AIAA JOURNAL
Vol. 43, No. 8, August 2005

Pointwise Bias Error Bounds and Min–Max Design for Response Surface Approximations

Melih Papila* and Raphael T. Haftka†
University of Florida, Gainesville, Florida 32611-6250
and
Layne T. Watson‡
Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061-0106

Bounds on response surface approximation errors as a result of model inadequacy (bias error) are presented, and a design of experiments minimizing the maximal bias error is proposed. The bias-error bounds are considered as a tool to identify locations in the design space where the accuracy of the approximation fitted on a given design of experiments might be poor. Two approaches to characterize the bias error assume that the functional form of the true model is known and seek, at each point in design space, worst-case bounds on the absolute error. The first approach is implemented before data generation. This data-independent error bound can easily be implemented in a search for a design of experiments that minimize the bias error bound as it requires very little computation. The second approach is to be used posterior to the data generation and provides tightened error bound consistent with the data. This data-dependent error bound requires the solution of two linear-programming problems at each point. The data-independent error bound for design of experiments of two-variable examples is demonstrated. Randomly generated polynomials in two variables are then used to validate the data-dependent bias-error bound distribution. Although the two approaches are used in conjunction in the given examples, the data-independent error bound design of experiment is not a prerequisite for the application of the data-dependent error bounds in search for the high bias-error regions.

Nomenclature

A	=	alias matrix
$\boldsymbol{b}$	=	vector of estimated coefficients of basis functions
b_j	=	estimated coefficients of basis functions
$\boldsymbol{c}^{(2)}$	=	vector of bounds for coefficients of the basis functions
$\boldsymbol{E}_b$	=	vector of true prediction errors caused by bias at the data points
$e_b(x)$	=	true bias error at design point x
$\lvert e_b^D(x)\rvert$	=	data-dependent worst-case bias error bound at design point x
$\lvert e_b^I(x)\rvert$	=	data-independent worst-case bias error bound at design point x
$\lvert e_b^I\rvert_{\text{av}}$	=	average of data-independent bias-error bounds over the design domain
$\lvert e_b^I\rvert_{\max}$	=	data-independent maximum absolute bias-error bound over the design domain
$e_{\text{es}}(x)$	=	estimated standard error at design point x
$[\bar{e}_{\text{es}}]_{\max}$	=	maximum of normalized estimated standard error over the design domain
$\boldsymbol{F}(x), \boldsymbol{F}^{(1)}(x), \boldsymbol{F}^{(2)}(x)$	=	vectors of basis functions at x
$f_j(x)$	=	basis functions
N	=	number of data points
n_1	=	number of basis functions in the regression model
n_2	=	number of missing basis functions in the regression model
R	=	domain for the response surface approximation
s^2	=	error mean square
$X, X^{(1)}, X^{(2)}$	=	Gramian (design) matrices
x	=	design point (vector of design variables)
$x_1, x_2, \ldots, x_n$	=	design variables
$x_1^{(i)}, x_2^{(i)}, \ldots, x_n^{(i)}$	=	design variables for ith design point
$\hat{\boldsymbol{Y}}$	=	vector of predictions
$\boldsymbol{y}$	=	vector of observed responses
$\hat{y}(x)$	=	response surface approximation at a design point x
$\alpha_1, \alpha_2, \ldots, \alpha_m$	=	design of experiment parameters
$\boldsymbol{\beta}, \boldsymbol{\beta}^{(1)}, \boldsymbol{\beta}^{(2)}$	=	vectors of basis function coefficients
$\beta_j, \beta_j^{(1)}, \beta_j^{(2)}$	=	coefficients of basis functions
$\eta(x)$	=	true mean response at x
σ^2	=	noise variance

Received 26 March 2004; revision received 10 February 2005; accepted for publication 15 February 2005.

*Research Associate, Mechanical and Aerospace Engineering; currently Assistant Professor, Faculty of Engineering and Natural Sciences, Sabancı University Orhanlı Tuzla, 34956 Istanbul, Turkey; mpapila@sabanciuniv.edu. Member AIAA.

†Distinguished Professor, Mechanical and Aerospace Engineering; haftka@ufl.edu. Fellow AIAA.

‡Professor, Departments of Computer Science and Mathematics; ltw@cs.vt.edu.

Background

RESPONSE-SURFACE-APPROXIMATION (RSA) techniques have become popular in engineering optimization based on computer simulations. Kaufman et al.,[1] for instance, fitted quadratic polynomial RSAs to the structural weight of the high-speed civil transport (HSCT), obtained with multiple structural optimizations via the commercial code GENESIS. Balabanov et al.[2,3] investigated RSA construction for the wing bending material weight of the HSCT based on structural optimization results of a number of different configurations and use of the RSA in the configuration optimization problem. Papila and Haftka[4,5] also constructed RSAs for weight equations based on structural optimizations. Shyy et al.[6] employed RSA methodology for rocket-engine injector design optimization. Papila et al.[7] constructed RSAs based on computational-fluid-dynamics simulations for supersonic turbine-blade design optimization. Redhe et al.[8] determined an efficient number of data points when using the RSA methodology in crashworthiness problems,

where numerical simulations were carried out by using Livermore Software-Dynamic Solver (LS-DYNA). A list of more RSA applications in aerospace and mechanical engineering design can be found in the survey paper by Simpson et al.[9]

This popularity has brought attention to ways of increasing the accuracy of RSA and protecting designers from misleading predictions. The accuracy of RSA is mainly affected by the following three factors: 1) limitations on the number of data points caused by the cost of data generation, 2) noise in the data, and 3) inadequacy of the fitting model (bias errors).

The prediction error of an RSA varies from point to point. There is an established statistical tool for characterizing noise-related prediction errors: prediction variance.[10,11] However, in many computer simulations numerical noise or the random effects are not the major problem,[9] and RSA accuracy suffers, mostly because of systematic errors such as bias errors. There has been a fair amount of work on design of experiments for minimizing the mean squared error averaged over the design domain that combines variance and bias errors.[12–15] The bias component of the averaged or integrated mean squared error was also minimized to obtain so-called minimum bias designs. The fundamentals of minimizing integrated mean squared error and its components can be found in Myers and Montgomery[10] and Khuri and Cornell.[11] Venter and Haftka[16] developed an algorithm implementing a minimum-bias criterion, necessary for an irregularly shaped domain where no closed-form solution exists for minimum-bias experimental design. They compared minimum-bias and D-optimal experimental designs for two problems with two and three variables. The minimum-bias experimental design was found to be more accurate than D-optimal for the two problems for the average error, but not for the maximum error. Montepiedra and Fedorov[17] investigated experimental designs minimizing the bias component of the integrated mean square error subject to a constraint on the variance component or vice versa. Fedorov et al.[18] studied later design of experiments via weighted regression prioritizing regions, where the approximation is desired to predict the response. Their approach considered both the variance and bias components of the estimation error. Palmer and Tsui[19] studied minimum-bias Latin hypercube experimental design for sampling from deterministic process simulators. Recently Qu et al.[20] have implemented Gaussian quadrature-based minimum-bias design. They have also presented minimum bias central composite designs for up to six variables.

Objectives

Bias error, averaged over the domain of interest, has been studied extensively in the approximation theory and numerical analysis literature.[10,11] However, very little has been done on the point-to-point variation of bias errors. This variation is of interest for two reasons. First, when using an approximation, it can be important to identify regions where the bias error can be large. Second, when selecting a design of experiments, it might be desirable to minimize the maximal bias error (min–max bias error) rather than the average. An approach for estimating a bound on RSA squared bias errors was previously presented by Papila and Haftka.[21] The traditional decomposition of the mean squared error into variance and the square of the bias was followed, but point to point rather than averaged over the domain. Their bound on the squared bias error does not depend on the response data as determined by the assumed true function form, locations of the design points [design of experiment (DOE)], and the prescribed bounds on the coefficients that are not included in the approximation [Papila and Haftka[21] and also in Eq. (10)]. Such an error bound can be used to identify regions in design space where the accuracy of the RSA might be poor for a given design of experiments. Moreover, the error bound can provide a tool for designing experiments against the bias error prior to data generation. The first objective of this paper is to demonstrate how the error bound based on absolute bias error can be used to obtain designs of experiments that minimize the maximal absolute bias.

Once data are available, the error bound can be tightened, and the second objective of the paper is to develop a point-to-point error bound for a given set of data. This error bound can identify regions of possible large errors and can be used to adaptively refine a design of experiments or to steer clear of such regions.

The next two sections, respectively, present the bias-error expression at a given design point and derive the bias-error bounds for a given design of experiment before data generation (data-independent bounds). This is followed by demonstration of a min–max bias design of experiments via data-independent bounds. Next, the data-dependent error bound is derived and demonstrated. The demonstrations are based on two-dimensional polynomial examples. The last section offers concluding remarks.

Bias Error in Response Surface Approximations

Response surface approximations, for which the fundamentals are reviewed in Appendix A, fit numerical or physical experimental data with an analytical model. The true function generating the data at a given design point x or the true mean response $\eta(x)$ is represented here as

$$\eta(x) = \boldsymbol{F}^{(1)}(x)^T \boldsymbol{\beta}^{(1)} + \boldsymbol{F}^{(2)}(x)^T \boldsymbol{\beta}^{(2)} \tag{1}$$

where $\boldsymbol{F}^{(1)}(x)$ and $\boldsymbol{F}^{(2)}(x)$ are the vectors of basis functions f_j [see Appendix A, Eq. (A2)], and $\boldsymbol{\beta}^{(1)}$ and $\boldsymbol{\beta}^{(2)}$ are the coefficient vectors associated with the basis function vectors. The superscript (1) denotes the sets of basis functions in the linear regression procedure. The superscript (2) denotes terms in the true function, Eq. (1), that are missing from the regression (often because of the cost of a comprehensive data matrix, in particular for high-dimensional problems). As the present paper deals with the bias error caused by partial characterization of the analytical model, the predicted response at a given design point x is given as

$$\hat{y}(x) = \boldsymbol{F}^{(1)}(x)^T \boldsymbol{b} \tag{2}$$

where $\boldsymbol{b}$ is the estimate of the coefficient vector $\boldsymbol{\beta}^{(1)}$, which is biased by the presence of the $\boldsymbol{\beta}^{(2)}$ in the true function. The true error in the approximation at a design point x is

$$e_b(x) = \eta(x) - \hat{y}(x) \tag{3}$$

In this study the true function is assumed to be a polynomial. Polynomials are popular in RSA, and in addition, assuming smoothness locally, all functions look like polynomials according to Taylor's theorem. Because one usually does not know the functional form of the true response, it is often assumed to be a higher-order polynomial. For instance, for a quadratic fitting model in two variables when the true function is cubic,

$$\begin{aligned}
\boldsymbol{F}^{(1)}(x) &= \begin{pmatrix} 1 & x_1 & x_2 & x_1^2 & x_1x_2 & x_2^2 \end{pmatrix}^T \\
\boldsymbol{\beta}^{(1)} &= \begin{pmatrix} \beta_1^{(1)} & \beta_2^{(1)} & \beta_3^{(1)} & \beta_4^{(1)} & \beta_5^{(1)} & \beta_6^{(1)} \end{pmatrix}^T \\
\boldsymbol{F}^{(2)}(x) &= \begin{pmatrix} x_1^3 & x_1^2x_2 & x_1x_2^2 & x_2^3 \end{pmatrix}^T \\
\boldsymbol{\beta}^{(2)} &= \begin{pmatrix} \beta_1^{(2)} & \beta_2^{(2)} & \beta_3^{(2)} & \beta_4^{(2)} \end{pmatrix}^T
\end{aligned} \tag{4}$$

Assuming that there will be no noise in the data y and the error is only associated with the modeling, the true response at the data points is

$$\eta = \boldsymbol{y} = X^{(1)}\boldsymbol{\beta}^{(1)} + X^{(2)}\boldsymbol{\beta}^{(2)} \tag{5}$$

where $X^{(1)}$ and $X^{(2)}$ are the design matrices corresponding to the basis functions in $F^{(1)}$ and $F^{(2)}$. The coefficient vector $\boldsymbol{b}$ in Eq. (2), that is, the biased estimates of the coefficient vector $\boldsymbol{\beta}^{(1)}$, becomes [after substitution of Eq. (5) into Eq. (A7) with $X = X^{(1)}$]

$$\boldsymbol{b} = \boldsymbol{\beta}^{(1)} + A\boldsymbol{\beta}^{(2)} \tag{6}$$

where $A = [(X^{(1)})^T X^{(1)}]^{-1}(X^{(1)})^T X^{(2)}$ is called the alias matrix. Substitution of Eq. (6) into Eq. (2) yields

$$\hat{y}(x) = F^{(1)}(x)^T \left(\beta^{(1)} + A\beta^{(2)}\right) \tag{7}$$

Unlike some previous work,[21,22] used here is the absolute bias error (rather than the squared error)

$$|e_b(x)| = \left|\boldsymbol{F}^{(1)}(x)^T\boldsymbol{\beta}^{(1)} + \boldsymbol{F}^{(2)}(x)^T\boldsymbol{\beta}^{(2)} - \boldsymbol{F}^{(1)}(x)^T\left(\boldsymbol{\beta}^{(1)} + A\boldsymbol{\beta}^{(2)}\right)\right|$$

$$|e_b(x)| = \left|m^T\boldsymbol{\beta}^{(2)}\right| \tag{8}$$

where

$$m = \boldsymbol{F}^{(2)}(x) - A^T\boldsymbol{F}^{(1)}(x) \tag{9}$$

Recall that one of the reasons why bias-error variation is of interest is to be able to identify points where the bias error can be large. Given $\boldsymbol{\beta}^{(2)}$, Eq. (8) is the exact bias error at a point of interest. Therefore, defining or assuming a bound on $\boldsymbol{\beta}^{(2)}$, which is unknown, can provide a tool for comparing the worst magnitude of $|e_b(x)|$ at the design points and determine relatively which regions might suffer the most as a result of an insufficient model.

Point-to-Point Bias-Error Bound Prior to Data Generation

The first approach dealing with the absolute bias error Eq. (8) makes use of the information in a given DOE and the assumed true function form, but does not use the function data. That is, the error bound developed here can be used before data generation. For this reason, it is called the data-independent worst-case bias-error bound.

It is assumed that whatever terms are missing from the fitting model are known, but there will not be enough data (for instance, because of computational or experimental cost) to calculate the corresponding coefficients $\boldsymbol{\beta}^{(2)}$. If one bounds the terms in $\boldsymbol{\beta}^{(2)}$, it is possible to formulate a maximization problem for the largest absolute bias error, Eq. (8), that can be experienced at any given design point for the worst possible $\boldsymbol{\beta}^{(2)}$ of the bound:

$$\max_{-c^{(2)} \le \beta^{(2)} \le c^{(2)}} |e_b(x)| \tag{10}$$

where $\boldsymbol{c}^{(2)} \ge 0$ is the bound on the $\boldsymbol{\beta}^{(2)}$ terms. As the bias error, Eq. (8), is expressed by the scalar product of the two vectors $\boldsymbol{m}$ [Eq. (9)] and $\boldsymbol{\beta}^{(2)}$, the solution of the maximization problem is easy to compute and given as

$$\left|e_b^I(x)\right| = \sum_{j=1}^{n_2} \operatorname{sgn}(m_j) m_j c_j^{(2)} \tag{11}$$

where superscript I labels the data-independent approach, m_j is the jth component of the vector $\boldsymbol{m}$ calculated at x, and n_2 is the size of the vectors, that is, the number of missing coefficients. Then the worst possible vector is $\boldsymbol{\beta}^{(2)} = [\operatorname{sgn}(m_1)c_1^{(2)}, \ldots, \operatorname{sgn}(m_{n_2})c_{n_2}^{(2)}]^T$, which characterizes the coefficients of the missing basis functions resulting in the largest bias error when fitting only with b. The worst $\boldsymbol{\beta}^{(2)}$ and therefore the true function corresponding to the worst possible bias error vary from point to point. The solution $|e_b^I(x)|$ reflects the assumed form of the true function with the basis function coefficients $\boldsymbol{\beta}^{(2)}$ [among all the possible combinations such that $-\boldsymbol{c}^{(2)} \le \boldsymbol{\beta}^{(2)} \le \boldsymbol{c}^{(2)}$] causing the largest error. It is viewed as a bound on the bias error given the assumptions on the true model and a bound $\boldsymbol{c}^{(2)}$ on the terms in $\boldsymbol{\beta}^{(2)}$. If the coefficient bound is scaled, $|e_b^I(x)|$ will simply scale in proportion without changing the locations identified as having high error potential. That is, the magnitude of coefficients is immaterial in terms of a relative or qualitative comparison over the domain. Here, the terms of the bound vector are taken as $\boldsymbol{c}_j^{(2)} = 1$.

Use of Error Bound as a Design of Experiment Criterion: Min–Max Bias Design

The data-independent error measure is obtained before data generation for a given DOE. That means it can easily be used to compare different choices of DOEs. Ultimately, it can be used to search for the DOE minimizing the maximum point-to-point bias-error bound $|e_b^I|_{\max}$ by using Eq. (11):

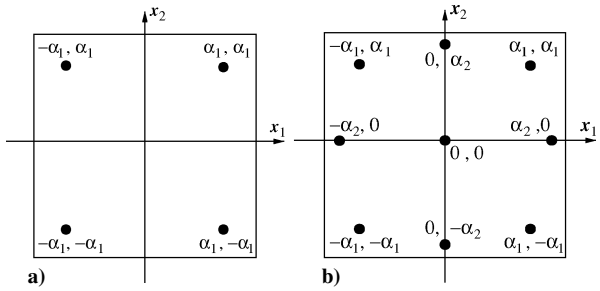

Fig. 1 Design of experiment in coded square domain $-1 \le x_1 \le 1$, $-1 \le x_2 \le 1$: a) example 1, linear-quadratic example; and b) example 2, quadratic-cubic example.

$$\left|e_b^I\right|_{\max} = \max_R \left|e_b^I(x)\right| \tag{12}$$

where R defines the domain for the RSA. The experimental designs derived here minimize the maximum bound given by Eq. (12) instead of the average square error, as is done with minimum bias designs.[10,11,16] The difference is analogous to H^∞ control vs H^2 control.

Consider a design of experiments when the domain R is an n-dimensional cube, $-1 \le x_1, x_2, \ldots, x_n \le 1$, for which the data points are determined by parameters $0 \le \alpha_1, \alpha_2, \ldots, \alpha_m \le 1$. For instance, two-level factorial points associated with the parameter α_m in the cube are 2^n combinations at $x_i = \pm\alpha_m$ for $i = 1, \ldots, n$. Figure 1 shows two examples for DOE as a function of the parameters α_m in two dimensions.

The minimum of the maximum absolute bias error given by Eq. (12) is referred to as min–max bias design and can be obtained by solving

$$\min_{0 \le \alpha_1, \alpha_2, \ldots, \alpha_m \le 1} \left|e_b^I\right|_{\max} \tag{13}$$

The min–max bias design from Eq. (13) is compared with the standard DOEs: minimum variance design and minimum bias design that minimizes the average square error. The metrics or error measures for the comparison of the experimental designs are 1) maximum absolute bias error $|e_b^I|_{\max}$, as given in Eq. (12) [obtained from Eq. (11) with $\boldsymbol{c}^{(2)} = 1$]; 2) average of absolute bias error $|e_b^I|_{\text{av}}$ [from Eq. (11) with $\boldsymbol{c}^{(2)} = 1$] over the domain $-1 \le x_1, x_2, \ldots, x_n \le 1$,

$$\left|e_b^I\right|_{\text{av}} = \frac{\int \left|e_b^I(x)\right| \mathrm{d}x}{\int \mathrm{d}x} \tag{14}$$

3) maximum normalized estimated standard error (square root of the prediction variance that predicts sensitivity to noise), $[\bar{e}_{\text{es}}]_{\max}$ [see Appendix A, maximum of Eq. (A9) normalized by s]:

$$[\bar{e}_{\text{es}}]_{\max} = \max_{-1 \le x_1, x_2, \ldots, x_n \le 1} e_{\text{es}}(x)/s \tag{15}$$

Two-dimensional polynomial examples that permit visualization of the error fields are presented for demonstration. Two cases are considered: first-order polynomial fit for a two-dimensional quadratic polynomial (linear-quadratic example), and quadratic fit for a two-dimensional cubic polynomial (quadratic-cubic example).

Example 1: Min–Max Bias Design of Linear Fit for Two-Dimensional Quadratic Polynomial

A two-variable quadratic true function fitted by a first-order polynomial is considered first:

$$\eta(x) = \beta_1^{(1)} + \beta_2^{(1)}x_1 + \beta_3^{(1)}x_2 + \beta_1^{(2)}x_1^2 + \beta_2^{(2)}x_1x_2 + \beta_3^{(2)}x_2^2$$

$$\hat{y}(x) = b_1 + b_2x_1 + b_3x_2 \tag{16}$$

Table 1 Design of experiments, example 1: two-dimensional linear-quadratic example (Fig. 1a)[a]

Design of experiment	α_1	$[\bar{e}_{es}]_{max}$[b]	$\lvert e_b^I\rvert_{av}$ for $c_j^{(2)}=1$[c]	$\lvert e_b^I\rvert_{max}$ for $c_j^{(2)}=1$[d]
Minimum noise variance	1.000	0.866	1.550	2.000
Minimum average bias	0.577	1.323	0.811	2.333
Minimum maximum bias	0.866	0.957	1.163	1.500

[a]In $-1 \le x_1 \le 1$, $-1 \le x_2 \le 1$. First-order polynomial fit at four data points $(-\alpha_1, -\alpha_1)$, $(-\alpha_1, +\alpha_1)$, $(+\alpha_1, -\alpha_1)$, and $(+\alpha_1, +\alpha_1)$.
[b]Maximum normalized standard error $[\bar{e}_{es}]_{max}$ [from Eq. (15)].
[c]Average of absolute bias errors $|e_b^I|_{av}$ [from Eq. (14) for $c_j^{(2)}=1$].
[d]Maximum absolute bias error [from Eq. (11) with $c_j^{(2)}=1$].

The four data points (Fig. 1a) are defined by a single variable α_1 that determines the experimental design within the coded square domain of $-1 \le x_1 \le 1$, $-1 \le x_2 \le 1$. The design matrices are given as

$$X^{(1)} = \begin{bmatrix} 1 & -\alpha_1 & -\alpha_1 \\ 1 & -\alpha_1 & \alpha_1 \\ 1 & \alpha_1 & -\alpha_1 \\ 1 & \alpha_1 & \alpha_1 \end{bmatrix}, \qquad X^{(2)} = \begin{bmatrix} \alpha_1^2 & \alpha_1^2 & \alpha_1^2 \\ \alpha_1^2 & -\alpha_1^2 & \alpha_1^2 \\ \alpha_1^2 & -\alpha_1^2 & \alpha_1^2 \\ \alpha_1^2 & \alpha_1^2 & \alpha_1^2 \end{bmatrix} \tag{17}$$

The minimum noise variance design is at the four vertices of the square domain ($\alpha_1 = 1$), while the standard minimum bias design (that minimizes the average square error) is at $(x_1, x_2) \approx (\pm 0.577, \pm 0.577)$ with $\alpha_1 = \sqrt{3}/3$ (Myers and Montgomery,[10] p. 415). In contrast, the min–max bias design, as the solution of Eq. (13), is found at $\alpha_1 = \sqrt{3}/2 \approx 0.866$. In other words, the maximum potential bias error in a linear model for a true quadratic response is minimized if the experimental design is $(x_1, x_2) \approx (\pm 0.866, \pm 0.866)$. Table 1 compares the minimum noise variance, minimum bias, and minimum-maximum bias error (min–max bias) experimental designs in terms of the three error measures. It shows that the maximum of the data-independent bound $|e_b^I|_{max}$, Eq. (12), is highest for the minimum bias design and lowest for the design where it was minimized in the domain. Average absolute bias error $|e_b^I|_{av}$, Eq. (14), reported in Table 1, is lowest for the minimum bias design as it minimizes the bias error in an average sense. The contour plots of the data-independent bound are presented in Fig. 2. Comparison of Figs. 2b and 2c shows that the data-independent error bound levels are generally lower for the minimum bias design, but the maximum error can be high on the boundary. This is because of the data points being much farther from the boundary for the minimum average bias design. In addition, the min–max bias design is much less sensitive to noise variance, as measured by $[\bar{e}_{es}]_{max}$, Eq. (15), than the minimum bias design (similar to an H^∞ optimal controller).

These observations indicate that min–max bias design handles well the tradeoff between sensitivity to noise variance and maximum bias error and provides a good compromise between the minimum noise variance and minimum bias designs.

Example 2: Min–Max Bias Design of Quadratic Fit for Two-Dimensional Cubic Polynomial

The second two-variable problem is a cubic polynomial fitted by a quadratic polynomial:

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

$$\hat{y}(x) = b_1 + b_2x_1 + b_3x_2 + b_4x_1^2 + b_5x_1x_2 + b_6x_2^2 \tag{18}$$

The experimental design now includes, in addition to center point (0,0), four factorial and four axial points determined by parameters

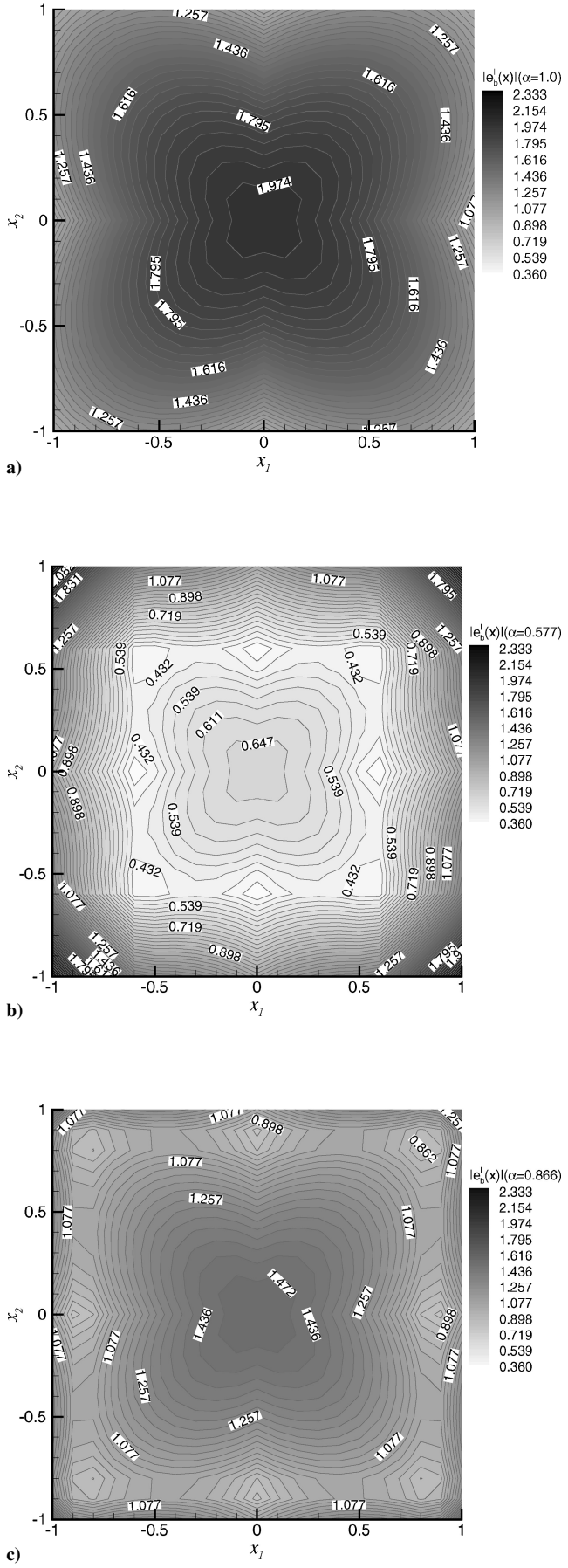

Fig. 2 Data-independent error bound in $-1 \le x_1 \le 1$, $-1 \le x_2 \le 1$ for linear-quadratic example (Fig. 1a): a) minimum noise variance design ($\alpha_1 = 1.0$), b) minimum average bias design ($\alpha_1 = 0.577$), and c) minimum maximum bias design ($\alpha_1 = 0.866$).

Table 2 Design of experiments, example 2: two-dimensional quadratic-cubic example (Fig. 1b)[a]

Design of experiment	α_1, α_2	$[\bar{e}_{es}]_{max}$[b]	$\lvert e_b^I\rvert_{av}$ for $c_j^{(2)} = 1$[c]	$\lvert e_b^I\rvert_{max}$ for $c_j^{(2)} = 1$[d]
Minimum noise variance	1.000,1.000	0.898	0.860	1.147
Minimum average bias	0.700,0.770	1.944	0.656	2.331
Minimum maximum bias	0.949,0.949	0.994	0.793	1.000

[a]In $-1 \le x_1 \le 1$, $-1 \le x_2 \le 1$.
[b]Maximum normalized standard error $[\bar{e}_{es}]_{max}$ [from Eq. (15)].
[c]Average of absolute bias errors $\lvert e_b^I\rvert_{av}$ [from Eq. (14) for $c_j^{(2)} = 1$].
[d]Maximum absolute bias error [from Eq. (11) with $c_j^{(2)} = 1$].

α_1, and α_2, respectively (Fig. 1b). The design matrices are given as

$$X^{(1)} = \begin{bmatrix} 1 & -\alpha_1 & -\alpha_1 & \alpha_1^2 & \alpha_1^2 & \alpha_1^2 \\ 1 & -\alpha_2 & 0 & \alpha_2^2 & 0 & 0 \\ 1 & -\alpha_1 & \alpha_1 & \alpha_1^2 & -\alpha_1^2 & \alpha_1^2 \\ 1 & 0 & -\alpha_2 & 0 & 0 & \alpha_2^2 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & \alpha_2 & 0 & 0 & \alpha_2^2 \\ 1 & \alpha_1 & -\alpha_1 & \alpha_1^2 & -\alpha_1^2 & \alpha_1^2 \\ 1 & \alpha_2 & 0 & \alpha_2^2 & 0 & 0 \\ 1 & \alpha_1 & \alpha_1 & \alpha_1^2 & \alpha_1^2 & \alpha_1^2 \end{bmatrix}$$

$$X^{(2)} = \begin{bmatrix} -\alpha_1^3 & -\alpha_1^3 & -\alpha_1^3 & -\alpha_1^3 \\ -\alpha_2^3 & 0 & 0 & 0 \\ -\alpha_1^3 & \alpha_1^3 & -\alpha_1^3 & \alpha_1^3 \\ 0 & 0 & 0 & -\alpha_2^3 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \alpha_2^3 \\ \alpha_1^3 & -\alpha_1^3 & \alpha_1^3 & -\alpha_1^3 \\ \alpha_2^3 & 0 & 0 & 0 \\ \alpha_1^3 & \alpha_1^3 & \alpha_1^3 & \alpha_1^3 \end{bmatrix} \tag{19}$$

The minimum noise variance design is achieved at $\alpha_1 = \alpha_2 = 1$. The standard minimum-bias design, minimizing the average square error, is with $\alpha_1 = 0.700$ and $\alpha_2 = 0.770$ (from Ref. 20). The min–max bias design is now obtained as the solution of Eq. (13) at $\alpha_1 = \alpha_2 = 0.949$. Table 2 compares the minimum noise variance, minimum average bias, and minimum maximum bias error (min–max bias) experimental designs in terms of three error measures.

As in example 1, the maximum of the data-independent bound $\lvert e_b^I\rvert_{max}$, Eq. (12), is the highest for the minimum bias design and lowest for the design where it was minimized in the domain. The contour plots for the data-independent bound are presented in Fig. 3. Comparisons of Figs. 3b and 3c reveal the fact that the minimum-bias design minimizes the bias error in an average sense (see Table 2 for average $\lvert e_b^I\rvert_{av}$), and so the data-independent error bound levels are generally lower, but the maximum error can be high. The two examples demonstrate that each of three design criteria produces a design that is optimal with respect to the criterion that produced it. The min–max bias experimental design, however, is a good compromise between the minimum noise variance and minimum-bias designs: it is much less sensitive to noise variance as measured by $[\bar{e}_{es}]_{max}$ than the minimum-bias design (Tables 1 and 2).

The squared error approach of Papila and Haftka[21] was previously integrated in a design space windowing procedure,[22] in which the regions of interest were zoomed by the prediction of a global RSA. Then, additional design points in the zoomed area were selected using a squared-error-based data-independent error bound. The

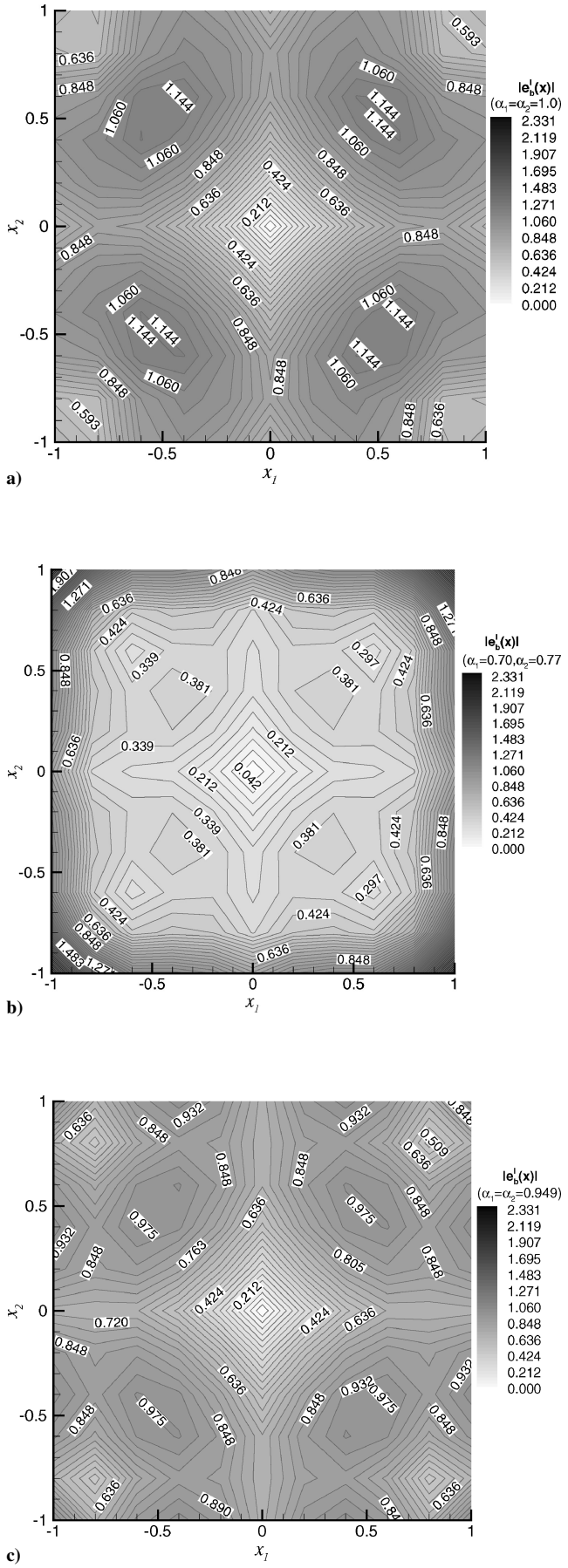

Fig. 3 Data-independent error bound in $-1 \le x_1 \le 1$, $-1 \le x_2 \le 1$ for quadratic-cubic example (Fig. 1b): a) minimum noise variance design ($\alpha_1 = \alpha_2 = 1.0$), b) minimum average bias design ($\alpha_1 = 0.700$, $\alpha_2 = 0.770$), and c) minimum-maximum bias design ($\alpha_1 = \alpha_2 = 0.949$).

present bias-error bounds are equally applicable for such adaptive design of experiment applications where more data are progressively generated.

Point-to-Point Bias-Error Bound Posterior to Data Generation

Once data are generated for a given design of experiment, the data-dependent worst-case bias-error bounds can be obtained. They are tighter than the data-independent error bounds for the same design of experiment. That is, sharper bounds on the bias error can now be obtained by constraining $\boldsymbol{\beta}^{(2)}$ to satisfy the data. This restricts consideration to all possible polynomials with coefficients that interpolate the given data points $(x^{(i)}, y_i)$ and result in the approximation $\hat{y}(x)$ associated with the data. Note that a data-independent error bound is not a prerequisite for the data-dependent bound. The two types, however, can be used in conjunction as suggested in this paper through the min–max bias design.

The bias error of the approximation $\hat{y}(x)$ at a design point x is given by Eq. (8). This relation along with Eq. (9) leads to the prediction error vector caused by bias at the data points $x^{(i)}$:

$$\boldsymbol{E}_b = \boldsymbol{y} - \hat{\boldsymbol{Y}} = \left[X^{(2)} - X^{(1)}A\right]\boldsymbol{\beta}^{(2)} \tag{20}$$

where $\hat{\boldsymbol{Y}}$ is the vector of predictions $\hat{y}_i$ by the response surface approximation at the data points.

The optimization problem now is to find the maximum absolute error at a design point subject to constraints on the magnitude of the coefficients and $\eta(x)$ matching the given data. As the latter constraint is equivalent to matching the prediction error vector at the data points, Eq. (20), the maximization of the error can be formulated as a linear program (LP):

$$\max_{-c^{(2)} \le \beta^{(2)} \le c^{(2)}} |e_b(x)| \qquad \text{such that} \qquad \left[X^{(2)} - X^{(1)}A\right]\boldsymbol{\beta}^{(2)} = y - \hat{Y} \tag{21}$$

The solution of the system (21) requires solving two LP problems, one for the minimum and one for the maximum error $e_b(x)$. The maximum of the absolute values of the two solutions to Eq. (21) is the data-dependent error bound $|e_b^D(x)|$ for data $\boldsymbol{y}$ (superscript D denotes the data-dependent bound). Like the data-independent error bound, this is also a worst-case search (and hence a bound on the bias error). The computational effort required for the data-dependent bound can grow large as the number of coefficients $\boldsymbol{\beta}^{(2)}$ that are variables for the optimization (LP) problems increases with the dimension of the problem.

Selection of the bounds on the coefficients is an important part of the present approach. The data-independent error bound does not make use of information on the function values, and so it makes sense to have the same coefficient bounds $\boldsymbol{c}^{(2)}$ [Eq. (10)] on all of the coefficients, and the magnitude is immaterial in terms of a relative or qualitative comparison over the domain. In contrast, the choice of the bounds $\boldsymbol{c}^{(2)}$ in the data-dependent error bound [Eq. (21)] must be compatible with the data. If variables are normalized or coded so that all of the data satisfy $\|x\|_\infty \le 1$, one logical selection criterion is to use the error magnitudes at the data points as an indication of the magnitude of the missing coefficients. This fails, however, when the errors at the data points are all small. Another indication is the magnitude of the coefficients $\boldsymbol{b}$. With normalized (coded) variables, one could assume that the coefficients $\boldsymbol{\beta}^{(2)}$ have magnitudes that are a fraction of the largest coefficient of $\boldsymbol{b}$. Finally, one could assume that the missing coefficients are a fraction of the maximum function value minus the value of the constant coefficient b_1. The choice of coefficient bounds $\boldsymbol{c}^{(2)}$ for the examples of this paper is explained in Appendix B.

Note that the LP problem with the equality constraints as given in Eq. (21) might not always be feasible because for many standard design of experiments the rows of the matrix $[X^{(2)} - X^{(1)}A]$ might be linearly dependent, and so the matrix is rank deficient. With such linear dependence, if the data are not exactly consistent with the assumed true model (because of noise or small deviations of the true function from the assumed true model) the equalities cannot be satisfied. This can be resolved by relaxing the equality constraint into a pair of inequality constraints with a small tolerance. Such rank-deficient DOEs have the advantage that they provide information on the magnitude of deviation from the assumed true model.

The two polynomial examples of the preceding section are used for demonstration along with the min–max bias designs developed. Note, however, that such experimental designs are not prerequisite or essential for the use of data-dependent error bounds. Also, recall that the data-dependent error bounds $|e_b^D(x)|$ are worst-case measures as they characterize the worst possible polynomials point to point subject to available data. Therefore, effective demonstration of the usefulness of the error bounds must consider a large number of quadratic and cubic polynomials in the first and second examples, respectively.

For the next two example cases, five different sets of data and 100 different polynomials for each set were randomly selected. This gives 100 randomly selected polynomials per data set to evaluate the data-dependent bound. The actual RSA error at x for the lth polynomial with the kth data set is denoted by

$$|e_b(x)|_{kl} = |\eta(x) - \hat{y}(x)|_{kl} \tag{22}$$

where the first index k denotes the data set (A, B, C, D, or E) and the second index $l = 1, 2, 3, \ldots, 100$ denotes the polynomial. Then the worst-case error at x for the kth data set and the overall worst-case error at x, respectively, are

$$|e_b(x)|_{k\,\max} = \max_l [|\eta(x) - \hat{y}(x)|_{kl}] \tag{23}$$

$$|e_b(x)|_{\text{overall}} = \max_k [|e_b(x)|_{k\,\max}] \tag{24}$$

Example 1: Linear Fit for Two-Dimensional Quadratic Polynomial Using the Min–Max Bias Design

For this example, with only bias error, substituting Eq. (17) into Eq. (20), the absolute prediction error at the four data points depends on $\beta_2^{(2)}$ alone [equal to $\alpha_1^2\beta_2^{(2)}$]. As a result, selecting the data randomly is equivalent to random selection of coefficients b_1, b_2, b_3, and $\beta_2^{(2)}$. Once the data are selected, different polynomials that interpolate these data can be obtained by generating random $\beta_1^{(2)}$ and $\beta_3^{(2)}$, which also determine $\beta^{(1)}$ because the RSA coefficients, from Eq. (6), are expressed as

$$b_1 = \beta_1^{(1)} + \alpha_1^2\left[\beta_1^{(2)} + \beta_3^{(2)}\right], \qquad b_2 = \beta_2^{(1)}, \qquad b_3 = \beta_3^{(1)} \tag{25}$$

Appendix B describes the details of the selection of data sets and associated random polynomials and the bounds $\boldsymbol{c}^{(2)}$ on the coefficients.

The data-dependent error bound $|e_b^D(x)|$ was computed using Eq. (21) for data sets A, . . ., E (associated 100 random polynomials each) and compared to the actual error distribution. The contour plots for data sets C and E presenting the actual worst-case errors (over associated 100 polynomials each) and the data-dependent error bound are given in Figs. 4 and 5, respectively. They show good agreement and suggest that the error bound determines successfully the map of the worst-case errors in the design space as a result of the available approximation. It suggests that the RSA might be vulnerable to prediction error at the center of the design domain. The corresponding coefficients of correlation between the error bounds and worst errors are calculated over 11×11 uniform grid points and given in Table 3.

Note that the total of 500 polynomials (100 for each data set A–E) is indeed a subset of the quadratic polynomials of the $-\boldsymbol{c}^{(2)} \le \boldsymbol{\beta}^{(2)} \le \boldsymbol{c}^{(2)}$ [$\boldsymbol{c}^{(2)} = 1$ as described in Appendix B]. Therefore, the maximum of the errors among all 500 polynomials can be characterized by the data-independent error bound $|e_b^I(x)|$ for the min–max bias experimental design [from Eq. (11) as shown also in Fig. 3c]. Correlations of data-independent error bounds with the worst, actual errors for the kth individual data set $|e_b(x)|_{k\,\max}$ and the worst of all 500 polynomials $|e_b(x)|_{\text{overall}}$, Eq. (24), are also presented in Table 3. As expected, coefficients of correlation of $|e_b^I(x)|$ with actual error for individual data sets are generally lower than the correlations obtained for the data-dependent bound $|e_b^D(x)|$. The

Table 3 Linear-quadratic example: coefficients of correlation between error bounds and the actual worst errors calculated on a uniform 11 × 11 grid

Bias error bound	$\lvert e_b(x)\rvert_{A\max}$	$\lvert e_b(x)\rvert_{B\max}$	$\lvert e_b(x)\rvert_{C\max}$	$\lvert e_b(x)\rvert_{D\max}$	$\lvert e_b(x)\rvert_{E\max}$	$\lvert e_b(x)\rvert_{\text{overall}}$
$\lvert e_b^D(x)\rvert$	0.9953	0.9374	0.9206	0.9744	0.9966	N/A
$\lvert e_b^I(x)\rvert$	0.7365	0.9417	0.9537	0.8877	0.5959	0.9650

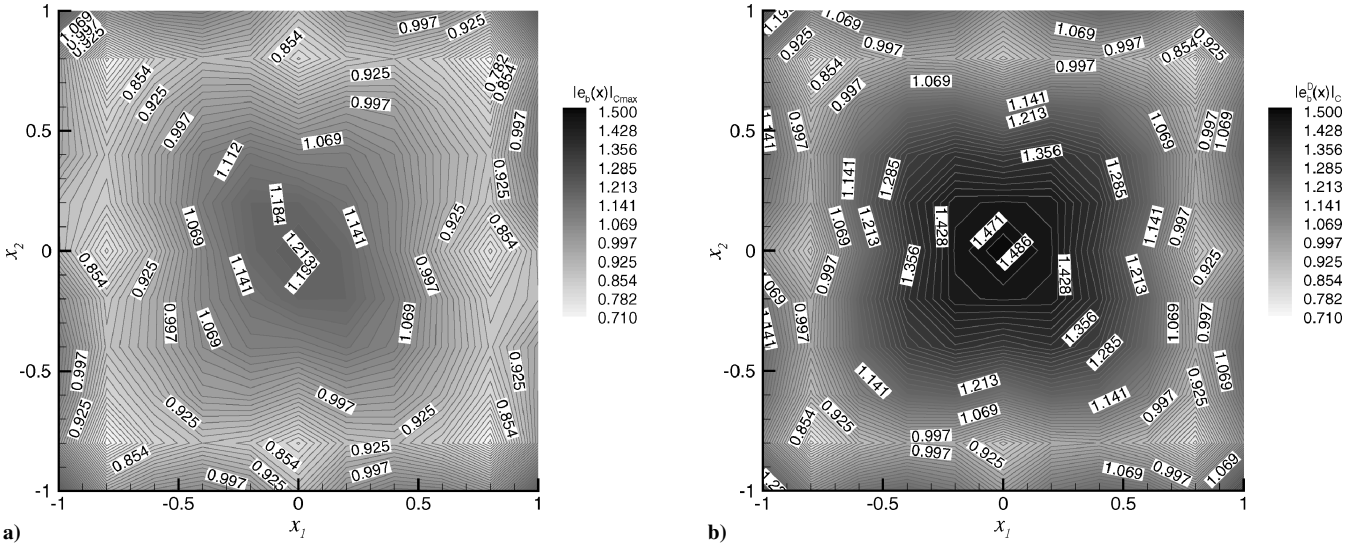

Fig. 4 Comparison of absolute worst-case errors for data set C in example 1: a) worst of 100 polynomials and b) data-dependent bound $|e_b^D(x)|$.

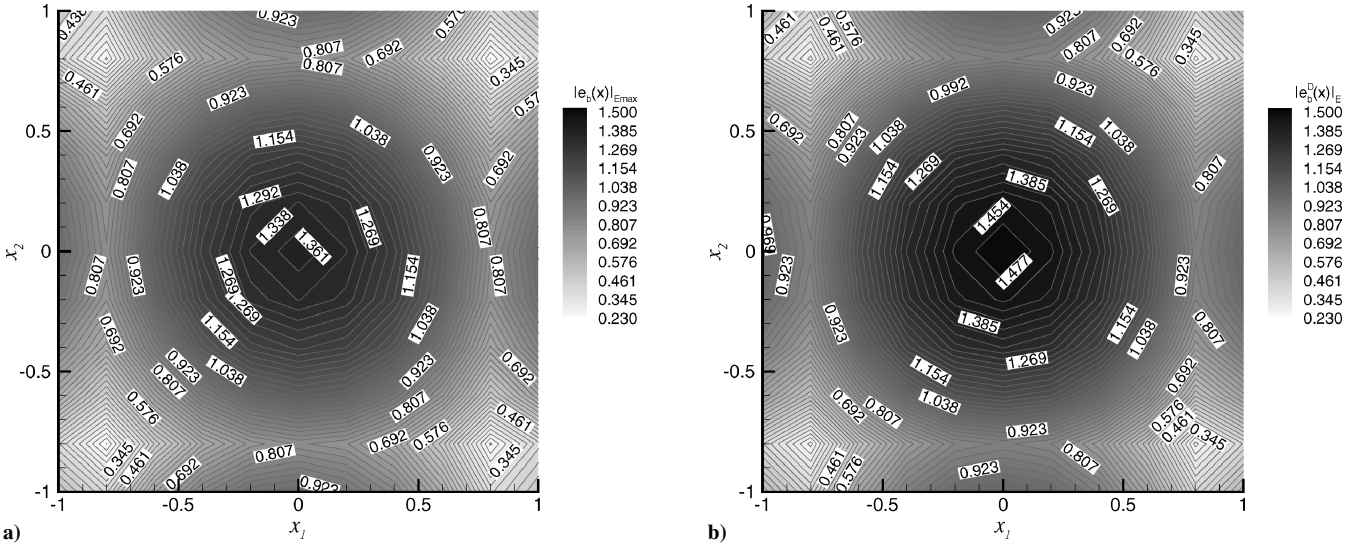

Fig. 5 Comparison of absolute worst-case errors for data set E in example 1: a) worst of 100 polynomials and b) data-dependent bound $|e_b^D(x)|$.

coefficient of correlation increases for the worst of all 500 polynomials, as the data-matching constraint averages out. The contour plots for $|e_b(x)|_{\text{overall}}$ and the data-independent bound $|e_b^I(x)|$ are in good agreement, as shown in Fig. 6.

Example 2: Quadratic Fit for Two-Dimensional Cubic Polynomial Using the Min–Max Bias Design

The example of DOE at $\alpha_1 = \alpha_2 = \alpha = 0.949$, with only bias error, Eq. (20), results in the prediction error at the data points as a function of only $\beta_2^{(2)}$ and $\beta_3^{(2)}$, and the RSA coefficients via Eq. (6) will be

$$b_1 = \beta_1^{(1)}, \qquad b_2 = \beta_2^{(1)} + \alpha^2\left[\beta_1^{(2)} + \tfrac{2}{3}\beta_3^{(2)}\right]$$

$$b_3 = \beta_3^{(1)} + \alpha^2\left[\tfrac{2}{3}\beta_2^{(2)} + \beta_4^{(2)}\right]$$

$$b_4 = \beta_4^{(1)}, \qquad b_5 = \beta_5^{(1)}, \qquad b_6 = \beta_6^{(1)} \tag{26}$$

Similar to example 1, randomly selecting the data is equivalent to random selection of the coefficients $b_1, b_2, b_3, b_4, b_5, b_6, \beta_2^{(2)}$, and $\beta_3^{(2)}$. The details of the selection of data sets and associated random polynomials and the bounds $\boldsymbol{c}^{(2)}$ on the coefficients in this example are also presented in Appendix B.

The correlations between the data-dependent error bound $|e_b^D(x)|$, Eq. (21), and the worst actual error distribution $|e_b(x)|_{k\max}$, Eq. (23), on a 11×11 grid are given in Table 4. The data-dependent error bound and actual worst error correlated very well. Figures 7 and 8 present the contour plots for data sets B and D, respectively, presenting the actual worst errors (over associated 100 polynomials each) and the data-dependent error bound. They also suggest that the data-dependent error bound successfully determines the worst-case error variation that can be considered as map of potential high bias-error locations.

The maximum of the errors among all 500 polynomials (100 for each data set A–E) can again be characterized by the

Table 4 Quadratic-cubic example: coefficients of correlation between error bounds and the actual worst errors calculated on a 11 × 11 grid

Bias error bound	$\|e_b(x)\|_{A\max}$	$\|e_b(x)\|_{B\max}$	$\|e_b(x)\|_{C\max}$	$\|e_b(x)\|_{D\max}$	$\|e_b(x)\|_{E\max}$	$\|e_b(x)\|_{\text{overall}}$
$\|e_b^D(x)\|$	0.9786	0.9639	0.9869	0.9932	0.9826	N/A
$\|e_b^I(x)\|$	0.7152	0.6743	0.7258	0.6403	0.7289	0.9701

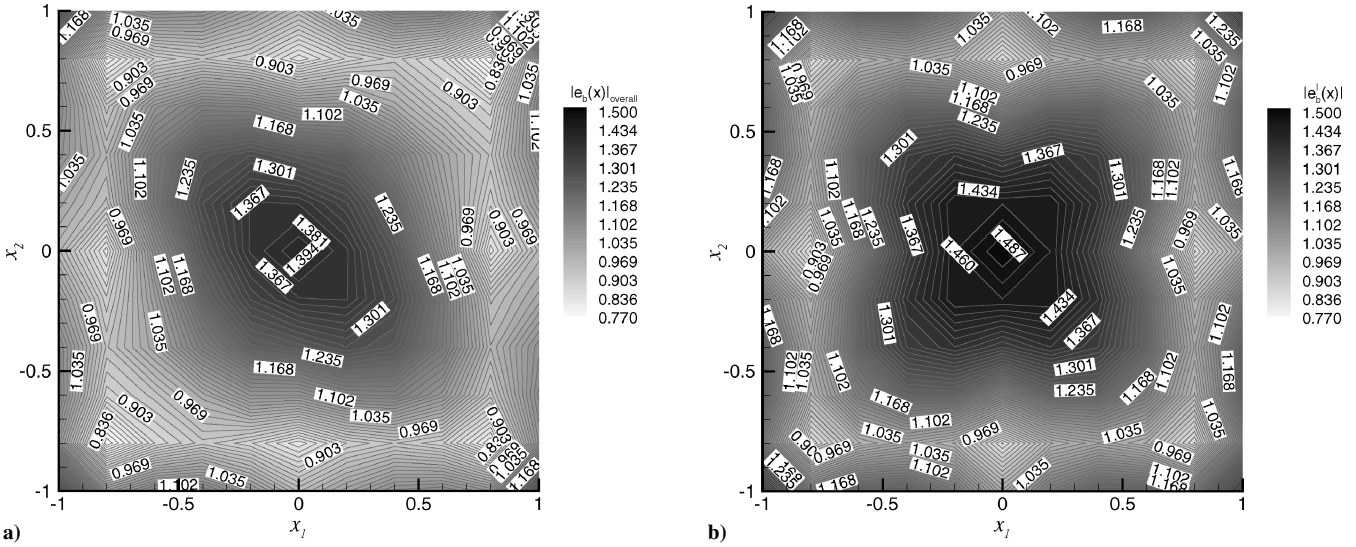

Fig. 6 Comparison of absolute worst-case errors for polynomials of data sets A–E in example 1: a) worst of 500 polynomials and b) data-independent bound $|e_b^I(x)|$.

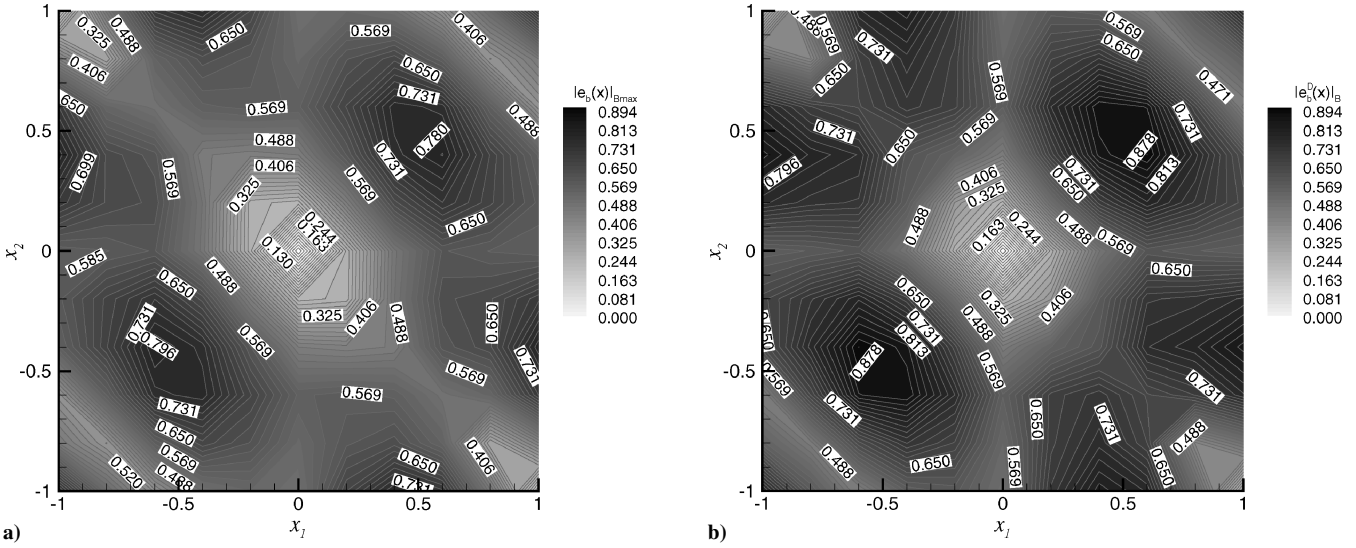

Fig. 7 Comparison of absolute worst-case errors for data set B in example 2: a) worst of 100 polynomials and b) data-dependent bound $|e_b^D(x)|$.

data-independent error bound $|e_b^I(x)|$ for the min–max bias experimental design [from Eq. (11) as shown also in Fig. 3c]. Table 4 also presents the correlations for the data-independent error bound. Lower correlations with the worst actual errors $|e_b(x)|_{k\max}$ for the individual data sets demonstrate the benefit of taking data into account. As expected, correlation of the data-independent error bound with the actual worst errors increases when polynomials for different data are included. It is about 0.97 for the worst of 500 polynomials $|e_b(x)|_{\text{overall}}$, Eq. (24). The contour plots for $|e_b(x)|_{\text{overall}}$ and the data-independent bound $|e_b^I(x)|$ are shown in Fig. 9.

The benefit of data available to tighten the bounds can be observed by comparing Figs. 7b, 8b, and 9b. The data-independent error bounds in Fig. 9b suggest, for instance, design points at $(-0.5, -0.5)$, $(-0.5, +0.5)$, $(0.5, -0.5)$, and (0.5,0.5) can experience high errors. This seems to be also true for the data set D as in Fig. 8b. Figure 7b for data set B, however, shows the bounds at points $(-0.5, +0.5)$ and $(0.5, -0.5)$ are tightened. That is, much less danger is predicted, as also observed in 100 polynomials of data B.

Note that the consistency in selection of the bounds for the uniform distributions and the bounds for coefficients used in optimization provides a quantitative assessment by the error bounds in the present polynomial examples (see Appendix B). Their use, however, should be considered qualitative, in general.

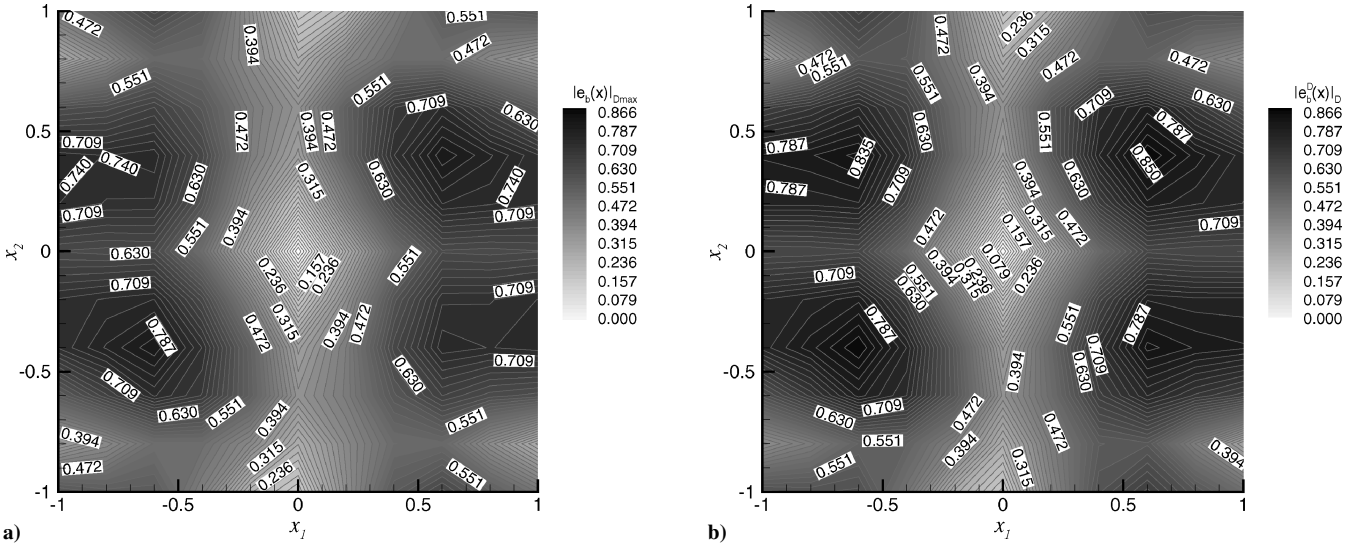

Fig. 8 Comparison of absolute worst-case errors for data set D in example 2: a) worst of 100 polynomials and b) data-dependent bound $|e_b^D(x)|$.

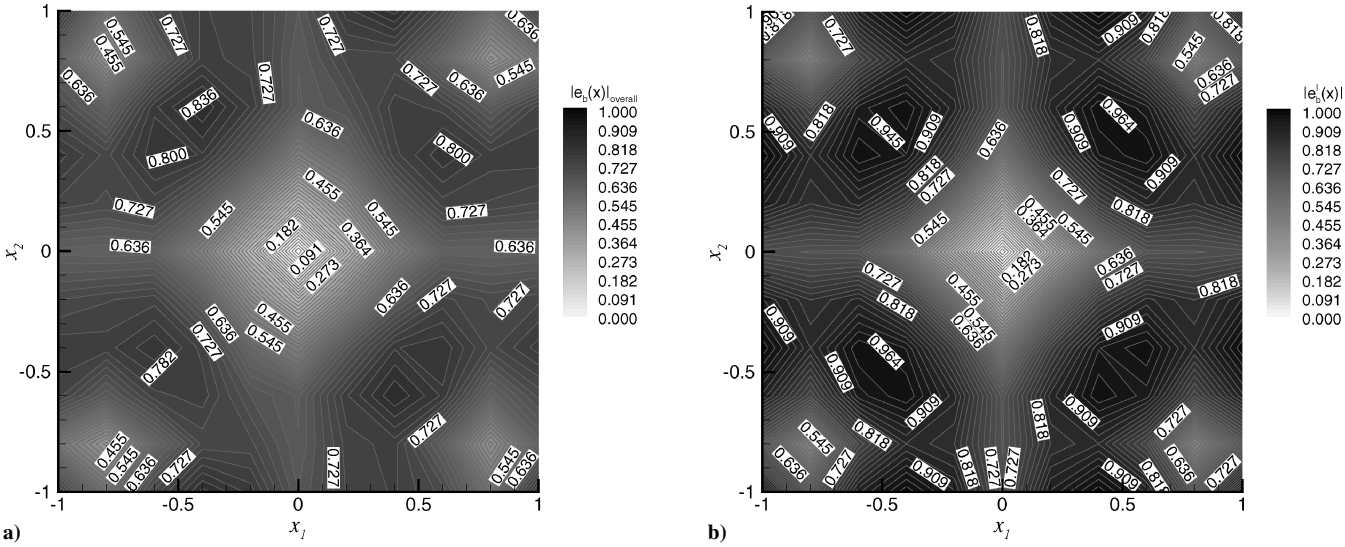

Fig. 9 Comparison of absolute worst-case errors for polynomials of data sets A–E in example 2: a) worst of 500 polynomials and b) data-independent bound $|e_b^I(x)|$.

Conclusions

Two error bounds were developed to characterize the point-to-point variation of error caused by an incorrect data model (bias error). A data-independent error bound is computed prior to data generation and depends only on the design of experiments, the true and least-square fit models. It seeks the worst-case absolute error, which varies point to point. The data-independent bound can be used to obtain experimental designs that minimize the maximum bias error. The min–max bias experimental design provided a good compromise between traditional minimum noise variance and minimum-bias designs as demonstrated in the examples (Tables 1 and 2).

The data-dependent bound also considers the worst-case absolute errors, but it is tightened by the available data. That is, the essential difference between the two error bounds is whether the coefficients of the permissible true functions are bounded without and with the data constraint. Comparisons of the actual worst errors and the error bounds for randomly selected polynomial examples showed that the data-dependent error bound successfully determines the map of the worst-case prediction errors in the design space caused by the available approximation. This information can be used to adaptively refine a design of experiments, or to steer clear of the identified potential high bias-error regions in the design domain.

The examples here are limited to polynomial true functions. The tools proposed, however, are still relevant. Even if the true function is not a polynomial, it is very well approximated locally by a cubic function, for instance, that is a better approximation than a quadratic to the true function. Therefore, these bias-error bounds, locally in particular, can be a valuable source of information about the true bias, even when the correct form of the true function is unknown.

Several suggestions associated with the use of the error bounds follow:

1) If several design of experiments are considered, make a choice among the candidates by using data-independent error bound and normalized estimated standard error distributions that do not need the data.

2) Before data generation, consider the data-independent bounds for a min–max experimental design to start with.

3) For the selected experimental design, evaluate the data-dependent error bounds.

4) If no resources are available for additional data generation and available RSA need to be used, consider the data-dependent error bounds to steer clear of the identified potential high bias-error regions.

5) If possible, refine the design of experiment by adding the new points where the error bounds are high.

Appendix A: Response Surface Methodology

A scalar function $\eta(x)$, $x \in E^n$ (E^n is real n-dimensional Euclidean space) is observed at N points $x^{(i)}$, $i = 1, \ldots, N$, and the observations y_i are contaminated by random variations ε_i such as measurement error, and so

$$y_i = \eta_i + \varepsilon_i \tag{A1}$$

The random errors ε at different points $x \in E^n$ are assumed to be uncorrelated and normally distributed random variables with zero mean and variance σ^2 that is the same at all points. The true mean response is assumed to be given in terms of basis functions $f_j(x)$ and associated coefficients β_j as

$$\eta(x) = \sum_{j=1}^{n_1} \beta_j f_j(x) = F(x)^T \beta \tag{A2}$$

where $F(x)^T = [f_1(x), f_2(x), \ldots, f_{n_1}(x)]$ and $\beta^T = (\beta_1, \beta_2, \ldots, \beta_{n_1})$. The n_1 basis functions f_j are often monomials. The least-squares approximation based on $N > n_1$ observations is

$$\hat{y}(x) = \sum_{j=1}^{n_1} b_j f_j(x) = F(x)^T b \tag{A3}$$

where the expected value of b_j (vector $\boldsymbol{b}$) is the β_j (vector $\boldsymbol{\beta}$). The difference (residual) between the data y and the estimate $\hat{y}(x)$ defined in Eq. (A3) at point $x^{(i)}$ is

$$e_i = y_i - \hat{y}_i \tag{A4}$$

For N observations y_i, $i = 1, \ldots, N$, the residuals can be written in matrix form:

$$e = y - X\boldsymbol{b} \tag{A5}$$

where $e^T = (e_1, \ldots, e_N)$, and X is the Gramian matrix with terms $X_{i,j} = f_j[x^{(i)}]$. For instance, for a quadratic model in two variables x_1, x_2 with N data points $[x_1^{(1)}, x_2^{(1)}], \ldots, [x_1^{(N)}, x_2^{(N)}]$,

$$X = \begin{bmatrix} 1 & x_1^{(1)} & x_2^{(1)} & \left(x_1^{(1)}\right)^2 & x_1^{(1)}x_2^{(1)} & \left(x_2^{(1)}\right)^2 \\ 1 & x_1^{(2)} & x_2^{(2)} & \left(x_1^{(2)}\right)^2 & x_1^{(2)}x_2^{(2)} & \left(x_2^{(1)}\right)^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_1^{(i)} & x_2^{(i)} & \left(x_1^{(i)}\right)^2 & x_1^{(i)}x_2^{(i)} & \left(x_2^{(i)}\right)^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_1^{(N)} & x_2^{(N)} & \left(x_1^{(N)}\right)^2 & x_1^{(N)}x_2^{(N)} & \left(x_2^{(N)}\right)^2 \end{bmatrix} \tag{A6}$$

The coefficient vector $\boldsymbol{b}$ in Eq. (A5) that minimizes $\|e\|_2^2$ is (assuming X has full rank)

$$b = (X^T X)^{-1} X^T y \tag{A7}$$

When the fitting model Eqs. (A1) and (A2) is valid (no bias error), an unbiased estimator s^2 (error mean square) for variance σ^2 of the noise is given as

$$s^2 = MS_E = \frac{y^T y - b^T X^T y}{N - n_1} \tag{A8}$$

The positive square root of the prediction variance is usually used as an estimate of the prediction error at a design point x, also called estimated standard error, and is defined by Myers and Montgomery[10]:

$$e_{es}(x) = s\sqrt{F(x)^T (X^T X)^{-1} F(x)} \tag{A9}$$

where the estimator s [Eq. (A8)] is often called in the engineering literature the root-mean-square-error predictor. The estimated standard error depends on the location of the design point. Furthermore, as Eq. (A2) is, in general, only an assumption about the true function $\eta(x)$, s will contain not only noise error, but also modeling (bias) error.

Appendix B: Selection of Random Polynomials and Coefficient Bounds for the Examples

In example 1, selecting the data randomly is equivalent to random selection of coefficients b_1, b_2, b_3, and $\beta_2^{(2)}$. Consider a uniform distribution $U(-1, +1)$ for coefficients b_2, b_3, and $\beta^{(2)}$. For the intercept or constant coefficient b_1, the $U(5, 10)$ distribution is selected to obtain reasonable relative errors at the data points. The coefficients and associated distributions corresponding to the linear-quadratic example are summarized in Table B1. The randomly selected coefficients for data sets A to E are presented in Table B2. It is seen that the maximum absolute error at the data points $|e^{(i)}|_{\max}$ is equal to the absolute value of $\alpha_1^2 \beta_2^{(2)}$ ($\alpha_1 = 0.866$). The coefficients $\beta^{(2)}$ are bounded conservatively at ± 1 [$c^{(2)} = 1$] by considering $|e^{(i)}|_{\max}$ from Table B2.

In example 2, randomly selecting the data is equivalent to random selection of the coefficients $b_1, b_2, b_3, b_4, b_5, b_6, \beta_2^{(2)}$, and $\beta_3^{(2)}$. Uniform distributions used in this example are also given in Table B1. The data sets A to E characterized by b, $\beta_2^{(2)}$, and $\beta_3^{(2)}$ are presented in Table B3. The maximum absolute errors at the data points $|e^{(i)}|_{\max}$ are also shown in the last column of the table. Then selecting $\beta_1^{(2)}$ and $\beta_4^{(2)}$ randomly provides 100 polynomials for each data set satisfying the data. The coefficients $\beta^{(2)}$ are bounded conservatively at ± 1 [$c^{(2)} = 1$] based on $|e^{(i)}|_{\max}$ from Table B3.

Table B1 Description of uniformly distributed random coefficients for the example problems

Uniform distribution	Examples	
	Linear-quadratic	Quadratic-cubic
$U(5, 10)$	b_1	b_1
$U(-1, +1)$	b_2, b_3	b_2, b_3, b_4, b_5, b_6
$U(-1, +1)$	$\beta_1^{(2)}, \beta_2^{(2)}, \beta_3^{(2)}$	$\beta_1^{(2)}, \beta_2^{(2)}, \beta_3^{(2)}, \beta_4^{(2)}$

Table B2 Linear-quadratic example: randomly selected coefficients for the data sets and the maximum absolute error at the data points

Data set	b_1	b_2	b_3	$\beta_2^{(2)}$	$\lvert e^{(i)}\rvert_{\max}$
A	5.11	−0.40	0.79	0.37	0.28
B	8.57	0.53	0.89	0.94	0.71
C	5.10	−0.90	0.80	−0.84	0.63
D	9.88	−0.93	−0.84	−0.66	0.50
E	6.80	0.88	−0.08	−0.07	0.05

Table B3 Quadratic-cubic example: randomly selected coefficients for the data sets and the maximum absolute error at the data points

Data set	b_1	b_2	b_3	b_4	b_5	b_6	$\beta_2^{(2)}$	$\beta_3^{(2)}$	$\lvert e^{(i)}\rvert_{\max}$
A	7.99	−0.45	−0.62	0.90	0.03	0.61	−0.96	0.64	0.54
B	7.72	0.64	−0.05	−0.78	−0.01	0.69	0.68	0.79	0.45
C	8.34	0.59	−0.81	0.62	0.38	0.67	−0.48	0.81	0.46
D	8.67	−0.83	−0.40	0.47	0.50	−0.22	0.13	0.88	0.50
E	6.78	−0.07	0.66	−0.18	−0.47	−0.01	−0.33	0.51	0.29

Acknowledgments

This research was supported by NASA Grant NAG1–02042, Air Force Office of Scientific Research Grant F49620-02-1-0090, and National Science Foundation Grant DMI-0423280.

References

[1]Kaufman, M., Balabanov, V., Burgee, S. L., Giunta, A. A., Grossman, B., Haftka, R. T., Mason, W. H., and Watson, L. T., "Variable-Complexity Response Surface Approximations for Wing Structural Weight in HSCT Design," *Computational Mechanics*, Vol. 18, No. 2, 1996, pp. 112–126.

[2]Balabanov, V. O., Giunta, A. A., Golovidov, O., Grossman, B., Mason, W. H., Watson, L. T., and Haftka, R. T., "Reasonable Design Space Approach to Response Surface Approximation," *Journal of Aircraft*, Vol. 36, No. 1, 1999, pp. 308–315.

[3]Balabanov, V., Kaufman, M., Knill, D. L., Haim, D., Golovidov, O., Giunta, A. A., Haftka, R. T., Grossman, G., Mason, W. H., and Watson, L. T., "Dependence of Optimal Structural Weight on Aerodynamic Shape for a High Speed Civil Transport," *Proceedings of the 6th AIAA/NASA/USAF Symposium on Multidisciplinary Analysis and Optimization*, Pt. 1, AIAA, Reston, VA, 1996, pp. 599–612.

[4]Papila, M., and Haftka, R. T., "Uncertainty and Wing Structural Weight Approximations," *Proceedings of the 40th AIAA/ASME/ASCE/ASC Structures, Structural Dynamics, and Material Conference*, Vol. 2, AIAA, Reston, VA, 1999, pp. 988–1002.

[5]Papila, M., and Haftka, R. T., "Response Surface Approximations: Noise, Error Repair and Modeling Errors," *AIAA Journal*, Vol. 38, No. 12, 2000, pp. 2336–2343.

[6]Shyy, W., Tucker, P. K., and Vaidyanathan, R., "Response Surface and Neural Network Techniques for Rocket Engine Injector Optimization," *Journal of Propulsion and Power*, Vol. 17, No. 2, 2001, pp. 391–401.

[7]Papila, N., Shyy, W., Griffin, L., and Dorney, D. J., "Shape Optimization of Supersonic Turbines Using Global Approximation Methods," *Journal of Propulsion and Power*, Vol. 18, No. 3, 2002, pp. 508–518.

[8]Redhe, M., Forsberg, J., Jansson, T., Marklund, P. O., and Nilsson, L., "Using the Response Surface Methodology and the D-Optimality Criterion in Crashworthiness Related Problems—An Analysis of the Surface Approximation Error Versus the Number of Function Evaluations," *Structural and Multidisciplinary Optimization*, Vol. 24, No. 3, 2002, pp. 185–194.

[9]Simpson, T. W., Peplinski, J. D., Koch, P. N., and Allen, J. K., "Metamodels for Computer-Based Engineering Design: Survey and Recommendations," *Engineering with Computers*, Vol. 17, No. 2, 2001, pp. 129–150.

[10]Myers, R. H., and Montgomery, D. C., *Response Surface Methodology—Process and Product Optimization Using Designed Experiments*, Wiley, New York, 1995, pp. 208–279.

[11]Khuri, A. I., and Cornell, J. A., *Response Surfaces: Designs and Analyses*, 2nd ed., Marcel Dekker, New York, 1996, pp. 207–247.

[12]Box, G. E .P., and Draper, N. R., "The Choice of a Second Order Rotatable Design," *Biometrika*, Vol. 50, No. 3, 1963, pp. 335–352.

[13]Draper, N. R., and Lawrence, W. E., "Designs Which Minimize Model Inadequacies: Cuboidal Regions of Interest," *Biometrika*, Vol. 52, No. 1–2, 1965, pp. 111–118.

[14]Kupper, L. L., and Meydrech, E. F., "A New Approach to Mean Squared Error Estimation of Response Surfaces," *Biometrika*, Vol. 60, No. 3, 1973, pp. 573–579.

[15]Welch, W. J., "A Mean Squared Error Criterion for the Design of Experiments," *Biometrika*, Vol. 70, No. 1, 1983, pp. 205–213.

[16]Venter, G., and Haftka, R. T., "Minimum-Bias Based Experimental Design for Constructing Response Surfaces in Structural O:ptimization," *Proceedings of the 38th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Pt. 2, AIAA, Reston, VA, 1997, pp. 1225–1238.

[17]Montepiedra, G., and Fedorov, V. V., "Minimum Bias Designs with Constraints," *Journal of Statistical Planning and Inference*, Vol. 63, No. 1, 1997, pp. 97–111.

[18]Fedorov, V. V., Montepiedra, G., and Nachtsheim, C. J., "Design of Experiments for Locally Weighted Regression," *Journal of Statistical Planning and Inference*, Vol. 81, No. 2, 1999, pp. 363–383.

[19]Palmer, K., and Tsui, K. L., "A Minimum Bias Latin Hypercube Design," *Institute of Industrial Engineers Transactions*, Vol. 33, No. 9, 2001, p. 793.

[20]Qu, X., Venter, G., and Haftka, R. T., "New Formulation of Minimum-Bias Central Composite Experimental Design and Gauss Quadrature," *Structural and Multidisciplinary Optimization*, Vol. 28, No. 4, 2004, pp. 231–242.

[21]Papila, M., and Haftka, R. T., "Uncertainty and Response Surface Approximations," AIAA Paper 2001-1680, April 2001.

[22]Papila, M., Papila, N., Shyy, W., Haftka, R. T., and Fitz-coy, N., "Error-Based Design Space Windowing," AIAA Paper 2002-0539, Jan. 2002.

A. Messac
Associate Editor