V European Conference on Computational Fluid Dynamics
ECCOMAS CFD 2010
J. C. F. Pereira and A. Sequeira (Eds)
Lisbon, Portugal,14-17 June 2010

SURROGATE MODELS BASED ON FUNCTION AND DERIVATIVE VALUES FOR AERODYNAMIC GLOBAL OPTIMIZATION

Manuel Bompard*, Jacques Peter† and Jean-Antoine Désidéri††

*†Office National des Etudes et Recherches Aérospatiales (ONERA),
Chatillon (France)
e-mail: manuel.bompard@onera.fr / jacques.peter@onera.fr
††Project-Team OPALE,
INRIA Centre de Sophia Antipolis - Mditerrane (France)
e-mail: Jean-Antoine.Desideri@sophia.inria.fr

Key words: Surrogate models, Adjoint vector method, Global Shape Optimization, Kriging, Radial Basis Function, Support Vector Regression

Abstract. *The application of global algorithms in aerodynamic shape optimization is still today very limited due to the required CPU time. To avoid this limitation, one method is to replace the objective function by a surrogate model. One important achievement of the recent years is the development of accurate and robust surrogate models, minimizing the quantity of information (computed via costly CFD simulations) required for their construction.*

In this article, we propose to use derivatives in addition to function values to determine the meta model. With adjoint methods, these derivatives may be obtained at a reduced cost, independent of the number of design variables. Three distinct meta model constructions including derivatives are presented and adapted. Then some aerodynamic design applications are presented. The benefit of including derivatives in the surrogate models is clearly evident.

1 INTRODUCTION

Optimum shape design in aerodynamics has been an active field for the last years. Gradient based methods allow to find an optimum at low computational cost in the neighborhood of the initial shape[35]. For complex objective functions, they can converge to a local optimum different from the global optimum. In this case, global optimization is necessary.

Global optimization methods seek for the optimum shape in the whole parametric domain, and require a large number of function evaluations. In aerodynamics, the computational cost of a 3D turbulent flow analysis using Computational Fluid Dynamics (CFD) is very high. Thus, global optimization is computationally very CPU demanding in complex industrial cases and it can be interesting to build a surrogate mathematical model from a sample of data, and to conduct the global optimization using both exact and surrogate estimates of the cost function[26].

Derivatives of aerodynamic functions with respect to (w.r.t.) design parameters can be obtained by the adjoint equation approach at a reduced computational cost, independent of the number of design variables[12]. In this study, we propose to enrich the inputs of three classical surrogate models by providing the gradient values to improve the accuracy of the meta-model and reduce the cost of their construction.

The article is organized as follows.

In section 2, three classical surrogate models are presented: Radial Basis Function, Kriging and Support Vector Regression (SVR). Since the accuracy of a model is very sensitive to the value of internal parameters, an automatic adjustment method is proposed.

The construction of each model is modified (section 3) to use derivatives. A mathematical test case is considered. Results are discussed in the end of the section.

In section 4, an algorithm combining surrogate models and global optimization is given. Again mathematical test cases are considered.

Finally, applications to aerodynamic shape optimization are presented in section 5.

2 CONSTRUCTION OF SURROGATE MODELS

In a context of shape optimization, we search to replace a function $f : x \in \mathcal{D} \subset \mathcal{R}^d \rightarrow \mathcal{R}$ by a mathematical model, called surrogate model, noted $\hat{f}$, built from a sample $\mathcal{E}$ of N responses $f(x^i) = y^i \in \mathcal{D}$ and from internal parameters P. The goal of the substitution is the reduction of computational cost by limiting the number of necessary CFD computations.

2.1 Design of experiments (DoE)

Building approximations of the objective function involves first to choose the design of experiments that will sample the input space $\mathcal{D}$. Several methods have been developed and compared during the last years : Latin Hypercubes[23], Hammersley Sequence Sampling[16], Orthogonal Arrays[25], Uniform Designs[7]. Hammersley Sequence Sam-

pling seems to be a good general N-point sampling of $[0,1]^d$[38]. The points are defined by:

$$x^{(N)} = 1 - \left[\frac{n}{N}, \Phi_{R_1}(n), \Phi_{R_2}(n), \ldots, \Phi_{R_{d-1}}(n)\right]^T \quad \forall n \in \{1 \ldots N\} \tag{1}$$

where $R_1, \ldots, R_{d-1}$ are the first $d-1$ prime numbers and $\Phi_R(n) = n_0 R^{-1} + n_1 R^{-2} + \ldots + n_m R^{-m-1}$ where $n = n_0 + n_1 R + n_2 R^2 + \ldots + n_m R^m$.

2.2 Descriptions of three classical surrogate models

2.2.1 Kriging

Finding their origin in works of Krige for mining and geostatistical applications[17][21], Kriging models have been introduced in literature by Sacks[29] in 1989. They have been rarely used in the 90s for industrial applications until the work of Simpson et al.[30] for aerodynamic design.

Kriging is based on the hypothesis that the data to interpolate are spatially correlated. So, it combines a global deterministic model $m(x)$ and local variations, represented by a statistical model $\epsilon(x)$:

$$\hat{f}(x) = m(x) + \epsilon(x). \tag{2}$$

As the data are correlated, the output at an unknown point can be written as a linear combination of the observed values:

$$\hat{f}(x^*) = \sum_{i=1}^{N} \lambda^i(x^*) f(x^i). \tag{3}$$

The vector of the λ^i is computed by finding the Best Linear Unbiased Predictor (BLUP), which is unbiased and minimizes mean quadratic error. In practice, the dependence structure of $f(x)$ is never known and is approximated by the following covariance function:

$$\phi(x^i, x^j) = Cov(f(x^i), f(x^j)) = \exp\left[-\frac{1}{2}\sum_{k=1}^{d}(\frac{x_k^i - x_k^j}{\sigma_k})^2\right]. \tag{4}$$

With this choice, the method can be applied to a pure deterministic problem. Different types of Kriging implementation have been proposed. Ordinary kriging, based on the hypothesis that the mean of the stochastic process is constant in space is used in this study:

$$\mu(x) = \mu. \tag{5}$$

Finding BLUP under unbiased condition ($\sum_{i=1}^{N} \lambda^i = 1$) leads to following system to be solved for λ :

$$\begin{bmatrix} \phi(x^1,x^1) & \cdots & \phi(x^1,x^N) & 1 \\ \vdots & & \vdots & 1 \\ \phi(x^N,x^1) & \cdots & \phi(x^N,x^N) & 1 \\ 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda^1 \\ \vdots \\ \lambda^N \\ \nu \end{bmatrix} = \begin{bmatrix} \phi(x^1,x^*) \\ \vdots \\ \phi(x^N,x^*) \\ 1 \end{bmatrix}. \tag{6}$$

2.2.2 Radial Basis Function

Radial Basis Function are elementary one-hidden-layer neural networks using RBF functions as activation functions[2]. In this work, the same radius is used for all activation functions. The number of cells in the hidden layer is fixed to the number of points in the learning sample (actually, the centers c^i coincide with the observations x^i). This leads to the following square linear system:

$$\begin{bmatrix} \phi(x^1,x^1) & \cdots & \phi(x^1,x^N) \\ \vdots & & \vdots \\ \phi(x^N,x^1) & \cdots & \phi(x^N,x^N) \end{bmatrix} \begin{bmatrix} w^1 \\ \vdots \\ w^N \end{bmatrix} = \begin{bmatrix} y^1 \\ \vdots \\ y^N \end{bmatrix}. \tag{7}$$

The outputs of the network are then computed by:

$$\hat{f}(x^*) = \sum_{i=1}^{N} w^i \phi(x^*, x^i). \tag{8}$$

2.2.3 Support Vector Regression

Support Vector Regression is a powerful method for the estimation of a function from a learning sample. Developed in the 90s by Vapnik[36] from Learning Theory of Vapnik and Chervonenkis[37], it is considered today as one of the most accurate surrogate model. A very complete presentation has been made by Smola[32] and Gunn[10]. In the linear case, SVR consists in finding the smoothest linear function $f(x) = \langle w, x \rangle + b$ which approximates all observations with a given precision ϵ_0:

$$\begin{aligned} \min_{w,b,\xi,\xi^*} \quad & \frac{1}{2}\|w\|^2 + C_0 \sum_{i=1}^{N} (\xi^{(i)} + \xi^{(i)*}) \\ \text{with} \quad : \quad & \left\langle w, x^{(i)} \right\rangle + b - y^{(i)} \leq \epsilon_0 + \xi^{(i)} \quad \forall i \in \{1 \dots N\} \\ & y^{(i)} - \left\langle w, x^{(i)} \right\rangle - b \leq \epsilon_0 + \xi^{(i)*} \quad \forall i \in \{1 \dots N\} \\ & \xi^{(i)}, \xi^{(i)*} \geq 0 \quad \forall i \in \{1 \dots N\} \end{aligned} \tag{9}$$

By applying KKT conditions, dual formulation can be derived:

$$\begin{aligned} \min_{\alpha,\alpha^*} \quad & \frac{1}{2} \left[\sum_{i,j=1}^{N} (\alpha^{(i)*} - \alpha^{(i)})(\alpha^{(j)*} - \alpha^{(j)}) \left\langle x^{(i)}, x^{(j)} \right\rangle \right] + \epsilon_0 \sum_{i=1}^{N} (\alpha^{(i)} + \alpha^{(i)*}) + \sum_{i=1}^{N} (\alpha^{(i)} - \alpha^{(i)*}) y^{(i)} \\ \text{with} \quad : \quad & \sum_{i=1}^{N} (\alpha^{(i)} - \alpha^{(i)*}) = 0 \\ & \alpha^{(i)}, \alpha^{(i)*} \in [0, C_0] \quad \forall i \, \{1 \dots N\} \end{aligned} \tag{10}$$

where α and α^* are Lagrange multipliers associated with the inequality constraints.

This optimization problem is quadratic in $(\alpha^{(i)}, \alpha^{(i)*})$. It can be solved easily by a quadratic programming method (QP) as the LOQO method of Vanderbei[34], which was used for this work. Then, the output is computed by the following formula:

$$\hat{f}(x^*) = \langle w, x^* \rangle + b = \sum_{i=1}^{N} (\alpha^{(i)*} - \alpha^{(i)}) \left\langle x^{(i)}, x^* \right\rangle + b, \tag{11}$$

where b is computed as $(y^i - \left\langle w, x^{(i)} \right\rangle - \epsilon_0)$ for $\alpha^{(i)} \in]0, C[$ or $(y^i - \left\langle w, x^{(i)} \right\rangle + \epsilon_0)$ for $\alpha^{(i)*} \in]0, C[$.

This method can be extended easily to a nonlinear case by introducing a kernel function $k(x, y)$ which simulates dot products in a high dimensional feature space[32]:

$$\begin{aligned} \min_{\alpha,\alpha^*} \quad & \frac{1}{2} \left[\sum_{i,j=1}^{N} (\alpha^{(i)*} - \alpha^{(i)})(\alpha^{(j)*} - \alpha^{(j)}) k(x^{(i)}, x^{(j)}) \right] + \epsilon_0 \sum_{i=1}^{N} (\alpha^{(i)} + \alpha^{(i)*}) + \sum_{i=1}^{N} (\alpha^{(i)} - \alpha^{(i)*}) y^{(i)} \\ \text{with} \quad : \quad & \sum_{i=1}^{N} (\alpha^{(i)} - \alpha^{(i)*}) = 0 \\ & \alpha^{(i)}, \alpha^{(i)*} \in [0, C_0] \quad \forall i \, \{1 \dots N\} \\ & k(x, y) = \exp \left[-\frac{1}{2} \sum_{k=1}^{d} (\frac{x_k - y_k}{\sigma_k})^2 \right] \end{aligned} \tag{12}$$

The model value in x^* is then given by:

$$\hat{f}(x^*) = \sum_{i=1}^{N} (\alpha^{(i)*} - \alpha^{(i)}) k(x^{(i)}, x^*) + b, \tag{13}$$

2.3 Optimization of internal parameters

For each of the three meta models presented above, the choice of internal parameters is a complex and critical issue. Parametric studies have been made for the three methods (RBF[24], Kriging[1], SVR[3][22]). They conclude that the best choice of parameter values is problem-dependent. We propose here to automatically tune the meta model by minimizing the *leave-one-out* error in the space of internal parameters:

$$\min_{P} e_{loo} = \frac{1}{N} \sum_{i=1}^{N} (\hat{f}_i(x^i) - y^i)^2 \tag{14}$$

An efficient computation of the *leave-one-out* error is used for Kriging and RBF method applying Rippa's method[28]:

$$e_{loo} = \frac{y^T H^{-2} y}{N diag(H^{-2})}, \tag{15}$$

where H is the matrix of the Kriging (or RBF) system and $diag(H^{-2})$ is a vector of diagonal elements of H^{-2}. Optimization is carried out by global methods because the objective function is multi-modal. Both Particle Swarm Optimization or Differential Evolution are used.

3 USE OF DERIVATIVES IN SURROGATE MODELS

In different fields of numerical simulation, derivatives of real functions w.r.t. design parameters may be inexpensively computed by the adjoint vector method. It is proposed here to use these derivatives in the construction of the surrogate models to improve their accuracy. In this part, the three meta models are successively considered in this respect.

3.1 Gradient-assisted surrogate models

3.1.1 Co-Kriging

In the original *Kriging* theory, only the values the function to be approximated (primary informations) are used to build the model. It can be useful to use other informations, called secondary information, to improve the accuracy of the estimation. For this kind of applications, *Co-Kriging* has been applied[11].

Actually *Co-Kriging* covers two different methods, called *Indirect Co-Kriging* and *Direct Co-Kriging*[20]. The first one consists to use the derivative for enrich the learning sample by applying a Taylor approximation. A Kriging model is then built on this augmented sample. Laurenceau et al.[18] have shown that the direct method outperforms this approach. *Direct Co-Kriging* is the *true Co-Kriging* method, in which derivatives are used as secondary informations[4]. This version is used in this study. The surrogate function is still a linear in all input data ($y^i_{x_k}$ is the value of derivative of f in x^i w.r.t. the kth coordinate):

$$\hat{f}(x^*) = \sum_{i=1}^{N} \lambda^i(x^*) y^i + \sum_{k=1}^{d} \sum_{i=1}^{N} \lambda^i_k(x^*) y^i_{x_k}. \tag{16}$$

Weights are computed as the solution of following $[N(d+1)+1]^2$ linear system:

$$\begin{bmatrix} \Phi & \frac{\partial \Phi}{\partial y_1} & \cdots & \frac{\partial \Phi}{\partial y_d} & 1 \\ \frac{\partial \Phi}{\partial x_1} & \frac{\partial^2 \Phi}{\partial x_1 \partial y_1} & \cdots & \frac{\partial^2 \Phi}{\partial x_1 \partial y_d} & 1 \\ \vdots & \vdots & & \vdots & 1 \\ \frac{\partial \Phi}{\partial x_d} & \frac{\partial^2 \Phi}{\partial x_d \partial y_1} & \cdots & \frac{\partial^2 \Phi}{\partial x_d \partial y_d} & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} \Lambda \\ \Lambda_1 \\ \vdots \\ \Lambda_d \\ \nu \end{bmatrix} = \begin{bmatrix} \phi \\ \frac{\partial \phi}{\partial x_1} \\ \vdots \\ \frac{\partial \phi}{\partial x_d} \\ 1 \end{bmatrix}, \tag{17}$$

where Φ is the $N \times N$ matrix of $[\phi(x^i, x^j)]_{(i,j)}$ and ϕ_k is the vector of $[\phi(x^*, x^i)]_{(i)}$.

3.1.2 Gradient-assisted Radial Basis Function (GA-RBF)

The use of derivatives as auxiliary informations in radial basis function networks has been proposed by Giannakoglou et al.[9]. The structure of the network does not change:

$$\hat{f}(x^*) = \sum_{i=1}^{N_c} w^i \phi^i(x^*, c^i). \tag{18}$$

Giannakoglou proposes to express weights w^i as combination of $d+1$ new coefficients b^i and $a^i_k, \quad k \in [1, \cdots, d]$:

$$w^i = b^i + \sum_{k=1}^{d} a^i_k (1 + (x^*_k - c^i_k)) \tag{19}$$

The coefficients are determined by solving the following $[N(d+1)]^2$ linear system:

$$\begin{bmatrix} \Phi & \frac{\partial \Phi}{\partial y_1} & \cdots & \frac{\partial \Phi}{\partial y_d} \\ \frac{\partial \Phi}{\partial x_1} & \frac{\partial^2 \Phi}{\partial x_1 \partial y_1} & \cdots & \frac{\partial^2 \Phi}{\partial x_1 \partial y_d} \\ \vdots & \vdots & & \vdots \\ \frac{\partial \Phi}{\partial x_d} & \frac{\partial^2 \Phi}{\partial x_d \partial y_1} & \cdots & \frac{\partial^2 \Phi}{\partial x_d \partial y_d} \end{bmatrix} \begin{bmatrix} b \\ a_1 \\ \vdots \\ a_d \end{bmatrix} = \begin{bmatrix} \phi \\ \frac{\partial \phi}{\partial x_1} \\ \vdots \\ \frac{\partial \phi}{\partial x_d} \end{bmatrix}, \tag{20}$$

3.1.3 Gradient-assisted Support Vector Regression (GA-SVR)

In 2005, Lazaro et al.[19] proposed to add constraints in the primal problem of SVR to force approximation of derivatives. In 2006, Jayadeva et al.[13] modified the SVR method to take in count derivatives by resolving two linear systems. Here we apply Lazaro's approach (ie derivatives conditions are added to the initial system):

$$\begin{aligned} \min_{w,b,\xi,\xi^*,\tau,\tau^*} \quad & \frac{1}{2}\|w\|^2 + C_0 \sum_{i=1}^{N} (\xi^{(i)} + \xi^{(i)*}) + \sum_{k=1}^{d} C_k \sum_{i=1}^{N} (\tau_k^{(i)} + \tau_k^{(i)*}) \\ \text{with} \quad : \quad & \left\langle w, \Phi(x^{(i)}) \right\rangle + b - y^{(i)} \leq \epsilon_0 + \xi^{(i)} \quad \forall i \in \{1 \dots N\} \\ & y^{(i)} - \left\langle w, \Phi(x^{(i)}) \right\rangle - b \leq \epsilon_0 + \xi^{(i)*} \quad \forall i \in \{1 \dots N\} \\ & \left\langle w, \frac{\delta \Phi(x^{(i)})}{\delta x_k} \right\rangle - \frac{\delta y^{(i)}}{\delta x_k} \leq \epsilon_k + \tau_k^{(i)} \quad \forall i \in \{1 \dots N\}, \quad \forall k \in \{1 \dots d\} \\ & \frac{\delta y^{(i)}}{\delta x_k} - \left\langle w, \frac{\delta \Phi(x^{(i)})}{\delta x_k} \right\rangle \leq \epsilon_k + \tau_k^{(i)*} \quad \forall i \in \{1 \dots N\}, \quad \forall k \in \{1 \dots d\} \\ & \xi^{(i)}, \xi^{(i)*}, \tau_k^{(i)}, \tau_k^{(i)*} \geq 0 \quad \forall i \in \{1 \dots N\}, \quad \forall k \in \{1 \dots d\} \end{aligned} \tag{21}$$

The dual problem becomes:

$$\begin{aligned} \min_{\alpha,\alpha^*,\lambda,\lambda^*} \quad & \frac{1}{2} \begin{bmatrix} \alpha \\ \alpha^* \\ \Lambda_1 \\ \Lambda_1^* \\ \vdots \\ \Lambda_d \\ \Lambda_d^* \end{bmatrix}^T \begin{bmatrix} S(K) & S\left(\frac{\delta K}{\delta y_1}\right) & \cdots & S\left(\frac{\delta K}{\delta y_d}\right) \\ S\left(\frac{\delta K}{\delta x_1}\right) & S\left(\frac{\delta^2 K}{\delta x_1 y_1}\right) & \cdots & S\left(\frac{\delta^2 K}{\delta x_1 y_d}\right) \\ \vdots & \vdots & & \vdots \\ S\left(\frac{\delta K}{\delta x_d}\right) & S\left(\frac{\delta^2 K}{\delta x_d y_1}\right) & \cdots & S\left(\frac{\delta^2 K}{\delta x_d y_d}\right) \end{bmatrix} \begin{bmatrix} \alpha \\ \alpha^* \\ \Lambda_1 \\ \Lambda_1^* \\ \vdots \\ \Lambda_d \\ \Lambda_d^* \end{bmatrix} + \begin{bmatrix} \epsilon_0 + Y \\ \epsilon_0 - Y \\ \epsilon_1 + \frac{\delta Y}{\delta x_1} \\ \epsilon_1 - \frac{\delta Y}{\delta x_1} \\ \vdots \\ \epsilon_d + \frac{\delta Y}{\delta x_d} \\ \epsilon_d - \frac{\delta Y}{\delta x_d} \end{bmatrix}^T \begin{bmatrix} \alpha \\ \alpha^* \\ \Lambda_1 \\ \Lambda_1^* \\ \vdots \\ \Lambda_d \\ \Lambda_d^* \end{bmatrix} \\ \text{with} \quad : \quad & \begin{bmatrix} 1 \\ -1 \end{bmatrix}^T \begin{bmatrix} \alpha \\ \alpha^* \end{bmatrix} = 0, \quad S(M) = \begin{bmatrix} M & -M \\ -M & M \end{bmatrix} \\ & \alpha^{(i)}, \alpha^{(i)*} \in [0, C_0] \quad \forall i \{1 \dots N\} \\ & \lambda_k^{(i)}, \lambda_k^{(i)*} \in [0, C_k] \quad \forall i \{1 \dots N\}, \quad \forall k \in \{1 \dots d\} \end{aligned} \tag{22}$$

The new surrogate model is then:

$$f(x^*) = \langle w, \Phi(x^*)\rangle + b = \sum_{i=1}^{N}(\alpha^{(i)*} - \alpha^{(i)})k(x^{(i)}, x^*) + \sum_{k=1}^{d}\sum_{i=1}^{N}(\lambda_k^{(i)*} - \lambda_k^{(i)})\frac{\delta k}{\delta x_k}(x^{(i)}, x^*) + b, \tag{23}$$

3.2 Optimization of internal parameters for gradient assisted methods

To optimize internal parameters for gradient-assisted methods, the notion of *leave-one-out* error is extended by the following formula:

$$e_{galoo} = \frac{1}{N(1+d)}\sum_{i=1}^{N}\left[(\hat{f}_i(x^i) - y^i)^2 + \sum_{k=1}^{d}(\frac{\partial \hat{f}_i^k}{\partial x_k}(x^i) - y_k^i)^2)\right], \tag{24}$$

This error can still be computed by Rippa method for GA-RBF and Co-Kriging method.

3.3 Experiments using a mathematical function

The goal of these experiments is to discuss the benefit of using gradients in the construction of the meta models. The Rastrigin function, defined for a space dimension $d \geq 1$, is considered:

$$f(x) = \sum_{k=1}^{d}\left[x_k^2 - 10\left(\cos 2\pi x_k - 1\right)\right] \tag{25}$$

A 2D input space is chosen ($[0,5] \times [0,5]$). For all examples in this article, the design space is scaled in $[0,1]^d$. For different sizes of learning sample, each model is built and compared (using L_2 error) with the exact function on another, finer, sample of points (figure 1). The stopping criterion is fixed to 1% of the function mean value.

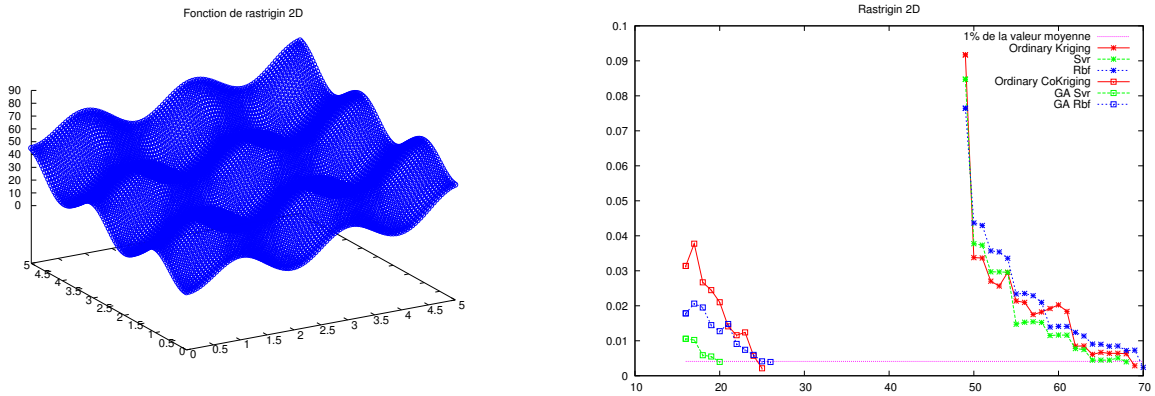

Figure 1: Accuracy of surrogate model for Rastrigin function in 2D

For the global approximation of a function, the use of derivatives seems to give promising results. If the cost of computing the derivative using an adjoint like method is equivalent to the cost of computing a function value, the use of one such model allows to reduce the computational cost required to approximate the function at the same accuracy. In the next part, surrogate models are used for global optimization.

4 SURROGATE-MODEL-BASED DESIGN OPTIMIZATION

The best method to combine surrogate models and global optimization algorithms is an active research topic since the article of Sacks[29]. Additionally, Simpson et al.[31] have given an overview of this research area.

Two different approaches must be distinguished. The first consists in building local models for optimization near each point of interest. In the second approach, a global model of the function to minimize is built, and optimization methods are applied to this model[27][8][26]. According to Praveen et Duvigneau[26], this last approach is more efficient. At each step of optimization, the model is enriched to improve its accuracy, as described in figure 2.

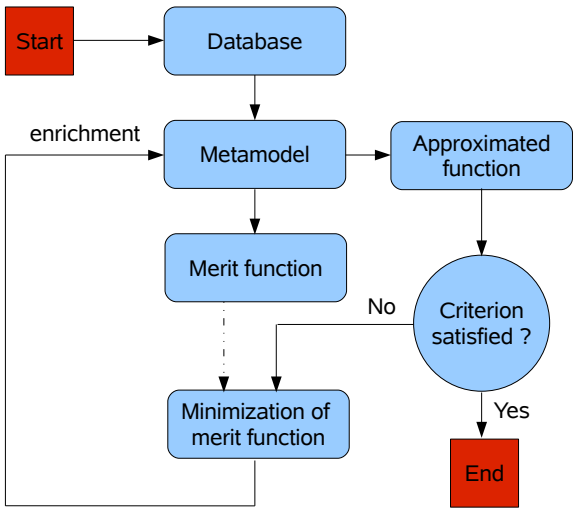

Figure 2: Global optimization assisted by surrogate model

4.1 Sampling enrichment

Several criteria have been proposed to identify most promising points[14][15]. The last article of Jones[15] seems to be the reference for the domain. He proposes two kinds of enrichment:

1. *Two-stage approach:* the criterion is computed from the model after its construction.

2. *One-stage approach:* the criterion is computed from mathematical properties of the model independently of its construction.

In CFD, the meta model computation cost is insignificant compared to flow simulations. This favours the choice of the two-stage approach. A simple and intuitive criterion is to enrich the sample by the minimum of the meta model. A simple example allows to understand that this choice can conduct to retain only one local optimum and this can cause the convergence to a minimum that is not global. To circumvent this difficulty, Jones[14] proposes to build a merit function from a statistic estimation of error in order to localise the most promising points. Several merit functions have been proposed [15][33][26]. The best choice seems to be a good compromise between exploitation capacity (its faculty to use identified interesting points) and exploration capacity (its faculty to sweep the entire design space). This is how *Weighed Expected of Improvement*[33] is defined by:

$$WEIF(x) = \begin{cases} w(y_{min} - \hat{f}(x))\Psi(\frac{y_{min}-\hat{f}(x)}{s(x)}) + (1-w)s(x)\psi(\frac{y_{min}-\hat{y}(x)}{s(x)}) & si \quad s(x) > 0 \\ 0 & si \quad s(x) = 0 \end{cases} \tag{26}$$

where :

- $y_{min} = \min(y^i, \quad i \in [1, N])$;
- Ψ is the cumulative distribution function of standard normal distribution;
- ψ is the probability density function of standard normal distribution;
- $s(x)$ is the standard deviation of the model statistical error which can be computed easily for the Kriging or Radial Basis Function model;
- $w \in [0, 1]$ is a weight factor between exploitation ($w = 0$) and exploration ($w = 1$).

4.2 Experiments using a mathematical function

The Rosenbrock function is a well known mathematical function used in test cases and defined for $d \geq 2$ by:

$$f(x) = \sum_{k=1}^{d-1} \left[(1 - x_k)^2 + 100(x_{k+1} - x_k^2)^2\right] \tag{27}$$

The initial sample, composed of 5 points, is computed by Hammersley method. Then, progressive enrichment, with *Weighted Expected of Improvement* criterion is performed. The optimization process is stopped when the minima is reached with ϵ accuracy. The number of iterations required is presented in the table below.

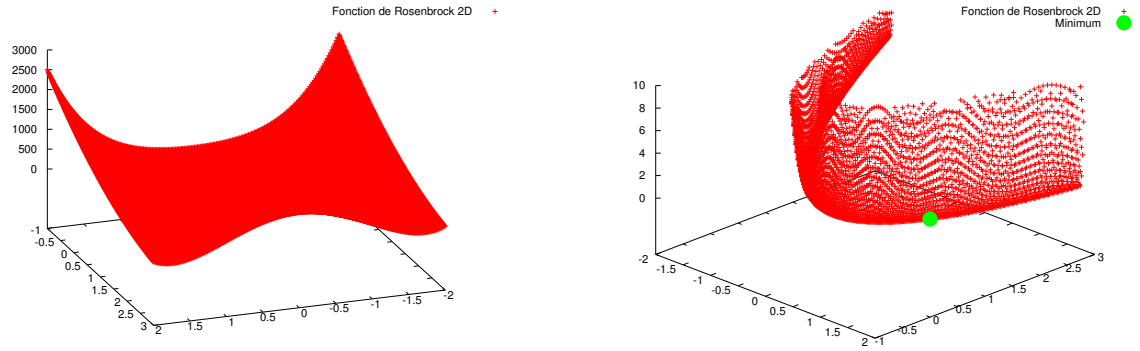

Figure 3: Rosenbrock function in 2D

ϵ	Kriging	RBF	Co-Kriging	GA-RBF
0.1	14	10	4	5
0.0001	18	20	6	15

Table 1: Results of surrogate assisted optimization on Rosenbrock 2D

5 APPLICATIONS TO SHAPE OPTIMIZATION PROBLEM

5.1 Nominal Geometry

The considered test case is related to drag minimization of an airfoil, starting from the classical NACA0012 geometry. A two-domain mesh, including 32148 points, is used (presented in figure 4). The flow characteristics are: $M = 0.73$, $\alpha = 3°$ and $Re = 6.10^6$.

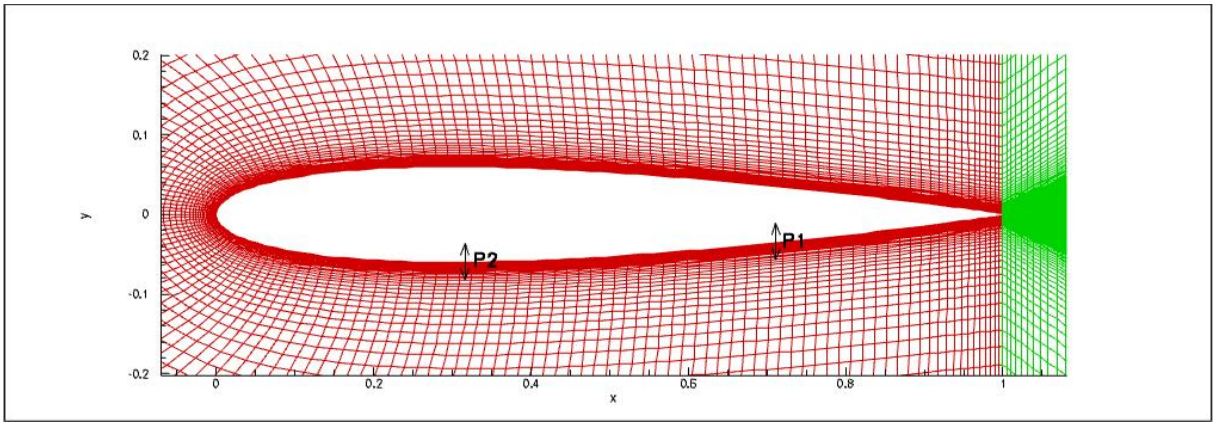

Figure 4: NACA0012 mesh

5.2 Flow Computation

The Reynolds averaged Navier-Stokes (RaNS) equations are considered. The turbulent viscosity is defined by the Wilcox's k-ω model[39]. The seven-equation non-linear system

is solved numerically by the ONERA finite-volume cell-centered code for structured mesh, *elsA*[6]. Second order Roe-flux (using the MUSCL approach with the Van Albada limiting function) is used for mean flow convective terms, centered fluxes with interface-centered evaluation of gradients are used for both diffusive terms. A centered formula is used for the source term of turbulent variables equations.

5.3 Design Space

As it can be viewed above, only two shape parameters, located on the lower surface, are used for this test case. These two parameters are two pilot points of a Free-Form Deformation (FFD) of the volume mesh[5]. The nominal configuration corresponds to parameter values of 0.5. These parameters are both allowed to vary within the interval $[0.25, 0.75]$.

A sampling of 11×11 design points is considered. All corresponding flows are computed with the same set of numerical parameters. The drag coefficient C_D and the lift coefficient C_L are determined in all the design space. The reconstruction of these functions is depicted in figures 5 and 6.

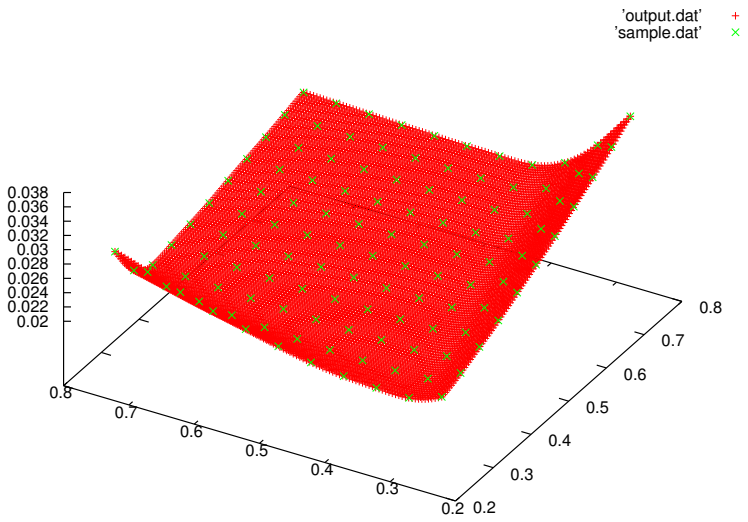

Figure 5: Drag coefficient in the design space

At each point, the derivatives of the aerodynamic functions are computed both by finite differences and adjoint equation. These methods provide somewhat different results due to the constant-turbulent viscosity assumption made in the adjoint approach. Their comparison is given in figure 7.

The next results are given by considering only the drag coefficient which is the most critical of these functions. A subset of the initial sample is selected by Hammersley

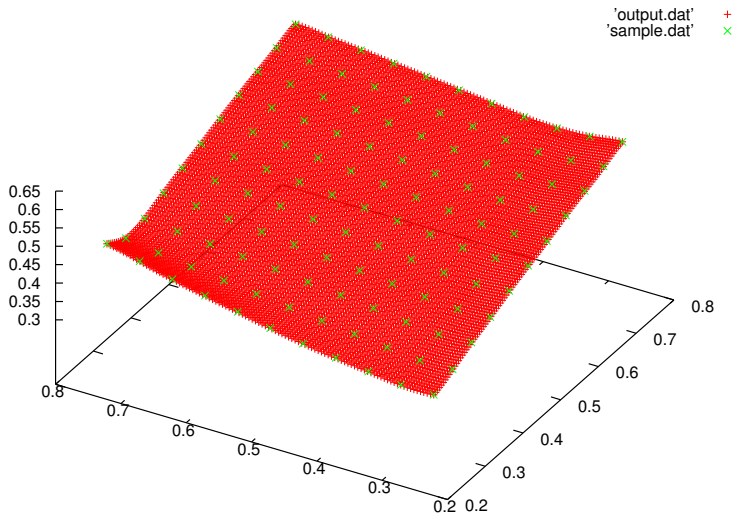

Figure 6: Lift coefficient in the design space

method, and a surrogate model is built from these points. The accuracy of the different models is compared in the table below. The sample size required to reduce the L_2 error to 5 percent of the function mean value is given for each surrogate model. For gradient-assisted models, the derivatives are given here by finite centered differences (DF) or approximated by the adjoint vector method (ADJ).

Kriging	RBF	Co-Kriging^{df}	GA-RBF^{df}	Co-Kriging^{adj}	GA-RBF^{adj}
21	21	8	8	10	12

Table 2: Results of surrogate assisted optimization on Rosenbrock 2D

With the exact derivatives, gradient-assisted methods outperforms the classical surrogate models. But, in practice, the cost of computation of a derivative by finite differences is similar to the cost of the computation of $2d$ function values. The interest of the use of gradients is effective if the derivatives are computed by the adjoint vector method. In this case, the cost of computation of a gradient and of a function value is similar. The benefit of including derivatives is likely to be more substantial with increasing dimension. Several additional applications must be made in the next months to confirm these results.

6 CONCLUSIONS AND FUTURE WORK

The applications of the previous sections seem to show a benefit to include derivatives in the computation of meta models. First of all, this benefit must be confirmed on different aerodynamic test cases with more irregular aerodynamic functions. Then, performances must be checked on higher dimensional spaces, because the higher the dimension, the greater the expected benefit of the adjoint vector method.

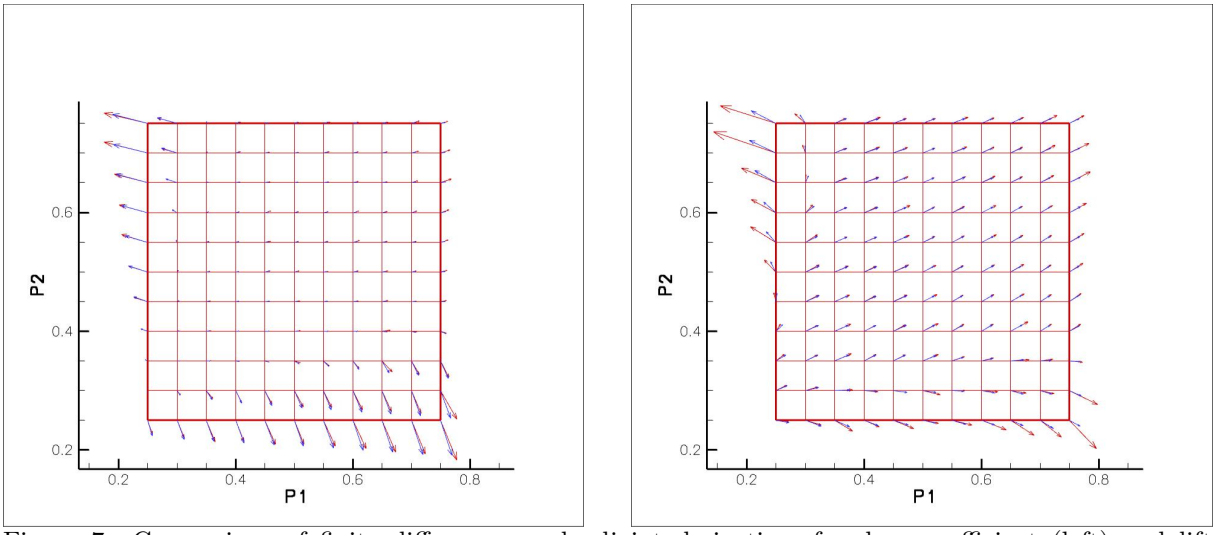

Figure 7: Comparison of finite differences and adjoint derivatives for drag coefficient (left) and lift coefficient (right) on design space

The survey of literature concerning the use of Support Vector Regression in the context of surrogate assisted global optimization has shown a very limited use. In particular, there is today no enrichment criteria developed from statistical assumptions for SVR. This subject can be one part of the future work.

Finally, the improvement of tolerance to gradient noise is one important issue following our work. Indeed, the derivatives computed by the adjoint vector method with the elsA solver have a bias of between 5 percents and 20 percents due to the hypothesis required for derivating the adjoint equation for (RaNS) flows. On-going work is currently devoted to evaluate the impact of this bias in the accuracy of the model.

REFERENCES

[1] N. Benoudjit, C. Archambeau, A. Lendasse, J. Lee, M. Verleysen, Width optimization of the Gaussian kernels in Radial Basis Function Networks, *European Symposium on Artificial Neural Networks proceedings*, 425–432 (2002)

[2] D. S. Broomhead and D. Lowe, Multi-variable functional interpolation and adaptive networks, *Complex Systems*, **2**, 321 (1988)

[3] V. Cherkassky and Y. Q. Ma, Selection of meta-parameters for support vector regression, *ICANN '02: Proceedings of the International Conference on Artificial Neural Networks*, 687–693, London, UK, Springer-Verlag (August 2002)

[4] H. S. Chung and J. J. Alonso, Using gradients to construct cokriging approximation models for high-dimensional design optimization problems, *AIAA Paper*, **317**, 14–17 (2002)

[5] J. A. Désidéri and A. Janka, Multilevel Shape Parametrization for Aerodynamic Optimization - Application to Drag and Noise Reduction of a Transonic/Supersonic Business Jet, *Proceedings of ECCOMAS 2004*, Jyvsakyla (2004)

[6] L. Cambier and M. Gazaix, elsA: An Efficient Object-Oriented Solution to CFD Complexity, *AIAA Paper 2002-0108* (2002)

[7] K. T. Fang, Experimental Design By Uniform Distribution, *Acta Mathematice Applicatae Sinica*, **3**, 363–372 (1980)

[8] A. Forrester and A. Keane, Recent advances in surrogate-based optimization, *Progress in Aerospace Sciences*, (January 2009)

[9] K. C. Giannakoglou, D. I. Papadimitriou and I. C. Kampolis, Aerodynamic shape design using evolutionary algorithms and new gradient-assisted metamodels, *Computer Methods in Applied Mechanics and Engineering*, **195**, 44-47, 6312–6329, Elsevier (2006)

[10] S. R. Gunn, Support Vector Machines for Classification and Regression, *University of Southampton* (1998)

[11] E. H. Isaaks and R. M. Srivastava, An introduction to applied geostatistics, *Oxford University Press New York* (1989)

[12] A. Jameson, Aerodynamic design via control theory, *Journal of Scientific Computing*, **3**, 233–260 (1988)

[13] Jayadeva, Reshma Khemchandani and Suresh Chandra, Regularized Least Squares Twin SVR for the Simultaneous Learning of a Function and its Derivative, *IJCNN*, 1192-1197 (2006)

[14] D. R. Jones, M. Schonlau and W. J. Welch, Efficient Global Optimization of Expensive Black-Box Functions, *Journal of Global Optimization*, **13**, 4, 455-492, Kluwer Academic Publishers, Hingham, MA, USA (1998)

[15] D. R. Jones, A Taxonomy of Global Optimization Methods Based on Response Surfaces, *Journal of Global Optimization*, **21**, 4, 345–383, Kluwer Academic Publishers, Hingham, MA, USA (2001)

[16] J. R. Kalagnanam and U. M. Diwekar, An Efficient Sampling Technique for Off-Line Quality Control, *Technometrics*, **39**, 3, 308–319 (1997)

[17] D. G. Krige, A statistical approach to some mine valuations and allied problems at the Witwatersrand, *University of Witwaterstrand*(1951)

[18] J. Laurenceau and P. Sagaut, Building Efficient Response Surfaces on Aerodynamic Functions with Kriging and Cokriging, *AIAA Journal*, **46**, 2, 498–507, (2008)

[19] M. Lazaro, I. Santamaría, F. Perez-Cruz and A. Artés-Rodríguez, Support vector regression for the simultaneous learning of a multivariate function and its derivatives, *Neurocomputing*, **69**, 1-3, 42–61, Elsevier (2005)

[20] W. Liu and S. M. Batill, Gradient-enhanced response surface approximations using Kriging models, *AIAA Paper 2002-5456, AIAA/ISSMO Symposium and Exhibit on Multidisciplinary Analysis and Optimization*, Atlanta, Georgia (September 2002)

[21] G. Matheron, La thorie des variables rgionalises, et ses applications, *Centre de Morphologie Mathmatique de Fontainebleau* (1970)

[22] D. Mattera and S. Haykin, Support vector machines for dynamic reconstruction of a chaotic system, *Advances in kernel methods: support vector learning*, 211–241, MIT Press, Cambridge, MA, USA (1999)

[23] M. D. McKay, R. J. Beckman and W. J. Conover, A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code, *Technometrics*, **21**, 2, 239–245 (1979)

[24] M. J. L. Orr, Recent advances in Radial Basis Function Networks, *Institute for Adaptative and Neural Computation* (1999)

[25] A. B. Owen, Orthogonal Arrays for Computer Experiments, Integration and Visualization, *Statistica Sinica*, **2**, 439–452 (1992)

[26] C. Praveen and R. Duvigneau, Study of some strategies for global optimization using Gaussian process models with application to aerodynamic design, *INRIA Report*, **6964**, INRIA (2009)

[27] N. V. Queipo, R. T. Haftka, W. Shyy, T. Goel and R. Vaidyanathan, Surrogate-based analysis and optimization, *Progress in Aerospace Sciences*, **41**, 1, 1–28 (2005)

[28] S. Rippa, An algorithm for selecting a good value for the parameter c in radial basis function interpolation, *Advances in Computational Mathematics*, **11**, 2-3, 193–210 (1999)

[29] J. Sacks, W. J. Welch, T. J. Mitchell and H. P. Wynn, Design and analysis of computer experiments, *Statistical science*, 409–423, Institute of Mathematical Statistics (1989)

[30] T. W. Simpson, T. M. Mauery, J. K. Korte and F. Mistree, Kriging Models for Global Approximation in Simulation-Based Multidisciplinary Design Optimization, *AIAA Journal*, **39**, 12, 2233–2241 (December 2001)

[31] T. W. Simpson, V. Toropov, V. Balabanov and F. A. C. Viana, Design and Analysis of Computer Experiments in Multidisciplinary Design Optimization: A review of how far we have come–or not, *12 th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference*, American Institute of Aeronautics and Astronautics, 1801 Alexander Bell Drive, Suite 500, Reston, VA, 20191-4344, USA (2008)

[32] A. J. Smola and B. Schlkopf, A tutorial on Support Vector Regression, *NeuroCOLT2* (1998)

[33] A. Sóbester, S. J. Leary and A. J. Keane, On the Design of Optimization Strategies Based on Global Response Surface Approximation Models, *Journal of Global Optimization*, **33**, 1, 31–59, Kluwer Academic Publishers, Hingham, MA, USA (2005)

[34] R. J. Vanderbei, LOQO: An interior point code for quadratic progamming, *Statistics and Operations Research, Princeton University* (1994)

[35] G. N. Vanderplaats and R. M. Hicks, Numerical airfoil optimization using a reduced number of design coordinates, *NASA TMX73 151* (July 1976)

[36] V. Vapnik, The Nature of Statistical Learning Theory, *Springer, New York* (1995)

[37] V. Vapnik and A. Chervonenkis, Theory of pattern recognition, *Nauka, Moscow* (1974)

[38] G. G. Wang and S. Shan, Review of Metamodeling Techniques in Support of Engineering Design Optimization, *ASME Transactions, Journal of Mechanical Design* (2007)

[39] D. C. Wilcox, Re-assessment of the scale-determining equation for advanced turbulence models, *AIAA Journal*, **26**, 1414–1421 (1988)