Pitfalls of using a single criterion for selecting experimental designs

Tushar Goel[1, *, †], Raphael T. Haftka[1], Wei Shyy[2] and Layne T. Watson[3]

[1]*Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL 32611, U.S.A.*
[2]*Department of Aerospace Engineering, University of Michigan, Ann Arbor, MI 48109, U.S.A.*
[3]*Departments of Computer Science and Mathematics, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061, U.S.A.*

SUMMARY

For surrogate construction, a good experimental design (ED) is essential to simultaneously reduce the effect of noise and bias errors. However, most EDs cater to a single criterion and may lead to small gains in that criterion at the expense of large deteriorations in other criteria. We use multiple criteria to assess the performance of different popular EDs. We demonstrate that these EDs offer different trade-offs, and that use of a single criterion is indeed risky. In addition, we show that popular EDs, such as Latin hypercube sampling (LHS) and D-optimal designs, often leave large regions of the design space unsampled even for moderate dimensions. We discuss a few possible strategies to combine multiple criteria and illustrate them with examples. We show that complementary criteria (e.g. bias handling criterion for variance-based designs and *vice versa*) can be combined to improve the performance of EDs. We demonstrate improvements in the trade-offs between noise and bias error by combining a model-based criterion, like the D-optimality criterion, and a geometry-based criterion, like LHS. Next, we demonstrate that selecting an ED from three candidate EDs using a suitable error-based criterion helped eliminate potentially poor designs. Finally, we show benefits from combining the multiple criteria-based strategies, that is, generation of multiple EDs using the D-optimality and LHS criteria, and selecting one design using a pointwise bias error criterion. The encouraging results from the examples indicate that it may be worthwhile studying these strategies more rigorously and in more detail. Copyright © 2007 John Wiley & Sons, Ltd.

*Correspondence to: Tushar Goel, Livermore Software Technology Corporation, Livermore, CA 95441, U.S.A.
†E-mail: tusharg@ufl.edu

1. INTRODUCTION AND LITERATURE REVIEW

Polynomial response surface approximations (RSAs) are widely adopted for solving optimization problems with high computational or experimental cost as they offer a computationally less expensive way of evaluating designs. It is important to ensure the accuracy of RSAs before using them for design and optimization. The accuracy of an RSA, constructed using a limited number of simulations, is primarily affected by two factors: (i) noise in the data and (ii) inadequacy of the fitting model (called modeling error or bias error). In experiments, noise may appear due to measurement errors and other experimental errors. Numerical noise in computer simulations is usually small, but it can be high for ill-conditioned problems, or if there are some unconverged solutions such as those encountered in computational fluid dynamics or structural optimization. The true model representing the data is rarely known, and due to limited data available, usually a simple model is fitted to the data. For simulation-based RSAs, modeling/bias error due to an inadequate model is mainly responsible for the error in prediction. Different designs of experiment techniques are widely used to improve the prediction performance of RSAs and other surrogates. There have been some studies on the influence of different experimental designs (EDs) on predictions [1–4].

In the design of experiment techniques, sampling of the points in design space seeks to reduce the effect of noise and reduce bias errors simultaneously. However, these objectives (noise and bias error) often conflict. For example, noise rejection criteria, such as D-optimality, usually produce designs with more points near the boundary, whereas the bias error criteria tend to distribute points more evenly in design space. Thus, the problem of selecting an ED is a multi-objective problem with conflicting objectives (noise and bias error). The solution to this problem would be a Pareto optimal front of EDs that yields different trade-offs between noise and bias errors. Seeking the optimal EDs considering only one criterion, though popular, may yield minor improvements in the selected criterion with significant deterioration in other criteria.

In the past, the majority of the work related to the construction of EDs was done by considering only one design objective [5–22]. When noise is the dominant source of error, there are a number of EDs that minimize the effect of variance (noise) on the resulting approximation, for example, the D-optimal design that minimizes the variance associated with the estimates of coefficients of the response surface model. Traditional variance-based designs minimize the effect of noise and attempt to obtain uniformity (ratio of maximum to minimum error) over the design space, but they do not address bias errors.

Classical minimum-bias designs consider only space-averaged or integrated error measures [5] in EDs. The bias component of the averaged or integrated mean-squared error was minimized to obtain the so-called minimum-bias designs. The fundamentals of minimizing integrated mean-squared error and its components can be found in Myers and Montgomery [5, Chapter 9] and Khuri and Cornell [6, Chapter 6]. Venter and Haftka [7] developed an algorithm implementing a minimum-bias criterion for irregularly shaped design spaces where no closed-form solution exists for ED. They compared minimum-bias and D-optimal EDs for two problems with two and three variables. The minimum-bias ED was found to be more accurate than D-optimal for average error but not for maximum error. Qu *et al*. [8] implemented Gaussian quadrature-based minimum-bias design and presented minimum-bias central composite designs (CCDs) for up to six variables.

There are some works reported on developing EDs by minimizing the integrated mean-squared error accounting for both variance and bias errors [9–12]. Box and Draper [9] minimized integrated mean-squared errors averaged over the design space by combining average weighted variance and average bias errors. Draper and Lawrence [10] minimized the integrated mean-squared error to

account for model inadequacies. Kupper and Meydrech [11] specified bounds on the coefficients associated with the assumed true function to minimize the integrated mean-squared error. Welch [12] used a linear combination of variance and bias errors to minimize the mean-squared error. Montepiedra and Fedorov [13] investigated EDs minimizing the bias component of the integrated mean-squared error subject to a constraint on the variance component or *vice versa*. Fedorov *et al*. [14] later studied design of experiments *via* weighted regression prioritizing regions where the approximation is needed to predict the response. Their approach considered both variance and bias components of the estimation error.

Bias error averaged over the design space has been studied extensively, but there is a relatively small amount of work to account for pointwise (spatial) variation of bias errors because of inherent difficulties. An approach for estimating bounds on bias errors in RSAs by a pointwise decomposition of the mean-squared error into variance and the square of bias was developed by Papila and Haftka [15]. They used the bounds to obtain EDs that minimize the maximum absolute bias error. Papila *et al*. [16] extended the approach to account for the data and proposed data-dependent bounds. They assumed that the true model is a higher degree polynomial than the approximating polynomial and that it satisfies the given data exactly. Goel *et al*. [17] generalized this bias error bounds estimation method to account for inconsistencies between the assumed true model and actual data. They demonstrated that the bounds can be used to develop adaptive EDs to reduce the effect of bias errors in the region of interest. Recently, Goel *et al*. [18] presented a method to estimate pointwise root mean square (RMS) bias errors in approximation prior to the generation of data. They applied this method to construct EDs that minimize maximum RMS bias error (min–max RMS bias designs).

Since minimum-bias designs do not achieve uniformity, designs that distribute points uniformly in design space (space-filling designs like Latin hypercube sampling (LHS)) are popular even though these designs have no claim to optimality. Since LHS designs can create poor designs, as illustrated by Leary *et al*. [22], different criteria such as maximization of minimum distance between points or minimization of correlation between points are used to improve its performance. We will demonstrate in this paper that even optimized LHS designs can occasionally leave large holes in design space, which may lead to poor predictions. Thus, there is a need to consider multiple criteria. Some previous efforts of considering multiple criteria are as follows. In an effort to account for variance, Tang [19] and Ye [20] presented orthogonal array-based LHS designs that were shown to be better than the conventional LHS designs. Leary *et al*. [22] presented strategies to find optimal orthogonal array-based LHS designs in a more efficient manner. Palmer and Tsui [21] generated minimum-bias Latin hypercube EDs for sampling from deterministic simulations by minimizing integrated squared bias error. A combination of face-centered cubic design (FCCD) and LHS designs is quite widely used [23].

The primary objective of this paper is to demonstrate the risks associated with using a single criterion to construct EDs. Firstly, we compare LHS and D-optimal designs and demonstrate that both these designs can leave large unsampled regions in design space that may potentially yield high errors. Besides, we illustrate the need to consider multiple criteria to construct EDs, as single criterion-based designs may represent extreme trade-offs among different criteria. Min–max RMS bias design, which yields a small reduction in the maximum-bias error at the cost of a huge increase in the maximum variance, is used as an example. While the above issue of trade-off among multiple criteria requires significant future research effort, we explore several strategies for the simultaneous use of multiple criteria to guard against selecting EDs that are optimal according to one criterion but may yield very poor performance on other criteria. In this context, we firstly

discuss which criteria can be simultaneously used meaningfully; secondly, we explore how to combine different criteria. We show that complementary criteria may cater to competing needs of EDs. Next, we demonstrate improvements by combining a geometry-based criterion (LHS) and a model-based D-optimality criterion, to obtain EDs. We also show that poor EDs can be filtered out by creating multiple EDs and selecting one of them using an appropriate error-based (pointwise) criterion. Finally, we combine the above-mentioned strategies to construct EDs.

This paper is organized as follows: Section 2 summarizes different error metrics used in this paper. Major results of this study are given in Section 3. We illustrate the issues associated with single-criterion-based EDs and show a few strategies to accommodate multiple criteria. We recapitulate the major findings of this work in Section 4.

2. ERROR MEASURES FOR EDs

Let the true response $\eta(\mathbf{x})$ at a design point $\mathbf{x}$ be represented by a polynomial $\mathbf{f}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta}$, where $\mathbf{f}(\mathbf{x})$ is the vector of basis functions and $\boldsymbol{\beta}$ is the vector of coefficients. The vector $\mathbf{f}(\mathbf{x})$ has two components: $\mathbf{f}^{(1)}(\mathbf{x})$ is the vector of basis functions used in the RSA or fitting model and $\mathbf{f}^{(2)}(\mathbf{x})$ is the vector of additional basis functions that are missing in the linear regression model (assuming that the true model is a polynomial). Similarly, the coefficient vector $\boldsymbol{\beta}$ can be expressed as a combination of vectors $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$ that represent the true coefficients associated with the basis function vectors $\mathbf{f}^{(1)}(\mathbf{x})$ and $\mathbf{f}^{(2)}(\mathbf{x})$, respectively. Precisely,

$$\eta(\mathbf{x}) = \mathbf{f}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta} = [\mathbf{f}^{(1)\mathrm{T}} \, \mathbf{f}^{(2\mathrm{T})}] \begin{bmatrix} \boldsymbol{\beta}^{(1)} \\ \boldsymbol{\beta}^{(2)} \end{bmatrix} = (\mathbf{f}^{(1)}(\mathbf{x}))^{\mathrm{T}}\boldsymbol{\beta}^{(1)} + (\mathbf{f}^{(2)}(\mathbf{x}))^{\mathrm{T}}\boldsymbol{\beta}^{(2)} \tag{1}$$

Assuming normally distributed noise ε with zero mean and variance σ^2 ($N(0, \sigma^2)$), the observed response $y(\mathbf{x})$ at a design point $\mathbf{x}$ is given as

$$y(\mathbf{x}) = \eta(\mathbf{x}) + \varepsilon \tag{2}$$

The predicted response $\hat{y}(\mathbf{x})$ at a design point $\mathbf{x}$ is given as a linear combination of approximating basis functions vector $\mathbf{f}^{(1)}(\mathbf{x})$ with the corresponding estimated coefficient vector $\hat{\boldsymbol{\beta}}^{(1)}$:

$$\hat{y}(\mathbf{x}) = (\mathbf{f}^{(1)}(\mathbf{x}))^{\mathrm{T}}\hat{\boldsymbol{\beta}}^{(1)} \tag{3}$$

The estimated coefficient vector $\hat{\boldsymbol{\beta}}^{(1)}$ is evaluated using the data ($\mathbf{y}$) for N_s design points as [5, Chapter 2]:

$$\hat{\boldsymbol{\beta}}^{(1)} = (X^{(1)\mathrm{T}}X^{(1)})^{-1}X^{(1)\mathrm{T}}\mathbf{y} \tag{4}$$

where $X^{(1)}$ is the Gramian matrix constructed using $\mathbf{f}^{(1)}(\mathbf{x})$.

The error at a general design point $\mathbf{x}$ is the difference between the true response and the predicted response, $e(\mathbf{x}) = \eta(\mathbf{x}) - \hat{y}(\mathbf{x})$. When noise is dominant, standard error $e_{\mathrm{es}}(\mathbf{x})$, used to appraise error, is given as [5]

$$e_{\mathrm{es}}(\mathbf{x}) = \sqrt{\mathrm{var}[\hat{y}(\mathbf{x})]} = \sqrt{\sigma_a^2 \mathbf{f}^{(1)\mathrm{T}}(\mathbf{x})(X^{(1)\mathrm{T}}X^{(1)})^{-1}\mathbf{f}^{(1)}(\mathbf{x})} \tag{5}$$

where σ_a^2 is the estimated variance of the noise.

If there is no noise, the true response $\eta(\mathbf{x})$ is the same as the observed response $y(\mathbf{x})$. Then the true response for N_s design points $(\mathbf{x}^{(i)},\ i = 1, \ldots, N_s)$ in matrix notation is

$$\mathbf{y} = X\boldsymbol{\beta} = [X^{(1)}\ X^{(2)}] \begin{bmatrix} \boldsymbol{\beta}^{(1)} \\ \boldsymbol{\beta}^{(2)} \end{bmatrix} = X^{(1)}\boldsymbol{\beta}^{(1)} + X^{(2)}\boldsymbol{\beta}^{(2)} \tag{6}$$

where $X^{(2)}$ is constructed using the missing basis functions corresponding to $\mathbf{f}^{(2)}(\mathbf{x})$. Substituting for $\mathbf{y}$ from Equation (6) in Equation (4) gives

$$\hat{\boldsymbol{\beta}}^{(1)} = (X^{(1)\mathrm{T}}X^{(1)})^{-1}X^{(1)\mathrm{T}}[X^{(1)}\boldsymbol{\beta}^{(1)} + X^{(2)}\boldsymbol{\beta}^{(2)}] \tag{7}$$

that can be rearranged as

$$\boldsymbol{\beta}^{(1)} = \hat{\boldsymbol{\beta}}^{(1)} - A\boldsymbol{\beta}^{(2)} \quad \text{where } A = (X^{(1)\mathrm{T}}X^{(1)})^{-1}X^{(1)\mathrm{T}}X^{(2)} \tag{8}$$

is called the alias matrix. Note that this relation is valid only if Equation (6) is satisfied (no noise).

When bias error is dominant, $e(\mathbf{x}) \approx e_{\mathrm{b}}(\mathbf{x})$, where $e_{\mathrm{b}}(\mathbf{x})$ is the bias error at design point $\mathbf{x}$. Substituting values from Equations (1) and (4) give

$$e_{\mathrm{b}}(\mathbf{x}) \approx \eta(\mathbf{x}) - \hat{y}(\mathbf{x}) = (\mathbf{f}^{(1)}(\mathbf{x}))^{\mathrm{T}}\boldsymbol{\beta}^{(1)} + (\mathbf{f}^{(2)}(\mathbf{x}))^{\mathrm{T}}\boldsymbol{\beta}^{(2)} - (\mathbf{f}^{(1)}(\mathbf{x}))^{\mathrm{T}}\hat{\boldsymbol{\beta}}^{(1)} \tag{9}$$

When data are consistent with the assumed true model, Equation (9) can be rearranged as

$$e_{\mathrm{b}}(\mathbf{x}) \approx [(\mathbf{f}^{(2)}(\mathbf{x})) - A^{\mathrm{T}}(\mathbf{f}^{(1)}(\mathbf{x}))]^{\mathrm{T}}\boldsymbol{\beta}^{(2)} \tag{10}$$

We assume that $\boldsymbol{\beta}$ is a random vector. Then the RMS of bias error $e_{\mathrm{b}}^{\mathrm{rms}}(\mathbf{x})$ at design point $\mathbf{x}$ is obtained by computing its L_2 norm $\sqrt{E_\beta(e_{\mathrm{b}}^2(\mathbf{x}))}$ ($E_\beta(\mathbf{x})$ denotes the expected value with respect to β), where

$$E_\beta(e_{\mathrm{b}}^2(\mathbf{x})) = E_\beta(e_{\mathrm{b}}(\mathbf{x})e_{\mathrm{b}}^{\mathrm{T}}(\mathbf{x})) \approx E_\beta([\mathbf{f}^{(2)\mathrm{T}}(\mathbf{x}) - \mathbf{f}^{(1)\mathrm{T}}(\mathbf{x})A]\boldsymbol{\beta}^{(2)}([\mathbf{f}^{(2)\mathrm{T}}(\mathbf{x}) - \mathbf{f}^{(1)\mathrm{T}}(\mathbf{x})A]\boldsymbol{\beta}^{(2)})^{\mathrm{T}}) \tag{11}$$

Note that the term $[\mathbf{f}^{(2)}(\mathbf{x}) - A^{\mathrm{T}}\mathbf{f}^{(1)}(\mathbf{x})]$ depends only on the design point $\mathbf{x}$ and the ED; hence, the above equation can be rewritten as

$$E_\beta(e_{\mathrm{b}}^2(\mathbf{x})) \approx E_\beta([\mathbf{f}^{(2)\mathrm{T}}(\mathbf{x}) - \mathbf{f}^{(1)\mathrm{T}}(\mathbf{x})A]\boldsymbol{\beta}^{(2)}\boldsymbol{\beta}^{(2)\mathrm{T}}[\mathbf{f}^{(2)}(\mathbf{x}) - A^{\mathrm{T}}\mathbf{f}^{(1)}(\mathbf{x})])$$

$$= [\mathbf{f}^{(2)\mathrm{T}}(\mathbf{x}) - \mathbf{f}^{(1)\mathrm{T}}(\mathbf{x})A]E_\beta(\boldsymbol{\beta}^{(2)}\boldsymbol{\beta}^{(2)\mathrm{T}})[\mathbf{f}^{(2)}(\mathbf{x}) - A^{\mathrm{T}}\mathbf{f}^{(1)}(\mathbf{x})] \tag{12}$$

Then,

$$e_{\mathrm{b}}^{\mathrm{rms}}(\mathbf{x}) = \sqrt{E_\beta(e_{\mathrm{b}}^2(\mathbf{x}))} \approx \sqrt{[\mathbf{f}^{(2)\mathrm{T}}(\mathbf{x}) - \mathbf{f}^{(1)\mathrm{T}}(\mathbf{x})A]E_\beta(\boldsymbol{\beta}^{(2)}\boldsymbol{\beta}^{(2)\mathrm{T}})[\mathbf{f}^{(2)}(\mathbf{x}) - A^{\mathrm{T}}\mathbf{f}^{(1)}(\mathbf{x})]} \tag{13}$$

Equation (13) shows that the RMS bias error at any point depends on the coefficient vector $\boldsymbol{\beta}^{(2)}$. This expression to estimate pointwise RMS bias error is easy to evaluate once the true polynomial basis $\mathbf{f}(\mathbf{x})$ and the distribution of coefficient vector $\boldsymbol{\beta}^{(2)}$ are known.

Prior to generation of data, the order of the fitting polynomial and the true model are assumed to be known, but the coefficient vector $\boldsymbol{\beta}^{(2)}$ is unknown. With no data on the distribution, the principle of maximum entropy [24] was followed by assuming that all components of the coefficient vector $\boldsymbol{\beta}^{(2)}$ have a uniform distribution between $-\gamma$ and γ (γ is a constant). Then, with simple algebra, it can be shown that $E_\beta(\boldsymbol{\beta}^{(2)}\boldsymbol{\beta}^{(2)\mathrm{T}}) = \gamma^2/3 I$, where I is an $n_2 \times n_2$ identity matrix. Substituting this in Equation (13), the pointwise RMS bias error is

$$e_{\mathrm{b}}^{\mathrm{rms}}(\mathbf{x}) \approx \sqrt{[\mathbf{f}^{(2)\mathrm{T}}(\mathbf{x}) - \mathbf{f}^{(1)\mathrm{T}}(\mathbf{x})A](\gamma^2/3I)[\mathbf{f}^{(2)}(\mathbf{x}) - A^\mathrm{T}\mathbf{f}^{(1)}(\mathbf{x})]}$$

$$= \gamma/\sqrt{3}\sqrt{[\mathbf{f}^{(2)\mathrm{T}}(\mathbf{x}) - \mathbf{f}^{(1)\mathrm{T}}(\mathbf{x})A][\mathbf{f}^{(2)}(\mathbf{x}) - A^\mathrm{T}\mathbf{f}^{(1)}(\mathbf{x})]} \tag{14}$$

It is clear from Equation (14) that the RMS bias error at a design point $\mathbf{x}$ can be obtained from the location of data points (defines alias matrix A) used to fit the response surface, the form ($\mathbf{f}_1$ and $\mathbf{f}_2$) of the assumed true function (which is a higher-order polynomial than the approximating polynomial), and the constant γ. Since γ has a uniform scaling effect, prior to generation of data it can be taken as unity for the sake of simplicity.

Goel *et al.* [18] demonstrated with examples that this error prediction model gives good estimates of actual error fields both when the true function is polynomial and when it is non-polynomial. Two representative examples of a polynomial true function and a non-polynomial true function (trigonometric function with multiple frequencies) are presented in Appendices A and B, respectively.

Many criteria used to construct/compare different EDs are presented in the literature. A few commonly used error metrics as well as new bias error metrics are listed below.

1. Maximum standard error in the design space (Equation (5) with $\sigma_a = 1$)

$$(e_{\mathrm{es}})_{\max} = \max_V e_{\mathrm{es}}(\mathbf{x}) \tag{15}$$

2. Space-averaged standard error (Equation (5) with $\sigma_a = 1$)

$$(e_{\mathrm{es}})_{\mathrm{avg}} = \frac{1}{\mathrm{vol}(V)} \int_V e_{\mathrm{es}}(\mathbf{x})\,\mathrm{d}\mathbf{x} \tag{16}$$

3. Maximum absolute bias error bound [16] in the design space

$$(e_{\mathrm{b}}^{\mathrm{I}})_{\max} = \max_V(e_{\mathrm{b}}^{\mathrm{I}}(\mathbf{x})) \tag{17}$$

and

$$e_{\mathrm{b}}^{\mathrm{I}}(\mathbf{x}) = \sum_{j=1}^{n_2} \left| f_j^{(2)}(\mathbf{x}) - \sum_{i=1}^{n_1} A_{ij} f_i^{(1)}(\mathbf{x}) \right| \tag{18}$$

4. Maximum RMS bias error in the design space (Equation (14) with $\gamma = 1$)

$$(e_{\mathrm{b}}^{\mathrm{rms}})_{\max} = \max_V(e_{\mathrm{b}}^{\mathrm{rms}}(\mathbf{x})) \tag{19}$$

5. Space-averaged RMS bias error (Equation (14) with $\gamma = 1$)

$$(e_{\mathrm{b}}^{\mathrm{rms}})_{\mathrm{avg}} = \frac{1}{\mathrm{vol}(V)} \int_V e_{\mathrm{b}}^{\mathrm{rms}}(\mathbf{x}) \, \mathrm{d}\mathbf{x} \tag{20}$$

This criterion is the same as space-averaged bias error.

Among all the above criteria, the standard error-based criteria are the most commonly used. For all test problems, a design space coded as an N_v-dimensional cube $V = [-1, 1]^{N_v}$ is used and bias errors are computed following the common assumption that the true function and the response surface model are cubic and quadratic polynomials, respectively.

Besides the above error-metric-based criteria, the following criteria are also frequently used:

6. D-efficiency [5, p. 393]

$$D_{\mathrm{eff}} = (|M|/\max|M|)^{1/n_1}, \quad |M| = |X^{(1)\mathrm{T}} X^{(1)}|/N_s^{n_1} \tag{21}$$

Here $\max|M|$ in Equation (21) is taken as the maximum of all EDs. This criterion is primarily used to construct D-optimal designs. A high value of D-efficiency is desired to minimize the variance of the estimated coefficients $\hat{\boldsymbol{\beta}}$.

7. *Radius of the largest unoccupied sphere* ($r_{\max}$): We approximate the radius of the largest sphere that can be placed in the design space such that there are no ED points inside this sphere. A large value of $r_{\max}$ indicates large holes in the design space and hence a potentially poor ED. This criterion is not used to construct EDs, but it allows us to measure the space-filling capability of any ED.

3. TEST PROBLEMS AND RESULTS

This section is divided into two parts. In the first subsection, we compare the widely used EDs, such as LHS designs, D-optimal designs, CCDs and their minimum-bias error counterparts. We show that different designs offer trade-offs among multiple criteria and EDs based on a single error criterion may be susceptible to high errors on other criteria. In the second subsection, we discuss a few possible strategies to simultaneously accommodate multiple criteria. Specifically, we present two strategies: (i) a combination of a geometry-based criterion (LHS) with a model-based criterion (D-optimality), and (ii) simultaneous use of multiple EDs combined with pointwise error estimates as a filtering criterion to seek protection against poor designs.

3.1. Comparison of different EDs

3.1.1. Space-filling characteristics of D-optimal and LHS designs. Popular EDs, such as LHS designs that cater to bias errors by evenly distributing points in design space or numerically obtained D-optimal designs that reduce the effect of noise by placing the design points as far apart as possible, can occasionally leave large holes in the design space due to the random nature of the design (D-optimal) or due to convergence to local optimized LHS designs. This may lead to poor approximation. Firstly, we demonstrate that for D-optimal and optimized LHS designs, a large portion of design space may be left unsampled even for moderate dimensional spaces. For demonstration, we consider two- to four-dimensional spaces $V = [-1, 1]^{N_v}$. The number of points
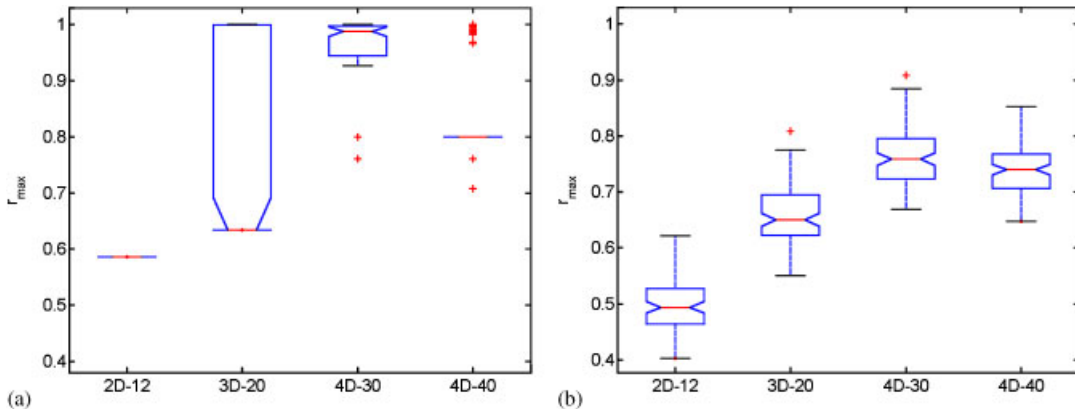
Figure 1. Boxplots (based on 100 designs) of radius (r_{max}) of the largest unoccupied sphere inside the design space $[-1, 1]^{N_v}$ (where N_v is the number of variables). x-axis shows the dimensionality of the design space and the corresponding number of points in the experimental design. Smaller r_{max} is desired to avoid large unsampled regions. (a) D-optimal designs are selected using MATLAB routine '*candexch*' (we specify a uniform grid of points and a maximum of 50 000 iterations for optimization). (b) LHS designs are generated using MATLAB routine '*lhsdesign*' with a maximum of 40 iterations for optimization.

in each ED is twice the number of coefficients in the corresponding quadratic polynomial, that is, 12 points for two dimensions, 20 points for three dimensions, and 30 points for four-dimensional design spaces. We also create EDs with 40 points in four-dimensional design space. We generate 100 designs in each group to alleviate the effect of randomness.

D-optimal designs were obtained using the MATLAB routine *candexch* that selects points from a candidate set using the D-optimality criterion. A uniform grid of 11×11 for two dimensions, $11 \times 11 \times 11$ for three dimensions, and 6^4 for four-dimensional case points was used as the candidate set and a maximum of 50 000 iterations were allowed to find a D-optimal design. Different random ordering of points in the candidate set resulted in different D-optimal designs.[‡] We explain the difficulties in finding D-optimal designs in Appendix C. In this study, we did not allow duplicate points because duplicate points are not useful to approximate data from deterministic functions or numerical simulations. LHS designs were constructed using the MATLAB routine *lhsdesign* with the *maximin* criterion that maximizes the minimum distance between points. We allocated 40 iterations for optimization.

For each ED, we estimated the radius (r_{max}) of the largest unsampled sphere that fits inside the design space and summarized the results with the help of boxplots in Figure 1. The box encompasses the 25–75th percentiles and the horizontal line within the box shows the median value. The notches near the median represent the 95% confidence interval of the median value. The whiskers of the box are placed at 1.5 times the inter-quartile range or the most extreme data, whichever appears earlier. Data points outside the extent of whiskers are denoted by a '+' sign.

It is obvious from Figure 1 that r_{max} increases with the dimensionality of the problem, i.e. the distribution of points in high-dimensional spaces tends to be sparse. As expected, an increase in

---

[‡]The results indicate the presence of a large number of local optimal solutions and hence, the difficulty in finding a true D-optimal design.
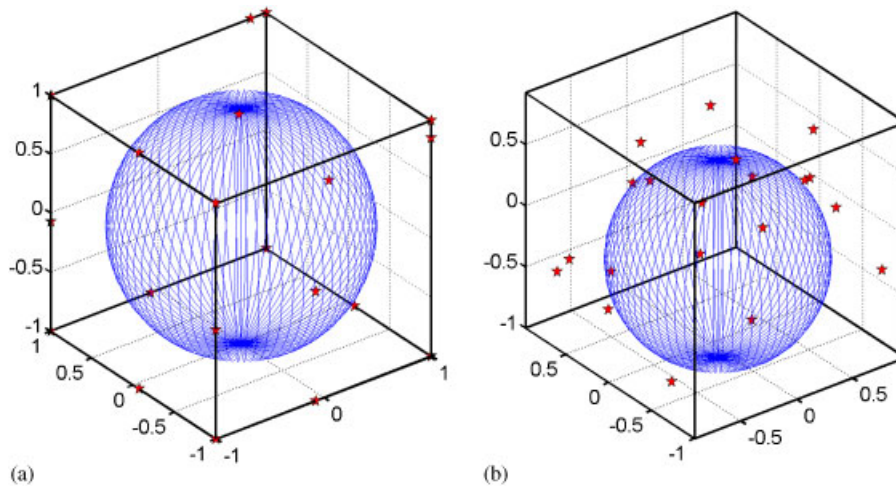
Figure 2. Illustration of the largest spherical empty space inside the three-dimensional design space $[-1, 1]^3$ (20 points): (a) D-optimal design and (b) LHS design.

the density of points reduced r_{max} (compare four-dimensional space with 30 and 40 points). The reduction in r_{max} was more pronounced for D-optimal designs than LHS designs. LHS designs had a less sparse distribution compared with D-optimal designs; however, the median r_{max} of approximately 0.75 units in four-dimensional space for LHS designs indicated that a very large region in the design space remained unsampled and data points are quite far from each other. The sparse distribution of points in the design space is illustrated with the help of a three-dimensional example with 20 points in Figure 2, where the largest unsampled sphere is shown. For both D-optimal and LHS designs, the large size of the sphere clearly demonstrates the presence of large gaps in the design space that makes the surrogate predictions susceptible to errors. This problem is expected to become more severe in high-dimensional spaces. The results indicate that a single criterion- (D-optimality for D-optimal designs and max–min distance for LHS designs) based ED may lead to poor performance on other criteria.

3.1.2. Trade-offs among various EDs. Next, we illustrate trade-offs among different EDs by comparing min–max RMS bias design (given in Appendix A), FCCD, D-optimal design (obtained using JMP [25], Table I), and LHS design (generated using MATLAB routine *lhsdesign* with *maximin* criterion, and allocating 1000 iterations to get a design, Table II) in four-dimensional space. Note that all EDs, except FCCD, optimize a single criterion, i.e. D-optimal designs optimize D-efficiency, LHS designs maximize the minimum distance between points, and min–max RMS bias designs minimize the influence of bias errors. On the other hand, FCCD is an intuitive design obtained by placing the points on the faces and vertices.

The designs were tested using a uniform 11^4 grid in the design space $[-1, 1]^4$ and different metrics are documented in Table III. We observed that no single design (used in the generic sense, meaning a class of designs) outperformed other designs on *all* criteria. The D-optimal and the FCCD had high D-efficiency; the min–max RMS bias design and the LHS design had low D-efficiency. The min–max RMS bias design performed well on bias error-based criteria but

Table I. D-optimal design (25 points, four-dimensional space) obtained using JMP.

# No.	x_1	x_2	x_3	x_4
1	−1	−1	−1	−1
2	−1	−1	−1	1
3	−1	−1	0	−1
4	−1	−1	1	0
5	−1	−1	1	1
6	−1	0	0	1
7	−1	0	1	−1
8	−1	1	−1	−1
9	−1	1	−1	1
10	−1	1	1	−1
11	−1	1	1	1
12	0	−1	−1	1
13	0	−1	1	−1
14	0	0	−1	−1
15	0	1	0	0
16	1	−1	−1	−1
17	1	−1	−1	0
18	1	−1	0	1
19	1	−1	1	−1
20	1	−1	1	1
21	1	0	1	0
22	1	1	−1	−1
23	1	1	−1	1
24	1	1	1	−1
25	1	1	1	1

caused a significant deterioration in standard error-based criteria, due to the peculiar placement of axial points near the center. While the D-optimal design was a good ED according to standard error-based criteria, but a large hole in the design space ($r_{max} = 1$) led to poor performance on bias error-based criteria. Since LHS designs neglect boundaries, they resulted in very high maximum standard error and bias errors. However, LHS designs yielded the least space-averaged bias error estimate. While high errors in LHS and D-optimal designs can be anticipated, the magnitude of errors was remarkably high. The FCCD design, which does not optimize any criterion, performed reasonably on all the metrics. However, we note that FCCD designs in high-dimensional spaces are not practical due to the high ratio of the number of simulations to the number of polynomial coefficients.

We used polynomial examples to illustrate the risks in using EDs constructed with a single criterion. Firstly, we considered a simple quadratic function $F_1(\mathbf{x})$ (Equation (22) in four-dimensional space) with normally distributed noise ε (*zero* mean and *unit* variance),

$$F_1(\mathbf{x}) = \eta(\mathbf{x}) + \varepsilon$$
$$\eta(\mathbf{x}) = 10(1 + x_1 + x_3 + x_2^2 + x_4^2) \tag{22}$$

We construct a quadratic polynomial RSA using the data at 25 points sampled with different EDs (min–max RMS bias design, FCCD, D-optimal, and LHS, Tables I and II) and compute actual absolute error in approximation at a uniform grid of 11^4 points in the design space $V = [-1, 1]^4$.

Table II. LHS designs (25 points, four-dimensional space) were obtained using the MATLAB [26] routine *lhsdesign* with *maximin* criterion and a maximum of 1000 iterations were allowed for optimization.

# No.	x_1	x_2	x_3	x_4
1	−0.970	−0.344	0.055	−0.809
2	0.679	0.123	0.709	0.520
3	−0.093	−0.233	−0.929	−0.659
4	−0.578	0.659	0.943	−0.531
5	0.798	−0.657	−0.829	0.355
6	0.328	0.282	−0.619	0.051
7	0.908	−0.059	−0.207	−0.714
8	−0.142	0.027	0.879	0.271
9	−0.679	0.427	−0.507	0.700
10	0.427	0.265	0.552	−0.900
11	0.470	0.090	0.604	−0.160
12	0.226	−0.474	0.504	0.006
13	−0.224	0.873	0.150	0.872
14	−0.717	−0.812	−0.125	0.396
15	−0.376	0.986	−0.686	−0.463
16	0.002	−0.361	−0.101	−0.936
17	−0.464	−0.862	−0.589	−0.256
18	0.972	−0.547	0.424	−0.411
19	0.689	0.557	−0.324	0.979
20	0.591	−0.175	0.010	0.471
21	0.065	0.457	0.255	−0.281
22	0.160	−0.935	−0.901	−0.075
23	−0.839	0.702	0.329	0.618
24	−0.916	−0.691	0.777	0.808
25	−0.327	0.806	−0.408	0.142

Table III. Comparison of RMS bias CCD, FCCD, D-optimal, and LHS designs for four-dimensional space (all designs have 25 points).

Experimental design	D-efficiency	$(e_{es})_{max}$	$(e_{es})_{avg}$	$|e_b^I|_{max}$	$(e_b^{rms})_{max}$	$(e_b^{rms})_{avg}$	r_{max}
RMS bias CCD	0.148	70.71	35.22	6.996	1.155	0.927	0.65
FCCD	0.932	0.877	0.585	6.208	1.176	0.827	0.67
LHS	0.256	3.655	1.032	21.48	3.108	0.588	0.83
D-optimal	1.000	0.933	0.710	12.00	1.996	1.004	1.00

Note: Errors are computed on a uniform 11^4 grid in space $V = [-1, 1]^4$.

The accuracy measures based on the data that are most commonly used are the adjusted coefficient of determination R_{adj}^2 and the standard error normalized by the range of the function (RMSE) in Table IV. Very high values of normalized maximum, RMS, and average absolute errors (normalized by the range of the data for the respective ED) in Table IV indicate that the min–max RMS bias design (also referred to as RMS bias CCD) is indeed a poor choice of ED when the error is due to noise, though all approximation accuracy measures (R_{adj}^2, standard error) suggested otherwise. That is, the high errors come with no warning from the fitting process! High values of the ratio of space-averaged, RMS, or maximum actual errors to standard error indicate the risks associated

Table IV. Prediction performance of different 25-point experimental designs in approximation of example functions F_1 (Equation (22)), F_2 (Equation (24)), and F_3 (Equation (25)) in four-dimensional spaces.

ED	R^2_{adj}	Standard error	Average error	RMS error	Maximum error
F₁ (noise)					
RMS bias CCD	1.00	0.021	**0.84**	**1.05**	**3.22**
FCCD	1.00	0.018	0.0081	0.010	0.044
LHS	0.99	0.032	0.032	0.043	0.26
D-optimal	0.99	0.022	0.023	0.025	0.042
F₂ (LHS sphere)					
RMS bias CCD	1.00	0.035	**0.15**	**0.18**	0.36
FCCD	0.91	0.094	0.090	0.11	0.25
LHS	0.99	0.024	0.042	0.072	**1.29**
D-optimal	0.90	0.090	**0.25**	**0.31**	**0.82**
F₃ (D-optimal sphere)					
RMS bias CCD	1.00	0.00065	**0.17**	**0.18**	0.22
FCCD	0.96	0.062	0.077	0.088	0.18
LHS	0.99	0.024	0.035	0.055	**0.69**
D-optimal	0.90	0.075	**0.17**	**0.22**	**0.65**

Note: We use quadratic polynomial response surface approximations and errors are computed on a uniform 11^4 grid in space $V = [-1, 1]^4$. R^2_{adj} indicates the quality of approximation. The true standard error, space-averaged actual absolute error, root mean square error in space, and maximum actual absolute error are normalized by the range of the data at respective experimental designs. D-optimal and LHS designs are given in Tables I and II, respectively. The high values of errors compared with the standard errors are shown in bold face.

with relying on measures such as R^2_{adj} to determine the accuracy of approximations. Among the other EDs, LHS design has a high normalized maximum error near the corners, where no data are sampled. FCCD and D-optimal designs performed reasonably, with FCCD being the best design.

Secondly, we illustrate errors due to large holes in design space observed in the previous section (Section 3.1.1). A simple function that is likely to produce large errors would be a cosine with maximum at the center of the sphere. However, to limit ourselves to a polynomial true function, we use a truncated Maclaurin series expansion of a translated radial cosine function $\cos(k\|\mathbf{x} - \mathbf{x}^c_{\text{lhs}}\|)$, namely

$$F(\mathbf{x}) = 20(1 - r^2/2 + r^4/4!), \quad r = k\|\mathbf{x} - \mathbf{x}^c_{\text{ed}}\| \tag{23}$$

where $\mathbf{x}^c_{\text{ed}}$ is a fixed point in design space and k is a constant. We consider two instances of Equation (23) by specifying the center of the largest unoccupied sphere associated with LHS design (Table II) and D-optimal design (Table I) as the fixed point

$$F_2(\mathbf{x}) = 20(1 - r^2/2 + r^4/4!), \quad r = k_2\|\mathbf{x} - \mathbf{x}^c_{\text{lhs}}\|$$
$$\mathbf{x}^c_{\text{lhs}} = [-0.168, -0.168, -0.141, 0.167], \quad k_2 = 1.1 \tag{24}$$

$$F_3(\mathbf{x}) = 20(1 - r^2/2 + r^4/4!), \quad r = k_3\|\mathbf{x} - \mathbf{x}^c_{\text{dopt}}\|$$
$$\mathbf{x}^c_{\text{dopt}} = [0, 0, 0, 0], \quad k_3 = 1.18 \tag{25}$$

The constants k_2 and k_3 are obtained by maximizing the normalized maximum actual absolute error in approximation of $F_2(\mathbf{x})$ using LHS ED and approximation of $F_3(\mathbf{x})$ using D-optimal ED, respectively, subject to a reasonable approximation (determined by the condition $R^2_{\text{adj}} \geqslant 0.90$) of the two functions by all considered EDs (FCCD, D-optimal, LHS, and RMS bias CCD). As earlier, the errors are normalized by dividing the actual absolute errors by the range of data values used to construct the experimental design.

We use quadratic polynomials to approximate $F(\mathbf{x})$ and errors are evaluated at a uniform grid of 11^4 points in the design space $V = [-1, 1]^4$. The quality of fit, maximum, RMS, and average actual absolute errors in approximation for each ED are summarized in Table IV. We observe that despite good metrics of fit (high R^2_{adj} and low normalized standard error), the normalized maximum actual absolute errors are high for all EDs. In particular, the approximations constructed using data sampled at the D-optimal and the LHS designs perform very poorly. This means that the accuracy metrics, though widely trusted, can mislead the actual performance of the ED. The high maximum error in approximations using the LHS designs occurs at one of the corners that was not sampled (thus extrapolation error); however, we note that LHS designs yield the smallest normalized space-averaged and RMS errors in approximation. On the other hand, the maximum error in approximations using the D-optimal designs appeared at a test point closest to the center $\mathbf{x}^c_{\text{lhs}}$ in the case of $F_2(\mathbf{x})$ and near $\mathbf{x}^c_{\text{dopt}}$ in the case of $F_3(\mathbf{x})$. Besides, high normalized space-averaged errors indicate poor approximation of the true function $F(\mathbf{x})$. The other two EDs, FCCD and RMS bias CCD, performed reasonably on maximal errors. The relatively poor performance of RMS bias CCD for average and RMS errors is explained by recalling that the ED is constructed by assuming the true function to be a cubic polynomial, whereas $F(\mathbf{x})$ is a quartic polynomial.

An important characteristic of all EDs is the ratio of space-averaged, RMS, or maximum actual absolute error to the standard error. When this ratio is large, the errors are unexpected and therefore potentially damaging. The FCCD design provided a reasonable value of the ratio of actual to standard errors; however, RMS bias design performed very poorly as the actual errors were much higher than the standard error. This means that the standard error misleads about the actual magnitude of the error. Similarly, for all functions, the ratio of maximum actual absolute error to the standard error for LHS designs (29–52) was much higher than for D-optimal designs (about 9). The surprise element is also evident by the excellent values of R^2_{adj} of 0.99 and 1.00 compared with 0.90 for the D-optimal design.

The results presented here clearly suggest that different EDs were non-dominated with respect to each other and offered multiple (sometimes extreme) trade-offs, and that it might be dangerous to use a single criterion-based ED without thorough knowledge of the problem (which is rarely the case in practice).

3.1.3. An extreme example of risks in single criterion-based design: min–max RMS bias CCD. A more pathological case, demonstrating the risks in developing EDs using a single criterion, was encountered for moderate dimensional cases while developing the CCD counterpart of the minimum-bias design, i.e. minimizing the maximum RMS bias error (Section 3.1.2). The performance of the min–max RMS bias designs constructed using two parameters (refer to Appendix A) for two- to five-dimensional spaces on different metrics is given in Table V. For two- and three-dimensional spaces, the axial points (given by α_2) were located at the face and the vertex points (given by α_1) were placed slightly inwards to minimize the maximum RMS bias errors. The RMS bias designs performed very reasonably on all the error metrics. A surprising result was obtained for

Table V. Min–max RMS bias central composite designs for two- to five-dimensional spaces and corresponding design metrics (N_v is the number of variables).

N_v	α_1	α_2	$(e_{es})_{max}$	$(e_{es})_{avg}$	$\lvert e_b^I \rvert_{max}$	$(e_b^{rms})_{max}$	$(e_b^{rms})_{avg}$
2	0.954	1.000	0.973	0.688	1.029	0.341	0.269
3	0.987	1.000	0.913	0.607	2.832	0.659	0.518
4*	1.000	1.000	0.877	0.585	6.208	1.176	0.827
4	1.000	0.100	70.71	35.22	6.996	1.155	0.927
5	1.000	0.100	77.46	41.60	12.31	1.826	1.200

Note: Errors are computed on a uniform 41^2 grid ($N_v = 2$), 21^3 grid ($N_v = 3$), and 11^{N_v} grid ($N_v > 3$), in space $V = [-1, 1]^{N_v}$. Refer to Appendix A for details of creating CCD.
*FCCD for four variables (not min–max RMS bias design).

optimal designs for four- and five-dimensional spaces: while the parameter corresponding to vertex points (α_1) was at its upper limit (1.0), the parameter corresponding to the location of axial points (α_2) hits the corresponding lower bound (0.1). This means that to minimize the maximum RMS bias error, the points should be placed near the center. The standard error was expectedly very high for this design. Contrasting FCCD for four-dimensional cases with three- and four-dimensional min–max RMS bias designs (Table V) isolated the effect of dimensionality and the change in ED (location of axial points) on different error metrics. The increase in bias errors (bounds and RMS error) was attributed to increase in dimensionality (small variation in bias errors with different EDs in four-dimensional design space) and the increase in standard error for min–max RMS bias design was the outcome of the change in ED (the location of axial points given by α_2). This unexpected result for four- and higher-dimensional cases is supported by theoretical reasoning[§] [27] and very strong agreement between the predicted and the actual RMS bias errors for the min–max RMS bias design and the face-centered CCD (Appendix A).

To further illustrate the severity of the risks in using a single criterion, we show the trade-offs among the maximum errors (RMS bias error, standard error, and bias error bound) for a four-dimensional design space $[-1, 1]^4$, obtained by varying the location of the axial points (α_2) from near the center ($\alpha_2 = 0.1$, min–max RMS bias design) to the face of the design space ($\alpha_2 = 1.0$, CCD), while keeping the vertex locations ($\alpha_1 = 1.0$) fixed. The trade-off between maximum RMS bias error and maximum standard error is shown in Figure 3. Moving the axial point away from the center reduced the maximum-bias error bound and the maximum standard error but increased the maximum RMS bias error. The relatively small variation in maximum RMS bias error compared with the variation in maximum standard error and maximum-bias error bound demonstrated the low sensitivity of maximum RMS bias error with respect to the location of axial points (α_2) and explains the success of the popular CCDs ($\alpha_2 = 1.0$) in handling problems with bias errors. While

[§]The ED is a compromise to minimize errors due to the cubic terms: x_i^3, $x_i x_j^2$, and $x_i x_j x_k$, which are modeled by x_i, cx_i (constant c depends on the location of axial and vertex points), and zero, respectively. Vertex points favor $c = \alpha_1^2$ and axial points favor $c = 0$. The $x_i x_j x_k$ terms dominate the choice of EDs because these terms generate large bias errors (modeled by zero) that peak at the all-positive or all-negative corners, and the number of these terms increases with the dimensionality of the problem (for $N_v = 2$, zero out of four terms; for $N_v = 3$, one out of 10 terms; for $N_v = 4$, four out of 20, and so on). If axial points are placed inside the design space, the term $x_i x_j^2$ contributes to the error at all-positive and all-negative corners ($c \leqslant \alpha_1^2$). However, when the axial points (hence α_2) go to zero, we can have $c \to \alpha_1^2$, which minimizes additional errors at the all-positive and all-negative corners.
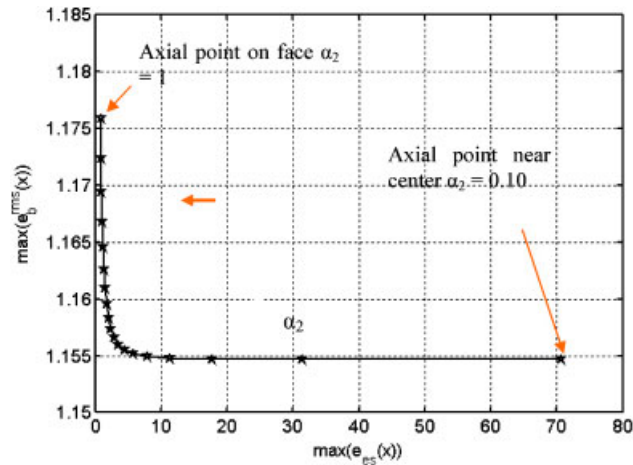
Figure 3. Trade-offs between maximum standard error $(e_{es})_{max}$ and maximum RMS bias error $(e_b^{rms})_{max}$ for four-dimensional space. Twenty-five points were used to construct central-composite experimental designs with vertex location fixed at $\alpha_1 = 1.0$.

we noted that each design on the curves in Figure 3 corresponds to a non-dominated (trade-off) point, a small increase in maximum RMS bias error permits a large reduction in maximum standard error, or in other words, the minimization with respect to a single criterion (here maximum RMS bias error) may lead to small gains at a cost of significant loss with respect to other criteria. Trade-off between maximum-bias error bound and maximum RMS bias error also reflected similar results, though the gradients were relatively small.

The most important implication of the results presented in this section is that it may not be wise to select EDs based on a single criterion. Instead, trade-off between different metrics should be explored to find a reasonable ED. While detailed exploration of this issue requires significantly more research, our initial attempts to simultaneously accommodate multiple criteria are illustrated next.

3.2. Strategies to address multiple criteria for EDs

As discussed in the previous section, the EDs optimized using a single criterion may perform poorly on other criteria. While a bad ED can be identified by visual inspection in low-dimensional spaces, we need additional measures to filter out bad designs in high-dimensional spaces. We explored several strategies to simultaneously accommodate multiple criteria in an attempt to avoid poor EDs. In this context, we discuss two issues: (i) which criteria are meaningful for different EDs and (ii) how can we combine different criteria.

Since the EDs are constructed to minimize the influence of bias error and noise, a sensible choice of suitable criteria for any ED should seek balance among the two sources of errors, i.e. bias and noise. Consequently, if we select an ED that primarily caters to one source of error, for example, noise, the secondary criterion should be introduced to address the other source of error, bias error in this case, and *vice versa*. We elaborate on this idea in Section 3.2.2.

Once we have identified criteria to construct EDs, we seek ways to combine different criteria. Taking inspiration from multi-objective optimization problems, we can accommodate multiple

criteria according to several methods, for example, (i) optimize the ED to minimize a composite function that represents the weighted sum of criteria, (ii) optimize the ED to minimize the primary criterion while satisfying constraints on the secondary criterion, and (iii) solve a multi-objective optimization problem to identify different trade-offs and then select a design that suits the requirements the most.

Here, we show two ways to avoid poor EDs using a four-dimensional example. Firstly, we present a method to combine the model-based D-optimality criterion that caters to noise with the geometry-based LHS criterion that distributes points evenly in design space and reduces space-averaged bias errors. Secondly, we demonstrate that selecting one out of several EDs according to an appropriate pointwise error-based criterion reduces the risk of obtaining poor EDs. Further, we show that the coupling of multiple criteria and multiple EDs may be effective to avoid poor designs.

3.2.1. Combination of model-based D-optimality criterion with geometry-based LHS criterion. We used an example of constructing an ED for a four-dimensional problem with 30 points (response surface model and assumed true model were quadratic and cubic polynomials, respectively). Three sets of EDs were generated as follows. The first set comprised 100 LHS EDs generated using the MATLAB routine *lhsdesign* with the *maximin* criterion (a maximum of 40 *iterations* were assigned to find an ED). The second set comprised 100 D-optimal designs obtained using the MATLAB routine *candexch* that selects points from a candidate set of a uniform grid of 6^4 points using the D-optimality criterion. A maximum of 50 000 iterations were allowed to find a D-optimal design. As before, different random ordering of points in the candidate set resulted in different D-optimal designs. The third set of (combination) EDs was obtained by combining D-optimal (model-based criterion) and LHS designs (geometry-based criterion). We selected 30 design points from a 650-point[¶] LHS design (*lhsdesign* with *maximin* criterion and a maximum of 100 *iterations* for optimization) using the D-optimality criterion (*candexch* with a maximum of 50 000 iterations for optimization). For each design, we computed the radius r_{max} of the largest unsampled sphere, D-efficiency, maximum and space-averaged RMS bias and standard error using a uniform 11^4 grid in the design space $[-1, 1]^4$.

We show the trade-off among different criteria for D-optimal, LHS, and combination designs in Figure 4. As can be seen from Figure 4(a), the D-optimal designs were the best and LHS designs were the worst with respect to the maximum standard and RMS bias errors. Compared with the LHS designs, the combination designs significantly reduced the maximum standard error with marginal improvement on the maximum RMS bias error (Figure 4(a)) and improved D-efficiency without sacrificing r_{max} (Figure 4(d)). The advantages of using combination designs were more obvious in Figure 4(b), where we compared space-averaged bias and standard errors. We see that D-optimal designs performed well on space-averaged standard errors but yielded high space-averaged RMS bias errors. On the other hand, the LHS designs had low space-averaged RMS bias errors but high space-averaged standard errors. The combination designs simultaneously yielded low space-averaged RMS bias and standard errors. This result was expected because the LHS criterion allows

---

[¶]The number of points in the uniform grid used to generate D-optimal designs was 1296. So to provide a fair comparison while keeping the computational cost low, we obtain 650 points using LHS and use this set of points to develop combination EDs. We note that the optimal number of points in the candidate LHS sample should be carefully selected because a small sample would not allow sufficient options for the D-optimality criterion and a large sample may degenerate into a 'D-optimal like' (most points near the boundary) ED.
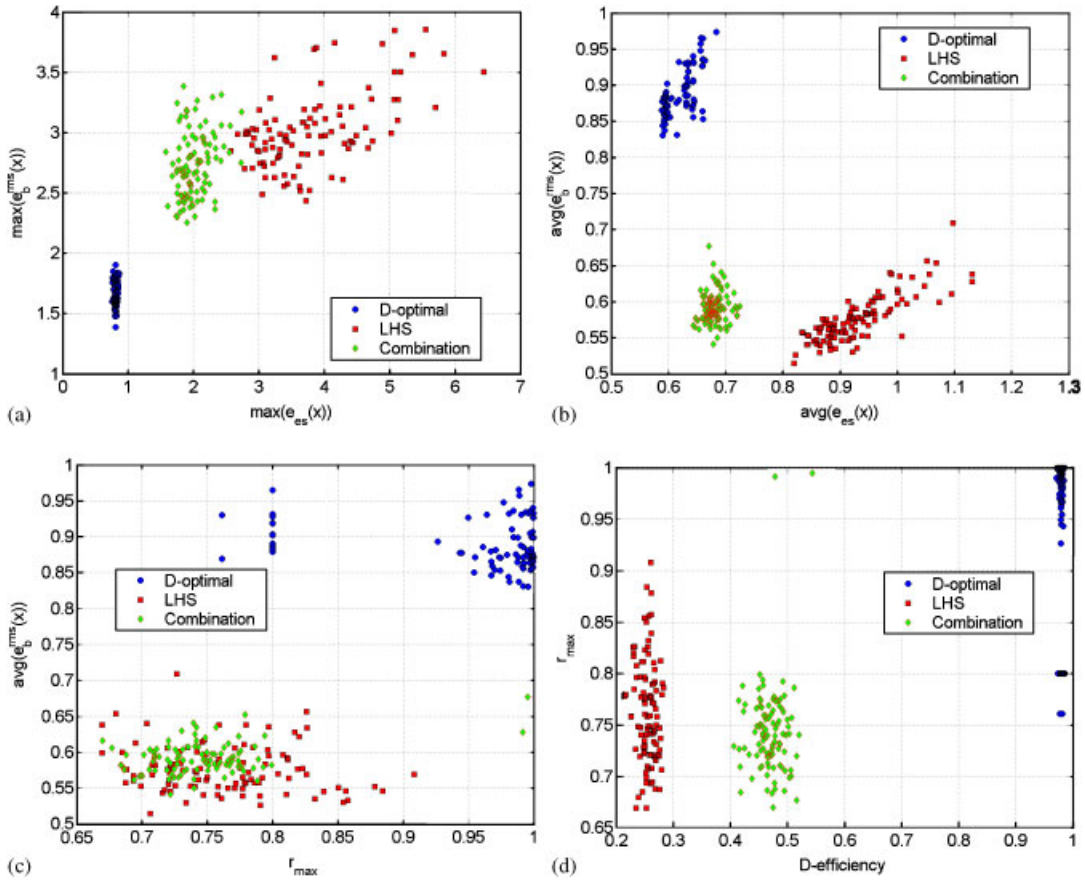
Figure 4. Comparison of 100 D-optimal, LHS, and combination (D-optimality+LHS) of experimental designs in four-dimensional space (30 points) using different metrics. r_{max}, radius of the largest unsampled sphere; $e_b^{rms}(\mathbf{x})$, RMS bias error; $e_{es}(x)$, standard error; $(\cdot)_{max}$, maximum of the quantity inside parenthesis, $(\cdot)_{avg}$, spatial average of the quantity inside parentheses. All metrics, except D-efficiency (D_{eff}), are desired to be low: (a) $(e_{es})_{max}$ vs $(e_b^{rms})_{max}$; (b) $(e_{es})_{avg}$ vs $(e_b^{rms})_{avg}$; (c) r_{max} vs $(e_b^{rms})_{avg}$; and (d) D_{eff} vs r_{max}.

relatively uniform distribution of the points by constraining the location of points that are used to generate combination designs using the D-optimality criterion. Similarly, we observed that unlike D-optimal designs, combination EDs performed very well on the space-averaged RMS bias error and the r_{max} criterion (refer to Figure 4(c)), and the performance was comparable to that of the LHS designs.

Mean and coefficient of variation (COV) of different metrics for the three sets of EDs are tabulated in Table VI. D-optimal designs outperformed LHS designs in terms of the ratio of maximum to average error (stability), D-efficiency, maximum RMS bias error, and maximum standard error. Also, for most metrics, the variation in results due to sampling (COV) was the least among the three. As seen before, LHS designs performed the best on space-averaged RMS bias errors.

Table VI. Mean and coefficient of variation (based on 100 instances) of different error metrics for various experimental designs in four-dimensional space (30 points).

ED	$(e_{es})_{max}$	$(e_{es})_{avg}$	$(e_b^{rms})_{max}$	$(e_b^{rms})_{avg}$	r_{max}	D_{eff}
Mean						
LHS	3.82	0.94	3.02	0.57	0.76	0.26
D-optimal	0.80	0.62	1.67	0.89	0.94	0.98
Combination	2.02	0.68	2.67	0.59	0.75	0.47
COV						
LHS	0.209	0.074	0.105	0.058	0.066	0.056
D-optimal	0.024	0.025	0.105	0.031	0.087	0.003
Combination	0.126	0.026	0.102	0.038	0.064	0.053

Note: Errors are computed on a uniform 11^4 grid in space $V = [-1, 1]^4$. D_{eff}, D-efficiency; r_{max}, radius of the largest unsampled sphere; e_b^{rms}, RMS bias error; e_{es}, standard error; $(\cdot)_{max}$, maximum of the quantity inside parentheses; $(\cdot)_{avg}$, average of the quantity inside parentheses.

The designs obtained by combining two criteria (D-optimality and LHS) were substantially closer to the best of the two except for $(e_b^{rms})_{max}$. Thus, they reduced the risk of large errors. Furthermore, the variation with samples (COV) is also reduced. The results suggested that though different EDs were non-dominated (trade-offs) with respect to each other, simultaneously considering multiple criteria by combining the model-based D-optimality criterion and the geometry-based LHS criterion may be effective in producing more robust EDs with a reasonable trade-off between bias errors and noise.

3.2.2. Multiple EDs combined with pointwise error-based filtering. Next, we demonstrate the potential of using multiple EDs to reduce the risk of finding poor EDs. The main motivation is that the cost of generating EDs is not high, hence, we can construct two or three EDs using LHS, or D-optimality, or a combination of the two criteria, and pick the best according to a complementary criterion. To illustrate the improvements by using three EDs over a single ED, each of the two criteria–maximum RMS bias error and maximum standard error–were used to select the best (least error) of the three EDs. For illustration, 100 such experiments were conducted with LHS designs, D-optimal designs, and the combination of LHS and D-optimal designs (as described above).

Actual magnitudes of maximum RMS bias error and maximum standard error for all 300 designs and the 100 designs obtained after filtering using min–max RMS bias or maximum standard error criteria are plotted in Figure 5 for three sets of (100) EDs. As is evident by the shortening of the upper tail and the size of the boxplots in Figure 5, both the min–max RMS bias and maximum standard error criteria helped eliminate poor EDs for all three sets. Filtering had more impact on the maximum error estimates than the space-averaged error estimates. The numerical quantification of improvements in actual magnitude of maximum and space-averaged error based on 100 experiments is summarized in Table VII. We observed that the pointwise error-based (min–max RMS bias or standard error) filtering significantly reduced the mean and COV of maximum errors. We also noted improvements in the individual EDs using multiple criteria. LHS designs were most significantly improved by picking the best of three based on estimated maximum standard error. D-optimal designs combined with the min–max RMS bias error-based filtering criterion helped eliminate poor designs according to the RMS bias error criterion. It can be concluded from this exercise that potentially poor designs can be filtered out by considering a small number of EDs with an
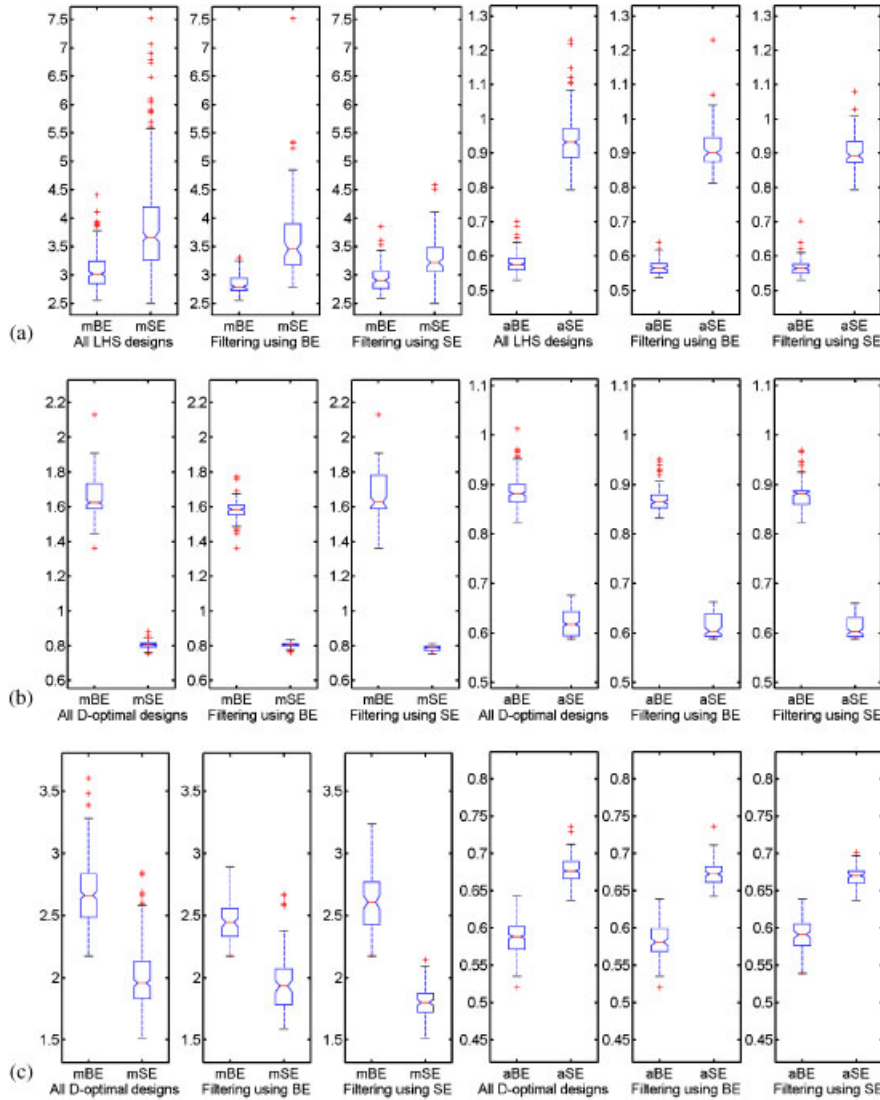
Figure 5. Simultaneous use of multiple experimental design concept, where one out of three experimental designs is selected using appropriate criterion (filtering). Boxplots show maximum and space-averaged RMS bias and standard errors in four-dimensional design space for; considering all 300 designs, 100 designs filtered using min–max RMS bias error as a criterion and 100 designs filtered using min–max standard error as a criterion (mBE denotes maximum RMS bias error, mSE denotes maximum standard error, aBE denotes space-averaged RMS bias error, aSE denotes space-averaged standard error). Each experimental design has 30 points and errors are computed in design space $[-1, 1]^4$ on 11^4 uniform grid. (a) LHS experimental designs; (b) D-optimal designs; and (c) D-optimal designs selected from a larger LHS set.

Table VII. Illustration of reduction in errors by considering multiple experimental designs and picking
one experimental design using appropriate criterion (filtering).

	LHS designs			D-optimal designs			Combination designs		
	All-ED	BE	SE	All-ED	BE	SE	All-ED	BE	SE
Mean-$(e_{es})_{max}$	3.85	3.62	3.29	0.80	0.81	0.79	2.00	1.96	1.80
COV- $(e_{es})_{max}$	0.21	0.19	0.11	0.025	0.018	0.020	0.12	0.12	0.066
Mean-$(e_{es})_{avg}$	0.94	0.92	0.90	0.62	0.62	0.61	0.68	0.67	0.67
COV-$(e_{es})_{avg}$	0.071	0.063	0.053	0.041	0.039	0.034	0.024	0.024	0.019
Mean-$(e_b^{rms})_{max}$	3.08	2.83	2.95	1.66	1.58	1.68	2.68	2.46	2.60
COV-$(e_b^{rms})_{max}$	0.099	0.053	0.083	0.063	0.040	0.071	0.094	0.066	0.091
Mean-$(e_b^{rms})_{avg}$	0.58	0.57	0.57	0.89	0.87	0.88	0.59	0.58	0.59
COV-$(e_b^{rms})_{avg}$	0.049	0.036	0.043	0.035	0.030	0.031	0.037	0.037	0.036

Note: We summarize actual magnitudes of maximum and space-averaged errors for LHS/D-optimal/combination
designs in four-dimensional space with 30 points. All-ED refers to error data corresponding to all 300 LHS/D-
optimal/combination designs. BE and SE refer to error data from 100 EDs selected using min–max RMS bias
error and min–max standard error criterion, respectively. Errors are computed on a uniform 11^{N_v} grid in space
$V = [-1, 1]^{N_v}$. Mean and coefficient of variation (COV) are based on 300 (ALL-ED) or 100 (BE/SE) EDs.

appropriate (min–max RMS bias or maximum standard) error criterion. The filtering criterion
should be complementary to the criterion used for construction of the ED, i.e. if a group of EDs
are constructed using a variance-based criterion, then the selection of an ED from the group should
be based on bias error criterion and *vice versa*.

Results presented in this section indicate that the use of multiple criteria (LHS and D-optimality)
and multiple EDs helps reduce maximum and space-averaged bias and standard errors. Imple-
menting the above findings, we can obtain EDs with reasonable trade-off between bias error and
noise in three steps as follows: (i) generate a large number of LHS ED points, (ii) select a D-optimal
subset within the LHS design (combine model-based and geometry-based criteria), (iii) repeat
steps (i) and (ii) three times and select the design that is the best according to one of the min–max
RMS bias or maximum standard error criteria (filtering using pointwise error-based criterion).

4. CONCLUDING REMARKS

In this paper, we demonstrated the risks of using a single criterion to construct EDs. We showed
that constructing EDs by combining multiple (model, geometry, and error based) criteria and/or
using multiple EDs reduces the risk of using a poor ED.

For four-dimensional space, comparison of computed LHS and D-optimal designs, that involve
random components and may yield poor approximation due to random components or convergence
to local optima, revealed that the D-optimal designs were better for maximum errors and LHS
designs were better for space-averaged RMS bias errors. Both designs were susceptible to leaving
large spheres in design space unsampled. A comparison of popular EDs (FCCD, min–max RMS
bias design, D-optimal design, and LHS design) revealed the non-dominated (trade-off among
different criteria) nature of different designs. The min–max RMS bias design, obtained by placing
the axial points close to the center, performed the best in reducing maximum RMS bias error,

but was the worst design for standard error metrics and D-efficiency. LHS design gave the best performance in terms of space-averaged bias errors. However, FCCD, which is an intuitive design, yielded a reasonable trade-off between bias error and noise reduction on all metrics. The same conclusions were supported by approximation of three example polynomials that highlighted the susceptibility of different EDs to the nature of the problem, despite the fact that the accuracy metrics suggested a very good fit for each example. We concluded that different EDs, constructed using one error criterion, do not perform the best on all criteria. Instead, they offer trade-offs.

Even in moderate dimensional spaces these single criterion-based designs can often lead to extreme trade-offs, particularly by using the maximum RMS bias error measure as a design criterion, such that small gains in the desired criterion are achieved at the cost of significant deterioration of performance in other criteria. A trade-off study, conducted to study the variation of different error metrics with the location of axial points in CCDs, illustrated the perils of using a single criterion to construct EDs and emphasized the need to consider multiple criteria to trade-off bias error and noise reduction.

To address the risk of using a poor ED by considering a single criterion, we explored a few strategies to accommodate multiple criteria. We demonstrated that the ED obtained by combining two criteria, the D-optimality criterion with LHS design, offered a reasonable trade-off between space-averaged RMS bias and standard error and space-filling criteria. Specifically, combination designs significantly improved the poor EDs. We showed that the risk of getting a poor ED could be further reduced by choosing one out of three EDs using a pointwise error-based criterion, e.g. min–max RMS bias or maximum standard error criterion. The combination of D-optimal designs and min–max RMS bias error was particularly helpful in reducing bias errors. Finally, we adopted selection of EDs by combining the D-optimality criterion with LHS design and selecting one out of three such combination designs to cater to both bias error and noise reduction. However, since these results are based on a limited number of examples, we note the need of future research to address the issues related to accommodating multiple criteria while constructing EDs.

APPENDIX A: CONSTRUCTION AND VERIFICATION OF MIN–MAX RMS BIAS EDs

We construct the minimum-bias error counterpart of the CCD, a popular minimum variance design, in an N_v-dimensional cube $V = [-1, 1]^{N_v}$ for N_v ranging from 2 to 5. For N_v dimensions, the CCD has $2^{N_v} + 2N_v + 1$ points. Minimum-bias experimental designs are constructed using two parameters $0 \leqslant \alpha_1, \alpha_2 \leqslant 1$ that define the coordinates of sampled data points. Points corresponding to vertices are placed at $x_i = \pm \alpha_1$, for $i = 1, \ldots, N_v$, and axial points are located at $x_i = \pm \alpha_2, x_j = 0$, for all $j \neq i$, for each $i = 1, \ldots, N_v$. In addition, one point is placed at the origin. Figure A1 shows an example of an ED as a function of parameters α_1, α_2 in two dimensions.

The min–max RMS bias design can be obtained by minimizing $(e_b^{rms})_{max}$ (Equation (19)) that is a function of α_1, α_2. Mathematically, this is a two-level optimization problem:

$$e_{min}^b = \min_{\alpha_1, \ldots, \alpha_m} (e_b^{rms})_{max} = \min_{0 \leqslant \alpha_1, \ldots, \alpha_m \leqslant 1} (\max_V |e_b^{rms}(\mathbf{x})|) \tag{A1}$$

The inner problem is the maximization of RMS bias error over the design space (this is a function of α_i and pointwise errors are estimated by Equation (14)), and the outer-level problem is minimization over the parameters α_i.
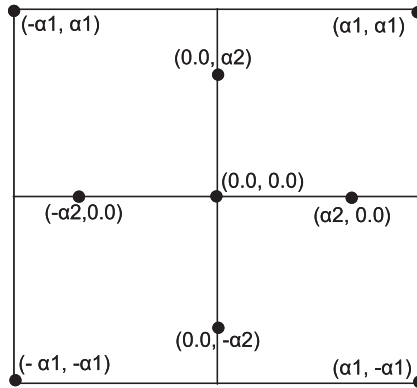
Figure A1. Two-dimensional illustration of central composite experimental design constructed using two parameters α_1 and α_2.

For the two-dimensional example problem, let the true model be a cubic polynomial

$$\eta(\mathbf{x}) = \beta_1^{(1)} + \beta_2^{(1)}x_1 + \beta_3^{(1)}x_2 + \beta_4^{(1)}x_1^2 + \beta_5^{(1)}x_1x_2 + \beta_6^{(1)}x_2^2 + \beta_1^{(2)}x_1^3 + \beta_2^{(2)}x_1^2x_2$$

$$+ \beta_3^{(2)}x_1x_2^2 + \beta_4^{(2)}x_2^3 \tag{A2}$$

and the fitted model be a quadratic polynomial

$$\hat{y}(\mathbf{x}) = \hat{\beta}_1 + \hat{\beta}_2 x_1 + \hat{\beta}_3 x_2 + \hat{\beta}_4 x_1^2 + \hat{\beta}_5 x_1 x_2 + \hat{\beta}_6 x_2^2 \tag{A3}$$

for which $\mathbf{f}^{(1)}(\mathbf{x}) = [1, x_1, x_2, x_1^2, x_1x_2, x_2^2]^T$, $\mathbf{f}^{(2)}(\mathbf{x}) = [x_1^3, x_1^2x_2, x_1x_2^2, x_2^3]^T$, $\boldsymbol{\beta}^{(1)} = [\beta_1^{(1)}, \beta_2^{(1)}, \beta_3^{(1)}, \beta_4^{(1)}, \beta_5^{(1)}, \beta_6^{(1)}]^T$, and $\boldsymbol{\beta}^{(2)} = [\beta_1^{(2)}, \beta_2^{(2)}, \beta_3^{(2)}, \beta_4^{(2)}]^T$.

Note that knowledge of $\hat{\boldsymbol{\beta}}, \boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}$ is not required to generate an ED. Different standard designs available in the literature–minimum variance design ED 1 (minimizes maximum variance, Myers and Montgomery [5]), minimum space-averaged bias design ED 2 (minimizes space-averaged bias error, Qu *et al.* [8]), min–max bias error bound design ED 3 (minimizes maximum-bias error bound, Papila *et al.* [16]), and min–max RMS bias design ED 4 (minimizes maximum RMS bias error, Goel *et al.* [18])—were compared using metrics defined in the paper. Errors were computed using a uniform 41×41 grid. Results are summarized in Table AI.

Interestingly, the design based on the most commonly used bias error metric, the minimum space-averaged bias design (ED 2), performed very poorly on all the metrics except the space-averaged bias errors (RMS and bound). Since for ED 2 the sampled points were located in the interior, there was a significant extrapolation region that explained the very high maximum errors. This suggested that spatial averaging of bias error may not be the best criterion to create EDs. On the other hand, all other designs gave comparable performance on all metrics. As expected, the EDs based on bias errors (min–max bias error bound design (ED 3) and min–max RMS bias design (ED 4)) performed better on bias errors and the ED based on min-variance (ED 1) reduced standard error more than the other designs. The differences between the min–max bias error bound design (ED 3) and the min–max RMS bias error design (ED 4) were small. The results show that a

Table AI. Comparison of different experimental designs for two dimensions.

ED	α_1	α_2	$(e_{es})_{max}$	$(e_{es})_{avg}$	$(e_b^I)_{max}$	$(e_b^I)_{avg}$	$(e_b^{rms})_{max}$	$(e_b^{rms})_{avg}$
(1) Min–max variance	1.000	1.000	0.898	0.670	1.170	0.892	0.385	0.302
(2) Min–avg BE	0.700	0.707	1.931	0.869	2.364	0.506	0.690	0.168
(3) Min–max BEB	0.949	0.949	0.993	0.681	1.001	0.764	0.351	0.261
(4) Min–max RMSBE	0.954	1.000	0.973	0.688	1.029	0.793	0.341	0.269

Note: (1) Minimum maximum variance design. (2) Minimum space-averaged bias design. (3) Min–max bias error bound design. (4) Min–max RMS bias design. Metrics of measurement were maximum standard error, space-averaged standard error, maximum-bias error bound (BEB), space-averaged bias error bound, maximum RMS bias error, and space-averaged RMS bias error (BE). Errors are computed on a uniform 41×41 grid in space $V = [-1, 1]^2$. α_1 and α_2 define the location of sampled points in Figure A1.

single criterion-based design does not perform the best on all metrics. Instead, we obtain different non-dominated trade-off solutions.

We also constructed min–max RMS bias EDs in two- to five-dimensional spaces by solving Equation (A1) with the conventional assumption of a quadratic RSA and a cubic polynomial as the true model. Different error metrics corresponding to these min–max RMS bias designs are given in Table V.

To verify the results, we compared the predicted RMS bias errors with the actual RMS bias errors for four-dimensional min–max RMS bias design ($\alpha_1 = 1.0$, $\alpha_2 = 0.1$) and CCD ($\alpha_1 = 1.0$, $\alpha_2 = 1.0$). A uniform grid of 11^4 points was used to compute the errors. Pointwise RMS bias error was estimated using Equation (14) with constant $\gamma = 1$.

To compute actual RMS bias errors, a large number of true polynomials were generated by randomly selecting the coefficient vectors $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$ from a uniform distribution over $[-1, 1]$. For each true polynomial, the true response was obtained using Equation (1) and the vector $\hat{\boldsymbol{\beta}}^{(1)}$ was evaluated using Equation (4). The actual error at a point due to each true polynomial was computed by taking the difference between actual and predicted response as

$$e_c(\mathbf{x}) = y(\mathbf{x}) - \hat{y}(\mathbf{x}) = (\mathbf{f}^{(1)}(\mathbf{x}))^T \boldsymbol{\beta}_c^{(1)} + (\mathbf{f}^{(2)}(\mathbf{x}))^T \boldsymbol{\beta}_c^{(2)} - (\mathbf{f}^{(1)}(\mathbf{x}))^T \hat{\boldsymbol{\beta}}_c^{(1)} \tag{A4}$$

where the subscript c represents an instance of the true polynomial. Pointwise actual RMS bias errors $e_{act}^{rms}(\mathbf{x})$ were estimated by averaging actual errors over a large number of polynomials in the RMSs sense as

$$e_{act}^{rms}(\mathbf{x}) = \sqrt{\sum_{c=1}^{N_P} e_c^2(\mathbf{x}) \bigg/ N_P} \tag{A5}$$

Here, we used $N_P = 100\,000$. The maximum and space-averaged errors along with the correlations between the actual and predicted error estimates are given in Table AII. Maximum and space-averaged values of the actual and predicted RMS bias errors compared very well. The correlations between pointwise actual and predicted RMS bias errors were also high. The small change in maximal actual RMS bias error with the location of axial design point confirmed the outcome that maximal RMS bias errors had low variation with respect to the parameter α_2 and placing the axial points close to the center minimized the maximal actual RMS error.

Table AII. Comparison of actual and predicted RMS bias errors for 30-point experimental
designs in four-dimensional space.

α_1	α_2	$(e_{\mathrm{b}}^{\mathrm{rms}})_{\max}$	$(e_{\mathrm{act}}^{\mathrm{rms}})_{\max}$	$(e_{\mathrm{b}}^{\mathrm{rms}})_{\mathrm{avg}}$	$(e_{\mathrm{act}}^{\mathrm{rms}})_{\mathrm{avg}}$	$r(e_{\mathrm{b}}^{\mathrm{rms}}, e_{\mathrm{act}}^{\mathrm{rms}})$
1.000	0.100	1.155	1.158	0.927	0.927	1.000
1.000	1.000	1.176	1.180	0.827	0.827	1.000

Note: Errors are computed on a uniform 11^4 grid.

APPENDIX B: RMS BIAS ERROR ESTIMATES FOR NON-POLYNOMIAL EXAMPLE

Bias error estimates are sometimes criticized because of the assumption that a higher-order polynomial than the fitting model is the true function. We demonstrate that this assumption on the true model is practically useful if it captures the main characteristics of the true unknown function. Suppose the true function is a trigonometric polynomial given by

$$\eta(\mathbf{x}) = \sum_{\substack{j+k \leqslant 6 \\ j,k \geqslant 0}} \mathrm{Real}(-\mathrm{i}a_{jk}e^{ijx_1}e^{ikx_2}), \quad x_1, x_2 \in [-1, 1] \tag{B1}$$

The coefficients a_{06}, a_{60} were assumed to be uniformly distributed over $[-10, 10]$ and the remaining a_{jk} were assumed to be uniformly distributed over $[-1, 1]$. This function represents the cumulative response of *sine* waves of different wavelengths. For the given range of parameters j, k, x_1, and x_2, the highest frequency (shortest wavelength corresponding to $j=6$ and $k=6$) components in the true function have more than one cycle in design space. To estimate actual errors, 10 000 ($N_p = 10000$) combinations of a_{jk} were used.

This function was approximated using a cubic polynomial and bias errors were estimated assuming the true model to be quintic. A uniform 4×4 grid was used for sampling (16 points) and a uniform 21×21 grid was used to estimate errors. The distribution of actual RMS bias error and predicted RMS bias error in design space is shown in Figure B1. The predicted RMS bias error was scaled by a factor of 55 to compare with the actual errors in design space. Note that prior to generation of data, the actual magnitude of the error was not important as the estimated errors were scaled by an 'unknown' factor γ (refer to Equation (14)) that was arbitrarily set to one. The predicted RMS bias errors correctly identified the presence of high error zones along the diagonals. However, predictions were inaccurate near the center and close to the boundary, where the effect of high-frequency terms in the true function was significant. The correlation between the actual RMS bias errors and the predicted RMS bias errors was 0.68. The error contours and correlation coefficient indicated a reasonable agreement between actual RMS bias error and predicted RMS bias error estimates considering the fact that the true function was a high-frequency trigonometric polynomial.

For this example problem, the actual distribution of the coefficient vector $\boldsymbol{\beta}^{(2)}$ can be obtained from the distribution of constants a_{jk} in the true function by expanding each *sine* term in Maclaurin series as

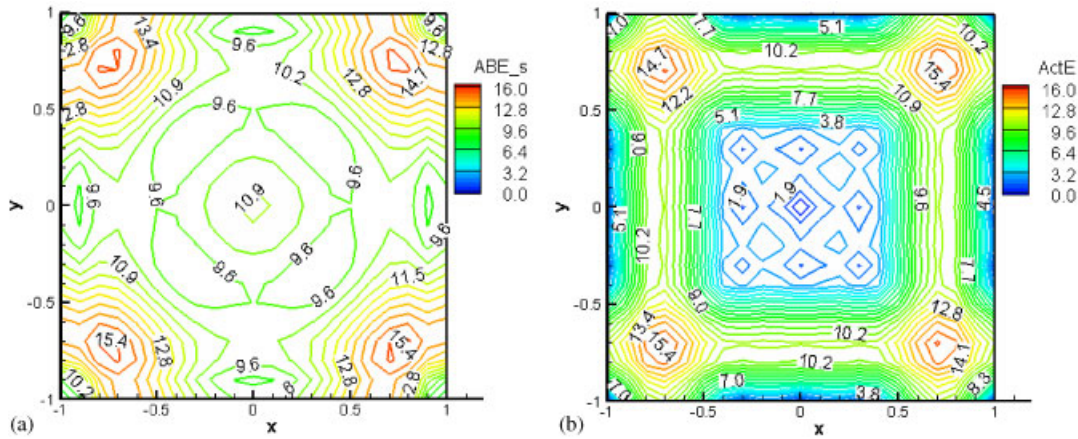$$\sin(x) = \sum_{n=1}^{\infty} \frac{(-1)^{n-1}}{(2n-1)!} x^{2n-1} \tag{B2}$$

Figure B1. Contours of scaled predicted RMS bias error (a) and actual RMS bias error (b) when assumed true model to compute bias error was quintic (same distribution assumed for all coefficients) while the true model was trigonometric (Equation (B1)) (scaled bias error$=55*$predicted RMS bias error).
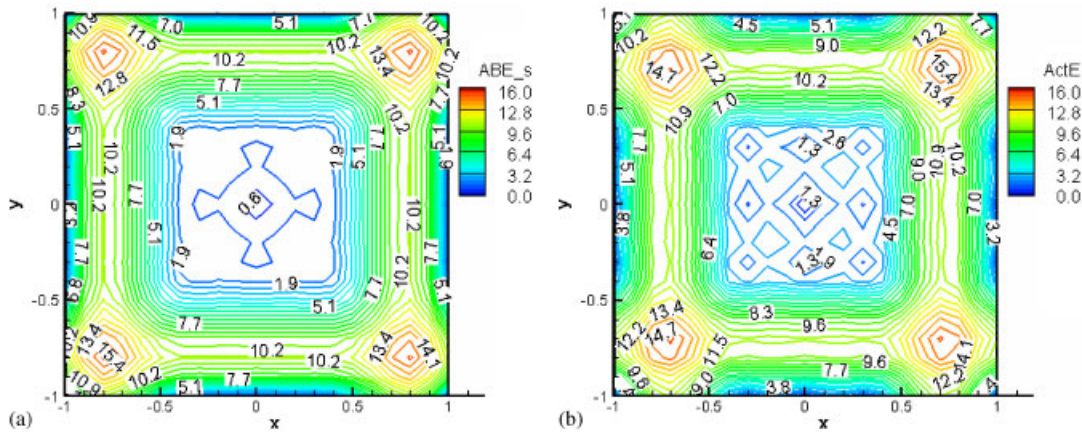

Figure B2. Contours of scaled predicted RMS bias error (a) and actual RMS bias error (b) near-actual distributions of $\boldsymbol{\beta}^{(2)}$ were specified (assumed true model was quintic while the true model was trigonometric Equation (B1)), (scaled bias error$=0.177*$RMS bias error).

and observing the coefficients of the quintic terms ($\eta(\mathbf{x})$ is odd, so there are no quartic terms). It was found that the coefficients $\beta_6^{(2)}$ and $\beta_{11}^{(2)}$ (corresponding to x_1^5 and x_2^5, respectively) approximately followed a uniform distribution with range $[-700, 700]$ and the coefficients $\beta_7^{(2)}$, $\beta_8^{(2)}$, $\beta_9^{(2)}$, and $\beta_{10}^{(2)}$ (corresponding to $x_1^4 x_2$, $x_1^3 x_2^2$, $x_1^2 x_2^3$, and $x_1 x_2^4$, respectively) followed approximately a uniform distribution with range $[-70, 70]$. Using the modified distribution of coefficients $E(\boldsymbol{\beta}^{(2)}\boldsymbol{\beta}^{(2)\mathrm{T}})$, RMS bias errors were estimated. The corresponding actual RMS bias error and scaled predicted RMS bias error (scaled by 0.177) contours are shown in Figure B2. For this case, the agreement between the actual RMS bias error and the predicted RMS bias error improved significantly (compare Figures B1(a) and B2(a)) and the correlation between the actual RMS bias errors and the

predicted RMS bias errors increased to 0.94. This indicates that supplying the correct distribution of the coefficient vector for the unmodeled basis functions $f^{(2)}(\mathbf{x})$ is helpful to assess the errors.

APPENDIX C: DIFFICULTIES IN FINDING D-OPTIMAL EDS

Identification of the D-optimal design requires identifying the coordinates of design points that maximize the determinant of the Gramian design matrix. All statistical packages (including Matlab) use a greedy exchange algorithm to compute D-optimal designs. Unless the objective function (a determinant in this case) is convex, all such greedy algorithms will converge to a point (design) at which no further exchanges will improve the objective function. However, there may be many other designs with a better objective function value that are not reachable from the converged point *via* the specified algorithm exchanges. In this sense, the spurious optimal design is a 'local' optimum point, where a 'neighborhood' of a design is defined as all designs (in the grid) obtainable by a single algorithmic exchange (and 'exchange' can mean any rule giving other points, as long as a single exchange does not include the entire grid).

To illustrate the local optima, we consider a problem of selecting 15 points in a three-dimensional $[-1, 1]^3$ design space (45 variables for the D-optimal problem). We simplify the problem by fixing eight points at the corners and six points on the faces of the cube of the design space. One point is allowed to vary on a face of the cube such that there are only two variables to identify the D-optimal design. To avoid duplication of the points (not practically useful for simulations), we limit the range of these two variables to $[-0.9, 0.9]$. We plot the determinant of the Gramian matrix (quadratic polynomial response surface) in Figure C1. The presence of many local maxima is clearly evident. The number of local optima increases significantly when there are more variables (more design points, high-dimension space). Our experience is that it is not realistic to expect convergence to a global optimum with EDs with even a modest number of points. We note that with the Matlab algorithm, different starting points usually result in different locally D-optimal design. Hence, to account for the variability in the creation of EDs, we present the results based on 100 samples.
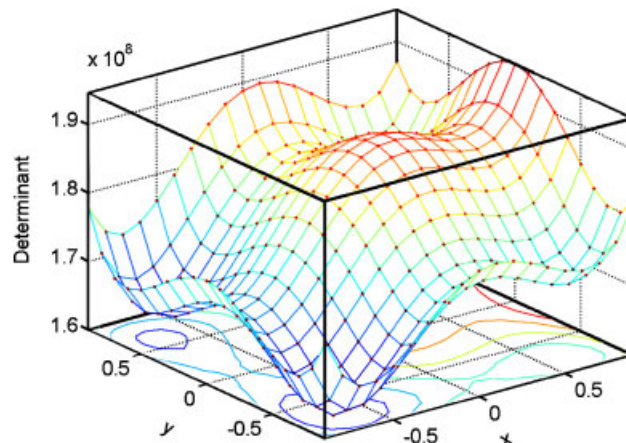

Figure C1. Variation of the determinant of a quadratic Gramian design matrix of 15 points in three variables. The points are placed at eight corners and seven faces of the cube. Two coordinates of one face point are allowed to vary.

NOMENCLATURE

A	alias matrix
$\mathbf{c}^{(2)}$	vector of bounds on the coefficient vector $\boldsymbol{\beta}^{(2)}$
D_{eff}	D-efficiency of an experimental design
ED	experimental design
$E_x(f(x))$	expected value of $f(\mathbf{x})$ with respect to random variable $\mathbf{x}$
$\mathbf{e}$	vector of true errors at the data points
$e(\mathbf{x})$	error at design point $\mathbf{x}$
$e_{\text{act}}^{\text{rms}}(\mathbf{x})$	actual RMS bias error at design point $\mathbf{x}$
$e_{\text{b}}^{\text{rms}}(\mathbf{x})$	predicted RMS bias error at design point $\mathbf{x}$
$e_{\text{b}}^{\text{I}}(\mathbf{x})$	data-independent bias error bound at design point $\mathbf{x}$
$e_{\text{b}}(\mathbf{x})$	true bias error at design point $\mathbf{x}$
$e_{\text{es}}(\mathbf{x})$	standard error at design point $\mathbf{x}$
$F(\mathbf{x}), F_i(\mathbf{x})$	true polynomial example functions
$\mathbf{f}(\mathbf{x}), \mathbf{f}^{(1)}(\mathbf{x}), \mathbf{f}^{(2)}(\mathbf{x})$	vectors of basis functions at design point $\mathbf{x}$
$\mathbf{f}_j(\mathbf{x})$	basis functions
LHS	Latin hypercube sampling
M	moment matrix
$N(\mu, \sigma^2)$	normal distribution with mean μ and standard deviation σ
N_s	number of design data points
N_v	number of design variables
n_1	number of basis functions in the regression model
n_2	number of missing basis functions in the regression model
RSA	polynomial response surface approximation
r_{max}	radius of the largest unoccupied sphere in the design space
V	design space
$X, X^{(1)}, X^{(2)}$	Gramian (design) matrices
$\mathbf{x}$	design point (vector)
$\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$	design variables
$x_1^{(i)}, x_2^{(i)}, \ldots, x_n^{(i)}$	design variables for ith design point
$\mathbf{y}$	vector of observed responses at design data points
$\mathbf{y}(\mathbf{x})$	observed response at design point $\mathbf{x}$
$\hat{\mathbf{y}}(\mathbf{x})$	predicted response at design point $\mathbf{x}$
$\alpha_1, \alpha_2, \ldots, \alpha_m$	parameters used to construct EDs
$\boldsymbol{\beta}, \boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}$	vectors of basis function coefficients
$\beta_j, \beta_j^{(1)}, \beta_j^{(2)}$	coefficients of basis functions
$\hat{\boldsymbol{\beta}}^{(1)}$	vector of estimated coefficients of basis functions
γ	range of uniform distribution of coefficient vector $\boldsymbol{\beta}^{(2)}$
ε	noise in the response
$\eta(\mathbf{x})$	true mean response at design point $\mathbf{x}$
σ^2	noise variance
σ_a	root mean square (rms) error
$\mathbf{1}$	vector of ones

$(\cdot)_{max}$ maximum value of the quantity over the design space
$(\cdot)_{avg}$ space-averaged value of the quantity

REFERENCES

1. Simpson TW, Lin DKJ, Chen W. Sampling strategies for computer experiments: design and analysis. *International Journal of Reliability and Applications* 2001; **2**(3):209–240.
2. Giunta AA, Wojtkiewicz SF, Eldred MS. Overview of modern design of experiments methods for computational simulations. *Forty-first AIAA Aerospace Sciences Meeting and Exhibit*, Reno, NV, 6–9 January 2003; AIAA-2003-0649.
3. Simpson TW, Booker AJ, Ghosh D, Giunta AA, Koch PN, Yang R-J. Approximation methods in multidisciplinary analysis and optimization: a panel discussion. *Structural and Multidisciplinary Optimization* 2004; **27**(5):302–313.
4. Queipo NV, Haftka RT, Shyy W, Goel T, Vaidyanathan R, Tucker PK. Surrogate based analysis and optimization. *Progress in Aerospace Sciences* 2005; **41**(1):1–28.
5. Myers RH, Montgomery DC. *Response Surface Methodology—Process and Product Optimization Using Designed Experiments*. Wiley: New York, 1995; 208–279.
6. Khuri AI, Cornell JA. *Response Surfaces*: *Designs and Analyses* (2nd edn). Marcel Dekker: New York, 1996; 207–247.
7. Venter G, Haftka RT. Minimum-bias based experimental design for constructing response surfaces in structural optimization. *Thirty-eight AIAA/ASME/ASCE/AHS/ASC Structures*, *Structural Dynamics*, *Materials Conference*, Kissimmee, FL, 1997; 1225–1238, AIAA-1997-1053.
8. Qu X, Venter G, Haftka RT. New formulation of a minimum-bias experimental design based on Gauss quadrature. *Structural and Multidisciplinary Optimization* 2004; **28**(4):231–242.
9. Box GEP, Draper NR. The choice of a second order rotatable design. *Biometrika* 1963; **50**(3):335–352.
10. Draper NR, Lawrence WE. Designs which minimize model inadequacies: cuboidal regions of interest. *Biometrika* 1965; **52**(1–2):111–118.
11. Kupper LL, Meydrech EF. A new approach to mean squared error estimation of response surfaces. *Biometrika* 1973; **60**(3):573–579.
12. Welch WJ. A mean squared error criterion for the design of experiments. *Biometrika* 1983; **70**(1):205–213.
13. Montepiedra G, Fedorov VV. Minimum bias designs with constraints. *Journal of Statistical Planning and Inference* 1997; **63**:97–111.
14. Fedorov VV, Montepiedra G, Nachtsheim CJ. Design of experiments for locally weighted regression. *Journal of Statistical Planning and Inference* 1999; **81**:363–383.
15. Papila M, Haftka RT. Uncertainty response surface approximations. *Forty-second AIAA/ASME/ASCE/ASC Structures*, *Structural Dynamics*, *and Material Conference*, Seattle, WA, 2001; AIAA-01-1680.
16. Papila M, Haftka RT, Watson LT. Pointwise bias error bounds and min–max bias design for response surface approximations. *AIAA Journal* 2005; **43**(8):1797–1807.
17. Goel T, Papila M, Haftka RT, Shyy W. Generalized pointwise bias error bounds for response surface approximation. *International Journal for Numerical Methods in Engineering* 2006; **65**(11):2035–2059.
18. Goel T, Haftka RT, Shyy W, Watson LT. Pointwise RMS bias error estimates for design of experiments. *Forty-fourth AIAA Annual Aerospace Science Meeting and Exhibit*, Reno, NV, January 2006; AIAA-2006-0724.
19. Tang B. Orthogonal array-based Latin hypercubes. *Journal of American Statistical Society* 1993; **88**:1392–1397.
20. Ye K. Orthogonal column Latin hypercubes and their application in computer experiments. *Journal of the American Statistical Association* 1998; **93**:1430–1439.
21. Palmer K, Tsui KL. A minimum bias Latin hypercube design. *Institute of Industrial Engineers Transactions* 2001; **33**(8):793–808.

22. Leary S, Bhaskar A, Keane A. Optimal orthogonal-array-based Latin hypercubes. *Journal of Applied Statistics* 2003; **30**(5):585–598.
23. Goel T, Zhao J, Thakur SS, Haftka RT, Shyy W. Surrogate model-based strategy for cryogenic cavitation model validation and sensitivity evaluation. *International Journal for Numerical Methods in Fluids*, submitted (an earlier version appeared in *Proceedings of 42nd AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit*, Sacramento, CA, July 2006; AIAA-2006-5047).
24. Jaynes ET. Information theory and statistical mechanics. In *Statistical Physics*, Ford K (ed.). Benjamin: New York, 1963; 181.
25. JMP®, *The Statistical Discovery Software*™, *Version 5*, © *1989–2003*. SAS Institute Inc., Cary, NC, U.S.A., 1989–2003.
26. MATLAB®. *The Language of Technical Computing*, *Version 6.5*, *Release 13*, © *1984–2002*. The MathWorks Inc., 1984–2002.
27. Goel T, Haftka RT, Shyy W, Watson LT. Combining bias and variance based criteria for selecting experimental designs. *2006 NSF Design*, *Manufacturing*, *and Innovation Grantees and Research Conference*, St. Louis, MO, July 2006.