Study of the influence of the initial a-priori training dataset size in the efficiency and convergence of surrogate-based evolutionary optimization

Daniel González-Juárez*
Fluid Dynamic Brach, National Institute for Aerospace Technology (INTA)
Ctra Ajalvir, km 4, 28850 Torrejón de Ardoz, Spain
Email: gonzalezjd@inta.es

Esther Andrés-Pérez
Engineering Department, Ingeniera de Sistemas para la Defensa de España S.A. (ISDEFE-INTA) & Universidad Politécnica de Madrid (UPM)
Ctra Ajalvir, km 4, 28850 Torrejón de Ardoz, Spain
Email:eandres@isdefe.es, esther.andres@upm.es

Summary

The development of an automatic geometry optimization tool for efficient aerodynamic shape design, supported by Computational Fluid Dynamic (CFD) methods is nowadays an attractive research field, as can be observed from the increasing number of scientific publications during the last years. Surrogate-based global optimization methods have demonstrated a huge potential to reduce the actual number of CFD runs, and therefore drastically speed-up the design process. Nevertheless, surrogates need initial high fidelity data sets to be built and to reach a proper accuracy. This work presents a study on the influence of the initial training dataset size in the proposed approach behavior. This approach is based on the use of Support Vector Machines (SVMs) as the surrogate model for estimating the objective function, in combination with an Evolutionary Algorithm (EA) and an adaptive sampling technique focused on optimization called the Intelligent Estimation Search with Sequential Learning (IES-SL). Several number of training points have been fixed to check the convergence, the accuracy and the objective function reached by the method.

Keywords: *aerodynamic shape design, evolutionary optimization, computational fluid dynamics, surrogate-based optimization, surrogate modelling.*

1 Introduction

Aerodynamic shape optimization by means of automatic tools is an industrial relevant field that has to breast several challenges. Some of these challenges are: how to handle deformations in certain regions (such as intersections between wing and fuselage or pylon/nacelle), how to reduce the number of CFD runs required for performing aerodynamic design optimization or how to tackle integrated components. Furthermore, surrogate-based optimization methods require several barriers to be broken when applied to complex configurations, such as the called "curse of dimensionality", the ability of surrogates to handle a high number of design parameters, efficient constraints handling[1], and the proper exploration and exploitation of the whole design space.

In the case of surrogate-based optimization (SBO) methods, the surrogate prediction is also highly influenced by training set size. A huge training set with a proper design space distribution ensures reaching a global optimum, but requires a vast computational cost to be built. On the other hand, a small training set is fast to be built but the accuracy is not enough for optimization purposes. A solution to this issue must be found for the suitable implementation of this method in the aeronautical industry.

In this work, Support Vector Machines (SVM) combined with Evolutionary Algorithms (EAs) and an adaptive sampling method, called Intelligent Estimation Search with Sequential Learning (IES-SL), is proposed. The approach is applied to the multipoint optimization of one typical test case, i.e., the transonic RAE 2822 airfoil. The aim of this work is to provide an analysis of the training set size influence in the behavior of the IES-SL approach proposed.

This paper is structured as follows. In Section 2, a review of the recent research efforts in SBO applied to aircraft design is presented. Section 3 presents the applied SBO strategy and Section 4 collects the study results. Finally, the conclusions extracted from the results are summarized in Section 5.

2 Literature review

2.1. Recent research efforts in SBO applied to aircraft design

Some recent efforts in SBO for aerodynamic shape design includes, e.g., a physics-based surrogates applied to the drag minimization of NACA 0012 and RAE 2822 airfoils in transonic flow conditions[2]. In this work, the geometries were parameterized using PARSEC involving 5 to 10 design parameters. SBO strategies were applied for the drag minimization of the NLF0416 airfoil using 10 design variables[3]. Variable-fidelity computational fluid dynamics (CFD) combined with shape optimization strategy was applied to the optimization of a transonic airfoil parameterized by the NACA 4-digit definition with three design variables[4].

A surrogate based on proper orthogonal decomposition (POD) applied to the aerodynamic shape optimization of an airfoil is presented by Iuliano[5]. The geometry was parameterized with 16 design variables defined with the CST method. An approach based on a combination of a genetic algorithm and an artificial neural network is presented by Jahangirian[6]. This approach was applied to the shape optimization of an airfoil, which was parameterized by a modified PARSEC involving 10 design variables.

Most of the SBO applications in aerodynamic shape optimization involve two-dimensional configurations, where the number of design variables is usually limited. Nevertheless, some applications to three-dimensional configurations can be found in literature. An investigation about SBO applied to a wing parameterized with 11 design variables was undertaken by Keane[7] .A multi-fidelity surrogate model applied to a three-dimensional wing optimization was addressed by Likeng[8] [8]. In this case, the design parameters were a combination of 12 variables using the CST method for three wing sections (root, hink and wing tip). Lukaczyk[9] , proposed a method based on an active subspace for effectively searching the whole design space. The method is applied to the optimization of the ONERA M6 transonic wing, which was parameterized with 50 FFD design variables. The aim was to discover a low-dimensional linear subspace of the input space that explained the majority of the variability in the drag and lift coefficients. An SBO application to the aerodynamic shape design of a wing parameterized with volumetric non-uniform rational B-splines (NURBS) was presented by current authors[10]. Also, in [11, 12] current authors present an application study about the influence of number and location of the design parameters in the behaviour of the IES-SL method applied to the aerodynamic shape optimization. The selected geometries, RAE 2822 airfoil and DPW-w1 wing, were parameterized with volumetric NURBS.

This work is within the aerodynamic shape design and optimization research line of INTA's Fluid Dynamics Branch.

3 Surrogate-based optimization strategy

This section introduces each of the components of the SBO approach applied in this study: geometry parameterization through volumetric NURBS, Evolutionary Algorithms (EAs), Support Vector Machines for Regression (SVR) and the Intelligent Estimation Search with Sequential Learning (IES-SL) as the strategy for adaptive sampling focused on optimization.

3.1. *Geometry parameterization*

Parameterization is a crucial step in an aerodynamic design optimization problem. NURBS have demonstrated to be able to accurately represent a large family of geometries. In aerodynamic design, NURBS provide smooth surfaces while maintaining some deformation locality[13]. In addition, the optimized surface at the end of the optimization process has the correct format to feed directly the CAD and grid generation applications. However, the use of surface NURBS can be impractical, because very frequently requires the additional effort to develop a surface representation that fits the original geometry, with an appropriated arrange of control points for the optimization. An alternative approach is to envelop the geometry in a volumetric NURBS[14], which maintain the deformation properties of a conventional 2-dimensional surface, but with the advantage that control points can be set up arbitrarily.

From a mathematical point of view, NURBS surfaces are defined as the tensor product of three NURBS curves, defining a volumetric region, where the deformation is governed by the movement of control points:

$$S(\xi,\eta,\mu)=\frac{\sum_i^I\sum_j^J\sum_k^K U_{i,n}(\xi)\,V_{i,n}(\eta)\,W(\mu)\,C_{ijk}}{\sum_i^I\sum_j^J\sum_k^K U_{i,n}(\xi)\,V_{i,n}(\eta)\,W(\mu)} \qquad (1)$$

where C are the control points, ξ, η, and μ are the parametric coordinates, and U, V, and W are the basis functions which are calculated using the following expression:

$$U_{i,1}(\xi)=\begin{cases}1 \; if\; u_i\le\xi< u_{i+1}\\ 0 \quad otherwise\end{cases}$$
$$U_{i,k}(\xi)=\frac{(\xi-u_i)U_{i,k-1}(\xi)}{u_{i+k-1}-u_i}+\frac{(u_{i+k}-\xi)U_{i+1,k-1}(\xi)}{u_{i+k}-u_{i+1}} \qquad (2)$$

The basis coefficients are calculated from the knot vectors $\overline{U}$, $\overline{V}$ and $\overline{W}$, and, which are a sequence of real numbers. Basis functions are equal to zero everywhere except for an interval delimited by the order of the NURBS, defining the area of influence of each control point[15]. The most common implementation of the control box is to employ uniform basis, which can be obtained with a knot sequence as:

$$\{\underbrace{0,\dots,0}_{p+1},\frac{1}{N}\dots,\frac{i}{N},\dots,\frac{N-1}{N},\underbrace{1,\dots,1}_{p+1}\} \qquad (3)$$

First order is equivalent to a linear interpolation, while second and third orders provide derivative and curvature continuity, respectively.

In this work, the airfoil is parameterized with third order volumetric NURBS, also called control box, and the design variables will be the vertical displacements (z axis) of the 14 control points. Figure 1 depicts the selected parameterization.

To clarify, there are additional control points at the trailing and leading edge that are kept fixed, in order to maintain the angle of attack; so these control points are not considered as design variables.

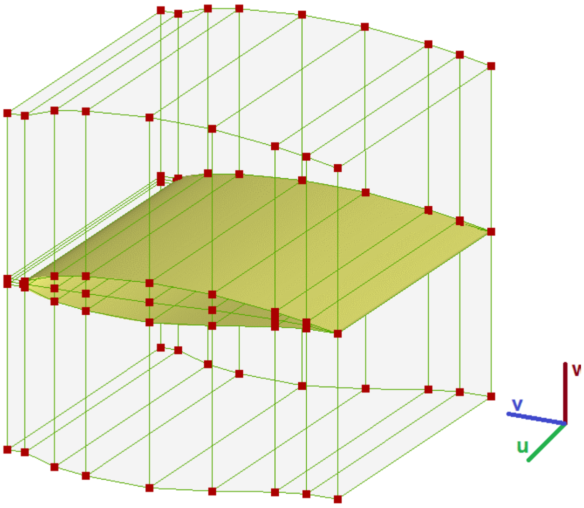

Figure 1. RAE 2822 control box parameterization

3.2. *Evolutionary Algorithm*

Evolutionary algorithms (EAs) are bio-inspired methods that clone the behaviour of natural evolution to solve complex optimization problems. The basic elements of an EA are the solution coding, the selection operator and the crossover and mutation operator.

In the design application to be considered in this work, each coding vector is composed by a given parameterization of a geometry, i.e.,$z = [cp1, cp2, cp3, \ldots, cpN]$, where cp is the vertical coordinates of each control point.

More details about the EA applied in this paper can be found in a previous work from the authors[17].

3.3. *Objective function approximation using Support Vector Machines (SVMs)*

Support vector machines acts as a meta-model to predict the objective function to be optimized, which in this case is given by the aerodynamic performance of de airfoil.

Support Vector machines for Regression are a powerful tool used on the machine learning field, and a modelling tool for a large amount of regression problems on engineering. The SVR can be solved as a convex optimization problem using kernel theory to face nonlinear problems. The SVR consider not only the prediction error but also the generalization of the model. To obtain the best performance, a search of the most suitable combination of the kernel parameters must be carried on, usually by using cross validation techniques over the training set. 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 by Ortiz-García et al.[16] which has proven to require pretty short search times.

More details about the SVR surrogate model applied in this paper can be consulted in a previous work from the authors[17].

3.4. *Flowchart of the proposed approach*

In this article, The Intelligent Estimation Search with Sequential Learning (IES-SL) method is applied. This method allows performing an efficient adaptive sampling guiding the optimization algorithm towards the most promising regions of the design space. The flowchart of the proposed approach is depicted in Figure 2. First, an initial set of randomly generated (including the baseline) geometries are selected and evaluated with CFD tool (DLR Tau code in this work). With this set, a first surrogate is built and linked within an evolutionary algorithm. The latter will search for the minimum of the surrogate in each of the optimization iterations, and the returned optima will be again evaluated using the high-fidelity CFD solver, and then incorporated to the surrogate model, which is rebuilt and more precise on each iteration. The process will end when a certain number of CFDs budget is reached.

The aim of this work is to study the influence of the initial training size in the precision of the surrogate and the convergence of the proposed approach.

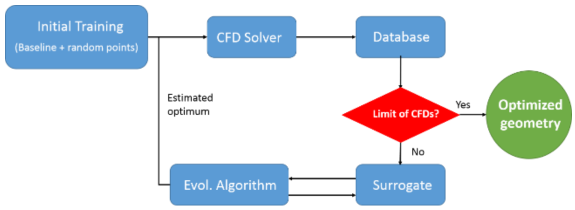

Figure 2. Flowchart of the proposed approach.

4 Numerical results

4.1. *Baseline geometry*

The selected geometry for this study was the well-known RAE2822 airfoil. The airfoil is a rear-loaded, sub-critical geometry, designed to exhibit a roof-top type pressure distribution at design conditions (Mach = 0.66, Cl = 0.56[18]). It has been tested in the RAE wind tunnel in 11 different flow conditions in the range of Mach numbers from 0.676 to 0.750 and at several Reynolds numbers[19].

Chord [m]	0.61
Maximum thickness-to-chord ratio	0.0121 at x/c=0.38
Maximum camber-to-chord ratio	0.0126 at x/c=0.76
Leading edge radius [m]	0.00827
Airfoil area [m^2]	0.0776
Trailing edge angle	9º

Table 1. Baseline airfoil features.

A 56k points unstructured grid was generated for this study.

4.2. *Test case definition*

The proposed approach is applied to 5 optimizations cases with 4, 8, 16, 32 and 64 initial random training points respectively. The multipoint optimization problem of the RAE 2822 is selected. The flow conditions for both design points 1 & 2 are:

	DP1	DP2
Mach	0.734	0.754
Re	6.5M	6.2M
Turb. Model	SA	κω TNT

The objective function selected was $Min\left(\frac{C_D}{C_L}\right)$ with some considerations. These are:

- Aerodynamics constraints and penalties:
  1. Prescribed minimum lift coefficient: $C_l^0|_k : C_l|_k \geq C_l^0|_k$
  2. Prescribed minimum pitching coefficient: $C_m^0|_k : C_m|_k \geq C_m^0|_k$
  3. Drag penalty: if constraint on minimum pitching moment is not satisfied, the penalty will be 1 drag count per 0.01 in ΔC_m
- Geometric constraints
  1. Limit: +/- 20% of the initial control points' values.
  2. Prescribed maximum thickness ratio $(t/c)_{max}$: $\max(t/c) = (t/c)_{max}$
  3. Prescribed minimum thickness ratio $(t/c)_{min}^{80}$ at $x = 0.8c$: $(t/c)^{80} \geq (t/c)_{min}^{80}$
  4. Prescribed minimum leading edge nose radius R_{min}^{le} : $R^{le} \geq R_{min}^{le}$

4.3. Sensitivity study results

In this section, the results of the present study are presented. Three issues are analysed. First, the influence of the initial training size in the convergence of the method. Next, the influence in the method precision of the initial data set. Finally, the value of the objective function reached in each case.

Regarding the first analysis, Figure 3 shows the convergence of the IES-SL for each test case. As can be seen, the five test cases have a huge oscillation during the "training period". This is the expected behaviour since the points in this data set are generated randomly. A lower size of initial training means the optimizer requires more iterations to reach the "optimum region". The reason is that the initial surrogate is more intelligent with a huge initial data set, but it requires more time to be built. At last, the five test cases reach the same optimum region (see Table 2).

Figure 4 illustrates the accuracy of the method with respect the initial training size. As expected, an initial surrogate with a vast number of points has an initial accuracy higher than one with a small set of points. This is in the same line that the convergence. Nevertheless, it requires more time to start the optimum seek.

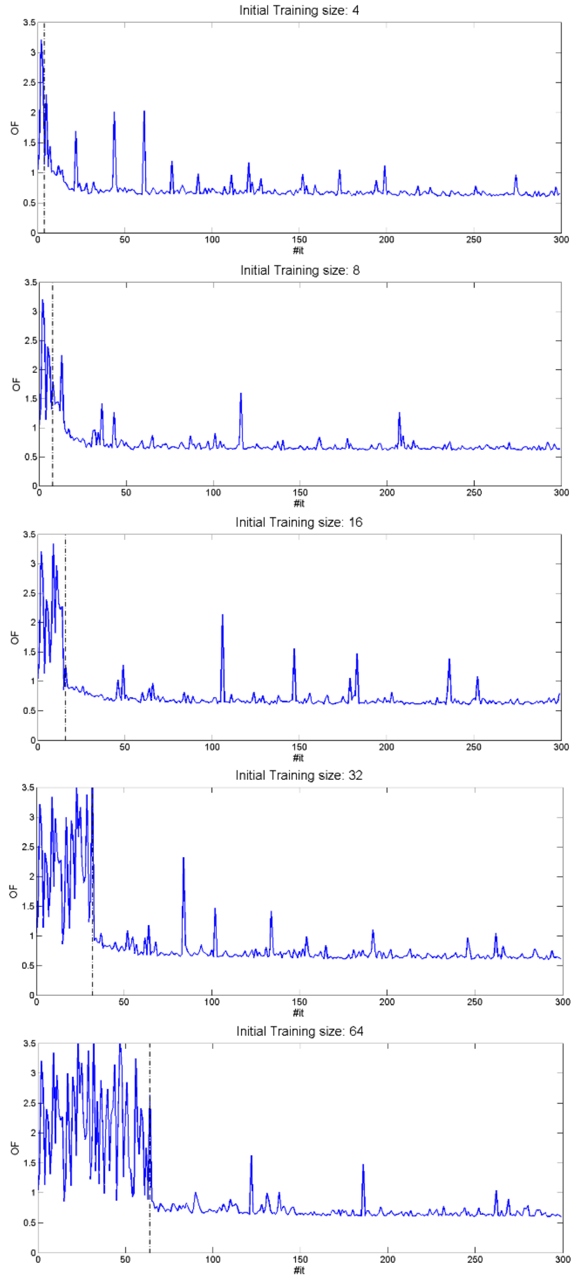

Figure 3. SBGO convergence vs. initial data set size

Last, but not least, Table 2 summarizes the value of the OF reached in each case. It can be seen that there is no influence of the initial training size in the final value of the OF (with a reasonable budget of iterations). This is the main advantage of the IES-SL proposed.

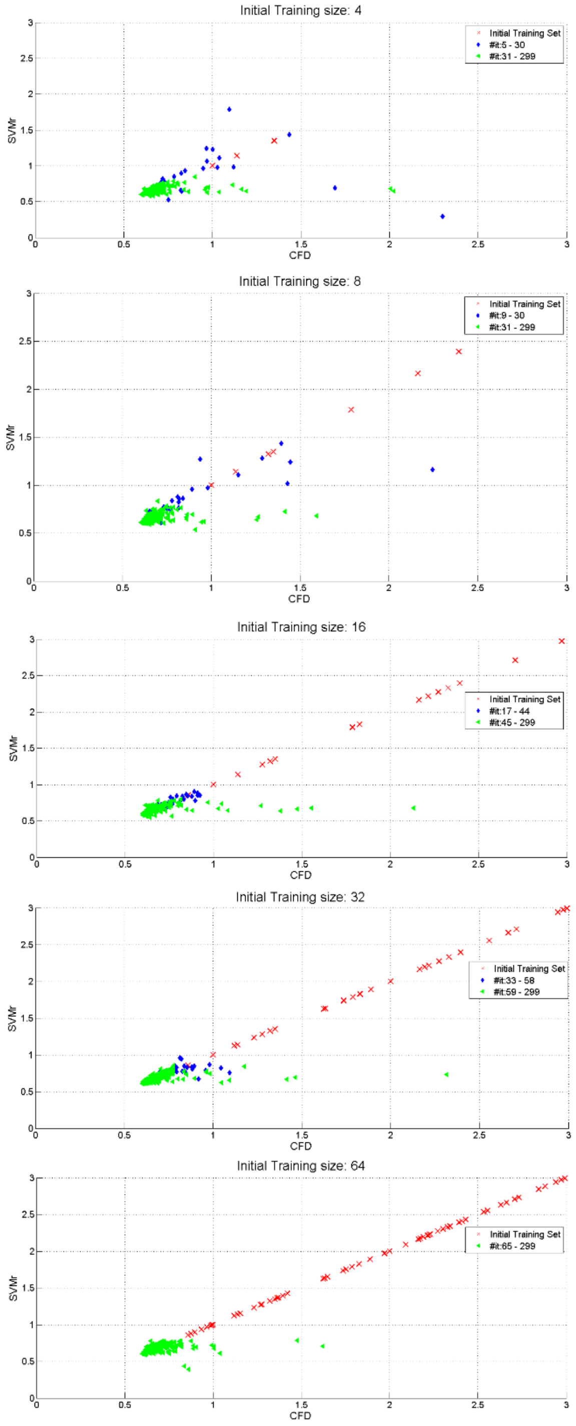

Figure 4. Approach accuracy for each initial training size

# Initial Training random points	Objective Function (OF)
4	0.6014
8	0.6059
16	0.6021
32	0.6016
64	0.5993

Table 2. OF evolution respect the initial training size

5 Conclusions

The aim of this work was to provide an analysis about how the initial training size of the surrogate affects the behaviour of the proposed IES-LS method. The following conclusions have been extracted from the solutions:

-- The optimum region reached is the same independently the training set size. A model with higher initial data set size requires less iterations to reach de optimum region, but it requires more computational time to be built, which is not feasible from the industry point of view.
-- In the same trend, the initial accuracy of the surrogate increases with the number of training samples, but the drawback is the same which is exposed in the previous point.
-- As summarized in Table 2, the training set size has no influence in the OF reached by the proposed IES-SL approach.

In summary, the main advantage of the proposed SBO method is that it can reach the global optimum with a small number of initial samples. This is feasible due to sequential learning allows the surrogate to become accurate each iteration. So, there's an important reduction of the initial computational cost that requires a standard offline SBO.

Acknowledgements

The research described in this work has been supported under INTA activity "Termofluidodinámica" (IGB99001) and the 'Rafael Calvo Rodés' scholarship.

6 References

[1] J. Parr, C. Holden, A. Forrester and A. Keane, "Review of Efficent Surrogate Infill Sampling Criteria with Constraint Handling," in 2nd International Conference on Engineering Optimization, Lisbon, Portugal, 2010.

[2] L. Leifsson, S. Koziel and Y. Tesfahungen, "Aerodynamic Design Optimization: Physics-based Surrogate Approaches for Airfoil and Wing Design," in AIAA SciTech, 2014.

[3] C. Li, J. Brezillon and G. S., "A Framework for Surrogate-Based Aerodynamic Optimization," in ONERA-DLR Aero-space Symposium, ODAS, 2001.

[4] S. Koziel and L. Leifsson, "Multi-Level Surrogate-Based Airfoil Shape Optimization," in 51st AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition, Grapevine (Dallas/Ft. Worth Region), Texas, 2013.

[5] E. Iuliano and D. Quagliarella, "Aerodynamic Shape Optimization via non-intrusive POD-based Surrogate Modeing," in IEEE Congress on Evolutionary Computation, Cancún, Mexico, 2013.

[6] A. Jahangirian and A. Shahrokhi, "Aerodynamic shape optimization using efficent evolutionary algorithms and

unstructured CFD solver," Computers & FLuids, vol. 46, pp. 270-276, 2011.

[7] A. Keane, "Wing Optimization Using Design of Experiment, Response Surface, and Data Fusion Methods," Journal of Aircraft, vol. 40, no. 4, pp. 741-750, 2003.

[8] H. Likeng and G. Zhenghong, "Wing-body Optimization Based on Multi-fidelity Surrogate Model," in International Congress of the Aeronautical Sciences ICAS, Brisbane, Australia, 2012.

[9] T. Lukaczyk, F. Palacios and A. JJ, "Active Subspaces for Shape Optimization," in AIAA SciTech, 2014.

[10] E. Andrés-Pérez and E. Iuliano, Application of Surrogate-Based Global Optimization to Aerodynamic Design, Springer Tracts in Mechanical Engineering, 2015.

[11] D. González-Juárez, E. Andrés-Pérez, M. Martin-Burgos, L. Carro-Calvo and S. Salcedo-Sanz, "Influence of geometry parameterization in aerodynamic shape design of aeronautical configurations by evolutionary algorithms," in 6th European Conference for Aeronautics and Space Sciences (EUCASS), Krakow, Poland, 2015.

[12] E. Andrés-Pérez, D. González-Juárez, M. Martin-Burgos, L. Carro-Calvo and S. Salcedo-Sanz, "Influence of the number and location of design parameters in the aerodynamic shape optimization of a transonic aerofoil and a wing through evolutionary algorithms and support vector machines," Engineering Optimization, vol. 48, 2016.

[13] A. Mousavi, P. Castonguay and S. Nadarajah, "Survey of shape parameterization techniques and its effect on three-dimensional aerodynamic shape optimization," in AIAA Computational Fluid Dynamics, Miami, 2007.

[14] M. Martin, E. Andrés, E. Valero and C. Lozano, "Gradients Calcultion for Arbitrary Parameteizations via Volumetric NURBS: The Control Box Approach," in EUCASS, 2013.

[15] L. Piegl and W. Tiller, The NURBS book, New York: Springer-Verlag Berlin Heidelberg, 1997.

[16] E. Ortiz-García, S. S. Sanz, Á. M. Pérez-Bellido and J. A. Portilla-Figueras, "Improving the training time of support vector regression algorithms through novel hyper-parameters search space reductions," Neurocomputing, 2009.

[17] E. Andrés, S. Salcedo-Sanz, F. Mongue and A. Pérez-Bellido, "Efficent aerdoynamic design through evolutionary programming and support vector regression algorithms," Expert Systems with Applications, vol. 39, pp. 10700-10708, 2012.

[18] ESDU, "Second-order method for estimating the subcritical preassure distribution on a two-dimensional aerofoil in compressible inviscid flow," 1973.

[19] P. Cook, M. Mcdoland and M. Firmin., "Aerofoil RAE 2822 - Pressure Distributions, and Boundary Layer and Wake Measurements," AGARD Report 138, 1979.