, 0 < y_i < 3 \times 3D_0$, where $i = 1, 2, \dots, 16$
Wind farm 4	$0 < x_i < 4 \times 7D_0, 0 < y_i < 4 \times 3D_0$, where $i = 1, 2, \dots, 25$
Product family 1	x_j^1, x_j^2 (see Table 3), $z_j \in \{0, 1\}$, where $j = 1, 2, \dots, 7$
Product family 2	x_j^1, x_j^2, x_j^3 (see Table 3), $z_j \in \{0, 1, 2, 4, 7\}$, where $j = 1, 2, \dots, 7$
Product family 3	$x_j^1, x_j^2, x_j^3, x_j^4$ (see Table 3), $z_j \in \{0, 1, 2, 4, 7, 8, 12, 16, 18, 28, 32, 33, 42, 52, 63\}$, where $j = 1, 2, \dots, 7$
Three-pane window	$259 \text{ K} < T_{out} < 309 \text{ K}, 0 < U < 21.5 \text{ m/s}, 0 < E_{solar} < 1000 \text{ W/m}^2$

Table 5 Parameter selection for the AHF

Method	Parameter value
E-RBF	$\lambda = 4.75, c = 0.9, t = 2$
RBF	$c = 0.9$
Kriging	$\theta_l = 0.1, \theta_u = 20$
Cross-validation	$q = 5$

B. Performance Criteria

The overall performance of the trained surrogate can be evaluated using both global and local error measure metrics. The most prominent approaches [2] include 1) split sample, 2) cross-validation, and 3) bootstrapping.

1. Split Sample

In a split sample strategy, the sample points are divided into training and test points. The former is used to construct the surrogate, and the latter is used to test the performance of the surrogate. The overall performance of the surrogate can be evaluated using three standard performance metrics: 1) root mean square error (RMSE) [8,41], which provides a global error measure over the entire design domain; 2) maximum absolute error (MAE), which is indicative of local deviations; and 3) relative accuracy error (RAE), which is also indicative of local deviations. To compare the results of different problems, we normalize the RMSE and the MAE measures using the actual function values. The RMSE is given by

$$\text{RMSE} = \sqrt{\frac{1}{n_t} \sum_{k=1}^{n_t} (f(x^k) - \tilde{f}(x^k))^2} \quad (25)$$

where $f(x^k)$ represents the actual function value at the test point x^k , $\tilde{f}(x^k)$ is the corresponding function value estimated by the surrogate, and n_t is the number of test points chosen for evaluating the error measure. The normalized RMSE (NRMSE) is given by

$$\text{NRMSE} = \sqrt{\frac{\sum_{k=1}^{n_t} (f(x^k) - \tilde{f}(x^k))^2}{\sum_{k=1}^{n_t} (f(x^k))^2}} \quad (26)$$

The MAE and NMAE are expressed as

$$\text{MAE} = \max_k |f(x^k) - \tilde{f}(x^k)| \quad (27)$$

$$\text{NMAE} = \frac{\max_k |f(x^k) - \tilde{f}(x^k)|}{\sqrt{\frac{1}{n_t} \sum_{k=1}^{n_t} (f(x^k))^2}} \quad (28)$$

The RAE is evaluated at each test point, as given by

$$\text{RAE}(x^k) = \frac{|\tilde{f}(x^k) - f(x^k)|}{f(x^k)} \times 100\% \quad (29)$$

2. Cross-Validation

Cross-validation is a technique that is used to analyze and improve the robustness of a surrogate model. Cross-validation error is the error estimated at a data point, when the response surface is fitted to a subset of the data points not including that point (also called the leave-one-out strategy). A vector of cross-validation errors, \tilde{e} , can be obtained, when the response surfaces are fitted to all other $p-1$ points. This vector is also known as the *prediction sum of squares* (the PRESS vector).

The leave-one-out strategy is computationally expensive for a large number of points, which can be overcome by the q -fold strategy. The q -fold strategy involves 1) splitting the data randomly into q (approximately) equal subsets, 2) removing each of these subsets in turn, and 3) fitting the model to the remaining $q-1$ subsets. A loss function L can be computed to measure the error between the predictor and the points in the subset that are set aside at each iteration; the contributions to L are then summed up over the q iterations.

More formally, when the mapping, $\zeta: 1, \dots, n \rightarrow 1, \dots, q$, describes the allocation of the n training points to one of the q , subsets, and $\hat{f}^{-\zeta(i)}(x)$ (of the predictor) is obtained by removing the subset $\zeta(i)$, the cross-validation measure is given by

$$\text{PRESS}_{\text{SE}} = \frac{1}{n} \sum_{i=1}^n [y^{(i)} - \hat{f}^{-\zeta(i)}(x^{(i)})]^2 \quad (30)$$

Hastie et al. [42] recommended using compromise values of $q = 5$ or $q = 10$. Using fewer subsets generally has an additional advantage of reducing the computational cost of the cross-validation process by reducing the number of models that have to be fitted.

C. Results and Discussion

In this section, the effectiveness of the AHF surrogate is explored by investigating: 1) the surrogate accuracy, 2) the impact of sampling technique and sample size; 3) the impact of problem dimensionality, and 4) the computational cost.

1. Surrogate Model Accuracy

The surrogate modeling results of the wind farm power generation, the product family, and the three-pane window are given in Tables A1–A5 (in the Appendix). We compare the values of the RMSE, MAE, and PRESS. The maximum and minimum values of the RMSE, MAE, and PRESS for each problem are presented in boldface and with underscore, respectively. The measured values with the LHS method are the average of 100 different experimental designs. The NRMSE and NMAE are shown in Fig. 5. In the case of wind farm power generation, we averaged the RMSE and MAE values of all four wind farms to get the average values of the NRMSE and NMAE. For the product platform planning problem, the average of the RMSE and MAE values of the 12 cases is presented. In Figs. 5a and 5b, each bar represents the NRMSE (or NMAE) value estimated for a problem when specific sampling methods and sample sizes are

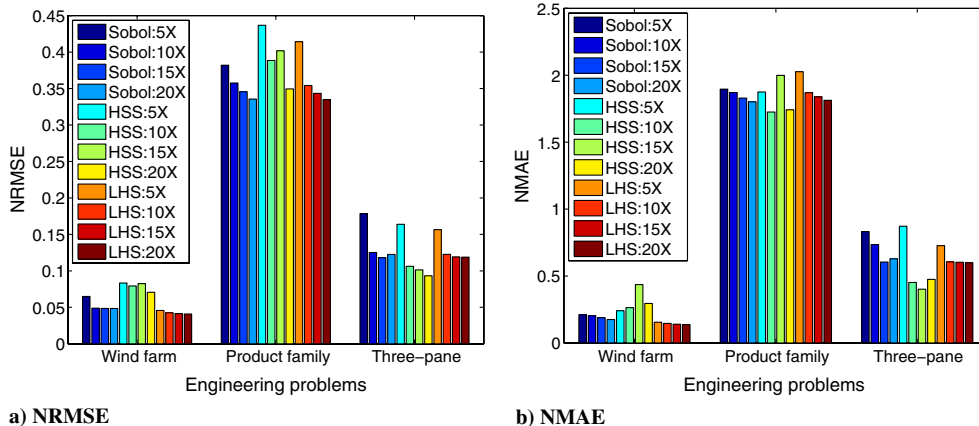


Fig. 5 Normalized error measures for the wind farm, product family, and three-pane window problems.

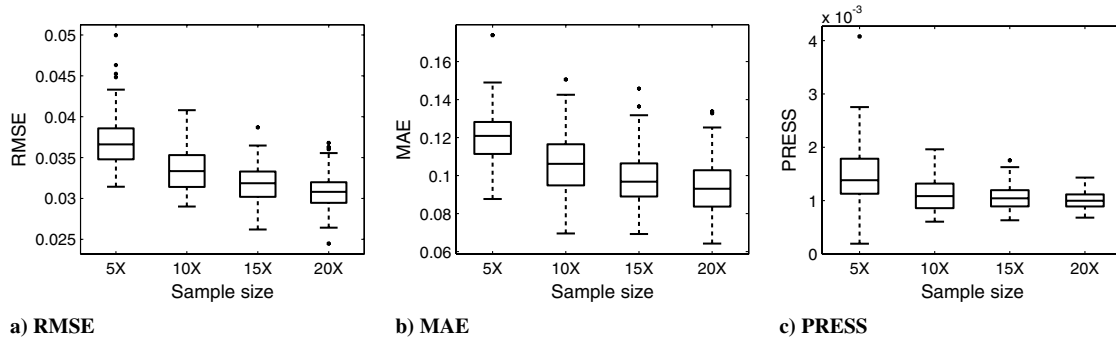


Fig. 6 Box plots of RMSE, MAE, and PRESS of surrogates for the four-turbine wind farm problem (100 experimental designs).

used. We make the following important observations from Tables A1–A5 and Fig. 5:

1) For the wind farm power generation problem (for 4, 9, 16, and 25 turbines), the maximum RMSE is 0.0709, the maximum MAE is 0.4725, and the maximum PRESS is 0.0048.

2) In the case of the product platform planning problem (for two, three, four products), the surrogates representing the first constraint function have relatively larger RMSE, MAE, and PRESS values than other surrogates representing the performance objective, the CI function, and the other constraint functions. This deviation can be attributed to the high nonlinearity of the first constraint function.

3) The NRMSE and NMAE values (in Fig. 5) show that the accuracy of the AHF surrogate for the wind farm and the three-pane window problems is higher than that for the product family problem.

Box plots are used to illustrate the variations of the 100 LHS experimental designs. Figures 6–8 show the box plots of the three performance metrics (RMSE, MAE, and PRESS) of surrogates for the three engineering problems (four-turbine wind farm, product family with two products, and three-pane window).

In a box plot, 1) the central mark is the median; 2) the edges of the box are the 25th and 75th percentiles; 3) the ends of the vertical lines indicate the minimum and maximum data values, or 1.5 times of the interquartile in each direction if the limit of the data fall beyond 1.5 times of the interquartile range; and 4) the points outside the ends of the lines are outliers.

2. Effect of Sampling Technique and Sample Size

Figures 9–11 illustrate the variations in the average RMSE, MAE, and PRESS values (across all functions and sampling techniques) as a function of the corresponding sample sizes. In the case of wind farm power generation, we averaged the values of RMSE, MAE, and PRESS computed for the four wind farms, when any particular sample size is used. The standard deviation of each error metric is also computed. The error bar at each point is 25% of the corresponding standard deviation value. For the product platform planning problem, the average RMSE, MAE, and PRESS values among the 12 cases are presented for this purpose. In this case, the error bar at each point is 2% of the corresponding standard deviation value. Expectedly, it is observed that the accuracies of the three problems are improved with an increase in the sample size; however, beyond a certain sample size, the improvement is marginal.

1) In the case of the three-pane window, it can be seen from Figs. 9a and 9b that the HSS technique performs better than the LHS and Sobol’s techniques.

2) For the wind farm power generation model, the LHS technique performs better than the other two sampling techniques (Figs. 10a and 10b).

3) For the product family problem (with universal motors), a conclusive comparison of the sampling technique performances could not be readily accomplished. In terms of the RMSE values

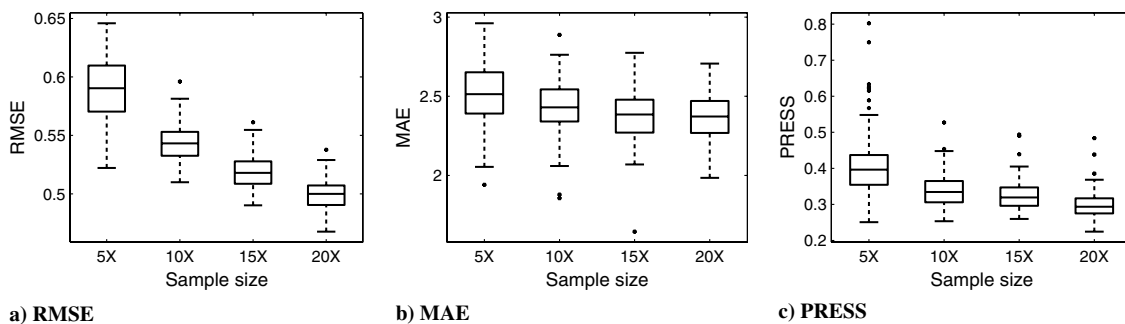


Fig. 7 Box plots of RMSE, MAE, and PRESS of surrogates for the objective of two-products family problem (100 experimental designs).

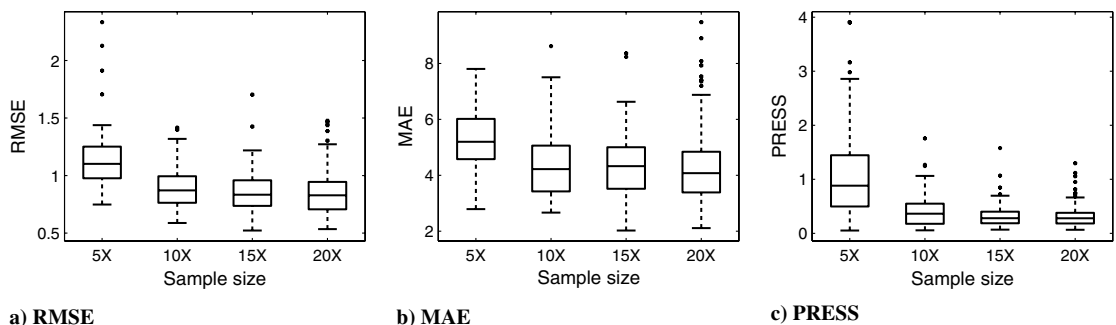


Fig. 8 Box plots of RMSE, MAE, and PRESS of surrogates for the three-pane window problem (100 experimental designs).

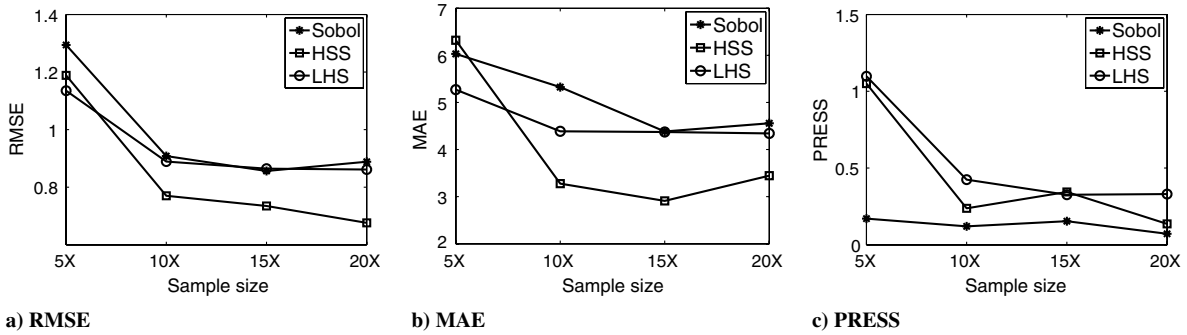


Fig. 9 Effect of sample technique and size on surrogate modeling accuracy for three-pane window model.

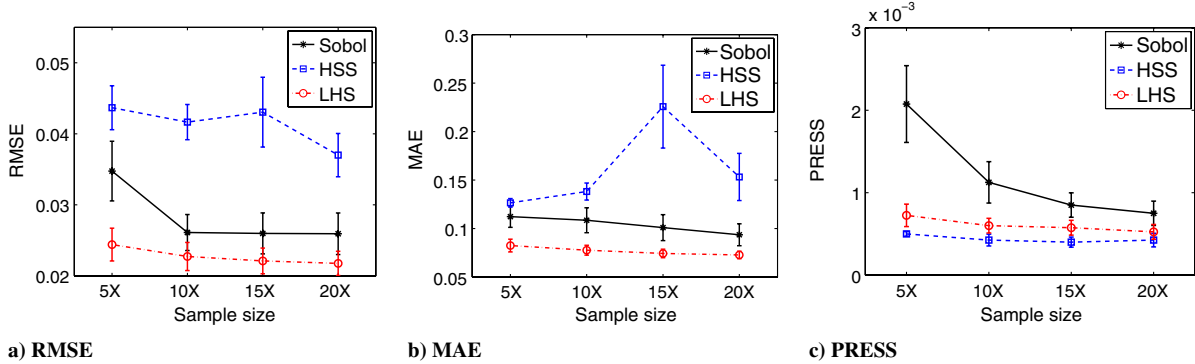


Fig. 10 Effect of sample technique and size on surrogate modeling accuracy for wind power problem.

(Fig. 11a), both the LHS and Sobol provide good performance. However, the HSS yields better (smaller) MAE values (Fig. 11b).

We observe that the standard deviation values of RMSE, MAE, and PRESS for the wind farm power generation and product family problems are relatively large. Such large standard deviation values can be attributed to the significant difference in the number of input variables.

3. Effect of Problem Dimensionality

We comment on the effect of problem dimensionality on the accuracy of the hybrid surrogate. For each wind farm (with different numbers of turbines), the mean and standard deviation values of RMSE, MAE, and PRESS for all four sampling sizes are estimated. Figure 12 shows the effect of increasing problem dimensionality on

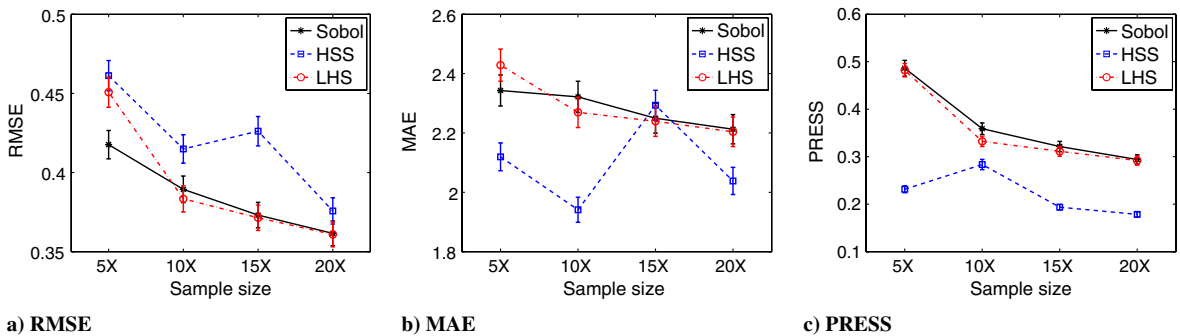


Fig. 11 Effect of sample technique and size on surrogate modeling accuracy for product family model.

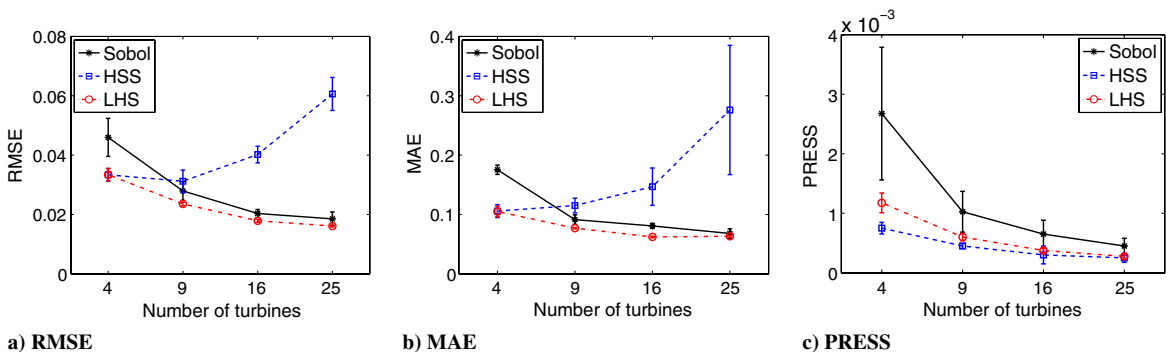


Fig. 12 Effect of problem dimensionality on surrogate modeling accuracy for wind power problems.

Table 6 Computational time of steps A–C in the AHF for the three-pane window problem

Sample size	Step A	Step B	Step C
15	43%	54%	3%
30	40%	58%	2%
40	35%	63%	2%
60	30%	69%	1%

Table 7 Experimental designs for the wind farm cost estimation

Parameter	Value
No. of variables	2
No. of training points	60
No. of test points	15
RMSE	0.2470
MAE	0.5916
PRESS	0.9565

the corresponding surrogate model performance. The mean values are represented by different markers. The error bar at each point is 75% of the standard deviation value. The number of input variables increases proportionally, when the number of turbines increases from 4 to 25. It is observed from Fig. 12 that:

- 1) In the case of the Sobol’s and the LHS sampling techniques, the values of RMSE, MAE, and PRESS decrease when the dimension increases.
- 2) In the case of the HSS technique, the AHF method has a high accuracy for the relatively lower dimensional problems (Figs. 12a and 12b).

4. Computational Cost

The total computational cost for constructing the AHF surrogate is higher than that for constructing the component surrogates (Kriging, RBF, or E-RBF), while the computational cost in terms of number of actual system evaluations is the same. Hence, the increase in the accuracy of the estimated output accomplished by the AHF approach does not demand an appreciable increase in the computational cost. Such positive attributes further illustrate the applicability of the AHF method to model a wide variety of complex systems.

The AHF surrogate is formulated by following three major steps (steps A–C in Fig. 1). To further investigate the computational cost, we compare the computational time of the three steps. Table 6 shows the computational time of the three steps for the three-pane window model. We observe that steps A and B take up most of the computational time. It is also observed that when the number

of training points increases, the relative computational time of step A decreases, and the relative computational time of step B increases.

5. Response Surface-Based Wind Farm Cost Model

The RS-WFC model development is not preceded by the design of experiments. There are 60 training points and 15 test points for the development of the wind farm cost model (Table 7). The values of RMSE, MAE, and PRESS are shown in Table 7. The PRESS value is estimated using all the training points and test points. The RAE value results show that the largest RAE value is approximately 0.48%, and the smallest RAE value is approximately 0.01% (estimated at test point 8). This example shows that the AHF surrogate can also be successfully employed to develop relationships between parameters using recorded or measured data.

VI. Conclusions

This paper presented applications of the Adaptive Hybrid Functions (AHF) to model complex engineering and economic systems. Three representative sampling techniques (Latin hypercube sampling [LHS], Sobol’s quasi-random sequence, and the Hammersley sequence sampling [HSS]) were selected to study their impact on the quality of the resulting surrogates. The influences of sample size and problem dimensionality on the performance of the resulting surrogate model was also investigated.

The results show that the accuracy of the surrogate model generally improves with an increase in the sample size (which is expected); however, a few exceptions to this trend was also observed. The AHF method maintains a relatively higher accuracy for high-dimensional problems when Sobol’s and the LHS sampling techniques are used, whereas the AHF method maintains a relatively higher accuracy for low-dimensional problems when the HSS technique is used. In addition, the increased accuracy of the AHF approach is not accompanied by a significant increase in computational cost. These case studies successfully establish the wide applicability of the AHF method.

The problems and challenges involved in the four problems are representative, and similarities can be found between the four problems and other engineering problems, such as in aerospace and automotive engineering. The implementation of surrogate-based optimization in other complex engineered and economic systems, using these surrogate models, should further establish the true potential of the AHF method. In addition, an exploration of how sensitive the surrogate accuracy is to the user-defined parameters in the individual surrogates and the AHF settings would also be an important topic for future work.

Table A1 Results of the AHF surrogate modeling for the wind farm power generation model

Problem	No. of training points	Sobol			HSS			LHS		
		RMSE	MAE	PRESS	RMSE	MAE	PRESS	RMSE	MAE	PRESS
Wind farm 1 4 Turbines	$5 \times n_{dim}$	0.0587	0.1774	0.0048	0.0356	0.1204	<u>0.0006</u>	0.0373	0.1208	0.0015
	$10 \times n_{dim}$	0.0403	0.1843	0.0026	0.0360	0.0950	0.0008	0.0336	0.1066	0.0011
	$15 \times n_{dim}$	0.0423	0.1797	0.0017	<u>0.0308</u>	0.1159	0.0007	0.0319	0.0987	0.0011
	$20 \times n_{dim}$	0.0426	0.1601	0.0016	0.0310	<u>0.0915</u>	0.0009	0.0309	0.0938	0.0010
Wind farm 2 9 Turbines	$5 \times n_{dim}$	0.0342	0.1034	0.0017	0.0340	0.1155	<u>0.0004</u>	0.0247	0.0790	0.0007
	$10 \times n_{dim}$	0.0264	0.0959	0.0009	0.0368	0.1379	0.0005	0.0235	0.0773	0.0006
	$15 \times n_{dim}$	0.0257	0.0875	0.0008	0.0278	0.1062	0.0005	0.0231	0.0760	0.0005
	$20 \times n_{dim}$	0.0254	0.0785	0.0007	0.0262	0.1006	<u>0.0004</u>	<u>0.0229</u>	<u>0.0752</u>	0.0006
Wind farm 3 16 Turbines	$5 \times n_{dim}$	0.0230	0.0871	0.0011	0.0441	0.1174	0.0006	0.0188	0.0644	0.0004
	$10 \times n_{dim}$	0.0199	0.0831	0.0006	0.0373	0.1387	<u>0.0002</u>	0.0178	0.0626	0.0004
	$15 \times n_{dim}$	0.0193	0.0755	0.0005	0.0427	0.2084	<u>0.0002</u>	0.0175	<u>0.0605</u>	0.0004
	$20 \times n_{dim}$	0.0191	0.0772	0.0004	0.0367	0.1229	<u>0.0002</u>	<u>0.0174</u>	0.0611	0.0003
Wind farm 4 25 Turbines	$5 \times n_{dim}$	0.0231	0.0814	0.0007	0.0610	0.1529	0.0004	0.0168	0.0657	0.0003
	$10 \times n_{dim}$	0.0178	0.0710	0.0004	0.0565	0.1812	<u>0.0002</u>	0.0160	0.0642	0.0003
	$15 \times n_{dim}$	0.0166	0.0612	0.0004	0.0709	0.4725	<u>0.0002</u>	0.0159	0.0622	0.0003
	$20 \times n_{dim}$	0.0166	<u>0.0585</u>	0.0003	0.0541	0.2978	<u>0.0002</u>	<u>0.0158</u>	0.0610	<u>0.0002</u>

Table A2 Results of the AHF surrogate modeling for the two-products family problem

Problem	No. of training points	Sobol			HSS			LHS		
		RMSE	MAE	PRESS	RMSE	MAE	PRESS	RMSE	MAE	PRESS
Two products Objective	$5 \times n_{\text{dim}}$	0.5814	2.4371	0.4256	0.6249	<u>1.9922</u>	0.2439	0.5898	2.5148	0.4111
	$10 \times n_{\text{dim}}$	0.5464	2.4196	0.3220	0.4660	2.2218	0.3432	0.5439	2.4338	0.3371
	$15 \times n_{\text{dim}}$	0.5252	2.4443	0.2950	0.4547	2.1015	0.2761	0.5190	2.3721	0.3253
	$20 \times n_{\text{dim}}$	0.5127	2.4201	0.2578	<u>0.4150</u>	2.1498	<u>0.2222</u>	0.5002	2.3646	0.2984
Two products Commonality	$5 \times n_{\text{dim}}$	0.0700	0.1813	0.0044	0.0489	<u>0.1241</u>	0.0038	0.0706	0.1772	0.0044
	$10 \times n_{\text{dim}}$	0.0686	0.1703	0.0039	0.0409	0.1329	0.0011	0.0688	0.1701	0.0040
	$15 \times n_{\text{dim}}$	0.0685	0.1691	0.0038	0.0396	0.1331	0.0007	0.0685	0.1678	0.0039
	$20 \times n_{\text{dim}}$	0.0681	0.1653	0.0038	<u>0.0359</u>	0.1309	<u>0.0005</u>	0.0678	0.1651	0.0038
Two products Constraint 1	$5 \times n_{\text{dim}}$	1.3669	7.3977	2.7425	1.4026	6.6388	1.2698	1.3476	7.4885	2.0555
	$10 \times n_{\text{dim}}$	1.2887	7.7741	2.0228	1.3560	<u>6.0880</u>	1.8955	1.2870	7.5353	1.8253
	$15 \times n_{\text{dim}}$	1.2572	7.1684	1.7952	1.2995	7.2284	0.9545	1.2500	7.4374	1.7262
	$20 \times n_{\text{dim}}$	1.2331	7.2238	1.6565	1.2534	6.7623	<u>0.8641</u>	<u>1.2174</u>	7.3513	1.6105
Two products Constraint 2	$5 \times n_{\text{dim}}$	0.0507	0.1586	0.0040	0.0629	0.1762	0.0028	0.0538	0.1845	0.0032
	$10 \times n_{\text{dim}}$	0.0492	0.1531	0.0029	0.0516	0.1688	0.0025	0.0506	0.1774	0.0028
	$15 \times n_{\text{dim}}$	0.0479	<u>0.1450</u>	0.0028	0.0485	0.1849	0.0021	0.0488	0.1746	0.0027
	$20 \times n_{\text{dim}}$	0.0463	0.1453	0.0024	<u>0.0460</u>	0.1846	<u>0.0017</u>	0.0474	0.1683	0.0025

Table A3 Results of the AHF surrogate modeling for the three-products family model

Problem	No. of training points	Sobol			HSS			LHS		
		RMSE	MAE	PRESS	RMSE	MAE	PRESS	RMSE	MAE	PRESS
Three products Objective	$5 \times n_{\text{dim}}$	0.6193	5.3032	0.2488	0.7285	4.5999	0.3069	0.8499	5.5758	0.8367
	$10 \times n_{\text{dim}}$	0.5520	4.9811	0.2089	0.5571	4.4519	0.1644	0.5451	4.8623	0.2173
	$15 \times n_{\text{dim}}$	0.5421	4.8942	0.1825	<u>0.5033</u>	<u>3.9131</u>	0.1480	0.5262	4.7993	0.1960
	$20 \times n_{\text{dim}}$	0.5088	4.7585	0.1629	0.5197	4.2039	<u>0.1439</u>	0.5071	4.7278	0.1849
Three products Commonality	$5 \times n_{\text{dim}}$	0.0587	0.1635	0.0028	0.0680	0.2024	0.0016	0.0848	0.2605	0.0088
	$10 \times n_{\text{dim}}$	0.0524	0.1589	0.0026	0.0557	0.1548	<u>0.0007</u>	0.0515	0.1521	0.0023
	$15 \times n_{\text{dim}}$	0.0512	0.1532	0.0024	0.0521	0.1535	<u>0.0007</u>	0.0507	0.1511	0.0022
	$20 \times n_{\text{dim}}$	0.0506	0.1808	0.0024	0.0503	<u>0.1445</u>	<u>0.0007</u>	<u>0.0502</u>	0.1501	0.0021
Three products Constraint 1	$5 \times n_{\text{dim}}$	0.9974	5.6228	1.4164	1.0173	4.7166	0.5065	1.1534	6.0866	1.6415
	$10 \times n_{\text{dim}}$	0.9273	5.3516	0.9859	0.9352	<u>4.2291</u>	0.5825	0.8831	5.3038	0.8942
	$15 \times n_{\text{dim}}$	0.8589	5.3763	0.9515	0.9902	4.9851	<u>0.4871</u>	0.8606	5.2719	0.8171
	$20 \times n_{\text{dim}}$	<u>0.8226</u>	4.8800	0.8331	0.9360	4.8892	0.5234	0.8347	5.0819	0.7717
Three products Constraint 2	$5 \times n_{\text{dim}}$	0.0281	0.0892	0.0011	0.0333	0.1560	<u>0.0004</u>	0.0319	0.1312	0.0012
	$10 \times n_{\text{dim}}$	0.0270	0.1036	0.0009	0.0329	0.1289	0.0005	0.0265	0.1098	0.0008
	$15 \times n_{\text{dim}}$	0.0269	0.1101	0.0008	0.0308	0.1201	0.0005	0.0259	0.1052	0.0007
	$20 \times n_{\text{dim}}$	0.0264	0.1051	0.0008	0.0281	<u>0.1013</u>	<u>0.0004</u>	<u>0.0255</u>	0.1041	0.0007

Table A4 Results of the AHF surrogate modeling for the four-products family model

Problem	No. of training points	Sobol			HSS			LHS		
		RMSE	MAE	PRESS	RMSE	MAE	PRESS	RMSE	MAE	PRESS
Four products Objective	$5 \times n_{\text{dim}}$	0.4305	3.0440	0.2001	0.5203	2.3004	0.1425	0.4097	3.0843	0.1969
	$10 \times n_{\text{dim}}$	0.3922	3.1585	0.1458	0.4824	<u>2.2122</u>	0.0989	0.3810	3.0316	0.1593
	$15 \times n_{\text{dim}}$	0.3730	3.0890	0.1370	0.5018	3.0059	0.0991	0.3664	2.9856	0.1516
	$20 \times n_{\text{dim}}$	0.3602	3.0410	0.1259	0.3788	2.7134	<u>0.0832</u>	<u>0.3544</u>	2.9188	0.1393
Four products Commonality	$5 \times n_{\text{dim}}$	0.0725	0.2511	0.0060	0.0782	0.2156	0.0020	0.0661	0.1956	0.0045
	$10 \times n_{\text{dim}}$	0.0610	0.1852	0.0045	0.0847	0.2653	0.0014	0.0635	0.1858	0.0040
	$15 \times n_{\text{dim}}$	0.0603	<u>0.1704</u>	0.0045	0.0859	0.3494	<u>0.0012</u>	0.0630	0.1841	0.0038
	$20 \times n_{\text{dim}}$	<u>0.0596</u>	0.1734	0.0042	0.0902	0.2354	0.0014	0.0627	0.1792	0.0038
Four products Constraint 1	$5 \times n_{\text{dim}}$	0.7158	3.4054	0.7837	0.9240	4.2043	0.2965	0.7315	3.3778	0.6218
	$10 \times n_{\text{dim}}$	0.6885	3.3460	0.6066	0.8930	3.1446	0.3117	0.6802	3.2089	0.5400
	$15 \times n_{\text{dim}}$	0.6474	3.2213	0.4805	1.0753	5.2102	0.3543	0.6592	3.1552	0.5064
	$20 \times n_{\text{dim}}$	<u>0.6317</u>	3.4038	0.4788	0.7328	<u>2.8579</u>	0.3002	0.6446	3.1720	0.4838
Four products Constraint 2	$5 \times n_{\text{dim}}$	0.0199	0.0630	0.0004	0.0255	0.1065	<u>0.0002</u>	0.0199	0.0647	0.0004
	$10 \times n_{\text{dim}}$	0.0191	0.0562	0.0004	0.0231	0.0946	<u>0.0002</u>	0.0189	0.0623	0.0004
	$15 \times n_{\text{dim}}$	0.0186	0.0540	0.0004	0.0315	0.1349	<u>0.0002</u>	<u>0.0184</u>	0.0602	0.0004
	$20 \times n_{\text{dim}}$	<u>0.0184</u>	<u>0.0532</u>	0.0004	0.0225	0.0845	<u>0.0002</u>	<u>0.0184</u>	0.0631	0.0004

Table A5 Results of the AHF surrogate modeling for the three-pane window model

Problem	No. of training points	Sobol			HSS			LHS		
		RMSE	MAE	PRESS	RMSE	MAE	PRESS	RMSE	MAE	PRESS
Three-pane window	$5 \times n_{\text{dim}}$	1.2942	6.0306	0.1703	1.1885	6.3210	1.0498	1.1352	5.2712	1.0974
	$10 \times n_{\text{dim}}$	0.9081	5.3257	0.1200	0.7702	3.2753	0.2372	0.8891	4.3869	0.4240
	$15 \times n_{\text{dim}}$	0.8562	4.3830	0.1534	0.7348	<u>2.9090</u>	0.3439	0.8644	4.3716	0.3256
	$20 \times n_{\text{dim}}$	0.8884	4.5566	<u>0.0713</u>	<u>0.6761</u>	3.4442	0.1359	0.8613	4.3423	0.3300

Appendix: Results of the AHF Surrogate Modeling

The surrogate modeling results of the wind farm power generation, the product family, and the three-pane window are given in Tables A1–A5.

Acknowledgments

Support from the National Science Foundation (NSF) Awards CMMI-1100948 and CMMI-0946765 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] Braha, D., Minai, A., and Bar-Yam, Y., *Complex Engineered Systems*, Springer-Verlag, New York, 2006, pp. 1–21.
- [2] 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. doi:10.1016/j.paerosci.2005.02.001
- [3] 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. 370–380. doi:10.1115/1.2429697
- [4] Jouhaud, J. C., Sagaut, P., Montagnac, M., and Laurenceau, J., “A Surrogate-Model Based Multidisciplinary Shape Optimization Method with Application to a 2D Subsonic Airfoil,” *Journal of Computers and Fluids*, Vol. 36, No. 3, 2007, pp. 520–529. doi:10.1016/j.compfluid.2006.04.001
- [5] Neufeld, D., and Chung, K. B., “Aircraft Wing Box Optimization Considering Uncertainty in Surrogate Models,” *Structural and Multidisciplinary Optimization*, Vol. 42, No. 5, 2010, pp. 745–753. doi:10.1007/s00158-010-0532-8
- [6] Madsen, J. I., Shyy, W., and Haftka, R. T., “Response Surface Techniques for Diffuser Shape Optimization,” *AIAA Journal*, Vol. 38, No. 9, 2000, pp. 1512–1518. doi:10.2514/2.1160
- [7] Myers, R., and Montgomery, D., *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 2nd ed., Wiley-Interscience, New York, 2002, pp. 13–72.
- [8] Forrester, A., and Keane, A., “Recent Advances in Surrogate-Based Optimization,” *Progress in Aerospace Sciences*, Vol. 45, Nos. 1–3, 2009, pp. 50–79. doi:10.1016/j.paerosci.2008.11.001
- [9] Wilson, B., Cappelleri, D., Simpson, T., and Frecker, M., “Efficient Pareto Frontier Exploration Using Surrogate Approximations,” *Optimization and Engineering*, Vol. 2, No. 1, 2001, pp. 31–50. doi:10.1023/A:1011818803494
- [10] Hardy, R. L., “Multiquadric Equations of Topography and Other Irregular Surfaces,” *Journal of Geophysical Research*, Vol. 76, No. 8, 1971, pp. 1905–1915. doi:10.1029/JB076i008p01905
- [11] Mullur, A., and Messac, A., “Extended Radial Basis Functions: More Flexible and Effective Metamodeling,” *AIAA Journal*, Vol. 43, No. 6, 2005, pp. 1306–1315. doi:10.2514/1.11292
- [12] Messac, A., and Mullur, A., “A Computationally Efficient Metamodeling Approach for Expensive Multiobjective Optimization,” *Optimization and Engineering*, Vol. 9, No. 1, 2008, pp. 37–67. doi:10.1007/s11081-007-9008-0
- [13] Simpson, T., Peplinski, J., Koch, P., and Allen, J., “Metamodels for Computer-Based Engineering Design: Survey and Recommendations,” *Engineering with Computers*, Vol. 17, No. 2, 2001, pp. 129–150. doi:10.1007/PL00007198
- [14] Basudhar, A., and Missoum, S., “Adaptive Explicit Decision Functions for Probabilistic Design and Optimization Using Support Vector Machines,” *Computers and Structures*, Vol. 86, Nos. 19–20, 2008, pp. 1904–1917. doi:10.1016/j.compstruc.2008.02.008
- [15] Yun, Y., Yoon, M., and Nakayama, H., “Multi-Objective Optimization Based on Meta-Modeling by Using Support Vector Regression,” *Optimization and Engineering*, Vol. 10, No. 2, 2009, pp. 167–181. doi:10.1007/s11081-008-9063-1
- [16] Zerpa, L., Queipo, N., Pintos, S., and Salager, J., “An Optimization Methodology of Alkaline-Urfactant-Polymer Flooding Processes Using Field Scale Numerical Simulation and Multiple Surrogates,” *Journal of Petroleum Science and Engineering*, Vol. 47, Nos. 3–4, 2005, pp. 197–208. doi:10.1016/j.petrol.2005.03.002
- [17] Goel, T., Haftka, R., Shyy, W., and Queipo, N., “Ensemble of Surrogates,” *Structural and Multidisciplinary Optimization*, Vol. 33, No. 3, 2007, pp. 199–216. doi:10.1007/s00158-006-0051-9
- [18] Zhang, J., Chowdhury, S., and Messac, A., “An Adaptive Hybrid Surrogate Model,” *Structural and Multidisciplinary Optimization*, Vol. 46, No. 2, 2012, pp. 223–238. doi:10.1007/s00158-012-0764-x
- [19] Zhang, J., Chowdhury, S., Messac, A., Zhang, J., and Castillo, L., “Surrogate Modeling of Complex Systems Using Adaptive Hybrid Functions,” *Proceedings of the ASME 2011 International Design Engineering Technical Conferences (IDETC)*, Vol. 5: 37th Design Automation Conference, Parts A and B, Washington, DC, Aug. 2011, pp. 775–790.
- [20] Sanchez, E., Pintos, S., and Queipo, N., “Toward an Optimal Ensemble of Kernel-Based Approximations with Engineering Applications,” *Structural and Multidisciplinary Optimization*, Vol. 36, No. 3, 2008, pp. 247–261. doi:10.1007/s00158-007-0159-6
- [21] Acar, E., and Rais-Rohani, M., “Ensemble of Metamodels with Optimized Weight Factors,” *Structural and Multidisciplinary Optimization*, Vol. 37, No. 3, 2009, pp. 279–294. doi:10.1007/s00158-008-0230-y
- [22] 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. doi:10.1007/s00158-008-0338-0
- [23] Acar, E., “Various Approaches for Constructing an Ensemble of Metamodels Using Local Measures,” *Structural and Multidisciplinary Optimization*, Vol. 42, No. 6, 2010, pp. 879–896. doi:10.1007/s00158-010-0520-z
- [24] Zhou, X., Ma, Y., and Li, X., “Ensemble of Surrogates with Recursive Arithmetic Average,” *Structural and Multidisciplinary Optimization*, Vol. 44, No. 5, 2011, pp. 651–671. doi:10.1007/s00158-011-0655-6
- [25] Zhang, J., Chowdhury, S., and Messac, A., “A New Robust Surrogate Model: Reliability Based Hybrid Functions,” *52nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, Denver, CO, April 2011.
- [26] Cressie, N., *Statistics for Spatial Data*, Wiley, New York, 1993, pp. 105–210.
- [27] Lophaven, S., Nielsen, H., and Sondergaard, J., “DACE—A Matlab Kriging Toolbox, Version 2.0,” Technical Univ. of Denmark TR-IMM-REP-2002-12, 2002.
- [28] Cherrie, J. B., Beatson, R. K., and Newsam, G. N., “Fast Evaluation of Radial Basis Functions: Methods for Generalized Multiquadrics in R^n ,” *SIAM Journal on Scientific Computing*, Vol. 23, No. 5, 2002, pp. 1549–1571. doi:10.1137/S1064827500367609
- [29] 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. doi:10.1016/j.renene.2011.06.033
- [30] *GE Energy 1.5MW Wind Turbine Brochure*, General Electric, <http://www.gpower.com/>, 2010.
- [31] Simpson, T., Siddique, Z., and Jiao, R., *Product Platform and Product Family Design: Methods and Applications*, Springer-Verlag, New York, 2006.
- [32] Chowdhury, S., Messac, A., and Khire, R. A., “Comprehensive Product Platform Planning (CP³) Framework,” *Journal of Mechanical Design*, Vol. 133, No. 10, 2011, p. 101004. doi:10.1115/1.4004969
- [33] Chowdhury, S., Messac, A., and Khire, R. A., “Comprehensive Product Platform Planning (CP³) Using Mixed-Discrete Particle Swarm Optimization and a New Commonality Index,” *ASME 2012 International Design Engineering Technical Conferences (IDETC)*, Chicago, IL, 12–15 Aug. 2012.
- [34] Zhang, J., Messac, A., Chowdhury, S., and Zhang, J., “Adaptive Optimal Design of Active Thermally Insulated Windows Using Surrogate Modeling,” *AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*, Orlando, FL, 12–15 April 2010.
- [35] Zhang, J., Chowdhury, S., Messac, A., and Castillo, L., “A Response Surface-Based Cost Model for Wind Farm Design,” *Energy Policy*, Vol. 42, 2012, pp. 538–550. doi:10.1016/j.enpol.2011.12.021

- [36] McKay, M., Conover, W., and Beckman, R., "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code," *Technometrics*, Vol. 21, No. 2, 1979, pp. 239–245.
- [37] Sobol, I., "Uniformly Distributed Sequences with an Additional Uniform Property," *USSR Computational Mathematics and Mathematical Physics*, Vol. 16, No. 5, 1976, pp. 236–242. doi:10.1016/0041-5553(76)90154-3
- [38] Bratley, P., and Fox, B., "Algorithm 659: Implementing Sobol's Quasirandom Sequence Generator," *ACM Transactions on Mathematical Software*, Vol. 14, No. 1, 1988, pp. 88–100. doi:10.1145/42288.214372
- [39] Kalagnanam, J., and Diwekar, U., "An Efficient Sampling Technique for Off-Line Quality Control," *Technometrics*, Vol. 39, No. 3, 1997, pp. 308–319. doi:10.1080/00401706.1997.10485122
- [40] Goldberg, M., *Jobs and Economic Development Impact (JEDI) Model*, National Renewable Energy Lab., Golden, CO, Oct. 2009.
- [41] Jin, R., Chen, W., and Simpson, T., "Comparative Studies of Metamodelling Techniques Under Multiple Modelling Criteria," *Structural and Multidisciplinary Optimization*, Vol. 23, No. 1, 2001, pp. 1–13. doi:10.1007/s00158-001-0160-4
- [42] Hastie, T., Tibshirani, R., and Friedman, J., *The Elements of Statistical Learning*, Springer-Verlag, New York, 2001.

J. Martins
Associate Editor