49th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials
7 - 10 April 2008, Schaumburg, IL

AIAA 2008-1716

Conservative Predictions Using Surrogate Modeling

Victor Picheny[1], Nam-Ho Kim[2] and Raphael T. Haftka[3]
University of Florida, Gainesville, Florida, 32611

And

Nestor V. Queipo[4]
University of Zulia, Maracaibo, Venezuela

[Abstract] Conservative prediction refers to calculations or approximations that tend to estimate safely the response of a system. The aim of this study is to explore and compare the alternatives to produce conservative predictions when using surrogate models. We propose four different approaches: empirical approaches (Safety factors and margins), biased fitting approaches, that constrain the surrogate to be on one side of the training points, statistic-based approaches that use the prediction errors of the surrogates, and indicator kriging, that provides probabilities to exceed some cut-off values. Since the more conservative estimators tend to overestimate the true values, the problem can be considered as a multi-objective optimization, and results are presented in the form of Pareto fronts: accuracy vs. conservativeness. The best approach is the one that provide the best chance to be on the conservative side with the least impact on accuracy. Two surrogate models, polynomial response surface and universal kriging, are evaluated through two test problems: a simple analytical function and a structural analysis that uses finite elements modeling. Results show that using safety factors is the least efficient method, while the other methods are equivalent. Using safety margins results with the least variability, but statistical-based methods prevent better from large unconservative errors. The relative equivalence of safety margin and error distribution allows us to use the error distribution to accurately choose the margin corresponding to a certain level of conservativeness.

Nomenclature

$y(x)$	=	Actual response (1), (6)
$\hat{y}(x)$	=	Unbiased surrogate predictor (3), (7)
$\hat{y}_{cons}(x)$	=	Conservative surrogate predictor
PRS	=	Polynomial Response Surface (1)-(5)
IK	=	Indicator Kriging (21)-(25)
CSF	=	Constant Safety Factor (11)
CSM	=	Constant Safety Margin (12)
BF	=	Biased fitting (15)
ED	=	Error Distribution(19), (20)
RMSE	=	Root Mean Square Error (27)
MaxUE	=	Maximum Unconservative Error (29)

I. Introduction

CONSERVATIVE modeling refers to calculations or approximations that tend to safely estimate the response of a system. In many engineering problems, there is an incentive to obtain approximations that are expected to be

[1] PhD Candidate, Mechanical and Aerospace Engineering, Student Member AIAA.
[2] Assistant Professor, Mechanical and Aerospace Engineering, Member AIAA.
[3] Distinguished Professor, Mechanical and Aerospace Engineering, Fellow AIAA.
[4] Professor, Applied Computing Institute, Faculty of Engineering

as close as possible but on the safe side of the actual response. In structural analysis, such response can be a maximum stress or strain value that must not be underestimated in order to avoid failure. In this paper, we call conservative estimates that are higher than the true response. Hence, conservative estimations tend to overestimate target values; so, each conservative estimator is a trade-off between accuracy and conservativeness.

Surrogate modeling has been widely used to model complex engineering systems [1]. It consists of constructing an approximation to system response based on its value at selected designs. Most surrogates are designed to be unbiased, that is, there is a 50% chance that the prediction will be higher than the real value. In this paper, we consider the alternatives to push this percentage to the conservative side with the least impact on accuracy.

Several conservative strategies have been developed over the years. FAA defines conservative material property (A-basis and B-basis) as the value of a material property exceeded by 99% (for A-basis, 90% for B-basis) of the population with 95% confidence. FAA recommends the use of A-basis for material properties and a safety factor of 1.5 on the loads. Acar et al. [2] studied the effects of safety measures on the design of airplanes. Acar et al. [3] and Picheny et al. [4] used biased fitting of distribution functions and bootstrap methods to obtain conservative estimates of probabilities of failure. Starnes and Haftka [5] defined a convex linearization method (CONLIN) that provides first order, conservative approximations to the objective function and to the constraints.

The most widely used method is to bias the prediction response by a multiplicative or additive constant. Such approaches are called empirical because the choice of the constant is somehow arbitrary and based on previous knowledge of the engineering problem considered. The first alternative we consider is to modify the fitting of the surrogate in order to bias the predictions to be conservative. The second alternative is to use the statistical knowledge from the surrogate fitting (prediction variance) to build one-sided confidence intervals on the prediction.

In this paper, we consider two types of surrogate models: polynomial response surfaces (PRS) and kriging. Classical regression provides confidence intervals for PRS. Two types of kriging methods are used: Universal Kriging (UK) and Indicator Kriging (IK). The methods differ in a sense that one assumes a particular type of distribution (UK), while the other does not rely on a pre-specified distribution model. UK provides prediction variance that can be used to compute confidence intervals, while IK returns directly probabilities to exceed a certain threshold.

In the first part, we describe the different surrogate models and methods to obtain conservative estimations. The second part describes the test functions, error metrics and numerical procedure. Finally, the results are presented and the different approaches compared.

II. Conservative Predictors

A. Surrogates

Polynomial Response Surface

The polynomial response surface model defines the response as the sum of a linear component plus measurement error:

$$y(\mathbf{x}) = \sum_{j=0}^{p} \beta_j \xi_j(\mathbf{x}) + \varepsilon(\mathbf{x}) \tag{1}$$

Where:
- $\mathbf{x}$ is the design vector
- $y(\mathbf{x})$ the system response
- $\xi_j(\mathbf{x})$ are the polynomial basis functions
- β_j are the weights
- $\varepsilon(\mathbf{x})$ is an error measure

Given a set of design points: $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n\}$, the polynomial response surface model, in matrix notation, is defined as follow:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{2}$$

where:

$$\boldsymbol{\xi}(\mathbf{x}) = \left[\xi_0(\mathbf{x}), \xi_1(\mathbf{x}), ..., \xi_p(\mathbf{x})\right] \quad \mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} \boldsymbol{\xi}(\mathbf{x}_1) \\ \boldsymbol{\xi}(\mathbf{x}_2) \\ \vdots \\ \boldsymbol{\xi}(\mathbf{x}_n) \end{bmatrix} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} \quad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Given an estimate $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$, the estimate of y at an unsampled location $\mathbf{x}_{new}$ is:

$$\hat{y}(\mathbf{x}_{new}) = \boldsymbol{\xi}(\mathbf{x}_{new})\hat{\boldsymbol{\beta}} \tag{3}$$

$\hat{\boldsymbol{\beta}}$ is chosen to minimize the mean square error (MSE) between the estimates and the actual function values:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}\left[\hat{y}(\mathbf{x}_i) - y(\mathbf{x}_i)\right]^2 \tag{4}$$

The value of $\hat{\boldsymbol{\beta}}$ that minimizes the MSE is given by:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \tag{5}$$

Under the classical assumptions of linear regression (the residuals are independent and follow the same distribution), this estimator is BLUE (Best Linear Unbiased Estimator) under the normal error distribution assumption.

Kriging (Universal Kriging)

Kriging is an interpolating technique named after the pioneering work of D.G. Krige (a South African mining engineer) and formally developed by Matheron in 1963. This method has been widely developed in geostatistics ([5], [6]) and has recently become popular in many engineering fields [1].
In kriging, the response is modeled as a linear component + systematic departure + measurement error. Mathematically, this is expressed as:

$$y(\mathbf{x}) = \sum_{j=0}^{p}\beta_j\xi_j(\mathbf{x}) + Z(\mathbf{x}) + \varepsilon(\mathbf{x}) \tag{6}$$

The kriging estimate is a weighted sum of the observed values:

$$\hat{y}(\mathbf{x}_{new}) = \sum_{i=1}^{n} w_i(\mathbf{x}_{new})y(\mathbf{x}_i) \tag{7}$$

Where $\hat{y}(\mathbf{x}_{new})$ is the kriging prediction at the design $\mathbf{x}_{new}$, w_i the weights, $\mathbf{x}_i$ the DOE locations and $y(\mathbf{x}_i)$ the corresponding observed values.

Given the data and a correlation model (structure and parameters) the estimate can be shown to be given by the following expressions:

$$\hat{y}(\mathbf{x}_{new}) = \boldsymbol{\xi}^T(\mathbf{x}_{new})\hat{\boldsymbol{\beta}} + \mathbf{v}_x^T\mathbf{V}^{-1}\left(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}\right) \tag{8}$$

Where:

$\hat{\boldsymbol{\beta}} = \left(\mathbf{X}^T\mathbf{V}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}\mathbf{V}^{-1}\mathbf{Y}$

$V = \left[\text{cov}\left(y_i, y_j\right)\right]_{i=1..n,\, j=1..n}$

$\mathbf{v}_x^T = \left[V(y_1, y(\mathbf{x})), \ldots, V(y_n, y(\mathbf{x}))\right]$

$\boldsymbol{\xi}(\mathbf{x})$, $\mathbf{Y}$ and $\mathbf{X}$ are as defined in (2).

Note that an explicit specification of the weights is not necessary. They can be obtained though through the following expression:

$$\mathbf{w}^T\mathbf{Y} = \left[\xi^T\left(\mathbf{x}_{new}\right), \mathbf{v}_x^T\right]\begin{bmatrix} 0 & \mathbf{X}^T \\ \mathbf{X} & \mathbf{V} \end{bmatrix}^{-1}\begin{pmatrix} 0 \\ \mathbf{Y} \end{pmatrix} \tag{9}$$

Computation of the correlation model involves an optimization problem that is not described here. Most kriging packages provide correlation structure estimation; such estimation is the main source of error in the model.
Kriging provides an estimate of the prediction error variance:

$$MSE\left(\hat{y}(\mathbf{x}_{new})\right) = \sigma_z^2 - \left[\xi^T\left(\mathbf{x}_{new}\right), \mathbf{v}_x^T\right]\begin{bmatrix} \mathbf{0} & \mathbf{X}^T \\ \mathbf{X} & \mathbf{V} \end{bmatrix}^{-1}\begin{pmatrix} \xi\left(\mathbf{x}_{new}\right) \\ \mathbf{v}_x \end{pmatrix} \tag{10}$$

B. Conservative strategies

Empirical estimators

Empirical conservative estimators are obtained by multiplying or adding a constant to the unbiased estimator:

$$\hat{y}_{CS_f}\left(\mathbf{x}_{new}\right) = \hat{y}\left(\mathbf{x}_{new}\right) * S_f \quad \text{(safety factor)} \tag{11}$$

$$\hat{y}_{CS_m}\left(\mathbf{x}_{new}\right) = \hat{y}\left(\mathbf{x}_{new}\right) + S_m \quad \text{(safety margin)} \tag{12}$$

We call these estimators Constant Safety Factor (CSF) and Constant Safety Margin (CSM) estimators, respectively.

Biased-fitting estimators

The second strategy is to include a bias during the fitting process; the coefficients $\hat{\boldsymbol{\beta}}_{cons}$ are still found by minimizing the MSE, but we constrain the predicted response to be on one side of the DOE responses (that is, the error between prediction and actual response is positive at DOE points). Since kriging is an interpolation approximation, the error is null at data points, so it is only possible to build a biased fitting for polynomial response surface.
The vector $\hat{\boldsymbol{\beta}}_{cons}$ is the solution of the following constrained optimization model:

$$\begin{aligned} &\underset{\boldsymbol{\beta}}{Min} \quad MSE = \frac{1}{n}\sum_{i=1}^{n}\left[\hat{y}\left(\mathbf{x}_i\right) - y\left(\mathbf{x}_i\right)\right]^2 \\ &s.t. \quad for \quad i = 1,...,n, \quad \hat{y}\left(\mathbf{x}_i\right) - y\left(\mathbf{x}_i\right) \geq 0 \end{aligned} \tag{13}$$

In matrix notation, the problem model stated in (13) can be written as:

$$\begin{aligned} &\underset{\boldsymbol{\beta}}{Min} \quad MSE = \frac{1}{n}\left(\mathbf{X}\boldsymbol{\beta} - \mathbf{Y}\right)\left(\mathbf{X}\boldsymbol{\beta} - \mathbf{Y}\right)^T \\ &s.t. \quad \mathbf{X}\boldsymbol{\beta} - \mathbf{Y} \geq 0 \end{aligned} \tag{14}$$

Then, the biased fitting conservative estimate is given by:

$$\hat{y}_{BF}\left(\mathbf{x}_{new}\right) = \xi\left(\mathbf{x}_{new}\right)\hat{\boldsymbol{\beta}}_{cons} \tag{15}$$

Note that unlike the empirical estimates, it is not possible to control the level of bias. To do so, we propose two alternatives: the first is constraint relaxation that allows a given amount of constraint violation denoted as δ:

$$\begin{aligned} &\underset{\boldsymbol{\beta}}{Min} \quad MSE = \frac{1}{n}\sum_{i=1}^{n}\left[\hat{y}\left(\mathbf{x}_i\right) - y\left(\mathbf{x}_i\right)\right]^2 \\ &s.t. \quad for \quad i = 1,...,n, \quad \hat{y}\left(\mathbf{x}_i\right) - y\left(\mathbf{x}_i\right) + \delta \geq 0 \end{aligned} \tag{16}$$

A positive δ will reduce the bias in the fitting; δ can also be chosen negative in order to be more conservative than with no constraint relaxation.

The second alternative is to reduce the number of constraints, that is, to constrain the errors to be positive only at a selected number of points (constraint selection). The constraints selected are those *a priori* easier to be satisfied, that is, where the error from the unbiased fit is minimal. The procedure to select these points is as follow:
1- Compute the unbiased estimates using classical regression
2- Compute the errors and sort them by ascending order
3- Select the points corresponding to the k smallest errors
4- Solve the following optimization problem:

$$\begin{aligned}&\underset{\boldsymbol{\beta}}{Min} \quad MSE = \frac{1}{n}\sum_{i=1}^{n}\left[\hat{y}(\mathbf{x}_i) - y(\mathbf{x}_i)\right]^2 \\ &s.t. \quad for \quad i = 1,...,k, \quad \hat{y}(\mathbf{x}_i) - y(\mathbf{x}_i) \geq 0\end{aligned} \tag{17}$$

Where the $\mathbf{x}_i$ are sorted as described above and $1 \leq k \leq n$.

In the following, the two above-referenced biased-fitting alternatives are entitled constraint relaxation and constraint selection, respectively.

Estimator based on error distribution

Conservative estimates can also be obtained assuming the error distribution is known as provided by the surrogate analysis.

Classical regression provides a confidence interval for the predicted model. A unilateral confidence interval of level α for the response y_{new} when $\mathbf{x} = \mathbf{x}_{new}$ is given by:

$$CI = \left]-\infty, \quad \xi(\mathbf{x}_{new})\hat{\boldsymbol{\beta}} + t_{n-p-1}^{-1}(1-\alpha)s_{new}\right] \tag{18}$$

Where:

$$s_{new} = \hat{\sigma}\sqrt{1 + \xi(\mathbf{x}_{new})^T(\mathbf{X}'\mathbf{X})^{-1}\xi(\mathbf{x}_{new})}$$

$$\hat{\sigma}^2 = \frac{1}{n-p-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

t_{n-p-1} is the Student's law with *n-p-1* degrees of freedom.

We define the conservative estimator of level (1- α) as the upper bound of the confidence interval:

$$\hat{y}_{ED}(\mathbf{x}_{new}) = \xi(\mathbf{x}_{new})\hat{\boldsymbol{\beta}} + t_{n-p-1}^{-1}(1-\alpha)s_{new} \tag{19}$$

Note that this conservative estimator has the form of a margin added to the unbiased prediction. However, the margin is not constant but depends on the prediction location and the design of experiment.

Kriging assumes that the prediction is normally distributed, with mean equals to the expected prediction and variance equals to the MSE. Then, we define the conservative estimator of level (1-α) as the (1-α) percentile of the prediction distribution:

$$\hat{y}_{ED}(\mathbf{x}_{new}) = F^{-1}\left(1-\alpha; \quad \hat{y}(\mathbf{x}_{new}), MSE\left(\hat{y}(\mathbf{x}_{new})\right)^{1/2}\right) \tag{20}$$

Where $F^{-1}(p;\mu,\sigma)$ is the inverse normal cumulative distribution function of mean μ and standard deviation σ.

In the following, these estimators are called ED (error distribution) estimator.

C. Indicator Kriging

Instead of estimating the response at an unsampled location, Indicator Kriging (IK) estimates at an unsampled location the probability that the response exceeds a given value (cut-off). In other words, IK provides an estimate of the conditional cumulative distribution function (CCDF) at a particular cut-off c.

The key idea of IK is to code the observed responses into probabilities of exceeding the cut-off. Since the responses are deterministic, these probabilities are 0 or 1. The indicator coding at a sampled location $\mathbf{x}_i$ is written:

$$I_i = I\left(c;\mathbf{x}_i \,|\, \{y_1, y_2, ..., y_n\}\right) \quad = \mathrm{P}\left(y(\mathbf{x}_i) \le c\right)$$
$$= \begin{cases} 1 & if \quad y(\mathbf{x}_i) > c \\ 0 & otherwise \end{cases} \tag{21}$$

At an unsampled location $\mathbf{x}_{new}$, the probability is estimated by the kriging prediction based on the indicator data:

$$\hat{\mathrm{P}}\left(y(\mathbf{x}_{new}) \le c\right) = \hat{y}_{IK}\left(\mathbf{x}_{new}\right) \tag{22}$$

Where $\hat{y}_{IK}$ is the kriging estimate based on $\{I_1, I_2, ..., I_n\}$ instead of $\{y_1, y_2, ..., y_n\}$.

For a given set of cut-offs $\{c_1, c_2, ..., c_m\}$ and prediction location $\mathbf{x}_{\mathrm{new}}$, we obtain a corresponding set of probabilities $\{P_1, P_2, ..., P_m\}$. We use these discrete probabilities to fit a continuous approximation of the CCDF of the response at $\mathbf{x}_{\mathrm{new}}$ and build confidence intervals. IK is often qualified as a 'non-parametric' approach since it does not rely on a pre-specified distribution model. Note that it is an expensive procedure since it may require a large number of kriging models.

Post-processing is necessary to transform the IK set of values into a usable discrete CDF. Indeed, there is no constraint during the procedure to have values only inside [0, 1] or that CDF estimates vary monotonically with cut-offs. We use here one of the methods proposed in the GSLIB user's guide [8]. First, values out of the interval [0, 1] are replaced by 0 or 1. Then, the original IK-derived percentiles are perturbed by running an optimization that minimizes the perturbation while ensuring all order relations.

Finally, we fit a continuous model to the discrete data. Here, we choose to fit a first order logistic regression model. The model is defined as followed:

$$f\left(u\right) = \frac{e^{\beta_0 + \beta_1 u}}{1 + e^{\beta_0 + \beta_1 u}} \tag{23}$$

Then, the probability of the response exceeding a threshold c is given by:

$$P\left(y\left(x_p\right) \ge c\right) = f\left(c\right) = \frac{e^{\beta_0 + \beta_1 c}}{1 + e^{\beta_0 + \beta_1 c}} \tag{24}$$

The $(1-\alpha)\%$ conservative estimator is the $(1-\alpha)^{th}$ percentile, given by the inverse of the logistic regression function:

$$\hat{y}_{IK}\left(\mathbf{x}_{new}\right) = f^{-1}\left(1-\alpha\right) = \frac{1}{\beta_1}\left[\ln\left(\frac{1-\alpha}{\alpha}\right) - \beta_0\right] \tag{25}$$

D. Table of conservative estimators

Acronym	Meaning	Principle	Surrogate
CSF	Constant Safety Factor	The surrogate response is multiplied by a constant	PRS and kriging
CSM	Constant Safety Margin	A constant is added to the surrogate response	PRS and kriging
BF	Biased fitting	The surrogate is constrained to be above the training points	PRS only
ED	Error distribution	Error distribution is used to build confidence intervals	PRS and kriging
IK	Indicator Kriging	The estimate is a percentile	Multiple kriging

III. Case Studies

A. Comparison Metrics and Numerical Procedure

As discussed in introduction, conservative estimates are biased, and a higher level of conservativeness can only be obtained at a price in accuracy. Thus, the quality of a method can only be measured as a trade-off between conservativeness and accuracy.

In order to assess a global performance of the methods, we propose to define an accuracy index and a conservativeness index.

We define the indexes as follow:
Let $y(\mathbf{x})$ be the actual response at x and $\hat{y}_{cons}(\mathbf{x})$ its conservative estimate.
Given a set of m test points $\left\{\mathbf{x}_{test\_1}, \mathbf{x}_{test\_2}, \ldots, \mathbf{x}_{test\_m}\right\}$:

- the conservativeness index is equal to the proportion of conservative estimates that are greater than the actual response:

$$P\left(\hat{y}_{cons}\right) = \frac{\sum_{i=1}^{m} I\left[\hat{y}_{cons}\left(\mathbf{x}_{test\_i}\right) \geq y\left(\mathbf{x}_{test\_i}\right)\right]}{m} \quad (26)$$

Where, $I[\gamma]$ is the indicator function, which equals 1 if γ is true and 0 if γ is false.

- the accuracy index is taken as the root mean square error (RMSE) between the actual response and conservative estimate:

$$RMSE\left(\hat{y}_{cons}\right) = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left[\hat{y}_{cons}\left(\mathbf{x}_{test\_i}\right) - y\left(\mathbf{x}_{test\_i}\right)\right]^2} \quad (27)$$

In order to reduce the variability due to different DOEs, we normalize this index by the index of the corresponding unbiased surrogate (respectively unbiased response surface (5) and kriging (8)). If $\hat{y}(\mathbf{x})$ is the unbiased surrogate corresponding to $\hat{y}_{cons}(\mathbf{x})$, the normalized accuracy index of $\hat{y}_{cons}(\mathbf{x})$:

$$RMSE_{norm}\left(\hat{y}_{cons}\right) = \frac{RMSE\left(\hat{y}_{cons}\right) - RMSE\left(\hat{y}\right)}{RMSE\left(\hat{y}\right)} \times 100 \quad (28)$$

Where:

$$RMSE\left(\hat{y}\right) = \sqrt{\frac{1}{m}\sum_{i=1}^{m}\left[\hat{y}\left(\mathbf{x}_{test\_i}\right) - y\left(\mathbf{x}_{test\_i}\right)\right]^2}$$

The normalized accuracy index represents the percent increase of the root mean square error of the conservative estimator compared to the BLUE estimator. In other words, it represents the 'price' to pay to be more conservative.

The conservativeness index gives the probability to be conservative. However, it does not inform by how much we are unconservative when predictions are unconservative. Thus, an alternate measure of conservativeness is the maximum unconservative error[5] MaxUE:

$$MaxUE = \left|\min_{i}\left(\hat{y}_{cons}\left(\mathbf{x}_{test\_i}\right) - y\left(\mathbf{x}_{test\_i}\right)\right)\right| \quad (29)$$

[5] One would want to use the mean or the median of the unconservative errors instead of the maximum for more stability. However, the maximum error decreases monotonically when conservativeness is increased, while mean and median can increase when we increase conservativeness, for instance when we have initially very small and very large errors.

This index can also be normalized by the index of the unbiased estimator:

$$MaxUE_{norm}\left(\hat{y}_{cons}\right)=\frac{MaxUE\left(\hat{y}_{cons}\right)-MaxUE\left(\hat{y}\right)}{MaxUE\left(\hat{y}\right)}\times 100 \tag{30}$$

A value of $MaxUE_{norm}$ of 50% means that the maximum unconservative error is reduced by 50% compared to the BLUE estimator.

These indices require a reasonably large number of test points to be accurate. In the absence of test points, if the DOE is large, one can use the cross-validation statistics (PRESS error) instead.

For each method, we can modify the level of bias by changing:

- the value of the safety factor and margin
- the relaxation value or the number of selected constraints for BF
- the level of the confidence interval (1- α).

Then, for a given method, we take different levels of bias, and for each level, we compute the indices described above. Hence, we can draw trade-off curves (or Pareto front) between two indices. This allows us to compare the different methods by looking at partial or global dominations. A detailed example of trade-off curve generation is proposed in Appendix 1.

Finally, a crucial performance of the statistical-based predictors is their adequacy to the expected level of conservativeness. To analyze this performance, we draw the QQ-plot of the target conservativeness (1-α) vs. the actual conservativeness $P\left(\hat{y}_{cons}\right)$.

In the test problems we consider, the DOEs are generated randomly. Thus, the procedure is repeated a large number of times. We present the results as the average over these repetitions. In addition, we use error bars to represent the 95% confidence interval on the accuracy index for a given level of the conservativeness index.

The unbiased polynomial response surfaces are computed using MatLab function *regress*; the biased fitting optimization is done using MatLab function *fmincon*. The universal kriging and indicator kriging estimates are computed using the GPML toolbox for MatLab.

B. Test Problems

The Branin-Hoo function

The first test function we consider is a deterministic 2D function, which is often used to test the global optimization methods (Dixon-Szegö, 1978):

$x \in [-5, 10]$, $y \in [0, 15]$

$$f(x,y)=\left(y-\frac{5.1x^2}{4\pi^2}+\frac{5x}{\pi}-6\right)^2+10\left(1-\frac{1}{8\pi}\right)\cos(x)+10 \tag{18}$$

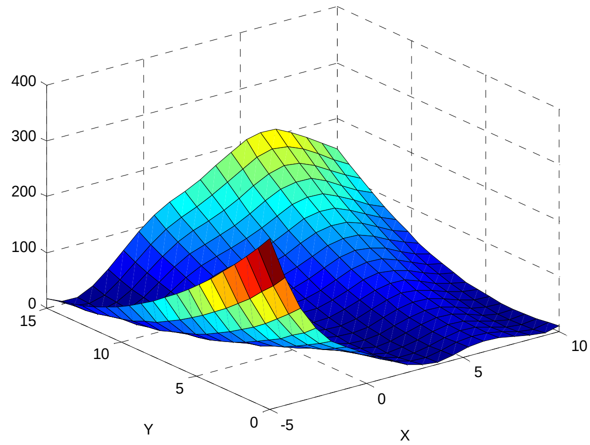

Figure 1: The Branin-Hoo function

The range of the function is between zero and 300. Large values are located on the bounds of the domain.

The Torque Arm Analysis

This second example was originally presented by Bennett and Botkin [9]. It consists of the design of a particular piece from automotive industry called a torque arm. The model, pictured in Figure 2, is under a horizontal and vertical load, F_x= -2789 N and F_y= 5066 N respectively, transmitted from a shaft at the right hole, while the left hole is fixed. The torque-arm consists of a material with Young's modulus, E= 206.8 GPa; Poisson's ratio, ν=0.29.

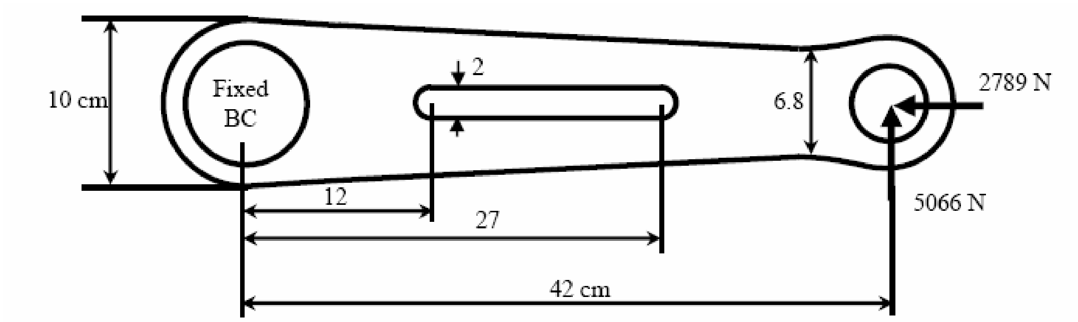

Figure 2: Initial design of the Torque Arm

The objective of the analysis is to minimize the weight with a constraint on the maximum stress. Seven design variables are defined to modify the initial shape. Figure 3 and Table 1 show the design variables and their lower and upper bounds, respectively.

Table 1: Range of the design variables (cm)

Design variable	Lower bound	Upper bound
1	-2.0	3.5
2	-0.2	2.5
3	-2.0	6.0
4	-0.2	0.5
5	-0.1	2.0
6	-1.5	2.0
7	-0.1	2.0

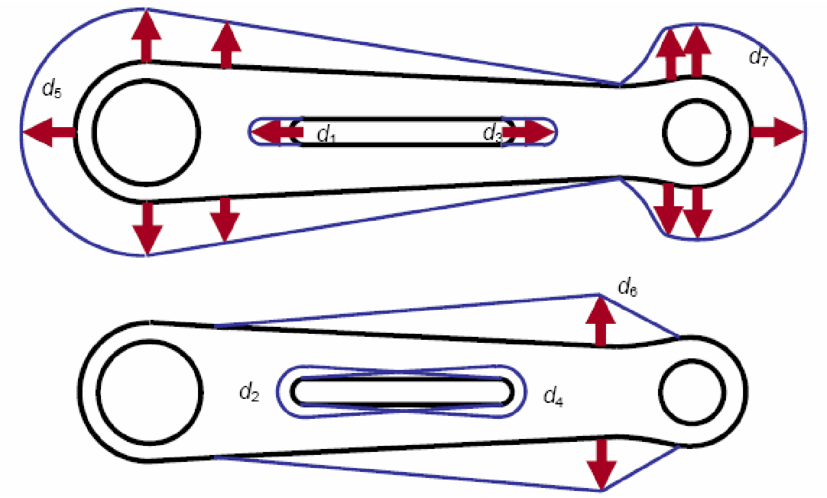

Figure 3: Design variables used to modify the shape.

The stress analysis is done by Finite Element Analysis (FEA) using the software ANSYS. In order to achieve a reasonable accuracy, a particular attention was given to the meshing technique. The model is divided into three regions, and the mesh is refined only around the areas which require it due to stress concentration. The model is built so that the mesh density can be defined along the lines, as depicted in Figure 4.

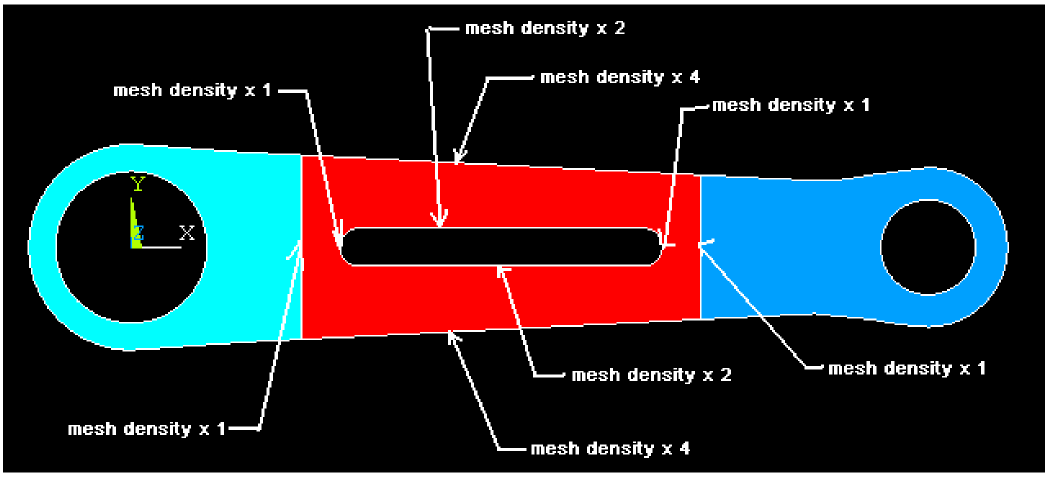

Figure 4: Non-uniform meshing process: the density of elements is defined along the lines.

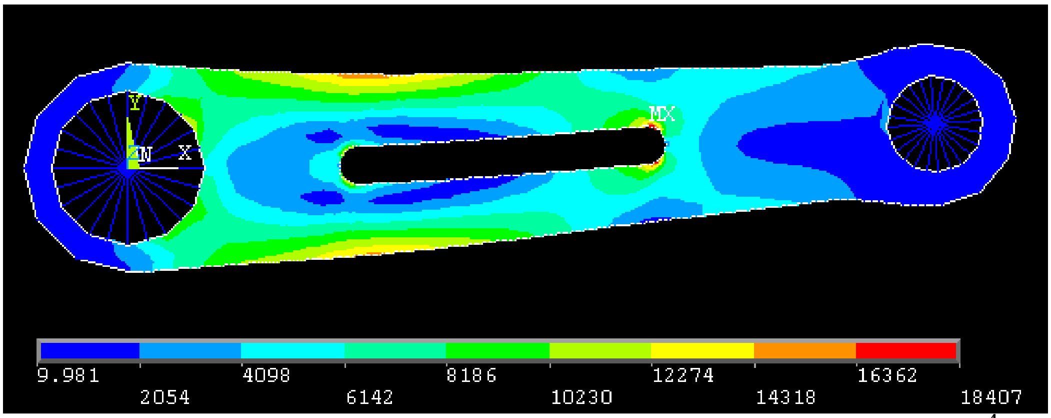

Figure 5: Typical Von Mises stress profile given by FEA (Color bar indicates values in 10^4 Pa).

Figure 5 shows a typical stress profile returned by FEA. The contour lines represent the Von Mises stress. For this design, the Von Mises stress range is between 100 and 184 MPa.

The FEA is computationally expensive; thus, a surrogate model is used to approximate the maximum stress on the design domain. Here, we do not focus on the optimization but on the quality of the surrogate prediction.

IV. Results and Discussion

A. The Branin-Hoo function

The results are presented for the following configuration:
The design of experiment consists of the four corners of the design region plus 13 points generated using Latin Hypercube Sampling (LHS) with maximum minimum distance criterion, for a total of 17 DOE points. The test points are generated using a 32x32 uniform grid (total 1024 points). The RMSE is weighted such that the points inside the domain have a weight 1, the points on the edges a weight ½ and the points on the corners ¼. The size of the DOE is chosen because it provides equivalent expectation of the RMSE for the unbiased response surface and kriging models.

The response surface is a cubic polynomial. For the kriging estimators, the covariance function used is a rational quadratic covariance function with Automatic Relevance Determination (ARD) distance measure[6], as provided by the GPML Toolbox. Indicator Kriging is not implemented for this function since the method requires a substantially higher number of training points to be accurate.

Since the DOE is generated randomly, the procedure is repeated 1000 times. We present the results as the average over these 1000 repetitions. In addition, we use error bars to represent the 95% confidence interval on the accuracy index for a given level of conservativeness.

Comparing empirical estimates
First, we want to determine the best strategy for the empirical estimators, that is, between using and additive constant (CSM) or a multiplicative constant (CSF). In order to draw the Pareto fronts, we choose the following ranges: $S_f \in [1 \quad 2]; S_m \in [0 \quad 15]$. Results are shown in Figure 6.

[6] Automatic Relevance Determination (ARD) (MacKay, 1992; Neal, 1996) is a hierarchical Bayesian approach where there are hyperparameters which explicitly represent the relevance of different input features. ARD optimizes these hyperparameters to discover which inputs are relevant.

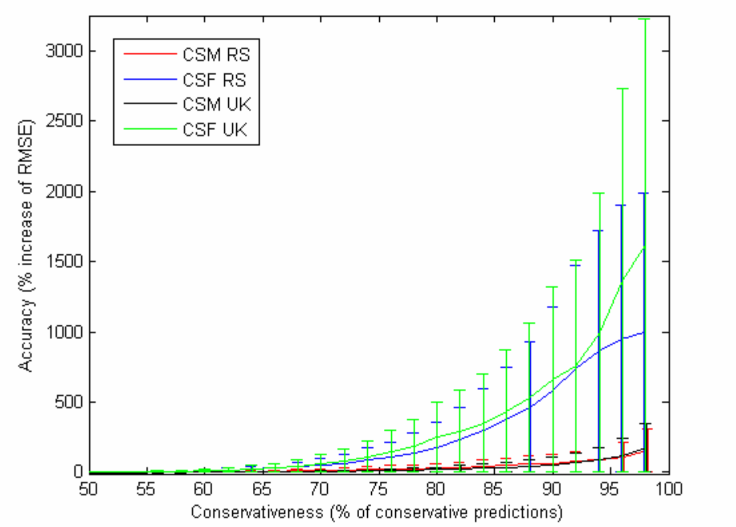

Figure 6: Pareto fronts of empirical conservative estimates for RS and Kriging. Using safety factors substantially increases the RMSE and its variability compared to using safety margins.

For both Kriging and PRS, using a safety margin is much more efficient than a safety factor. To reach the same level of conservativeness, the RMSE is much higher when using safety factors. The variability is also much higher. The reason to such difference is simply that safety factors substantially increase the bias where the predicted response is high and only a little where it is low, while the error does not depend on the response value. On the contrary, safety margins increase the bias constantly among the design region.

Comparing biased fitting estimates

Now, we compare the two strategies for biased fitting estimates. The range of the constraint relaxation is chosen as [-3; 15]; the proportion of selected constraints is chosen between 0 and 1.

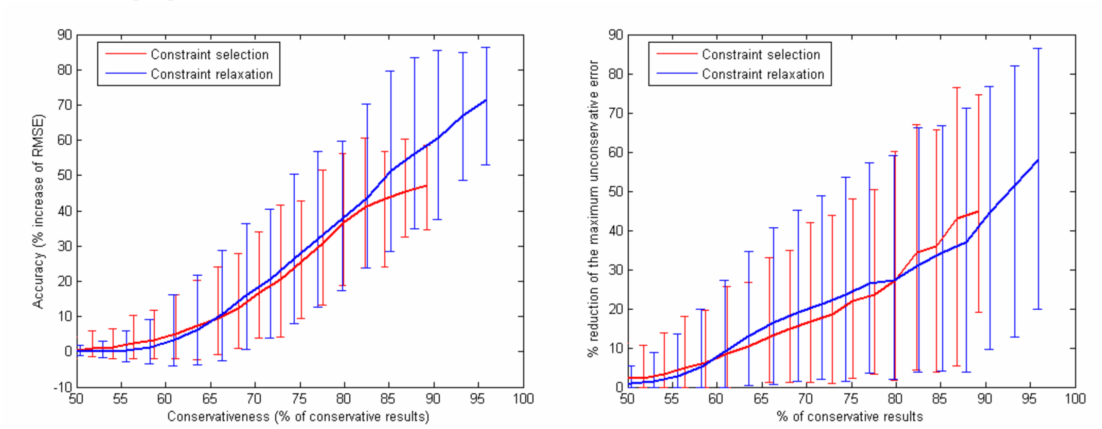

Figure 7: Biased fitting estimates with constraint relaxation (blue) and constraint selection (red). The left figure represents the trade-off between conservativeness and accuracy, the right figure shows the relationship between percentage of conservativeness and maximum unconservative error reduction.

Both methods seem to provide similar results; the difference for high percentiles on the left figure is here due to numerical noise. Using constraint selection does not allow obtaining very conservative estimates: indeed, using all the 17 constraints leads to a 85% conservativeness. On the other hand, with constraint relaxation, using a negative shift allows to be more conservative. On the right graph, we see that a high proportion of conservative estimates does not prevent for having large unconservative errors: for instance, for a 90% conservativeness, the maximum unconservative error is reduced by 40% only.

Comparing CSM and ED estimators

Now, we compare, for each metamodel, the empirical and the statistical-based approaches. We show the results of the best empirical estimator only, which is the CSM. Figure 8 shows the results corresponding to response surface, Figure 9 to kriging. For both graphs, the range of the safety margin is chosen in the interval [0; 15]; the target conservativeness is chosen between 50% and 97%.

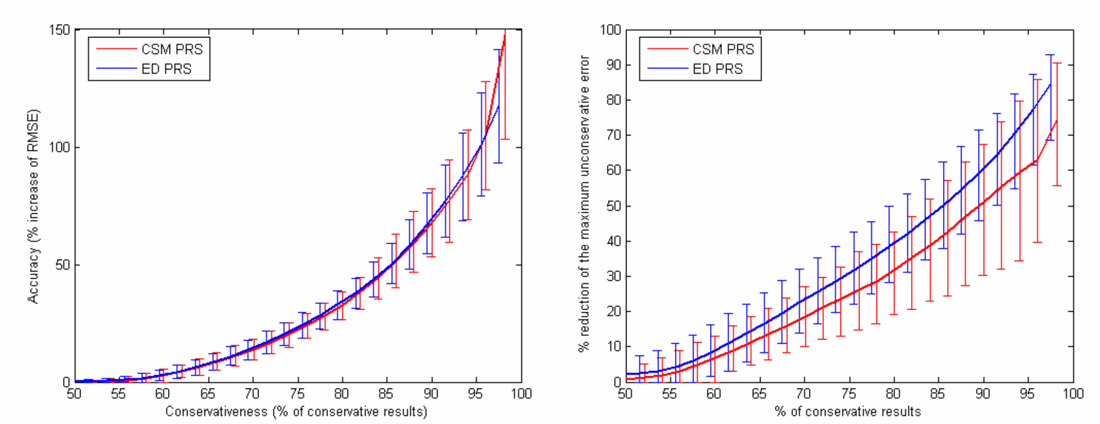

Figure 8: CSM (blue) and ED estimators (red) for RS.

For the response surface, we see on the left figure that the two methods are equivalent in terms of accuracy and variability. 95% conservativeness is obtained for an RMSE twice as big as the RMSE of the unbiased response surface. However, there is a substantial difference for the maximum unconservative error: for the same proportion of conservative results, this error is more reduced with ED than with CSM.

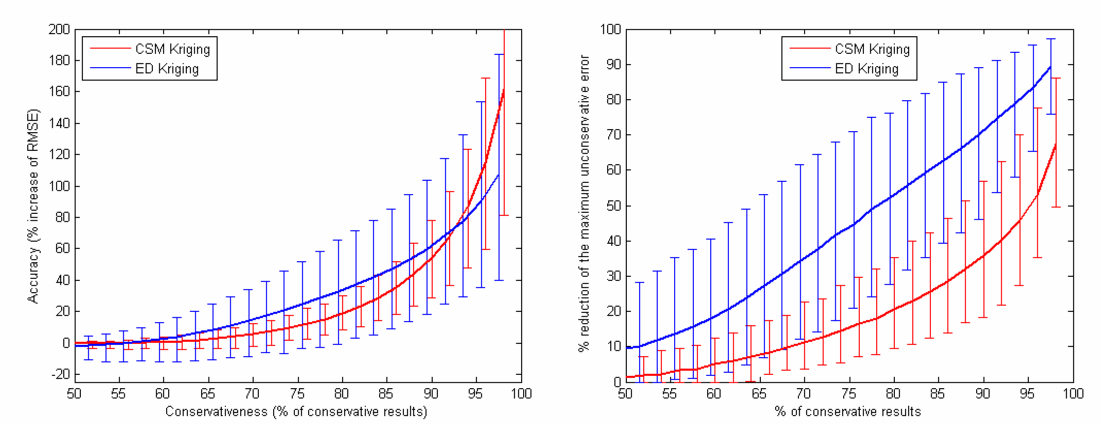

Figure 9: CSM (blue) and ED estimators (red) for Kriging.

For the kriging, the CSM estimator is clearly better than the ED estimator in terms of accuracy vs. conservativeness (left figure). The same level of conservativeness is obtained with less effect on the RMS error, especially for the 70-90% range. Moreover, the variability is much larger for the statistical estimator.

A remarkable result is the flat portion of the curve for the low conservative levels of the empirical estimator: up to 70% conservativeness can be achieved with very little effect on the RMS error. This behavior can be imputed to the nature of the kriging model: since it is an interpolation, errors are very small at the vicinity of the training points; thus, adding a small constant is sufficient to be conservative but has a little effect on the accuracy.

On the other hand, ED estimator appears to be much better to reduce the maximum unconservative error: on the right figure, the blue curve is always higher than the red one, which means, for a equivalent proportion of conservative results, the maximum unconservative error is more reduced with ED than with CSM.

When comparing response surface and kriging, we see that the mean values are equivalent, except for the flat portion. However, the variability is much bigger for kriging than for response surface. Indeed, for the Branin-Hoo function, kriging appears to be much more sensitive to the DOE values than response surface. Thus, we want to see if the observed behavior is consistent when we increase significantly the DOE size. Figure 10 shows the results for a 34-point DOE.

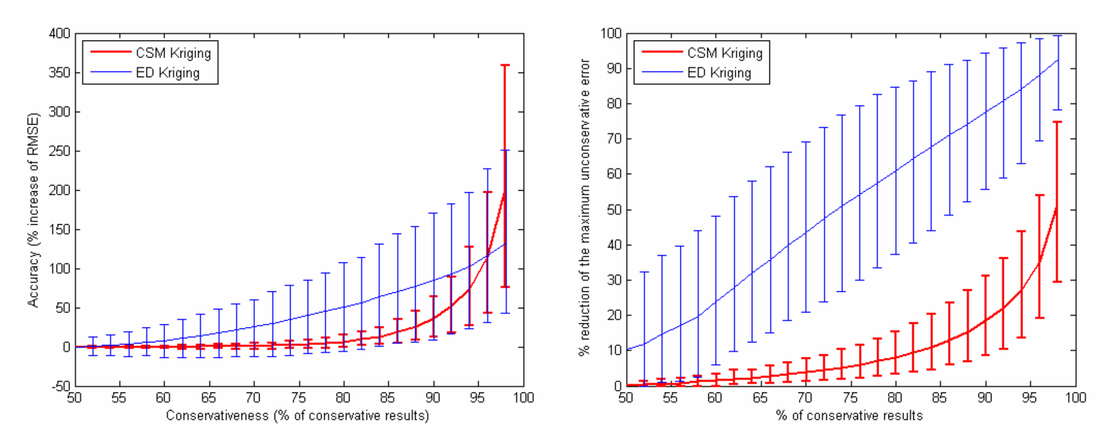

Figure 10: CSM and ED kriging estimators for 34-point DOEs.

The behavior observed for 17-point DOEs is much clearer for 34-point DOEs. The CSM estimator curve has a very flat portion up to 75% conservativeness, while the ED estimator curve increases rapidly. A longer flat portion is logical since the region where errors are small, at the vicinity of the DOE points, is larger since there are more points. The variability is much higher for the ED estimator.
On the other hand, the CSM estimator does not prevent at all from large unconservative errors: the maximum error remains almost the same even for large proportions of conservative results. The ED estimator performs a lot better.
A detailed analysis of error distribution is proposed in Appendix 2.

ED estimators: QQ-plots
The Pareto fronts showed the relation between conservativeness and accuracy. Another central measure is the fidelity to the target conservativeness. Figure 11 represents the QQ-plots for both polynomial response surface and kriging. The x-axis represents the target conservativeness, the y-axis the actual conservativeness. The error bars show the 95% confidence interval on the actual conservativeness for a given target conservativeness.

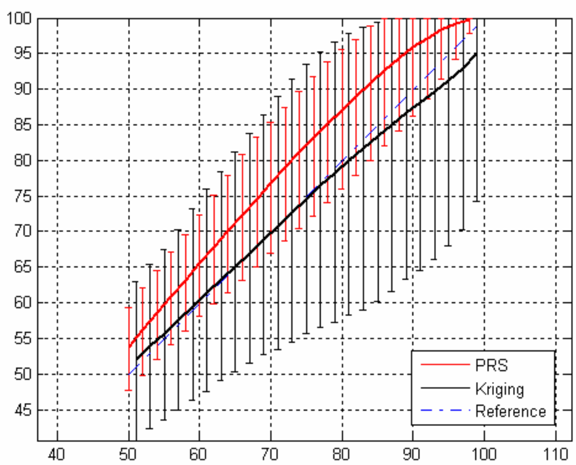

Figure 11: QQ-plots of target and actual conservativeness for PRS (red) and Kriging (black).

In average, kriging shows a very good fidelity since the trend is almost equal to the straight line. For the higher levels, the actual conservativeness is a less than the target. This is due to the fact that kriging assume normality of errors, while for the Branin-Hoo function this assumption may be at a certain degree violated. The response surface results are biased: actual conservativeness is (in average) always more than the target. Again, the assumptions of regression may be violated with the Branin-Hoo function; for the unbiased response surface, the actual conservativeness is 52% where 50% is expected.
When looking at the variability, the difference between response surface and kriging is very significant. Indeed, the confidence interval is about plus and minus 10% for the response surface, which can be considered reasonable, but it is plus or minus 20% for kriging. In particular, when 97% conservativeness is expected, the actual conservativeness can be as low as 70% only.

Comparing biased fitting and CSM response surface estimates
Finally, we compare empirical strategies with biased fitting for the response surface. Figure 12 shows the Pareto fronts for the CSM estimator and the biased fitting estimator with constraint relaxation.

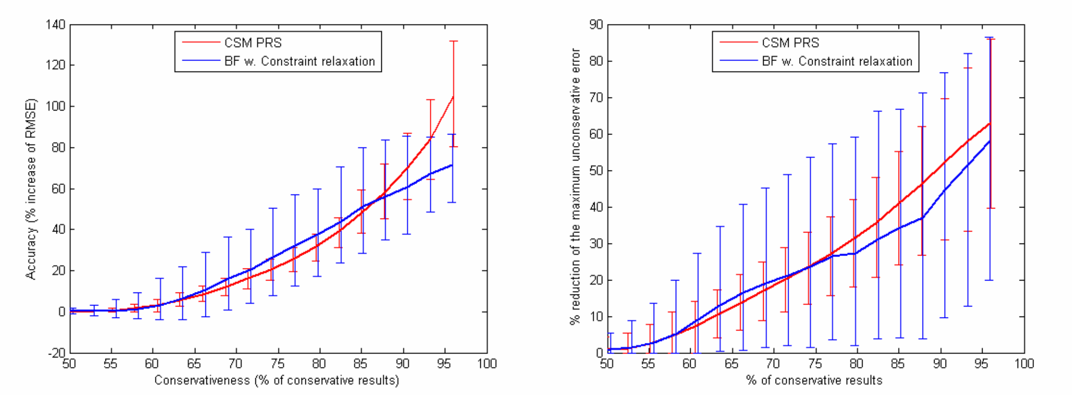

Figure 12: Biased fitting and CSM estimators.
The two estimators have similar trends, but biased fitting results with higher variability. Indeed, this method is a lot more sensitive to the DOE, since a single constraint can have a large influence on the shape of the response.

B. The Torque Arm Analysis

In this part, we consider the analysis of the Torque Arm. The analysis is set up as follow:
- the DOE consists of 300 points generated from LHS with maximum minimum distance criterion
- 1000 test points are generated using LHS
- For each point, the ANSYS code is run and returns the overall maximum stress. The values are in MPa; their range is of the order of 10^2.

The stress values strongly violate the hypothesis of normality of residuals of the regression analysis. Thus, we performed the analysis on the natural logarithm of the stress instead of the stress itself. Using such transformation, the hypothesis of normality is verified. From an engineering point of view, looking at the logarithm of the stress might not seem relevant; indeed, we focus here on the quality of the surrogate prediction rather than on the mechanical problem. However, it is to be noticed that adding a margin to the logarithm is equivalent to multiplying the stress by a safety factor.

The response surface used is a second order polynomial. For the kriging estimators, the covariance function used is the sum of a rational quadratic covariance function with ARD and white noise. For the IK, a rational quadratic function with ARD is used; the parameters of the function are re-estimated for each cut-off. A total of 100 cutoffs are used for the distribution estimation.

For the Branin-Hoo function, we chose the DOE size in order to have comparable RMS errors for the unbiased estimators. Table 2 shows the performances of the unbiased surrogates for the Torque Arm analysis (the unbiased IK corresponds to the 50^{th} percentile given by this method). Kriging performs better than polynomial response surface;

indicator kriging performs poorly compare to the other methods. However, IK has the lowest ratio between RMSE and maximum unbiased error, which means that there is less risk of having an outlier in the unconservative side. Distribution of errors are given in Appendix 3.

Table 2: Performances of unbiased surrogates for the Torque Arm.

Surrogate	RMSE	MaxUE	MaxUE / RMSE
PRS	0.0589	0.4956	8.41
Kriging	0.0415	0.2835	6.83
IK	0.1413	0.6049	4.28

Since the finite element analysis computationally expensive, it is not possible to generate a large number of DOEs for variability analysis, as we did for the Branin-Hoo function. In order to obtain confidence intervals, we randomly choose 300 points out of the 1300 points generated (300 training + 1000 test points), and use the remaining 1000 as test points. This procedure is repeated 500 times.

Comparing CSM and ED estimators

Figure 13 shows the results of the CSM and ED estimators for the polynomial response surface.

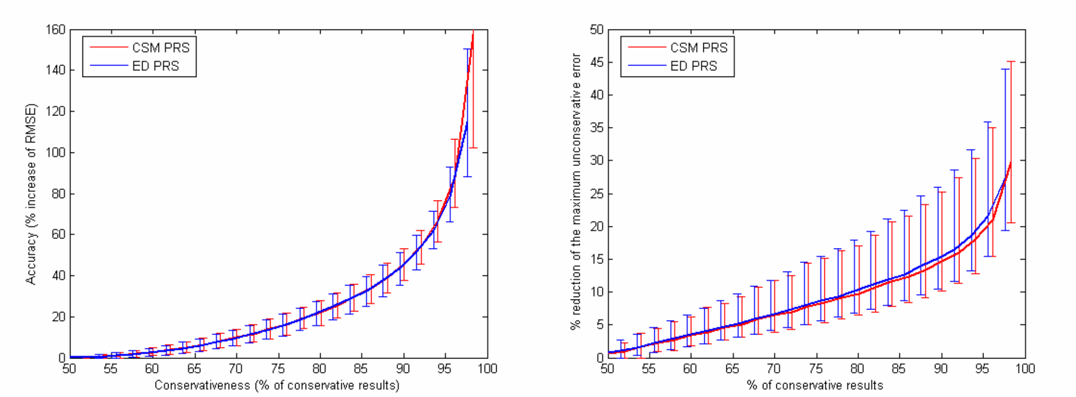

Figure 13: CSM and ED PRS estimator performances for the Torque Arm analysis.

For the torque arm analysis, using safety margins and error distribution lead to almost identical results for all our three indices. One reason to explain such equivalence is that the width of the confidence interval does not vary a lot. Figure 14 shows the histogram of all the confidence interval widths at the 1000 test points for a 80% target conservativeness. We see that the range of the size of the confidence interval is [0.042, 0.051], which means that the confidence interval range vary by less than 11% around its mean value. For a 95% target conservativeness, this value increase to 14% only.

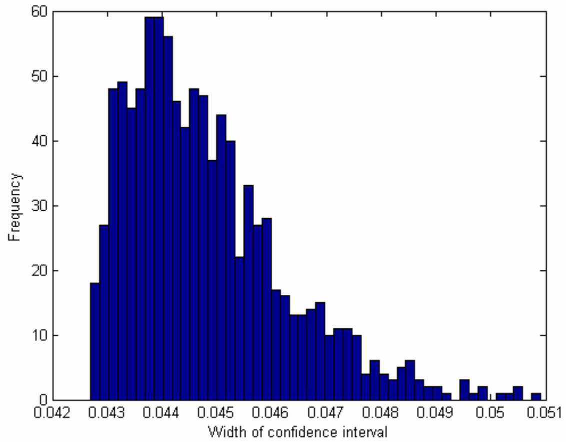

Figure 14: Histogram of the confidence interval widths at the 1000 test points for a 80% target conservativeness with PRS.

Figure 14 shows the results of the CSM and ED estimators for kriging.

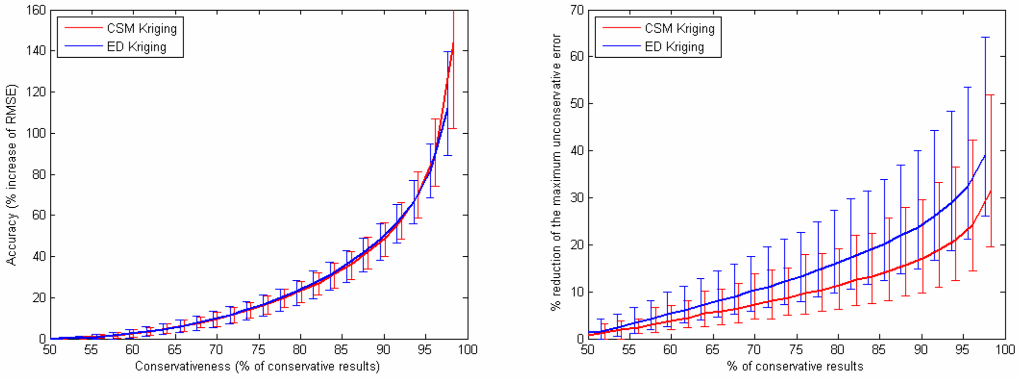

Figure 15: CSM and ED kriging estimator performances for the Torque Arm analysis.

Here, the CSM and ED estimators are equivalent for the percentage of conservativeness and RMSE. For the Branin-Hoo function, we found that CSM performed a lot better than ED, particularly when the DOE size was large. Here, the DOE size is 300, which is not large since there are seven dimensions. Thus, the flat behavior of the Pareto front due to small local errors is here negligible.
Results are different for the reduction of maximum unconservative error: here, like for the Branin-Hoo example, the ED estimator reduces a lot more the unconservative errors. Since the results are equivalent in terms of accuracy and variability, we can conclude that the kriging ED estimator outperforms the CSM estimator for this example.

Figure 16 shows the histogram of all the confidence interval widths at the 1000 test points for a 80% target conservativeness. The range is a lot larger than for PRS. There are a few very large values that may have a large impact on the RMSE.

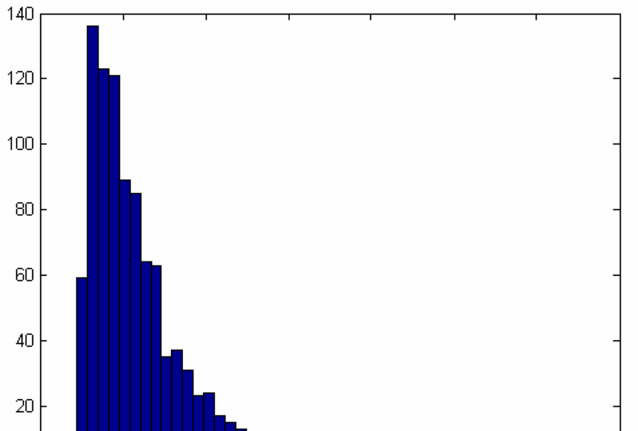

Figure 16: Histogram of the confidence interval widths at the 1000 test points for a 80% target conservativeness with kriging.

Comparing biased fitting and CSM response surface estimates

Figure 17 shows the results for the biased fitting and CSM estimators.

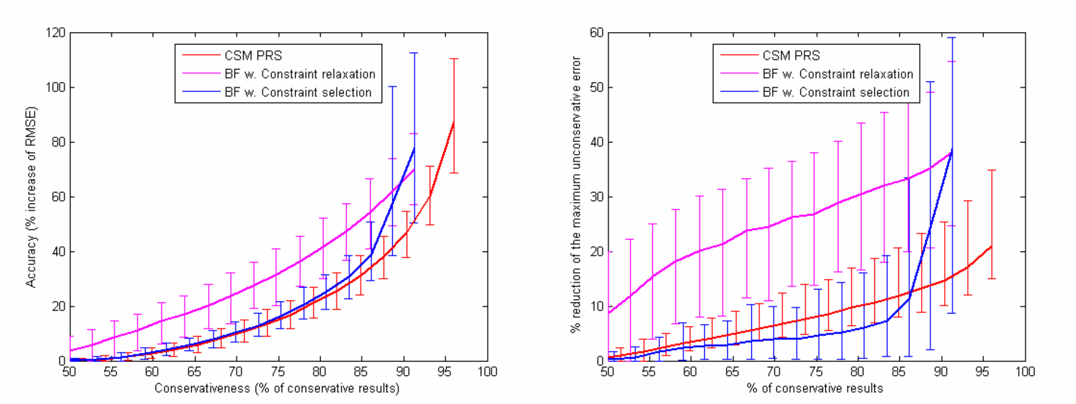

Figure 17: CSM and BF estimator performances for the Torque Arm analysis.

On the left figure, we see that constraint selection gives equivalent results to CSM, but constraint relaxation is less efficient. Indeed, its RMSE is increased for the lowest levels of conservativeness: that means without any effect on the conservativeness, the constraints change the shape of the response surface. With constraint selection, the error increase only for high levels of conservativeness. That means, only a few constraints are responsible for large errors in the model fit. These few constraints may correspond to outliers. Hence, constraint selection is an efficient way to get rid of outliers by being conservative where it is not expensive to be, while on the contrary, constraint relaxation increases error without significant effect on the conservativeness.

When looking at the reduction of unconservative errors, we see that biased fitting with constraint relaxation outperforms by far the other methods, while constraint selection is not as good as CSM. This is logical since by construction constraint relaxation tries to correct all the unconservative errors, while constraint selection tries to correct the smallest unconservative errors first.

Comparing all statistical-based estimators

Here, we want to compare the performances of the different surrogates. We saw that for PRS, CSM and ED are equivalent, and for kriging, ED performs better. Thus, we use for comparison only the ED estimators, and indicator kriging (IK).

First, we compare the absolute performances of the different surrogates, so we do not normalize our indices and use RMSE and MaxUE as defined in (27) and (29). Results are shown in Figure 18.

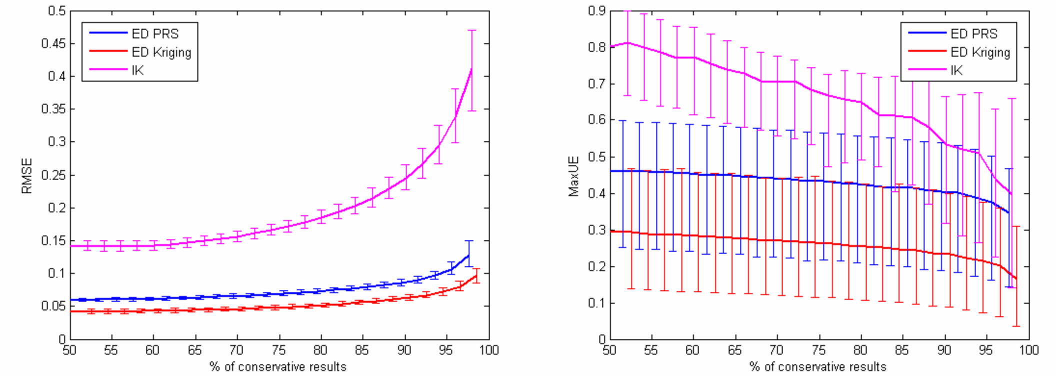

Figure 18: ED and IK estimator (not normalized) performances for the Torque Arm analysis.

On the left graph, we see that the conservative strategies do not change the order of the surrogate performances: kriging performs better than the others, and IK performs poorly. However, when looking at the maximum unconservative error, we see that IK reduces a lot more the error than the other methods, and for the highest levels of conservativeness, it is almost equivalent to PRS.

Figure 19 shows the normalized results for the three surrogates.

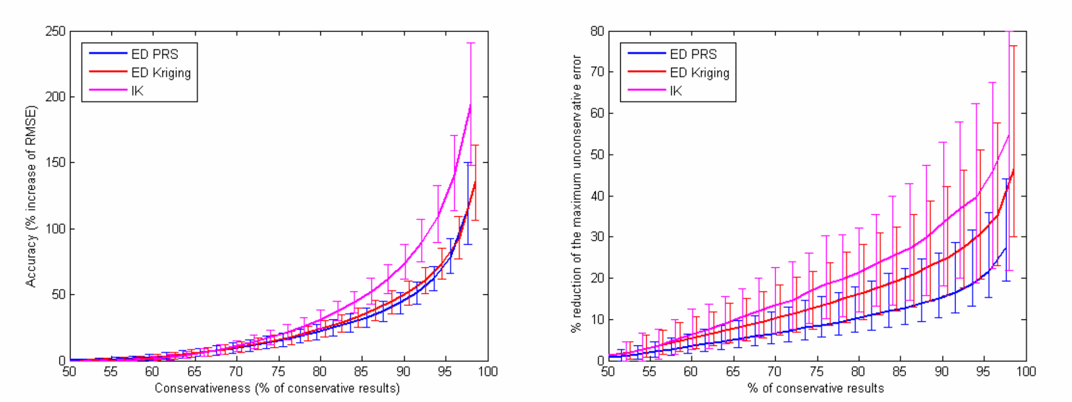

Figure 19: ED and IK estimator performances for the Torque Arm analysis.

On the left graph, Indicator kriging is accurate until 75% conservativeness; for the highest percentiles, the error increases rapidly. Indeed, indicator kriging fails to accurately predict high percentiles [10]. Kriging and PRS show very comparable results; for both, a 95% conservativeness is at a price of an 80% increase of the RMS error. A change of curvature on the Pareto front occurs around 90% conservativeness; that means, low levels of conservativeness are relatively cheap to obtain, while more than 90% increase a lot the error.
On the right graph, we see that IK performs better than the others. Kriging performs better than PRS, which is not unexpected since the kriging prediction variance is more accurate than the PRS one. For all three surrogates, we notice that the maximum unconservative error is never fully compensated, even for the highest levels of conservativeness. That means, even with a high percentage of conservative results, unconservative errors remain large.

Additional details, including distribution of errors for a 95% target conservativeness, are given in Appendix 3.

ED estimators: QQ-plots
Finally, Figure 20 shows the graph of target conservativeness vs. actual conservativeness for the statistical-based estimators. Both parametric (ED RS and kriging) and non-parametric (IK) estimators show a very good fidelity, even for high percentiles. The poor accuracy of the IK for the high percentiles does not affect the overall fidelity to the target level of conservativeness. For the parametric estimators, it means that the hypotheses of normality are not violated.

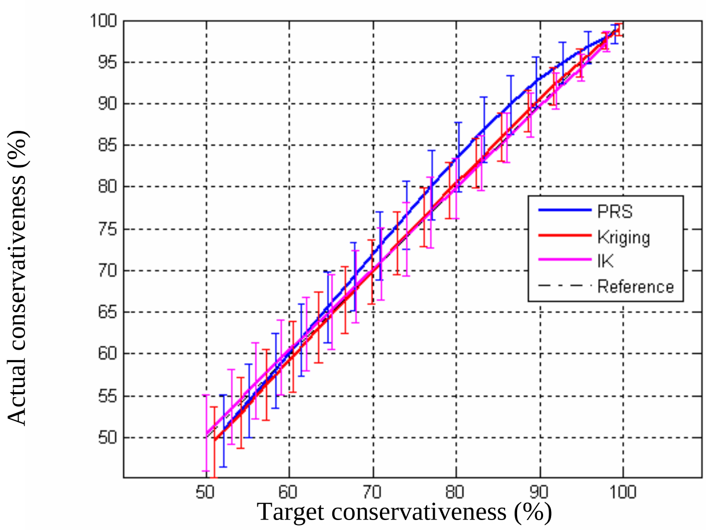

Figure 20: QQ-plots of the ED estimators

Figure 21 show the QQ-plots when the surrogates are fitted to the stresses directly and not the natural logarithm of the stresses. The violation of normality hypothesis affects a lot the fidelity of the PRS. Kriging results are not as good as in the previous case but remain acceptable. As expected, IK is the least sensitive to violation of normality.

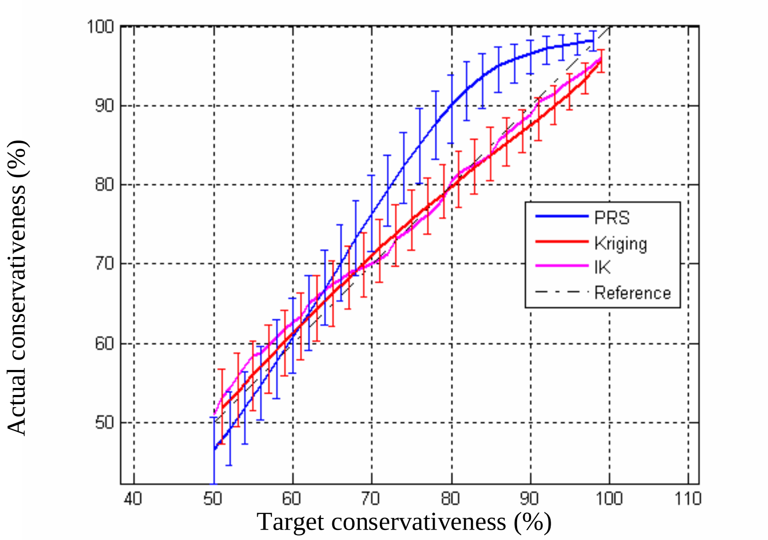

Figure 21: QQ-plots of the ED estimators when hypothesis of normality of residuals are violated.

V. Conclusion

In this report, we explored the alternatives to obtain conservative predictions using surrogate modeling. First, we showed that conservativeness could be obtained by empirical methods, such as constant safety factors or margin, by adding constraints during the fitting to obtain a biased model, or by taking advantage of error distribution measures given by the model. Then, a Pareto front methodology was introduced to measure the quality of the different methods. Finally, the methods are implemented for two tests problem: one analytical and one based on Finite Element Analysis. Results showed that:

- safety margins are much more efficient that safety factors, in particular when the range of the response is large
- the use of safety margins and error distribution lead to very comparable results when looking at an accuracy and conservativeness trade-off
- although, ED estimators, for an equivalent level of conservativeness and accuracy, prevent better than CSM from the risk of large unconservative errors
- for biased fitting, constraint selection is better than constraint relaxation, in particular in presence of outliers
- Indicator kriging is reliable in terms of level of conservativeness but shows poor accuracy, in particular for high percentiles
- ED estimators are based on assumptions that may not be violated in order to obtain acceptable fidelity of conservativeness level.

More generally, it has been shown that measuring the conservativeness of a method is not easy and can be very problem-dependant. Indeed, the chance of being conservative and the risk of large unconservative errors are two measures of conservativeness that do not behave identically. One may choose a conservative strategy based on the trade-off between these two quantities and the global measure of accuracy.

The results for the ED estimators were overall disappointing, since they provide local estimates of error, while safety margin uses a single constant an perform as well for the torque arm problem, and better for the Branin-Hoo function in terms of accuracy. However, MSE estimates do not depend on the response values but only on the DOE and prediction location. Since we used equally maximum minimum distance criterion for the DOE, results may not be as interesting as for clustered DOE. Besides, the local measure of error seems to actually detect the regions of unconservativeness (hence a better reduction of the large unconservative errors), but this is counterbalanced by

assessing large uncertainties to regions where prediction is accurate, which leads to large overestimations that penalize the global measure of accuracy.
The equivalence of safety margin and error distribution is interesting since the challenge with empirical estimators is to choose the appropriate margin. Thus, one can use the statistical information to accurately choose the margin corresponding to a certain level of conservativeness.

Appendix 1: Trade-off curve of a single kriging modeling of the Branin-Hoo function

We show here the results of a typical kriging model of the Branin-Hoo function. DOE and test points are chosen as described in section III-1. The DOE and unbiased kriging prediction are represented in Figure 22. For this model, the RMSE is 8.1 and the percentage of conservative predictions is 49.1%.

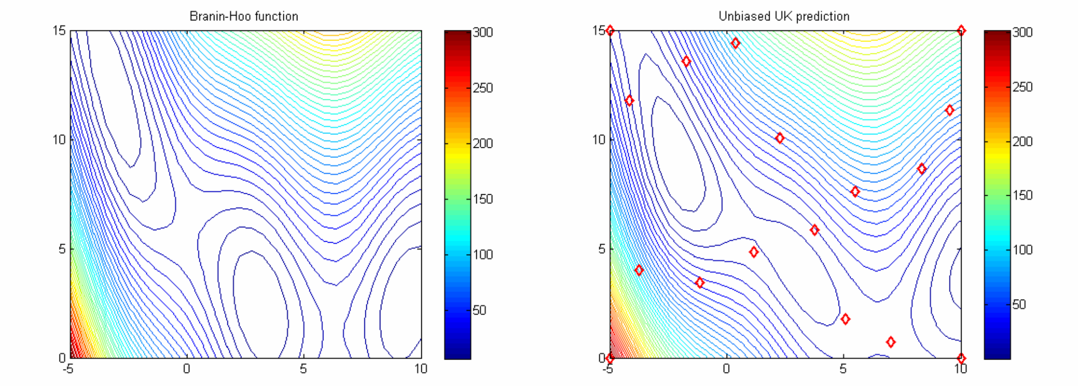

Figure 22: Branin-Hoo function (left) , DOE and unbiased kriging prediction (right).

We choose five values for the CSM: [0, 1.5, 4, 6, 12] and five levels for the confidence intervals (1-α): [0.5, 0.6, 0.7, 0.8, 0.9]. For each, we compute the different indices as described in section II-1. The results are reported in Table 3 and Table 4. In Figure 23, we represent for the two estimators the curves of the percentage of conservative predictions vs. the percentage increase of RMSE, and the percentage of conservative predictions vs. the percentage reduction of MaxUE.

Table 3: Results for the CSM estimator.

CSM	% cons. predictions	RMSE	MaxUE	$RMSE_{norm}$	$maxUE_{norm}$
0	49.1	8.1	15.5	0	0
1.5	63.5	8.4	14.0	3.2	9.7
4	76.6	9.3	11.5	15.0	25.9
6	82.2	10.5	9.5	29.1	38.8
12	95.8	15.0	3.5	85.3	77.7

Table 4: Results for the ED estimator.

(1 - α)	% cons. predictions	RMSE	MaxUE	$RMSE_{norm}$	$maxUE_{norm}$
0.5	49.1	8.1	15.5	0	0
0.6	62.3	9.1	12.0	12.7	22.1
0.7	72.8	11.1	8.5	37.1	45.1
0.8	82.1	14.1	4.5	73.7	70.6
0.9	95.8	18.7	1.7	131.3	88.9

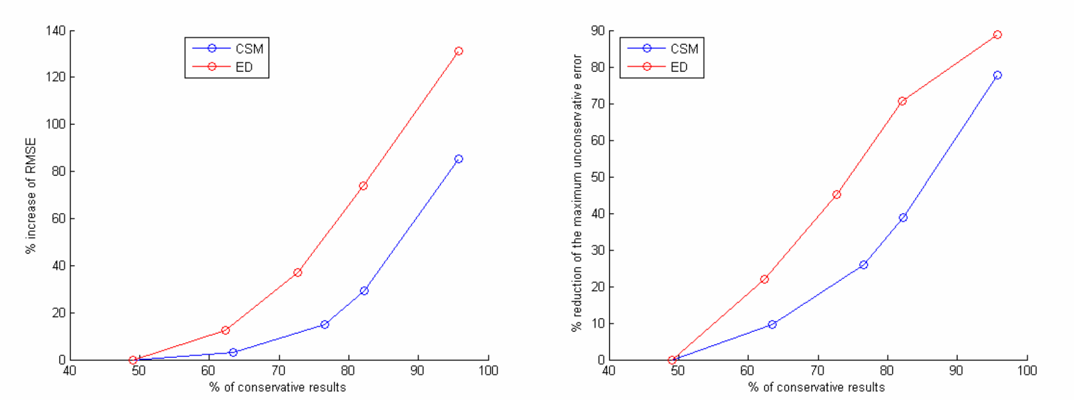

Figure 23: Results for CSM and ED estimators.

Here, we see that for an equivalent proportion of conservative results, the RMSE is more increased when using the ED estimator than with the CSM estimator. On the other hand, the maximum unconservative error is better reduced with the ED estimator.

Appendix 2: Error analysis of a single kriging modeling of the Branin-Hoo function

Here, we analyze the spatial error distribution of the kriging conservative predictors for the Branin-Hoo function. We compare the errors for three models: the unbiased kriging, CSM kriging with a margin of 5.97 and ED kriging with a confidence of 80%. For these two conservative estimators, the percentage of conservative prediction is the same, equal to 82.1%. Their indices are reported in Table 5.

Table 5: Results for unbiased, CSM and ED kriging estimators.

	% cons. predictions	**RMSE**	**MaxUE**	**RMSE$_{norm}$**	**MaxUE$_{norm}$**
Unbiased	49.1	8.1	15.5	0	0
CSM = 5.97	82.1	10.4	9.5	28.9	38.6
(1 - α) = 0.8	82.1	14.1	4.5	73.7	70.6

We see that, for an equivalent level of conservativeness, the RMSE is higher for the ED estimator, but the reduction of maximum unconservative error is higher. In order to explain this in detail, we draw the error contour plots over the design region and the histograms of the 1024 errors at data points.

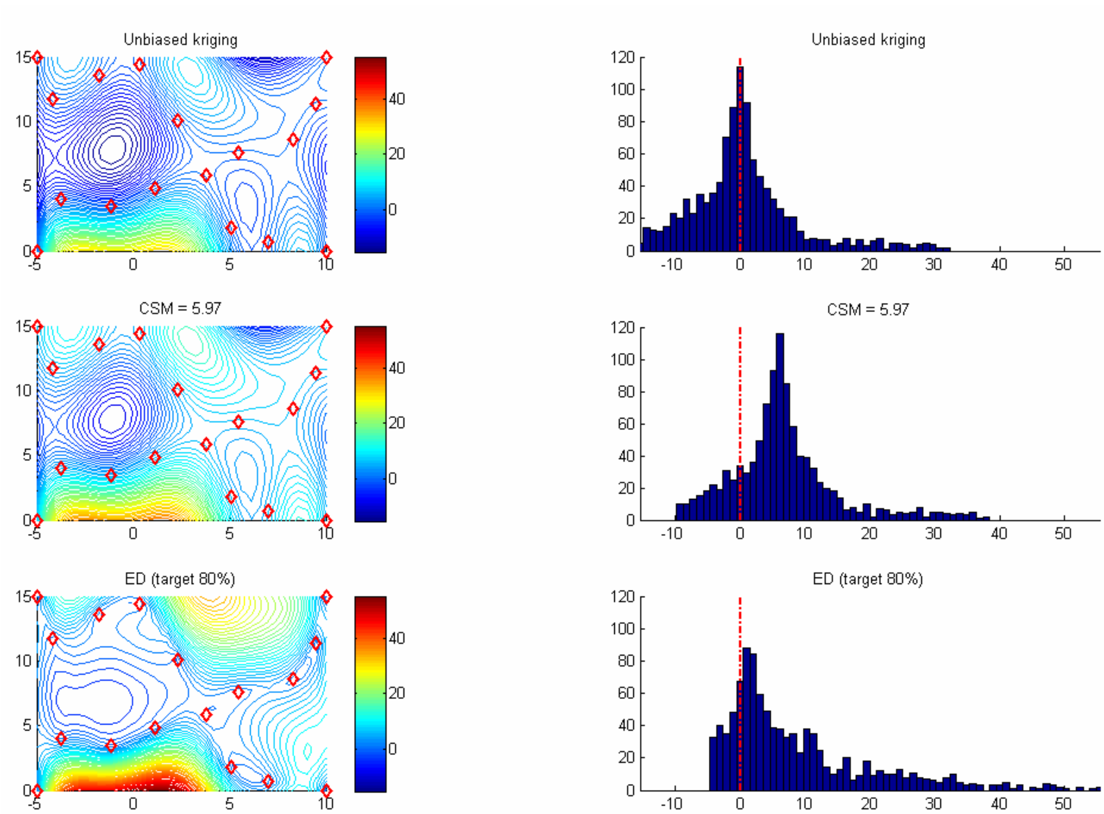

Figure 24: Error contour plots and histograms of errors at data points for unbiased, CSM and ED kriging estimators.

On the top left figure, we see that, since kriging is an interpolation, it leads to regions of small errors (pale blue) that are around the training points, and large errors far from the training points (yellow, at the bottom left, and dark blue, center left and top right). Thus, the histogram (top right figure) shows a pick around zero, and large tails on both sides.

When using a CSM (middle figures), all the errors are switched up by the constant; hence, the histogram has the exact same shape but is moved toward the right. Since the pick of small errors is not large, a small margin is enough to move it to the positive values, and increase a lot the percentage of conservative predictions with a small effect on the RMSE. However, if the maximum unconservative error is large compared to the margin, it will not be reduced substantially.

The ED estimator increases the prediction values in the regions of high uncertainty: center left, top right, bottom left. The two first correspond to the region of high unconservativeness. Thus, the unconservative errors are reduced. When comparing the CSM and ED histograms, we see that the distribution of the unconservative errors (the part of the histogram on the left of the red line) is very different: both correspond to 18% of the total errors, but most errors are close to zero for ED while the tail is heavier for CSM. However, although the ED estimator captures the region of unconservativeness, it also leads to very large overestimations, in particular in the bottom left. As a consequence, the distribution of errors has a very long right tail that affects the RMSE value.

In section II-1, we saw that this behavior is amplified when the number of points is increased. In Table 6 and Figure 25, we present the results for a 34-point DOE.

Table 6: Results for unbiased, CSM and ED kriging estimators based on a 34-point DOE.

	% cons. predictions	RMSE	MaxUE	$RMSE_{norm}$	$MaxUE_{norm}$
Unbiased	57.1	0.0502	0.296	0	0

CSM = 0.0135	76.4	0.0518	0.282	3.2	4.6
(1 - α) = 0.8	76.2	0.0737	0.1713	46.8	42.0

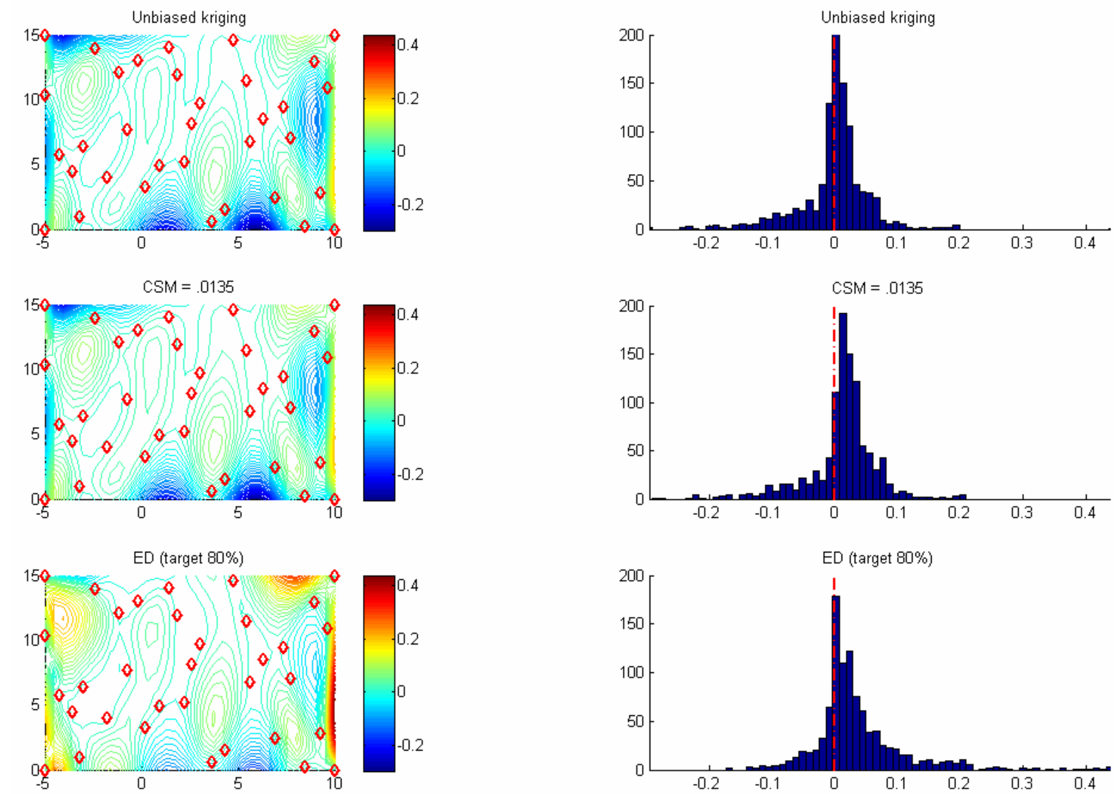

Figure 25: Error contour plots and histograms of errors at data points for unbiased, CSM and ED kriging estimators based on a 34-point DOE.

For the unbiased estimator, most of the errors are very small, but the errors are locally large in regions where there are not training points. Hence, the histogram has the form of a thin pick centered on zero, with a couple of large values on both sides.

On the second histogram, we see that a very small margin is sufficient to move the entire pick to the positive side. As a consequence, the proportion of conservative estimations is increased to 76.4% with almost no effect on the RMSE. On the other hand, the maximum unconservative error remains almost the same.

On the third histogram, the pick is also moved to the right, but the maximum unconservative error is also reduced by 42%. On the other hand, large positive errors have appeared that affect the RMSE: on the right edge and top right of the contour graph.

Appendix 3: Error analysis of the statistical-based estimators for the Torque Arm analysis

Here, we report the histograms of the residuals for kriging, PRS and IK estimators. The DOE is the original DOE generated; the residuals are computed at the 1000 data points. Table 7 and Table 8 show the results for the unbiased estimators and the conservative estimators with a target conservativeness level of 95%, respectively.

Table 7: Statistics for unbiased estimators.

Surrogate	Pcons	RMSE	MaxUE
PRS	49.7	0.0589	0.4956
Kriging	47.8	0.0415	0.3106
IK	49.8	0.1411	0.6049

Table 8: Statistics for a 90% target conservativeness estimators.

Surrogate	Pcons	RMSE	MaxUE
PRS	82.7	0.1059	0.4011
Kriging	95.1	0.0732	0.2272
IK	90.9	0.2908	0.3434

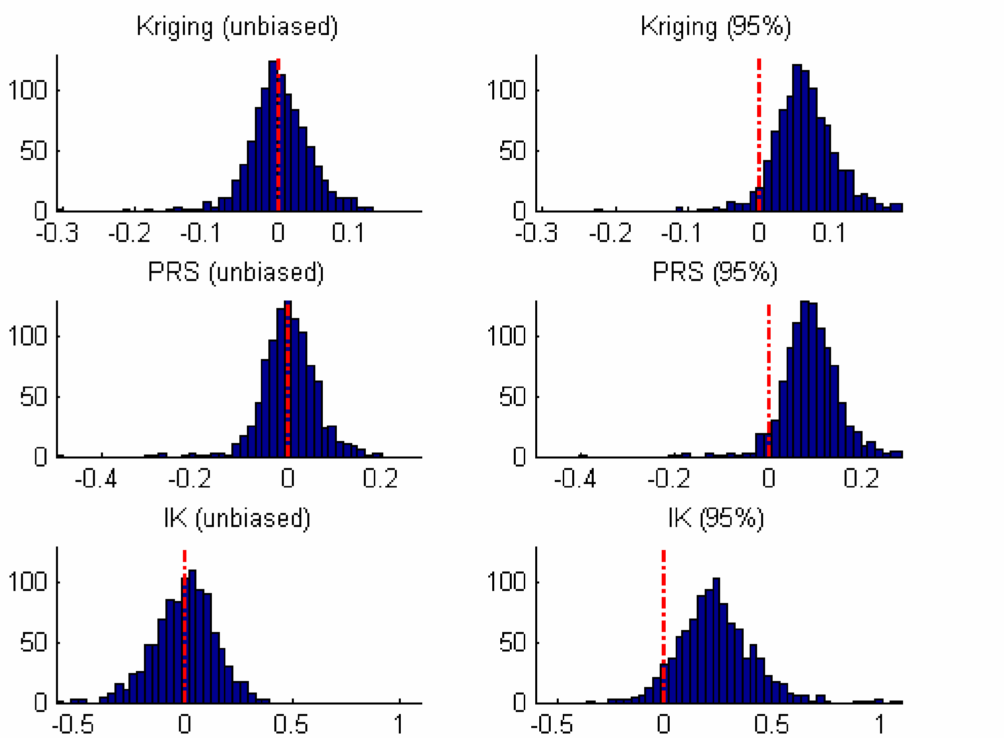

Figure 26: Histograms of residuals for the statistical-based estimators, unbiased and 95% target conservativeness.

For the unbiased estimators, kriging and PRS show the presence of a few outliers (5-6), while IK does not. These large unconservative values are not significantly decreased for the 95%. Comparatively, IK reduces a lot more the MaxUE. However, we see on the bottom right histogram that there are a few large overestimations.

Acknowledgments

Funding for this research was provided by National Science Foundation (Grant # 0423280). The authors greatly acknowledge this support.

References

[1] Queipo, N. V.,Haftka, R. T., Shyy, W., Goel, T., Vaidyanathan, R., Tucker, K., Surrogate-based Analysis and Optimization, Progress in Aerospace Sciences Vol. 41, pp. 1-28, 2005

[2] Acar, E., Kale, A., and Haftka, R.T., "Comparing Effectiveness of Measures that Improve Aircraft Structural Safety," in press, ASCE Journal of Aerospace Engineering, 2007.

[3] Acar, E., Haftka, R.T. and Johnson, T.F., "Tradeoff of Uncertainty Reduction Mechanisms for Reducing Structural Weight," ASME Journal of Mechanical Design, Vol. 129, No. 3, 2007, pp. 266-274.

[4] 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, 2006, AIAA-2006-7038

[5] Starnes, J.H., and Haftka, R.T. 'Preliminary Design of Composite Wings for Buckling, Stress and Displacement Constraints', Journal of Aircraft, vol. 16, 1979, pp 564-570.

[6] J. Sacks, W.J. Welch, T.J. Mitchell, H.P. Wynn: 'Design and Analysis of Computer Experiments', Statistical Science, Vol. 4, No. 4 (Nov. 1989), 409-423

[7] E.H. Isaaks, R.M. Srivastava, 'An Introduction to applied geostatistics', Oxford University Press, 1989

[8] C.V. Deutsch and A.G. Journel: 'GSLIB: geostatistical software library and user's guide', 1998, New-York, Oxford University press

[9] Bennett, J. A., and Botkin, M. E., The optimum shape, Plenum Press, New York, N.Y., 1986.

[10] J. Emery, J.M. Ortiz: 'Shortcomings of multiple indicator kriging for assessing local distributions', Applied Earth Science (Trans. Inst. Min. Metall. B), December 2004, Vol.133