Building Efficient Response Surfaces of Aerodynamic Functions with Kriging and Cokriging

J. Laurenceau *
CERFACS, Toulouse, 31057, France

P. Sagaut †
F-75252 Paris cedex 5, France

January 24, 2008

Abstract

In this paper, we compare the global accuracy of different strategies to build response surfaces by varying sampling methods and modeling techniques. The final application of the response surfaces being aerodynamic shape optimization, the test cases are issued from CFD simulations of aerodynamic coefficients. For comparisons, a robust strategy for model fit using a new efficient initialization technique followed by a gradient optimization was applied. Firstly, a study of different sampling methods proves that including 'a posteriori' information on the function to sample distribution can improve accuracy over classical space filling methods like Latin Hypercube Sampling. Secondly, comparing Kriging and gradient enhanced Kriging on two to six dimensional test cases shows that interpolating gradient vectors drastically improves response surface accuracy. Although direct and indirect Cokriging have equivalent formulations, the indirect Cokriging outperforms the direct approach. The slow linear phase of error convergence when increasing sample size is not avoided by Cokriging. Thus, the number of samples needed to have a globally accurate surface stays generally out of reach for problems considering more than four design variables.

Nomenclature

$CV(x)$	Leave one out cross validation error
$\mathcal{D}$	Domain of design variables
$dKdS(x)$	Leave one out sensitivity error
$f(x)$	Regression vector $[n_p]$
F	Regression matrix $[N \times n_p]$
L	Model likelihood estimate
N	Order of the correlation matrix
n_{dv}	Number of design variables
n_p	Dimension of regression vectors
n_s	Number of samples
p	SCF power coefficients $[n_{dv}]$
R	Correlation matrix $[N \times N]$

*Ph.D. Student, CFD, 42 Avenue Coriolis.
†Professor, Institut Jean Le Rond d'Alembert, Université Pierre et Marie Curie, 4 place Jussieu.

$r(x)$	Correlation vector $[N]$
$\mathcal{S}$	Domain of sample points
$SCF(.,.)$	Spatial correlation function
$S(x)$	Standard error
s^i	i-th sample
x	Vector of design variables $[n_{dv}]$
$Y_{exa}(x)$	Exact function to approximate
Y_s	Exact function at samples $[N]$
$Y(x)$	Surrogate function
$Z(x)$	Gaussian process
β	Regression coefficients $[n_p]$
σ^2	Model variance
θ	SCF correlation coefficients $[n_{dv}]$

Subscripts

k	$\in [1, n_{dv}]$
l	$\in [1, n_{dv}]$
m	$\in [1, n_{dv}]$

Superscripts

i	$\in [1, n_s]$
$-i$	i-th sample left out for model build
j	$\in [1, n_s]$

1 Introduction

NUMERICAL shape optimization in the field of detailed aircraft design is more and more used since high performance computers enable RANS flow simulation around complete aircraft in a short time scale. Before building a complex optimization process around a CFD suite considering a lot of design variables, some time must be spent to identify an efficient optimizer.

When dealing with computer expensive simulation, the main requirement is to converge to an optimal shape in as few simulations as possible excluding most of the global optimizers such as Genetic Algorithm or Simulated Annealing that are generally applied to conceptual or preliminary design. Assuming that the sensitivity of the functions with respect to the design variables is available at a comparative computational cost of a direct function evaluation (with an adjoint method), the quickest process is then gradient based. Using this strategy, the high-fidelity aerodynamic shape optimization suite *OPTaliA* of *Airbus* showed good results [1]. Despite their efficiency, gradient algorithms suffer from some limitations. They solve only single objective problems (multiple objective functions must be combined) and are easily blocked by local optima whereas aerodynamic functions are generally multi-modal. Moreover, heavily constrained problems are hard to solve efficiently.

Promising alternative approaches use response surfaces (or surrogate models) to approximate the expensive computer code on the whole domain by inexpensive models. Numerous strategy exist [2] and have been applied to perform aerodynamic shape optimization [3, 4, 5]. Firstly, response surfaces can lead the optimization process. The surrogate management framework [6] and trust-region framework [7] are generally the most favored methods, but refining the response surface at its predicted optimum is also efficient particularly when taking into account uncertainty (Expected Improvement or Lower Confidence Bounding Function criteria) [8, 9, 10, 4, 5]. Response surfaces can be reliable discipline independent models applied to solve Multi-Disciplinary Optimization (MDO) problems when no code coupling is possible [2]. When the model is sufficiently accurate, the predicted optimum on the response surface can be taken as the true optimum [11, 12]. Secondly, the convergence of existing optimizers can be improved by switching between optimization on expensive model and on response surface (variable fidelity) [3]. Thirdly, the exploration capabilities of local optimizers can be improved by using the optimum found on the response surface as initialization. Accordingly to these different applications, two concepts of response surfaces are defined. The global response surface is considered an accurate surrogate model of the true function on the whole domain, whereas the local response surface is reliable only near sampled locations but gives information on the overall tendencies.
In this study we wanted to define how to build accurate response surfaces and assess their domain of validity before any applications to shape optimization. Kriging model was chosen to approximate the objective function for its accuracy and robustness with small datasets [13]. An efficient methodology is detailed within this article including numerical issues and a new technique to fit the model (section 2).
The sampling quality being essential to obtain an accurate model [14], section 4 compares different distribution methods. For a given domain of study, classical sampling are space filling and locations are computed without knowing the function 'a priori'. Beside other modeling techniques, the uncertainty predictions available with Kriging are very useful. Thus, methods of 'a posteriori' sequential sampling based not solely on information on the domain but also on information on the function through predicted errors are also described. In addition to the standard error of Kriging and leave-one-out cross validation errors, two new predicted error criteria are introduced (section 2-2.2-2.2.2) and tested (section 4-4.2). Methods are compared by varying sample size for the reconstruction of a bidimensional aerodynamic test function. 'A posteriori' samplings were already applied to polynomial test cases [10, 15] without in-depth comparison with classical methods. As 'a posteriori' samplings are tuned to the function studied, their performance is closely related to the test case. That is why it was necessary to perform the comparison directly on an aerodynamic function. More than a sampling criterion, these predicted error are very important when using the surrogate model as optimizer in order to have a balancing between exploration of unknown locations and exploration of predicted optima on the model.
As a very powerful gradient computation method was available (adjoint state solver [16, 17, 1]), it was also interesting to evaluate Cokriging (gradient enhanced) model. General comparison between Kriging and Cokriging existed only on analytic functions [18, 19]. In the fields of aerodynamic, it was proven that refining iteratively at the predicted minimum of a Cokriging response surface improved the shape of a supersonic business jet depending on up to 15 design variables [20, 21], but the gradient information was issued from finite differences analysis and no comparison with optimization on a Kriging-based response

surface was made. A graphical validation of Cokriging versus Kriging was made on the same aerodynamic test case but considering only two design variables and small datasets [21]. The benefits of Cokriging on global response surfaces accuracy cannot be assessed by only varying the sample size as the information added to the model through gradients is dimensional. Section 5 compares Cokriging and Kriging accuracy on an aerodynamic test case, the number of design variables increasing from two to six. A total of 5800 CFD computations were necessary to build reference surfaces.

2 Kriging and Cokriging for Computer Experiments

2.1 Sampling

A surrogate model only needs a sample distribution to approximate a function, thus it must concentrate as much information as possible. In this way, the sampling must be adapted to : known specificities of the function, modeling formulation and the final application of the surrogate. Without information on the function, design optimal distribution methods using for example entropy criterion [22] (or DETMAX) give space filling locations. Numerous methods exist to evenly distribute samples.

2.1.1 'A priori' Sampling Methods

Four 'a priori' methods requiring only information on the domain and assuming that the function is unknown a priori were used: grid, Halton, Sobol and Latin Hypercube (LHS) sampling (source code of J. Burkardt `http://www.scs.fsu.edu/~burkardt/`).
The grid distribution is the most common in engineering and was thus included in the benchmark. However, it should be used with great cares as truncated grids are to be avoided and grids in high dimension require unreachable sample sizes. The LHS method is also well known. It roughly corresponds to random sampling with guaranteed space filling properties even for small sampling size. Among the four methods presented here only LHS is applicable to high dimensions.
Quasi-random sequences enable methods such as generalized Halton or Sobol [23] to give ordered space filling sampling. These methods are very efficient on low-dimensionnal problems but unapplicable to dimensions higher than 10 without special care. Halton and Sobol are sequential methods. A distribution of n_s+1 samples inherits from the previously computed n_s samples. Latin Hypercube and grid sampling are not sequential methods.

2.1.2 'A posteriori' Sequential Sampling Methods

Maximum error sequential sampling includes information on the true function in sample distribution, explaining the denomination of 'a posteriori' sampling.
It consists in iteratively refining the sample datasets where the model exhibits its maximum of error. By construction it ensures inheritance of samples. Its main drawback is that all intermediate models must be computed. To compute a sampling of a given size the CFD simulations must be launched one at a time, whereas with 'a priori' methods the n_s computations can be launched in parallel. The only requirement is the availability of an error criterion null at already sampled locations. 'A posteriori' samplings are tuned to the function studied.

2.2 Kriging

The French mathematician Georges Matheron developed in the early sixties the theory of Kriging from the seminal work of D.G. Krige on mining data. In the field of computer experiments, the Kriging method refers to the DACE formulation [22] (fitted by optimization of the likelihood estimate instead of using covariogram) described in the late eighties.

2.2.1 Function Prediction

Kriging method is a statistical prediction of a function at untried inputs. It requires fitting the correlation parameters of the model to each sample distribution by solving an optimization problem.
Since computer codes are deterministic, and therefore not subject to measurement error, the usual measures of uncertainty derived from least-square residuals have no obvious meaning. Consequently, statisticians have suggested approximating responses as a combination of a polynomial regression plus localized deviations interpolating samples

$$Y(x) = f^t(x)\beta + Z(x) \approx Y_{exa}(x) \tag{1}$$

$f^t(x)\beta$ is a low-order (constant, linear or quadratic) polynomial regression model. Solving a least-square regression problem gives

$$\beta = (F^t R^{-1} F)^{-1} F^t R^{-1} Y_s \tag{2}$$

In this paper a constant regression model is used, so $n_p = 1$, $f(x) = 1$ and $F = 1$.
$Z(x)$ is the realization of a normally distributed random process with mean zero, variance σ^2,

$$\sigma^2 = \frac{1}{n_s}(Y_s - F\beta)^t R^{-1}(Y_s - F\beta) \tag{3}$$

and co-variance

$$cov[Z(s^i), Z(s^j)] = \sigma^2 R_{ij} \tag{4}$$

The matrix of correlation between samples R is determined by a Spatial Correlation Function (SCF).

$$R_{ij} = SCF(s^i, s^j) = \prod_k scf_k(|s_k^i - s_k^j|) \tag{5}$$

This matrix is dense symmetric positive definite with ones along diagonal and become ill-conditioned when samples are too close. The order of the Kriging correlation matrix depends only on the number of samples considered ($N = n_s$) and not on the number of variables.
The SCF can be any function reflecting the characteristics of the output function. Exponentials or cubic splines are the most common functions. The exponential function is adapted to a wide range of physical applications and is described below

$$scf_k(x) = \exp\big(-\theta_k x^{p_k}\big), \theta_k > 0, p_k \in [1, 2] \tag{6}$$

The $2n_{dv}$ correlation parameters (θ_k, p_k) need to be fitted through an optimization process (see section 2-2.5-2.3.1)

The cubic spline function was also used

$$scf_k(x) = \begin{cases} 1 - 6(x\theta_k)^2 + 6(x\theta_k)^3 \,,\, x < \frac{1}{2\theta_k} \\ 2(1 - x\theta_k)^3 \,,\, \frac{1}{2\theta_k} \leq x < \frac{1}{\theta_k} \\ 0 \,,\, x \geq \frac{1}{\theta_k} \end{cases} \tag{7}$$

Only n_{dv} correlation parameters $\theta_k > 0$ must be fitted. Commonly, cubic splines are less adapted to very chaotic functions than exponentials.
To evaluate Kriging at an unknown location a vector of correlation r between sample points and unknown is computed

$$r_i(x) = SCF(x, s^i) \tag{8}$$

Finally, the Kriging prediction equation is

$$Y(x) = f^t(x)\beta + r^t(x)R^{-1}(Y_s - F\beta) \tag{9}$$

Practically, the Kriging model can be evaluated as a black-box function by storing vectors β and $rhs = R^{-1}(Y_s - F\beta)$. Then, the evaluation of the model at an unknown location costs only two dot products.

2.2.2 Error Prediction

With the statistical prediction of the function is also associated an uncertainty prediction. As by construction the model interpolates the function at samples, its error is null at samples.
The standard error $S(x)$ of the Kriging model grows with distance from samples, its maximum value being σ.

$$\begin{aligned} S^2(x) = \quad & \sigma^2 \Big(1 - r^t(x)R^{-1}r(x) \\ & + u^t(x)(F^tR^{-1}F)^{-1}u(x) \Big) \end{aligned} \tag{10}$$

with $u(x)$ vector $n_p \times 1$

$$u(x) = F^tR^{-t}r(x) - f(x) \tag{11}$$

Other estimated errors can be computed with a technique of leave-one-out cross validation [9]. The idea behind cross validation errors is to study the sensitivity of the model with respect to sample dataset by altering it of one point. It implies n_s intermediate Kriging constructions and evaluations. Two predicted errors based on leave-one-out cross validation named sample sensitivity error ($dKdS(x)$) and cross validation error ($CV(x)$) are defined

$$dKdS(x) = \frac{1}{n_s} \sum_i |Y(x) - Y^{-i}(x)| \tag{12}$$

$$CV(x) = \frac{1}{n_s} \sum_i S^{-i}(x) \tag{13}$$

Kriging sensitivity error $dKdS(x)$ grows at locations where sample points give more information and is small at location where samples are redundant. Kriging standard sensitivity

error $CV(x)$ is very close to the standard error. As cross validation errors are a mean of models built on altered sample datasets, they are not null at samples.
We decided to make a geometric mean between cross validation and standard errors to set them null at samples. That is why errors $dKdS(x)$ and $CV(x)$ are mixed with the standard error, and two new error predictions are introduced

$$dKdS_{mix}(x) = \sqrt{S(x)\, dKdS(x)} \tag{14}$$

$$CV_{mix}(x) = \sqrt{S(x)\, CV(x)} \tag{15}$$

Predicted errors are very useful to improve iterative process. When minimizing a function, iteratively refining at the minimum of the response surface usually leads to premature convergence [6]. This can be improved by including exploration in the process using either a domain exploration ('Poll' step of the surrogate management framework [6]), either an infill cirterion based on predicted error [10]. When minimizing global error of a response surface, predicted errors can be used as sampling criteria (see MAXVAR [10], Application-driven Sequential Designs [15] and 4-4.2).

2.3 Model Correlation Parameters Fit

Kriging spatial correlation parameters (θ_k, p_k) from equations 6 or 7 must always be fitted to the samples. Otherwise, if the correlation strength is underestimated (θ_k too large), the local deviation function $Z(x)$ looks like a peak function interpolating each sample and only the polynomial regression model gives information.

$$for\ \theta \to +\infty, \begin{cases} Z(x) \to 0 \\ Y(x) \to f^t(x)\beta \end{cases} \tag{16}$$

2.3.1 Maximization of Model Likelihood

The correlation parameters (θ_k, p_k) of the model are fitted by Maximizing its Likelihood Estimate (MLE method).

$$MLE = max_{\theta,p}\big(\ln(L)\big) \tag{17}$$

$$\begin{gathered} \ln(L(\theta,p)) = \\ -\tfrac{1}{2}\Big(n_s\big(\ln(\sigma^2) + \ln(2\pi) + 1\big) + \ln(|R|)\Big) \end{gathered} \tag{18}$$

The model likelihood L evaluation is usually very quick compared to a high-fidelity CFD computation, it only costs to compute the determinant of the correlation matrix R. Maximization of the likelihood estimate can sometimes lead to an incorrect model identifiable by high standard error or by comparing MLE to its asymptotical value [24]. This is due to the nature of the MLE function: quasi-null gradient on some zones, and multiple local optima [25]. This sometimes high-dimensional problem (from n_{dv} to $2n_{dv}$ parameters according to the correlation function) can be efficiently solved using a gradient optimization algorithm. By using a correct initialization of the correlation parameters as described in the next section (2-2.3-2.3.2), the optimizer converges to the global optimum in most cases. The gradient of the determinant of the correlation matrix is computed with a forward finite differences method.

2.3.2 Initialization of the Spatial Correlation Parameters

Estimating the Likelihood requires inverting the correlation matrix and for large datasets, as it is the case in 5, the MLE problem becomes computer expensive. A method was developed to easily find an initial guess and the optimization boundaries for the correlation parameters. It enables to optimize the correlation parameters with a gradient optimizer (local) requiring few evaluations of the Likelihood. The parameters θ_k govern the directional strength of the correlation between two locations. The correlation decreases with distance between points. The log-Likelihood function is multimodal and has quasi null gradients for large θ_k [25]. The main idea is to define what a large correlation parameter (θ_k) is by introducing a scaling factor depending on the samples. It is then possible to overestimate the correlation strength with a small θ_k in order to directly study the region of the log-Likelihood exhibiting large gradients.
Considering the exponential SCF (Eqn. 6), p_k values close to 2 are best suited to smooth functions but less numerically robust that is why the initial value is $p_k = 1$.
First, assuming that all control points are correlated, the minimum correlation factor c is defined as

$$\begin{aligned} &\forall (x^i, x^j) \in \mathcal{D}^2,\ \exists\, c \in]0;1[\ / \\ &\exp\big(-\theta_k |x_k^i - x_k^j|\big)\ \geq\ c \end{aligned} \tag{19}$$

Thus, by fixing the strength of the correlation c between the two most directional-distant sample points $(s_k^{\tilde{i}}, s_k^{\tilde{j}})$

$$\begin{aligned} &\exists \theta_k > 0\ / \\ &\exp\big(-\theta_k |s_k^{\tilde{i}} - s_k^{\tilde{j}}|\big) = c \end{aligned} \tag{20}$$

Finally giving

$$\theta_k = \frac{-\ln(c)}{|s_k^{\tilde{i}} - s_k^{\tilde{j}}|} \tag{21}$$

According to our physical functions a good correlation factor is $c = 0.2$. Experience shows that underestimated θ_k parameters usually give better approximation than overestimated parameters, that is why a strong correlation factor was chosen. The initial parameters are determined by

$$\theta_k = \frac{\ln(5)}{|s_k^{\tilde{i}} - s_k^{\tilde{j}}|} \tag{22}$$

The optimization boundaries are $[10^{-3}\theta_k,\ 10^{+3}\theta_k]$ around the initial guess.
The same method can be applied to cubic spline functions.

2.4 Cokriging - Gradient Enhanced Kriging

A Cokriging model [19] interpolates the function and the gradient at each sample location. As Cokriging models include more information on the true function than Kriging models, they need fewer samples to achieve a given level of accuracy. Moreover, the vectorial information provided by the gradient should be more beneficial for high-dimensional problems.
Two different formulations exist for Cokriging: direct or indirect. When using the same correlation parameters both formulations give the same results.

2.4.1 Indirect Cokriging

The indirect Cokriging method does not require Kriging source code modification. In fact, the original Kriging formulation is used with an augmented sample database. For each sample point, one point determined by a first order Taylor approximation is added in each direction. The size of the augmented sample set is $n_s(n_{dv}+1)$. The step chosen to add those points must not be too low in order to avoid ill-conditioning of the correlation matrix. A good compromise is $s^{n_s+ik} = s^i + 10^{-4} range_k(\mathcal{S})$.

$$y(s^{n_s+ik}) = y(s^i) + \frac{\partial y(s^i)}{\partial x_k} 10^{-4} range_k(\mathcal{S}) \tag{23}$$

The indirect Cokriging applied to very dense datasets is expected to be problematic. If the delta used for the Taylor expansion is of the same order of magnitude than point density, the Taylor augmented datasets could degrade the model. Hopefully, it does not happen when approximating expensive computer codes.

2.4.2 Direct Cokriging

The direct Cokriging approach is equivalent to make one Kriging approximation for each direction of the gradient, and the link between function and gradient is conserved through an heritage of the correlation function and parameters. The vectors and matrix of size n_s are augmented to a size of $n_s(n_{dv}+1)$.

$$\begin{aligned} Y_s = \quad & \Big[y(s^1), y(s^2), \cdots, y(s^{n_s}), \\ & \frac{\partial y(s^1)}{\partial x_1}, \frac{\partial y(s^1)}{\partial x_2}, \cdots, \frac{\partial y(s^1)}{\partial x_{n_{dv}}}, \\ & \frac{\partial y(s^2)}{\partial x_1}, \cdots, \frac{\partial y(s^{n_s})}{\partial x_{n_{dv}}} \Big] \end{aligned} \tag{24}$$

For a constant regression model

$$\begin{aligned} f(x) = \quad & \Big[1, 1, \cdots, 1, \\ & 0, 0, \cdots, 0, \\ & 0, \cdots, 0 \Big] \end{aligned} \tag{25}$$

The augmented correlation matrix is assembled through 4 blocks, representing covariance of functions between functions, of functions between gradients, of gradients between functions and of gradients between gradients. Whereas the indirect formulation approximates the function on an augmented database, the direct formulation keeps the original database unchanged but approximates more functions $(n_{dv}+1)$.

$$\begin{cases} Cov\big(y(s^i), y(s^j)\big) = \sigma^2 SCF(s^i, s^j) \\ Cov\big(y(s^i), \dfrac{\partial y(s^j)}{\partial x_k}\big) = \sigma^2 \dfrac{\partial SCF(s^i, s^j)}{\partial x_k} \\ Cov\big(\dfrac{\partial y(s^i)}{\partial x_k}, y(s^j)\big) = -\sigma^2 \dfrac{\partial SCF(s^i, s^j)}{\partial x_k} \\ Cov\big(\dfrac{\partial y(s^i)}{\partial x_k}, \dfrac{\partial y(s^j)}{\partial x_l}\big) = \sigma^2 \dfrac{\partial^2 SCF(s^i, s^j)}{\partial x_k \partial x_l} \end{cases} \tag{26}$$

One should notice that the augmented correlation matrix is dense non-symmetric and its order $N = n_s(n_{dv}+1)$ increases with the number of variables and the number of samples. The 2 blocks of the new correlation vector are

$$r_i(x) = \begin{cases} SCF(x, s^i) \ , \ [1, n_s] \\ \dfrac{\partial SCF(x, s^i)}{\partial x_k} \ , \ [n_s+1, n_s(n_{dv}+1)] \end{cases} \tag{27}$$

With the exponential function the distance (L1-norm) is elevated to a power lesser than two and is not twice differentiable. That is why cubic spline correlation functions are used with direct Cokriging. Derivatives of the cubic spline (Eqn. 7) are

$$\frac{\partial SCF(s^i, s^j)}{\partial x_k} = \begin{cases} sign(s_k^i - s_k^j) \\ \times \big(18\,\theta_k^3 |s_k^i - s_k^j|^2 - 12\,\theta_k^2 |s_k^i - s_k^j|\big) \\ \times \prod_{m \neq k} \Big[1 - 6(|s_m^i - s_m^j|\theta_m)^2 \\ +6(|s_m^i - s_m^j|\theta_m)^3\Big] \\ \\ -6\,\theta_k sign(s_k^i - s_k^j) \\ \times (1 - \theta_k |s_k^i - s_k^j|)^2 \\ \times \prod_{m \neq k} \Big[2(1 - 6|s_m^i - s_m^j|\theta_m)^3\Big] \\ \\ 0 \end{cases} \tag{28}$$

Assuming $k \neq l$,

$$\frac{\partial^2 SCF(s^i, s^j)}{\partial x_k \partial x_l} = \begin{cases} \prod_{m \in \{k,l\}} \Big[sign(s_m^i - s_m^j) \\ \times \big(18\,\theta_m^3 |s_m^i - s_m^j|^2 \\ -12\,\theta_m^2 |s_m^i - s_m^j|\big)\Big] \\ \times \prod_{m \notin \{k,l\}} \Big[1 - 6\big(|s_m^i - s_m^j|\theta_m)^2 \\ +6(|s_m^i - s_m^j|\theta_m\big)^3\Big] \\ \\ \prod_{m \in \{k,l\}} \Big[-6\,\theta_m^2 sign(s_m^i - s_m^j) \\ \times \big(1 - \theta_m |s_m^i - s_m^j|\big)^2\Big] \\ \times \prod_{m \notin \{k,l\}} \Big[2(1 - 6|s_m^i - s_m^j|\theta_m)^3\Big] \\ \\ 0 \end{cases} \tag{29}$$

$$\frac{\partial^2 SCF(s^i, s^j)}{\partial x_k^2} = \begin{cases} \big(-12\,\theta_k^2 + 36\,\theta_k^3 |s_k^i - s_k^j| \big) \\ \times \prod_{m \neq k} \Big[1 - 6(|s_m^i - s_m^j| \theta_m)^2 \\ +6(|s_m^i - s_m^j| \theta_m)^3 \Big] \\ \\ 12\,\theta_k^2 \big(1 - \theta_k |s_k^i - s_k^j| \big) \\ \prod_{m \neq k} \Big[2(1 - 6|s_m^i - s_m^j| \theta_m)^3 \Big] \\ \\ 0 \end{cases} \tag{30}$$

2.4.3 From Kriging to Cokriging

All Kriging functionalities are available when using Cokriging including error predictions. Cokriging models are also fitted by MLE (Eqn. 17). The MLE correlation parameters of Kriging, indirect Cokriging and direct Cokriging are different in general.

2.5 Numeric Implementation

2.5.1 Correlation Matrix Inversion and Robustness

Considering SCF formulation, close sample points can lead to quasi-identical columns in the correlation matrix R the matrix being then ill-conditioned [26] (especially with low θ_k values). The optimization of Kriging parameters implies computing very small determinants, and in order to avoid underflow truncature error the $\ln |R|$ term is computed as a sum of logarithms of eigenvalues with a LU method from LAPACK [27].
Typically, one can obtain a fitted Kriging with $cond(R) \approx 10^{16}$ and still a good interpolation of samples. That is why ill-conditioned matrix is not a stopping error criteria, but sample interpolation is checked after each correlation matrix inversion.
We noticed that when using the same correlation parameters the direct Cokriging correlation matrix (non-symmetric) is better conditioned than the indirect Cokriging correlation matrix (symmetric). It can be explained by the fact that the augmented sampling of the indirect method exhibits samples very close from each other.

2.5.2 Memory and Computation Time

The computational cost to build the model depends only on the correlation matrix order N. On a standard Intel Pentium4 2.8GHz processor, the estimated computation time is:

$$\begin{array}{c} t_{user} \approx 3.10^{-9} N^3 \ seconds \\ memsize \approx 8.10^{-6} N^2 \ Mbytes \end{array} \tag{31}$$

For example, saying that the model must be built in one minute with 300 computations of the likelihood for the MLE fit, the maximum matrix order is then $N = 400$. For models approximating high-fidelity CFD solver the typical maximum number of computations authorized for an optimization is close to $n_s = 200$. The Kriging computation cost depending only on the number of samples ($N_{kriging} = n_s$) it is effectively negligible here. On

the contrary, the computation cost of Cokriging models can be limiting on high dimensional problems ($N_{cokriging} = n_s(n_{dv} + 1)$). Considering an optimization problem with a hundred variables, the computation time for a single correlation matrix inversion would be $t_{user} > 6$ hours.
The ScaLAPACK user guide [28] (parallel version of LAPACK) recommends a maximum matrix order of 1000 per CPU.
In term of memory, a Kriging code only requires to allocate the inverted correlation matrix plus some vectors in double precision.

3 Definition of the Aerodynamic Test Case

The non-linearties observed on aerodynamic global coefficients when varying shape cannot be represented by an analytical test function. This is why a typical CFD test case was chosen. The aerodynamic function considered within this article is the far-field wave drag [29] of a RAE2822 airfoil in an Euler flow at a Mach number $M = 0.78$ and an angle of attack $AOA = 0.6^o$. The flow analysis was performed with the elsA CFD code [30] on a structured multiblock C-mesh (33×130). The second order Roe's upwind scheme with the Van Albada limiter is used as spatial scheme coupled with an implicit time resolution. By modifying the shape of the transonic airfoil with Hicks-Henne sinusoidal bumps, the objective function exhibits multiple local optima and a non-linear behavior (Figure 1).

$$HH_k(u) = X_k sin^4\left(\pi u^{\frac{\log 0.5}{\log \alpha_k}}\right) \; ; \; u \in [0,1] \tag{32}$$

The six design variables considered are amplitudes X_k of Hicks-Henne bumps on the upper part at $\alpha_k = 10\%, 80\%, 38\%, 66\%, 24\%, 52\%$ of the chord line so as the two first parameters correspond to deformations of the leading edge and of the zone were the shock occurs. The distribution is chosen so as when incrementally increasing the number of design variables, the bumps stay equidistributed on the upper surface. The amplitude of each bump is in the domain $[-5; +5]mm$, and the chord length of the airfoil is one meter.
At each update of the design variables the initial mesh is deformed using an improved integral method (iso-topology). The mesh deformation module is also able to compute the sensitivity of the mesh with respect to the shape variables needed to compute the gradient of the objective function.
The gradient information used to build the Cokriging models was computed using the discrete adjoint method [16, 1]. In term of computational cost, a gradient calculation with the adjoint method costs about the same as one direct CFD computation, so a Cokriging model built with n_s samples costs the same as a Kriging model on $2n_s$ samples for any number of design variables. It should be noticed that the discrete adjoint formulation used within this paper ensures a coherence between direct flow computation and gradient computation and thus a correlation between numerical residual errors at convergence.
The dimension of the problem varies from two to six, and for each problem a reference function considered exact is built using respectively 104, 208, 272, 1056 and 4160 samples. The indirect Cokriging formulation was used to build the reference functions, but as the five dimensional reference was already expensive to build with a matrix order $N = 6336$ the six dimensional reference was built with a Kriging formulation instead. The Cokriging code should be made parallel to extend this study to higher dimensions.

Due to the important number of CFD evaluations required to establish references, a coarse mesh was used. It is worth noting that this is not detrimental, since we are addressing inviscid flow simulations and that aerodynamic objective functions on finer meshes or more realistic problems exhibit the same kind of non-linearities.
A Sobol technique was used for the space filling sample distributions and the corners of the box domain were added to the distributions.
Only the first test surface with two variables can be plotted, but about the same level of complexity is expected when increasing the number of design variables.

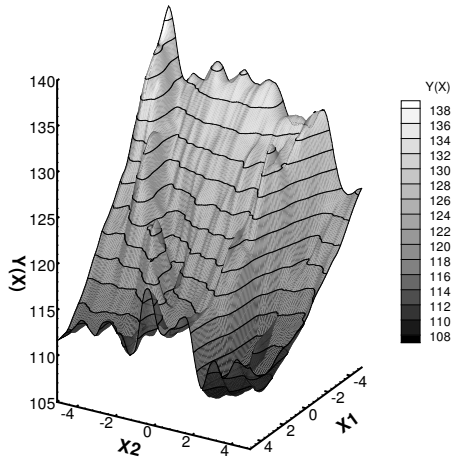

Figure 1: Aerodynamic reference surface for $n_{dv} = 2$

4 Influence of Sampling Methods on Global Accuracy of Response Surfaces

As predicted errors are available with Kriging we wanted to determine if using them as 'a posterioi' sampling criteria could outperform 'a priori' methods. Maximum error 'a posteriori' sampling were already applied to polynomial test cases [10, 15] , but without comparison with classical space filling designs.
The different techniques presented in 2-2.1 were compared to reconstruct our first reference function depending on two variables (Figure 1) with as few samples as possible. Figures 2 and 3 show the evolution of the exact relative root mean squared error (RMSE) when increasing the number of samples from 6 to 64. The target is the dashed line corresponding to the sampling refinement at maximum of exact error.

4.1 Performance of 'a priori' Sampling Methods (grid, Halton, Sobol, LHS)

The fact that LHS and grid method are not sequential explains why on Figure 2 the error converges unsmoothly when increasing sampling size. Otherwise, the lack of coherence on the convergence can also be explained by problems of correlation parameter optimization.

The optimal parameters in the sense of MLE are not always the optimal parameters in the sense of minimum exact error. More particularly, grid sampling is unadapted to build Kriging because optimization of the directional correlation parameters is biased by the linear dependence existing within the spatial distribution [31, 32].
Anyhow, regarding Figure 2 one can observe that the convergence of the error seems composed of two phases. Firstly, we observe a transitory phase of strong error reduction for small datasets with a negative power trend until reaching $n_s = 20$. Secondly, the error seems to reach a linear phase with a low negative slope. This linear phase is very costly if a low level of error is required.
The Halton method was taken as the reference for 'a priori' methods for its good overall performance with small and large datasets. However, the only method applicable to high dimensions is LHS.

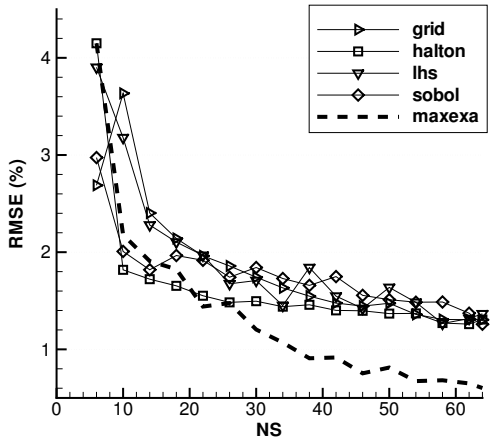

Figure 2: Evolution of error with sampling size for 'a priori' sampling methods, $n_{dv} = 2$

4.2 Performance of 'a posteriori' Sequential Sampling Methods

Looking at Figure 3 it appears that all methods are quasi-equivalent to the Halton reference. The linear phase of convergence is not avoided using sampling at maximum predicted error.
Regarding 'a posteriori' methods, the $max(dKdS_{mix})$ error refinement is the best. It is understandable as the $dKdS_{mix}$ error is based on $dKdS$ error, the Kriging sensitivity to the samples. It performs better than the Halton reference for large datasets (Table 1). This is coherent with the fact that as the number of samples increases, the model is more precise and the predicted error is also more precise.
On small datasets, sampling refinement at maximum of error (even $max(Exa)$) does not outperform Halton but still produces very satisfying distributions.
When ignoring quasi-random sequence samplings, 'a posteriori' methods outmatch LHS and grid for small and large sampling size.
The fact that 'a posteriori' methods outperform only slightly space filling designs shows that the aerodynamic test function complexity is evenly distributed in the design space. In

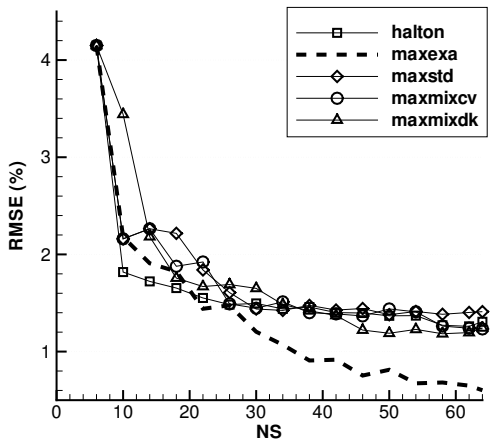

Figure 3: Evolution of error with sampling size for 'a posteriori' sampling methods, $n_{dv} = 2$

our sense this is always the case in aerodynamic optimization and global reponse surface are generally unaffordable.

method	$n_s = 6...25$	$n_s > 25$	$n_s \geq 6$
$grid$	2.43	1.49	1.92
$Halton$	1.96	1.39	1.76
LHS	2.45	1.57	2.01
$Sobol$	2.11	1.61	1.85
$max(Exa)$	2.17	0.89	1.47
$max(S)$	2.36	1.43	1.87
$max(CV_{mix})$	2.32	1.38	1.84
$max(dKdS_{mix})$	2.43	1.36	1.86

Table 1: Integrated relative RMSE for all sampling methods, $n_{dv} = 2$

5 Influence of Dimension on Global Response Surface Accuracy and Benefits of Cokriging

No general rule can be made to assess a priori the number of samples needed to achieve a given accuracy as it depends on the complexity of the unknown function. It is generally admitted that to obtain a correct global response surface, 10 samples per direction is a good starting sample size ($n_s \approx 10^{n_{dv}}$). The sample size increasing exponentially it becomes then impossible to build a global response surface on high dimensional problems. The benefits of Cokriging beside Kriging cannot be assessed by only varying the sample size [18, 19, 21] because the information added to the model through gradients is dimensional. Thus the accuracy of Kriging and Cokriging will be compared by varying the sample size and the

number of design variables. The dependence between number of samples and dimension is expected to be reduced with Cokriging. When dealing with approximation of expensive computer analysis, the dataset size is generally very limiting and the reconstruction was made with less than 300 samples. The comparison was conducted on an aerodynamic function (section 3) in order to use the same solvers as in aerodynamic shape optimization and gradient computed with the adjoint state formulation.

5.1 Results

Figures 4-8 show the evolution of the exact (with respect to the reference surface) root mean squared error (RMSE in percent) when increasing the number of samples for the five problems.

Firstly it appears that the indirect Cokriging method is the best. Despite the mathematical appeal of direct Cokriging formulation its performance are deceiving. Overall, it is worse than Kriging on the two, three and five dimensional tests (Table 2). The spikes occuring on direct Cokriging convergence correspond to wrong MLE correlation parameters. The MLE problem has more local minima representing optimal correlation parameters for each of the $n_{dv}+1$ functions (the function itself and its derivatives Eqn. 26). In fact, the MLE of direct Cokriging sometimes gives the same parameters as MLE of Kriging and sometimes gives the same parameters as MLE of indirect Cokriging, for example on Figure 7 between 150 and 200 samples. This does not discard the method as direct and indirect Cokriging models give the same results when using the same correlation parameters, but shows that direct Cokriging models fitted by MLE are easily ill-fitted.

The convergence in two phases described previously (in section 4) appears on all tests with Kriging or Cokriging models.

The number of samples needed to switch from transitory to linear phase is roughly the same for Kriging and Cokriging (Figures 4-8). The benefits of the Cokriging models on error is obvious during the transitory phase (Table 2).

During the second phase, the low negative linear slope is the same for Kriging and Cokriging models. However, the indirect Cokriging is still more efficient due to the advance taken during the first phase (Table 2).

In term of computation cost, the indirect Cokriging does effectively need fewer samples to reach a given accuracy. As the samples are twice more expensive to compute for Cokriging (cost of the adjoint solver), they should use at least twice less samples than Kriging to reach a given accuracy to be interesting. This seems true in general. By interpolating the gradient vector at each sample the surrogate models interpolate hyperspheres instead of points and as expected the best gain is achieved on the higher dimensional problem with small datasets.

It appears on table 2 that global error increases with the number of variables. It was expected since it was not possible to have a sampling size of ten samples per direction for all tests.

n_{dv}	$n_s = 3...25$			$n_s > 25$			$n_s \geq 3$		
	Krig.	D.Cok.	I.Cok.	Krig.	D.Cok.	I.Cok.	Krig.	D.Cok.	I.Cok.
2	3.45	2.87	2.42	2.20	1.76	1.57	2.57	2.09	1.82
3	5.23	6.18	3.36	2.44	2.60	1.66	2.84	3.11	1.91
4	8.40	8.16	3.91	3.14	3.58	2.57	3.75	4.11	2.72
5	17.9	15.7	8.68	5.59	6.05	4.73	6.62	6.85	5.05
6	22.8	20.0	11.6	7.25	6.71	6.71	8.45	7.74	7.08

Table 2: Integrated relative RMSE for all formulations and tests

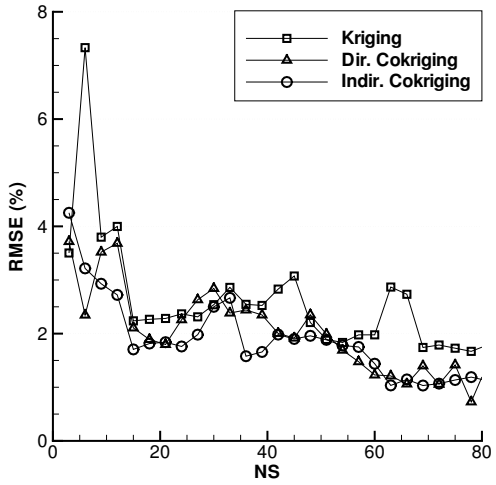

Figure 4: Evolution of Kriging and Cokriging errors with sampling size, $n_{dv} = 2$

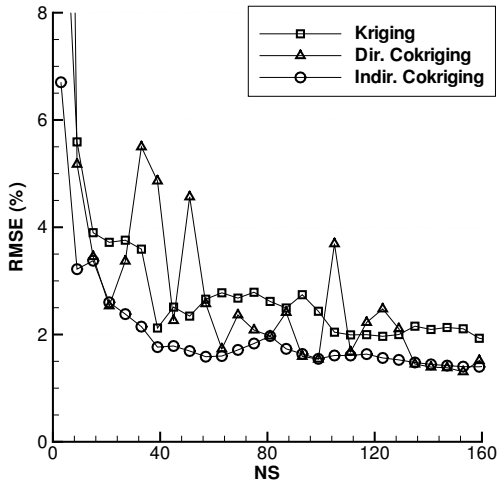

Figure 5: Evolution of Kriging and Cokriging errors with sampling size, $n_{dv} = 3$

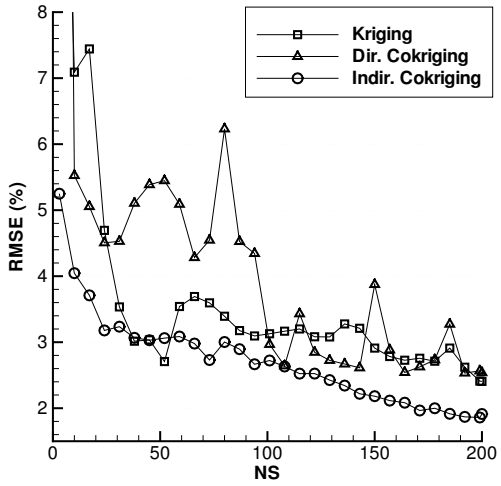

Figure 6: Evolution of Kriging and Cokriging errors with sampling size, $n_{dv} = 4$

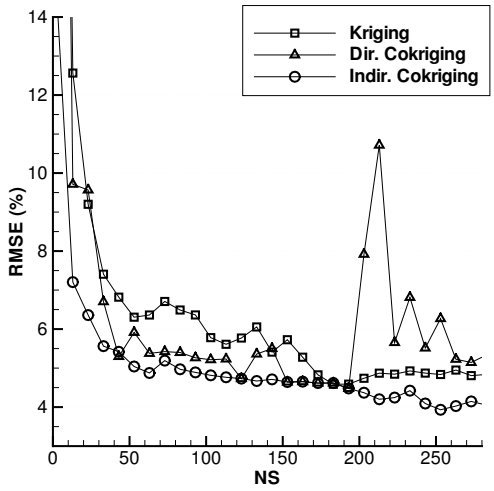

Figure 7: Evolution of Kriging and Cokriging errors with sampling size, $n_{dv} = 5$

6 Conclusion

An efficient framework to fit Kriging correlation parameters by likelihood optimization was described. It consisits in departing from a good initial guess for the parameters and then apply a gradient optimizer. It shows good overall robustness, even during the fit of computer expensive Kriging (up to order $N = 6336$).

Usualy, comparison of modeling techniques are made on analytical test functions. Aerodynamic functions being highly non-linear, anaytical test cases cannot represent the true level of complexity encountered in aerodynamic shape optimization.

Classical response surface frameworks use space filling 'a priori' sampling. It is possible to tune the sample distribution to the function studied using 'a posteriori' refinement techniques to concentrate more information into the sampling. Compared to space fill-

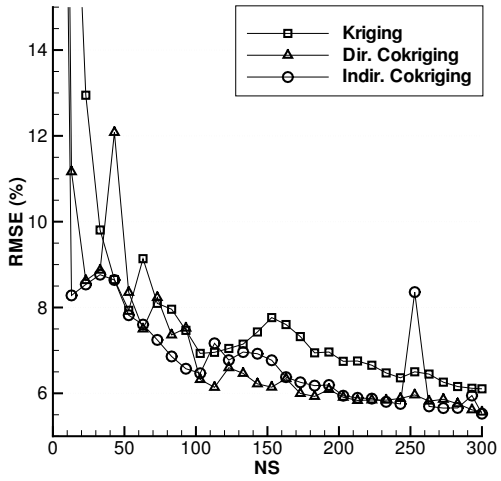

Figure 8: Evolution of Kriging and Cokriging errors with sampling size, $n_{dv} = 6$

ing designs, 'a posteriori' sampling at the maximum of exact error significantly increases accuracy of the model. Kriging being an interpolating model, any error criterion null at samples (standard error) can be used for sampling refinement. Two new predicted error criteria based on cross-validation, but satisfying this property were introduced.
The convergence of the global error when increasing the number of samples proceeds in a transitory phase of fast decrease followed by a linear phase of slow decrease. The sample size needed to reach this second phase cannot be foretold, but increases with dimension and complexity of the function studied.
Among 'a priori' techniques, quasi-random Halton give best results. Moreover, it also gives ordered samples. The new sequential sampling methods based on refinement at maximum error show fair results compared to classical sampling methods, but do not break the slow linear convergence as was expected regarding performance of refinement at maximum of exact error. From all predicted errors, the new leave one out sensitivity error introduced within this article enables to have the smallest exact error on large datasets, but needs to build all intermediate models. The fact that 'a posteriori' methods outperform only slightly space filling designs shows that the aerodynamic test function complexity is evenly distributed in the design space. In our sense this is always the case in aerodynamic optimization and global reponse surface are generally unaffordable.
The use of adjoint gradient information to build Cokriging models drastically improves global accuracy, but is unsufficient to break the slow convergence during the second linear phase. The computation cost induced by Cokriging formulation makes the response surface expensive to build when considering lots of design variables, but as the optimization process usually takes place on parallel computer architecture, it is possible to lessen this cost.
As was stated in previous papers[19], indirect and direct Cokriging models are identical when built with the same correlation parameters. However, the sample database augmentation used to build indirect Cokriging induces bad condition number of the correlation matrix that can be avoided using the direct formulation. Despite its better conditioning, the direct Cokriging model cannot be easily fitted. In our view, the best framework to

build gradient enhanced Kriging reuses correlation parameters from the indirect formulation to build a direct Cokriging. The best gain of Cokriging is observed during the transient phase, on small datasets with a large number of design variables.
Regarding the global error of reconstruction on all test cases considered within this paper, it appears that a global response surface accurate at one percent is out of reach with less than a few hundred samples except for dimension lesser than four. That is why, global response surfaces should be used only when it is possible to identify a very small set of interesting design variables or with low-fidelity solvers and large sample database.
However, physical functions generally exhibit a simple global tendency plus non-linear perturbations and it is still possible to find an optimum by iteratively refining the model at its predicted optimum. To overcome early convergence of this process it is interesting to reuse predicted error to introduce some exploration. The general tendencies found on the low-accuracy global response surface can also be used to initialize a local optimizer closer to the global optimum. As the domain of accuracy around sample locations is increased when interpolating gradient, an optimization process refining iteratively a local Cokriging response surface around interesting locations should be investigated on a high dimensional problem.

Acknowledgments

The authors would like to acknowledge M. Meaux (Mathieu.Meaux@airbus.com) and M. Montagnac (Marc.Montagnac@cerfacs.fr) for their support and suggestions. We also thank *Airbus France* for providing their shape optimization suite *OPTaliA*. We used the computer implementation of J. Burkardt (`http://www.scs.fsu.edu/~burkardt/`) for 'a priori' sampling methods and thank him for sharing his source code.

References

[1] M. Meaux, M. Cormery, and G. Voizard. Viscous aerodynamic shape optimization based on the discrete adjoint state for 3d industrial configurations. European Congress on Computational Methods in Applied Sciences and Engineering, Jul. 2004.

[2] Jaroslaw Sobieszczanski-Sobieski and Raphael T. Haftka. Multidisciplinary aerospace design optimization: Survey of recent developments. In *4th Aerospace Science Meeting & Exhibit*, number AIAA-1996-0711, 1996.

[3] S. Tursi. Transonic wing optimization combining genetic algorithm and neural network. In *21st Applied Aerodynamics Conference*, number AIAA 2003-3787.

[4] S. Pierret, H. Kato, R. Coelho, and A. Merchant. Multi-objective and multidisciplinary shape optimization. In *EUROGEN*, 2005.

[5] S. Jeong, M. Murayama, and K. Yamamoto. Efficient optimization design method using kriging model. *Journal of Aircraft*, 42(2):413–419, 2005.

[6] Andrew J. Booker, J. E. Dennis, Jr., Paul D. Frank, David B. Serafini, Virginia Torczon, Trosset, and W. Michael. A rigorous framework for optimization of expensive

functions by surrogates. Technical Report TR-98-47, Center for Research on Parallel Computation, 1998.

[7] Natalia Alexandrov, J. E. Dennis, Jr., Robert Michael Lewis, Torczon, and Virginia. A trust region framework for managing the use of approximation models in optimization. Technical Report TR-97-50, 1997.

[8] D. Cox and S. John. Sdo: A statistical method for global optimization. In *Multidisciplinary design optimization (Hampton, VA, 1995)*, pages 315–329. SIAM, Philadelphia, PA, 1997.

[9] D. R. Jones, M. Schonlau, and W. J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13:455–492, 1998.

[10] M. J. Sasena. *Flexibility and Efficiency Enhancements for Constrained Global Design Optimization with Kriging Approximations*. PhD thesis, University of Michigan, 2002.

[11] J.C. Jouhaud, P. Sagaut, and B. Labeyrie. A kriging approach for CFD/Wind tunnel data comparison. *Journal of Fluid Engineering*, 128:847–855, 2006.

[12] J.-C. Jouhaud, P. Sagaut, M. Montagnac, and J. Laurenceau. A surrogate model based multi-disciplinary shape optimization method with application to a 2d subsonic airfoil. *Computers and Fluids*, 36(3):520–529, 2007.

[13] T. W. Simpson, J. D. Peplinski, P. N. Koch, and J. K. Allen. Metamodels for computer-based engineering design: Survey and recommendations. *Engineering with Computers*, 17:129–150, 2001.

[14] T. W. Simpson, D. K. Lin, and W. Chen. Sampling strategies for computer experiments: Design and analysis. *Journal of Reliability and Applications*, 2:209–240, 2001.

[15] W. C. M. van Beers. *Kriging Metamodeling for Simulation*. PhD thesis, Universiteit van Tilburg, 2005.

[16] A. Jameson. Aerodynamic design via control theory. *Journal of Scientific Computing*, 3(3), 1988.

[17] J. Peter, Chi-Tuan Pham, and F. Drullion. Contribution to discrete implicit gradient and discrete adjoint method for aerodynamic shape optimisation. ECCOMAS, 2004.

[18] R. M. Lewis. Using sensivity information in the construction of kriging models for design optimization. *AIAA Journal*, (AIAA-98-4799):730–737, 1998.

[19] W. Liu. *Development of Gradient-Enhanced Kriging Approximations for Multidisciplinary Design Optimization*. PhD thesis, University of Notre Dame, 2003.

[20] H.-S. Chung and J. J. Alonso. Using gradients to construct cokriging approximation models for high dimensional design optimization problems. In *40th AIAA Aerospace Science Meeting & Exhibit*, number AIAA-2002-0317, 2002.

[21] H.-S. Chung and J. J. Alonso. Design of a low-boom supersonic business jet using cokriging approximation models. In *9th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, number AIAA-2002-5598, 2002.

[22] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn. Design and analysis of computer experiments. *Statistical Science*, 4(4):409–435, 1989.

[23] L. Kocis and W. J. Withen. Computational investigations of low-discrepancy sequences. *ACM Transactions on Mathematical Software*, 23(2):155–294, 1997.

[24] J. D. Martin. *A Methodology for Evaluating System-Level Uncertainty in the Conceptual Design of Complex Multidisciplinary Systems*. PhD thesis, Pennsylvania State University, 2005.

[25] K. V. Mardia and A. J. Watkins. On multimodality of the likelihood in the spatial linear model. *Biometrika*, 2:289–295, 1989.

[26] G. J. Davis and M. D. Morris. Six factors which affect the condition number of matrices associated with kriging. *Mathematical Geology*, 5:669–683, 1997.

[27] E. Anderson, Z. Bai, C. Bischof, S. Blackford, J. Demmel, J. Dongarra, J. Du Croz, A. Greenbaum, S. Hammarling, A. McKenney, and D. Sorensen. *LAPACK Users' Guide*, third edition, 1999.

[28] L. S. Blackford, J. Choi, A. Cleary, E. Dázevedo, J. Demmel, I. Dhillon, J. Dongarra, S. Hammarling, G. Henry, A. Petitet, K. Stanley, D. Walker, and R. C. Whaley. *ScaLAPACK Users' Guide*, 1997.

[29] D. Destarac. Far-field / near-field drag balance and applications of drag extraction in cfd. In *CFD-Based Aircraft Drag Prediction and Reduction*. von Karman Institute for Fluid Dynamics, 2003.

[30] L. Cambier and M. Gazaix. elsa: an efficient object-oriented solution to cfd complexity. In *40th AIAA Aerospace Science Meeting & Exhibit*, number AIAA-2002-0108, 2002.

[31] M. Meckesheimer, R. B. Barton, T. W. Simpson, F. Limayem, and B. Yannou. Metamodeling of combined discrete/continuous responses. *AIAA Journal*, 39(10):1950–1959, 2001.

[32] B. Wilson, D. J. Cappelleri, T. W. Simpson, and M. I. Frecker. Efficient pareto frontier exploration using surrogate approximations. In *8th AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, number AIAA 2000-4895.