AIAA JOURNAL
Vol. 48, No. 7, July 2010

Nonparametric Fitting of Aerodynamic Data Using Smoothing Thin-Plate Splines

Ernesto Benini* and Rita Ponza
University of Padova, 35131 Padova, Italy

DOI: 10.2514/1.J050028

This paper introduces a nonparametric fitting method for the interpolation of aerodynamic observations over a large range of multiple angles of attack. The method is based on the employment of smoothing thin-plate spline class functions, a well-renewed mathematical tool for multivariate data mining based on the generalization of the univariate natural cubic splines, in which a roughness penalty criterion is used to produce very smooth predictive hypersurfaces. Compared with other methods, such as parametric or even conventional nonparametric methods, the use of a smoothing thin-plate spline is more effective, in that the predictive surface comes directly from the observed points, thus minimizing any intervention of the analyst aimed at introducing model parameters. This forms the basis for a very reliable fitting technique, in which model construction can be relatively easy to implement. An application of the method is carried out on a case study representative of some experimental data coming from a wind-tunnel campaign on a typical three-dimensional fuselage-shaped body, aimed at the acquisition of its aerodynamic coefficients over a rather extensive attitude range. Specifically, the application is focused on the body lift coefficient as a function of both angle of attack and sideslip angle. The data set is also interpolated using concurrent response-surface methods: namely, a linear model, a bivariate spline, a radial basis function network, a support vector regression technique, a regression kriging, and a moving-least-squares approach, alternatively known as local polynomial regression. Results of data fitting are assessed using a cross-validation approach and reveal a clear superiority of smoothing thin-plate spline over the other methods, leading to a more regular fitted surface and a more reliable prediction tool, even when some observations are omitted. This is important per se, but acquires even more significance when an aerodynamic test campaign is to be planned with the minimum number of experimental observations.

Nomenclature

$\mathbf{a}$ = vector of unknown B spline or thin-plate spline coefficients
$B_{i,h}$ = ith B spline of order h for a given knot sequence s
CL = lift coefficient
CV = cross-validation score function
D^2 = second derivative operator
$E(t)$ = functional, parametric in t
$f(\mathbf{x})$ = systematic component of the functional relationship between $\mathbf{y}$ and $\mathbf{x}$
h = bandwidth in moving least squares
$J(f)$, $J_m(f)$ = functional in function f
$K(\mathbf{x})$ = Epanechnikov kernel in moving least squares
$K_h(.)$ = d-variate kernel function in moving least squares
$m(\cdot)$, $m^{(k)}(\cdot)$ = model and its kth-order partial derivative in moving least squares
n, m = number of observations
p = smoothing parameter in B splines, degree of the local polynomials in moving least squares
R = determination coefficient
$S(f)$, $S_{md}(f)$ = functional in function f
$\mathbf{s}$, $\mathbf{t}$ = vectors of knots for B spline construction
$\mathbf{w}$ = vector of weights
$\mathbf{X}$ = regression matrix
$\mathbf{x}$ = vector of independent variables, independent parameters, and predictors
$\mathbf{y}$ = vector of observed values or output variables
$\hat{\mathbf{y}}$ = estimated response vector
$\hat{y}^{(-i)}$, $\hat{f}^{(-i)}$ = estimated response vector after cross-validation, obtained omitting ith observation
z = $\phi(\|\mathbf{x}-\mu\|)$, radial basis function
α = smoothing parameter in thin-plate spline, angle of attack
$\boldsymbol{\beta}$ = vector of parameters to be estimated, vector of weights in radial basis functions
β = sideslip angle
$\boldsymbol{\delta}$ = vector of weights in thin-plate spline
ε = threshold in thin-plate spline
$\boldsymbol{\varepsilon}$ = vector of random errors between the estimated and the observed $\mathbf{y}$
η = function of the Euclidean distance in the independent variable space
μ = radial basis function center
$\boldsymbol{\sigma}$ = vector of widths in radial basis functions
τ = generic knot or break for B spline construction

I. Introduction

ONE of the major concerns with experimental aerodynamics when handling data coming from both wind-tunnel campaigns and in-flight tests is the identification of a useful and sound fitting technique for model building, especially when the aerodynamic coefficients over bodies at high incidence angles are investigated. Actually, a robust and accurate modeling approach may help gaining a deeper insight into the physical interpretation of the phenomena under analysis; moreover, the need to feed the flight mechanics tools with reliable data can not be addressed unless experimental acquisitions are properly postprocessed.

Fitting of aerodynamic data is currently carried out using both parametric and nonparametric techniques.

Received 26 June 2009; revision received 30 March 2010; accepted for publication 30 March 2010.

*Research Professor, Department of Mechanical Engineering; ernesto.benini@unipd.it. Member AIAA (Corresponding Author).

A parametric technique is the one that obeys the rules of the conventional statistical regression analysis [1], and is often used within the realm of physical experiments as it typically smoothes out the random errors that inevitably affect the tests. It basically consists in predefining a form of a response surface, usually of the low-order polynomial type because of their intrinsic physical meaning, the unknown coefficients of which are determined using a generalized least-squares regression criterion to fit the response-surface-predicted values to the observed data. The result is an approximating function or hypersurface that mimics the functional relationship between the response y and the input variables xs.

This technique is often referred to as statistical response-surface methodology (RSM) [2–4]. Examples of application of RSM techniques are uncountable in aeronautics and usually refer to fitting wind-tunnel or numerical data samples in a rather restricted range of incidence angles. Among others, in [5] the aerodynamic characterization of a complex aircraft configuration is carried out using a RSM approach in which the selected input variables are the vehicle's attitude (defined by pitch and yaw angles, which vary in a limited range) and the control surfaces' settings; the reliability of the parametric model is evaluated using the statistical analysis of variance (ANOVA). In [6], an approach to preliminary design of aircraft is presented in which RSM is used to accelerate the search for optimal configurations of vehicles. RSM in combination with modern design of experiments (MDOE) is used in [7,8] for the numerical estimation of the drag polars at different attitudes (although still within a restricted range) of a joined wing concept aircraft, in which the parametric model is eventually used to provide reliable data to the flight mechanics tools. A response surface for pressure distribution prediction built on some experimental data over a model aircraft is instead created in [9], in which pitch and yaw angles are varied within a relatively small range (up to 30 deg) to cover the flight envelope under investigation.

However, in usual wind-tunnel tests, it is rather unlikely for a unique low-order polynomial-type response surface to adequately model experimental data over the whole attitudes' domain of interest. Hence, it becomes necessary to split the parameters' exploratory space into a number of subdomains over which polynomial-type modeling functions can correlate the data in a satisfactory way: the global response results then from the combination of different submodels and looks like a piecewise continuous function, consisting of a series of adjacent multiple response surfaces [9]. Nevertheless, it clearly appears that the identification of proper subdomains implies some knowledge of the response trend and hence needs to rely on empirical criteria coming from previous experiments on similar configurations in order to be effective. To this purpose, in [10] the application of an RSM approach is illustrated, aimed at an exploratory analysis of a model-scaled aircraft with highly nonlinear aerodynamic behavior at high angles of attack. Different parameters' subdomains are selected, representative of two spaces (namely, high- and low-incidence angles), and a response model is developed over adjacent subspaces on the global design space, defined by proper limits on model attitudes and deflections of the aircraft control surfaces. Also, in [5] the need for identifying a priori the limits of the parameters' domain over which a RSM approach can be successfully adopted is emphasized as being one of the most critical features of such methods when applied to wind-tunnel data modeling. However, the identification of meaningful application limits is always left to the analyst's judgment: while this could be acceptable for limited attitudes' ranges or in regions of known behavior, it could lead to inadequate representation of nonlinear phenomena, like stall and poststall of complex configurations.

As a matter of fact, RSM might not be satisfactory for handling the level of complexity typically encountered in fitting aerodynamic data, especially when robust inter- or extrapolation is needed over a wide range of angles of attack. In fact, in practical applications in which complex flight envelopes require to account for large variations in the aerodynamic coefficients due to the extensive range of attitudes under consideration (e.g., helicopter fuselages, in which the vehicle is intrinsically subject to extremely variable flight conditions resulting in a wide range of angles of attack), a polynomial regression technique may result in misleading responses [5]. Actually, when low-order polynomials are used to model a limited number of sampled points, an apparent lack of fit is likely to occur as a result of the model structure deficiency: this typically leads to a response that does not capture the phenomenon under consideration, being far away from the observed points. On the other hand, a more complex high-order polynomial model could be desirable as it better fits the available experimental data; however this often results in very complicated response surfaces with poor or undecipherable engineering relevance. Moreover, the criterion upon which to base the choice of the polynomial model and its order may not to be generalized and is almost invariably left to the analyst's judgment.

A nonparametric approach, in which no hypothesis is made a priori about the form of the function relating the response variable to the independent variables, is an effective and useful alternative to RSM [11–13]. Unlike parametric statistical inference techniques, in a nonparametric approach the response function is not assumed to belong to a specified parametric class of functions: on the contrary, it is only supposed to obey to a few and rather general smoothness conditions. The very attractive feature of this approach is that data to be fitted is not forced into a prescribed mathematical structure in order for the unknown model parameters to be determined, but it is left free to build the statistical model on its own without being trapped into a predefined, constrained formulation. In other words, the response function is identified only on the basis of the assigned data, and its determination becomes actually the final goal of the model identification. In this sense, it is called nonparametric, i.e., not because it is parameterless, but because the goal of the regression is now to estimate the regression function f directly rather than the parameters.

Some misunderstanding exists among engineers on what is to be regarded as a nonparametric model, in that it is erroneously believed that this property should characterize only models that produce an exact, or interpolative, fit through all of the observed data. Instead, a nonparametric approach is focused on determining the correct function according to some optimization criteria regardless the local property of passing *through* or *near* the observations. In fact, similar to parametric regression, a weighted sum of the y observations is used to obtain the fitted values. However, instead of using equal weights as in ordinary least-squares or weights proportional to the inverse of variance as is often the case in weighted least-squares, a different rationale determines the choice of weights in nonparametric regression, as will be specified later.

Some relevant examples of nonparametric methods for data fitting or metamodel construction and optimization in aeronautics have already been documented that include mainly multivariate adaptive regression splines [14,15], neural networks [16–18], and radial basis functions (RBF) [which are, de facto, a particular type of neural network, thus leading to the popular expression radial basis functions networks (RBN)] [19–21], each of which has some advantages and drawbacks [22,23]. Other more recently developed nonparametric techniques, whose application to aeronautics is far less diffused, encompass support vector regression [24–26], regression kriging [27–30], and moving least-squares or local polynomial regression [31–35].

In this paper, a technique for robust fitting of aerodynamic data is proposed that takes advantage of all the strengths of a nonparametric approach while minimizing its weaknesses. In particular, an innovative application to aerodynamic data fitting is introduced that is based on a very robust, reliable, and physically sound nonparametric methodology. Such methodology stems from the deterministic formulation referred to as smoothing thin-plate spline technique that, as will be evident in the paper, perfectly suits for the purposes of fitting data over a wide range of angles of attack without incurring in over- or underfitting-related problems.

The paper is organized in the following way: after a brief discussion of parametric models, both the main conventional and innovative nonparametric techniques for model building are described, including multivariate adaptive regression splines, neural networks, radial basis functions, support vector regression, regression kriging, and moving-least-squares technique. Then the methodology based on the use of smoothing thin-plate splines is presented and an application to some sample aerodynamic data is illustrated. Results are discussed and compared with those coming from both parametric and traditional techniques, as well as more recent nonparametric techniques.

II. Parametric Models

RSM consists in a collection of statistical and mathematical techniques for parametric model building, aimed at developing a reliable model that exhibits the highest correlation with observations, while keeping the number of explanatory variables to a minimum.

In the experimental aerodynamic field, the response surface usually turns out to be expressed in the form of a statistical linear model in which the estimation function of the output variable **y** is linear in the parameters:

$$E(y_i) = \alpha + \beta_1 x_{1i} + \ldots + \beta_p x_{pi} \tag{1}$$

where $(\alpha, \beta_1, \ldots, \beta_p)$ is the vector of parameters to be estimated; $(x_{1i}, \ldots, x_{pi})$ is a vector of predictors for the ith of n observations; $\mathbf{y} = f(\mathbf{x}) + \varepsilon$, where $f(\mathbf{x})$ the systematic component of the functional relationship between the response and the independent variables; ε_i is the sum of measurement (random) errors and lack of fit between the estimated and the observed ith values; and ε is assumed to be normally, identically, and independently distributed, with zero mean and constant variance.

The general approach of RSM includes first some screening trials focused at selecting those input factors that are most influential to the response variable being investigated: this is essential in order to reduce the complexity of the model while enhancing its prediction accuracy and is usually addressed using MDOE [36–38]. The statistical significance of each term is usually assessed through ANOVA: terms in the polynomial function having large variance may be dropped from the model with negligible effect on the fidelity of the response-surface fit.

Once the most significant terms are identified, the estimation problem can be formulated as $\hat{\mathbf{y}} = \mathbf{X}\beta$, where $\hat{\mathbf{y}}$ is the estimated response vector and $\mathbf{X}$ is the regression matrix. Based on the least-squares regression approach, the solution to the above formulated problem is $\beta = (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{y}$.

Finally, some well-known statistical techniques exist through which both the approximation model accuracy and uncertainty levels may be evaluated through the use of several metrics that quantify the modeling error: i.e., the discrepancy between the response values given by the model and the actual observations. A well-established practice consists, for instance, in evaluating the overall correlation coefficient, examining the residual distribution and quantifying the global variance between observed and predicted values.

III. Concurrent Nonparametric Methods for Data Fitting

In the following, a brief summary of the most popular and effective methods of nonparametric modeling is outlined in order to help understanding both their benefits and weaknesses.

While polynomial models can be regarded for as global models, in which both the observations near to (in the Euclidean distance sense) and far from a location **x** in the input parameters' domain equally influence the predicted response over **x**, nonparametric approaches have a somewhat local character [39]. Specifically, the closer the available observations to **x**, the higher their weight in the determination of the predicted response $\hat{f}(\mathbf{x})$. This seems particularly attractive when the unknown response function is highly multimodal, as is the case for aerodynamic coefficients of bodies at high incidence angles.

The general nonparametric regression model is written in a similar manner to the parametric one, but the function f is left unspecified:

$$E(y_i) = f(x_{1i}, \ldots, x_{pi}) + \varepsilon_i \tag{2}$$

Most methods of nonparametric regression implicitly assume that f is a smooth, continuous function. As in parametric regression, it is standard practice to assume that the errors ε_i are normally distributed with zero mean and constant variance.

A. Multivariate Adaptive Regression Splines

Multivariate splines are piecewise-polynomial functions of given smoothness [40,41]. They constitute the generalization of the well-known 1-D (or univariate) cubic spline curve in an n-dimensional space. When a two-dimensional space of independent variables is considered, as the case in this work, they are often referred to as bivariate splines.

A bivariate spline in B form is obtained from a univariate spline by using the tensor product construction:

$$\begin{aligned} f(x_1, x_2) &= \sum_{i=1}^{m}\sum_{j=1}^{n} B_{i,h}^{x_1} B_{j,k}^{x_2} a_{ij} \\ &= \sum_{i=1}^{m}\sum_{j=1}^{n} B(x_1|s_i, \ldots, s_{i+h}) B(x_2|t_j, \ldots, t_{j+k}) a_{ij} \end{aligned} \tag{3}$$

where $B_{i,h}^{x_1} = B(x_1|s_i, \ldots, s_{i+h})$ is the ith B spline of order h for the given knot sequence s in the x_1 direction: i.e., the B spline with knots $s_i, \ldots, s_{i+h}$ in the basic interval $(s_1, \ldots, s_{n+h})$. Similarly, $B_{j,k}^{x_2} = B(x_2|t_j, \ldots, t_{j+k})$ is the jth B spline of order k for the given knot sequence $\mathbf{t}$ in the x_2 direction: i.e., the B spline with knots $t_j, \ldots, t_{j+k}$ in the basic interval $[t_1, \ldots, t_{j+k}]$. It is worth recalling that, for example, $B_{j,k}^{x_2}$ is a piecewise polynomial of degree $<k$, with breaks $t_j, \ldots, t_{j+k}$, is zero outside its basic interval, and is normalized such that $\sum_{j=1}^{n} B_{j,k}^{x_2} = 0$ for $[t_k, \ldots, t_{n+1}]$.

Spline smoothness is governed by the nodal surface derivatives, in particular the second partial derivative, therefore is linked to multiplicity of knots. If the number τ occurs exactly r times in the sequence $t_j, \ldots, t_{j+k}$, then $B_{j,k}$ and its first k-r-1 derivatives are continuous across the break τ, while the (k-r)th derivative has a jump at τ.

Unknown coefficients a_{ij} are determined according to the type of approach used for data approximation. In the framework of a nonparametric approach, the spline does not fit observed data $\mathbf{y} = (y_1, \ldots, y_{m+n})$ exactly but is rather derived as a best interpolant, e.g., as the function having the smallest Euclidean distance from all observed data (least-squares criterion). This approach is often referred to as variational approach. The surface is then determined by solving the following variational problem:

$$\begin{aligned} &\text{minimize } pE(t) + (1-p)F(t) = p\sum_i w_i |y_i - f_i(x_1, x_2)|^2 \\ &\quad + (1-p)\int_{a_1}^{b_1}\int_{a_2}^{b_2} |D^2 f(t)|^2 \, \mathrm{d}t \end{aligned} \tag{4}$$

where p is a smoothing parameter, which is determined to make $F(t)$ as small as possible subject to the condition that $E(t)$ is no greater than a prescribed, small, tolerance.

Multivariate smoothing spline techniques have several advantages, such as simplicity and robustness, when applied to fitting noisy data, since they are able to produce physically realistic interpolation with reasonable accuracy and smoothness. They are very flexible for approximating known or unknown functions or any given data sets. Moreover, from the numerical point of view, routines for spline construction are very consolidated and reliable, since the linear systems to be solved for this are fairly simple.

However, the main drawback with this approach relies on the fact that there is no universal criterion for determining the number of knots upon which to construct the fitting model. In particular, if the number of spline sites and provided knots obey to the so-called Schoenberg–Whitney conditions, i.e.,

$$\begin{aligned} &\text{knots}(j) < x_{1,2}(j) < \text{knots}(j+k) \\ &\quad \text{for} \quad j = 1, \ldots, \text{length}(x) = \text{length}(\text{knots}) - k \end{aligned} \tag{5}$$

then there is a unique spline passing through the observed points: i.e., $y_i = f_i(x_1, x_2)$ for all j. This results in a purely nonparametric approach (exact fitting) that may lead to erroneous predictions (e.g., overfitting) within the observed data, especially for nonnumerous sets of data.

B. Radial Basis Functions and Radial Basis Networks

As it is known, a generic radial basis function can be expressed in the form [42]:

$$z(\mathbf{x}) = \phi(\|\mathbf{x} - \mu\|) \tag{6}$$

where $\mathbf{x}$ is an n-dimensional input vector, μ is called center, $\| \cdot \|$ denotes the Euclidean distance, and ϕ is a univariate function, which is often referred to as the profile function. Typically, a fitting model is set up as a linear combination of N radial basis functions having N distinct centers. This is equivalent to build a linear neural network having a number of inputs corresponding to the number of input vectors, primitive nodes whose transfer function is given by Eq. (6), and a single output (which corresponds to the values to be fitted) [43,44]. For instance, when an input vector $\mathbf{x}$ is given to the network, the output of the RBF network (or RBN) is the activity (or neural) vector:

$$\hat{y}(\mathbf{x}) = \sum_{j=1}^{N} \beta_j z_j(\mathbf{x}) \tag{7}$$

where β_j is the weight associated with the jth radial basis function centered at μ_j.

A variety of radial basis functions are used in practice [45], such as Gaussian, Cauchy, multiquadrics, and others. A particular type of RBF is often and generically referred to as a thin-plate spline (TPS), which is the main subject of this paper. The general form of a TPS is the following:

$$\phi(\mathbf{x}) = \left(\frac{\mathbf{x}}{\boldsymbol{\sigma}}\right)^2 \log\left(\frac{\mathbf{x}}{\boldsymbol{\sigma}}\right) \tag{8}$$

However, the meaning of this designation in this paper is intended to be significantly different from the one described in the framework of the radial basis function community, to which the expression above refers. The reasons for this are explained below.

First, RBFs (and TPS in the form given above) need the centers to be specified. At present, no general rule exists for selecting them [46], even though some criteria have been developed [47], such as the regularized orthogonal least-squares (ROLS) procedure, in which the centers are chosen one at a time using a forward-selection procedure from a candidate set consisting of all the data points or a subset thereof. However, it is almost impossible to build automatic selection schemes, so the aerodynamicist is often obliged to select them by trial and error until a satisfactory fitting is obtained.

Moreover, all of the radial basis functions have an associated width parameter $\boldsymbol{\sigma}$, which is related to the spread of the function around its center. Again, such parameters must be decided by the user, since there is no rigorous criterion for its choice. A heuristic option is given in [46], in which the width is the average over the centers of the distance of each center to its nearest neighbor. However, this holds true for Gaussians RBFs, and it is only a rough guide that provides a starting value. Some algorithms exist for the width selection [48], including generalized cross-validation (GCV), but basically all of them proceed from a tentative value and test several widths values equally spaced between specified initial upper and lower bounds, then the width value minimizing log10 (GCV) is selected.

C. Support Vector Regression

Proper versions of support vector machines (SVM) [24] suitable for interpolation and regression have been developed recently that are called support vector regression (SVR) [25]. In SVR, a nonlinear function is learned by a linear learning machine in a kernel-induced feature space, while the capacity of the system is controlled by a parameter that does not depend on the space dimensionality. As in the classification case, the learning algorithm minimizes a convex functional and its solution is sparse. The model produced by SVR only depends on a subset of the training data, because the cost function for building the model does not care about training points that lie beyond the margin and moreover because the cost function for building the model ignores any training data close to the model prediction (within a threshold ε). SVM springs to mind as the most prominent method that balances data fitting and model simplicity. It is particularly suited to data from physical experiments as the level of experimental error can be used as the width ε.

Despite the numerous advantages related to the kernel-induced mapping, the use of SVM in fitting aerodynamic data seems to be constrained by the cardinality of the observed data: in fact, as in neural networks, the predictive capability of such technique is largely dependent on how numerous and dense are observations used to learn the machine. Moreover, the biggest limitation when using a support vector approach probably lies in the choice of both the kernel function parameters (e.g., for Gaussian kernels the width parameter) and the value of ε in the ε-insensitive loss function, which are left to the analyst's judgment [26].

D. Regression Kriging

Regression kriging (RK) is a spatial interpolation technique that combines a regression of the dependent variables on auxiliary variables with simple kriging of the regression residuals [27]. It is mathematically equivalent to the interpolation method variously called *universal kriging* and *kriging with external drift*, in which auxiliary predictors are used directly to solve the kriging weights. From a different point of view, RK belongs to the family of linear least-squares estimation algorithms and, as such, is basically a regressing Gaussian RBF.

Advantages of kriging as an interpolation technique have been well understood and relate mainly to every estimate being accompanied by a corresponding kriging standard deviation. Thus, for any set of predicted values, a quantitative measure of confidence can be produced [28]. This makes kriging uniquely different from other interpolation methods. The estimation variance can also be used to determine when more information is needed if future sampling is planned. Other advantages in using RK have been underlined in [29] and deals with the fact that kriging weights depend not only on the distances between observational points and estimation locations but on the mutual distances among observational points as well. As a result, kriging has two interesting and unique properties: declustering and screen effect. With the declustering property, several observational points close to each other will have collectively the weight of a single observational point located near the centroid of the cluster. With screen effect, the influence of an observational point will be reduced by addition of one or more observational points at the intermediate locations between the original observational point and the estimation location. As the screen effect makes the influence of distant observational points negligible, the use of sampling subset in kriging is a safe practice compared with other weighting methods.

A remarkable drawback associated with RK is the estimation of semivariogram [30], which measures the degree of spatial correlation among observational data points in a study area as a function of the distance and direction vector between observational data points. It controls the way that kriging weights are assigned to data points during interpolation, and consequently affects the quality of the results. It is not always easy to ascertain whether a particular estimate of the semivariogram is, in fact, a true estimator of the spatial correlation in an area. The reasons for choosing a particular semivariogram to fit the given data set are often difficult to explain in terms of physical processes. They can only be rationalized in terms of a least-squares on maximum likelihood fit to the data set.

E. Moving Least Squares or Local Polynomial Regression

Moving-least-squares technique (MLS), often referred to as LPR, [31,32], was developed to overcome well-known drawbacks of traditional multivariate polynomials, such as excessive smoothing that makes them not flexible enough to achieve an adequate fit, as well as their attitude to exalt individual observations influence on remote parts of the fitted hypersurface. In LPR, instead of using unique polynomials to estimate the model that gives the best prediction, a multivariate polynomial is fitted locally for each independent variable value.

The technique assumes that the mean function is a smooth function and the amount of smoothness required is inherently decided by the degree of the local polynomial used. In a multivariate formulation, LPR is constructed in the following way. Polynomial estimation of model $m^{(\cdot)}$ or one of its partial derivatives is denoted by

$$m^{(k)}(\cdot) = \frac{\partial^K}{\partial x_1^{k_1}, \dots, \partial x_d^{k_d}} m(\cdot) \tag{9}$$

where $\mathbf{k} = (k_1, \dots, k_d)$, $k_j \geq 0$, for $j = 1, \dots, d$, and $K = k_1 + \cdots + k_d$ involves minimization of

$$\sum_{i=1}^{n} \left\{ Y_i - \sum_{j=0}^{p} \beta_{j,x}(X_i - x)^j \right\}^2 K_h(X_i - x) \tag{10}$$

with respect to $\boldsymbol{\beta}_{\cdot,x} = (\beta_{0,x}, \dots, \beta_{p,x})^\top$, where $K_h(\cdot)$ is a d-variate kernel function. The local polynomial estimation of $m^{(k)}(\mathbf{x})$ is given by

$$\begin{aligned} \hat{m}_k(\mathbf{x}; \mathbf{H}, p) &= \prod_{j=1}^{d} (k_j!) \hat{\beta}_{k,x} = \prod_{j=1}^{d} (k_j!) \hat{\boldsymbol{\beta}}_{\cdot,x} \\ &= (\mathbf{X}_{p,x}^\top \mathbf{W}_{H,x} \mathbf{X}_{p,x})^{-1} \mathbf{X}_{p,x}^\top \mathbf{W}_{H,x} \mathbf{Y} \end{aligned} \tag{11}$$

where

$$\mathbf{W}_{H,x} = \text{diag}\{K_H(\mathbf{X}_1 - \mathbf{x}), \dots, K_H(\mathbf{X}_n - \mathbf{x})\}$$

In local polynomial regression there are at least three parameters that are important for the estimation: namely, the degree p of the local polynomials, the kernel function $K(\cdot)$, and the bandwidth h such that $\mathbf{H} = h\mathbf{I}_d$, where $\mathbf{I}_d$ is the d-dimensional identity matrix. To achieve good estimation results it is important to choose some of these parameters carefully. The kernel function is actually the least important of them. The most common choice is a kernel supported on $[-1, 1]$ (though also the Gaussian kernel is not unusual), since a compactly supported kernel conveys nice asymptotic properties to the resulting estimate. A natural choice is the Epanechnikov d-dimensional kernel:

$$K(\mathbf{x}) = \frac{d(d+2)}{2S_d} \max\{(1 - \|\mathbf{x}\|^2), 0\} \tag{12}$$

where $S_d = 2\pi^{d/2}/\Gamma(d/2)$, since it is optimal in the sense that it minimizes the asymptotic mean squared error (MSE) and mean integrated squared error at interior points and is nearly optimal at most boundary points for all choices of p and k [33].

More important for the performance of the estimator is the choice of p. Here the bias-variance tradeoff is more tangible. To decrease the bias one can increase the value of p, though this might in turn increase the variability, since more local parameters are used and vice versa. As noted in [34], there is no loss in terms of asymptotic variance when moving from an even value of p-k to an odd value. Therefore, it is recommended to use odd values of p-k: generally, polynomials of order $p = k + 1$. If $k = 2$ [i.e., estimation is required for a C-2 class (existence and continuity of first-order derivative) of polynomials], a cubic polynomial ($p = 3$) is to be used.

The most critical parameter for the estimation result is the bandwidth, h, since it controls the size of the local neighborhood of the response function. The choice of h is a tradeoff between variance and bias. By choosing a large bandwidth the local estimate is influenced by many observations and thus the variance is small. On the other hand, the influence of remote observations might increase the squared bias. Many different techniques have been proposed in the literature for bandwidth selection. For instance, the conditional MSE minimization with respect to $h(\mathbf{x})$ gives a criterion for bandwidths estimation (and the result is often referred to as asymptotically optimal local bandwidths), as discussed in [34]. However, the minimization procedure requires some constants to be introduced depending in turn on p and k; therefore, some intervention from the analyst is still required. Instead, a very common criterion is to use cross-validation to obtain reasonable values for h or an iterative procedure, called the plug-in method [35].

IV. Motivation of the Work

In this paper, an effective formulation of TPS, alternative and distinct from the one presented above, is described and applied that drastically improves the predictive capability of both parametric and conventional nonparametric models. This holds true especially when a sparse or small database of aerodynamic data coming from either wind-tunnel or in-flight tests is available that cover wide ranges of angles of attack.

The advantages of the method presented hereafter over concurring approaches rely mainly on the deterministic result of its application, which does not imply any choice of model parameters.

In particular, the following aspects are worth mentioning:

1) Opposite of linear models, the method presented in this paper overtakes the choice of base functions and therefore keeps the analyst close to the physical meaning of the problem being considered.

2) Knot number and type, as required in multivariate splines, do not need to be defined.

3) Centers and width parameters in radial basis function networks are no longer needed, as well as kernel function parameters in support vector regression techniques.

4) Contrary to kriging regression, no criteria for weight assignment to data points is needed; therefore, selection of the semivariogram is not necessary, thus keeping the analyst close to the physical problem under consideration.

5) Choice of degree of local polynomials, type of kernel function $K(.)$ and bandwidth, as done in local polynomial regression, is definitely overcome.

Moreover, the proposed approach will be shown to significantly improve the fitting quality with respect to other techniques, in terms of both reduced data overfitting and augmented robustness.

Finally, a significant improvement of the method described below is that the obtained fitting is better than the one produced using other techniques, in that it leads to more robust and reliable predictions at high angles of attack.

In the following, the theoretical basis of the smoothing splines will be outlined, starting from the idea of a roughness penalty approach for determination of the response function in the one-dimensional case. The natural extension to multivariate regression will eventually lead to the introduction of the TPS that, to the authors' best knowledge, have never been explored for interpolating multivariate aerodynamic data over a broad range of angles of attack.

V. Thin-Plate Splines

TPS are known since quite a long time in the field of applied mathematics, in which they were originally introduced for geometric design: specifically, the name *thin-plate spline* refers to a physical analogy involving the bending of a thin sheet of metal. What is still unexplored is their application to the fitting of experimental aerodynamic data, which is the subject of the present work. However, the thin-plate splines that will be used to this scope must not be confused with those already introduced within RBFs: to this purpose, the latter will be referred to as radial basis thin-plate splines (RB-TPS), while the term TPS will indicate only those functions with the properties defined below. In particular, we will refer to what is often called a smoothed thin-plate spline (STPS) in the works by Wahba [49] and Green and Silverman [13].

In the following, a two-dimensional space of input variables will be treated: nevertheless, TPS and STPS are inherently multivariate in nature and hence their generalization to higher-order domains is straightforward [13], as will be described later.

Since it is assumed that not all of the aerodynamicists may be familiar with these topics, STPS and their properties will be widely discussed in this paper.

First, TPS are an important class of functions defined as follows:

$$f(\mathbf{x}) = \sum_{i=1}^{n} \delta_i \eta(\|\mathbf{x} - \mathbf{x}_i\|) + \sum_{j=1}^{3} a_j \phi_j(\mathbf{x}) \tag{13}$$

where $\mathbf{x}_1, \ldots, \mathbf{x}_n$ is an assigned set of input vectors in the two-dimensional space, η is a function of the Euclidean distance of the generic vector point $\mathbf{x}$ to $\mathbf{x}_i$,

$$\eta(\|\mathbf{x} - \mathbf{x}_i\|) = \frac{1}{16\pi} \|\mathbf{x} - \mathbf{x}_i\|^2 \log(\|\mathbf{x} - \mathbf{x}_i\|^2) \quad \text{if } \|\mathbf{x} - \mathbf{x}_i\| > 0,$$
$$\eta(0) = 0 \tag{14}$$

and

$$\phi_1(x_1, x_2) = 1, \qquad \phi_2(x_1, x_2) = x_1, \qquad \phi_3(x_1, x_2) = x_2 \tag{15}$$

where (x_1, x_2) are the coordinates of $\mathbf{x}$; finally, δ_i and a_j are constants with appropriate values. Specifically, under the condition $\mathbf{X}\boldsymbol{\delta} = 0$, where $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_n)$,

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \end{bmatrix}$$

and f is called natural thin-plate spline.

A. Smoothing Thin-Plate Splines

The smoothing TPS approach for multivariate regression aimed at model building stems from the idea of penalizing a response surface based on its roughness: basically, this method originates from the relaxation of some of the assumptions needed when building a model with the classical polynomial regression techniques.

When searching for an approximation function $\hat{\mathbf{y}} = f(\mathbf{x})$ using a least-squares approach, only the surface fitting to the observed data is to be maximized, and usually no additional constraint is imposed to f. Obviously, the least-squares sum goes down to zero if f exactly interpolates data; however, many different surfaces exist passing through the observed values and, even if some generic smoothness conditions are imposed to f, in order to make it continuous up to a specified derivative order, the resulting surface may not be satisfactory, being potentially subject to excessive fluctuations (as illustrated in Fig. 1a for a one-dimensional curve). Actually, a good data fitting should not be the unique goal in the identification of a response model: a further and often conflicting objective should be to obtain a surface free from undesired fluctuating behavior.

The roughness penalty approach tries to quantify this rather qualitative notion: hence, the problem of identifying an adequate model of the observed data may be stated so as to clearly highlight the need for a tradeoff between the aforementioned objectives.

A function f is considered sufficiently smooth when it is twice differentiable over its definition domain. Given a generic smooth surface f, there are many more or less intuitive ways of quantifying its smoothness, the most immediate being the calculation of the functional $J(f)$, defined as follows:

$$J(f) = \iint_{\Re^2} \left\{ \left(\frac{\partial^2 f}{\partial x_1^2}\right)^2 + 2\left(\frac{\partial^2 f}{\partial x_1 \partial x_2}\right)^2 + \left(\frac{\partial^2 f}{\partial x_2^2}\right)^2 \right\} \mathrm{d}x_1 \, \mathrm{d}x_2 \tag{16}$$

for a given point $\mathbf{x}$ of coordinates (x_1, x_2) in $\Re^2$. The finite character of $J(f)$ is guaranteed, given that the squares of f second derivatives are integrable over $\Re^2$. Furthermore, it can be demonstrated that $J(f)$ is positive definite and goes to zero if and only if f is a linear function. $J(f)$ offers a useful measure of the amount of roughness of f: in fact, it is somewhat intuitive that $J(f)$ will grow if the function f has pronounced local curvatures, since this corresponds to high values of the second derivatives. Moreover, it allows the quantification of a surface roughness not to be influenced by the addition of constant

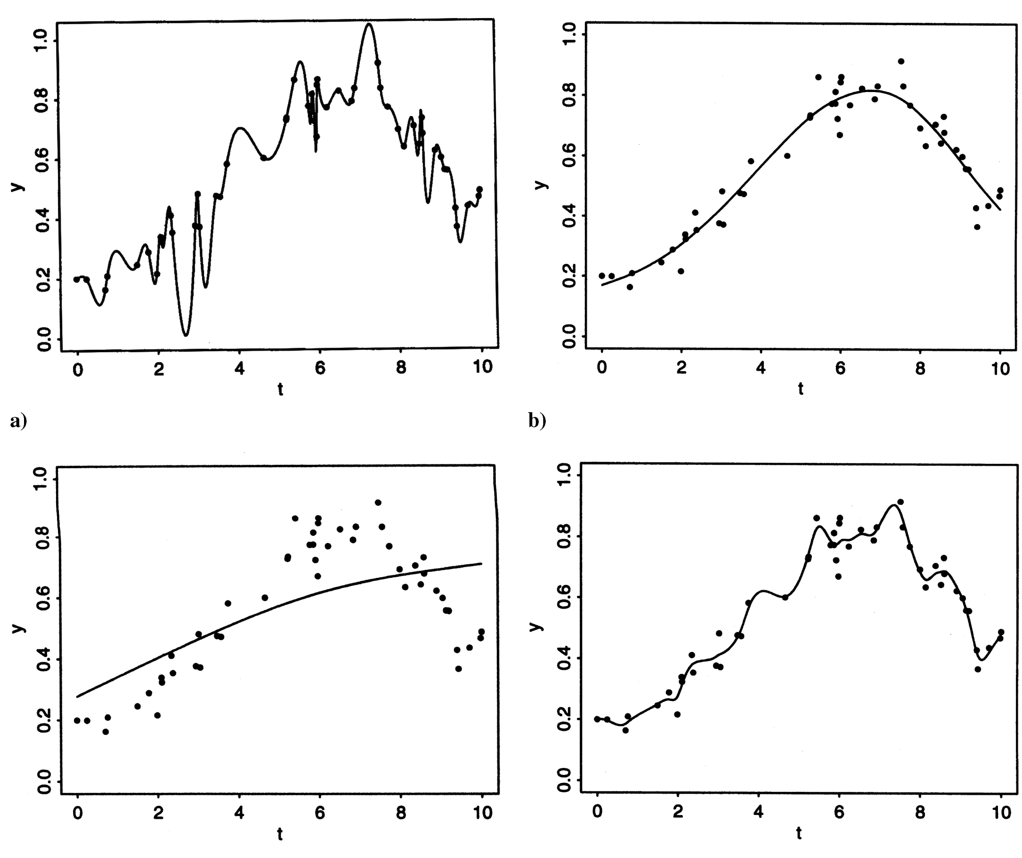

Fig. 1 Interpolation of data pairs by means of a curve a) with continuous second derivative, b) that minimizes $S(f)$ with a smoothing parameter equal to 1 and that minimizes $S(f)$ over the data pairs with c) a high value, and d) a small value of the smoothing parameter α.

terms or linear functions, as required in statistical regression. Obviously, other metrics can also be selected for roughness estimation, such as for instance the number of inflection points of the surface or the maximum absolute values of its second derivatives; however, $J(f)$ appears to be a more global measure and, in addition, shows some advantages from a computational standpoint.

In fact, from the definition of $J(f)$, a useful expression for the roughness penalty approach could be formulated in a straightforward way; given the observations' vector $\mathbf{y}$ over n input vectors $\mathbf{x}_i$, the following penalized sum of squared residuals can be defined:

$$S(f) = \sum_i \{y_i - f(\mathbf{x}_i)\}^2 + \alpha J(f) \tag{17}$$

where $\alpha > 0$ is a smoothing parameter and the functional $S(f)$ combines a term quantifying f goodness of data fitting to another one through which a penalization is assigned based on the response-surface roughness. The penalized least-squares model $\hat{f}$ is the one that, for a given value of α, minimizes the functional $S(f)$ over the class of functions f twice differentiable in $\Re^2$, giving the best compromise between data fitting and curve smoothness (see Fig. 1b for the one-dimensional case).

It clearly appears that the smoothing parameter α acts as a balancing factor between the two objectives: minimization of residual errors and abatement of the local fluctuations. High values of α make the penalty term become the dominant one in the definition of $S(f)$, so that the minimizer $\hat{f}$ will feature low curvature (Fig. 1c), approaching the linear regression model if $\alpha \to \infty$. On the other hand, when α relatively small, greater importance is given to the sum of squared residuals: as a consequence, the surface $\hat{f}$ will fit the data as close as possible, regardless of the potential variability introduced in the shape of the function (Fig. 1d). Some criteria that could drive the choice of the smoothing parameter value will be discussed in the following.

The main result of the roughness penalty approach is that the problem of minimizing $S(f)$ for a given α admits a unique solution that corresponds to a natural thin-plate spline. It is worth noting that such a formulation makes the identification of the response surface a fully deterministic problem, in which nothing needs to be imposed by the analyst except for the smoothing parameter, which, on the other hand, may be automatically determined, as will be illustrated below.

Another aspect that deserves to be mentioned is that no predefined shape (particularly, a TPS shape) has been a priori imposed to $\hat{f}$, as is the case for RB-TPS, but this rather comes out as a natural consequence of choosing $J(f)$ to quantify the surface roughness. Knowing that the minimizer of $S(f)$ has the form of a natural thin-plate spline is of outstanding importance, in that it allows to limit the search for $\hat{f}$ to a restricted class of functions, rather than to the whole space of smooth functions twice differentiable in $\Re^2$. Moreover, it can be demonstrated that a solution to the minimization problem over the domain of natural thin-plate splines does exist and can be univocally determined by solving a linear system of equations.

B. Choice of the Smoothing Parameter

The need for choosing a proper smoothing parameter (which can make the analyst feel quite uncomfortable, at least initially, with the roughness penalty approach) is actually not peculiar to this method, being rather intrinsic to model building from a given data set. For instance, choosing the polynomial degree in polynomial regression is substantially equivalent to identifying a value for α. The only difference here is that this parameter is inherent to the STPS approach, and hence it appears explicitly since the very early formulation of the problem. In one hand α is regarded for as a useful additional degree of freedom in model building, through which features with different dimensional scales can be successively investigated in the available data; thus, the value of the smoothing parameter can be selected on a rather subjective basis, according to those characteristics the analyst may need to emphasize. On the other hand, due to the somewhat arbitrary nature of α, the need could rise for an automatic procedure upon which the identification of its optimal value for an assigned observation set can be devolved; automatic procedures become even mandatory when using TPS routinely for handling huge amounts of data.

Many different procedures exist for determination of the smoothing parameter value, the most widely known of them being based on a cross-validation approach [13], which privileges the predictive capabilities of the model rather than data fitting. A regression surface $\hat{f}$ is optimal in the predictive sense if, under the hypothesis of zero mean random error, the estimated value $\hat{f}(\mathbf{x})$ over an unsampled input vector $\mathbf{x}$ is as close as possible to the observed value $\mathbf{y} = f(\mathbf{x})$. Hence, a good choice for $\hat{f}$ is the one that minimizes the mean squared error $\{\mathbf{y} - \hat{f}(\mathbf{x})\}^2$ for any vector point in the input domain.

When applying smoothing TPS to a single data set, usually no additional observations are available over which the cross-validation approach may be applied: this obstacle can be moved around if each of the assigned data is regarded for as a potential new observation. Specifically, a response surface, denoted as $\hat{f}^{(-i)}(\mathbf{x};\alpha)$, may be derived from the whole data set, except for the pair $\mathbf{x}_i, y_i$. This surface is the one that minimizes

$$\sum_{j \neq i} \{y_j - f(\mathbf{x}_j)\}^2 + \alpha J(f)$$

Now the predicted capability of $\hat{f}^{(-i)}(\mathbf{x};\alpha)$ may be judged over the input vector point $\mathbf{x}_i$ that has been omitted from the data when building the model. Extending this procedure to all the input point vectors while keeping the value of the smoothing parameter fixed gives a global measure of the model predictive efficiency. Specifically, a cross-validation score function may be defined as proposed by Green and Silverman [13]:

$$CV(\alpha) = n^{-1} \sum_{i=1}^{n} \{y_i - \hat{f}^{(-i)}(\mathbf{x}_i;\alpha)\}^2 \tag{18}$$

This represents the target function to be minimized in order to find the optimal value for α. The minimization problem formulated in Eq. (18) seems to require that n distinct smoothing functions $\hat{f}^{(-i)}$ are to be determined. However, the property of the generic smoothing spline $\hat{f}$ to be linearly dependent on the observations y_i can be demonstrated to result in a dramatic reduction of the complexity of the problem: actually, it is found that $CV(\alpha)$ can be expressed as

$$CV(\alpha) = n^{-1} \sum_{i=1}^{n} \left\{ \frac{y_i - \hat{f}(\mathbf{x}_i)}{1 - A_{ii}(\alpha)} \right\}^2 \tag{19}$$

where A_{ii} are the diagonal elements of the matrix $\mathbf{A}$ that maps the observations y_i over their estimated values $\hat{f}(\mathbf{x}_i)$. It follows that the cross-validation score function can be estimated from the residuals $\{y_i - \hat{f}(\mathbf{x}_i)\}$ of the regression smoothing spline $\hat{f}$ determined over the complete data set; hence, a unique smoothing problem needs to be solved to find the proper value for α.

A further approach for choosing the smoothing parameter is the so-called generalized cross-validation, which is actually a modified version of the cross-validation metric [49]. It stems from the idea of approximating each of the diagonal elements of the matrix $\mathbf{A}$ with their mean value. The generalized cross-validation score function to be minimized then becomes

$$\mathrm{GCV}(\alpha) = n^{-1} \frac{\sum_{i=1}^{n} \{y_i - \hat{f}(\mathbf{x}_i)\}^2}{\{1 - n^{-1} \operatorname{tr} A(\alpha)\}^2} \tag{20}$$

Even if the GCV approach was originally introduced due to its higher computational efficiency over classical cross-validation techniques, the distinction between the two methods is rather of a statistical nature. Actually, GCV seems more suitable for predictive purposes: in fact, in statistical regression, the diagonal elements of $\mathbf{A}$ are called leverage values, because they determine how much influential each observation y_i is in the determination of the estimated value $\hat{f}(\mathbf{x}_i)$. The estimated response in a point with a high leverage must be regarded carefully, because it turns out to be very sensitive to the actual value observed over this point. Unlike the cross-validation score function (which sums up with the same weight all the residuals of the estimated values obtained by omitting a point from the whole data set) the GCV score function can be written as

$$\mathrm{GCV}\,(\alpha) = n^{-1}\sum_{i=1}^{n}\left\{\left(\frac{1-A_{ii}(\alpha)}{1-n^{-1}trA(\alpha)}\right)^2\{y_i-\hat{f}^{(-i)}(\mathbf{x}_i)\}^2\right\} \quad (21)$$

From Eq. (21) it clearly appears that the residuals' weights over high leverage-value points are, in a sense, counterbalanced by a common term.

Another feature of GCV is the number of equivalent degrees of freedom, which gives an indication on the effective number of parameters that should be determined if the problem was solved using a parametric approach for a specified value of the smoothing parameter.

C. Multivariate STPS

The penalization of a response surface based on its roughness, from which the smoothing TPS approach originates, can be extended quite straightforwardly to functions of more than two variables. Consider the problem of finding out a proper, sufficiently smooth, estimator f to build up a model of the form $y_i = f(\mathbf{x_i}) + \text{error}$ $(i = 1, \ldots, n)$ for n observations y_i, where $\mathbf{x_i}$ is now a d-dimensional vector.

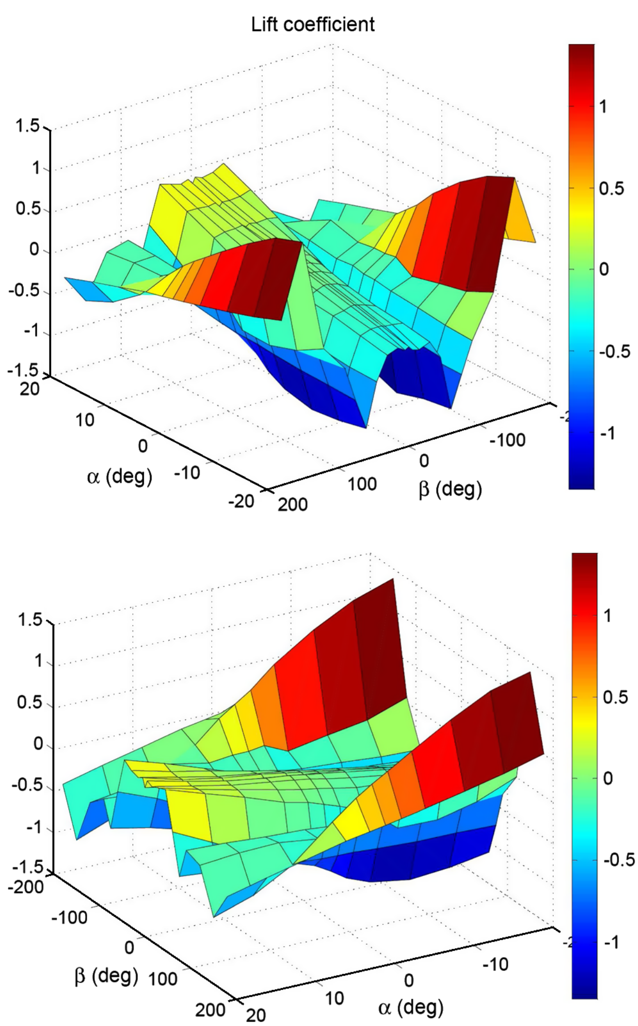

Fig. 2 Typical distribution of lift coefficient for a 3-D-shaped body as a function of angles of attack α and sideslip β.

As for the two-dimensional case, f is the solution of a minimization problem, in which the function to be minimized is a penalized sum of squares,

$$S_{md}(f) = \sum_{i=1}^{n}\{y_i - f(\mathbf{x_i})\}^2 + \alpha J_m(f) \quad (22)$$

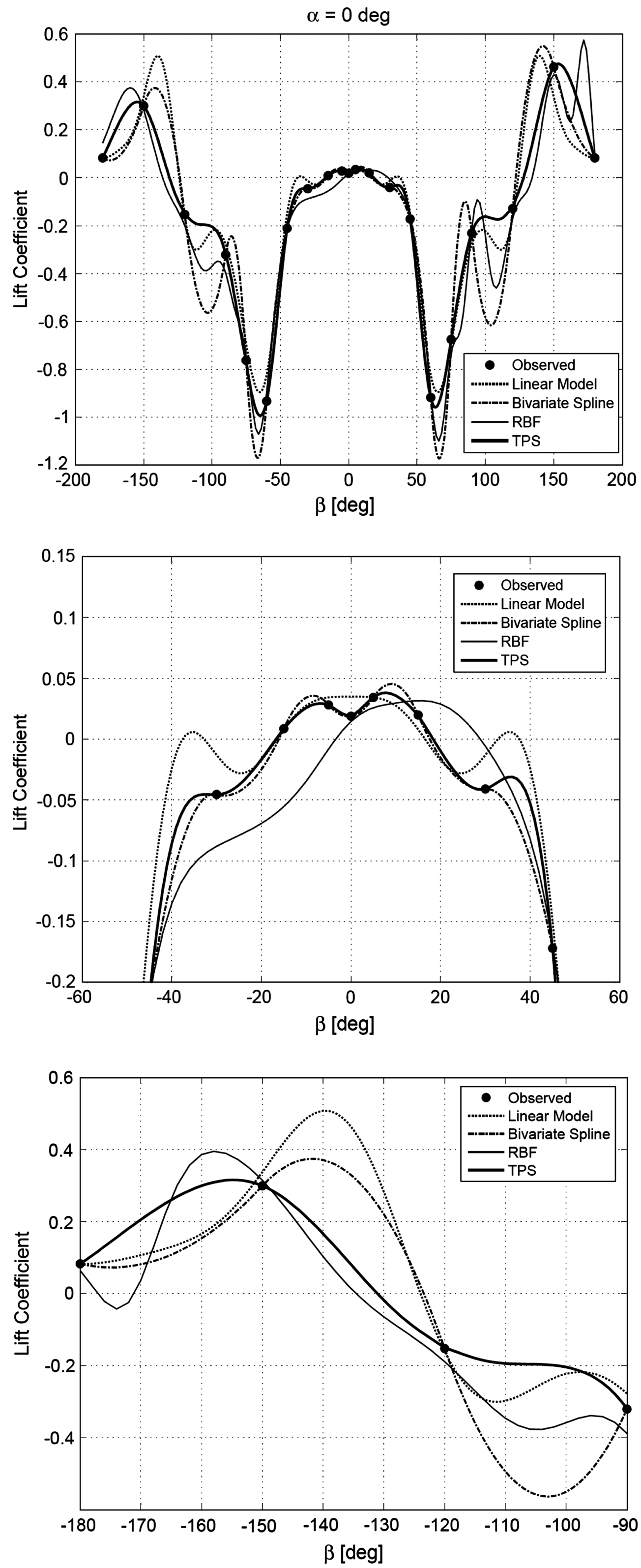

Fig. 3 Comparison among fitted values of the lift coefficient for $\alpha = 0°$. Overall view (top); zoom in the range $-60° < \beta < +60°$ (middle); zoom in the range $-180° < \beta < -90°$ (bottom). models used: LM, BS, RBF, and TPS.

and m is the order of the derivative (higher than the second) used to measure roughness.

The expression for $J_m(f)$ in Eq. (22) that indicates a penalty in d dimensions based on the mth derivative is

$$J_m(f) = \int \cdots \int_{\mathbb{R}^d} \sum \frac{m!}{\nu_1!, \ldots, \nu_d!} \left(\frac{\partial^m f}{\partial x_1^{\nu_1}, \ldots, \partial x_d^{\nu_d}} \right)^2 \mathrm{d}x_1, \ldots, \mathrm{d}x_d \tag{23}$$

where the sum within the integral is extended over all the nonnegative integers $\nu_1, \nu_2, \ldots, \nu_d$ such that $\nu_1 + \nu_2 + \cdots + \nu_d = m$. According to this definition, the only surfaces for which $J_m(f) = 0$ are polynomials of degree less than m. Moreover, it is necessary to impose the condition $2m > d$, so that roughness functionals J_m based on integrated first derivatives can be used only for one-dimensional problems, those based on integrated second derivatives only for three or less dimensions, and so on. The reason for that is expressed mathematically in terms of Beppo Levi and Sobolev spaces [50].

As usual, we can face the optimization problem expressed in Eq. (22) by considering a peculiar finite-dimensional class of functions f. Let us define a function η_{md}:

$$\eta_{md}(r) = \begin{cases} \theta r^{2m-d} \log r & \text{if } d \text{ is even} \\ \theta r^{2m-d} & \text{if } d \text{ is odd} \end{cases} \tag{24}$$

where the constant of proportionality θ has the following expression:

$$\theta = \begin{cases} (-1)^{m+1+\frac{d}{2}} 2^{1-2m} \pi^{-\frac{d}{2}} (m-1)!^{-1} \left(m - \frac{d}{2} \right)!^{-1} & \text{if } d \text{ is even} \\ \Gamma\left(\frac{d}{2} - m \right) 2^{-2m} \pi^{-\frac{d}{2}} (m-1)!^{-1} & \text{if } d \text{ is odd} \end{cases} \tag{25}$$

Finally, we define

$$M = \binom{m+d-1}{d}$$

and we focus our attention on a class $\{\phi_j, j = 1, 2, \ldots, M\}$ of linearly independent polynomials spanning the M-dimensional space of polynomials in $\mathbb{R}^d$ with degree less than m.

A function f on $\mathbb{R}^d$ is called a natural thin-plate spline of order m if it has the form

$$f(\mathbf{x}) = \sum_{i=1}^{n} \delta_i \eta_{md}(\|\mathbf{x} - \mathbf{x_i}\|) + \sum_{j=1}^{M} a_j \phi_j(\mathbf{x}) \tag{26}$$

and the coefficient vector δ satisfies the condition $\mathbf{T}\delta = 0$, where $T_{ij} = \phi_i(\mathbf{x_j})$.

Again, it can be demonstrated that, provided that the points $\mathbf{x_i}$ are distinct and sufficiently dispersed to determine a unique least-squares polynomial surface of degree $m - 1$, and under the condition $2m > d$, the function f that minimizes $J_m(f)$ under the constraint $f(\mathbf{x_i}) = y_i$ is a natural thin-plate spline of order m.

It follows that the function that minimizes S_{md} is a natural thin-plate spline of order m, with coefficient vectors $\mathbf{a}$ and δ uniquely identified by

$$\begin{bmatrix} \mathbf{E} + \alpha \mathbf{I} & \mathbf{T}^T \\ \mathbf{T} & 0 \end{bmatrix} \begin{pmatrix} \delta \\ a \end{pmatrix} \begin{pmatrix} \mathbf{y} \\ 0 \end{pmatrix} \tag{27}$$

where the generic $\mathbf{E}$ matrix element $E_{ij} = \eta_{md}(\|\mathbf{x_i} - \mathbf{x_j}\|)$.

VI. Example of Application

The case study that is analyzed hereafter is representative of some experimental data coming from a wind-tunnel campaign on a typical 3-D fuselage-shaped body, aimed at the acquisition of its aerodynamic coefficients over a rather extensive attitudes' range. Specifically, the application is focused on the body lift coefficient CL as a function of both angle of attack α and sideslip angle β (Fig. 2). The CL coefficient is chosen because of its pronounced multimodality, especially when the dependence on β is considered, with

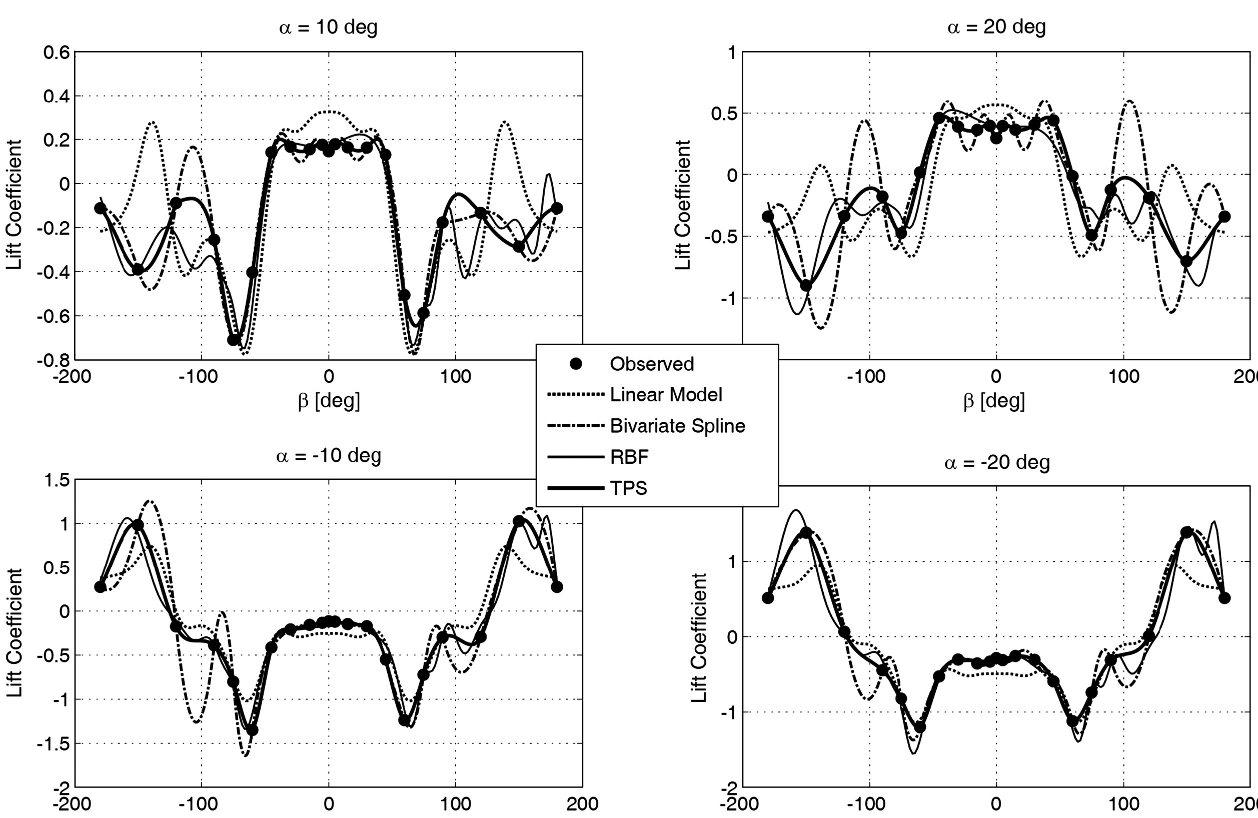

Fig. 4 Comparison among fitted values of the lift coefficient for $\alpha = \pm 10$ and $\pm 20°$. Models used: LM, BS, RBF, and TPS.

multiple function value inversions and sudden gradient changes. This makes the selected function an ideal candidate for a model building technique to be tested, especially when its predictive capability over unknown points is of interest. However, extension to other aerodynamic coefficients is straightforward.

The experimental data samples are closer at low-incidence angles and sparser at higher attitudes, as is typically the case for wind-tunnel acquisitions. In fact, costs related to wind-tunnel campaigns oblige to limit the total number of tests to be performed, so that more resources are commonly dedicated to understanding the aerodynamic behavior at those attitudes that are more usual in-flight conditions.

In the following, results of the implementation of the various techniques discussed above are presented and compared with each other. The resulting two-dimensional response surface is visualized in Figs. 3–6, in the form of a series of sectional cuts along the direction of constant α for different values of the angle of attack. In fact, the functional dependence of the lift coefficient on the sideslip angle β makes it possible to better appreciate the remarkable variability of the observations to be fitted.

A higher-order response-surface application, e.g., CL as function of three variables (two angles and Mach number), could also be considered in principle. However, for the sake of simplicity and the necessity for an immediate physical understanding, we shall refer to a bivariate example only.

First, a linear model, whose analytic expression is reported in Eq. (28), is applied to fit the experimental data:

$$\begin{aligned}CL = &-0.27782112 - 0.89774327\cos(\beta) - 0.02717918\sin(\alpha)\\ &- 3.25607\cos^2(\beta) + 0.94288\cos^3(\beta) - 1.36882\cos^3(\beta)\sin^3(\alpha)\\ &+ 78.53349\cos^6(\beta) - 204.80132\cos^8(\beta) + 192.82\cos^{10}(\beta)\\ &- 62.96\cos^{12}(\beta) - 0.06851\cos^{13}(\beta) + 1.73567\cos(\beta)\sin(\alpha)\end{aligned} \tag{28}$$

The corresponding determination coefficient is

$$R^2 = \frac{\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} = 0.8528$$

and the residual distribution is visualized in Fig. 7.

The model presented here derived from the elaboration of a series of parametric linear models having progressively increasing values of the determination coefficient as a result of their growing functional complexity. Actually, in spite of the apparent correlation of the measured data with the trigonometric functions in α and β, it is not possible to achieve a determination coefficient larger than that mentioned above, unless higher-order terms were introduced into the model. However, complicating further the model leads to extremely limited gains in terms of correlation with the observed data; moreover, a high functional complexity does not seem to be sound from an engineering standpoint. Actually, the need for a high value of the determination coefficient often contrasts with the requirement to keep to a minimum the complexity of the calculated functions, due to the emerging risk of data overfitting, as emphasized by the appearance of multiple local peaks (either maxima or minima) that do not make sense when a reasonable physical interpretation of the model is searched for. This is much more evident as the greatest attitude angles are considered.

As far as the bivariate-spline response surface is concerned, the number and location of the knot sequence are selected so as to correspond to the observed data, since reducing the number of knots results into response surfaces of poorer quality. It follows that the model presented in Figs. 3 and 4 actually represents the most satisfactory one among those analyzed in the class of bivariate splines. As can be observed, data fitting at low angles is acceptable for all values of α, while at higher sideslip angles, some undesirable local peaks appear (Fig. 3 bottom view and Fig. 4): even if the function values over the unsampled points are unknown, practical considerations seem to suggest that this is not a plausible physical behavior.

Regarding the RBF model, several options were investigated. The proposed model adopts a Gaussian basis function; however, it was found that the choice of basis function has negligible effects on the response function, as well documented in the literature. What is most influential is the choice of centers and widths. As far as the centers are concerned, two approaches were chosen: first, a stepwise procedure

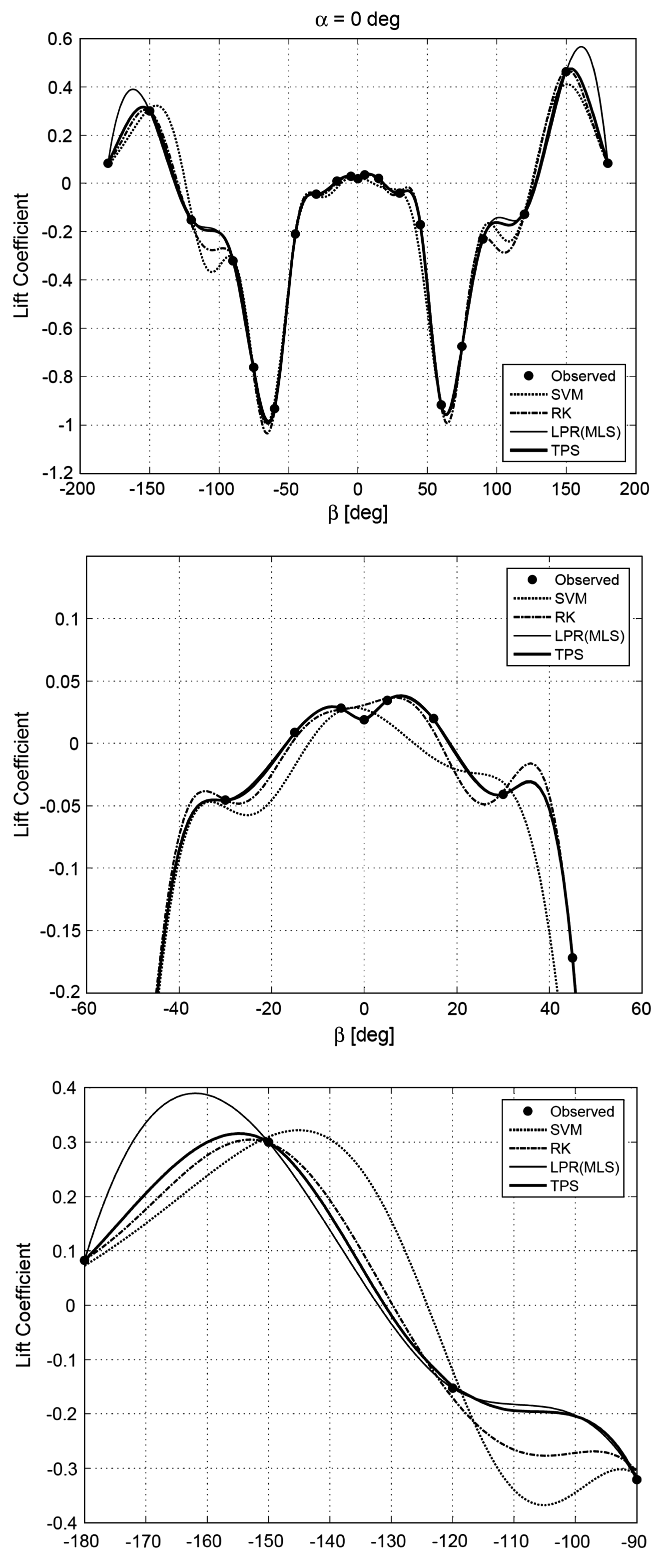

Fig. 5 Comparison among fitted values of the lift coefficient for $\alpha = 0°$. Overall view (top); zoom in the range $-60° < \beta < +60°$ (middle); zoom in the range $-180° < \beta < -90°$ (bottom). Models used: SVM, RK, LPR, and TPS.

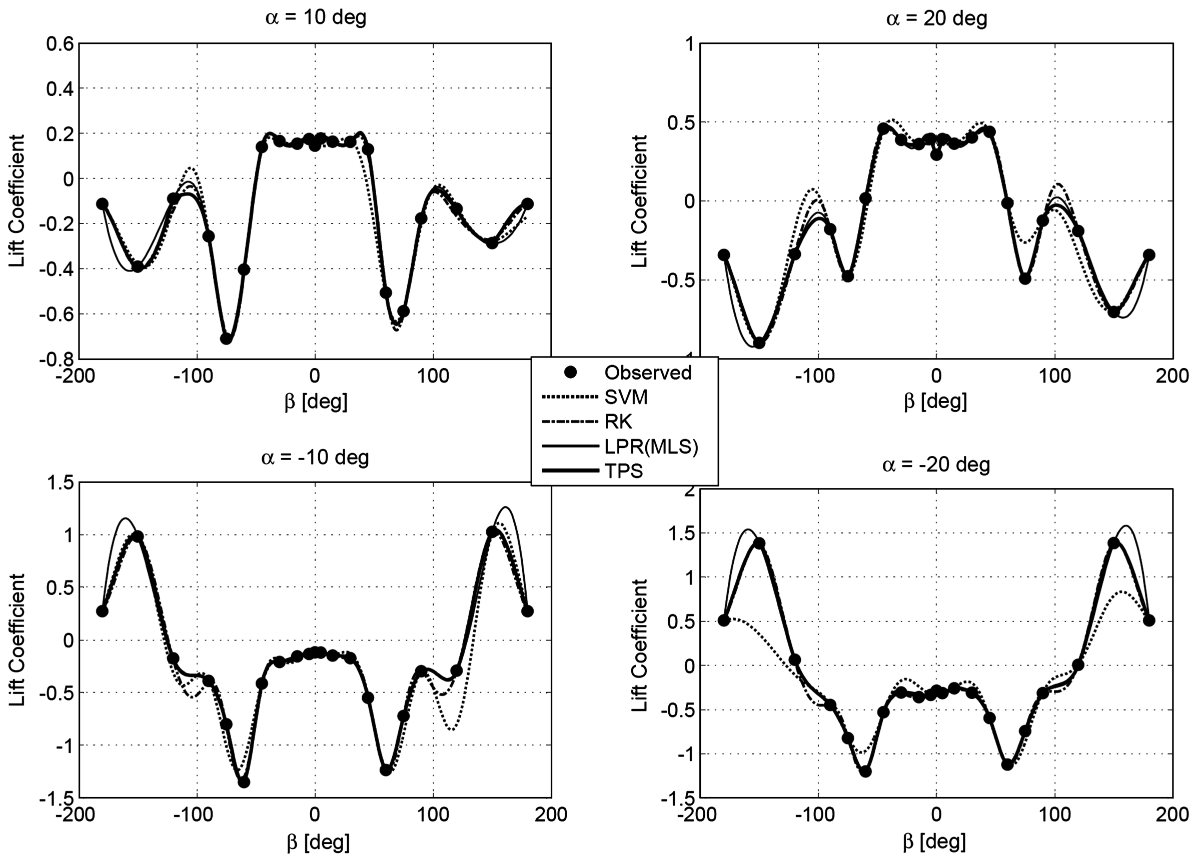

Fig. 6 Comparison among fitted values of the lift coefficient for $\alpha = \pm 10$ and $\pm 20°$. Models used: SVM, RK, LPR, and TPS.

was followed: i.e., starting from the whole group of observed data, different points were progressively excluded from the centers' sequence; second, a sequence of several clustering techniques (including the ROLS algorithm [47], mean-tracking clustering algorithm [51], and Fisher ratio class separability measure [52]) was adopted. Using the first approach, the best correlation at both low and high attitude angles was observed when all the experimental data were retained in the centers' sequence. When the others algorithms were used, the resulting surface was not satisfactory, in that it did not adequately replicate the experimental trend; moreover, predicted values were far from the true observations over the points that were excluded from the centers sequence. For determination of widths, a vector was created whose components were calculated on the basis of the Euclidean distance between successive point pairs and the corresponding optimal multiplying coefficients were searched for. When widths larger than the exact distances between points were used, the resulting response surface was excessively smooth and unable to capture the function variations adequately. On the other hand, if widths were selected as fractions of the distance the model is undesirably unstable. Thus, the final solution presented in Figs. 3 and 4 has widths exactly equal to the relative distances between points. It is worth noting that the multidimensionality of the problem renders quite difficult to achieve a satisfactory correlation over both directions in the input variables' domain. In fact, at $\alpha = 0°$ correlation at low β values is quite poor: in particular, the RBF model does not capture the symmetric behavior of the curve around $\beta = 0°$ with two local maxima, while this is not the case for the other values of incidence angles that are visualized. However, at $\alpha = 0°$ and when β approaches its extremity values (see bottom of Fig. 3), the appearance of doubtful and presumably meaningless local maxima/minima occurs. On the other hand, the correlation at higher sideslip angles seems more reasonable over all the investigated values of α. It is again worth underlying that the results presented using RBF are the best obtained by the authors in terms of adherence to the observations. When criteria others than the ones mentioned above were chosen for centers and width selection, the response surface was far away from being acceptable.

In Figs. 5 and 6, the results obtained using the more recent nonparametric fitting techniques are presented and compared again to those generated using the STPS technique in order to highlight the main differences as functions of the two independent variables. At a first glance, an apparent improvement in the predictive capability of the models implemented is found.

SVM regression was constructed using the following kernel function parameters: kernel function was always radial (in fact, the type of kernel function seemed to have negligible influence on the predictive model) and the ε-insensitive parameter was chosen as a function of the actual number and type of the training data so as to minimize the MSE [53,54]. In our case, several subsets of randomly selected training data were used featuring different cardinality. In particular, the following rules were used at the very beginning of model construction: observed data set was split into three subsets (in the ratio 50, 25, and 25%), each of which was used for training, validating, and testing the model, respectively. After this, we tried several subset cardinalities and ε values until the tested results produced the lowest possible MSE, as well as were the closest to the actual validation set outputs. As a result, we found that these goals were obtained using a random selection of 150 data points out of 231 observations (the whole set) for training the SVM and an ε value equal to 0.01. The remaining data were split into two halves and used for validating and testing the model, respectively. When larger training set were used, an undesirable overfitting was produced. In case of lower number of training data, the model begun to be very far away from the observed points. Finally, when adjusting the value of ε for a given training set, even very close to the one giving the lowest MSE, we registered a poor model prediction. From Figs. 5 and 6, a qualitative judgment of obtained results can be drawn. At null sideslip angles, the model is able to follow the trend given by the experimental values but it produces an unsatisfactory prediction overall, especially close to $\beta = 0$. Moreover, at higher model attitudes (particularly at $\alpha = -20°$) the prediction at higher angles of attack is completely erroneous. The reason for this seems clear: the training phase of a SVM requires some data to be excluded (either randomly or using some still-arbitrary criterion) so that the resulting

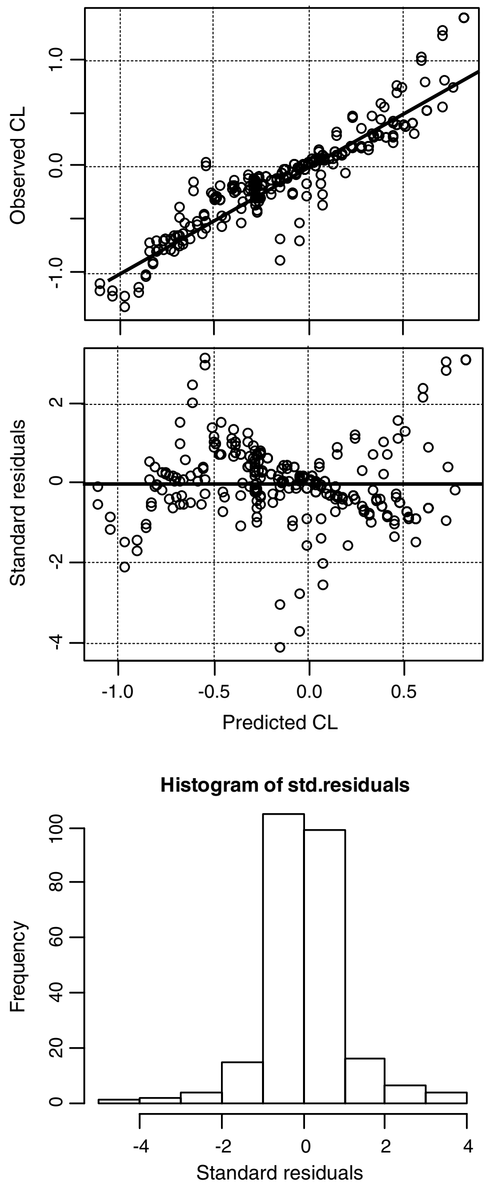

Fig. 7 Residual distribution for the linear model.

model will perform badly just in the neighborhood of the excluded points. This is crucial in cases, as ours, when number of available data is relatively low. From this point of view, all predictive methods based on supervised learning would probably behave poorly.

RK was implemented as a regressing Gaussian radial basis function using the Model Based Calibration Toolbox in MATLAB version 13 [55]. The procedure was the following: after choosing the kernel type, the selection of centers was performed using a reduced-error algorithm, in which a forward-selection procedure is performed until the largest data points residual was achieved; then a trial-widths algorithm was used to choose proper model widths, in which several widths values between lower and upper bounds were tested and iteratively selected to get the lowest GCV index using a zonal approach, such that the different widths are finally selected according to the local density of observed values; finally, an iterate Ridge criterion was adopted to select the optimal regularization parameter (which is, in fact, iterated until the minimum GCV index is obtained). Results depicted in Figs. 5 and 6 show a very good data fitting compared with the preceding methods even though predictions was missed at $\alpha = \beta = 0°$, a very important condition from an experimental point of view. Nevertheless, the model behaved reasonably well at both low and high angles, and seemed able to adjust to the experimental data with a noticeable degree of physical smoothness.

In LPR, a two-dimensional Epanechnikov kernel was chosen, i.e., $K(\alpha, \beta) = 2\max\{(1 - \alpha^2 - \beta^2), 0\}$, then $k = 2$ (existence and continuity of first-order derivative) polynomials were selected and, therefore, a cubic polynomial ($p = 3$) was used; finally, a cross-validation approach was adopted to obtain $h = 5.1$. As apparent from Figs. 5 and 6, the accuracy of LPR estimation is very good and somehow qualitatively equivalent to what has been obtained using STPS (discussed later). Someone could perhaps say that the quality of interpolation is superior to the one obtained using the methods presented so far. Equipollency of LPR and STPS is even surprising at $\alpha = 0$ in the range $-60° < \beta < +60°$ (see Fig. 5, middle), in which the two curves are almost overlapped. At lower values of the incidence angle (especially for $\beta < -150°$), in which observed points are sparser, LPR is less smoothed such that a presumable nonphysical local maximum occurs around $\beta < -163°$ (see Fig. 5, bottom). The same happens at very high sideslip angles ($\alpha = 0$ and $\beta > +150°$), in which the overestimation in the lift coefficient is even amplified. It is worth noting that for $\alpha = 0, -120° < \beta < -90°$, and $+90° < \beta < +120°$, the different local partial derivative of the response surface along β at the interval extremes leads to the appearance of a saddle point that seems a direct consequence of a mathematical soundness and robustness of both methods. In fact, within these intervals, the other methods cause the predicted function to have a, perhaps nonphysical or at least doubtful, local minimum. The same behavior is observed at different incidence angles α–s.

Finally, the STPS model is generated in a completely automated manner, as discussed before: the corresponding value of the smoothing parameter is $\alpha = 1.854 \times 10^{-6}$, as obtained by the GCV algorithm. The superiority of this model over the others is apparent (having in mind the arguments presented above, especially regarding LPR), in terms of both goodness of fit and regularity of the global behavior of the response surface.

Cross-validation tests were performed as described below. When a predictive model is built, validating its performance is perhaps the most delicate issue to be addressed. In fact, when selecting a validation technique, it is vital to keep in mind the purpose of such validation: i.e., to estimate the level of performance we may expect from generated models when such models are run on different data set of independent variables.

For this purpose, we proposed two kinds of distinct criteria: the first was based on overall data resampling, and the second was a local cross-validation test.

In data resampling, we used the well-known k-fold cross-validation test, sometimes referred to as simply cross-validation [56]. k-fold cross-validation stems from the idea of holdout testing in a clever way by rotating data through the process. To this purpose, data were randomly split into k equal-sized subsets, so that a train-test process was repeated k times, each time leaving a different segment of the data out, as the test set. Of the k subsamples, in fact, a single

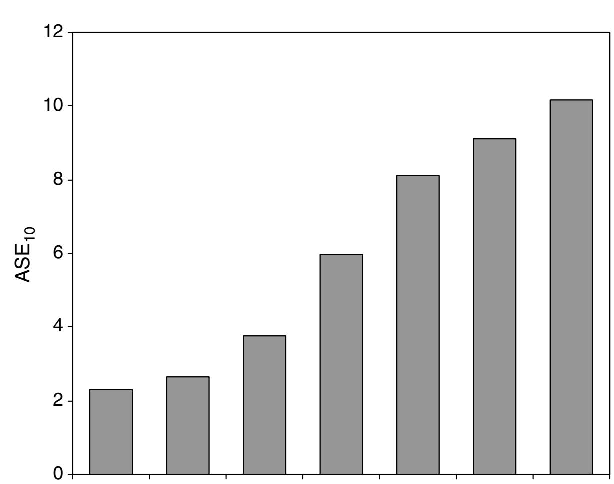

Fig. 8 Model validation using tenfold cross-validation: values of ASE_{10} for prediction models.

Table 1 Cases used in the cross-validation procedure and corresponding model errors

	Omitted point[a]	
	$\beta = -150°, \alpha = -15°$	$\beta = +15°, \alpha = +15°$
	Case 1	Case 2
LM	44.0	60.7
RBF	5.1	17.6
BS	39	30.2
SVM	37.5	10.4
RG	4.0	1.6
LPR	43.0	0.1
TPS	1.3	4.6

[a]Prediction error over the omitted point is $\varepsilon_i = |(y_i - \hat{f}(\mathbf{x}_i))/y_i|(\%)$.

subsample was retained as the validation data for testing the model, and the remaining $k - 1$ subsamples were used as training data. Specifically, we used $k = 10$, resulting in tenfold cross-validation such that observations were randomly assigned to 10 groups. The final performance measurement was taken as the absolute squared error (ASE10) across all 10 trials. Results are summarized graphically in Fig. 8. As can be seen, according to this performance metric the TPS approach gives the best prediction compared with the other models. It is worth noting that RBF is comparable in terms of goodness and, contrary to what is qualitatively judged from the response cuts, LM still behaves as a statistically sound model. This is due to the fact that in the parametric model, predictions are almost insensitive to the type of learning and test subsets. This does not mean that the model is good per se, but only that it is stable when predictions over new set of data are to be performed. In our case, this consideration does not apply to nonparametric models, with the exception of TPS and RBF, for which the predictive capability tends to degrade when samples of observations are omitted.

A local validation test was also performed: one single point was excluded at a time from the observed data and the models based on the formulations cited above were built on the modified observations' group. Then the resulting response surfaces were used to predict the response over the excluded point and the local percentage error between the estimated and the true value was evaluated as

$$\varepsilon_i = |(y_i - \hat{f}(\mathbf{x}_i))/y_i|(\%)$$

Unlike global metrics typical of classical statistical regression, this local measure seems to be suitable to the specific purpose of the paper, since models of different nature (parametric vs nonparametric) are to be compared. Moreover, when omitting the experimental acquisition over some specified attitudes during a wind-tunnel campaign, it is of practical interest to the analyst that the response surface is as robust as possible, giving an estimated value close to what would be really observed if the whole test matrix was sampled.

A series of points were selected to be omitted from the assigned data: for the sake of brevity, only two of them ($\alpha = -15°$ with $\beta = -150°$ and $\alpha = +15°$ with $\beta = +15°$), representative of the low and high sideslip-angle values, respectively, are reported here. The results are summarized in Table 1, in which the different methods are ranked adopting the metric cited above.

In the case of high sideslip-angle absolute values ($\alpha = -15°$ and $\beta = -150°$), the predictive performance of the STPS model is by far the most satisfactory among the investigated models, with percentage errors up to 1 order of magnitude less than those coming out from other techniques. A visual representation of the response-surface cuts obtained after omitting the aforementioned point is given in Figs. 9 and 10. On the other hand, at lower sideslip angles ($\alpha = +15°$ and $\beta = +15°$), both RK and LPR perform better than STPS (Figs. 11 and 12): however, this seems to be dependent on the peculiar criterion selected for cross-validation and in any case the percentage predictive error of STPS over the omitted point is less than 5%, which is judged satisfactory.

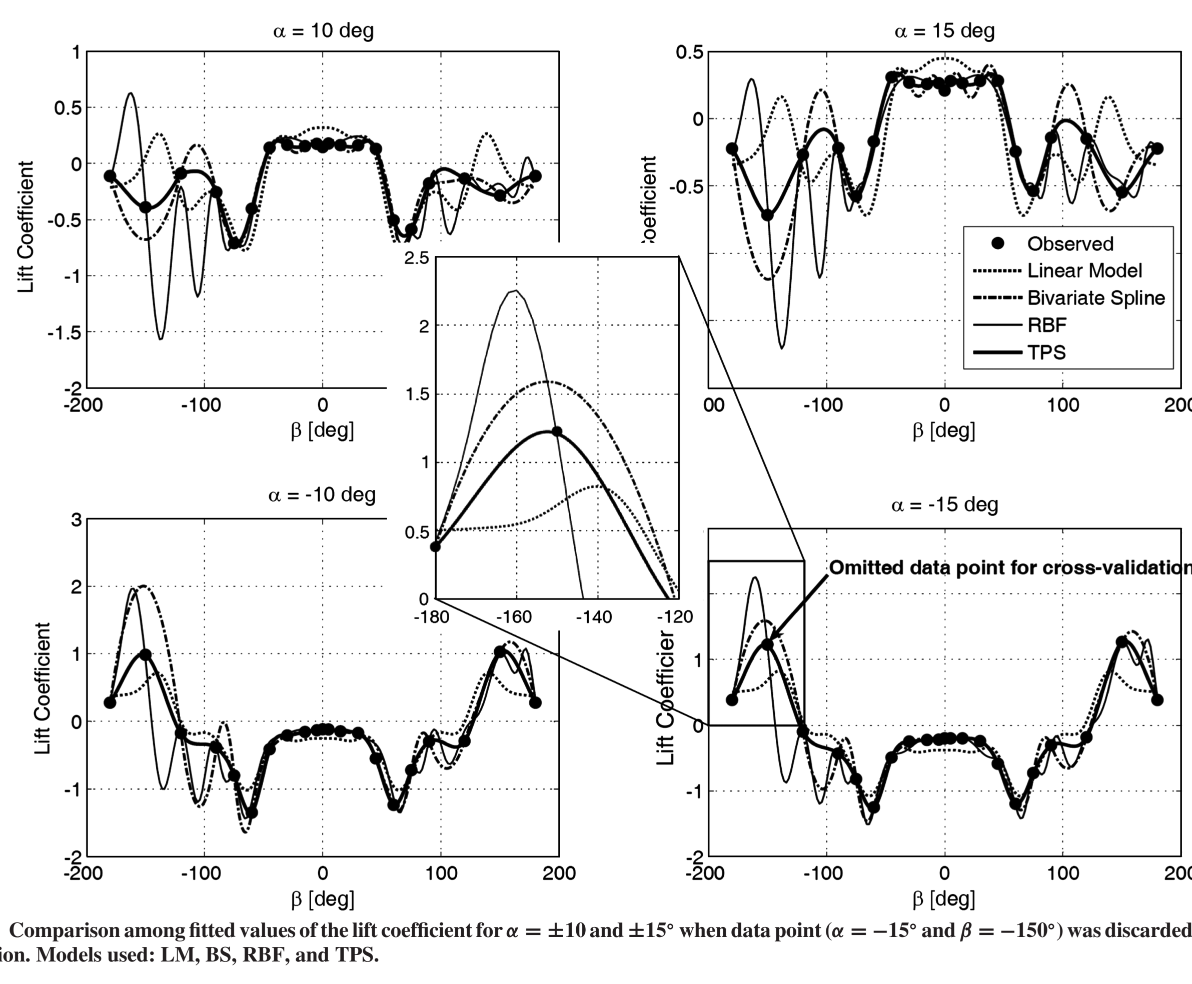

Fig. 9 Comparison among fitted values of the lift coefficient for $\alpha = \pm 10$ and $\pm 15°$ when data point ($\alpha = -15°$ and $\beta = -150°$) was discarded for cross-validation. Models used: LM, BS, RBF, and TPS.

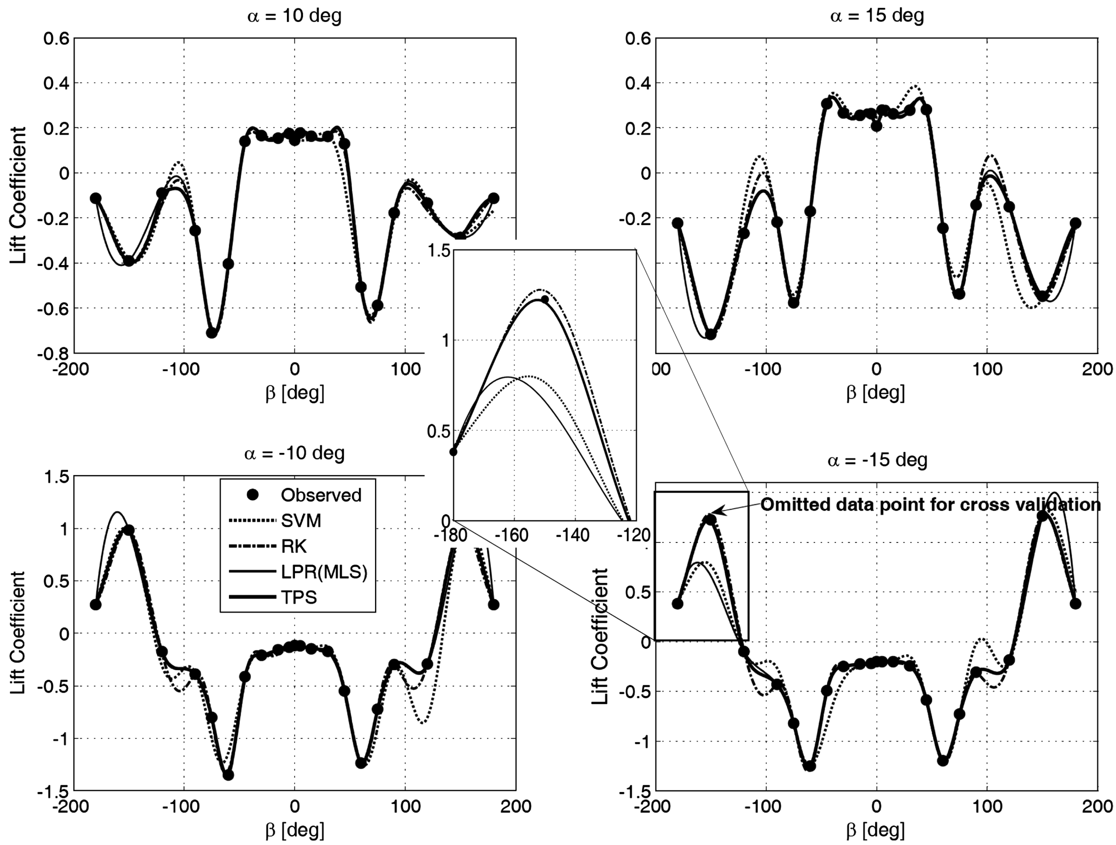

Fig. 10 Comparison among fitted values of the lift coefficient for $\alpha = \pm 10$ and $\pm 15°$ when data point ($\alpha = -15°$ and $\beta = -150°$) was discarded for cross-validation. Models used: SVM, RK, LPR, and TPS.

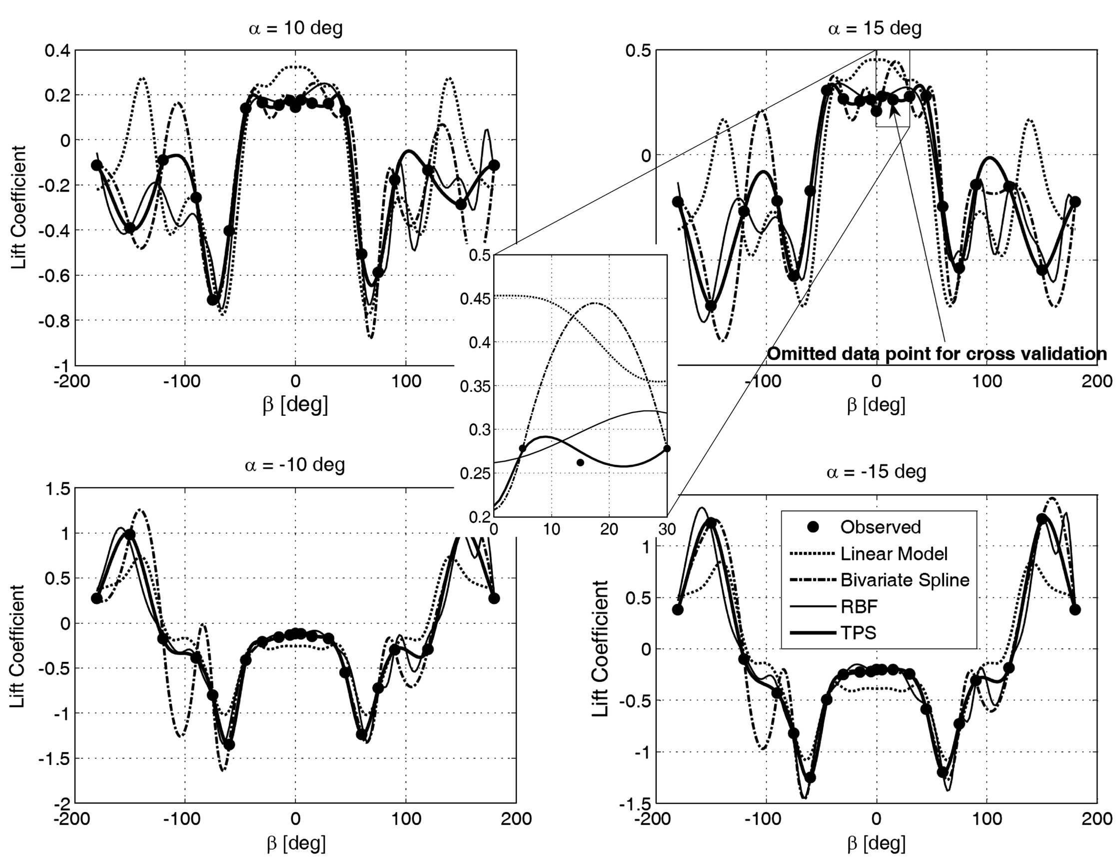

Fig. 11 Comparison among fitted values of the lift coefficient for $\alpha = \pm 10$ and $\pm 15°$ when data point ($\alpha = 15°$; $\beta = 15°$) was discarded for cross-validation. Models used: LM, BS, RBF, and TPS.

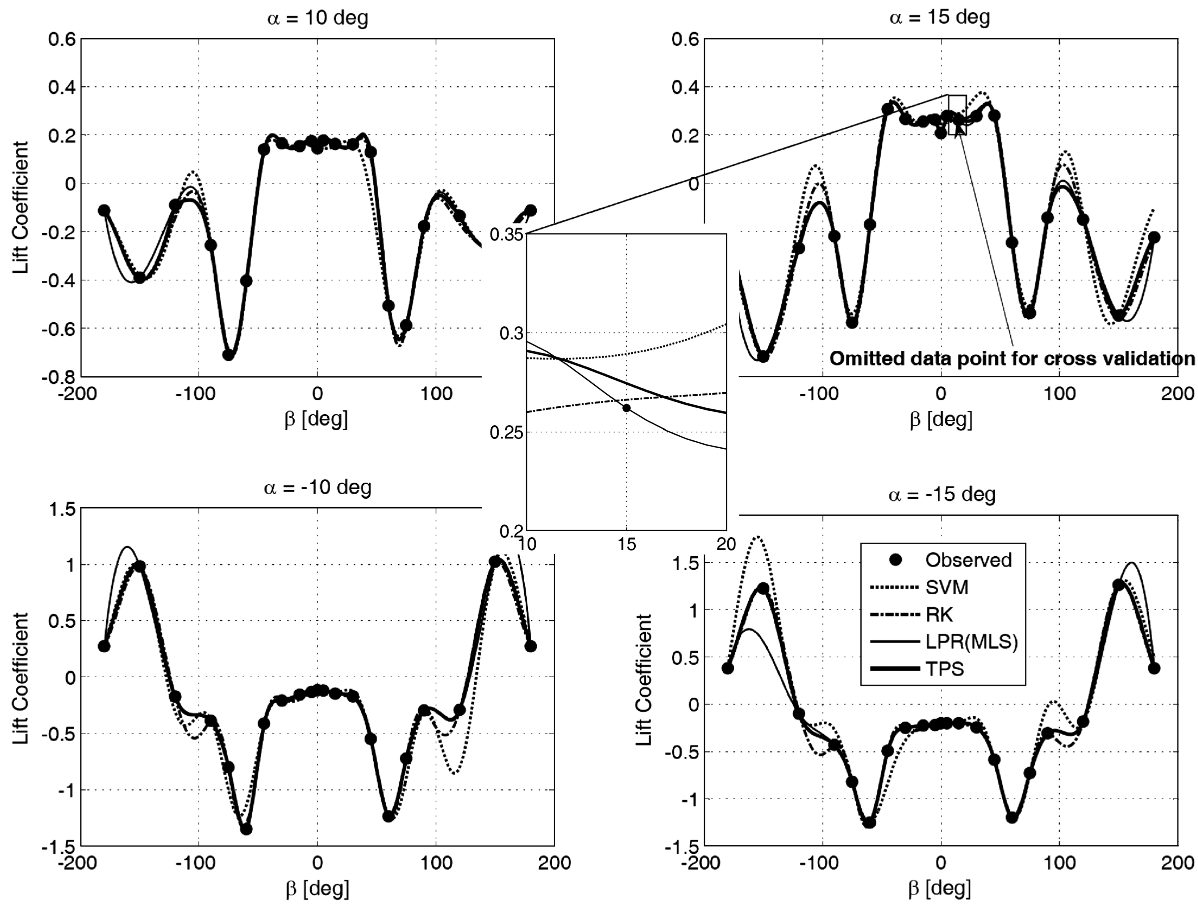

Fig. 12 Comparison among fitted values of the lift coefficient for $\alpha = \pm 10$ and $\pm 15°$ when data point ($\alpha = 15°$ and $\beta = 15°$) was discarded for cross-validation. Models used: SVM, RK, LPR, and TPS.

Moreover, while LPR exhibits the lowest percentage error in this second case, prediction over the preceding omitted point ($\alpha = -15°$ and $\beta = -150°$) is very unsatisfactory, with an error up to 43%; it follows that STPS gives a much better compromise for prediction over both low and high values of β. This confirms that STPS have significantly more pronounced stability and reliability with respect to the other techniques, especially when high angles of attack are considered.

VII. Conclusions

In this paper, a nonparametric method for robust fitting of aerodynamic data coming from either wind-tunnel or in-flight tests was presented. The formulation based on the multivariate smoothing thin-plate spline approach is relatively simple and easy to implement and straightforward to generalize to any arbitrary order. The main advantages over both parametric and other nonparametric approaches rely mainly in the deterministic result of its application, which does not imply any subjective choice of model parameters.

In particular, opposite to parametric linear models, the method presented does not require any arbitrary, and often biasing, choice of base functions for model construction. Also, contrary to other concurrent nonparametric methods (i.e., multivariate splines, radial basis function networks, support vector machines, regression kriging and local polynomial regression), knots number, centers, and width parameters are no longer needed, therefore keeping the aerodynamicist almost free of any subjectivity in building his/her model while staying close to the physical meaning of the problem being considered.

An application of the method was then proposed in a typical two-dimensional result matrix coming from a sample wind-tunnel campaign, in which the lift coefficient values of a representative 3-D streamlined body were to be properly fitted.

From the obtained results, the proposed approach was demonstrated both qualitatively (using graphical and engineering considerations) and quantitatively (by means of a cross-validation procedure) to significantly improve the fitting quality with respect to other techniques, in terms of reduced data overfitting and augmented robustness.

Finally, a meaningful improvement of the method is that the fitting produced when omitting one point at a time is significantly better than the one produced using other techniques, in that it leads to more robust and reliable predictions, at least at high angles of attack. This has also an important practical implication, in that the method could be successfully used to enforce the potentialities of an MDOE approach when an aerodynamic test campaign is to be planned with the minimum number of experimental observations.

References

[1] Hill, T., and Lewicki, P., *Statistics Methods and Applications*, StatSoft, Tulsa, OK, 2007.

[2] Myers, R. H., and Montgomery, D. C., *Response Surface Methodology*, 2nd ed., Wiley-Interscience, New York, 2002.

[3] Box, G. E., and Draper, N. R., *Empirical Model-Building and Response Surfaces*, Wiley, New York, 1987.

[4] Hussain, M. F., Barton, R. R., and Joshi, S. B., "Metamodeling: Radial Basis Functions, Versus Polynomials," *European Journal of Operational Research*, Vol. 138, No. 1, 2002, pp. 142–154.

[5] Landman, J., Simpson, D. V., and Parker, P., "Efficient Methods for Complex Aircraft Configuration Aerodynamic Characterization Using Response Surface Methodologies," 44th AIAA Aerospace Sciences Meeting and Exhibit, Reno, NV, AIAA Paper 2006-922, Jan. 2006.

[6] DeLaurentis, D., Mavris, D. N., and Schrage, D. P., "System Synthesis in Preliminary Aircraft Design Using Statistical Methods," *20th International Council of the Aeronautical Sciences (ICAS) Congress*, Sorrento, Italy, Sept. 1996.

[7] Rallabhandi, S. K., Cagaty, E., and Mavris, D. N., "An Improved Procedure for Prediction of Drag Polars of a Joined Wing Concept

Using Physics-Based Response Surface Methodology," Society of Automotive Engineers Paper 2001-01-3015, 2001.
[8] Mavris, D. N., and Qiu, S., "An Improved Process for the Generation of Drag Polars for Use in Conceptual/Preliminary Design," Society of Automotive Engineers, Paper 1999-01-5641, 1999.
[9] Dowgwillo, R. M., and DeLoach, R., "Using Modern Design of Experiments to Create a Surface Pressure Database from a low Speed Wind Tunnel Test," 24th AIAA Aerodynamic Measurement Technology and Ground Testing Conference, Portland, OR, AIAA Paper 2004-2200, June 2004.
[10] Landman, D., Simpson, J., Mariani, R., Ortiz, F., and Britcher, C., "A High Performance Aircraft Wind Tunnel Test Using Response Surface Methodologies," AIAA Paper 2005-7602, Dec. 2005.
[11] Eubank, R. L., *Nonparametric Regression and Spline Smoothing*, Marcel Dekker, New York, 1999.
[12] Bowman, A. W., and Azzalini, A., *Applied Smoothing Techniques for Data Analysis*, Clarendon, Oxford, 1997.
[13] Green, P. J., and Silverman, B. W., *Nonparametric Regression and Generalized Linear Models*, Chapman and Hall, London, 1994.
[14] Friedman, J. H., "Multivariate Adaptive Regression Splines," *Annals of Statistics*, Vol. 19, No. 1, 1991, pp. 1–67.
doi:10.1214/aos/1176347963
[15] Klein, V., "Estimation of Aircraft Aerodynamic Parameters from Flight Data," *Progress in Aerospace Sciences*, Vol. 26, No. 1, 1989, pp. 1–77.
doi:10.1016/0376-0421(89)90002-X
[16] Raisinghani, S. C., Ghosh, A. K., and Kalra, P. K., "Two New Techniques for Parameter Estimation Using Neural Networks," *The Aeronautical Journal*, Vol. 102, No. 1011, 1998, pp. 25–29.
[17] Raisinghani, S. C., Ghosh, A. K., and Khubchandani, S., "Estimation of Aircraft Lateral-Directional Parameters Using Neural Networks," *Journal of Aircraft*, Vol. 35, No. 6, 1998, pp. 876–881.
doi:10.2514/2.2407
[18] Vijaykumar, M., Omkar, S. N., Ganguli, R., Sampath, P., and Suresh, S., "Identification of Helicopter Dynamics Using Recurrent Neural Networks and Flight Data," *Journal of the American Helicopter Society*, Vol. 51, No. 2, 2006, pp. 164–174.
doi:10.4050/JAHS.51.164
[19] Mullur, A. A., and Messac, A., "Extended Radial Basis Functions: More Flexible and Effective Metamodeling," *AIAA Journal*, Vol. 43, No. 6, 2005, pp. 1306–1315.
doi:10.2514/1.11292
[20] Rocha, H., Li, W., and Hahn, A., "Principal Component Regression for Fitting Wing Weight Data of Subsonic Transports," *Journal of Aircraft*, Vol. 43, No. 6, 2006, pp. 1925–1936.
doi:10.2514/1.21934
[21] Praveen, C., and Duvigneau, R., "Radial Basis Functions and Kriging Metamodels for Aerodynamic Optimization," Institut National de Recherche en Informatique et en Automatique, Rept. 6151, Paris, 2007.
[22] Mullur, A. A., and Messac, A., "Extended Radial Basis Function: More Flexible and Effective Metamodeling," *AIAA Journal*, Vol. 43, No. 6, 2005, pp. 1306–1315.
doi:10.2514/1.11292
[23] Forrester, A. J., and Keane, A. J., "Recent Advances in Surrogate-Based Optimization," *Progress in Aerospace Sciences*, Vol. 45, 2009, pp. 50–79.
doi:10.1016/j.paerosci.2008.11.001
[24] Cristianini, N., and Shawe-Taylor, J., *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*, Cambridge Univ. Press, New York, 2000.
[25] Smola, A. J., and Schölkopf, B., "A Tutorial on Support Vector Regression," *Statistics and Computing*, Vol. 14, No. 3, 2004, pp. 199–222.
doi:10.1023/B:STCO.0000035301.49549.88
[26] Suykens, J. A. K., Horvath, G., Basu, S., Micchelli, C., and Vandewalle, J. (eds.), "Advances in Learning Theory: Methods, Models and Applications," *NATO Science Series III: Computer & Systems Sciences*, Vol. 190, IOS Press Amsterdam, 2003.
[27] Cressie, N. A. C., "The Origins of Kriging," *Mathematical Geology*, Vol. 22, 1990, pp. 239–252.
doi:10.1007/BF00889887
[28] Clark, I., and Harper, W. V., *Practical Geostatistics 2000*, Ecosse, Columbus, OH, 2000.
[29] Forrester, A. I. J., Keane, A. J., and Bressloff, N. W., "Design and Analysis of 'Noisy' Computer Experiments," *AIAA Journal*, Vol. 44, No. 10, 2006, pp. 2331–2339.
doi:10.2514/1.20068
[30] Hu, J., "Methods of Generating Surfaces In Environmental GIS Applications," *Proceedings of the 1995 ESRI Conference* [CD-ROM], Palm Springs, CA, 1995, http://proceedings.esri.com/library/userconf/proc95/to100/p089.html.
[31] Toropov, V. V., Schramm, U., Sahai, A., Jones, R. D., and Zeguer, T., "Design Optimization and Stochastic Analysis Based on the Moving Least Squares Method," *6th World Congress of Structural and Multidisciplinary Optimization* [CD-ROM], Rio de Janeiro, 2005.
[32] Cleveland W. S., "Robust Locally Weighted Regression and Smoothing Scatterplots," *Journal of the American Statistical Association*, Vol. 74, 1979, pp. 829–836.
[33] Fan, J., Gasser, T., Gijbels, I., Brockmann, M., and Engel, J., "Local Polynomial Regression: Optimal Kernels And Asymptotic Minimax Efficiency," *Annals of the Institute of Statistical Mathematics*, Vol. 49, No. 1, 1997, pp. 79–99.
[34] Fan, J., and Gijbels, I., *Local Polynomial Modelling and Its Applications*, Chapman and Hall, Boca Raton, FL, 1996.
[35] Ruppert, D., Sheather, S. J., and Wand, M. P., "An Effective Bandwidth Selector for Local Least Squares Regression," *Journal of the American Statistical Association*, Vol. 90, No. 432, 1995, pp. 1257–1270.
[36] DeLoach, R., "MDOE Perspectives on Wind Tunnel Testing Objectives," 22nd AIAA Aerodynamic Measurement Technology and Ground Testing Conference, St. Louis, MO, AIAA Paper 2002-2796, June 2002.
[37] DeLoach, R., and Berrier, B. L., "Productivity and Quality Enhancement in a Configuration Aerodynamic Test Using the Modern Design of Experiments," 42nd AIAA Aerospace Sciences Meeting and Exhibit, Reno, NV, AIAA Paper 2004-1145, Jan. 2004.
[38] DeLoach, R., "The Modern Design of Experiments for Configuration Aerodynamics: A Case Study," 44th AIAA Aerospace Sciences Meeting and Exhibit, Reno, NV, AIAA Paper 2006-0923, Jan. 2006.
[39] Giunta, A. A., and Watson, L. T., "A Comparison of Approximation Modeling Techniques: Polynomial Versus Interpolating Models," *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis & Optimization*, Vol. 1, AIAA, Reston, VA, Sept. 1998, pp. 392–404.
[40] de Boor, C., *B-Form Basics, Geometric Modeling: Algorithms and New Trends*, edited by G. Farin, SIAM Publication, Philadelphia, 1987, pp. 131–148.
[41] de Boor, C., and Höllig, K., "Approximation Power of Smooth Bivariate PP Functions," *Mathematische Zeitschrift*, Vol. 197, 1988, pp. 343–363.
doi:10.1007/BF01418335
[42] Powell, M. J. D., "Radial Basis Functions for Multivariable Interpolation: A Review," *Algorithms for Approximation*, edited by J. C. Mason, and M. G. Cox, Clarendon, Oxford, 1987, pp. 143–167.
[43] Broomhead, D. S., and Lowe, D., "Multivariate Functional Interpolation and Adaptive Networks," *Complex Systems*, Vol. 2, 1988, pp. 321–355.
[44] Hassoun, M., *Fundamentals of Artificial Neural Networks*, MIT Press, Cambridge, MA, 1995.
[45] Orr, M. J. L., "Introduction to Radial Basis Function Networks," Univ. of Edinburgh, 1996, http://www.anc.ed.ac.uk/rbf/rbf.html#pprs.
[46] Orr, M., "Regularisation in the Selection of Radial Basis Function Centers," *Neural Computation*, Vol. 7, No. 3, pp. 954–975, 1995.
[47] Chen, S., Chng, E. S., and Alkadhimi, "Regularized Orthogonal Least Squares Algorithm for Constructing Radial Basis Function Networks," *International Journal of Control*, Vol. 64, No. 5, 1996, pp. 829–837.
doi:10.1080/00207179608921659
[48] Orr, M., "Optimizing the Widths of Radial Basis Functions," *Proceedings of the 5th Brazilian Symposium on Neural Networks*, Belo Horizonte, Brazil, pp. 26–29, 1998.
[49] Wahba, G., *Spline Models for Observation Data*, Society for Industrial and Applied Mathematics, Philadelphia, 1990.
[50] Meinguet, J., "Multivariate Interpolation of Arbitrary Points Made Simple," *Zeitschrift fur Angewandte Mathematik und Physik*, Vol. 30, 1979, pp. 292–304.
doi:10.1007/BF01601941
[51] Sutanto, E. L., Mason, J. D., and Warwick, K., "Mean-Tracking Clustering Algorithm for Radial Basis Function Centre Selection," *International Journal of Control*, Vol. 67, No. 6, 1997, pp. 961–977.
doi:10.1080/002071797223884
[52] Mao, K. Z., "RBF Neural Network Center Selection Based on Fisher Ratio Class Separability Measure," *IEEE Transactions on Neural Networks*, Vol. 13, No. 5, 2002, pp. 1211–1217.
doi:10.1109/TNN.2002.1031953
[53] Schölkopf, B., Smola, A. J., Williamson, R. C., and Bartlett, P. L., "New Support Vector Algorithms," *Neural Computation*, Vol. 12, No. 5, 2000, pp. 1207–1245.
doi:10.1162/089976600300015565

[54] Schölkopf, B., Williamson, R. C., Smola, A. J., and Shawe-Taylor, J., "SV Estimation of a Distribution's Support," *Neural Information Processing Systems 12*, edited by S. A. Solla, T. K. Leen, and K.-R. Müller, MIT Press, Cambridge, MA, 2000.

[55] MATLAB, Software Package, Ver. 13, The Mathworks, Inc., Natick, MA, 2002.

[56] Picard, R., and Cook, D., "Cross-Validation of Regression Models," *Journal of the American Statistical Association*, Vol. 79, No. 387, 1984, pp. 575–583.

A. Tumin
Associate Editor