Reliability analysis performed with high-fidelity simulations is usually expensive. As a result, a popular strategy is to replace the costly computer models by surrogates built using few simulations (sampled prefentially near the limit state). Unfortunately, estimation based on surrogates introduces uncertainty in the probability of failure. In this paper, we use kriging to study the uncertainty in probability of failure estimation. Kriging models the response as a Gaussian process. For given set of kriging parameters, the prediction at any point follows a Gaussian distribution. This makes the failure region uncertain and consequently the estimated probability of failure. Additionally, the choice of kriging parameters adds another layer of uncertainty. We studied both sources of uncertainty and found that: (a) with a single set of parameters (e.g., maximum likelihood estimate), only a portion of the probability of failure uncertainty is accounted for, and (b) uncertainty in probability of failure correlates very well with kriging accuracy.

I. Introduction

When performing reliability analysis, designers usually rely on approximation techniques because high-fidelity physics-based models are very expensive. For linear or quadratic limit states, well established techniques such as the first order and second order reliability methods\cite{1,2} can be used. However, when the limit state is highly nonlinear, it is common to couple an approximation and a sampling method such as in the polynomial chaos expansion\cite{3,4} and in surrogate-based reliability analysis\cite{5,6}.

In this paper, we focus on the case in which surrogates fitted to the constraints replace expensive simulations in numerical estimation of the probability of failure.\cite{7-10} The estimate of the probability of failure has uncertainty that comes from:

- Finite sample: this portion is quantifiable. Here, we do not focus on it because surrogate predictions are fast to the point that a large number of samples are used (one can enhance computation with techniques such as importance sampling\cite{11}).

- Surrogate modeling: which relies on statistical assumptions behind the surrogate technique and few observed data.\cite{10,12-14} The assumptions might not always hold. For example, polynomial response surface\cite{15} assumes that the true function is exactly a polynomial of the assumed order, that the data is noisy with the same standard deviation everywhere, and the that noise at data points is uncorrelated. Ordinary kriging\cite{16,17} assumes that data comes from a Gaussian process with the correlation between points depending only on their relative positions. Additionally, surrogates are fitted based on few observations and chosen parameters might change with the choice of the loss and likelihood functions.\cite{18,19}

It is difficult to quantify the portion of the uncertainty in probability of failure due to unrealistic statistical assumptions of a given surrogate. However, given that we have selected a surrogate, we can estimate uncertainty due to fitting and predicted response. Here, we demonstrate it using kriging. From the Gaussian process, we sample possible outcomes for the response. Then, we propagate this uncertainty to the probability of failure estimate.

*Currently Mechanical Engineer at GE Global Research, AIAA Member.
†Distinguished Professor, Department of Mechanical and Aerospace Engineering, haftka@ufl.edu, Fellow AIAA.
The remaining of the paper is organized as follows. Section II reviews kriging presenting the Gaussian process formulation and distribution of kriging parameters conditional on data. Section III presents our scheme for quantifying the uncertainty in the probability of failure. Section IV presents the numerical experiments, results and discussion. The paper is closed with section V by recapitulating salient points and concluding remarks.

II. Kriging overview

We denote by \( g \) the response of a numerical simulator or function of interest

\[
g : D \subset \mathbb{R}^d \to \mathbb{R}
\]

\[
x \to g(x)
\]

Kriging\(^{16,17}\) models \( g(x) \) as a realization of a regression model \( \mathcal{F} \) and a stochastic process \( z \)

\[
g(x) = \mathcal{F}(\beta, x) + z(x),
\]

\[
\mathcal{F}(\beta, x) = f\beta , \text{ and }
\]

\[
E[z(w)z(x)] = \sigma^2_{PV} R(\theta, w, x),
\]

where \( \beta \) are the regression coefficients and \( f = f(x)^T \beta \) is the regression vector (\( f(x) = 1 \) is a popular choice). The stochastic process is assumed to have zero mean, process variance \( \sigma^2_{PV} \), and correlation model \( R(\theta, w, x) \) between \( z(w) \) and \( z(x) \) with parameters \( \theta \).

The response at a point \( x \) is a random variable with mean and variance defined as

\[
E[g(x)] = \hat{g}(x) = F\beta + r^T R^{-1} (g - F\beta),
\]

\[
\text{Var}[g(x)] = s^2(x) = \sigma^2_{PV} \left[ 1 + u^T (F^T R^{-1} F)^{-1} u - r^T R^{-1} r \right],
\]

where \( F \) is the matrix of linear equations constructed using the regression function at the experimental design, \( \beta \) is the vector of coefficients, \( r \) is the vector of correlations between the point \( x \) and the points of the experimental design, \( R \) is the matrix correlations between \( z \)'s at design sites (i.e., for two given points of the experimental design \( x(i) \) and \( x(j) \), \( R_{ij} = R(\theta, x(i), x(j)) \)), \( g = [g_1 \ldots g_p]^T \) is the vector of \( p \) observations, and \( u = F^T R^{-1} r - f \), and \( f = [f_1 \ldots f_p]^T \). If Gaussian correlation is used, then

\[
R(\theta, x(i), x(j)) = \prod_{k=1}^d \exp \left( -\theta_k \times (x^{(i)}_k - x^{(j)}_k)^2 \right), \text{ and }
\]

\[
G(x|\beta, \theta, \sigma^2_{PV}) \sim \mathcal{N}(\hat{g}(x), s^2(x)).
\]

Given the experimental design \( X = [x^{(1)} \ldots x^{(p)}]^T \) and observed response \( g = [g_1 \ldots g_p]^T \), the regression coefficients \( \beta \), the correlation parameters \( \theta \) and the process variance \( \sigma^2_{PV} \) can be estimated with aid of the likelihood of \( \gamma = (\beta, \theta, \sigma^2_{PV}) \) given the observed data\(^a\)

\[
L(\gamma|X, y) = \frac{1}{\sqrt{(2\pi \sigma^2_{PV})^p}} \exp \left( -\frac{(g - F\beta)R^{-1}(g - F\beta)}{2\sigma^2_{PV}} \right).
\]

The maximum likelihood estimates of the regression coefficients and the process variance are

\[
\hat{\beta} = (F^T R^{-1} F)^{-1} F^T R^{-1} g , \text{ and }
\]

\[
\hat{\sigma}^2_{PV} = \frac{1}{p} (g - F\hat{\beta}) R^{-1} (g - F\hat{\beta}).
\]

\(^a\)Other definitions are available. For example, Lophaven et al.\(^{20}\) and Bichon et al.\(^{21}\) adopt \( L(\gamma|X, g) = \frac{1}{|R|^{|\theta|}} \).

American Institute of Aeronautics and Astronautics
The correlation parameters $\theta$ are obtained through maximization of Eq. (4) using $\hat{\beta}$ and $\hat{\sigma}^2_{PV}$ from Eq. (5).

Alternatively, a Bayesian analysis of kriging$^{19,22,23}$ can be used to determine the distribution of $\gamma = (\beta, \theta, \sigma^2_{PV})$ given the observed data

$$f_{\gamma}(\gamma|X, y) \propto L(X, y|\gamma) \times f_\gamma(\gamma),$$

where $L(X, y|\gamma) = L(\gamma|X, y)$ and $f_\gamma(\gamma)$ is the prior distribution of the kriging parameters. $\beta$ different than the one obtained with Eq. (5a) prevents kriging from interpolating the observed data. To avoid that, and to constrain our study to the impact of the correlation parameters, $\theta$. We adopt the estimates of the regression coefficients and the process variance from Eq. (5).

III. Probability of failure uncertainty quantification

A. Probability of failure estimation

One of the goals of reliability analysis is to estimate the probability that a component, device, or system will fail in service. The system response is random because of uncertainty in material properties, manufacturing, etc. The system capacity is random because of uncertainty in operating conditions, strength, etc. In this paper, failure is assumed to occur when

$$g(x) > 0,$$

where $g(x) = 0$ defines the limit state. The probability of failure $p_f$ is then defined by

$$p_f = \int_{g(x)>0} f_X(x)dx,$$

where $f_X(x)$ is the joint probability density function of the random variables $x$.

Nonlinearities in the response make the integration over the failure region very difficult. Here, we compute $p_f$ through numerical integration using $n_g$ points in a grid

$$p_f \approx \frac{V}{n_g} \sum_{i=1}^{n_g} f_X(x_i) \times I(x_i),$$

where $I(x_i) = 1$ if $g(x_i) > 0$, and 0 otherwise, and $V$ is the volume of the grid.

B. Uncertainty in probability of failure using kriging models

When the response is obtained via expensive computer simulations, direct numerical integration becomes infeasible. Here, we overcome this limitation by modeling the response with kriging built from few observations. Since kriging predictions are very inexpensive, we can reduce uncertainty associated with Eq. (9). Figure 1(a) illustrates the reliability analysis of a design defined by $X \sim N(\mu = 8, \sigma = 0.05)$. The system fails if the response $g(x) - 5 > 0$. When we use the actual response, the calculated probability of failure is $p_f = 0.029$. However, because of limited computational budget, we construct a surrogate for $g(x)$ based on few observations, as shown in Fig. 1(b). Then, we replace $g(x)$ in Eq. (9) by the kriging mean (Eq. (2)) and obtain $p_f = 0.055$, as in Fig. 1(c).

We can estimate uncertainty in $p_f$ using the kriging mean and variance. Equations (2) and (3) characterize a Gaussian distribution from which we can sample possible outcomes of the response. Figure 2(a) illustrates this uncertainty turns into uncertain definition of the limit state. Ultimately, uncertainty in the limit state is synonym of uncertainty in the probability of failure. Figure 2(b) shows a histogram of probability of failure computed sampling from the Gaussian process.

There is also uncertainty in $p_f$ because of chosen kriging parameters. Figure 3(a) shows how the kriging prediction at $x = 0.9$ changes with this choice. If one constructs $n$ kriging models with $n$ parameters $\theta$ sampled from Eq. (6) and then samples possible values of the response, the distribution of the prediction tends to the Gaussian mixture of the kriging predictors equally weighted

$$f_G(x)(g(x)) = \sum_{i=1}^{n} w_i f_{\gamma_i}(g(x)), w = 1/n.$$
The response is given by $g(x) = (6x^2)^2 \sin(2 \times (6x^2))$. The current design is defined by $X \sim N(\mu = 8, \sigma = 0.05)$. Failure occurs when $g(x) - 5 > 0$. If the $g(x)$ can be used, as in Fig. 1(a), $p_f = 0.029$. When $g(x)$ is expensive, a kriging model can be fitted to few observations, as shown in Fig. 1(b) (shaded area illustrates uncertainty about the response – mean plus or minus one standard deviation). When only then mean value of the kriging model is used, as in Fig. 1(c), $p_f = 0.055$.

The response is given by $g(x) = (6x^2)^2 \sin(2 \times (6x^2))$. The current design is defined by $X \sim N(\mu = 8, \sigma = 0.05)$. Failure occurs when $g(x) - 5 > 0$. If only then mean value of the kriging model is used (solid line), $p_f = 0.055$. However, as Fig. 2(a) illustrates, there is uncertainty in the limit state because kriging mean and variance, given by Eqs. (2) and (3). Figure 2(b) shows the resulting histogram for $p_f$. 

Figure 1. Probability of failure estimation. The response is given by $g(x) = (6x^2)^2 \sin(2 \times (6x^2))$. The current design is defined by $X \sim N(\mu = 8, \sigma = 0.05)$. Failure occurs when $g(x) - 5 > 0$. If the $g(x)$ can be used, as in Fig. 1(a), $p_f = 0.029$. When $g(x)$ is expensive, a kriging model can be fitted to few observations, as shown in Fig. 1(b) (shaded area illustrates uncertainty about the response – mean plus or minus one standard deviation). When only then mean value of the kriging model is used, as in Fig. 1(c), $p_f = 0.055$.

Figure 2. Quantification of uncertainty in $p_f$ due to kriging predictor. The current design is defined by $X \sim N(\mu = 8, \sigma = 0.05)$. Failure occurs when $g(x) - 5 > 0$. If only then mean value of the kriging model is used (solid line), $p_f = 0.055$. However, as Fig. 2(a) illustrates, there is uncertainty in the limit state because kriging mean and variance, given by Eqs. (2) and (3). Figure 2(b) shows the resulting histogram for $p_f$. 

4 of 10

American Institute of Aeronautics and Astronautics
Figure 3(b) illustrates $f_G(x)$ when 2000 samples are drawn from Eq. (6). It is expected that accounting for uncertainty in the kriging parameters increases the uncertainty in the probability of failure. This can be seen first in the limit state, as shown in Fig. 3(c). This uncertainty is propagated to the probability of failure. Figure 3(d) shows a histogram of probability of failure computed sampling from the Gaussian mixture.

**Figure 3.** Quantification of uncertainty in $p_f$ due to kriging parameters. The current design is defined by $X \sim N(\mu = 8, \sigma = 0.05)$. Failure occurs when $g(x) - 5 > 0$. Figures 3(a) and 3(b) illustrates the distribution for the response at $x = 0.9$ when using three different correlation parameters and when the correlation parameters are sampled from Eq. (6), respectively. Figures 3(c) and 3(d) show the reflect in the limit state and uncertainty about $p_f$, respectively.

### IV. Numerical experiments

#### A. Branin-Hoo function

The nonlinear response is defined by the Branin-Hoo function, illustrated in Fig 4(a)

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

and $-5 \leq x_1 \leq 10$, and $0 \leq x_2 \leq 15$.

Failure is defined by

$$g_{BH}(x) - 50 > 0 .$$

The design of interest is defined by $X \sim N(\mu_X, \Sigma_X)$, where $\mu_X = [2.8]^T$ and $\Sigma_X = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$. With that, the limit state is shown in Fig 4(b) and the actual probability of failure is $p_f = 0.078$.

---

bPosterior is obtained via Monte Carlo integration using uniform priors (which support is the same search space used to solve the maximum likelihood estimation).
Figure 4. Illustration of the Branin-Hoo function. Failure occurs when \( g_R(x) - g_C(x) > 0 \), shaded area in Fig. 4(b). If \( X \sim \mathcal{N}(\mu, \Sigma) \), where \( \mu = [2 8]^T \), \( \Sigma = \sigma \), \( \sigma_i = 1 \), and \( \sigma_{ij} = 0.5 \) the probability of failure \( p_f = 0.078 \).

We study the influence of point density in the probability of failure estimates by comparing sampling Branin-Hoo with 15 points to sampling with 30 points. To average out the influence of the design of experiments, we repeat the experiments with 20 different Latin hypercube designs.\(^{26}\) The experimental designs are created by the SURROGATES Toolbox\(^{27}\) function \texttt{srgtsDOEOLHS}, set with the “ESEA” option with 10 iterations (running the enhanced stochastic evolutionary algorithm\(^{28}\)). Table 1 details the setup used to create the kriging models.

<table>
<thead>
<tr>
<th>Component</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression model</td>
<td>Constant trend function, i.e., ( f(x) = 1 )</td>
</tr>
<tr>
<td>Correlation model</td>
<td>Gaussian with correlation parameters ( 0.1 \leq \theta_i \leq 5, i = 1, 2 )</td>
</tr>
</tbody>
</table>

We can also check the accuracy of the created kriging models by using misclassification fraction\(^{29}\)

\[
m_f = \frac{\sum_i I(g(x_i) \leq \bar{g}) \oplus I(\hat{g}(x_i) \leq \bar{g})}{n_g},
\]

where \( I(\cdot) \) is the indicator function, \( \oplus \) represents the exclusive disjunction (which is 1 if only one of the operands is true and 0 otherwise), and \( n_g = 49729 \) (here, we used a \( 223 \times 223 \) grid). \( m_f \) indicates the fraction of the design space the surrogate misclassifies as either feasible or infeasible. That is, it is an indicator of how accurate the surrogate defines the limit state.

B. Results and discussion

Data density changes the uncertainty about the kriging parameters, as shown in Fig. 5. Figures 5(a) and 5(b) show the contour of the likelihood (computed with Eq. 4) when using one of the experimental designs of 15 and 30 points, respectively. Figures 5(c) and 5(d) show the marginal posterior distribution of the parameters when using one of the experimental designs of 15 and 30 points, respectively. Both the likelihood and the posterior distribution get narrow with more data.

Figure 6 shows how uncertainty in kriging prediction (which comes from uncertainty in parameters and Gaussian response) impact the probability of failure estimation. Figures 6(a) and 6(b) show the uncertainty in the probability of failure when using the MLE kriging (i.e., kriging with the parameters found by maximum likelihood estimation). In such case, the response has uncertainty because it is modeled as a Gaussian process (Eqs. 2, 3, and 4). Figures 6(c) and 6(d) show the uncertainty in the probability of failure when using the Bayesian kriging (Eqs. 4, 6, and 10). Bayesian kriging accounts for uncertainty in parameters. Both in MLE and Bayesian kriging, increasing number of points reduces the misalignment between observations and kriging statistical assumptions. The improved accuracy of kriging makes the estimates of the probability
of failure better. However, MLE kriging estimates only a portion of the total uncertainty (with very large data sets — definition of “large” depends on the number of variables — the differences between MLE and Bayesian kriging might be reduced). As a result, the probability of failure distribution is wider in Bayesian kriging than in the MLE kriging.

Increasing the number of points improves the global accuracy of the kriging model. Figure 7(a) illustrates the probability of failure versus the misclassification fraction obtained with the mean (Eq. 2) of the MLE kriging. The probability of failure estimates becomes better with more accurate definition of the limit state. That is, when moving from 15 to 30 points, the misclassification fraction reduces significantly. As a result, the probability of failure estimation is more accurate. Uncertainty in the probability of failure can be related with the uncertainty of the misclassification fraction. Figs 7(b) to 7(e) show the standard deviations of the probability of failure and misclassification fraction for the different variants of kriging. Even when comparing results with the same number of points, uncertainties on misclassification fraction and probability of failure are highly correlated. This is obviously also true when considering the improvements that increased number of points bring. The results is very encouraging because while one is not likely to have access to the misclassification fraction, one can always compute the uncertainty in the probability of failure.

V. Summary and conclusions

In this paper, we studied the influence of uncertainty in kriging in the estimation of the probability of failure. Uncertainty comes from the fact that kriging models the response as a Gaussian process and kriging parameters are only estimated. With growing computer power, we can afford using Monte Carlo to estimate the effect of both sources of uncertainty in the estimation of the probability of failure. Here, we used a two-dimensional problem to demonstrate the idea. We found that:

- when considering only a single set of kriging parameters (such as those obtained via maximum likelihood), only a portion of the probability of failure uncertainty is accounted for, and
- uncertainty in probability of failure correlates very well with kriging accuracy. Accurate kriging models
Figure 6. Uncertainty in probability of failure. Circles represents the median of the probability of failure over 50,000 samples. Bars are the 25% and 75% percentiles. As expected, increasing density of observations reduces the uncertainty about the probability of failure. Also, with the MLE kriging model only a portion of the probability of failure uncertainty is estimated.

Figure 7. Uncertainty in $p_f$ versus uncertainty in $m_f$ (defined in Eq. 11). Reduced $m_f$ shows that increasing the number of points improves definition of the limit state (as seen in the better probability of failure estimation).
provide accurate probability of failure estimates. Also, the smaller the uncertainty in the kriging root mean square error the smaller the uncertainty in the probability of failure.

As a next step, we would like to explore the use of the uncertainty estimates in sequential sampling. We believe that uncertainty in probability of failure can serve as stopping criterion and guide point selection.

Acknowledgments

This work was supported in part by U.S. Air Force Office of Scientific Research grant FA9550-09-1-0153. The views expressed here reflect the views of the authors alone, and do not reflect the views of any of their organizations or the funding agencies.

References

