



**HAL**  
open science

## Semiparametric estimation of plane similarities: application to fast computation of aeronautic loads

Edouard Fournier, Stéphane Grihon, Thierry Klein

► **To cite this version:**

Edouard Fournier, Stéphane Grihon, Thierry Klein. Semiparametric estimation of plane similarities: application to fast computation of aeronautic loads. *Statistics*, 2019, 53 (5), pp.1168-1186. 10.1080/02331888.2019.1632859 . hal-01879135v3

**HAL Id: hal-01879135**

**<https://hal.science/hal-01879135v3>**

Submitted on 4 Apr 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Semiparametric estimation of plane similarities: Application to fast computation of aeronautic loads

Edouard Fournier<sup>\*1,2,3</sup>, Stéphane Grihon<sup>†2</sup>, and Thierry Klein<sup>‡ 1,3</sup>

<sup>1</sup>Institut de Mathématique, UMR5219; Université de Toulouse;  
CNRS, UPS IMT, F-31062 Toulouse Cedex 9, France

<sup>2</sup>Airbus France 316, Route de Bayonne, Toulouse France

<sup>3</sup>ENAC - Ecole Nationale de l'Aviation Civile, Université de  
Toulouse, France

## Abstract

In the big data era, it is often needed to resolve the problem of parsimonious data representation. In this paper, the data under study are curves and the sparse representation is based on a semiparametric model. Indeed, we propose an original registration model for noisy curves. The model is built transforming an unknown function by plane similarities. We develop a statistical method that allows to estimate the parameters characterizing the plane similarities. The properties of the statistical procedure is studied. We show the convergence and the asymptotic normality of the estimators. Numerical simulations and a real life aeronautic example illustrate and demonstrate the strength of our methodology.

*Keywords:* Semiparametric model, Registration of curves, Statistical learning of a physical system

*MSC Classification:* 62F12, 62F30, 62P30

## 1 Introduction

It may be useful, when dealing with a large set of curves differing slightly from each other, to exhibit a parametric model of transformations. Indeed, usually a reduction taking into account some prior knowledge on the curves may lead

---

\*edouard.fournier@airbus.com

†stephane.grihon@airbus.com

‡thierry.klein@math.univ-toulouse.fr or thierry01.klein@enac.fr

to a better understanding of the variability within the population. One way to perform such reduction consists in considering that the set of curves has been obtained by deforming a template with parametric transformations. This point of view is usually called curve registration and has been widely studied in statistics. Two main statistical tasks are generally considered: the estimation of the template, and the estimation of transformation parameters.

For example, Härdel [12] considers the case of semiparametric models where the curves are nonparametric but are related in a parametric manner: the  $x$ -axis is shifted and the ordinate axis is rescaled. Regarding Golubev in [10], the idea is to estimate the period of an unknown signal with white noise conditions. In [8], Gasser and Kneip study the shape estimation of such templates, i.e structural features (extrema, inflection point) which can occur in a consistent manner among a sample of curves. The paper [16] introduces the shape invariant model (SIM). In this paper, the authors estimate the deformation parameters and the unknown model function at the same time for a parametric family. In [14], Kneip considers a general non linear parametric regression model with unknown template function. Many authors have worked on the SIM estimating either the template, the parameters or both. We refer to [14, 17, 12, 13, 7] for more details and examples on the SIM. More recently, these estimation topics have seen a resurgence pushed by applications in image and signal processing. We refer for example to [11, 19, 1, 7, 4, 5, 9, 26] for models and techniques related to signal processing problems and the multidimensional extension of SIM defined on the plane. Concerning non parametric approaches, we refer to [21, 15]. In these papers, a nonlinear regression method is developed. It allows to aligned features of curves through the use of monotone functions acting on the abscissa axis (also called Dynamic Time Warping in the literature).

Hence, the set of transformations often consists in a parametric family of operator acting on curves [3]. In this paper, the parametric transformations involve rotation and scaling parameters. This model has been studied for image registration see for example [5] or [19]. According to our knowledge, all the models studied in the literature considered transformations that act independently on the argument and on the value of the template function. In this work, we considered a family of parametric transformations that act *jointly* on the argument and on the value of the template function. We work on functions defined on  $[0, 1]$  and with plane similarities acting on the whole curve  $\tilde{C} := (x, \tilde{f}(x))_{x \in [0, 1]}$  representing the template.

Our aim is twofold:

- *Estimation of the deformation parameters*: we address the estimation of the deformation parameters and study the asymptotic properties of the estimator when the template is known on  $[0, 1]$ . Then, using these preliminary results we study the consistency and the asymptotic normality of our estimator in the more realistic case where the reference curve is

defined on the same grid as the set of curves to be registered. In this paper, we consider that the set of curves to be registered are noisy but not the template. By supposing the legitimate existence of this template, it allows us to estimate the deformation parameters through a  $M$ -estimation procedure (see [25]).

- *Building a meta-model*: we will use the estimation of the deformation parameters in a real world application - the prediction of aeronautic loads. The idea will be to use and predict the transformation parameters for different aircraft configurations. Using a proper representation of the problem, we get as good results as classical methods of dimension reduction techniques. Additionally, our methodology endows aerospace engineers with a better understanding of the variability within the set of load curves.

Our paper is organized as follow: in Section 2, we define our framework and the model of transformations. On the regression model, we use a  $M$ -estimation procedure to perform the estimation and study the asymptotic behavior of the estimators in two cases: when the reference curve  $\tilde{C}$  is known, and when it is defined on the same grid as the set of curves to be registered. Section 3 is devoted to examples. We first compare our procedure to the Dynamic Time Warping method studied in [21] and using the **fd**a package [22]. It appears, in particular, that for the class of functions we consider, our deformation strategy outperforms the warping method in the case where the noise does not exceed a certain threshold. Then, we apply our methodology to the prediction of aeronautic loads (see [6]): for different aircraft configurations we will predict the deformation parameters allowing us to compute the external constraints affecting the wing structure.

All the proofs are postponed to the last section.

## 2 Framework, model and analytic results

In this section, we describe the statistical model studied and give the asymptotic behavior of the  $M$ -estimators of the unknown parameters.

### 2.1 The observations

**Notation.** Let be  $x \in [0, 1]$ , and  $f : [0, 1] \rightarrow \mathbb{R}^+$ . We denote by  $C$  the curve

$$C := \begin{pmatrix} x \\ f(x) \end{pmatrix}_{x \in [0,1]}.$$

In our framework, we have at hand  $K + 1$  curves. One is the reference curve denoted by  $\tilde{C}$  that will also be called pattern. The  $K$  other curves

$C_j$ ,  $j = 1, \dots, K$  are assumed to be the images of  $\tilde{C}$  by the transformation model described below. The  $K$  curves are observed on the same random grid  $\mathcal{D}_N := \{X_1, \dots, X_N\}$  where  $(X_i)_{i=1, \dots, N}$  are iid random variables with uniform distribution on  $[0, 1]$ . Hence we have at our disposal

$$C_j^N = (X_i, f_j(X_i))_{i=1, \dots, N}, j=1, \dots, K.$$

We will consider the following two cases

- i)  $\tilde{C}$  is known everywhere:  $\tilde{C} = (x, \tilde{f}(x)), \forall x \in [0, 1]$ ,
- ii)  $\tilde{C}$  is only known on  $\mathcal{D}_N$ :  $\tilde{C} = (X_i, \tilde{f}(X_i))_{i=1, \dots, N}$ .

## 2.2 Transformation model

Before defining the transformation model linking the  $K$  observed curves to the pattern  $\tilde{C}$ , one must ensure that the functions  $f_j$ ,  $j = 1, \dots, K$  and  $\tilde{f}$  are "admissible". This is the aim of the next definition:

**Definition.** Let  $\theta_0 \in ]0, \frac{\pi}{2}[$ ,  $\mathcal{F}_{\theta_0}$  is the set of applications defined by:

$$\begin{aligned} \mathcal{F}_{\theta_0} := \{ & f : [0, 1] \rightarrow \mathbb{R}^+, f \text{ is differentiable on } [0, 1], \\ & f(0) > 0, f(1) = 0, \\ & \forall x \in [0, 1], f'(x) < 0, f'(1) = 0, f'(x) > -\cot \theta_0 \}. \end{aligned}$$

The class of curves  $\mathcal{C}_{\theta_0}$  is defined such that:

$$\mathcal{C}_{\theta_0} := \{(x, f(x)), x \in [0, 1], f \in \mathcal{F}_{\theta_0}\}.$$

Let us now define the parametric family of transformation that we consider.

**Notation.** For any function  $T_\alpha : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  depending on a parameter  $\alpha \in \Theta$ , we will denote by  $T_\alpha^1$  and  $T_\alpha^2$  its two coordinates.

In the following, we define our transformation model.

Set  $0 < \theta_0 < \theta_1 < \frac{\pi}{2}$ , and  $0 < \lambda_{min} < \lambda_{max}$ . For any  $\alpha := (\theta, \lambda) \in \Theta = [-\theta_0, \theta_1] \times [\lambda_{min}, \lambda_{max}]$ , we introduce the transformations

$$T_\alpha := H_\lambda \circ \tilde{R}_\theta.$$

Our parametric set of transformations is the compositions of a rotation  $\tilde{R}_\theta$  and a scaling on the second coordinate  $H_\lambda$ . More precisely

- $\tilde{R}_\theta$  is the rescaled rotation centered in  $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$  and of angle  $\theta$ . Note that if one performs just a rotation on a curve  $C$  of  $\mathcal{F}_{\theta_0}$ , the input space of the transformed curve is no longer  $[0, 1]$  but  $[a_{\theta,C}, 1]$  so we must rescale the input space to  $[0, 1]$ . That is

$$\begin{aligned} \tilde{R}_\theta : \mathcal{C}_{\theta_0} &\rightarrow \mathcal{C}_{\theta_0} \\ C &\rightarrow \tilde{R}_\theta(C), \end{aligned}$$

which transforms each point  $(x, f(x))$  of  $C$  to  $\begin{pmatrix} \frac{(x-1)\cos\theta - f(x)\sin\theta + 1 - a_{\theta,C}}{1 - a_{\theta,C}} \\ (x-1)\sin\theta + f(x)\cos\theta \end{pmatrix}$ , where  $a_{\theta,C} = -\cos\theta - f(0)\sin\theta + 1$ .

Notice that the elements of  $\mathcal{F}_{\theta_0}$  are functions that are decreasing but not too strongly. Indeed, if there is a (smooth)-vertical drop on the graph of a function in  $\mathcal{F}_{\theta_0}$  such that  $\forall x \in [0, 1], |f'(x)| < \cot\theta_0$ , when operating a negative rotation on the curve, it prevents the transformed curve to be still a function. As a matter of fact, if the vertical drop does not respect the constraint, the transformed curve's graph does not look like a graph function but more to a "zig-zag" going locally forward and backward.

- The scaling transformation of parameter  $\lambda > 0$  acting on the second coordinate of the curve:

$$\begin{aligned} H_\lambda : \mathcal{C}_{\theta_0} &\rightarrow \mathcal{C}_{\theta_0} \\ C &\rightarrow H_\lambda(C). \end{aligned}$$

which transforms each point  $(x, f(x))$  of  $C$  to  $\begin{pmatrix} x \\ \lambda f(x) \end{pmatrix}$ .

We now introduced the assumption that will be used:

**(A1):** The functions  $(f_j)_{j=1,\dots,K}$ , and  $\tilde{f}$  belong to  $\mathcal{F}_{\theta_0}$ .

**(A2):** For any  $\theta \in [-\theta_0, \theta_1], \forall C \in \mathcal{C}_{\theta_0}$ , all the second coordinate of  $\tilde{R}_\theta(C)$  are non-negative.

**(A3):**  $\Theta = [-\theta_0, \theta_1] \times [\lambda_{min}, \lambda_{max}]$ .

(A4):  $\tilde{C}$  is known on  $[0, 1]$ .

(A5):  $\tilde{C}$  is known on the grid  $\mathcal{D}_N$ .

Thus, the transformation model considered is as follows:

$$\begin{aligned} T_\alpha &: \mathcal{C}_{\theta_0} \rightarrow \mathcal{C}_{\theta_0} \\ C &\rightarrow T_\alpha(C) \end{aligned}$$

which transforms each point  $(x, f(x))$  of  $C$  to  $\left( \begin{array}{c} \frac{(x-1)\cos\theta - f(x)\sin\theta}{\cos\theta + f(0)\sin\theta} + 1 \\ \lambda((x-1)\sin\theta + f(x)\cos\theta) \end{array} \right)$ :  
the transformation model acts jointly on the argument and on the value of the template function.

### 2.3 Regression model

Reminding that we wish to adjust the reference curve  $\tilde{C}$  on the other curves  $C_j$  ( $j = 1, \dots, K$ ) by transformations belonging to  $\{T_\alpha, \alpha = (\theta, \lambda) \in \Theta\}$ . Notice that these transformations act on both axis. For any  $\alpha$ , we want to compare the value of the transformed curve  $(T_\alpha \tilde{C})(X_i)$  with  $f_j(X_i)$ . Since the abscissa points are affected by the transformation, we denote by  $X_i(\alpha)$  the point such that  $T_\alpha^1(X_i(\alpha)) = X_i$ . For that reason, we introduce the following definition:

**Definition.** We denote by  $x(\alpha, \tilde{f})$  the solution of the equation:

$$T_\alpha^1(u, \tilde{f}(u)) = x. \quad (1)$$

To ease the notation, we finally set  $x(\alpha) := x(\alpha, \tilde{f})$ , so  $X_i(\alpha) = X_i(\alpha, \tilde{f})$ . We consider the parametric regression model:

$$f_j(X_i) = T_{\alpha_j^*}^2(X_i(\alpha_j^*), \tilde{f}(X_i(\alpha_j^*))) + \epsilon_{j,i}, (j = 1, \dots, K). \quad (2)$$

Where:

- $(X_i)$  are iid,  $[0, 1]$ -uniformly distributed. It is the design on which we observe the curves  $C_j$ ;
- $\alpha_j^* = (\theta_j^*, \lambda_j^*)$  is the couple of true parameters for each curve ( $j = 1, \dots, K$ );
- $\epsilon_{j,i}$ ,  $\forall i = 1, \dots, N$ ,  $\forall j = 1, \dots, K$  are iid  $\mathcal{N}(0, \sigma^2)$  random variables. These variables are assumed to be independent of  $X_i$ .

## 2.4 Estimation

We consider the case where the noise applies on the set of curves to be registered only, and not on the reference curve. We legitimately suppose the existence of a reference curve. Under this last assumption, we will estimate the deformation parameters through a  $M$ -estimation procedure. The first subsection refers to the estimation of the deformation parameters and its analytical results when the reference curve is known on  $[0, 1]$ . Although this case is not realistic, these first results are necessary prerequisites to show the consistency and the asymptotic normality of the estimator in the more realistic case where the reference curve is defined on the same grid  $\mathcal{D}_N$  as the set of curves to be registered - this is the topic of the last subsection.

### 2.4.1 Estimation when $\tilde{C}$ is known on $[0, 1]$

For the sake of simplicity, let us fix  $j$ . Relying on a classical  $M$ -estimation procedure, we consider a semiparametric method to estimate the parameters and define consequently the following empirical contrast function to fit the reference curve  $\tilde{C}$  to  $C_j$  ( $j = 1, \dots, K$ ):

$$\begin{aligned} M_N^j(\alpha) &= \frac{1}{N} \sum_{i=1}^N (f_j(X_i) - T_\alpha^2(X_i(\alpha), \tilde{f}(X_i(\alpha))))^2 \\ &= \frac{1}{N} \sum_{i=1}^N m_\alpha^j(X_i). \end{aligned} \tag{3}$$

The random function  $M_N^j$  is non negative. Furthermore, intuitively, its minimum value should be reached close to the true parameter  $\alpha_j^*$ . Indeed, the following theorem gives the consistency of the  $M$ -estimator, defined by :

$$\hat{\alpha}_N^j = \underset{\alpha \in \Theta}{\operatorname{argmin}} M_N^j(\alpha). \tag{4}$$

Recall that our empirical contrast function enters in the general theory of  $M$ -estimator. The Central Limit Theorem will be shown by using  $M$ -estimator arguments.

**Theorem 1.** *Assume that A1, A2, A3 and A4 are satisfied. Then*

$$i) \hat{\alpha}_N^j \xrightarrow[N \rightarrow +\infty]{\mathbb{P}} \alpha_j^*, \tag{5}$$

$$ii) \sqrt{N}(\hat{\alpha}_N^j - \alpha_j^*) \xrightarrow[N \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \Gamma_{\alpha_j^*}). \tag{6}$$

*In particular, the covariance matrix has the following form*

$$\Gamma_{\alpha_j^*} = V_{\alpha_j^*}^{-1} 2\sigma^2, \tag{7}$$



with  $V_{\alpha_j^*} = 2\mathbb{E}[\dot{T}_{\alpha_j^*}^2 \dot{T}_{\alpha_j^*}^{2T}]$ , and  $\dot{T}_{\alpha_j^*}^2$  is the vector of partial derivatives of  $T_{\alpha_j}^2$  w.r.t elements of  $\alpha$  taken at  $\alpha_j^*$ .

#### 2.4.2 Estimation when $\tilde{C}$ is observed on $\mathcal{D}$

In this section, we consider the case where the reference curve  $\tilde{C}$  is observed on the same grid  $\mathcal{D} := (X_i)_{i=1,\dots,N}$  as the other curve  $C_j$ , i.e  $\tilde{C} = \left( \begin{matrix} x \\ \tilde{f}(x) \end{matrix} \right)_{x \in \mathcal{D}}$ . By applying the transformation  $T_\alpha$  to  $\tilde{C}$ , the transformed pattern  $T_\alpha \tilde{C}$  is no longer observable on  $\mathcal{D}$ . As a consequence, one must make use of an approximation process over  $\tilde{f}$ . Let  $\tilde{f}_N$  be the linear interpolate of  $\tilde{f}$ , defined by:

$$\tilde{f}_N(x) = \sum_{i=1}^N \Delta_i(x) \mathbf{1}_{x \in [X_{(i)}, X_{(i+1)})}, \quad (8)$$

where

$$\Delta_i(x) = \frac{\tilde{f}(X_{(i+1)}) - \tilde{f}(X_{(i)})}{X_{(i+1)} - X_{(i)}} x + \tilde{f}(X_{(i)}) - \frac{\tilde{f}(X_{(i+1)}) - \tilde{f}(X_{(i)})}{X_{(i+1)} - X_{(i)}} X_{(i)}. \quad (9)$$

It is easy to see that  $\tilde{f}_N$  belongs also to  $\mathcal{F}_{\theta_0}$ . Replacing  $\tilde{C}$  by  $\hat{C}$  in (3) we obtain

$$\hat{M}_N^j(\alpha) = \frac{1}{N} \sum_{i=1}^N (f_j(X_i) - T_\alpha^2(X_i(\alpha, N), \tilde{f}_N(X_i(\alpha, N))))^2,$$

where  $X_i(\alpha, N)$  is the solution to the equation  $T_\alpha^1(u, \tilde{f}_N(u)) = X_i$ .

Using the linear interpolate defined by (8), we show the consistency and asymptotic normality of our  $M$ -estimator defined as follow:

$$\hat{\alpha}_N^j = \underset{\alpha \in \Theta}{\operatorname{argmin}} \hat{M}_N^j(\alpha). \quad (10)$$

**Theorem 2.** *Assume that A1, A2, A3 and A5 are satisfied. Let  $\tilde{f}_N$  be defined by (8) and assume that  $\exists C > 0$  s.t  $\forall x \in [0, 1]$ ,  $\tilde{f}'_N(x) \leq C$ , and  $\tilde{f}'(x) \leq C$ , then*

$$i) \hat{\alpha}_N^j \xrightarrow[N \rightarrow +\infty]{\mathbb{P}} \alpha_j^*, \quad (11)$$

$$ii) \sqrt{N}(\hat{\alpha}_N^j - \alpha_j^*) \xrightarrow[N \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \Gamma_{\alpha_j^*}) \quad (12)$$

with  $\Gamma_{\alpha_j^*}$  such defined in (7).

*Remark.* Notice that the asymptotic variance of the estimator is the same as in Theorem 1.

### 3 Simulations and applications

In this section we illustrate the method on numerical applications. The first subsection is dedicated to some simulated example while the second to a real problem. The optimisation problems (4) and (10) will be numerically solved by using the BFGS algorithm [20]. In each case, in order to apply the results presented in the previous section, one shall define a reference curve. In the simulated example, the reference curve is known. In the real world application, we simply choose the average curve as the reference curve (note that choosing the curve with the lowest values would assure that (A2) is automatically satisfied).

#### 3.1 Simulated example

We consider the following model:

$$f_{\lambda,\theta}(x) = \lambda[(x - 1) \sin \theta + g(x) \cos \theta],$$

with  $g(x) = 2(\cos(\pi x) + 1)$ . We observe  $f_j(X_i) = f_{\lambda_j,\theta_j}(X_i) + \epsilon_{ij}$  for  $i = 1, \dots, N$ , for  $J = 25$  values of  $(\lambda, \theta)$  with some iid errors  $\epsilon_{ij}$ .

- The observations points  $X_i$ ,  $i = 1, \dots, N$  are iid random variables with uniform distribution on  $[0, 1]$ .
- The parameters are chosen randomly with the following arbitrary distribution  $(\lambda, \theta) \hookrightarrow \mathcal{U}([0, 15]) \times \mathcal{U}([-\frac{\pi}{3}, \frac{\pi}{30}])$ .
- The errors are assumed to be  $\mathcal{N}(0, \sigma^2)$ .

The numerical experiment will be conducted for different values of  $N = 50, 100, 1000$  and  $\sigma^2 = 0, 0.01, 0.05, 0.1, 0.5, 1$ . To avoid numerical issues, each curve is rescaled to  $[0, 1]$ . Some simulated data are shown in Figure 1. In this example, the reference curve is known. This is the function  $g$  displayed in blue. As shown in Figure 1 (b), one shall emphasize that by increasing the value of the noise variance we loss the properties that the functions  $f_j$ ,  $j = 1, \dots, J$  belong to  $\mathcal{F}_{\theta_0}$ , which is not in our best interests.

In the following, we compare the method developed in the previous section to the famous non-parametric method of curve registration defined by J.O. Ramsay and X. Li in [21]. This method is also called Dynamic Time Warping (DTW): a non-parametric technique is used to estimate the smooth monotone transformations  $h_i$  applied on the features (i.e the x-abcissa) by preserving their order. Hence, the features  $X_i$  are warped through  $h_i$  and the characteristic (peaks, valleys) of the reference curve and the target are aligned. In order to compare the two methodologies, we will use the following mean squared error (MSE) between the registered curve and  $f_{\lambda_j,\theta_j}$

$$MSE = \frac{1}{25} \sum_{j=1}^{25} \sum_{i=1}^N (f_{\lambda_j,\theta_j}(X_i) - \hat{f}_{\hat{\lambda}_j,\hat{\theta}_j}(X_i))^2$$

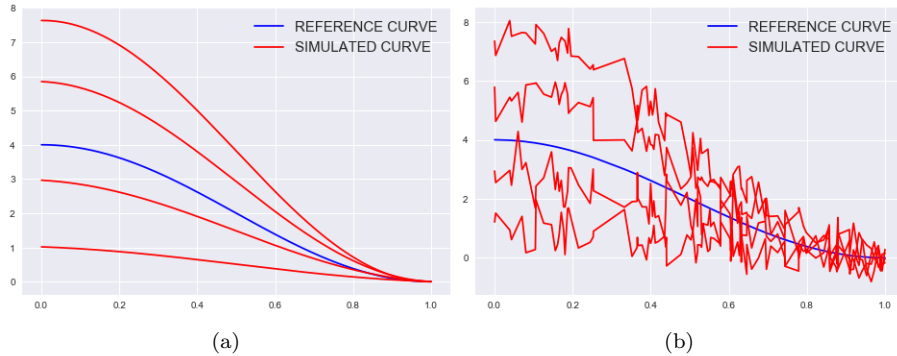


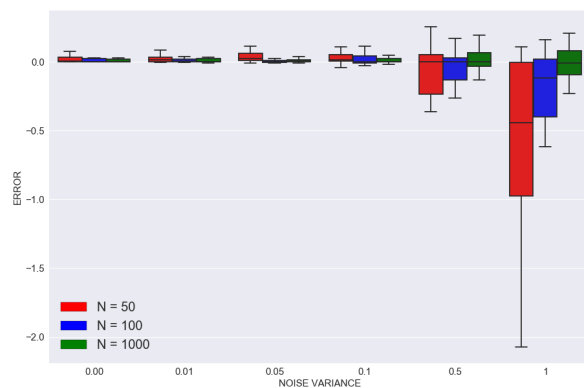
Figure 1: Simulated data with : (a)  $N = 100$  and  $\sigma^2 = 0$ , (b)  $N = 100$  and  $\sigma^2 = 0.5$

Results are displayed in Table 1. It comes at no surprise that increasing the noise's variance increases the MSE. Nevertheless, due to its transformation nature, Ramsay's method might take the augmentation of noise variance to its advantage contrary to our morphing strategy. However, when the noise is small, our methodology outperforms the other. This is due to the fact that the peaks and valleys of the curves are most of the time already aligned in this kind of function class inducing few deformations on the abscissa.

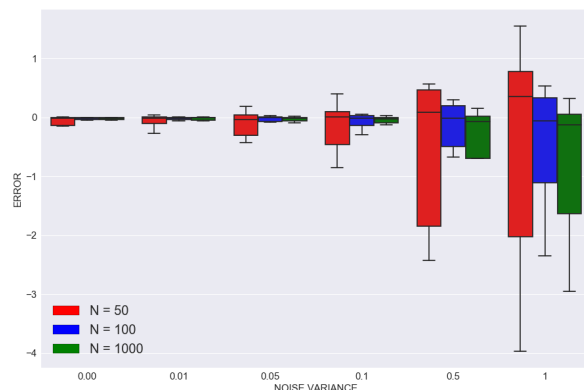
Table 1: Mean Squared Error (MSE) comparison between our morphing strategy and the DTW for different  $N$  and  $\sigma^2$ .

	Morphing			DTW		
	N = 50	N = 100	N = 1000	N = 50	N = 100	N = 1000
$\sigma^2 = 0.00$	<b>7.e-05</b>	<b>6.e-05</b>	<b>4.e-05</b>	0.0011	0.0014	0.0012
$\sigma^2 = 0.01$	<b>0.0006</b>	<b>0.0002</b>	<b>0.0001</b>	0.0061	0.0041	0.007
$\sigma^2 = 0.05$	<b>0.0032</b>	<b>0.0025</b>	<b>0.0024</b>	0.0076	0.0051	0.0092
$\sigma^2 = 0.10$	<b>0.0104</b>	0.0098	<b>0.0091</b>	0.012	<b>0.0086</b>	0.015
$\sigma^2 = 0.50$	0.2489	0.2449	0.2473	<b>0.1451</b>	<b>0.1413</b>	<b>0.1732</b>
$\sigma^2 = 1.00$	0.9942	0.9782	0.9892	<b>0.65</b>	<b>0.5974</b>	<b>0.79</b>

In Figure 2 are displayed the errors when estimating the deformation parameters for the different values of the noise variance  $\sigma^2$  and the number of observations  $N$ . The rotation parameter  $\theta$  is the one which is the most sensitive to the number of observations  $N$ , and in average our morphing strategy seems to be slightly optimistic in the estimation of  $\lambda$  for a small noise variance.



(a)



(b)

Figure 2: Boxplot of the errors for the estimation of (a)  $\lambda$  and (b)  $\theta$  depending on the noise variance  $\sigma^2$  and the number of observations  $N$  using the morphing method.

Examples of the registration process are displayed in Figure 3. Our methodology works better when the data we are dealing with are closed to the proposed model, else it tends to make more error undeniably. Nevertheless, compared to a nonparametric model, we can use the estimations of the deformation parameters to do meta-modeling as we will show in the next subsection.

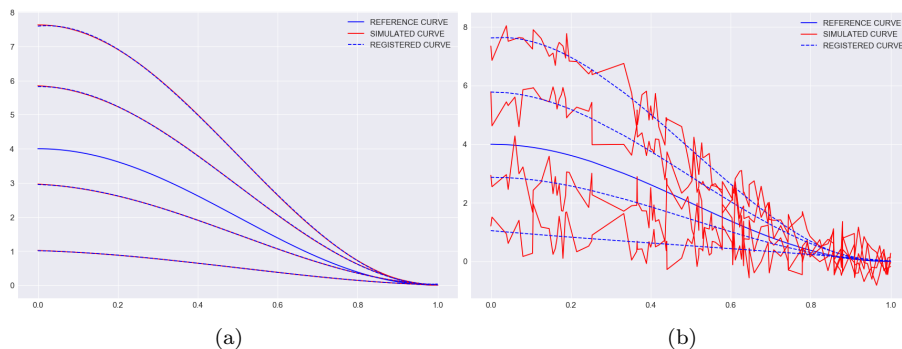


Figure 3: Results of the registration process with : (a)  $N = 100$  and  $\sigma^2 = 0$ , (b)  $N = 100$  and  $\sigma^2 = 0.5$

### 3.2 Aeronautic loads

An airframe structure is a complex system and its design is a complex task that today involves many simulation activities generating massive amounts of data. This is, for example, the process of loads and stress computations of an aircraft. That is the computations of the forces and the mechanical strains suffered by the structure. The overall process exposed in Figure 4 is run to identify load cases (i.e aircraft mission and configurations: maneuvers, speed, loading, stiffness...), that are critical in terms of stress endured by the structure and, of course, the parameters which make them critical. The final aim is to size and design the structure (and potentially to reduce loads in order to reduce the weight of the structure). Typically for an overall aircraft structure, millions of load cases can be generated and for each of these load cases millions of structural responses (i.e how structural elements react under such conditions) have to be computed. As a consequence, computational times can be significant.

In an effort to continuously improve methods, tools and ways-of-working, Airbus has invested a lot in digital transformation and the development of infrastructures allowing to treat data (newly or already produced). The main industrial challenge for Airbus is to reduce lead time in the computation and preliminary sizing of an airframe as well as extracting value from already calculated loads. In this paper, we focus on the external loads of a wing: the shear forces (transverse forces near to vertical arising from aerodynamic pressure and inertia) and the associated bending moments (resulting from the shear forces, they represent the flexion of the wing) are calculated for each load case. Examples of bending moments are displayed in Figure 5.

These external loads appeared to be extremely regular and one can legitimately suppose that there exists a link between all those curves. Indeed, it is natural to assume that there exists a reference bending moment (a reference

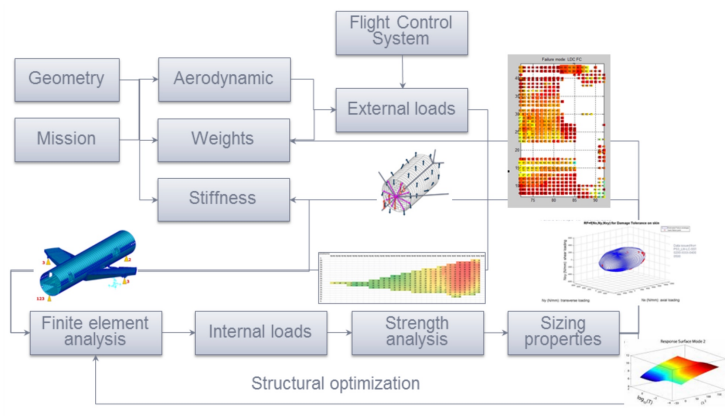


Figure 4: Flowchart for loads and stress analysis process

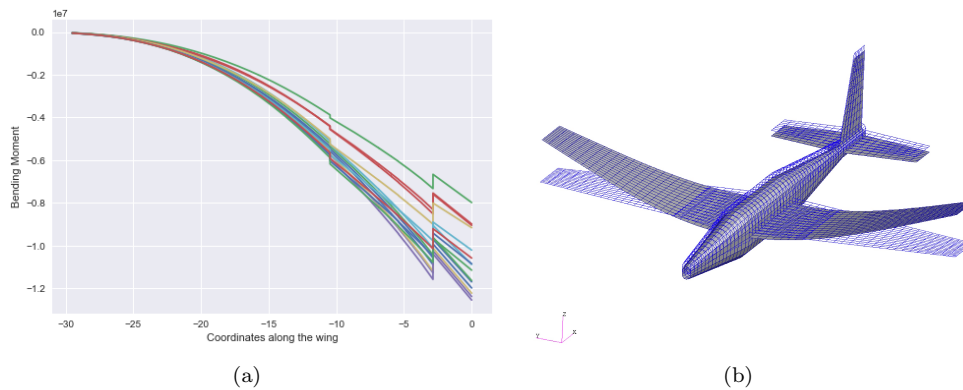


Figure 5: (a) Examples of bending moments along the wing for different load cases - (b) Finite element model of a generic aircraft representing the wing deformation [23]

curve) which can be morphed through a deformation model to give all the other curves.

In [6], the authors present an aeronautic model that computes the loads (forces and moments) on the wing of some aircraft model denoted by *ACM1*. They present several statistical methods in order to study these data. In this section, we will compare the method used in [6] with the model presented in Section 2 for a new aircraft model called *ACM2*. The data at our disposal represents bending moments of a wing (representing its flexion) of an aircraft calculated for 1152 different configurations (load cases). Each configuration is defined by 28 features (speed of the aircraft, mass, altitude, quantity of fuel, etc.), leading to a bending moment calculated on 45 stations along the wing. In a more formal

way, we observe the couple  $(X_j, Y_j)_{j=1, \dots, 1152}$ , where  $X_j = (X_j^1, \dots, X_j^{28})$  are the features and  $Y_j = (Y_j^1, \dots, Y_j^{45})$  is the bending moment. The idea is to predict the bending moment for different configurations. The data are represented in Figure 6.

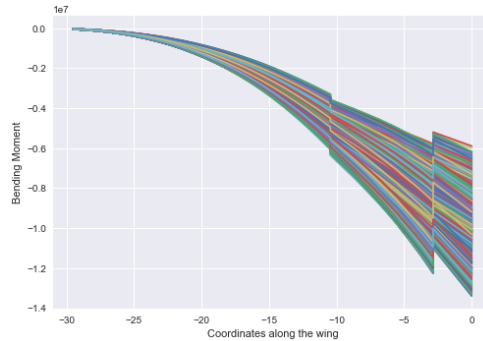


Figure 6: Representation of all the bending moments of a wing of our data base: the wing root is located at the zero origin, where the strains are maximum when the wing bends.

Due to the discontinuities at the 3<sup>rd</sup> and 20<sup>th</sup> stations, we apply our methodology to each section independently: we suppose, reasonably, that the average curve (of each section), can be used as the reference curve. Then, each section can be represented by its minimum and maximum values, and by its rotation and scaling coefficients  $\lambda_j$  and  $\theta_j$ . Figure 7 assess the quality of the matching process (the reference curve used is the average bending moment). As a comparison, we display in Table 2 the MSE between the DTW and our deformation method (note that the MSE values are high due to the fact that the curves have high values).

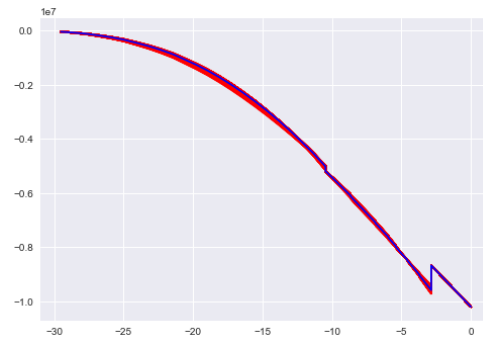


Figure 7: Results of the matching process

Table 2: Mean Squared Error comparison between our morphing strategy and the DTW for the real world application.

	Morphing	DTW
MSE	<b>2.2e+09</b>	3.1e+09

Thus the dimensional space of the outputs is reduce to 12 instead of 45. We compare our method to three other methods of [6] applied on the outputs: no transformation (we call it raw - we build 45 models, one per station); a PCA (the three first principal components represent 99,9% of the explained variance - 3 models instead of 45); a polynomial fitting per section (of degree 4 for the first section, of degree 2 for the second and of degree 1 for the third section) which leads to 10 models instead of 45. The Table 3 sums up the number of outputs to predict depending on the method used.

Table 3: Number of outputs to be predicted depending on the method used on the raw outputs: Raw, Deformation Model, PCA, polynomial fitting

	Number of outputs	Names of outputs
Raw	45	Bending moment value at station 0 to 44
Deformation Model	12	$\theta_1, \theta_2, \theta_3, \lambda_1, \lambda_2, \lambda_3, min_1, min_2, min_3, max_1, max_2, max_3$
PCA	3	Principal components 1 to 3
Polynomial fitting	10	Coefficients of polynomials

The significant advantage of the reduction dimension techniques used is that the response of the model would have a physical form contrary to the simple linear models performed on the raw data. To build our models, we use the Orthogonal Greedy Algorithm (OGA) also known as the Matching Pursuit Algorithm. Detailed explanations can be found in [2], [24] and [18]. Roughly speaking, we consider the problem of approximating a function by a sparse linear combination of inputs.

To assess the goodness of fit of our models, we defined for a curve of bending moment  $j$  the error rate as follows:

$$error(j) = \sqrt{\frac{\sum_{i=1}^{45} (\hat{y}(x_i) - y_j(x_i))^2}{\sum_{i=1}^{45} y_j^2(x_i)}}, j = 1, \dots, n_{test},$$

where  $n_{test}$  is the size of the sample of test. We compute the error rates on (the sample of test is of 25% the size of the total database). It gives an idea



of how accurate our predictions are. For this standpoint, we can easily compute the empirical cumulative distribution function (CDF):  $\forall j = 1, \dots, n_{test}$ , let  $\alpha \in [0, 1]$ . The empirical CDF is defined as:

$$\alpha \rightarrow G(\alpha) = \frac{1}{n} \sum_{j=1}^{n_{test}} \mathbb{1}_{(error(j) \leq \alpha)}$$

In Table 4, we give the values of  $G(\alpha)$  for  $\alpha = 1\%, 2\%, 5\%, 10\%$  and the mean error. In Figure 8 we give the plots of the function  $G(\alpha)$  for the different methods.

Table 4: Average estimated  $\mathbb{P}(error \leq 1\%)$ ,  $\mathbb{P}(error \leq 2\%)$ ,  $\mathbb{P}(error \leq 5\%)$ ,  $\mathbb{P}(error \leq 10\%)$ ,  $\mathbb{E}(error)$  calculated on several random test data set (25% of the size of the total dataset)

	Deformation Model	Polynomial Fitting	PCA	Raw
$\mathbb{P}(error \leq 1\%)$	17%	14%	16%	15%
$\mathbb{P}(error \leq 2\%)$	45%	45%	43%	51%
$\mathbb{P}(error \leq 5\%)$	88%	88%	86%	88%
$\mathbb{P}(error \leq 10\%)$	98%	97%	95%	98%
$\mathbb{E}(error)$	2.9%	2.9%	3%	2.8%

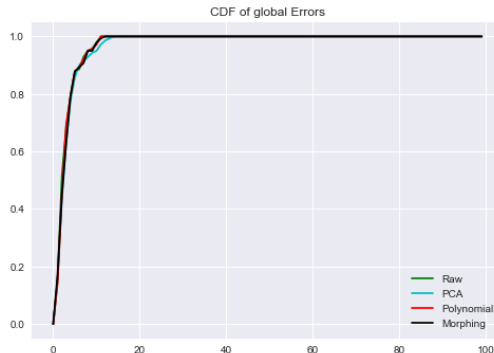


Figure 8: Empirical CDF of error rates ( $\mathbb{P}(error \leq \alpha)$ )

Concerning the approximation and prediction of loads, our model is equivalent in average to other tested methods, there are just slightly more observations with an error below 1%. Nevertheless, in our case, the linear models built through the deformation model are sparser than the other. Indeed, in average, 11 variables are chosen as optimal parameter of the greedy algorithm by cross-validation for the deformation model, 13 for the polynomial fitting, 15 for the

PCA and 14 for the raw outputs one.

Even though the prediction of loads with the deformation model is so likely equivalent to none transformation, it obtains better results than the polynomial fitting and the PCA. Besides, using this deformation model gives a physical response contrary to a simple linear model per station whose response could be irregular.

We have shown that, when dealing with a large set of curves, if we provide and use a good representation, we can have: as good results as state of the art classical methods of dimension reduction techniques or none; and a better understanding of the variability within the population at the same time. It is the last point which is important particularly: indeed, it exists many ways to have good prediction results (non-parametric regression, black-box models), but using reduction dimension techniques without an underlying engineering judgement is not of significant interest: from our methodology, besides having a good prediction accuracy, we can extract knowledge for aerospace engineers to better understand the variation of behaviors of the wing depending on the reference behavior (in other words, it gives to engineers a physical interpretation and idea of how the wing will react to new constraints). Hence, in this sense, our proposed methodology outperforms the others.

## 4 Proofs and technical result

### 4.1 Technical result

This section is dedicated to the technical result used in the proof of Theorem 2.

**Lemma 3.** *Let  $X_1, \dots, X_N$  be  $N$  independent and identically distributed random variables with uniform distribution on  $[0, 1]$  and let  $X_{(1)} \leq \dots \leq X_{(N)}$  be the reordered sample. Let  $a_N = O(\sqrt{N})$ , then:*

$$\mathbb{P}(a_N \sup_j |X_{(j+1)} - X_{(j)}| \geq \epsilon) \xrightarrow{N \rightarrow +\infty} 0.$$

*Proof of Lemma 3.* Let  $Z_1, \dots, Z_{N+1}$  be  $N$  independent and identically distributed random variables with exponential distribution with parameter 1. It is a well known fact that

$$\left( \frac{Z_1}{\sum_{k=1}^{N+1} Z_k}, \frac{Z_1+Z_2}{\sum_{k=1}^{N+1} Z_k}, \dots, \frac{Z_1+\dots+Z_N}{\sum_{k=1}^{N+1} Z_k} \right) \stackrel{(\mathcal{L})}{=} (X_{(1)}, \dots, X_{(N)}) \text{ and we have}$$

$$X_{(j+1)} - X_{(j)} \stackrel{(\mathcal{L})}{=} \frac{Z_{j+1}}{\sum_{k=1}^{N+1} Z_k}. \quad (13)$$

Now, for  $\epsilon > 0$

$$\begin{aligned}\mathbb{P}(\sup_j |X_{(j+1)} - X_{(j)}| \geq \epsilon) &\leq \sum_j \mathbb{P}(X_{(j+1)} - X_{(j)} \geq \epsilon) \\ &\leq N \max_j \mathbb{P}(X_{(j+1)} - X_{(j)} \geq \epsilon).\end{aligned}$$

By using (13), we have

$$\mathbb{P}(X_{(j+1)} - X_{(j)} \geq \epsilon) = (1 - \epsilon)^{N-1}.$$

Then,

$$\mathbb{P}(\sup_j |X_{(j+1)} - X_{(j)}| \geq \epsilon) \leq N(1 - \epsilon)^{N-1}.$$

The result follows replacing  $\epsilon$  by  $\frac{\epsilon}{a_N}$  and letting  $N \rightarrow +\infty$ .  $\square$

## 4.2 Proofs of Theorems 1 and 2

*Proof of Theorem 1.* To ease the notation, we do not display the dependency in  $j$ .

i) By (3) it is easy to see that  $M_N(\alpha)$  is an empirical mean of iid bounded random variables. Thus, by the Strong Law of Large Number (SLLN)

$$M_N(\alpha) \xrightarrow[N \rightarrow +\infty]{p.s.} M(\alpha),$$

with  $M(\alpha) = \mathbb{E}[\epsilon^2] + \mathbb{E}[(T_\alpha^2(X(\alpha), \tilde{f}(X(\alpha))) - T_{\alpha^*}^2(X(\alpha^*), \tilde{f}(X(\alpha^*))))^2]$ .

$M(\alpha)$  is continuous and has an obvious unique minimum  $\alpha^*$ . Since  $\Theta$  is compact, this implies that  $\inf_{\alpha: d(\alpha, \alpha^*) \geq \epsilon} M(\alpha) > M(\alpha^*)$  is satisfied (see Problem 27 p. 84 in [25]).

It remains to prove that  $\{m_\alpha : \alpha \in \Theta\}$  is a Glivenko-Cantelli class. Thanks to the remark following the proof of Theorem 5.9 in [25], this is an easy consequence of the continuity of  $\alpha \rightarrow m_\alpha$  and the fact that the function is bounded by a continuous and integrable function on  $[0, 1]$ . Indeed, it exists at least a function  $f^*$  in  $\mathcal{F}_{\theta_0}$  which bounds every other functions, and two constants  $K_1 > 0, K_2 > 0$  such that

$$m_\alpha(x) \leq K_1(f^*(x) + K_2)^2,$$

and

$$\sup_{\alpha \in \Theta} |M_N(\alpha) - M(\alpha)| \xrightarrow[N \rightarrow +\infty]{\mathbb{P}} 0. \quad (14)$$

The result follows from the Theorem 5.7 in [25].

ii) The Central Limit Theorem will be a consequence of Theorem 5.23 in [25]. Recall that

$$m_\alpha(x) = [f(x) - \lambda((x(\alpha) - 1) \sin \theta + \cos \theta \tilde{f}(x(\alpha)))]^2.$$

By the Implicit Function Theorem, that is easy to see that  $\alpha \rightarrow x(\alpha)$  is  $C^1$  on a compact set. This implies that the norm of the gradient of  $m_\alpha$  is uniformly bounded in  $\alpha$ . Hence  $\exists \dot{\phi}(x) \in L^1$  such that  $\|\nabla_\alpha m_\alpha(x)\| \leq \dot{\phi}(x)$  hence

$$|m_{\alpha_1}(x) - m_{\alpha_2}(x)| \leq \dot{\phi}(x) \times \|\alpha_1 - \alpha_2\|.$$

In order to give an explicit formula for the limit variance, we apply the results of Example 5.27 in [25] where  $f_\theta$  becomes in our case  $T_\alpha^2$  and hence, we have

$$\sqrt{N}(\hat{\alpha}_N^j - \alpha_j^*) \xrightarrow[N \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \Gamma_{\alpha_j^*}),$$

$$\text{with } \Gamma_{\alpha_j^*} = V_{\alpha_j^*}^{-1} 2\sigma^2 \text{ and } V_{\alpha_j^*} = 2\mathbb{E}[\dot{T}_{\alpha_j^*}^2 \dot{T}_{\alpha_j^*}^{2\mathbf{T}}].$$

□

*Proof of Theorem 2.*

i) To prove the consistency of  $\hat{\alpha}_N$  we have to show that

$$\sup_{\alpha \in \Theta} |\hat{M}_N(\alpha) - M(\alpha)| \xrightarrow[N \rightarrow +\infty]{\mathbb{P}} 0.$$

We have,

$$\sup_{\alpha \in \Theta} |\hat{M}_N(\alpha) - M(\alpha)| \leq \sup_{\alpha \in \Theta} |\hat{M}_N(\alpha) - M_N(\alpha)| + \sup_{\alpha \in \Theta} |M_N(\alpha) - M(\alpha)|.$$

It has been shown in the proof of Theorem 1 that

$$\sup_{\alpha \in \Theta} |M_N(\alpha) - M(\alpha)| \xrightarrow[N \rightarrow +\infty]{\mathbb{P}} 0.$$

It remains to prove that

$$\sup_{\alpha \in \Theta} |\hat{M}_N(\alpha) - M_N(\alpha)| \xrightarrow[N \rightarrow +\infty]{\mathbb{P}} 0.$$

To ease the notation, we write  $T_\alpha^2(i, N) = T_\alpha^2(X_i(\alpha, N), \tilde{f}_N(X_i(\alpha, N)))$ , and  $T_\alpha^2(i) = T_\alpha^2(X_i(\alpha), \tilde{f}(X_i(\alpha)))$ . Set

$$\begin{aligned} D_N(\alpha) &= M_N(\alpha) - \hat{M}_N(\alpha) \\ &= \frac{1}{N} \sum_{i=1}^N 2f(X_i) [T_\alpha^2(i, N) - T_\alpha^2(i)] - \frac{1}{N} \sum_{i=1}^N [T_\alpha^2(i, N) - T_\alpha^2(i)] [T_\alpha^2(i, N) + T_\alpha^2(i)]. \end{aligned}$$

As  $f$  and  $T_\alpha^2$  are continuous and bounded on  $\Theta \times [0, 1]$ , this implies that:

$$\begin{aligned}
|D_n(\alpha)| &\leq \left| \frac{1}{N} \sum_{i=1}^N 2f(X_i)[T_\alpha^2(i, N) - T_\alpha^2(i)] \right| + \left| \frac{1}{N} \sum_{i=1}^N [T_\alpha^2(i, N) - T_\alpha^2(i)][T_\alpha^2(i, N) + T_\alpha^2(i)] \right| \\
&\leq K \left( \frac{1}{N} \sum_{i=1}^N [T_\alpha^2(i, N) - T_\alpha^2(i)]^2 \right)^{\frac{1}{2}} \\
&\leq K' \left( \frac{1}{N} \sum_{i=1}^N [(X_i(\alpha, N) - X_i(\alpha))(1 + C) + (\tilde{f}_N(X_i(\alpha)) - \tilde{f}(X_i(\alpha)))^2] \right)^{\frac{1}{2}}.
\end{aligned}$$

By construction, there exists  $j$  such that  $X_{(j)} \leq X_i(\alpha) \leq X_{(j+1)}$ , and  $X_{(j)} \leq X_i(\alpha, N) \leq X_{(j+1)}$  which leads to:

$$X_i(\alpha, N) - X_i(\alpha) = \gamma(X_{(j+1)} - X_{(j)}).$$

Besides, there exists  $\gamma' > 0$  such that  $\tilde{f}_N(X_i(\alpha)) = \gamma' \tilde{f}(X_{(j+1)}) + (1 - \gamma') \tilde{f}(X_{(j)})$ .  $C$  and  $\gamma'$  being uniform constants, we have

$$\begin{aligned}
|\tilde{f}_N(X_i(\alpha)) - \tilde{f}(X_i(\alpha))| &\leq \gamma' |\tilde{f}(X_{(j+1)}) - \tilde{f}(X_{(j)})| \\
&\leq C\gamma' |X_{(j+1)} - X_{(j)}|,
\end{aligned}$$

and

$$\begin{aligned}
|D_n(\alpha)| &\leq K' \left( \frac{1}{N} \sum_{j=1}^N [X_{(j+1)} - X_{(j)}]^2 \right)^{\frac{1}{2}} \\
\sup_{\alpha \in \Theta} |D_n(\alpha)| &\leq K' \left( \frac{1}{N} \sum_{j=1}^N \sup_{\alpha \in \Theta} [X_{(j+1)} - X_{(j)}]^2 \right)^{\frac{1}{2}} \\
&\leq K' \sup_j |X_{(j+1)} - X_{(j)}|.
\end{aligned}$$

By Lemma 1  $\mathbb{P}(K' \sup_j |X_{(j+1)} - X_{(j)}| \geq \epsilon) \xrightarrow{N \rightarrow +\infty} 0$ . Hence  $D_N$  is bounded by an integrable and continuous function which goes to 0 in probability on  $\Theta$

$$\sup_{\alpha \in \Theta} |\hat{M}_N(\alpha) - M(\alpha)| \xrightarrow{N \rightarrow +\infty} 0.$$

So we may conclude.

ii) First, we use that  $\sqrt{N}(\hat{\alpha}_N - \alpha^*) = \sqrt{N}(\hat{\alpha}_N - \hat{\alpha}_N) + \sqrt{N}(\hat{\alpha}_N - \alpha^*)$ . By Theorem 3,  $\sqrt{N}(\hat{\alpha}_N - \alpha^*) \xrightarrow[N \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, \Gamma_{\alpha^*})$  with  $\Gamma_{\alpha^*}$  defined in (6). It remains to prove that  $\sqrt{N}(\hat{\alpha}_N - \hat{\alpha}) \xrightarrow[N \rightarrow +\infty]{\mathbb{P}} 0$ .

Using the same arguments as in the proof i), we have

$$\mathbb{P}(\sqrt{N} \sup_{\alpha \in \Theta} |\hat{M}_N(\alpha) - M_N(\alpha)| \geq \epsilon) \leq \mathbb{P}(K \sqrt{N} \sup_{\alpha \in \Theta} |X_{(j+1)} - X_{(j)}| \geq \epsilon) \quad (15)$$

The right hand side of (15) converges to 0 by Lemma 3. This implies that  $\sqrt{N}(\hat{\alpha}_N - \hat{\alpha}) \xrightarrow[N \rightarrow +\infty]{\mathbb{P}} 0$ .

□

## 5 Perspectives and conclusion

One of the main quality of our approach is that it is easy to implement and execute. The cost function being simple, we use a BFGS algorithm to find the optimal parameters, and because of the regularity of curves we deal with, the initial points for optimization can be easily defined. Furthermore, the search of the coordinate of the reference curve which is sent to the coordinate of the curve to fit can be easily implemented with a simple value search.

We have seen that our methodology performs very well when the data we are dealing with are close to the chosen model. Besides, the deformation parameters can be exploited through an explainable model such as the linear model used in the real world problem.

It seems that the deformation model is robust if the noise is controlled. An interesting extension of this work would be to study what is going on when the reference curve is noisy. A generalization of this work to less regular functions would be worthwhile. Finally, it would be interesting to include in the model a way to handle discontinuities in order to reduce the dimension and have a more global representation of the deformation.

## Acknowledgements

We are very much indebted to the referees and the Associate Editor for their constructive criticisms, comments and remarks that resulted in a major improvement of the original manuscript. We would also like to thank Fabrice Gamboa and Christian Bes for careful re-readings and thoughtful advice.

## References

- [1] Allasonniere, S., Bigot, J., Glaunes, J., Maire, F., F., R.: Statistical models for deformable templates in image and shape analysis. *Annales Mathematiques Blaise Pascal* **20**(1), 1–35 (2013)
- [2] Barron, A.R., Cohen, A., Dahmen, W., DeVore, R.A.: Approximation and learning by greedy algorithms. *Ann. Statist.* **36**(1), 64–94 (2008). DOI 10.1214/009053607000000631. URL <https://doi.org/10.1214/009053607000000631>
- [3] Bigot, J., Christophe, C., Gadat, S.: Random action of compact lie groups and minimax estimation of a mean pattern. *IEEE Transactions on Information Theory* **58**(6), 3509–3520 (2012)
- [4] Bigot, J., Gadat, S.: A deconvolution approach to estimation of a common shape in a shifted curves model. *Ann. Statist.* **38**(4), 2422–2464 (2010). DOI 10.1214/10-AOS800. URL <https://doi.org/10.1214/10-AOS800>
- [5] Bigot, J., Gamboa, F., Vimond, M.: Estimation of translation, rotation, and scaling between noisy images using the fouriermellin transform. *SIAM Journal on Imaging Sciences* **2**, 614–645. ISSN 2936-4954
- [6] Fournier, E., Grihon, S., Klein, T.: A case study : Influence of Dimension Reduction on regression trees-based Algorithms -Predicting Aeronautics Loads of a Derivative Aircraft (2018). URL <https://hal.archives-ouvertes.fr/hal-01700314>. Working paper or preprint
- [7] Gamboa, F., Loubes, J.M., Maza, E.: Semi-parametric estimation of shifts. *Electron. J. Statist.* **1**, 616–640 (2007). DOI 10.1214/07-EJS026. URL <https://doi.org/10.1214/07-EJS026>
- [8] Gasser, T., Kneip, A.: Searching for structure in curve samples. *Journal of the american statistical association* **90**(432), 1179–1188 (1995)
- [9] Gassiat, E., Lévy-Leduc, C.: Efficient semiparametric estimation of the periods in a superposition of periodic functions with unknown shape. *Journal of Time Series Analysis* **27**(6), 877–910 (2006)
- [10] Golubev, G.: Estimation of the period of a signal with an unknown form against a white noise background. *Problems Inform. Transmission* **24**(4), 288–299 (1988)
- [11] Grenander, U.: *General pattern theory*. Oxford Science Publications (1993)
- [12] Hardle, W., Marron, J.S.: Semiparametric comparison of regression curves. *The Annals of Statistics* pp. 63–89 (1990)
- [13] Ke, C., Wang, Y.: Semiparametric nonlinear mixed-effects models and their applications. *Journal of the American Statistical Association* **96**(456), 1272–1298 (2001)

- [14] Kneip, A., Engel, J.: Model estimation in nonlinear regression under shape invariance. *The Annals of Statistics* pp. 551–570 (1995)
- [15] Kneip, A., Li, X., MacGibbon, K., Ramsay, J.: Curve registration by local regression. *Canadian Journal of Statistics* **28**(1), 19–29 (2000)
- [16] Lawton, W., Sylvestre, E., Maggio, M.: Self modeling nonlinear regression. *Technometrics* **14**(3), 513–532 (1972)
- [17] Lindstrom, M.J.: Self-modelling with random shift and scale parameters and a free-knot spline shape function. *Statistics in medicine* **14**(18), 2009–2021 (1995)
- [18] Mallat, S.G., Zhang, Z.: Matching pursuits with time-frequency dictionaries. *IEEE Transactions on Signal Processing* **41**(12), 3397–3415 (1993). DOI 10.1109/78.258082
- [19] McGuire, M.: An image registration technique for recovering rotation, scale and translation parameters. Tech. rep., NEC Tech Report (1998). URL <http://casual-effects.com/research/McGuire1998ParameterRecovery/index.html>
- [20] Nocedal, J., Wright, S.J.: *Numerical Optimization*, second edn. Springer, New York, NY, USA (2006)
- [21] Ramsay, J.O., Li, X.: Curve registration. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **60**(2), 351–363. DOI 10.1111/1467-9868.00129
- [22] Ramsay, J.O., Wickham, H., Graves, S., Hooker, G.: *fda: Functional Data Analysis* (2018). URL <https://CRAN.R-project.org/package=fda>. R package version 2.4.8
- [23] Ritter, M., Dillinger, J.: Nonlinear numerical flight dynamics for the prediction of maneuver loads. *IFASD 2011* pp. 1–10 (2011)
- [24] Sancetta, A.: Greedy algorithms for prediction. *Bernoulli* **22**(2), 1227–1277 (2016). DOI 10.3150/14-BEJ691. URL <https://doi.org/10.3150/14-BEJ691>
- [25] Van der Vaart, A.W.: *Asymptotic statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press (1998)
- [26] Vimond, M.: Efficient estimation for homothetic shifted regression models. Université Paul Sabatier, Laboratoire de statistique et probabilités (2006)