ECCOMAS Congress 2016
VII European Congress on Computational Methods in Applied Sciences and Engineering
M. Papadrakakis, V. Papadopoulos, G. Stefanou, V. Plevris (eds.)
Crete Island, Greece, 5–10 June 2016

PERFORMANCE COMPARISON OF KRIGING AND SVR SURROGATE MODELS APPLIED TO THE OBJECTIVE FUNCTION PREDICTION WITHIN AERODYNAMIC SHAPE OPTIMIZATION

D. Viúdez-Moreiras[1], E. Andrés-Pérez[2], D. González-Juárez[1] and M. J. Martin Burgos[1]

[1] Fluid Dynamics Branch, Spanish National Institute for Aerospace Technology (INTA),
Ctra. Ajalvir, km. 4. 28850 Torrejón de Ardoz (Spain)
viudezmd@inta.es, dviudezmoreiras@gmail.com
{gonzalezjd,martinmj}@inta.es

[2] Engineering Department, Ingeniería de Sistemas para la Defensa de España S.A. (ISDEFE-INTA) &
Technical University of Madrid (UPM)
Ctra. de Ajalvir Km. 4.5. 28850 Torrejón de Ardoz (Spain)
eandres@isdefe.es, esther.andres@upm.es

Keywords: Surrogate Models, Kriging, Support Vector Regression, Aerodynamic coefficients prediction, computational fluid dynamics, surrogate-based global optimization

Abstract. *All optimization methods depend on some form of internal model of the problem space they are exploring. To build such a model when there are many variables can require a large number of analyses to be carried out. Because of these difficulties, it is now common in aerospace design to manage explicitly the building and adaptation of the internal surrogate model used during optimization. However, it is very difficult to know a priori which surrogate model is more suitable for a specific application. As described above, in this paper, the performance of two surrogate models, Kriging and Support Vector Regression is compared in order to choose the most suitable model targeting a future application within an aerodynamic shape optimization process. The selected test case was the DPW wing, from the AIAA Workshop on drag prediction, also used in the GARTEUR AD-AG52 on "Surrogate-based global optimization for aerodynamic shape design."*

1 INTRODUCTION

Surrogate modeling is referred to a group of techniques that make use of previously obtained sampled data in order to build surrogate models, which aims to predict the value of variables at untried points in the design space. These groups of techniques, also known as metamodeling, have been developed from many different disciplines including statistics, mathematics, computer science, and various engineering disciplines.

Different techniques have been studied in the literature, which can be classified in three categories [1] such as, (I) the response surface method (RSM) with optimization of coefficients for a base function, (II) the neural network approximation (NN) and (III) an estimation method using observed values at sampling locations to compute an estimated value at an optional location in a solution space. Although these all can be used practically in industry, each method has different features that have to be taken into account before the application to a particular problem. Several comparisons among those methods have been previously described in [2–5].

The RSM is one of the very effective approaches for modeling with small numbers of design variables and its solution space is not too complex, being successfully used in some optimization problems in engineering [6–9]. However, the RSM usually requires the assumption of the order of the approximated base function because the approximation process is performed using the least-square method for the function coefficients. Therefore, the knowledge of the qualitative trend of the entire design space is required by the designer, which will sometimes be difficult to determine. This problem will be highlighted as the number of design variables increases.

NN has been used for solving difficult modeling problems [10-11]. NN generally minimizes the sum of the approximation errors at sampling locations, so that the accuracy of the approximated value at a sampling location is relatively high. However, NN implies high computational cost incurred for learning stage and the need for the designer to be skilled or experienced in using NN [2].

Estimation methods such as Kriging (KR) [12–15], Radial Basis Functions (RBF) [16] or Support Vector Regression (SVR) [17,18] usually require more sample points in the solution space than RSM or NN within the training stage, in order to perform an accurate estimation [19,20]. However, they allow to build complex high non-linear models [21-24] which it is very difficult to achieve with RSM or too complex with NN. They are then found to be a valuable tool to support a wide scope of activities in modern engineering design such as chemical and materials engineering [25,26], and other fields such as agriculture and ecology [27,28], medicine [29-31] and economy [32].

One of the main model applications can be found within the aerospace field [17]. Thus, the use of long running expensive simulations in design leads to a fundamental problem when trying to compare and contrast various competing options: is very expensive from the computational resources point of view to analyze all combinations of variables in the design space. This problem is particularly highlighted when using optimization schemes. All optimization methods depend on some form of internal model of the problem space they are exploring. To build such a model when there are many variables can require large numbers of analyses to be carried out. Because of these difficulties it is now common in aerospace design to manage explicitly the building and adaptation of the internal model used during optimization (Surrogate Based Optimization, SBO).

Thus, the performance obtained by the surrogate model in SBO schemes is very important in order to minimize the number of iterations in the design process, which implies expensive Computational Fluid Dynamics (CFD) simulations in a high performance computer (HPC). However, it is very difficult to know a priori which surrogate model is more suitable for a specific application.

This paper presents a comparison between two surrogate techniques, KR and SVR, for the prediction of aerodynamic coefficients and objective functions for different aircraft's wing geometries, in order to have more information for choosing the most suitable surrogate for this kind of application. In order to perform this comparison, both KR and SVR have been applied to the DPW-wing [33, 34], used by the GARTEUR AD/AG-52 Group [35], with the same initial database composed of different geometries. This database is derived with high fidelity CFD simulations for use in the initial stage of the design process. V-fold cross validation [36, 37] has been used to compare both methods.

This paper is structured as follows: Section 2 briefly describes both surrogate model theory, sampling and model validation methodologies used in this research. Section 3 describes the database to be modeled by the surrogate models. Section 4 presents the comparative results between KR and SVR obtained in the application to the wing case described in the previous section. Finally, Section 5 presents the conclusions.

2 SURROGATE MODEL THEORY

This section briefly describes KR and SVR surrogate modelling. Then, the sampling and the model validation methodologies used in this research are stated.

2.1 Kriging (KR)

This section recalls the basic KR algorithm, as an interpolation technique. Full description of KR algorithm and modifications of basic implementation can be found in [1,16,19-21,24].

KR can be seen as a two-step process. First, a regression function f($\mathbf{x}$) is constructed based on the data, and, subsequently, a Gaussian process Z($\mathbf{x}$) is constructed through the residuals [1,38]. Thus, the prediction $\hat{y}(\boldsymbol{x})$ in $\mathbf{x}$ is derived by:

$$\hat{y}(\boldsymbol{x}) = f(\boldsymbol{x}) + Z(\boldsymbol{x}) \tag{1}$$

where f($\mathbf{x}$) is a regression function and Z is a Gaussian process with mean 0, variance σ^2 and a correlation matrix ψ.

Depending on the form of the regression function Kriging has been prefixed with different names. *Simple Kriging* assumes the regression function to be a known constant, i.e., f($\mathbf{x}$)=0. A more popular version is *Ordinary Kriging*, which assumes a constant but unknown regression function f($\mathbf{x}$), and *Universal Kriging*, which assumes other more complex trend functions such as linear or quadratic polynomials. In general, *Universal Kriging* treats the trend function as a multivariate polynomial:

$$f(x) = \sum_{i=1}^{p} \alpha_i \, b_i(x) \tag{2}$$

where $b_i(\boldsymbol{x})$ are $i = (1 \dots p)$ basis functions and $\alpha = (\alpha_1 \dots \alpha_p)$ denotes the coefficients. Therefore, the regression function captures the general trend of the data and the Gaussian Process interpolates the residuals. However, selecting the correct regression function is a difficult problem, hence, the regression function is often chosen constant.

Consider a set of n samples, X = { $\mathbf{x^1}, \dots, \mathbf{x^n}$ } in *d* dimensions and associated function values, $\mathbf{y}$ = { $y^1, \dots, y^n$ }. Essentially, the regression part is encoded in the n x p model matrix F:

$$F = \begin{pmatrix} b_1(\boldsymbol{x^1}) & \cdots & b_p(\boldsymbol{x^1}) \\ \vdots & \ddots & \vdots \\ b_1(\boldsymbol{x^n}) & \cdots & b_p(\boldsymbol{x^n}) \end{pmatrix}$$

while the stochastic process is mostly defined by the n x n correlation matrix ψ,

$$\psi = \begin{pmatrix} \psi(\boldsymbol{x^1}, \boldsymbol{x^1}) & \cdots & \psi(\boldsymbol{x^1}, \boldsymbol{x^n}) \\ \vdots & \ddots & \vdots \\ \psi(\boldsymbol{x^n}, \boldsymbol{x^1}) & \cdots & \psi(\boldsymbol{x^n}, \boldsymbol{x^n}) \end{pmatrix}$$

where $\psi(\boldsymbol{x^i}, \boldsymbol{x^j})$ is the correlation function. $\psi(\boldsymbol{x^i}, \boldsymbol{x^j})$ is parametrized by a set of hyperparameters $\boldsymbol{\theta}$, which are identified by Maximum Likelihood Estimation (MLE) [39,40]. Subsequently, the prediction mean and prediction variance of KR are derived, respectively, as,

$$\mu(\boldsymbol{x}) = M\alpha + r(\boldsymbol{x})\, \psi^{-1}\, (\mathbf{y} - \mathbf{F}\alpha) \tag{3}$$

$$s^2(\boldsymbol{x}) = \sigma^2 \left(1 - r(\boldsymbol{x})\, \psi^{-1} r(\boldsymbol{x})^T + \frac{1-(F^T\, \psi^{-1} r(x)^T}{F^T\, \psi^{-1}\, F} \right) \tag{4}$$

where $M = (\, b_1(\boldsymbol{x}) \dots b_p(\boldsymbol{x})\,)$ is the model matrix of the predicting point **x**, and $r(\boldsymbol{x}) = (\, \psi(\boldsymbol{x}, \boldsymbol{x^1}) \dots \psi(\boldsymbol{x}, \boldsymbol{x^n})\,)$ is an 1 x n vector of correlations between the point **x** and the samples X. The process variance σ^2 is given by:

$$\sigma^2 = \frac{1}{n} (\mathbf{y} - \mathbf{F}\alpha)^T\, \psi^{-1}\, (\mathbf{y} - \mathbf{F}\alpha) \tag{5}$$

and the coefficients of the regression function, α, are determined by Generalized Least Squares (GLS) by:

$$\alpha = (\, F^T\, \psi^{-1}\, F\,)^{-1}\, F^T\, \psi^{-1} \boldsymbol{y} \tag{6}$$

2.2 Support Vector Machines (SVR)

SVR can be solved as a convex optimization problem using kernel theory to face non-linear problems. Thus, SVR consider not only the prediction error but also the generalization of the model [17,41].

SVR consist of training a model with the form $y = w^T\phi(x) + b$ given a set of parameters $C = \{(x_i, y_i)\}, i = 1,2, \dots, l\}$, to minimize a general risk function of the form:

$$R[f] = \frac{1}{2}\left|\left|w\right|\right|^2 + \frac{1}{2} C \sum_{i=1}^{l} L\big(y_i, f(x)\big) \tag{7}$$

where w controls the smoothness of the model, $\phi(x)$ is a function of projection of the inputs space w to the feature space, b is a parameter of bias, x_i is a feature vector of the input space with dimension N, y_i is the output value to be estimated and $L\big(y_i, f(x)\big)$ is the loss function selected. In this study, the $L1$ support vector regression $(L1 - SVR)$ is used,

$$L\big(y_i, f(x)\big) = |y_i + f(x_i)|_\varepsilon \quad (8)$$

In order to train this model, it is necessary to solve the following optimization problem:

$$\min\left(\frac{1}{2}\left||w|\right|^2 + \frac{1}{2}C\sum_{i=1}^{l}\xi_i + \xi_i^*\right) \quad (9)$$

subject to:

$$\begin{gathered} \mathrm{y_i} - w^{\mathrm{T}}\,\phi(x) - b \leq\ \varepsilon + \xi_i, i = 1, \dots, l \\ -\mathrm{y_i} + w^{\mathrm{T}}\,\phi(x) + b \leq\ \varepsilon + \xi_i^*, i = 1, \dots, l \\ \xi_i\xi_i^* \geq 0, i = 1, \dots, l \end{gathered} \quad (10)$$

To do this, a dual form is usually applied, obtained from the minimization of the Lagrange function that joins the function to minimize and the constraints:

$$\begin{aligned} \max\Bigg(& -\frac{1}{2}\sum_{i,j=1}^{l}(\alpha_i + \alpha_i^*)\big(\alpha_j + \alpha_j^*\big)K\big(x_i + x_j\big) \\ & - \ \varepsilon\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) + \sum_{i=1}^{l}y_i(\alpha_i + \alpha_i^*)\Bigg) \end{aligned} \quad (11)$$

subject to the following constraint:

$$\sum_{i=1}^{l}(\alpha_i + \alpha_i^*) = 0, \alpha_i, \alpha_i^* \in [0, C] \quad (12)$$

In addition to the constraints, also must be taken in account the Karush-Kuhn-Tucker conditions and obtain the bias value. In the dual formulation, the apparition of the kernel function $K\big(x_i, x_j\big)$ must be emphasized, which is equivalent to the scalar product$\big\langle\phi(x_i\ )\big|\phi(x_j)\big\rangle$. In this case, the kernel function is a Gaussian function:

$$K = \exp(-\gamma \cdot \left|\left|x_i - x_j\right|\right|^2) \quad (13)$$

The final form of the regression model depends on the Lagrange multipliers α_i, α_i^*, following the expression:

$$f(x) = \sum_{i=1}^{l}(\alpha_i - \alpha_i^*)K(x_i, x) + b \quad (14)$$

In this way, SVR model depends on three parameters, ε, C and γ. (I) ε parameter controls the error margin permitted for the model, as can be seen in equations (9) and (10). (II) C parameter controls the number of outliers allowed on the optimization of the equation (3). Finally, (III) γ parameter determines the Gaussian variance for the kernel. Depending on the selection of these values, the model can have different performance. To obtain the best SVR performance, search of the most suitable combination of these three parameters must be performed, usually by means of cross validation techniques over the training set, as explained in Section 2.4. To reduce the computational time of this process, different methods have been proposed in the literature to reduce the search space related to these parameters. In this case, it has been applied the one developed in [41], which has proven to require pretty short search times.

2.3 DESIGN OF EXPERIMENTS (DOE) METHODOLOGY

Both KR and SVR surrogate models have been applied to the database generated with the DLR TAU solver [42] for a set of initial wing geometries. Therefore, a suitable Design of Experiments (DoE) technique, that envelopes the design space, is required.

When the initial database is produced by a deterministic computer code, as opposed to a physical experiment or stochastic analysis, a given input will always yield the same output, because there is no measurement error or other random sources of noise. Under these conditions, the DoE need only be space-filling [43,44] so that all regions of the design space Ω as a subset of $\mathbb{R}^{NS}$, being NS the number of independent variables of the design space, are sampled. A commonly used space-filling design is Latin hypercube sampling (LHS) [45], which has been used as sampling methodology in this study in order to perform the comparison between KR and SVR. In LHS, each input parameter is partitioned into N equally spaced sections. Each input parameter is sampled once in each section, resulting in a column vector $\boldsymbol{x}_{\mathrm{i}}$ containing NS different values of the input parameter. The column vectors for each input parameter are arranged side by side into a matrix and the components of the vectors are then randomly reordered. The resulting training set $X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_N\}$ is a matrix of size N x NS, known as a Latin hypercube, in which each row corresponds to a different training case defined by the input parameters.

In this particular case, NS columns correspond to the z-axis coordinates of different control points distributed by the wing surface to optimize. N sets of coordinates are derived, in order to obtain the aerodynamic coefficients by means of CFD and, in this way, perform the initial database in order to train in the same conditions both surrogate models.

2.4 METRICS FOR MODEL VALIDATION

In most real applications, only a limited amount of data is available, which leads to the idea of splitting the data: part of data, the *training sample*, is used for training the algorithm, and the remaining data, the *validation sample*, is used for evaluating the performance of the algorithm. The *validation sample* can play the role of “new data”.

A single data split yields a validation estimate of the risk, and averaging over several splits yields a cross-validation estimate. V-fold CV (VFCV) was introduced in [46] and now is widely used as a model validation technique within surrogate modelling. Thus, the methodology used in this paper to obtain the model performance in order to obtain the surrogate models comparison is, for a training set $X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_N\}$, the following:

a. Partition the training set X into K independent equal-sized subsets X_k, $k = 1, \dots, K$ such as $X = X_1 \cup X_2 \cup \dots \cup X_K$;
b. for $k = 1, \dots, K$
    i. train the prediction model on $X_{(-k)} = X \backslash X_k$;
    ii. test the prediction model on X_k

iii. $\forall\, \boldsymbol{x}_i \in X_k$ compute the values of $\hat{y}^{-k(i)}(\boldsymbol{x}_i)$,

iv. Derive the quadratic error between the real y_i and the estimated $\hat{y}^{-k(i)}(\boldsymbol{x}_i)$ values:

$$\text{SE}= \left(y_i - \hat{y}^{-k(i)}(\boldsymbol{x}_i)\right)^2 \tag{15}$$

c. Estimate the following metrics:

i. Mean Squared Error of $\hat{y}$

$$MSE(\hat{y}) = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \hat{y}^{-k(i)}(\boldsymbol{x}_i)\right)^2 \tag{16}$$

ii. R-squared

$$R^2 = 1 - \frac{\sum_i (y_i - y_i)^2}{\sum_i (y_i - \bar{y}_i)^2} \tag{17}$$

where $\bar{y}$ is the mean value of the true values y_i.

iii. Pearson's correlation coefficient

$$\rho = \frac{N\sum_{i=1}^{N} y_i \hat{y}^{-k(i)}(x_i) - \sum_{i=1}^{N} y_i \sum_{i=1}^{N} \hat{y}^{-k(i)}(x_i)}{\sqrt{N\sum_{i=1}^{N} y_i{}^2 - \left(\sum_{i=1}^{N} y_i\right)^2}\sqrt{N\sum_{i=1}^{N}\left[\hat{y}^{-k(i)}(x_i)\right]^2 - \left[\sum_{i=1}^{N}\hat{y}^{-k(i)}(x_i)\right]^2}} \tag{18}$$

The MSE metric gives an estimate of the expected test error by using the squared error as loss function. It ranges between zero and plus infinity, smaller values indicate smaller errors. The Pearson's correlation coefficient ranges between -1 and 1 and provides the ratio between the covariance of y and $\hat{y}$ and the product of their standard deviations. In other words, it measure the tendency of $\hat{y}$ to vary in function of y. If ρ is close to zero, y and $\hat{y}$ are weakly correlated and, hence, is expected that the prediction model $\hat{y}$ badly reproduces the variation of the function y. On the other hand, if ρ approaches the unity value, a strong correlation between the variables and the two datasets y_i and $\hat{y}^{-k(i)}(x_i)$ is obtained. If ρ is close to -1, anti-correlation exists and is expected that, for positive variation of y, negative variation of $\hat{y}$ is obtained.

3 AERODYNAMIC DATABASE USED FOR SURROGATE COMPARISON

This section describes how the database, to be used for comparison and validation purposes, was generated from the DPW-W1 wing baseline geometry [33], whose reference quantities for this wing are displayed in Table 1.

Reference quantity	Value
S_{ref} (wing reference area)	290,322 mm^2
C_{ref} (wing reference chord)	197.55 mm
X_{ref} (relative to the wing root leading edge)	154.24 mm
b/2 (semi span)	762 mm
AR (aspect ratio, $AR=b^2/S_{ref}$)	8

Table 1: Reference quantities for the DPW wing

As can be seen in Fig. 1, in order to generate the database, the DPW geometry was parameterized by a 3D control box, with 5 control points in direction **u**, 10 in direction **v** and 5 in direction **w**. The parametric **u** direction corresponds to the y-axis, the **v** direction to the x-axis, and the **w** direction to the z-axis. The vertical displacement of those control points, which correspond to the design space, set up on the aerodynamic surface. The wing is split in three profile sections and the transition between sections is linear. Each section has 6 active control points for the upper side and other 6 for the lower side, which are independent (the movement of a control point at the upper side does not modify the lower side and vice versa), with a total of NS=36 design variables for the whole wing.

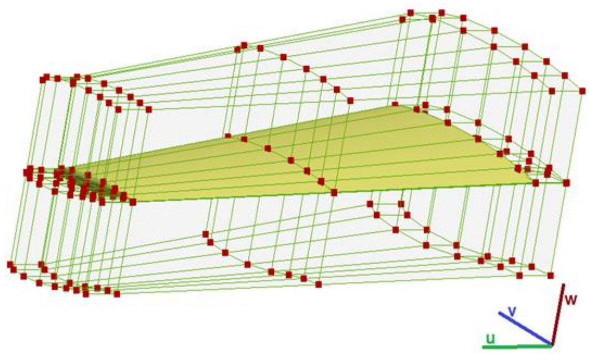

Figure 1: DPW wing parameterization

The initial database X was performed according to the methodology described in previous sections, executing 180 cases for position changes of the 36 design control points by means of DLR TAU solver in a High Performance Computer platform (HPC). Each design variable has been constrained by its minimum and maximum values that will be chosen as +/- 20% of their original value. In addition, other constraints, such as airfoil's maximum thickness and beam constraints have been defined, according to [34]. The flow conditions are Mach = 0.80, angle of attack 0 deg and Reynolds, $5*10^6$.

The design goal for the SBO is to achieve a geometry with the minimum drag, while maintaining the specified aerodynamic constraints:

1. Prescribed constant lift coefficient ($C_L = C_L^0$).
2. Minimum pitching moment: $C_M \geq C_M^0$.
3. Drag penalty: If constraint in minimum pitching moment is not satisfied, the penalty will be 1 drag count per 0.01 increment in C_M.

Therefore, an objective function was derived, based on the lift, drag and pitching moment aerodynamic coefficients obtained by the DLR TAU code, and the previously mentioned constraints [17]. This objective function will be modeled by the surrogate model in order to predict new values within the SBO scheme.

4 EXPERIMENTAL AND COMPARATIVE RESULTS

Surrogate modeling has been applied to the database X. As described in Section 3, this database has N=180 samples and NS=36 columns, with a +/-20% variation of the design variables from the base geometry. Thus, both KR and SVR methodologies have been applied in the same conditions, validating them with 10-fold CV as described in Section 2.4.

Figure 2 shows the R^2 comparison between KR and SVR. In the KR case, R^2 is 0.8239, 62.7% greater than the value obtained with SVR (0.5064). This difference can be seen to a lesser extent in the Pearson coefficient stated in Table 2, with a 27.6% of variation. In addition, the MSE and the RMSE have a difference of 58.3% and 35.4% respectively.

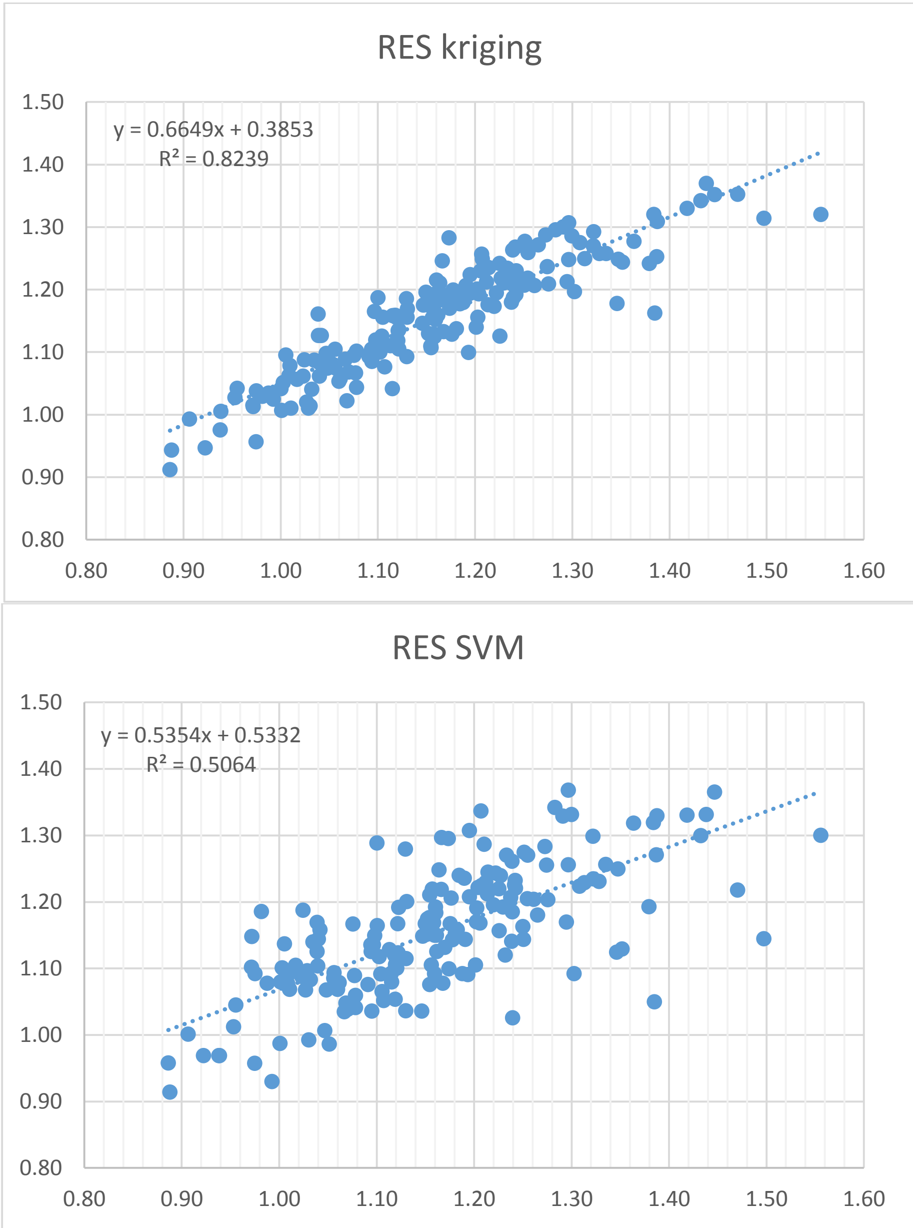

Figure 2: R^2 comparison between KR (above) and SVR (below).

Thus, KR seems to offer a better performance than SVR, applied to the DPW prediction of the objective function. This study allows to select the best surrogate model between SVR and KR based on experimental data, due to it is very difficult to know a priori which surrogate model is more suitable for a specific application.

	SVR	KRIGING	Difference
MSE	7.8093E-03	3.2580E-03	58.3%
RMSE	8.8370E-02	5.7079E-02	35.4%
Pearson	0.7116	0.9077	27.6%
R^2	0.5064	0.8239	62.7%

Table 2: Summary of performance comparison between KR and SVR.

5 CONCLUSIONS AND FUTURE WORKS

This paper describes the comparison between two different surrogate models, Kriging (KR) and Support Vector Regression Machines (SVR), for prediction of the objective function within different aerodynamic configurations, with the aim of aerodynamic optimization. In order to carry out this comparison, both KR and SVR have been applied to the same training database generated with CFD simulations for different geometries. V-fold cross validation has been used to compare both methods, showing the better performance obtained by KR methodology.

Future work will focus on studying the relationship between the model performance and the number of design variables and other parameters with influence, in order to minimize the required CFD simulations for the initial database, and therefore speed up the design process.

REFERENCES

[1] S. Sakata, F. Ashida and M. Zako, An efficient algorithm for Kriging approximation and optimization with large-scale sampling data, Computer methods in applied mechanics and engineering, 2003, pp. 386–404.

[2] W.C. Carpenter, J.F.M. Barthelemy, A Comparison of polynomial Approximations and Artificial Neural Nets as Response Surfaces, A collection of technical papers: the 33rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, AIAA, Dallas, TX, April 13–15, 1992, pp. 2474–2482.

[3] E. Nikolaidis, L. Long, Q. Ling, Neural networks and response surface polynomials for design of vehicle joints, AIAA-98-1777, 1998, pp. 653–662.

[4] N. Papila, W. Shyy, N. Fitz-Coy, R.T. Haftka, Assessment of neural net and polynomial-based techniques for aerodynamic applications, in: AIAA 17th Applied Aerodynamics Conference, AIAA-99-3167, 1999.

[5] R. Jin, W. Chen, T.W. Simpson, Comparative studies of metamodeling techniques under multiple modeling criteria, in: 8th AIAA/NASA/USAF/ISSMO Symposium on Multi-disciplinary Analysis and Optimization, Long Beach, CA, AIAA, 2000, AIAA-2000-4801.

[6] G.E.P. Box, K.B. Wilson. On the experimental attainment of optimum conditions, J. Roy. Statist. Soc. Ser. B 13,1951, 1–45.

[7] T. Kashiwamura, M. Shiratori, Q. Yu, Statistical optimization method, Comput. Aided Optimum Des. Struct. V,1997. 213–227.

[8] I.P. Sobieski, V.M. Manning, I.M. Kroo, Response surface estimation and refinement in collaborative optimization, AIAA-98-4753, 1998, pp. 4753–4770.

[9] S. Hosder, L.T. Watson, B. Grossman, W.H. Mason, H. Kim, R.T. Haftka, S.E. Cox, Polynomial response surface approximations for the multidisciplinary design optimization of a high speed civil transport, Technical Reports of NCSTRL at Virginia Tech CS (TR-01-03), 2001.

[10] S. Thamarai-Selvi, E. Mahendran and S. Rama. Neural Network-Based Interpolation and Extrapolation of Wind Tunnel Test Data Int. J. Comp. Intel. Appl. 08, 225 (2009. DOI: 10.1142/S1469026809002564.

[11] L. Berke, P. Hajela, Application of artificial neural nets in structural mechanics, Struct. Optim. 4,1992, pp. 90–98.

[12] J. Sacks, W.J. Welch, T.J. Mitchell, H.P. Wynn, Design and analysis of computer experiments, Statist. Sci. 4 (4),1989, pp. 409–435.

[13] A. Giunta, L.T. Watson, J. Koehler, A Comparison of approximation modeling techniques: polynomial versus interpolating models, in: 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis & Optimization, St. Louis, MI, September 2–4, AIAA-98-4758-CP, 1998.

[14] D.J. Cappelleri, M.I. Frecker, T.W. Simpson, Design of a PZT bimorph actuator using a metamodel-based approach, J. Mech. Des. Trans. ASME 124,2002, 354–357.

[15] S. Sakata, F. Ashida, M. Zako, Optimization of laminated composites for eigenfrequency problem using K-CSSL method, in: F.- K. Chang (Ed.), Proceedings of the Tenth US–Japan Conference on Composite Materials, DEStech Publishers, 2002, pp. 560–569.

[16] C. Praveen., R. Duvigneau. Radial Basis Functions and Kriging Metamodels for Aerodynamic Optimization. INRIA Technical report 2007.

[17] E. Andres, S. Salcedo-Sanz, F. Monge, A.M. Perez-Bellido. Efficient aerodynamic design through evolutionary programming and support vector regression algorithms. Expert Systems with Applications 39,2012. pp. 10700–10708.

[18] S. M. Clarke, J. H. Griebsch, and T. W. Simpson. Analysis of Support Vector Regression for Approximation of Complex Engineering Analyses. Transactions of ASME, Journal of Mechanical Design, 127(6), 2005, pp. 1077-1087.

[19] T.W. Simpson, T.M. Mauery, J.J. Korte, F. Mistree, Comparison of Response Surface and Kriging Models for Multidisciplinary Design Optimization, 7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis & Optimization, St. Louis, MI, September 2–4, AIAA-98-4755, 1998.

[20] S. Sakata, F. Ashida, M. Zako, Structural optimization using Kriging approximation, Comput. Methods Appl. Mech. Engrg. 192 (7–8), 2003, pp. 923–939.

[21] J.P.C. Kleijnen. Kriging metamodeling in simulation: A review. European Journal of Operational Research, 192(3), 2009, pp. 707–716.

[22] J. Sacks, S. B. Schiller, and W. J. Welch. Designs for computer experiments. Technometrics, 31(1):41–47, February 1989.

[23] J. Sacks, W. J. Welch, T. J. Mitchell, and H. P. Wynn. Design and analysis of computer experiments. Statistical Science, 4(4):409–423, November 1989.

[24] W. Shao, H. Deng, Y. Ma, and Z. Wei. Extended Gaussian kriging for computer experiments in engineering design. Engineering with Computers, 28(2):161–178, 2012.

[25] E. Içten, Z.K. Nagy, G.V. Reklaitis, Process control of a dropwise additive manufacturing system for pharmaceuticals using polynomial chaos expansion based surrogate model, *Computers & Chemical Engineering*, 83, 221-231,2015.

[26] J. Steuben, J. Micholopoulos, A. Iliopoilos, C. Turner, Inverse characterization of composite materials via surrogate modeling, *Composite Structures*, 132, 694-708, 2015.

[27] X. Wu, Y. Zheng, B. Wu, Y. Tian, F. Han, C. Zheng, Optimizing conjunctive use of surface water and groundwater for irrigation to address human-nature water conflicts: A surrogate modeling approach. *Agricultural Water Management*, 163, 380-392, 2016 (in press).

[28] J.C. Pierson, A. Mortelliti, P.S. Barton, P.W. Lane, D.B. Lindenmayer, Evaluating the effectiveness of overstory cover as a surrogate for bird community diversity and population trends, *Ecological Indicators*, 2015.

[29] M.K. Song, S.E. Ward, J.P. Fine, L.C. Hanson, F.C. Lin, G.A. Hladik, J.B. Hamilton, J.C. Bridgman, Advance Care Planning and End-of-Life Decision Making in Dialysis: A Randomized Controlled Trial Targeting Patients and Their Surrogates, 66 (5), 813-822, 2015.

[30] K. Bollaerts, M. Snock, K. Simons, S. Fierens, A. Poffijn, L.Van Bladel, D. Geraets, P. Gosselin, H. Van Oyen, J. Francart, A. Van Nieuwenhuyse, Thyroid cancer incidence around the Belgian nuclear sites: Surrogate exposure modelling, *Cancer Epidemiology*, 130, 12-26, 2014.

[31] A.M. Osiceanu, L.E. Murao, D. Kollanur, J. Swinnen, A.R. De Vleeschuwer, D.J. Lefebvre, K. De Clerq, J. Neyts, N. Goris, In vitro surrogate models to aid in the development of antivirals for the containment of foot-and-mouth disease outbreaks, *Antiviral Research*, 105, 59-63, 2014.

[32] Y. Yin, P. Shang, Modified cross sample entropy and surrogate data analysis method for financial time series, *Physica A: Statistical Mechanics and its Applications*, 433, 17-25, 2015.

[33] "AIAA 3rd Drag Prediction Workshop. URL: http://aaac.larc.nasa.gov/tsab/cfdlarc/aiaa-dpw/Workshop3/workshop3.html," [Online].

[34] B. Epstein, A. Jameson, S. Peigin, D. Roman, N. Harrison and J. Vassberg, "Comparative Study of 3D Wing Drag Minimization by Different Optimization Techniques.," in 46th AIAA Aerospace Sciences Meeting and Exhibit. AIAA paper 2, Reno, Nevada, 2008.

[35] GARTEUR AD/AG-52 Surrogate Based Global Optimization methods in preliminary Aerodynamic Design. Garteur group, 2015.

[36] Y. Bengio, and Y. Grandvalet. No unbiased estimator of the variance of K-fold cross-validation. J. Mach. Learn. Res., 5, 2004, pp.1089–1105

[37] P. Burman. A comparative study of ordinary cross-validation, v-fold cross-validation and the repeated learning-testing methods. Biometrika, 76(3), 1989, pp.503–514.

[38] I. Couckuyt, T. Dhaene, and P. Demeester. ooDACE toolbox: A flexible object-oriented kriging implementation. Journal of Machine Learning Research, 15:3183–3186, 2014.

[39] S.M. Stigler. Statistics on the table: the history of statistical concepts and methods. Harvard University Press, 1999. ISBN 0-674-83601-4.

[40] C. Le, L. Lucien and G.Yang. Asymptotics in statistics: some basic concepts (Second ed.). Springer, 2000. ISBN 0-387-95036-2.

[41] E. Ortiz-García, S. Salcedo Sanz, Á. Pérez-Bellido and J. Portilla-Figueras, "Improving the training time of support vector regression algorithms through novel hyper-parameters searchspace reductions," Neurocomputing, vol. 72, p. 3683–3691, 2009.

[42] DLR Institute of Aerodynamics and Flow Technology. "Technical Documentation of the DLR TAU-Code". Braunschweig, Germany, 2004.

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

[44] T. W. Simpson, A. J. Booker, D. Ghosh, , A. A. Giunta, , P. N. Koch, , and R. J. Yang, "Approximation Methods in Multidisciplinary Analysis and Optimization: A Panel Discussion," Structural and Multidisciplinary Optimization, Vol. 27, No. 5, July 2004, pp. 302–313. doi:10.1007/s00158-004-0389-9.

[45] 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, Vol. 21, No. 2, May 1979, pp. 239–245. doi:10.2307/1268522.

[46] S. Geisser. The predictive sample reuse method with applications. J. Amer. Statist. Assoc., 70, 1975. pp. 320–328.