AIAA JOURNAL
Vol. 48, No. 10, October 2010

Using Cross Validation to Design Conservative Surrogates

Felipe A. C. Viana,* Victor Picheny,† and Raphael T. Haftka‡
University of Florida, Gainesville, Florida 32611

DOI: 10.2514/1.J050327

The use of surrogates (also known as metamodels) for facilitating optimization and statistical analysis of computationally expensive simulations has become commonplace. Surrogate models are usually fit to be unbiased (i.e., the error expectation is zero). However, in certain applications, it might be important to safely estimate the response (e.g., in structural analysis, the maximum stress must not be underestimated in order to avoid failure). In this work we use safety margins to conservatively compensate for fitting errors associated with surrogates. We propose the use of cross validation for estimating the required safety margin for a desired level of conservativeness (percentage of safe predictions). The approach was tested on three algebraic examples for two basic surrogates: namely, kriging and polynomial response surface. For these examples we found that cross validation is effective for selecting the safety margin. We also applied the approach to the probabilistic design optimization of a composite laminate. This design under uncertainty example showed that the approach can be successfully used in engineering applications.

Nomenclature

CI_S	=	standard estimated confidence interval
CI_W	=	Wilson's estimated confidence interval
$e(\mathbf{x})$	=	prediction error ($e(\mathbf{x}) = \hat{y}(\mathbf{x}) - y(\mathbf{x})$)
e_{RMS}	=	root mean square error
$\mathbf{e}_{XV}$	=	vector of cross-validation errors
PRESS	=	prediction sum of squares
REG	=	relative error growth (loss in accuracy)
REG_{XV}	=	relative error growth (estimated loss in accuracy) based on cross-validation data
s	=	safety margin
$\hat{s}$	=	estimated safety margin
$y(\mathbf{x})$	=	actual function
$\hat{y}(\mathbf{x})$	=	surrogate model of the actual function
$\hat{y}_C(\mathbf{x})$	=	conservative surrogate model of the actual function
$\%c$	=	percentage of conservative (i.e., positive) errors

I. Introduction

SURROGATE-BASED analysis and optimization has become a major tool in engineering design of complex systems [1–6]. Usually, surrogate models are fit to be unbiased, so that predictions are equally likely to be below and above the actual value of the response (i.e., the error expectation is zero). However, in many applications, we might be interested in approximations that safely predict the actual response [7–10] (while expected to maintain accuracy). In constrained optimization (constraints being surrogate models) or in reliability-based design optimization (limit state composed by surrogate models), it can happen that after running the optimization the solution turns out to be infeasible due to surrogate errors [11,12]. There are different ways to address this issue: for example, 1) using conservative constraints [13–15] so that the optimization is pushed to the feasible region and 2) making the limit state more accurate near the boundary between of the feasible domain (usually through iterative update of the surrogate model [16–18]).

In this paper, we focus on conservative surrogates. Here, when estimates are higher than the true response we call them conservative. Hence, conservative estimations tend to overestimate the actual response, and as a consequence, there is a tradeoff between accuracy and conservativeness. One of the most widely used methods for conservative estimation is to bias the prediction response by additive or multiplicative constants (termed safety margin and safety factors, respectively). This practice is common in structural analysis, where stress or strain values must not be underestimated in order to avoid failure. The choice of the constant is often based on previous knowledge of the problem [19–21]. However, for surrogate-based analysis, there is no established practice for choosing the safety margin. Since conservative estimators tend to overestimate the true values, the design of the safety margin can be considered as a two-objective optimization problem. The results can be presented in the form of Pareto fronts: accuracy vs conservativeness.

One way of improving the conservativeness of a surrogate is to require it to conservatively fit the data points [13,14]. However, this approach does not allow tuning in the level of desired conservativeness. Kim and Choi [22] used prediction interval given by the weighted least-squares surrogate model as a safeguard against surrogate error. Several ways of dealing with surrogate uncertainty can be found in the literature. In their previous work [15,23], the authors explored and compared different alternatives to produce conservative predictions with surrogates. They found that safety margins and estimators based on the surrogate uncertainty distribution were comparable in performance, but the safety margin approach lacked a basis for selecting the magnitude of the margin.

In this paper we propose the use of cross validation for estimating the safety margin. For unbiased surrogates, there has been research pointing to the utility of cross validation [24–29] for selecting the most accurate surrogate from several fitted surrogates. We explore if cross validation allows selecting the safety margin for a given conservativeness level. Compared to other alternatives, the combined used of safety margin and cross validation offer two potential advantages:

1) The procedure does not depend on the statistical assumptions of any surrogate technique.

2) It does not depend on quality of the surrogate uncertainty model (very dependent on the problem).

We tested our approach using kriging [30,31] and polynomial response surface [32,33] surrogates.

The rest of the paper is organized as follows. Section II reviews the basis of conservative prediction using safety margins. Section III introduces the proposed approach for designing the safety margins

Received 13 November 2009; revision received 10 April 2010; accepted for publication 17 May 2010.

*Research Assistant, Department of Mechanical and Aerospace Engineering; fchegury@ufl.edu.

†Research Assistant, Department of Mechanical and Aerospace Engineering; vpicheny@ufl.edu.

‡Distinguished Professor, Department of Mechanical and Aerospace Engineering; haftka@ufl.edu.

based on cross validation. Section IV defines the numerical experiments and presents results and discussion. Section V applies the proposed methods to the design of a composite laminate. The paper is closed with Sec. VI by recapitulating salient points and concluding remarks. There are two appendices. In the first one we check whether we minimize the losses in accuracy associated with a conservative predictor by selecting between alternate surrogates using cross validation. The second one describes box plots (a visualization tool used in our statistical analysis).

II. Background

A. Conservative Surrogates

We denote by y the response of a numerical simulator or function that is to be studied:

$$y\colon D \subset \mathbb{R}^d \to \mathbb{R} \qquad \mathbf{x} \mapsto y(\mathbf{x}) \tag{1}$$

where $\mathbf{x} = \{x_1, \ldots, x_d\}^T$ is a d-dimensional vector of input variables.

When the response $y(\mathbf{x})$ is expensive to evaluate, we approximate it by a less expensive model $\hat{y}(\mathbf{x})$ (surrogate model), based on 1) assumptions on the nature of $y(\mathbf{x})$ and 2) on the observed values of $y(\mathbf{x})$ at a set of points, called the experimental design [34,35] (also known as design of experiment). Figure 1a shows a kriging model fitted to seven equally spaced data points of the $y(x) = (6x-2)^2 \sin(12x-4)$ function. In this example, the prediction errors are conservative in half of the design space, as illustrated by Fig. 1b.

In this paper, a conservative surrogate $\hat{y}_C(\mathbf{x})$ obtained by adding a safety margin s to an unbiased surrogate model $\hat{y}(\mathbf{x})$ is an empirical estimator of the type

$$\hat{y}_C(\mathbf{x}) = \hat{y}(\mathbf{x}) + s \tag{2}$$

When we check the accuracy of a conservative surrogate, we calculate the root mean square error:

$$e_{\mathrm{RMS}} = \sqrt{\frac{1}{V}\int_D e_C^2(\mathbf{x})\,\mathrm{d}\mathbf{x}} = \sqrt{\frac{1}{V}\int_D (\hat{y}_C(x) - y(x))^2\,\mathrm{d}\mathbf{x}} \tag{3}$$

where V is the volume of the design domain D.

The e_{RMS} is computed by Monte Carlo integration at a large number of p_{test} test points:

$$e_{\mathrm{RMS}} = \sqrt{\frac{1}{p_{\mathrm{test}}}\sum_{i=1}^{p_{\mathrm{test}}} e_{Ci}^2} \tag{4}$$

$$e_{Ci} = \hat{y}_{Ci} - y_i = \hat{y}_i - y_i + s \tag{5}$$

where $\hat{y}_{Ci}$ and y_i are values of the conservative prediction and actual simulation at the ith test point, respectively.

B. Conservativeness Level and Relative Error Growth

Although there are different measures of the conservativeness of an approximation (e.g., the average error or the maximum nonconservative error); for convenience, we use the percentage of conservative (i.e., positive) errors. With the actual prediction errors $e(\mathbf{x}) = \hat{y}(\mathbf{x}) - y(\mathbf{x})$ (at any point of the design space), we can use the cumulative distribution function of the prediction error $F_E(e)$ to find the fraction of the errors that are conservative. In fact, $F_E(e)$ gives the proportion of the errors that lies bellow any value. Therefore, the percentage of conservative errors $\%c$ is obtained by

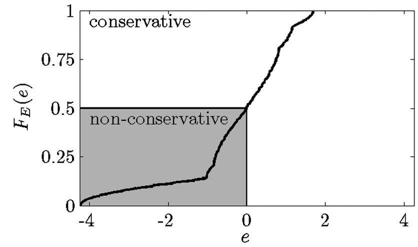


Fig. 2 Cumulative distribution function of the prediction error $F_E(e)$ of the surrogate shown in Fig. 1. In this example, 50% of the errors are conservative (positive).

$$\%c = 100 \times (1 - F_E(0)) \tag{6}$$

As illustrated in Fig. 2, with $F_E(e)$ we can easily obtain the conservativeness level of the surrogate. It might be worth noting that the distribution of the errors is not always symmetric with respect to $e = 0$. Both Figs. 1 and 2 show that although 50% of the errors are conservative, the nonconservative errors can have greater magnitude than the conservative ones.

The safety margin is selected to turn a desired percentage of the errors conservative. The safety margin s for a given conservativeness $\%c$ can be expressed in terms of $F_E(e)$ as

$$s = -F_E^{-1}\left(1 - \frac{\%c}{100}\right) \tag{7}$$

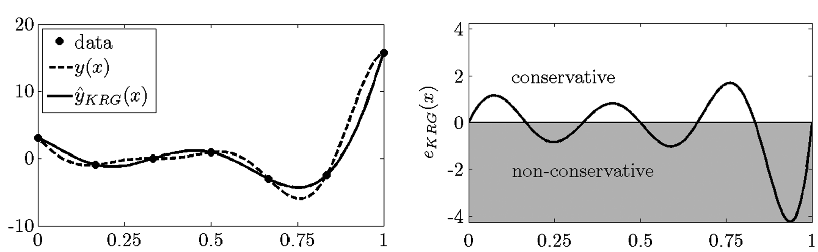
Figure 3 illustrates how we apply Eq. (7) to obtain the safety margin s that would lead to a conservativeness of $\%c = 90\%$ of the surrogate shown in Fig. 1.

As stated before, conservative estimators tend to overestimate the true values. As a consequence, the accuracy of the surrogate is degraded. We define the relative error growth (loss in accuracy) in terms of the e_{RMS} as

$$\mathrm{REG} = \frac{e_{\mathrm{RMS}}}{e_{\mathrm{RMS}}^*} - 1 \tag{8}$$

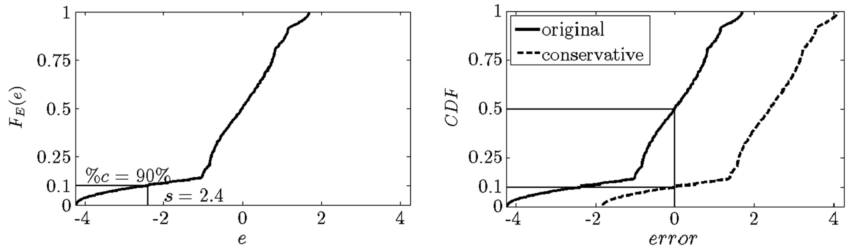
where e_{RMS} is taken at a given target conservativeness; and e_{RMS}^* is the e_{RMS} value of the surrogate without adding any safety margin.

Figure 4 illustrates the effect of using a safety margin to achieve $\%c = 90\%$ when fitting the data used in Fig. 1. Figure 4a shows the original kriging model and the conservative counterpart. Adding the safety margin $s = 2.4$ makes the e_{RMS} to increase from $e_{\mathrm{RMS}} = 1.4$ to $e_{\mathrm{RMS}} = 2.6$. That means that the error increased by approximately 86%. Figure 4b illustrates the shifting in the prediction errors. After adding the safety margin, 90% of the domain ($0 \le x \le 1$) becomes conservative.



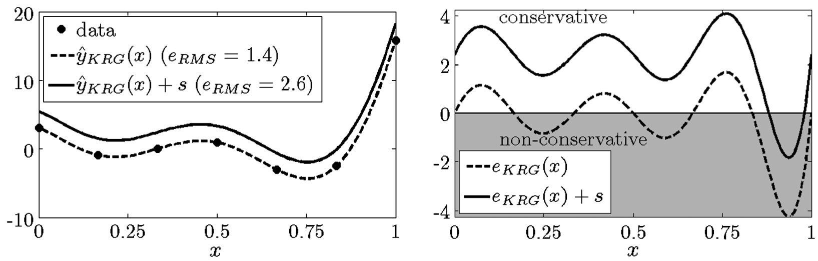
a) Kriging model **b) Prediction error of the kriging model**

Fig. 1 Illustration of surrogate modeling. Figure 1a shows a kriging model (KRG) for $y(x) = (6x-2)^2 \sin(12x-4)$ fitted with seven data points. Figure 1b shows the prediction error associated with this surrogate.



a) Finding the safety margins **b) Prediction error of the unbiased surrogate**

Fig. 3 Design of safety margin using $F_E(e)$. Figure 3a shows that $\%c = 90\%$ would be achieved with a safety margin of $s = 2.4$. Figure 3a illustrates that 90% of the errors are greater than zero for the conservative surrogate.



a) Effect of safety margin on prediction **b) Effect of safety margin on prediction error**

Fig. 4 Illustration of conservative kriging model ($\%c = 90\%$) for the $y(x) = (6x - 2)^2 \sin(12x - 4)$ function fitted with seven data points. Conservativeness comes at the price of losing accuracy.

C. Cross Validation

Cross validation is often used for both assessing accuracy and surrogate selection [24–29]. It is attractive because it does not depend on the statistical assumptions of a particular surrogate technique and it does not require extra expensive simulations (test points). Cross validation is a process of estimating errors by constructing the surrogate without one of the p points and calculating the error at the left out point (leave-one-out strategy). After repeating the process with all other $p - 1$ points, we obtain the vector of cross-validation errors $\mathbf{e}_{XV}$ (sometimes called the PRESS vector, where PRESS stands for prediction sum of squares). Figure 5 illustrates how we obtain the cross-validation errors for a kriging surrogate. With $\mathbf{e}_{XV}$, the e_{RMS} is estimated by

$$\mathrm{PRESS}_{\mathrm{RMS}} = \sqrt{\frac{1}{p}\mathbf{e}_{XV}^T \mathbf{e}_{XV}} \tag{9}$$

The leave-one-out strategy is computationally expensive for large number of points. We then use a variation of the k-fold strategy [24] to overcome this problem.§ According to the classical k-fold strategy, after dividing the available data (p points) into p/k clusters, each fold is constructed using a point randomly selected (without replacement) from each of the clusters. Of the folds, one is retained as the validation data for testing the model and the remaining are used as training data. The cross-validation process is then repeated with each fold used exactly once as validation data. Note that k-fold turns out to be the leave-one-out when $k = p$.¶

When we add a safety margin to a predictor, we do not need to repeat the costly process of cross validation to assess the new PRESS vector, because the vector of cross-validation errors associated with $\hat{y}_C(\mathbf{x})$, $\mathbf{e}_{XVC}$, is simply

$$\mathbf{e}_{XVC} = \mathbf{e}_{XV} + s \tag{10}$$

With that, the computation of $\mathrm{PRESS}_{\mathrm{RMS}}$ employs Eq. (9) with $\mathbf{e}_{XVC}$ replacing $\mathbf{e}_{XV}$ (without the need of any extra surrogate fitting).

III. Design of Safety Margin Using Cross-Validation Errors

In practice, $F_E(e)$ is unknown, so the safety margin that ensures a given level of conservativeness cannot be determined exactly. We propose to design the safety margin using cross validation. That is, we replace $F_E(e)$ by the cumulative distribution function of the cross-validation errors $F_{E_{XV}}(e_{XV})$. This way, the estimated safety margin for a given target conservativeness $\%c$ is

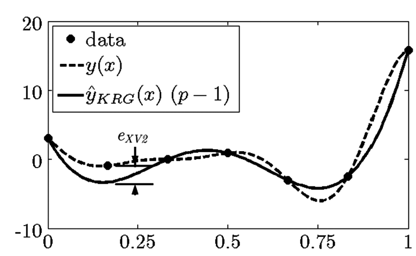


Fig. 5 Cross-validation error at the second point of the experimental design e_{XV_2} exemplified by fitting a kriging model to $p = 7$ data points of the $y(x) = (6x - 2)^2 \sin(12x - 4)$ function.

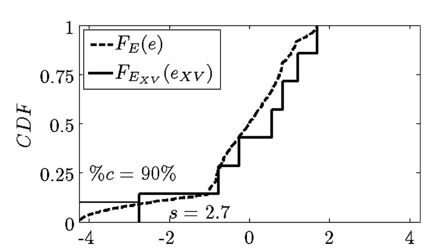


Fig. 6 Design of safety margin using $F_{E_{XV}}(e_{XV})$. Cross-validation errors would overestimate the safety margin suggesting $s = 2.7$ as opposed to $s = 2.4$ suggested by actual errors (see Fig. 3).

§Some surrogate techniques also offer other inexpensive strategies for computation of cross-validation errors. See [36–38] for further reading.

¶We implemented the strategy by 1) extracting p/k points of the set using a maxmin criterion (maximization of the minimum interdistance) and 2) removing these points from the set and repeating step 1 with the remaining points. Each set of extracted points is used for validation and the remaining for fitting.

$$\hat{s} = -F_{E_{XV}}^{-1}\left(1 - \frac{\%c}{100}\right) \tag{11}$$

Figure 6 illustrates how we would design the safety margin for $\%c = 90\%$ using cross validation in the example of Fig. 1. Using $F_{E_{XV}}(e_{XV})$, the safety margin for $\%c = 90\%$ is $s = 2.7$, which means that we would get an slightly more conservative surrogate (actual conservativeness would be $\%c = 91\%$).

Note that there are two error sources associated with using Eq. (11) for estimating s [as compared to the exact value defined by Eq. (7)]. The first is due to finite sampling and the second is due to the use of cross-validation errors instead of actual errors. We can estimate the uncertainty due to finite sampling by assuming that $I(e_{XVi})$ (which equal 1 if $e_{XVi} > 0$ and 0 otherwise) are independent and identically distributed random variables following a Bernoulli distribution with a probability of success of c. Although this is a strong assumption (since the prediction errors are spatially correlated), it might be acceptable if the points are far from each other (as in the case of the data points and cross-validation errors). With that, the standard interval estimation of the $(1-\alpha)$ confidence interval CI_S for the conservatives $\hat{c}$ is usually taken as [39]

$$\mathrm{CI}_S = \hat{c} \pm \kappa p^{-1/2}(\hat{c}\,\hat{q})^{1/2}, \qquad \hat{q} = 1 - \hat{c} \tag{12}$$

$$\kappa \equiv z_{\alpha/2} = \Phi^{-1}(1 - \alpha/2) \tag{13}$$

where p is the number of data points, $\Phi(z)$ is the standard normal distribution function, and α is some prespecified value between 0 and 1. CI_S is also known as the Wald interval.

However, as pointed out by the literature [40], Eq. (12) fails, especially for small p and high values of $\hat{c}$ (which are often those of interest). An example of a well-accepted alternative to the standard interval is the so-called Wilson interval [41]:

$$\mathrm{CI}_W = \frac{(\hat{c}p) + \kappa^2/2}{p + \kappa^2} \pm \frac{\kappa p^{1/2}}{p + \kappa^2}\left(\hat{c}\,\hat{q} + \frac{\kappa^2}{4p}\right)^{1/2} \tag{14}$$

In addition to designing the safety margin, we also use cross validation to estimate the relative error growth (i.e., to estimate of loss in accuracy) using $\mathrm{PRESS}_{\mathrm{RMS}}$ instead of e_{RMS}:

$$\mathrm{REG}_{XV} = \frac{\mathrm{PRESS}_{\mathrm{RMS}}}{\mathrm{PRESS}^*_{\mathrm{RMS}}} - 1 \tag{15}$$

When considering a single surrogate, $\mathrm{PRESS}^*_{\mathrm{RMS}}$ is the $\mathrm{PRESS}_{\mathrm{RMS}}$ value of the surrogate without adding any safety margin.

Although the focus of this paper is the selection of proper safety margin for a given conservativeness level, we also checked immediate benefits of multiple surrogates. Appendix A shows that by using cross validation, we can minimize the losses in accuracy by selecting between alternate surrogates.

IV. Analytical Examples

A. Basic Surrogates

Table 1 details the two surrogates used during the investigation. Because the strength of the method is that it is not tied to a particular surrogate, it is important to show that it works well for diverse surrogates. The DACE toolbox of Lophaven et al. [42], and the SURROGATES toolbox of Viana ** were used to execute the kriging and the polynomial response surface algorithms, respectively. The SURROGATES toolbox was also used for easy manipulation of the surrogates. There might be better implementations of each surrogate techniques readily available.††

Table 1 Information about the two generated surrogates

Surrogates	Details
prs	Polynomial response surface: full second-order model.
krg	Kriging model: set with zero-order polynomial regression model, respectively. In all cases, a Gaussian correlation and $\theta_{0i} = (p^{-1/n_v})$, and $10^{-3} \le \theta_i \le 2 \times \theta_{0i}$, $i = 1, 2, \ldots, n_v$ were used. We chose 3 different kriging surrogates by varying the regression model.

B. Test Problems

As test problems, we employed the two following widely used analytical benchmark problems [43].

Branin–Hoo (two variables):

$$y(\mathbf{x}) = \left(x_2 - \frac{5.1x_1^2}{4\pi^2} + \frac{5x_1}{\pi} - 6\right)^2 + \left(10 - \frac{10}{8\pi}\right)\cos(x_1) + 10$$
$$-5 \le x_1 \le 10, \qquad 0 \le x_2 \le 15 \tag{16}$$

Hartman (six variables):

$$y(\mathbf{x}) = -\sum_{i=1}^{4} a_i \exp\left(-\sum_{j=1}^{n_v} b_{ij}(x_j - d_{ij})^2\right), \qquad 0 \le x_j \le 1$$

$$j = 1, 2, \ldots, n_v, \qquad n_v = 6, \qquad \mathbf{a} = [1.0 \quad 1.2 \quad 3.0 \quad 3.2]$$

$$\mathbf{B} = \begin{bmatrix} 10.0 & 3.0 & 17.0 & 3.5 & 1.7 & 8.0 \\ 0.05 & 10.0 & 17.0 & 0.1 & 8.0 & 14.0 \\ 3.0 & 3.5 & 1.7 & 10.0 & 17.0 & 8.0 \\ 17.0 & 8.0 & 0.05 & 10.0 & 0.1 & 14.0 \end{bmatrix}$$

$$\mathbf{D} = \begin{bmatrix} 0.1312 & 0.1696 & 0.5569 & 0.0124 & 0.8283 & 0.5886 \\ 0.2329 & 0.4135 & 0.8307 & 0.3736 & 0.1004 & 0.9991 \\ 0.2348 & 0.1451 & 0.3522 & 0.2883 & 0.3047 & 0.6650 \\ 0.4047 & 0.8828 & 0.8732 & 0.5743 & 0.1091 & 0.0381 \end{bmatrix} \tag{17}$$

Rosenbrock (nine variables):

$$y(\mathbf{x}) = \sum_{i=1}^{n_v - 1}[(1 - x_i)^2 + 100(x_{i+1} - x_i^2)^2], \qquad -5 \le x_i \le 10$$
$$i = 1, 2, \ldots, \qquad n_v = 9 \tag{18}$$

The quality of fit, and thus the performance, depends on the experimental design. Hence, for all test problems, a set of different Latin hypercube designs [44] was used as a way of averaging out the experimental design dependence of the results. We used the MATLAB [45] function `lhsdesign` set with the `maximin` option with 1000 iterations to generate the experimental designs for fitting. For the Branin–Hoo and Hartman we used 1000 experimental designs and for Rosenbrock we used 100 (because of the high cost of cross validation). Table 2 details the data set generated for each test function. Naturally, the number of points used to fit surrogates increase with dimensionality. We also use the Branin–Hoo and the Hartman functions to investigate to investigate the effect of the point density. We fitted the Branin–Hoo function using 17 and 34 points and the Hartman function with 56 and 110 points. For the computation of the cross-validation errors, in most cases we use the leave-one-out strategy ($k = p$ in the k-fold strategy, see Sec. II.C). However, for the Rosenbrock function, due to the high cost of the leave-one-out strategy; we used the k-fold strategy with $k = 22$, instead. For all problems, we use a large Latin hypercube design for evaluating the actual conservativeness and the relative error growth of the surrogates [by Monte Carlo evaluation of Eqs. (6) and (8)]. These experimental designs are also created by the MATLAB Latin hypercube function `lhsdesign`, but set with the `maximin` option with 10 iterations.

**Data available online at http://sites.google.com/site/fchegury/ [retrieved 10 April 2010].

†† The interested reader can certainly find other packages (e.g., those available at http://www.kernel-machines.org, http://www.support-vector-machines.org, http://www.sumo.intec.ugent.be, and the free companion code of [5]) [retrieved 10 April 2010].

Table 2 Experimental design specifications for each test function

Test problem	No. design variables	No. points for fitting	No. points for test
Branin–Hoo[a]	2	17 and 34	10,000
Hartman[b]	6	56 and 110	10,000
Rosenbrock[c]	9	110	12,500

[a]For the Branin–Hoo function, we chose 17 points to have reasonable accuracy for the two surrogates (34 is just the double).
[b]For Hartman, we chose 56 points, because 56 is twice the number of terms in the full quadratic polynomial in six dimensions (110 is roughly the double).
[c]For Rosenbrock, we chose 110 points, because 110 is twice the number of terms in the full quadratic polynomial in nine dimensions.

C. Results and Discussion

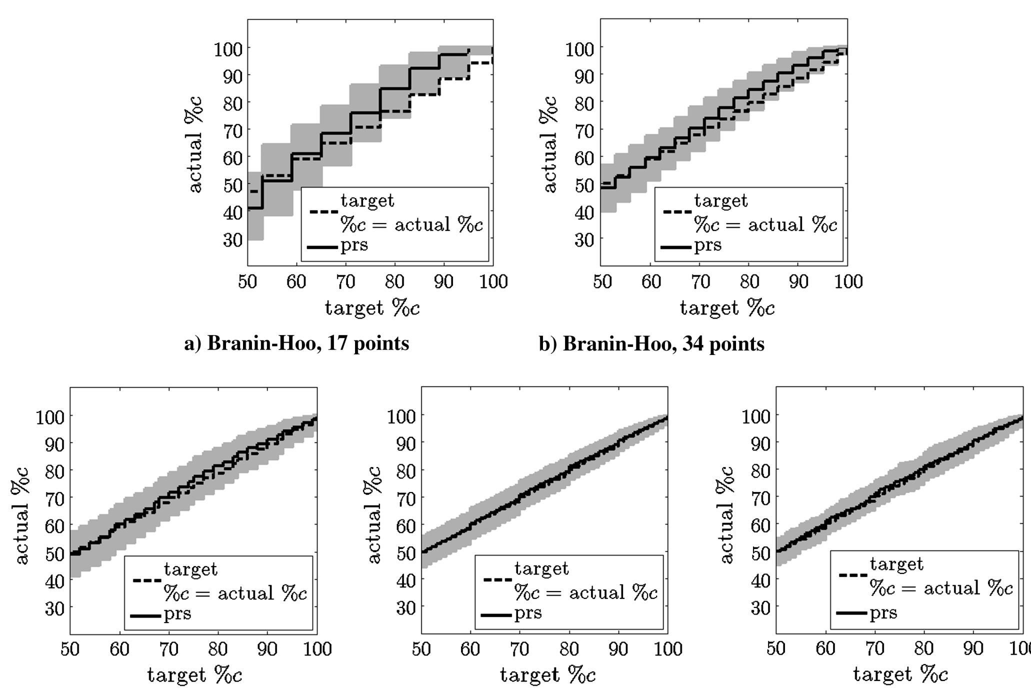
First, we checked whether cross validation is reliable for selecting the safety margin. Figure 7 illustrates the performance of the polynomial response surface model in estimating the conservativeness level. We designed the safety margin for a given target conservativeness using Eq. (11) (cross-validation errors). We checked the actual conservativeness using Eq. (6) (large set of test points). For small number of points (for example, Figs. 7a and 7b), we can incur large errors in the conservativeness level (due to wrong selection of safety margin). However, increasing the number of points allows better accuracy in selection of safety margin, as seen in Figs. 7c–7e (although it is difficult to know the precise required number of points for a given application). Note that sparseness does not seem to be important. Figures 7c–7e show that the Hartman and Rosenbrock functions with 56 and 110 points, respectively, do well even though there is less than one point per orthant. Additionally, we can make two observations:

1) Figure 7a shows that we underestimate the actual conservativeness. It means when we ask for 50% conservativeness, the distribution of cross-validation errors actually says that we should add a negative safety margin. Although this is not a safe practice, we observed that it happens more frequently in the low levels of conservativeness. In most applications, we are interested in high levels of conservativeness.

2) Figure 7e shows we could adequately estimate the safety margin even when we use the k-fold strategy for cross validation.

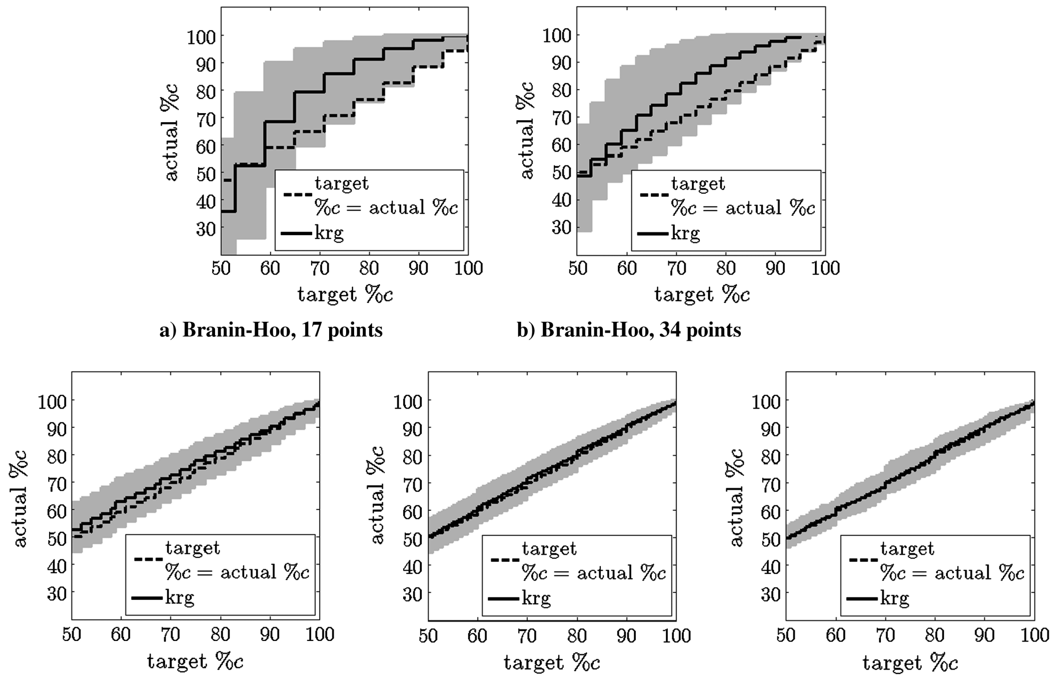
Figure 8 is the counterpart of Fig. 7 for the kriging model. Again, the safety margin is designed using cross-validation errors and actual conservativeness is checked with test points. Here, kriging strongly overestimates the safety margin for the Branin–Hoo. We believe that the reason for such behavior is the void created when we left out one point during cross validation. Figure 9a illustrates the volume fraction of the largest empty circle that can fit in between points when one point is removed for cross validation. On average, this void occupies a volume as large as 30 and 16% of the design space when 17 and 34 points are used, respectively. Figure 9b shows the ratio of the volume fraction of the voids of cross-validation data and the original experimental design. The volume of the voids in cross validation is about 50% times larger than the voids in the original data set. Surrogates that do interpolation of the data, such as kriging, are very sensitive to this. As a result, cross validation tends to overestimate the errors, as seen in Figs. 9c and 9d. Because of that, we close the discussions on kriging for Branin–Hoo function knowing that we overestimate the safety margins, which leads to overconservative (safer than desired) surrogates. From this point on, for Branin–Hoo, we only show the results for the polynomial response surface. For all other functions, both kriging and polynomial response surface benefits from increased number of points and the cross validation tends to estimate the errors efficiently.

Next, we estimate the errors due to limited sampling in the actual conservativeness using the Wilson interval (see Sec. III). Figure 10 shows the comparisons between the confidence limits of the actual



a) Branin-Hoo, 17 points
b) Branin-Hoo, 34 points
c) Hartman, 56 points
d) Hartman, 110 points
e) Rosenbrock, 110 points

Fig. 7 Estimation of actual conservativeness using cross-validation errors using polynomial response surface (prs). Solid line represents the median of the actual conservativeness over the experimental designs (1000 of them all functions but Rosenbrock, which has only 100). Gray area represents the [10–90] percentiles. Dashed lines represent perfect estimation (target$\%c$ = actual$\%c$). Increasing the number of points (in spite of sparsity) allows better accuracy in selection of safety margin.



a) Branin-Hoo, 17 points

b) Branin-Hoo, 34 points

c) Hartman, 56 points

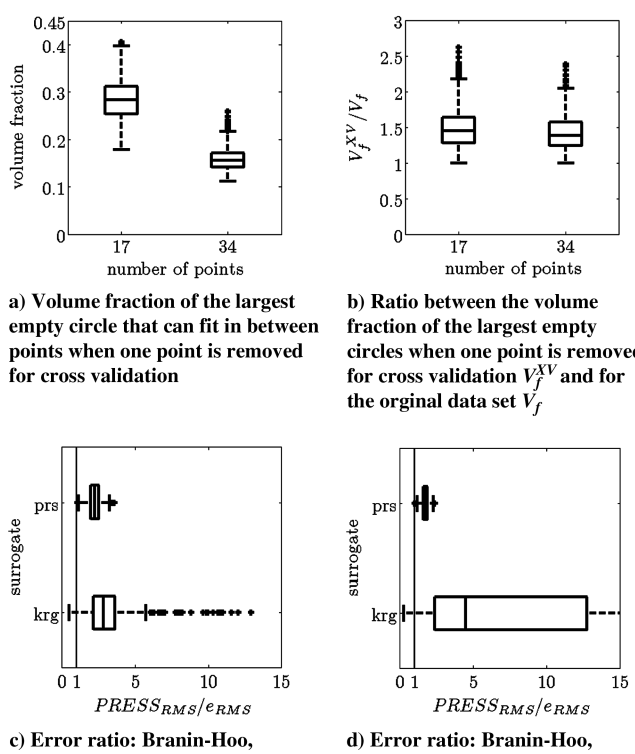
d) Hartman, 110 points

e) Rosenbrock, 110 points

Fig. 8 Estimation of actual conservativeness using cross-validation errors using kriging. Solid line represents the median of the actual conservativeness over the experimental designs (1000 for all functions but Rosenbrock, which has only 100). Gray area represents the [10–90] percentiles. Dashed lines represent perfect estimation (target$\%c$ = actual$\%c$).

conservativeness over the experimental designs and its estimation based on Eq. (14). With enough points, the Wilson interval gives a reasonable estimate of the confidence intervals for higher levels of conservativeness (see results for Hartman). For Rosenbrock, the results are not as good in the lower levels of conservativeness (we have to remember that, in practice, we are interested in the opposite range: that is, higher levels of conservativeness). We suspect that several factors contribute for that such as the use of k-fold cross validation (since a large number of points is left out, it is harder to estimate the smaller errors needed in the lower conservativeness range) and the small number of experimental designs (we used only 100, and we expect that by the time that we get to 1000 experimental designs, the figure would be less noisy). We believe that the use of the k-fold strategy is an interesting topic of future research.

Finally, we check whether cross validation allows estimation of relative error growth. Figure 11 compares the actual and the estimated relative error growth as a function of the target conservativeness. The actual relative error growth is checked with Eq. (8) (large set of test points) and using the e_{RMS} of the unbiased surrogates as reference. The estimated relative error growth is obtained with Eq. (15) and uses the $\mathrm{PRESS}_{\mathrm{RMS}}$ of the unbiased surrogates as reference. Once again, the estimates are poor for small number of points, but increasing the number of points permits better estimation. This trend is observed from Figs. 11a–11e.



a) Volume fraction of the largest empty circle that can fit in between points when one point is removed for cross validation

b) Ratio between the volume fraction of the largest empty circles when one point is removed for cross validation V_f^{XV} and for the orginal data set V_f

c) Error ratio: Branin-Hoo, 17 points

d) Error ratio: Branin-Hoo, 34 points

Fig. 9 Analysis of cross validation for Branin-Hoo (1000 experimental designs). Appendix B describes box plots. Cross validation might not always benefit from increased density because of the high impact of the voids in interpolators such as kriging.

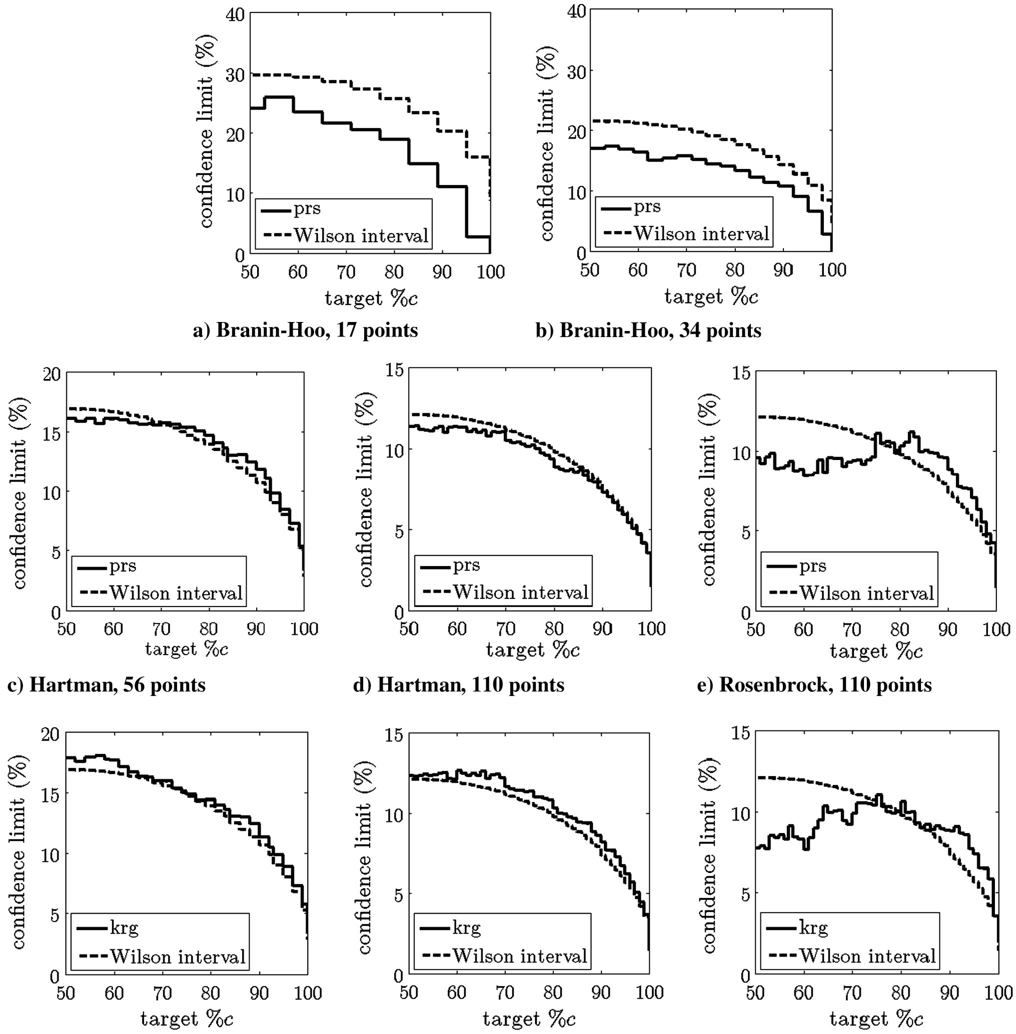
V. Application to the Design of a Cryogenic Tank

A. Problem Definition

In this section, conservative estimates are obtained for reliability measures of a composite laminated panel under mechanical and thermal loadings (Fig. 12). The panel is used for a liquid hydrogen tank. The cryogenic operating temperatures are responsible for large residual strains due to the different coefficients of thermal expansion of the fiber and the matrix, which is challenging in design.

The deterministic and probabilistic design optimizations of the barrel were performed by Qu et al. [46], using response surface approximations for probability of failure calculations. In this section, the probabilistic optimization problem of [46] is considered for investigating conservative surrogates. The geometry, material parameters, and loading conditions are taken from their paper.

The composite panel is subject to stress resultants caused by internal pressure ($N_x = 8.4 \times 10^5$ N/m and $N_y = 4.2 \times 10^5$ N/m) and thermal loading due to the operating temperature in the range of 20 to 300 K. The objective is to minimize the weight of the composite panel, which is a symmetric balanced laminate with two ply angles $[\pm\theta_1, \pm\theta_2]_s$ (that means an eight-layer composite). The design



a) Branin-Hoo, 17 points b) Branin-Hoo, 34 points

c) Hartman, 56 points d) Hartman, 110 points e) Rosenbrock, 110 points

f) Hartman, 56 points g) Hartman, 110 points h) Rosenbrock, 110 points

Fig. 10 Estimation of the error in the actual conservativeness versus target conservativeness: coverage probability of the [10–90] percent interval (results for 1000 experimental designs for all functions but Rosenbrock, which has only 100).

variables are the ply angles and the ply thicknesses $[t_1, t_2]$. The geometry and loading condition are shown in Fig. 12. The material used in the laminates composite is IM600/133 graphite epoxy, defined by the mechanical properties listed in Table 3.

The minimum thickness of each layer is taken as 0.127 mm, which is based on the manufacturing constraints as well as for preventing hydrogen leakage. The failure is defined when the strain values of the first ply exceed failure strains. The deterministic optimization problem is formulated as

$$\text{Minimize}_{\,t_1,t_2,\theta_1,\theta_2}\; h = 4(t_1 + t_2) \qquad \text{such that } t_1, t_2 \geq 0.127$$
$$\varepsilon_1^L \leq S_F \varepsilon_1 \leq \varepsilon_1^U, \qquad \varepsilon_2^L \leq S_F \varepsilon_2 \leq \varepsilon_2^U, \qquad S_F|\gamma_{12}| \leq \gamma_{12}^U \tag{19}$$

where the safety factor S_F is chosen at 1.4.

Given the material properties and the design variables, the ply strains can be calculated using classical lamination theory [47] using temperature-dependent material properties. E_2, G_{12}, α_1, and α_2 are functions of the temperature. Since the design must be feasible for the entire range of temperature, strain constraints are applied at 21 different temperatures, which are uniformly distributed from 20 to 300 K. Details on the analysis and the temperature dependence of the properties are given in [46]. The solutions for the deterministic optimization problem are summarized in Table 4. Three optima have equal total thickness but different ply angles and ply thicknesses.

B. Reliability-Based Optimization Problem

Because of the manufacturing variability, the material properties and failure strains are considered random variables. Hence, the deterministic constraints in Eq. (19) must be replaced by probabilistic constraints to ensure a desired level of reliability of the system. We assess the reliability of the system through the probability of failure. Here, the critical strain is known to be the transverse strain on the first ply (direction 2 in Fig. 12), with the effect of other strains on the probability of failure being negligible [46]. Hence, failure of the system occurs when the difference G between the critical strain and the failure strain is positive:

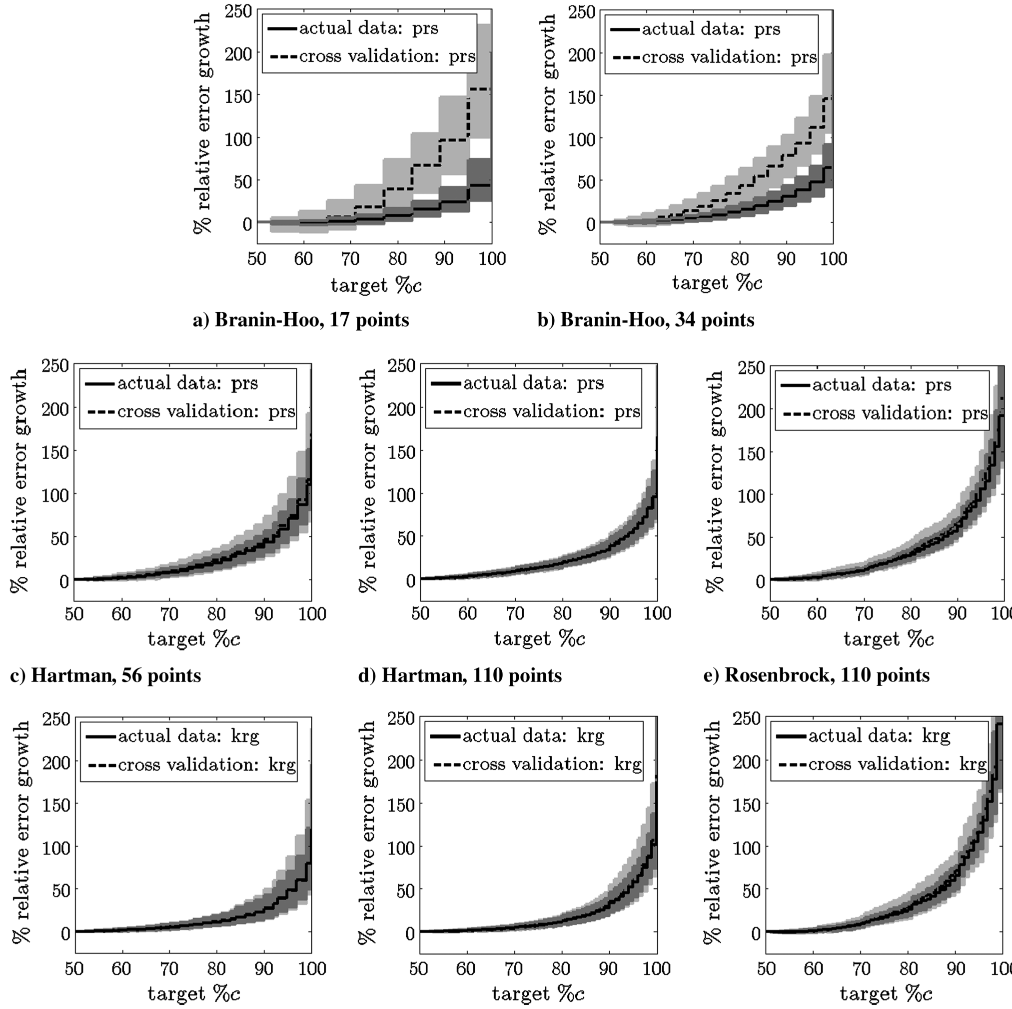
$$G = \varepsilon_2 - \varepsilon_2^U \tag{20}$$

Then the probability of failure is defined as

$$P_f = \text{Prob}(G \geq 0) \tag{21}$$

There are many methods for calculating the failure probability of a system [48]. Here, we chose to use Monte Carlo simulations. To limit the computational cost, we use separable Monte Carlo method [49] (SMC), which takes advantage of the additive form of the failure criterion to improve the accuracy of the Monte Carlo estimates.

All random variables are assumed to follow uncorrelated normal distributions. The coefficients of variation are given in Table 5. Since E_2, G_{12}, α_1, and α_2 are function of the temperature, the mean values of the random variables are calculated for a given temperature, and



a) Branin-Hoo, 17 points
b) Branin-Hoo, 34 points
c) Hartman, 56 points
d) Hartman, 110 points
e) Rosenbrock, 110 points
f) Hartman, 56 points
g) Hartman, 110 points
h) Rosenbrock, 110 points

Fig. 11 Estimation of REG equation (8) (1000 experimental designs for all functions but Rosenbrock, which has 100). Solid lines represent median of actual REG. Dashed lines represent median of estimation based on cross validation. Gray area represents the 10–90 percentiles. More points allow good estimation of REG.

then a set of random samples is generated according to their distributions.

The reliability-based optimization replaces the constraints on 2the strains in by a single constraint on the probability of failure. The target reliability of the cryogenic tank is chosen as 10^{-4}. Since the probability of failure can vary several orders of magnitude from one design to another, it is preferable to solve the problem based on the reliability index, which is denoted by β and related to the probability of failure as

$$\beta = -\Phi^{-1}(P_f) \tag{22}$$

where Φ is the cumulative distribution function of the standard normal distribution.

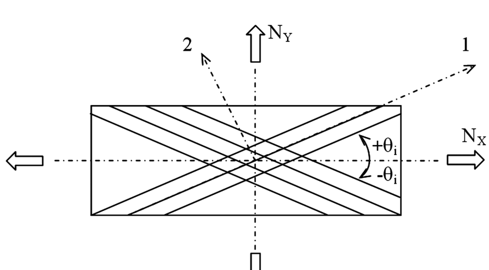


Fig. 12 Geometry and loading of the cryogenic laminate.

Table 3 Mechanical properties of IM600/133 material

Variables	Values
Elastic properties	
E_1, GPa	147
ν_{12}	0.359
E_2 [a], GPa	[8,14]
G_{12} [a], GPa	[4,8]
Coefficients of thermal expansion	
α_1 [a], K^{-1}	$[-1.5\times10^{-7}, -5\times10^{-7}]$
α_2 [a], K^{-1}	$[1\times10^{-5}, 3\times10^{-5}]$
Stress-free temperature	
T_{zero}, K	422
Failure strains	
ε_1^U	0.0103
ε_2^L	−0.013
ε_2^U	0.0154
γ_{12}^U	0.0138

[a]Temperature-dependent; the numerical values in the bracket are the range for T going from 20 to 300 K.

Table 4 Deterministic optima found by Qu et al. [46]

θ_1, deg	θ_2, deg	t_1, mm	t_2, mm	h, mm
27.04	27.04	0.254	0.381	2.540
0	28.16	0.127	0.508	2.540
25.16	27.31	0.127	0.508	2.540

Table 5 Coefficients of variation of the random variables

Variables	Values
$E_1, E_2, G_{12}, \nu_{12}$	0.035
α_1, α_2	0.035
T_{zero}	0.03
$\varepsilon_1^L, \varepsilon_1^U$	0.06
$\varepsilon_2^L, \varepsilon_2^U, \gamma_{12}^U$	0.09

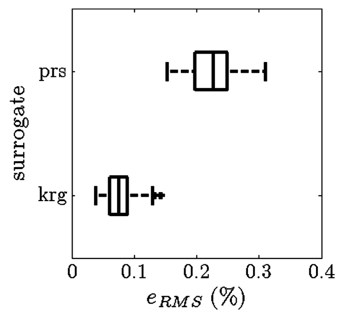


Fig. 13 Box plots of the PRESS $_{\mathrm{RMS}}$ for the reliability index surrogates. Appendix B describes box plots.

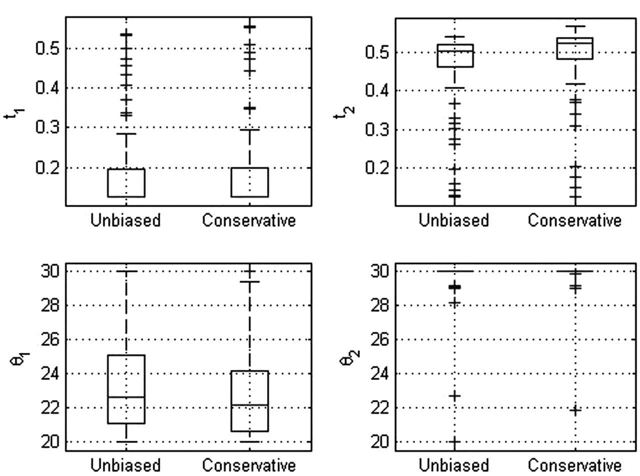


Fig. 14 Box plots of the variables of the optimal designs. Appendix B describes box plots. The main difference between the unbiased and conservative results can be seen in the mean value of t_2 (which explains the weight increase) and in the value of θ_1.

Table 6 Average of performance functions of the optimal designs

	Average objective function (thickness)	Average constraint (reliability index)	Proportion of conservative designs
Unbiased surrogate	2.59	3.64	26%
Conservative surrogate	2.67	3.75	72%

Since $-\Phi^{-1}$ is a monotonically decreasing function: a low probability corresponds to a high reliability index. Thus, a conservative estimation of β should not overestimate the true β (since a conservative estimation should not underestimate the true P_f).

The reliability-based optimization can be expressed as

$$\text{Minimize}_{\ t_1,t_2,\theta_1,\theta_2} \quad h = 4(t_1 + t_2) \quad \text{subject to } t_1, t_2 \geq 0.127$$
$$\beta(t_1, t_2, \theta_1, \theta_2) \geq \Phi^{-1}(10^{-4}) = 3.719 \tag{23}$$

C. Reliability-Based Optimization Using Unbiased and Conservative Surrogates

Solving with a sampling-based estimate of reliability is too expensive computationally. To address this issue, we choose to fit a surrogate model to approximate the reliability index and solve the optimization with the surrogate. The range of the variables is 19–30 for the angles and 0.127–0.800 mm for the ply thicknesses.

As in the previous section, a set of 100 different experimental designs is used as a way averaging out the data dependence of the results. Each experimental design and corresponding responses are generated as follow:

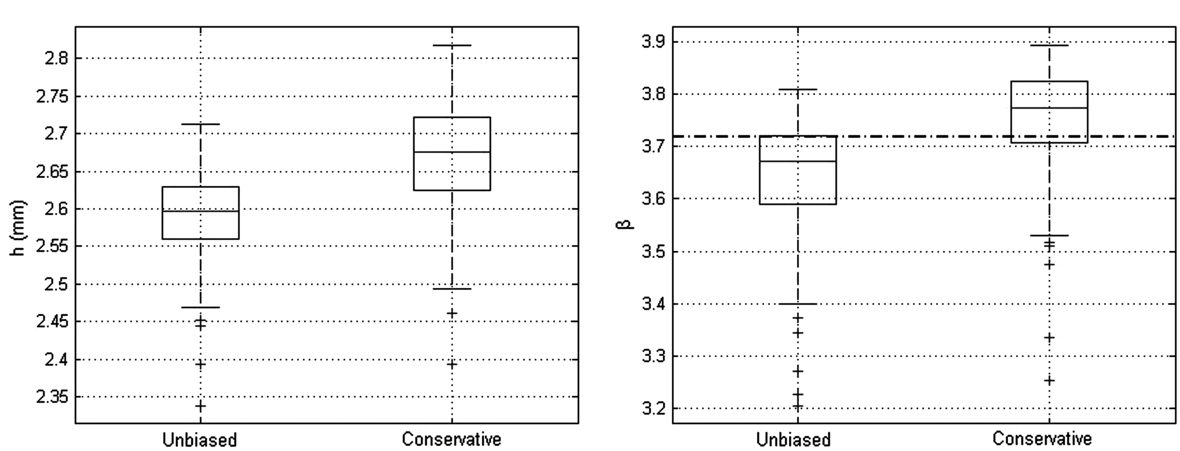
1) The experimental design consists of 100-point Latin hypercube with `maximin` criterion.

2) For each training point, the reliability index is calculated by SMC based on 1000 simulations, which leads to an observation noise of less than 1% in the reliability index.

3) The response has high local nonlinearities, which might create outliers in the experimental design. To overcome this problem, observations with responses lower than a reliability index of 1 are removed from the experimental design (approximately 3% of the data).

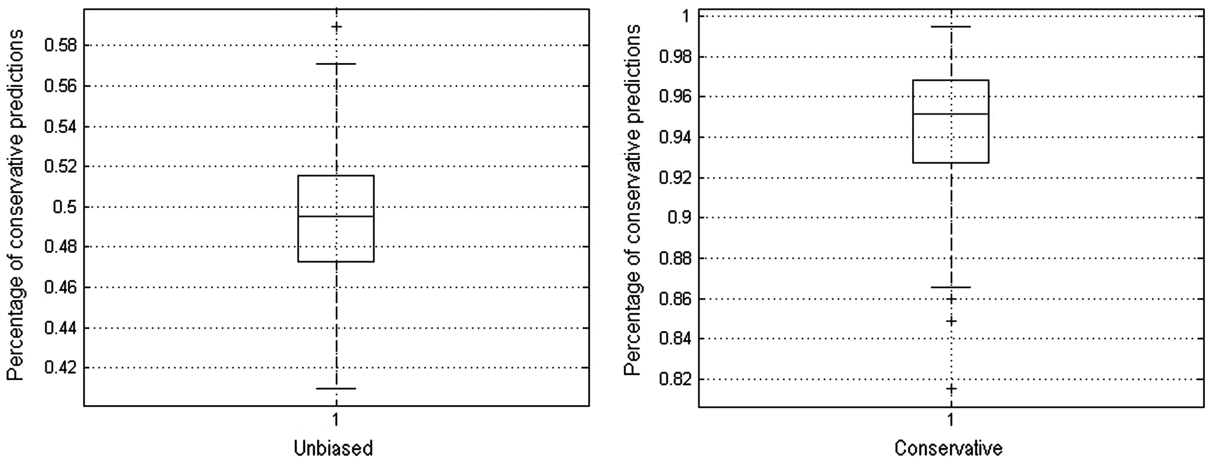
Then the two surrogates detailed in Table 1 are fitted to the 100 experimental designs. Figure 13 shows the box plot of the $\mathrm{PRESS}_{\mathrm{RMS}}$ for the surrogates with no safety margin. We choose kriging to conduct the probabilistic optimization study, because kriging is significantly better than the polynomial response surface and it presents smaller variability.

To assess the effectiveness of the conservative strategy, two optimizations are performed for each experimental design: one based



a) Objective function (thickness) **b) Constraint (reliability index)**

Fig. 15 Box plots of the performance functions of the optimal designs. Appendix B describes box plots.



a) Objective function (thickness)

b) Constraint (reliability index)

Fig. 16 Box plots of the proportion of conservative predictions at test points. Appendix B describes box plots.

on the unbiased surrogate, one based on the 95% conservative surrogate. At the optima found, the actual reliability is computed using 5000 SMC (which leads to an error of the order of 0.1% on the reliability index).

D. Optimization Results

We present the results for the two optimizations based on the unbiased and conservative KRG2. The optimization is performed using MATLAB function `fmincon`, with 40 runs based on different initial points to ensure global convergence. Figures 14 and 15 illustrate the results for the design variables and objective function and constraint, respectively. The mean values of constraint and objective functions as well as the proportions of conservative designs are given in Table 6.

The fact that an unbiased surrogate leads to only 26% of conservative results is not surprising. Indeed, optimization is biased to explore regions where the error is favorable: that is, where the constraint is underestimated. On the other hand, the safety margin seems insufficient here to provide the desired 95% conservativeness level for the same reason.

To check the overall conservativeness of the surrogate, for each design of experiment we used the remaining 99 experimental designs as test points. The proportion of conservative predictions at those 9900 test points is then computed. Results over all the experimental designs are presented in the box plots of Fig. 16. We can see that in contrast to the optimum points, on the entire design space, the unbiased surrogate leads to 50% conservative predictions on average, and the conservative surrogate to 95% conservative predictions.

Based on these results, it would appear that if the designer wants to recover the additional weight associated with the conservativeness, he may want to create a more accurate surrogate in the vicinity of the conservative optimum and repeat the process. However, this option is often not available because of cost and time constraints.

VI. Conclusions

We proposed using cross validation for designing conservative estimators and multiple surrogates for improved accuracy. The approach was tested on three algebraic examples for 10 basic surrogates including different instances of kriging, polynomial response surface, radial basis neural networks, and support vector regression surrogates. For these examples, we found the following:

1) The best surrogate changes with sampling points (density and location) and with target conservativeness.

2) Cross validation appears to be useful for both estimation of safety margin and selection of surrogate. However, it may not be accurate enough when the number of data points is small.

3) The uncertainty in the estimation of the actual conservativeness can be estimated using the Wilson interval.

4) Application to the design of composite laminate under uncertainty showed that the process of optimization may cause the design to be less conservative than expected, because optimization is biased toward regions where the surrogate is overly optimistic.

Appendix A: Cross Validation for Surrogate Selection

With advances in computer throughput, surrogates such as kriging [30,31], radial basis neural networks [50,51], and support vector regression [52,53] increasingly share a place with the traditional polynomial response surface [32,33]. The choice of the surrogate is a difficult problem, because it might hard to point out the best one just based on the data points. Furthermore, computer codes might differ in the algorithm for fitting the surrogate parameters (some codes might be more efficient than others). For all these reasons, the literature [54,55] has shown that the use of multiple surrogates may work as an insurance policy against poorly fitted surrogates. We could easily generate a large set of surrogates and select the most accurate based on cross validation [29]. The surrogate of choice is usually the one with smallest $\mathrm{PRESS}_{\mathrm{RMS}}$ value (we called it the BestPRESS surrogate).

In this appendix, we show the benefits of multiple surrogates in terms of the relative error growth. We do not expect to make the select the surrogate with best estimation of the safety margin (see the issues faced in the Branin–Hoo example while discussing Figs. 8 and 9). Instead, we expect to reduce the relative error growth due to inadequate choice of the surrogate.

Given a set of surrogates, we redefine the relative error growth by

$$\mathrm{REG} = \frac{e_{\mathrm{RMS}}}{e^{*}_{\mathrm{RMS}}} - 1 \tag{A1}$$

where e_{RMS} is taken at a given target conservativeness, and e^{*}_{RMS} is the smallest e_{RMS} of the set of surrogates without adding any safety margin.

Equation (A1) implies that even for the unbiased surrogates, there might be relative error growth. Only the most accurate surrogate (i.e., surrogate with smallest e_{RMS}, which we call BestRMSE) has $\mathrm{REG} = 0$. As we ask for conservative estimation, two things happen:

1) Because of Eq. (4), the identity of BestRMSE might change.

2) Even BestRMSE loses some accuracy (but the loss is minimal).

Thus, if the goal is to reduce the losses, we suggest to select the surrogate that has smallest estimated relative error growth. That is, for every target conservativeness, we pick the surrogate with smallest $\mathrm{PRESS}_{\mathrm{RMS}}$ [we update the $\mathrm{PRESS}_{\mathrm{RMS}}$ values using Eqs. (9) and (10)].

We employed the Hartman function sampled with 110 points and fitted to the surrogates shown in Table A1 to illustrate our strategy for surrogate selection. Again, we used the DACE toolbox of Lophaven et al. [42] and SURROGATES toolbox of Viana (see footnote **) to execute the kriging and polynomial response surface methods.

Table A1 Information about the set of 10 surrogates

Surrogates	Details
krg0, krg1, krg2	Kriging models: krg0, krg1, and krg2 indicate zero-, first-, and second-order polynomial regression model, respectively. In all cases, a Gaussian correlation and $\theta_{0i} = (p^{-1/n_v})$, and $10^{-3} \leq \theta_i \leq 2 \times \theta_{0i}$, $i = 1, 2, \ldots, n_v$ were used. We chose 3 different kriging surrogates by varying the regression model.
rbnn	Radial basis neural network: goal is $(0.05\bar{y})^2$ and spread is 1/3.
svr-grbf-full, svr-grbf-short, svr-poly-full, svr-poly-short	Support vector regression: GRBF and poly indicate the kernel function (Gaussian and second-order polynomial, respectively). All use ε-insensitive loss function. Full and short refer to different values for the regularization parameter, C, and for the insensitivity, ε. Full adopts $C = \infty$ and $\varepsilon = 1 \times 10^{-4}$, while Short uses the selection of values according to Cherkassky and Ma [56]: $\varepsilon = \sigma_y/\sqrt{k}$; and for both $C = \max(\lvert\bar{y} + 3\sigma_y\rvert, \lvert\bar{y} - 3\sigma_y\rvert)$, where $\bar{y}$ and σ_y are the mean value and the standard deviation of the function values at the design data, respectively. We chose four different surrogates by varying the kernel function and the parameters C and ε.
prs2, prs3	Polynomial response surface: full models of degree 2 and 3.

Additionally, we also used the native neural networks MATLAB toolbox [45] and the code developed by Gunn [57] to run radial basis neural network and support vector regression algorithms, respectively. We use multiple instances of different surrogates in the same fashion of Viana et al. [29] and Sanchez et al. [58]. This is possible because kriging allows different instances by changing parameters such as basis and correlation functions. As we said, although there might be better implementations, for the purpose of this paper it is advantageous to have some poor surrogates in the mix, because it is when surrogates are least accurate that compensating for their errors by conservativeness is most important. We use the same 1000 experimental designs to average out the dependency of the data points.

Figure A1 shows (in black) the frequency at which different unbiased surrogates appear as most accurate of the set out of 1000 experimental designs. The gray bars in Fig. A1 show the frequency at which the surrogates are most accurate for the experimental design no. 50 when the target conservativeness varies from 50 to 100%. We can extend what is known in the literature by saying that the accuracy will depend not only on the problem and experimental design but also on the conservativeness level. We hope that by selecting the surrogate using PRESS$_{\mathrm{RMS}}$ we can avoid further losses of a poorly fitted model.

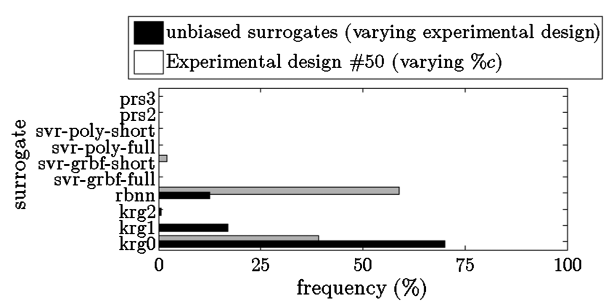


Fig. A1 e_{RMS} analysis for Hartman (110 points): 1000 experimental design and 51 target %c (from 50 to 100%). Most accurate surrogate changes not only with the design of experiment but also with the target %c.

Figure A2 gives the median over 1000 experimental designs of the relative error growth for the Hartman function fitted with 110 points. The actual relative error growth is checked with Eq. (A1) (large set of test points) using the e_{RMS} of the most accurate unbiased surrogate of the set as reference. This figure shows four surrogates:

1) The most accurate surrogate for target conservativeness close to 50% (unbiased case) is krg0.

2) The most accurate surrogate for high values of target conservativeness is rbnn.

3) The surrogate with the smallest PRESS$_{\mathrm{RMS}}$ at each conservativeness level [Eqs. (9) and (10)] is BestPRESS.

4) The surrogate with the smallest e_{RMS} at each conservativeness level [Eqs. (4) and (5)] is BestRMSE.

Figure A2a reinforces the idea that the loss in accuracy depends on the experimental design and on the conservativeness level. It shows that changing the target conservativeness may change the surrogate that offers the minimum relative error growth. For example, at target conservativeness close to 50%, the choice of krg0 leads to a median of 0% relative error growth, while the choice of rbnn leads to a relative error growth of 10%. However, for high levels of conservativeness (for example, 95%), the difference between the surrogates favors rbnn by 18%. So just by comparing at the unbiased surrogates, we would never know that the performance of the conservative surrogates may be different. We further see that cross validation successfully selects the surrogate for minimum loss of accuracy. Figure A2b shows that BestPRESS (selection based on data points) performs almost as well as BestRMSE (selection based on test points, not practical).

Appendix B: Box Plots

In a box plot, the box is defined by lines at the lower quartile (25%), median (50%), and upper quartile (75%) values. Lines extend

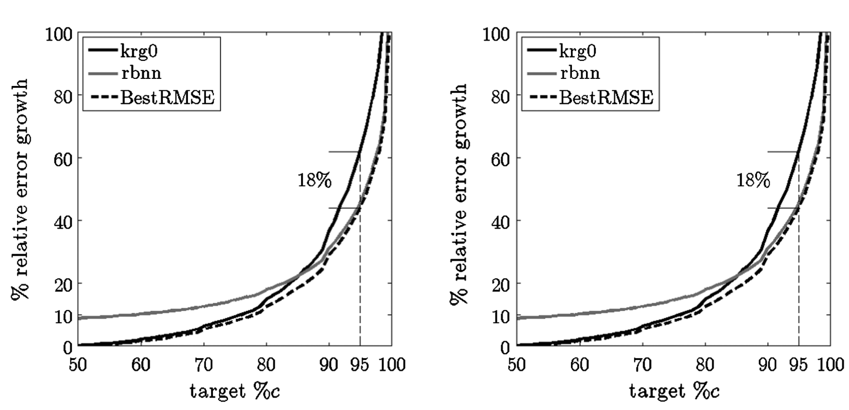


Fig. A2 Median of the actual relative error growth versus target conservativeness for the Hartman with 110 points. We can see that 1) most accurate surrogate changes with target %c and 2) cross validation successfully selects the best choice.

from each end of the box and outliers show the coverage of the rest of the data. Lines are plotted at a distance of 1.5 times the interquartile range in each direction or the limit of the data, if the limit of the data falls within 1.5 times the interquartile range. Outliers are data with values beyond the ends of the lines by placing a + sign for each point.

Acknowledgments

Authors would like to thank the U.S. Air Force Office of Scientific Research and the National Science Foundation for supporting this work under the grants FA9550-09-1-0153 and CMMI-0856431. We also thank the reviewers for their valuable comments and suggestions, which greatly improved the quality of this work.

References

[1] Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P., "Design and Analysis of Computer Experiments," *Statistical Science*, Vol. 4, No. 4, 1989, pp. 409–435.
doi:10.1214/ss/1177012413

[2] Toropov, V. V., "Simulation Approach to Structural Optimization," *Structural and Multidisciplinary Optimization*, Vol. 1, No. 1, 1989, pp. 37–46.
doi:10.1007/BF01743808

[3] Queipo, N. V., Haftka, R. T., Shyy, W., Goel, T., Vaidyanathan, R., and Tucker, P. K., "Surrogate-Based Analysis and Optimization," *Progress in Aerospace Sciences*, Vol. 41, 2005, pp. 1–28.
doi:10.1016/j.paerosci.2005.02.001

[4] Keane, A. J., and Nair, P. B., *Computational Approaches for Aerospace Design: The Pursuit of Excellence*, Wiley, Chichester, England, U.K., 2005.

[5] Forrester, A., Sobester, A., and Keane, A., *Engineering Design Via Surrogate Modelling: A Practical Guide*, Wiley, New York, 2008.

[6] Simpson, T. W., Toropov, V. V., Balabanov, V. O., and Viana, F. A. C., "Design and Analysis of Computer Experiments in Multidisciplinary Design Optimization: A Review of How Far We Have Come—Or Not," *12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, Victoria, BC, Canada, AIAA Paper 2008-5802, Sept. 2008.

[7] Starnes, J. H., Jr., and Haftka, R. T., "Preliminary Design of Composite Wings for Buckling, Stress and Displacement Constraints," *Journal of Aircraft*, Vol. 16, 1979, pp. 564–570.
doi:10.2514/3.58565

[8] Picheny, V., Kim, N-H., Haftka, R. T., and Peters, J., "Conservative Estimation of Probability of Failure," 11th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, AIAA Paper 2006-7038, Portsmouth, VA, Sept. 2006.

[9] Acar, E., Kale, A., and Haftka, R. T., "Comparing Effectiveness of Measures that Improve Aircraft Structural Safety," *Journal of Aerospace Engineering*, Vol. 20, No. 3, 2007, pp. 186–199.
doi:10.1061/(ASCE)0893-1321(2007)20:3(186)

[10] Acar, E., Haftka, R. T., and Johnson, T. F., "Tradeoff of Uncertainty Reduction Mechanisms for Reducing Structural Weight," *Journal of Mechanical Design*, Vol. 129, No. 3, 2007, pp. 266–274.
doi:10.1115/1.2406097

[11] Apley, D. W., Liu, J., and Chen, W., "Understanding the Effects of Model Uncertainty in Robust Design with Computer Experiments," *Journal of Mechanical Design*, Vol. 128, No. 4, 2006, pp. 945–958.
doi:10.1115/1.2204974

[12] Chen, S., Xiong, Y., and Chen, W., "Multiresponse and Multistage Metamodeling Approach for Design Optimization," *AIAA Journal*, Vol. 47, No. 1, 2009, pp. 206–218.
doi:10.2514/1.38187

[13] Lee, J., Jeong, H., Choi, D. H., Volovoi, V., and Mavris, D., "An Enhancement of Constraint Feasibility in BPN Based Approximate Optimization," *Computer Methods in Applied Mechanics and Engineering*, Vol. 196, 2007, pp. 2147–2160.
doi:10.1016/j.cma.2006.11.005

[14] Lee, J., Jeong, H., and Kang, S., "Derivative and GA-Based Methods in Metamodeling of Back-Propagation Neural Networks for Constrained Approximate Optimization," *Structural and Multidisciplinary Optimization*, Vol. 35, No. 1, 2008, pp. 29–40.
doi:10.1007/s00158-007-0110-x

[15] Picheny, V., Kim, N-H., Haftka, R. T., and Queipo, N. V., "Conservative Predictions Using Surrogate Modeling," *49th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Schaumburg, IL, AIAA Paper 2008-1716, April 2008.

[16] Audet, C., Dennis, J. E., Moore, D. W., Booker, A., and Frank, P. D., "A Surrogate-Model-Based Method for Constrained Optimization," 8th AIAA/NASA/USAF/ISSMO Symposium on Multidisciplinary Analysis and Optimization, AIAA Paper 2000-4891, Long Beach, CA, Sept. 2000.

[17] Picheny, V., Ginsbourger, D., Roustant, O., Haftka, R. T., and Kim, N-H., "Adaptive Design of Experiments for Accurate Approximation of a Target Region," *Journal of Mechanical Design* (submitted for publication).

[18] Bichon, B. J., Mahadevan, S., and Eldred, M. S., "Reliability-Based Design Optimization Using Efficient Global Reliability Analysis," 50th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, AIAA Paper 2009-2261, Palm Springs, CA, May 2009.

[19] Wu, Y-T., Shin, Y., Sues, R., and Cesare M "Safety-Factor Based Approach for Probability-Based Design Optimization," 42nd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, AIAA Paper 2001-1522, Seattle, WA, April 2001.

[20] Elishakoff, I., "Safety Factors and Reliability: Friends or Foes?," Kluwer Academic, Norwell, MA, 2004.

[21] Acar, E., Kale, A., and Haftka, R. T., "Comparing Effectiveness of Measures that Improve Aircraft Structural Safety," *Journal of Aerospace Engineering*, Vol. 20, No. 3, 2007, pp. 186–199.
doi:10.1061/(ASCE)0893-1321(2007)20:3(186)

[22] Kim, C., and Choi, K. K., "Reliability-Based Design Optimization Using Response Surface Method with Prediction Interval Estimation," *Journal of Mechanical Design*, Vol. 130, No. 12, 2008, Paper 121401.
doi:10.1115/1.2988476

[23] Picheny, V., "Improving Accuracy and Compensating for Uncertainty in Surrogate Modeling," Ph.D., Dissertation, Univ. of Florida, Gainesville, FL, 2009.

[24] Kohavi, R., "A Study of Cross-Validation and Bootstrap for Accuracy Estimation And Model Selection," *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, Morgan Kaufmann, San Francisco, 1995, pp. 1137–1143.

[25] Meckesheimer, M., Barton, R. R., Simpson, T. W., and Booker, A., "Computationally Inexpensive Metamodel Assessment Strategies," *AIAA Journal*, Vol. 40, No. 10, 2002, pp. 2053–2060.
doi:10.2514/2.1538

[26] Goel, T., Haftka, R. T., Shyy, W., and Queipo, N. V., "Ensemble of surrogates," *Structural and Multidisciplinary Optimization*, Vol. 33, No. 3, 2007, pp. 199–216.
doi:10.1007/s00158-006-0051-9

[27] Goel, T., Haftka, R. T., and Shyy, W., "Comparing Error Estimation Measures for Polynomial and Kriging Approximation of Noise-Free Functions," *Structural and Multidisciplinary Optimization*, Vol. 38, No. 5, 2009, pp. 429–442.
doi:10.1007/s00158-008-0290-z

[28] Viana, F. A. C., and Haftka, R. T., "Cross Validation Can Estimate How Well Prediction Variance Correlates with Error," *AIAA Journal*, Vol. 47, No. 9, 2009, pp. 2266–2270.
doi:10.2514/1.42162

[29] Viana, F. A. C., Haftka, R. T., and Steffen, V., "Multiple Surrogates: How Cross-Validation Errors Can Help Us to Obtain the Best Predictor," *Structural and Multidisciplinary Optimization*, Vol. 39, No. 4, 2009, pp. 439–457.
doi:10.1007/s00158-008-0338-0

[30] Stein, M. L., *Interpolation of Spatial Data: Some Theory for Kriging*, Springer, New York, 1999.

[31] Simpson, T. W., Mauery, T. M., Korte, J. J., and Mistree, F., "Kriging Models for Global Approximation in Simulation-Based Multidisciplinary Design Optimization," *AIAA Journal*, Vol. 39, No. 12, 2001, pp. 2233–2241.
doi:10.2514/2.1234

[32] Box, G. E. P., Hunter, J. S., and Hunter, W. G., *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, Wiley, New York, 1978.

[33] Myers, R. H., and Montgomery, D. C., *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, Wiley, New York, 1995.

[34] Simpson, T. W., Lin, D. K. J., and Chen, W., "Sampling Strategies for Computer Experiments: Design and Analysis," *International Journal of Reliability and Applications*, Vol. 2, No. 3, 2001, pp. 209–240.
doi:10.1.1.7.3757

[35] Giunta, A. A., Wojtkiewicz, S. F., and Eldred, M. S., "Overview of Modern Design of Experiments Methods for Computational Simulations," 41st AIAA Aerospace Sciences Meeting and Exhibit, AIAA Paper 2003-0649, Reno, NV, 6–9 Jan. 2003.

[36] Myers, R. H., *Classical and Modern Regression with Applications*, Duxbury Press, Belmont, CA, 2000.

[37] Congdon, C. D., and Martin, J. D., "On Using Standard Residuals as a Metric of Kriging Model Quality," 48th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, AIAA Paper 2007-1928, Honolulu, HI, April 2007.

[38] Goel, T., and Stander, N., "Comparing Three Error Criteria for Selecting Radial Basis Function Network Topology," *Computer Methods in Applied Mechanics and Engineering*, Vol. 198, Nos. 27–29, 2009, pp. 2137–2150.
doi:10.1016/j.cma.2009.02.016

[39] Stone, C. J., *A Course in Probability and Statistics*, Duxbury Press, Belmont, CA, 1995.

[40] Brown, L. D., Cai, T. T., and DasGupta, A., "Interval Estimation for a Binomial Proportion," *Statistical Science*, Vol. 16, No. 2, 2001, pp. 101–133.

[41] Wilson, E. B., "Probable Inference, the Law of Succession, and Statistical Inference," *Journal of the American Statistical Association*, Vol. 22, No. 158, 1927, pp. 209–212.
doi:10.2307/2276774

[42] Lophaven, S. N., Nielsen, H. B., and Søndergaard, J., *DACE—A MATLAB Kriging Toolbox*, Informatics and Mathematical Modelling, Rept. IMM-TR-2002-12, Technical University of Denmark, Lyngby, Copenhagen, Denmark, 2002.

[43] Dixon, L. C. W., Szegö, G. P., *Towards Global Optimization 2*, North Holland, Amsterdam, 1978.

[44] McKay, M. D., Beckman, R. J., and Conover, W. J., "A Comparison of Three Methods for Selecting Values of Input Variables from a Computer Code," *Technometrics*, Vol. 21, 1979, pp. 239–245.
doi:10.2307/1268522

[45] MATLAB, Software Package, Ver. 7.6 R2008a, The MathWorks, Inc., Natick, MA, 2008.

[46] Qu, X., Haftka, R. T., Venkataraman, S., and Johnson, T., "Deterministic- and Reliability-Based Optimization of Composite Laminates for Propeller Tanks," *AIAA Journal*, Vol. 41, No. 10, 2003, pp. 2029–2036.
doi:10.2514/2.1893

[47] Kwon, Y. W., and Berner, J. M., "Matrix Damage of Fibrous Composites: Effect of Thermal Residual Stresses and Layer Sequences," *Computers and Structures*, Vol. 64, Nos. 1–4, 1997, pp. 375–382.
doi:10.1016/S0045-7949(96)00432-4

[48] Melchers, R. E., *Structural Reliability Analysis and Prediction*, 2nd ed., Wiley, New York, 1999.

[49] Smarslok, B. P., Alexander, D., Haftka, R. T., Carraro, L., and Ginsbourger, D., "Separable Monte Carlo Applied to Laminated Composite Plates Reliability," *49th AIAA/ASME/ ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Schaumburg, IL, AIAA, 2008.

[50] Park, J., and Sandberg, I. W., "Universal Approximation Using Radial-Basis-Function Networks," *Neural Computation*, Vol. 3, No. 2, 1991, pp. 246–257.
doi:10.1162/neco.1991.3.2.246

[51] Cheng, B., and Titterington, D. M., "Neural Networks: A Review from a Statistical Perspective," *Statistical Science*, Vol. 9, No. 1, 1994, pp. 2–54.
doi:10.1214/ss/1177010638

[52] Smola, A. J., and Scholkopf, B., "A Tutorial on Support Vector Regression," *Statistics and Computing*, Vol. 14, No. 3, 2004, pp. 199–222.
doi:10.1023/B:STCO.0000035301.49549.88

[53] Scholkopf, B., and Smola, A. J., *Learning with Kernels*, MIT Press, Cambridge, MA, 2002.

[54] Viana, F. A .C., and Haftka, R. T., "Using Multiple Surrogates for Metamodeling," *7th ASMO-UK/ISSMO International Conference on Engineering Design Optimization* [CD-ROM], Bath, England, U.K., July 2008.

[55] Samad, A., Kim, K., Goel, T., Haftka, R. T., and Shyy, W., "Multiple Surrogate Modeling for Axial Compressor Blade Shape Optimization," *Journal of Propulsion and Power*, Vol. 24, No. 2, 2008, pp. 302–310.
doi:10.2514/1.28999

[56] Cherkassky, V., and, Ma., Y., "Practical Selection of SVM Parameters and Noise Estimation for SVM Regression," *Neural Networks*, Vol. 17, No. 1, 2004, pp. 113–126.
doi:10.1016/S0893-6080(03)00169-2

[57] Gunn, S. R., "Support Vector Machines for Classification and Regression," Image Speech and Intelligent Systems Research Group, Univ. of Southampton, Southampton, England, U.K., 1997.

[58] Sanchez, E., Pintos, S., and Queipo, N. V., "Toward an Optimal Ensemble of Kernel-Based Approximations with Engineering Applications," *Structural and Multidisciplinary Optimization*, Vol. 36, No. 3, 2008, pp. 247–261.
doi:10.1007/s00158-007-0159-6

A. Messac
Associate Editor