Functional Principal Component Analysis of Aircraft Trajectories

Florence Nicol

RAPPORT DE RECHERCHE

Mai 2013

Functional Principal Component Analysis of Aircraft Trajectories

N° : RR/ENAC/2013/02

ENAC

La référence aéronautique

www.enac.fr

École Nationale de l'Aviation Civile

Florence NICOL

Abstract: In Functional Data Analysis (FDA), the underlying structure of a raw observation is functional and data are assumed to be sample paths from a single stochastic process. Functional Principal Component Analysis (FPCA) generalizes the standard multivariate Principal Component Analysis (PCA) to the infinite-dimensional case by analyzing the covariance structure of functional data. By approximating infinite-dimensional random functions by a finite number of random score vectors, FPCA appears as a dimension reduction technique just as in the multivariate case and cuts down the complexity of data. This technique offers a visual tool to assess the main direction in which trajectories vary, patterns of interest, clusters in the data and outlier detection. This method is applied to aircraft trajectories and the problem of registration is discussed when phase and amplitude variations are mixed.

7, avenue Édouard Belin - BP 54005 - 31055 Toulouse cedex 4 - Tél. +33 (0)5 62 17 40 00 - Fax +33 (0)5 62 17 40 23

Functional Principal Component Analysis of Aircraft Trajectories

Florence Nicol *
ENAC-MAIAA, Toulouse.

Keywords. Functional Data Analysis, Principal Component Analysis, Random Variable, Karhunen-Loève Decomposition, Dimension Reduction, Registration Problem.

Introduction

Principal Component Analysis (PCA) was one of the first methods of multivariate statistical analysis to be generalized to functional data that are assumed to be drawn from a continuous stochastic process. This point of view differs from standard statistical approaches: the nature of observations is different as we assume that the underlying structure of a raw observation is functional. Rather than on a sequence of individual points or finite-dimensional vectors. In this work, we will focus on Functional Principal Component Analysis (FPCA) which is an useful tool, providing common functional components explaining the structure of individual trajectories. In Section 1, we briefly recall the main ideas of a standard PCA approach. The general framework for Functional Data Analysis (FDA) is presented in Section 2 and the FPCA approach is formalized in Section 3. In Section 4, the registration problem is considered when phase variation due to time lags and amplitude variation due to intensity differences are mixed. Finally, FPCA is applied to aircraft trajectories that can be viewed as functional data.

1 Multivariate Principal Component Analysis

Principal Component Analysis (PCA), proposed by Pearson [20] and developed by Hotelling [13], is a powerful exploratory statistical method which has become a major tool for the analysis of multivariate data. When we observe more than two numeric variables, their variance and covariance matrices can be difficult to interpret. PCA allows to analyze data variability by studying the structure of their covariance matrix.

The main idea of PCA relies on creating a small number of new variables, which preserve the maximum amount of variation in the original variables and will make it possible to synthetize the large quantity of information into an understandable form. To illustrate this, let us suppose that we want to extract a two-dimensional representation (e.g. a picture of my cat Filou) from a three-dimensional image. What is the best view that will better preserve the representation of the object (side view, front

*Corresponding Author: ENAC, MAIAA, 7 avenue Edouard Belin, CS 54005, F-31055 Toulouse, FRANCE; E-mail: nicol@recherche.enac.fr.

view or back view of Filou)? PCA has many advantages. Firstly, the number of these new variables may give some indication as to the complexity of data. Next, these new descriptors are required to be orthogonal so that they eliminate redundant information contained in the original variables. For this reason, PCA can be seen as a tool for compression and data reduction. Finally, PCA can also be used as an outlier detection method.

Let X be the centered random vector of the p original variables $X_1, \ldots, X_p$. PCA consists in looking for a linear subspace with smaller dimension on which the variance of the projected data is maximal. Let Δ_1 be the linear subspace spanned by the original variables defined by a unit vector u_1 and $Y_1 = \langle u_1, X \rangle = u_1^T X$ be the projection of X onto Δ_1. This new random variable is a linear combination of the p original variables. The vector of weights u_1 has to be chosen such that the projection Y_1 onto Δ_1 has a maximal variance (also called *inertia*)

$$\mathrm{Var}\left(\langle u_1, X \rangle \right) = u_1^T S u_1 \quad \text{subject to } u_1^T u_1 = 1,$$

where $S = \mathbf{E}\left[X X^T \right]$ represents the covariance matrix of the random vector X. The solution of such a quadratic maximization problem is the eigenvector of S corresponding to the largest eigenvalue λ_1. We then search the unit vector u_2 orthogonal to u_1 such that the variance of the projected variables onto this new direction is maximal. Similarly, this is solved by finding the eigenvector of S associated to the second largest eigenvalue λ_2.

Using a stepwise procedure, the subspace of q dimensions which we are looking for is spanned by the weight vectors $u_1, \ldots, u_q$ so that each unit vector u_j maximizes the inertia of the projection $Y_j = \langle u_j, X \rangle$

$$\mathrm{Var}\left(\langle u_j, X \rangle \right) = u_j^T S u_j \quad \text{subject to } \langle u_i, u_j \rangle = \delta_{ij},\ i \leq j,\ j = 1, \ldots, q.$$

This is solved by finding the eigenvectors of S corresponding to the non-increasing sequence of eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_q \geq 0$. Weight vectors $u_1, \ldots, u_q$ are called *principal component factors* and the new random variables Y_j are called *principal component scores*. As is noted in [25], the unit norm constraint on the weights is essential to make the problem well-defined because the variance could be made arbitrarily large. Moreover, the weights have to be orthogonal to those identified previously so that they indicate something new. The amount of variation will decline stepwise and at some point we expect to lose interest in modes of variation so that only a little number of components will be sufficient to seek the most important modes of variation.

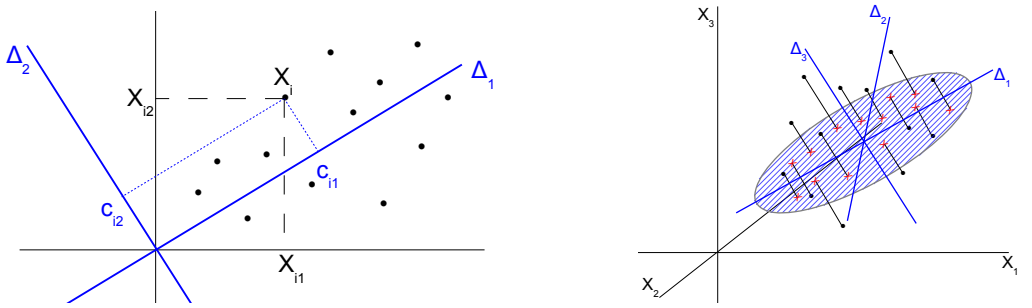

Figure 1: Scores of $p = 2$ variables (left panel) and $p = 3$ variables (right panel).

As the covariance matrix S is unknown, we need to replace it by the sample covariance matrix

$$\Sigma_n = \frac{1}{n} X^T X,$$

where X is the centered data matrix in which columns represent the p variables and lines represent the n individuals. The principal component score corresponding to the ith individual projected onto the jth factor direction can be expressed as follows

$$c_{ij} = u_j^T x_i, \quad j = 1 \ldots, p,\ i = 1, \ldots, n,$$

where x_i is the individual vector $(x_{i1}, \ldots, x_{ip})^T$ of the original data. We can express the score vector $c_j = Xu_j$ as a linear combination of the original variables, where u_j is reached by solving the maximization problem

$$\max_{u_j \in \mathbb{R}^p} \mathrm{Var}\,(\,c_j\,) = u_j^T \Sigma_n u_j, \quad \text{subject to } \langle u_i, u_j \rangle = \delta_{ij},\ i \leq j,\ j = 1, \ldots, p.$$

The solution is given by the eigenvector corresponding to the jth largest eigenvalue λ_j of the sample covariance matrix Σ_n. The maximum variance is then equal to λ_j.

The eigenvalues $\lambda_1, \ldots, \lambda_p$ measure the variation of the projected data onto the u_j direction, $j = 1, \ldots, p$, and their sum is equal to the total variance of the data. Then the cumulated ratio of variance explained by the first q principal component factors

$$\tau_q^C = \frac{\sum_{k=1}^{q} \lambda_k}{\sum_{k=1}^{p} \lambda_k}$$

measures the influence of these first q factors $u_1, \ldots, u_q$, i.e. the quality of representations onto the principal planes. Finally, the score vectors $c_1, \ldots, c_p$ contain the coordinates of the orthogonal projections of the data onto the axis defined by the factors u_j. Then, we can easily visualize the projected data by means of two dimensional scatter plots of scores and interpret outliers or some groups of individuals relatively to the principal component factors. Moreover, we may interpret the factor u_j according to the original variables by computing the linear correlation coefficients between principal component factors and original variables.

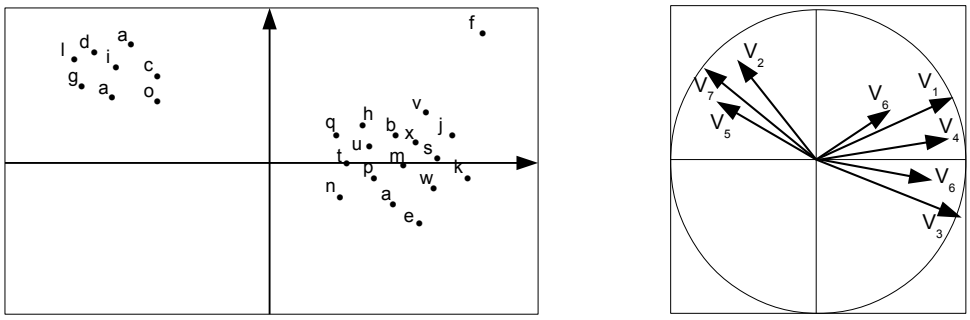

Figure 2: Representation of individuals (left panel) and variables (right panel) onto a plane defined by two principal component factors.

We can reconstruct the original data from the principal component factors $u_1, \ldots, u_p$ by means of the following reconstitution formula

$$X = \sum_{i=1}^{p} c_j u_j^T = \sum_{i=1}^{p} \sqrt{\lambda_j} z_j u_j^T,$$

where z_j and u_j respectively represent the eigenvectors of the matrices $\frac{1}{n} X X^T$ and $\frac{1}{n} X^T X$, associated to the same eigenvalue λ_j. We can then perform a data compression by using the q first components, $q \leq p$,

$$X^* = \sum_{i=1}^{q} c_j u_j^T \simeq X = \sum_{i=1}^{p} c_j u_j^T.$$

One of the advantages of PCA is to retain the first q terms in the above expansion as an approximation of the original data and hence achieve dimension reduction. The choice of the dimension q is then crucial for the quality of data compression.

To summarize, PCA is a linear factorial method. The quantity of information of the data is synthetized by creating new descriptors (in limited number) which are uncorrelated linear combinations of the originally correlated variables with maximal variance. Because the first principal components capture the maximum of the variability, PCA can be seen as a dimension reduction technique. In addition, PCA is a powerful tool for visualizing relations between variables and eventual groups of individuals, and for detecting possible outliers. There are many generalizations of PCA, mainly based on non-linear transformations of data, among others:

- Independant Component Analysis (ICA), in which components are not only orthogonal but independent,
- Curvilinear Component Analysis (CCA), in which we determine a subspace that best respects the distances,
- Kernel Principal Component Analysis, in which a non-linear projection into a high-dimensional space is followed by a standard PCA.

2 Functional Data Analysis and Random Function

Functional Data Analysis (FDA) consists in studying a sample of random functions generated from an underlying process. This point of view differs from standard statistical approaches: the nature of observations is different as we assume that the underlying structure of a raw observation is functional. Rather than on a sequence of individual points or finite-dimensional vectors as in a classical approach, we focus on problems raised by the analysis of a sample of functions (curves or images), when data are assumed to be drawn from a continuous stochastic process X. The sample of data consists of n functions $x_1(t), \ldots, x_n(t)$, $t \in J$, where J is a compact interval. Rather than a N-dimensional vector $(x_{i1}, \ldots, x_{iN})^T$ as in multivariate case, we entirely observe a function $x_i(t)$, $i = 1, \ldots, n$. This yields a vector of functions rather than a $n \times N$ data matrix, where each function x_i consists of infinitely many values $x_i(t)$, $t \in J$. General definitions of functional variables and functional data are given in [9] as follows.

Definition 2.1. *A random variable $X = \{X(t), t \in J\}$ is called funtional variable (f.v.) if it takes values in an infinite dimensional space (a functional space $\mathcal{H}$). An observation x of X is called a functional datum.*

Definition 2.2. *A functional dataset $x_1, \ldots, x_n$ is the observation of n functional variables $X_1, \ldots, X_n$ identically distributed as X (or n realizations of the f.v. X).*

Usually, the functional variable X can be viewed as a second order stochastic process and $\mathcal{H}$ as the separable Hilbert space $L^2(J)$ of square integrable functions defined on the interval J. The associated inner product for such functions is $\langle x, y \rangle = \int x(t)y(t)dt$ and the most common type of norm, called L^2-norm, is related to the above inner product through the relation $\|x\|^2 = \langle x, x \rangle = \int x^2(t)dt$. In a functional context, equivalence between norms fails and the choice of a preliminary norm becomes crucial that can be drawn by the shape of the functions, as noted in [9].

Let X be a square integrable functional variable with values in a separable Hilbert space $\mathcal{H}$. We can define a few usual functional characteristics of X, for all $s, t \in J$, as follows:

- the mean function $\mu(t) = \mathbf{E}[X(t)]$,
- the covariance function $\mathbf{Cov}_X(s,t) = \sigma(s,t) = \mathbf{E}[X(t)X(s)] - \mathbf{E}[X(t)]\,\mathbf{E}[X(s)]$,
- the variance function $\mathbf{Var}_X(t) = \sigma^2(t) = \mathbf{E}[X(t)^2] - (\mathbf{E}[X(t)])^2$,

In the following, we will assume that X is centered, i.e. $\mu = 0$, otherwise, subsequent results refer to $X - \mu$. In addition, the covariance operator induced by the covariance function, plays a crucial role in FDA, particularly in Functional Principal Component Analysis, as will be seen in the next section.

Definition 2.3. *The covariance operator* $\Gamma : \mathcal{H} \longrightarrow \mathcal{H}$ *is defined by*

$$\forall v \in \mathcal{H}, \quad \Gamma v(t) = \int_J \sigma(s,t)v(s)ds.$$

The covariance operator is a linear Hilbert-Schmidt operator in the functional space of square integrable functions $L^2(J)$ associated to the Hilbert-Schmidt kernel σ [5].

We can derive some empirical characteristics, among others, the sample mean function, the sample covariance function and the sample variance function as below

$$\overline{X}_n(t) = \frac{1}{n}\sum_{j=1}^{n} X_j(t),$$

$$\widehat{\sigma}_n(s,t) = \frac{1}{n}\sum_{j=1}^{n} X_j(t)X_j(s),$$

$$\widehat{\sigma}_n^2(t) = \widehat{\sigma}_n(t,t) = \frac{1}{n}\sum_{j=1}^{n} X_j(t)^2,$$

where $X_1, \ldots, X_n$ are independent functional variables identically distributed as X. The sample covariance function $\widehat{\sigma}_n$ induces the sample covariance operator $\widehat{\Gamma}_n$ as follows

$$\widehat{\Gamma}_n v(t) = \int_J \widehat{\sigma}_n(s,t)v(s)ds = \frac{1}{n}\sum_{j=1}^{n}\langle X_j, v\rangle X_j(t), \quad v \in \mathcal{H}.$$

Note that X is a functional space $\mathcal{H}$-valued random function and its observation x is a non-random function of $\mathcal{H}$. Usually, in practice, functional data $x_1, \ldots, x_n$ are observed discretely: we only observe a set of function values on a set of arguments that are not necessarily equidistant or the same for all functions. Because data are observed on a discretized grid, it could make sense to apply standard multivariate statistical tools where at each time value t_j, the observed vector-functions $(x_i(t_j))_{i=1,\ldots,n}$ can be viewed as variable vectors. Yet in recent years, advances in computing and data storage have increased the number of observations on ever finer grids. Standard methods of multivariate statistics have became inadequate, being plagued by the "curse of dimensionality", as the number of variables has became much more important than the number of individuals. As a result, statistical methods developed for multivariate analysis of random vectors are inoperative and FDA is a relevant alternative of multivariate statistical tools. As examples of FDA techniques, we can mention Functional Principal Component Analysis, Functional Discriminant Analysis and Functional Linear Models.

Discretized data have thereby to be transformed into functional data, as is requested in this framework, especially when observations are noisy. Most procedures developed in FDA are based on the use of interpolation or smoothing methods in order to estimate

the functional data from noisy observations. Examples of such methods outlined in [25] are, among others, kernel estimation, local polynomial estimation, smoothing splines, B-splines and basis function expansions such as a Fourier basis or wavelets. When the observed data are noisy, it may be important to combine smoothing techniques within functional data analysis. Finally, we can distinguish two important characteristics of functional data: data are intrinsiquely functional in nature (considered to be elements of an infinite-dimensional space) and the observed measurements are viewed as the values of sample paths with possibly measurement errors. Then, in FDA, two types of errors have often to be considered: sampling error in random functions generated from an underlying process, and measurement error when functions are unknown, discrete noisy data.

For illustrating, in Air Traffic Management (ATM), the aircraft trajectory data $f_i(t) = (x_i(t), y_i(t), z_i(t))$, $i = 1, \ldots, n$, collected over time are effectively producing three dimensional functions over the observed intervals $[0, T_i]$. There is no way to measure f_i at each time point, because aircraft trajectories are measured with radars. We only have access to values at given times with about N_i radar measurements for each trajectory made over slightly different intervals $[0, T_i]$. Time arguments at which

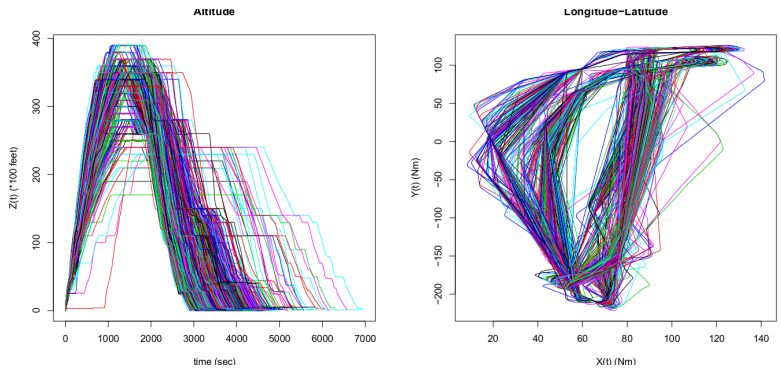

Figure 3: Sample aircraft trajectories (Paris-Toulouse)

trajectories are observed are not necessarily the same for each aircraft and may vary from one record to another. Although each observation could be viewed as N_i data points rather than one function, the collection of points possess a certain smoothness property that facilitates functional data interpretation. However, the assumption that all aircraft trajectories are sample paths from a single stochastic process defined on a time interval $[a, b]$ is clearly not satisfied: departure times are different and the time to destination is related to the aircraft type and the wind experienced along the flight. Ensuring a common starting time is very easy, just by assigning time 0 to the start of the flight and shifting accordingly all sample times along the trajectory as in Figure 3. The issue arising from aircraft type and exogeneous stochastic factors is more challenging. The remaining individual time variation, due to differences in dynamics, which occurs in addition to amplitude variation, is a more complex problem, well known in FDA as *registration problem*.

Aircraft trajectories exhibit high local variability both in amplitude and in dynamics. We could be interested in exploring the ways in which aircraft trajectories vary and highlight their characteristic features. Some of these features are expected to be there but other aspects may be surprising and can eventually be related to other variables such as wind, temperature, route or aircraft type. An extended problem is to bring out the common features between different routes. Visualization and classification of such trajectories may be another interesting problem in an exploratory analysis. One may identify aircrafts with outlying trajectories that may be studied or removed

before proceeding further analysis. In addition, a principal component analysis would be helpful to generate new aircraft trajectory samples. For instance, clustering and PCA techniques may be useful in direct and reverse edge bundling techniques applied to aircraft trajectories (work in progress).

Similar problems arise in many fields of applied research; for instance, biology and biomedicine with longitudinal growth studies, medicine with psychophysiological studies of electro-encephalogram (EEG) curves and brain images, applied economics with studies of production functions and economic panel data, lead to similar functional patterns. There are numerous examples in the literature, in various other fields, as in meteorology with mean temperature curves, or in speech recognition with log-periodogram curves for different phoneme classes. The goals of functional data analysis outlined in [25], are essentially the same as those of any other branch of statistics, as FDA aims to:

- represent the data in ways that facilitate further analysis;
- display the data so as to highlight various characteristics;
- study important sources of pattern and variation among the data;
- explain variation in a outcome or dependant variable by using input or independant variable information;
- compare two or more sets of data with respect to certain types of variation, where two sets of data can contain different sets of replicates of the same functions, or different functions for a common set of replicates.

3 Functional Principal Component Analysis

Multivariate Principal Component Analysis (PCA) is a powerful exploratory statistical method which synthetizes the quantity of data information by creating new descriptors when we observe more than two numeric variables [20, 13]. The main idea of PCA relies on creating a small number of new uncorrelated variables with maximal variance as linear combination of the originally correlated variables. PCA was one of the first methods of multivariate analysis to be generalized to the infinite-dimensional case. The first studies and tools developed for functional data were based on signal processing, principal component analysis and Karhunen-Loève decomposition. Rao [21] and Tucker [31] first introduced the earliest approach of PCA that linked factor analysis methods with growth curve models. Deville [8] generalized principal component analysis to stochastic processes and introduced the term of harmonic analysis for the orthogonal decomposition of a random function. Dauxois et al. [6] [7] proposed a mathematical framework and studied consistence and asymptotic properties for the principal component analysis of a vector random function. As for the covariance matrix in the multivariate standard case, the variance and covariance functions of functional variables are difficult to interpret and one goal is to analyze the variability of the functional data in a understandable manner. Functional Principal Component Analysis (FPCA) is an useful tool for studying functional data providing common functional components explaining the structure of individual trajectories. By approximating infinite-dimensional random functions by a finite number of scores, FPCA appears as a dimension reduction technique just as in the multivariate case and cuts down the complexity of the data. Finally, FPCA can be seen from two different points of view: a non-parametric point of view and a semi-parametric model, these two approaches being connected by the Karhunen-Loève decomposition.

3.1 Generalization to the infinite-dimensional case

In FDA, the counterparts of variable values $x_i = (x_{i1}, \ldots, x_{ip})^T$ are function values $x_i(t)$, $i = 1, \ldots, n$. Many properties of standard PCA can be generalized to infinite dimension, replacing matrices by linear operators, summations over j by integrations over t to define the inner product in the square integrable functional Hilbert space $\mathcal{H}$. In a non-parametric point of view, the variability of the sample is characterized by spectral decomposition of the sample covariance operator. Suppose that X is a centered square integrable functional variable of $\mathcal{H}$. As in multivariate PCA, we want to find weight functions γ_i such that the variance of the linear combination $\langle \gamma_i, X \rangle$ is maximal

$$\max_{\gamma_i \in \mathcal{H}} \mathrm{Var}\left(\langle \gamma_i, X \rangle \right) \quad \text{subject to } \langle \gamma_i, \gamma_k \rangle = \delta_{ik},\ k \leq i,\ i = 1, 2, \ldots, \tag{1}$$

or equivalently

$$\max_{\gamma_i \in \mathcal{H}} \int \int \gamma_i(s) \sigma(s,t) \gamma_i(t) ds dt \quad \text{subject to } \langle \gamma_i, \gamma_k \rangle = \delta_{ik},\ k \leq i,\ i = 1, 2, \ldots,$$

where $\sigma(s,t) = \mathbf{E}\left[X(s)X(t) \right]$ denotes the covariance function of X. The solutions are obtained by solving the Fredholm functional eigenequation,

$$\int_J \sigma(s,t) \gamma_i(t) dt = \lambda_i \gamma_i(s), \quad s \in J,$$

that can be expressed by means of the covariance operator Γ induced by the covariance function σ such that

$$\Gamma \gamma_i(s) = \lambda_i \gamma_i(s), \quad s \in J, \tag{2}$$

where γ_i is now an eigenfunction rather an eigenvector, corresponding to the eigenvalues λ_i of the covariance operator Γ, and the maximum variance is equal to λ_i. The eigenfunctions γ_i of the covariance operator Γ are called *functional principal components* or *principal component functions* and the random variables $\theta_i = \langle \gamma_i, X \rangle = \int \gamma_i(t) X(t) dt$ are called *principal component scores* of X into the γ_i-direction [25].

3.2 Estimation and Properties

When the covariance function is unknown, we can replace it by its sample version. The maximization problem (1) becomes

$$\max_{\gamma_i \in \mathcal{H}} \frac{1}{n} \sum_{j=1}^{n} \langle X_j, \gamma_i \rangle^2 = \max_{\gamma_i \in \mathcal{H}} \langle \gamma_i, \widehat{\Gamma}_n \gamma_i \rangle \quad \text{subject to } \langle \gamma_i, \gamma_k \rangle = \delta_{ik},\ k \leq i,\ i = 1, \ldots, n.$$

or equivalently

$$\max_{\gamma_i \in \mathcal{H}} \int \int \gamma_i(s) \widehat{\sigma}_n(s,t) \gamma_i(t) ds dt \quad \text{subject to } \langle \gamma_i, \gamma_k \rangle = \delta_{ik},\ k \leq i,\ i = 1, \ldots, n.$$

The solutions are obtained by solving the empirical version of the Fredholm eigenequation, for $i = 1, \ldots, n$,

$$\widehat{\Gamma}_n \widehat{\gamma}_i(s) = \widehat{\lambda}_i \widehat{\gamma}_i(s), \quad s \in J,$$

where $\widehat{\gamma}_1, \ldots, \widehat{\gamma}_n$ are the eigenfunctions of $\widehat{\Gamma}_n$, ordered by the corresponding eigenvalues $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \cdots \geq \widehat{\lambda}_n \geq 0$ and they form an orthogonal basis of the linear space spanned

by $X_1, \ldots, X_n$. The scores $\theta_{ij} = \langle \widehat{\gamma}_i, X_j \rangle$, $j = 1, \ldots, n$, into the γ_i-direction are centered and uncorrelated random variables such that, for all $i, k = 1, \ldots, n$, $i \neq k$,

$$\frac{1}{n}\sum_{j=1}^{n} \theta_{ij} = 0, \quad \frac{1}{n}\sum_{j=1}^{n} \theta_{ij}\theta_{kj} = 0, \quad \frac{1}{n}\sum_{j=1}^{n} \theta_{ij}^2 = \widehat{\lambda}_i.$$

Dauxois et al. [7] showed consistency and asymptotic properties of $\widehat{\Gamma}_n$, $\widehat{\gamma}_i$ and $\widehat{\lambda}_i$ under mild assumptions. Let $X_1, \ldots, X_n$ be an independent identically distributed random sample from a random process X, $\mathbf{E}\left[\, \|X\|^4 \,\right] < +\infty$. The norm of an operator $G : \mathcal{H} \longrightarrow \mathcal{H}$ is defined as

$$\|G\| = \sup_{\psi \in \mathcal{H}, \|\psi\|=1} \|G\psi\|.$$

In the literature, it is often assumed that $\mu = 0$ and all results presented here can be rewritten by replacing X by $X - \mu$. If the mean function μ is unknown, we can replace it by its sample version. It can be shown that the term that comes from using $\overline{X}_n$ instead of μ is negligible for asymptotic results.

Proposition 3.1. *1. The sequence of random operators $(\widehat{\Gamma}_n)_{n\in\mathbb{N}}$ is unbiased and converges almost surely to Γ.*

2. For any $v \in \mathcal{H}$ with $\|v\| = 1$, we obtain

$$\mathbf{E}\left[\, \|\widehat{\Gamma}_n - \Gamma\|^2 \,\right] \leq \frac{1}{n}\mathbf{E}\left[\, \|X\|^4 \,\right].$$

3. Let $Z = n^{1/2}(\widehat{\sigma} - \sigma)$ denote the scaled difference between the covariance functions $\widehat{\sigma}$ and σ. Then, the random function Z converges in distribution to a Gaussian process with mean 0.

The following inequalities about the eigenvalue estimators can also be verified

Proposition 3.2. *1.* $\sup_{j\geq 0} \left|\widehat{\lambda}_j - \lambda_j\right| \leq \|\widehat{\Gamma}_n - \Gamma\|$.

2. If $\lambda_1 > \lambda_2$, then

$$\|\widehat{\gamma}_1 - \gamma_1\| \leq \frac{2\sqrt{2}\|\widehat{\Gamma}_n - \Gamma\|}{\lambda_1 - \lambda_2}.$$

3. If for some $j > 1$, $\lambda_{j-1} > \lambda_j > \lambda_{j+1}$, then

$$\|\widehat{\gamma}_j - \gamma_j\| \leq \frac{2\sqrt{2}\|\widehat{\Gamma}_n - \Gamma\|}{\min(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1})}.$$

This proposition has several important implications. Firstly, it is clear from the preceeding results that the eigenvalue estimators $\widehat{\lambda}_j$ converge asymptotically to λ_j at the rate $O(n^{-1/2})$. Secondly, we have mean-squared convergence at the rate $O(n^{-1})$. Finally, the eigenfunction estimators $\widehat{\gamma}_j$ become sensitive to closely spaced eigenvalues when the dimension L is large. If we cannot choose a reasonnably small dimension L, then we should increase the sample size n in order to overcome this problem. But this leads to slower convergence rate of the estimators $\widehat{\lambda}_j$ and $\widehat{\gamma}_j$.

Dauxois et al. [7] showed another important result on the asymptotic distribution of eigenvalues and eigenfunctions.

Proposition 3.3. *Assume that eigenvalues λ_j are all distinct. For each $j \in \mathbb{N}$, $\sqrt{n}(\widehat{\lambda}_j - \lambda_j)$ and $\sqrt{n}(\widehat{\gamma}_j - \gamma_j)$ converge in distribution to a zero-mean Gaussian random variable. Particularly, if X is a Gaussian process, the eigenvalue estimator $\widehat{\lambda}_j$ is asymptotically Gaussian with*

$$\sqrt{n}(\widehat{\lambda}_j - \lambda_j) \xrightarrow{\mathcal{L}} \mathcal{N}\left(0;\, 2\lambda_j^2\right).$$

Several estimation methods of scores and principal component functions were developed for FPCA, some of them are presented in the next section.

3.3 From Karhunen-Loève Representation to Functional Principal Components

An important characterization of FPCA as a semi-parametric model directly results from the Karhunen-Loève decomposition. Indeed, the eigenfunctions $\gamma_1, \gamma_2, \ldots$ of the covariance operator Γ form an orthonormal basis of the functional space $\mathcal{H}$ so that

$$X(t) = \sum_{i=1}^{+\infty} \theta_i \gamma_i(t),$$

where the principal component scores $\theta_i = \langle \gamma_i, X \rangle$ are centered and uncorrelated random variables such that $\mathrm{Var}\,(\,\theta_i\,) = \lambda_i \geq 0$. This yields an important representation of individual trajectory X which provides a decomposition of X into orthogonal components with uncorrelated random coefficients. This series expansion of X converges in the L^2-sense and also pointwise. The eigenvalue λ_i measures the variability in X onto the γ_i-direction and the random coefficients θ_i are independant if X is a Gaussian process. The total variance satisfies

$$\int \sigma^2(t)dt = \sum_{i=1}^{+\infty} \lambda_i = \sum_{i=1}^{+\infty} \mathbf{E}\left[\,\theta_i^2\,\right] < +\infty.$$

Another important property for FPCA involves the decomposition of variance and the best L-term approximation property.

Proposition 3.4. *For any further orthogonal basis $\psi_1, \psi_2, \ldots$ of $\mathcal{H}$ and every $L \in \mathbb{N}$,*

$$\mathbf{E}\left[\, \|X - \sum_{i=1}^{L} \theta_i \gamma_i\|^2 \,\right] \leq \mathbf{E}\left[\, \|X - \sum_{i=1}^{L} \langle \psi_i, X \rangle \psi\|^2 \,\right].$$

This means that the finite expansion $\sum_{i=1}^{L} \theta_i \gamma_i$ is the best approximation of X with a given number L of common components $\gamma_1, \ldots, \gamma_L$ with varying strengths captured by centered uncorrelated coefficients θ_i. Then, the maximization problem in (1) is equivalent to the minimization problem of the mean integrated square error

$$\min_{\gamma_1,\ldots,\gamma_L} \mathbf{E}\left[\, \|X - \sum_{i=1}^{L} \theta_i \gamma_i\|^2 \,\right], \tag{3}$$

that is solved by the first L eigenfunctions $\gamma_1, \ldots, \gamma_L$ of the covariance operator Γ ordered by the corresponding eigenvalues $\lambda_1, \ldots, \lambda_L$. The two approaches have then been connected by the Karhunen-Loève decomposition for which the mean integrated square error in (3) is minimum if $\gamma_1, \gamma_2, \ldots, \gamma_L$ are the first L eigenfunctions of the covariance operator Γ and $\theta_i = \langle \gamma_i, X \rangle$.

By characterizing individual trajectories X_j through an empirical Karhunen-Loève decomposition,

$$X_j(t) = \sum_{i=1}^{n} \theta_{ij} \widehat{\gamma}_i(t), \quad j = 1, \ldots, n,$$

FPCA leads to the best empirical basis expansion of functional data in the sense of the mean integrated square error. Then, this may be better than alternative representations of functional data by fixed basis functions such as Fourier series, wavelets or B-splines. Indeed, the disadvantage of fixed basis functions is that a larger number of them may be needed to correctly represent a given sample of trajectories. In addition, the estimated coefficients are not uncorrelated and are less convenient for subsequent applications such as functional regression. The random scores $\theta_{ij} = \langle \widehat{\gamma}_i, X_j \rangle$, $j = 1, \ldots, n$, represent the coordinates of the projected sample onto the $\widehat{\gamma}_i$-direction and can be interpreted as proportionality factors that represent strengths of the representation of each individual trajectory by the ith principal component function. Furthermore, FPCA provides eigenfunction estimates that can be interpreted as "modes of variation". These modes have a direct interpretation and are of interest in their own right. They offer a visual tool to assess the main direction in which functional data vary. As in the multivariate case, pairwise scatterplots of one score against another may reveal patterns of interest and clusters in the data. In addition, these plots may also be used to detect outliers and explain individual behaviour relatively to modes of variation.

As in the multivariate standard PCA, we can easily measure the quality of the representation by means of the eigenvalue estimators. The ith eigenvalue estimator $\widehat{\lambda}_i$ measures the variation of the scores $\theta_{i1}, \ldots, \theta_{in}$ into the $\widehat{\gamma}_i$-direction. The percentage of total variation τ_i explained by the ith principal component and the cumulated ratio of variation τ_L^C explained by the first L principal components are then computed from the following ratio

$$\tau_i = \frac{\widehat{\lambda}_i}{\sum_{i=1}^{n} \widehat{\lambda}_i}, \quad \tau_L^C = \frac{\sum_{k=1}^{L} \widehat{\gamma}_k}{\sum_{i=1}^{n} \widehat{\gamma}_i}.$$

The amount of explained variation will decline on each step and we expect that a small number L of components will be sufficient to account for a large part of variation. Indeed, the total approximated error $\sum_{k=L+1}^{n} \widehat{\lambda}_k$ is expected to be sufficiently small so that $\widehat{X}_j = \sum_{i=1}^{L} \theta_{ij} \widehat{\gamma}_i$ is a good approximation of X_j for a relevant choice of L. Determining a reasonable number L of components is often a crucial issue in FDA. We have seen in the preceeding section that closely-spaced eigenvalues may cause problems and lead to instability of the principal component estimators. The problem becomes even more serious if approximations from discretized noisy functional data are involved. To show this, let us assume that we observe n noisy functional data Y_{jk} on a time grid (t_{jk}), $k = 1, \ldots, m$

$$Y_{jk} = x_j(t_{jk}) + \varepsilon_{jk}, \quad j = 1, \ldots, n, \; k = 1, \ldots, m, \tag{4}$$

where the error term ε_{jk} are independent and identically distributed random variables, with $\mathbf{E}[\varepsilon_{jk}] = 0$ and $\mathrm{Var}(\varepsilon_{jk}) = \sigma^2$. The best possible approximation of the $x_j(t)$ by linear combinations of L basis functions in a L^2-sense is

$$x_j(t) \simeq \sum_{i=1}^{L} \theta_{ij} \gamma_i(t).$$

We can then view the problem stated in (4) as the following semi-parametric regression problem

$$Y_{ij} = \sum_{i=1}^{L} \theta_{ij}\gamma_i(t_{jk}) + \varepsilon_{jk}, \quad j = 1, \ldots, n,\ k = 1, \ldots, m,$$

where the i-th principal component γ_i can be interpreted as a regressor and the corresponding scores $\theta_{i1}, \ldots, \theta_{in}$ as unknown coefficients to be estimated. If we assume that the principal components $\gamma_1, \ldots, \gamma_L$ are known, we can estimate the scores θ_{ij} by least squares such that $\widehat{x}_j(t) = \sum_{j=1}^{L} \widehat{\theta}_{ij} g_i(t)$. It is well known that the bias decreases with the number L of regressors but the variance of $\widehat{x}_j(t)$ increases as L increases. As this trade-off between bias and variance, choosing $L = n$ components may be inadequate and high values of L are associated with high frequency components which represent the sampling noise. A simple and fast method to choose the dimension L is the scree plot that plots the cumulated proportion of variance explained by the first L components against the number of included components L. Alternative procedures to estimate an optimal dimension can be found in [15], [17] and [3].

4 Estimation Methods

In this section, we present several estimation methods developed for FPCA. The earliest method applied to discretized functional data is based on numerical integration or quadrature rules. A more sophisticated method is based on expansion of the functional data on a known basis. This method will better take into account the functional nature of the data and implies to reduce the eigenequation to discrete or matrix form. Finally, we will focus on the regularized FPCA which involves smooth functional principal components. In the following, we will suppose that we have observed a sequence of n functional data $x_1(t), \ldots, x_n(t)$ to which the mean function was substracted.

4.1 Discretization Method

Rao [21] and Tucker [31] first introduced the earliest approach of PCA applied to functional data discretized to a fine grid of time arguments that span the interval J. When the design points are the same for all the observed functions $x_1, \ldots, x_n$, the functional eigenequation can be approximated by using quadrature rules. Usually, functions are observed at the same time arguments, no necessarily equally spaced. This yields an $n \times N$ data matrix

$$\begin{pmatrix} x_1(t_1) & x_1(t_2) & \ldots & x_1(t_j) & \ldots & x_1(t_N) \\ x_2(t_1) & x_2(t_2) & \ldots & x_2(t_j) & \ldots & x_2(t_N) \\ \vdots & \vdots & \ldots & \vdots & \ldots & \vdots \\ x_i(t_1) & x_i(t_2) & \ldots & x_i(t_j) & \ldots & x_i(t_N) \\ \vdots & \vdots & \ldots & \vdots & \ldots & \vdots \\ x_n(t_1) & x_n(t_2) & \ldots & x_n(t_j) & \ldots & x_n(t_N) \end{pmatrix}.$$

The discretization method stands on the approximation of the integrals by a sum of discrete values as

$$\int f(t)dt \simeq \sum_{j=1}^{N} \omega_j f(t_j),$$

where N is the number of time arguments, t_j are the time arguments called *quadrature points* and ω_j are the weights called *quadrature weights*. Several numerical quadrature

schemes can be used to involve a discrete approximation of the functional eigenequation

$$\Sigma_n W \widetilde{\gamma}_m = \widetilde{\lambda}_m \widetilde{\gamma}_m,$$

where $\Sigma_n = (\widehat{\sigma}_n(t_i, t_j))_{i,j=1,\ldots,N}$ is the sample covariance matrix evaluated at the quadrature points and W is a diagonal matrix with diagonal values being the quadrature weights. The solutions $\widetilde{\gamma}_m = (\widetilde{\gamma}_m(t_1), \ldots, \widetilde{\gamma}_m(t_N))$ are the eigenvectors associated with the eigenvalues $\widetilde{\lambda}_m$ of the matrix $\Sigma_n W$. The orthonormality constraints are now

$$\sum_{j=1}^{N} \omega_j \widetilde{\gamma}_l(t_j) \widetilde{\gamma}_m(t_j) = \widetilde{\gamma}_l^T W \widetilde{\gamma}_m = \delta_{lm}, \quad l, m = 1, \ldots, N.$$

The eigenvectors $\widetilde{\gamma}_m$ form an orthonormal system relatively to the metric defined by the weight matrix W. In general, the choice of interpolation functions is equivalent to the choice of a metric.

We can express the functional eigenequation in an equivalent symmetric eigenvalue problem

$$W^{1/2} \Sigma_n W^{1/2} u_m = \widetilde{\lambda}_m u_m \quad \text{subject to } u_l^T u_m = \delta_{lm}, \quad l, m = 1, \ldots, N,$$

where $u_m = W^{1/2} \widetilde{\gamma}_m$. Then the whole procedure runs as follows

Algorithm:

1. Choose the quadrature points t_j and the quadrature weigths ω_j for $j = 1, \ldots, N$.
2. Compute the eigenvalues $\widetilde{\lambda}_m$ and the corresponding eigenvectors u_m of the symmetric matrix $W^{1/2} \Sigma_n W^{1/2}$.
3. Calculate the discretized principal components $\widetilde{\gamma}_m = W^{-1/2} u_m$.
4. If needed, approximate the functional principal components from the discrete values by using any convenient smoothing technique or interpolation method.

Ramsay and Silverman [25] note that, if the discretization values t_j are closely spaced, the choice of the interpolation method should not have a great effect compared to sampling errors, even if the observations are corrupted by noise.

A naive approach consists in directly determining the eigenvectors of the discretized sample covariance matrix Σ_n. This may lead to determine wrong results because the resulting principal components may not form an orthonormal system in a functional sense, except if the metric W is the identity matrix. A particular case is the one of equally spaced design points. The eigenequation then becomes

$$\omega \Sigma_n \widehat{\gamma}_m = \widehat{\lambda}_m \widehat{\gamma}_m,$$

and the solutions are those of the standard multivariate PCA on the data matrix X

$$\Sigma_n u_m = \rho_m u_m,$$

where $\widehat{\lambda}_m = \omega \rho_m$ and u_m are the eigenvectors associated to the eigenvalues ρ_m of the sample covariance matrix Σ_n. These eigenvectors obtained by the naive approach are orthogonal in a functional sense but a normalization correction is needed. Usually, the above equally spaced method is used after applying spline smoothing to the data.

4.2 Basis Function Expansion

Another way to reduce the eigenequation problem to a discrete form is to represent the functional data as linear combinations of known basis functions such as a Fourier basis or spline functions. Functional data are estimated by their projections onto a linear functional space spanned by K known basis functions $\psi_1, \ldots, \psi_K$ such as

$$\widetilde{x}_i(t) = \sum_{k=1}^{K} \theta_{ik}\psi_k(t) = \theta_i^T \psi(t),$$

where the unknown coefficient vectors $\theta_i = (\theta_{i1}, \ldots, \theta_{iK})^T$ have to be estimated from the data and $\psi(t)$ denotes the vector-valued function $(\psi_1(t), \ldots, \psi_K(t))^T$. This method takes into account the functional nature of the data and makes it possible to discretize the problem by replacing the functional data $x_i(t)$ by its coefficient vector θ_i, $i = 1, \ldots, n$. The sample covariance function of the projected data

$$\widetilde{\sigma}_n(s,t) = \frac{1}{n}\sum_{i=1}^{n} \widetilde{x}_i(s)\widetilde{x}_i(t) = \psi(s)^T \Theta \psi(t),$$

can be expressed by means of the $K \times K$ matrix $\Theta = \frac{1}{n}\sum_{i=1}^{n} \theta_i\theta_i^T$ which represents the covariance matrix of the coefficient vectors. Consider now the basis expansion of the eigenfunctions $\widetilde{\gamma}_m(s) = b_m^T\psi(s)$ where $b_m = (b_{m1}, \ldots, b_{mK})^T$ is the unknown coefficient vector to be determined. This yields the discretized eigenequation

$$\Theta W b_m = \widetilde{\lambda}_m b_m,$$

where $W = (\langle \psi_i, \psi_j \rangle)_{i,j=1,\ldots,K}$ is the matrix of the inner products $\langle \psi_i, \psi_j \rangle = \int \psi_i(t)\psi_j(t)dt$ of the basis functions. The solutions b_m are the eigenvectors associated with the eigenvalues $\widetilde{\lambda}_m$ of the matrix ΘW. The orthonormality constraints on the functional principal components satisfy

$$\langle \widetilde{\gamma}_l, \widetilde{\gamma}_m \rangle = \int \widetilde{\gamma}_l(t)\widetilde{\gamma}_m(t)dt = b_l^T W b_m = \delta_{lm}, \quad l, m = 1 \ldots, K.$$

We can remark that this method looks like the discretization method for which the coefficient vectors $\theta_i = (\theta_{i1}, \ldots, \theta_{iK})^T$ play the role of the discretized function vectors $x_i = (x_i(t_1), \ldots, x_i(t_N))^T$. FPCA is then equivalent to a standard multivariate PCA applied to the matrix of coefficients with the metric defined by the inner product matrix $W = (\langle \psi_i, \psi_j \rangle)_{i,j=1,\ldots,K}$.

We can express the eigenvalue problem in a more standard form $W^{1/2}\Theta W^{1/2}u_m = \widetilde{\lambda}_m u_m$, where $b_m = W^{-1/2}u_m$. Then the whole procedure runs as follows

Algorithm:

1. Calculate the matrices Θ, W and $W^{1/2}$ (Cholesky decomposition).
2. Determine the eigenvalues $\widetilde{\lambda}_1 \geq \widetilde{\lambda}_2 \geq \cdots \geq \widetilde{\lambda}_K$ of the matrix $W^{1/2}\Theta W^{1/2}$ and the corresponding eigenvectors $u_1, \ldots, u_K$.
3. Determine the coefficient vectors $b_m = W^{-1/2}u_m$.
4. The estimations of λ_m and γ_m are given by $\widetilde{\lambda}_m$ and $\widetilde{\gamma}_m(t) = \sum_{k=1}^{K} u_{km}\psi_k(t)$.

Note that the inner product matrix W may be evaluated by any quadrature techniques

except for particular choices of the basis functions. For example, for the orthonormal Fourier series, W is the identity matrix. Moreover, the maximum number of eigenfunctions is now equal to the dimension K of the basis, possibly smaller than the number N of discretized time arguments as is done in the the discretization approach. As noted in [25], the number K of basis functions depends on many considerations, in particular how efficient the basis functions are in reproducing the behaviour of the original functions.

Two special cases have to be considered. Firstly, if the basis functions are orthonormal, i.e. $W = I$, then the discretized eigenequation becomes $\Theta b_m = \widetilde{\lambda}_m b_m$ and the FPCA problem finally amounts to solving a standard multivariate PCA on the coefficient matrix. Such is the case if we use a Fourier series basis. The second special case was introduced by Kneip and Utikal [17] for the case in which the functional data are density functions. The main idea consists in using the functional data themselves as their own basis functions. This approach overcomes the problem of the flexibility of the basis respect to the data and is particularly appropriate if the number n of observed functions is small relatively to the number K of basis functions. This empirical basis implies that the covariance matrix Θ of the coefficients vectors is equal to $\frac{1}{n}I$. Then the discretized eigenequation becomes $Wb_m = n\widetilde{\lambda}_m b_m$ and the FPCA problem is reduced to the spectral analysis of the symmetric matrix $W = (\langle x_i, x_j \rangle)_{i,j=1,\dots,n}$ whose entries are the inner products of the functional data. The solutions b_m are the eigenvectors associated with the eigenvalues μ_m of the matrix W. They are related to the eigenvalues $\widetilde{\lambda}_m$ and the eigenfunctions $\widetilde{\gamma}_m$ as follows

$$\begin{aligned}
\widetilde{\mu}_m &= m\widetilde{\lambda}_m, \\
\widetilde{\gamma}_m &= (\widetilde{\mu}_m)^{-1/2} \sum_{i=1}^{n} \widetilde{p}_{im} x_i = \frac{\sum_{i=1}^{n} \theta_{im} x_i}{\sum_{i=1}^{n} \theta_{im}^2},
\end{aligned}$$

and the principal score of the function x_i onto the mth principal component is given by $\sqrt{\widetilde{\mu}_m}\widetilde{p}_{im}$. As previously, the entries of the matrix W have to be estimated by some quadrature techniques or by an appropriate estimation method. However, if the functional data x_i are not observed at the same time arguments, the estimation of the inner products in W may be difficult. A detailed discussion can be found in [17].

4.3 Smoothed Principal Component Analysis

In many applications, functional data are assumed to be smooth, and yet, the estimated principal component functions may be rough and present important variability because of the sampling error, the measurement noise and the choice of the basis functions. In the unsmoothed FPCA approaches, the functional principal components are estimated by maximizing the sample variance $\langle \gamma_m, \widehat{\Gamma}_n \gamma_m \rangle$ of the projected data under some orthonormality constraints on the principal components. Rather than first smoothing the functional data before proceeding with FPCA [25], it makes sense to incorporate this smoothness assumption into the estimation procedure. The *smoothed FPCA* approaches, also called *regularized FPCA* in [25], are based on the well-known roughness penalty approaches. The basic idea consists in adding a penalty term $PEN(\gamma_m)$ into the maximization problem such that the form

$$\langle \gamma_m, \widehat{\Gamma}_n \gamma_m \rangle - \alpha_m PEN(\gamma_m) \quad \text{subject to } \|\gamma_m\|^2 = 1 \text{ and } \langle \gamma_l, \gamma_m \rangle = 0 \text{ for } l < m, \quad (5)$$

quantifies the trade-off between fidelity to the data (in this case, it is measured by the sample variance of the projected data in the γ_m-direction) and roughness as measured

by the roughness penalty. This method was developed by Rice and Silverman [27] for which a sequence of smoothing parameters α_m controls the relative importance of the roughness for each principal component γ_m.

Rather than penalizing the variance as is done in Rice and Silverman [27], Silverman [30] proposed to incorpore the roughness penalty into the orthonormality constraints by modifying the norm. In this approach, functional data and eigenfunctions are assumed to have continuous and square integrable second derivatives. For a given smoothing parameter α, Silverman [30] defined the modified inner product as follows

$$\langle x, y\rangle_\alpha = \langle x, y\rangle + \alpha\langle x^{(2)}, y^{(2)}\rangle, \quad x, y \in \mathcal{H},$$

with the corresponding squared norm $\|x\|_\alpha^2 = \|x\|^2 + \alpha\|x^{(2)}\|^2$. These are slight generalizations of the standard Sobolev inner products and norms. In the FPCA procedure, the L^2-orthonormality constraints are replaced by orthonormality constraints with respect to the modified inner product $\langle x, y\rangle_\alpha$ that takes into account the roughness of the functions. Then, the functional principal components are found by maximizing

$$\max_{\gamma_m \in \mathcal{H}} \langle \gamma_m, \widehat{\Gamma}_n \gamma_m\rangle \quad \text{subject to } \|\gamma_m\|_\alpha^2 = 1 \text{ and } \langle \gamma_l, \gamma_m\rangle_\alpha = 0 \text{ for } l < m. \tag{6}$$

Roughness is then penalized in a different manner than in (5): the orthonormality constraint is now composed by the usual L^2-norm term $\|\gamma_m\|^2$ and a roughness penalty term $\|\gamma_m^{(2)}\|^2$. Finally, as noted in [25], the problem (6) is equivalent to maximize

$$\frac{\langle \gamma_m, \widehat{\Gamma}_n \gamma_m\rangle}{\|\gamma_m\|_\alpha^2} = \frac{\langle \gamma_m, \widehat{\Gamma}_n \gamma_m\rangle}{\|\gamma_m\|^2 + \alpha\|\gamma_m^{(2)}\|^2} \quad \text{subject to } \langle \gamma_l, \gamma_m\rangle_\alpha = 0 \text{ for } l < m. \tag{7}$$

To see this, note that scaling any γ_m to satisfy the orthonormality constraint in (6) does not affect the value of the ratio (7), and so the maximum of the ratio is unaffected by the imposition of the constraint. Once the constraint is imposed, the denominator of (7) is equal to 1, and so maximizing the ratio subject to $\langle \gamma_l, \gamma_m\rangle_\alpha = 0$ is exactly the same as the original maximization problem (6).

In [30], this method has been easily implemented in the periodic case when γ_m has square integrable fourth derivative and when its second and third derivatives satisfy some boundary conditions on the interval J. In this case, it is easy to check that $\langle x^{(2)}, y^{(2)}\rangle = \langle x, y^{(4)}\rangle$ by integrating by parts twice. This implies that the modified inner product can be expressed as $\langle x, y\rangle_\alpha = \langle x, y + \alpha y^{(4)}\rangle$. Defining the operator S by $(I + \alpha Q)^{-1/2}$ where Q denotes the fourth derivative operator, the maximization problem becomes

$$\max_{\gamma_m \in \mathcal{H}} \langle \gamma_m, \widehat{\Gamma}_n \gamma_m\rangle \quad \text{subject to } \langle \gamma_l, S^{-2}\gamma_m\rangle = \delta_{lm} \text{ for } l \leq m.$$

The solutions are obtained by solving the generalized eigenequation

$$\widehat{\Gamma}_n \widetilde{\gamma}_m = \widetilde{\lambda}_m S^{-2}\widetilde{\gamma}_m = \widetilde{\lambda}_m (I + \alpha Q)\widetilde{\gamma}_m, \tag{8}$$

or equivalently $S\widehat{\Gamma}_n S g_m = \widetilde{\lambda}_m g_m$ where $\widetilde{\gamma}_m = S g_m$ and g_m is the eigenfunction with the associated eigenvalue λ_m of the operator $S\widehat{\Gamma}_n S$.

A simple algorithm for this method was easily implemented in [30] in the case of Fourier series. Indeed, the operator S can be interpreted as a smoothing operator and finding the eigenfunctions of the operator $S\widehat{\Gamma}_n S$ is equivalent to carrying out an unsmoothed FPCA of the smoothed data Sx_i. Next, the smoothing operator S is applied to the resulting eigenvectors. Note that the orthonormality constraints are

satisfied in the sense of the modified inner product $\langle x, y\rangle_\alpha$ and not in the L^2-sense as $g_l^T g_m = \gamma_l^T S^{-2} \gamma_m = \langle \gamma_l, \gamma_m \rangle_\alpha = \delta_{lm}$.

When Fourier expansions are no longer appropriate because fo the boundary conditions, the previous method can be easily extended to any suitable basis expansion such that $\widetilde{x}_i(t) = \theta_i^T \psi(t)$ and $\widetilde{\gamma}_m(t) = b_m^T \psi(t)$ where $\theta_i = (\theta_{i1}, \ldots, \theta_{iK})^T$, $b_m = (b_{m1}, \ldots, b_{mK})^T$ are the coefficient vectors and $\psi(t)$ denotes the vector $(\psi_1(t), \ldots, \psi_K(t))^T$ of K known basis functions. Then, the penalized sample covariance in (7) can be expressed by means of the basis functions as

$$\frac{b_m^T W \Theta W b_m^T}{b_m^T W b_m + \alpha\, b_m^T K b_m},$$

where $\Theta = \frac{1}{n} \sum_{i=1}^n \theta_i \theta_i^T$ is the covariance matrix of the data coefficient vectors, $W = (\langle \psi_i, \psi_j \rangle)_{i,j=1,\ldots,K}$ and $K = (\langle \psi_i^{(2)}, \psi_j^{(2)} \rangle)_{i,j=1,\ldots,K}$ are respectively the inner products matrices of the basis functions and of their second derivatives. This yields the discretized eigenequation

$$W \Theta W b_m = \lambda_m (W + \alpha K) b_m.$$

Performing a factorization $LL^T = W + \alpha K$ and defining $S = L^{-1}$, the equation can now be written as

$$(SW\Theta WS^T) u_m = \lambda_m u_m,$$

where $b_m = S^T u_m$ and u_m is the eigenfunction with associated eigenvalue $\tilde{\lambda}_m$ of the matrix $SW\Theta WS^T$ that represents the covariance matrix of the transformed coefficients vector $SW\theta_i$. This is equivalent to performing a basis expansion FPCA on the matrix of the new coefficients $\widetilde{\theta}_i = SW\theta_i$. Then the whole procedure runs as follows

Algorithm:

1. Calculate the matrices Θ, W and K.
2. Find L and $S = L^{-1}$ (Cholesky decomposition or SVD).
3. Carry out a PCA on the coefficient vectors $\widetilde{\theta}_i = SW\theta_i$ and determine the eigenvalues $\widetilde{\lambda}_m$ and the eigenvectors u_m of $SW\Theta WS^T$.
4. Apply the smoothing operator S^T, calculate the coefficient vectors $b_m = S^T u_m$ and renormalize them so that $b_m^T W b_m = 1$.
5. The estimations of λ_m and γ_m are given by $\widetilde{\lambda}_m$ and $\widetilde{\gamma}_m(t) = b_m^T \psi = u_m^T S \psi$.

The drawback of this technique is that the functional principal components $\widetilde{\gamma}_m$ are not orthogonal in the L^2-sense. Then, we may orthonormalize the coefficents b_m with respect to the matrix W, using a Gramm-Schmidt procedure. Silverman [30] showed that the estimates are consistent under some regularity conditions when $n \to +\infty$ and $\alpha \to 0$. The automatic choice of an appropriate smoothing parameter α can be solved by using a cross-validation approach. Moreover, some theoretical results suggest that the procedure developed by Silverman [30] gives better results than those of the Rice and Silverman procedure [27]. Another approach, followed by Besse and Ramsay [2] and Kneip [15], consists in smoothing the functional data first and then carrying out unsmoothed PCA on the smoothed data. The differences with the smoothed FPCA appear to be small but the comparison depends on the way in which the data are smoothed. In all cases, smoothing procedures clearly improve the quality of the estimation of the principal components.

5 The Registration Problem

The process of registration, well known in the field of functional data analysis [28, 11, 25], is an important preliminary step before further statistical analysis. Indeed, a serious drawback must be considered when functions are shifted, owing to time lags or general differences in dynamics. Phase variation due to time lags and amplitude variation due to intensity differences are mixed and it may be hard to identify what is due to each kind of variation. This problem due to such mixed variations can hinder even the simplest analysis of trajectories.

Firstly, standard statistical tools such as pointwise mean, variance and covariance functions, may not be appropriate. For example, a sample mean function may badly summarize sample functions in the sense that it does not accuratly capture typical characteristics as illustrated in Figure 4. In addition, more complex analysis such as

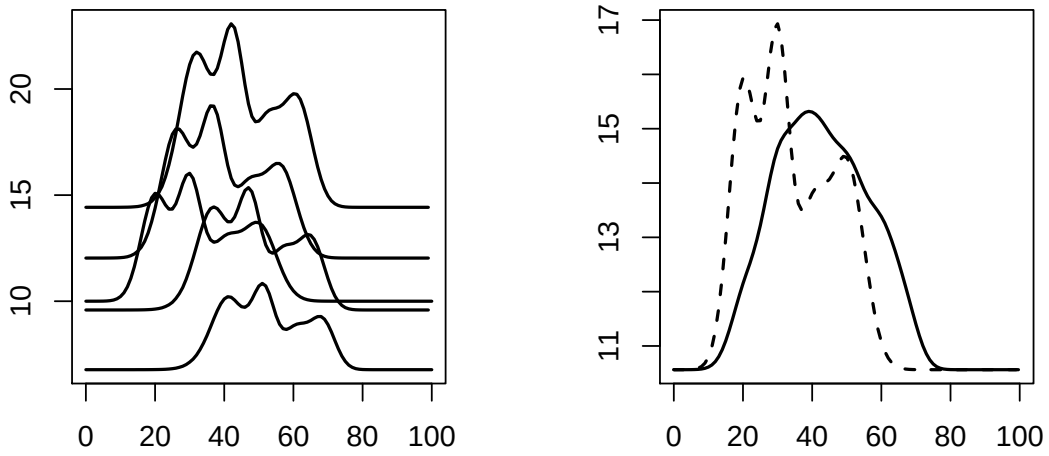

Figure 4: The left panel gives plot of simulated sample functions and the right panel displays mean functions of unregistered curves (solid line) and registered curves (dashed line).

trajectory clustering may be failed because distance between two similar trajectories may be wrongly inflated by phase variation. In the case of FPCA, some functional components may not correspond to effects added to a mean function but rather to a transformation of time arguments and they may be shifted from function to function. Then, FPCA may produce too many components and some components can be expressed as derivatives of others.

A registration method consists in aligning features of a sample of functions by non decreasing monotone transformations of time arguments, often called *warping functions*. These time transformations have to capture phase variation in the original functions and transform the different individual time scales into a common time interval for each function. Generally speaking, a non decreasing smooth mapping $h_i \colon [a, b] \to [c_i, d_i]$, with $[c_i, d_i]$ the original time domain of the trajectory, is used to map each trajectory y_i to a reference trajectory x, usually called *target* or *template function*, already defined on $[a, b]$. In this way, remaining amplitude differences between registered (aligned) trajectories $y_i \circ h_i$ can be analyzed by standard statistical methods. The choice of a template function is sometimes tricky and it may be simply selected among the sample trajectories as a reference with which we want to synchronize all other trajectories. Note that warping functions h_i have to be invertible so that for the same sequence of events, time points on two different scales correspond to each other uniquely. Moreover, we require that these functions are smooth in the sense of being differentiable a certain number of times.

Most of literature deals with two kinds of registration methods: *landmark registration* and *goodness-of-fit based registration* methods. A classical procedure called *marker* or *landmark registration* aims to align curves by identifying locations $t_{i1}, \ldots, t_{iK}$ of certain structural features such as local minima, maxima or inflexion points, which can be found in each curve [4, 14, 11]. Curves are then aligned by transforming time in such a way that marker events may occur at the same time $t_{01}, \ldots, t_{0K}$, giving $h_i(t_{0k}) = t_{ik}$, $k = 1, \ldots, K$. Complete warping functions h_i are then obtained by smooth monotonic interpolation. This non-parametric method is able to estimate possibly non-linear transformations. However, marker events may be missing in certain curves and feature location estimates can be hard to identify. Finally, phase variation may remain between too widely separated markers. An alternative method is based on goodness-of-fit by minimizing distance between registered trajectories $y_i \circ h_i$ and a template trajectory, with possible inclusion of a roughness penalty for h_i [22, 23]. A serious problem may appear when functions have important amplitude variation and distance minimizing methods may break down in this case. Indeed, the goodness-of-fit criterion will tend to explain main differences by phase variation and leads to locally constant registered trajectories as in Figure 5. Then it may be crucial to specify a procedure for estima-

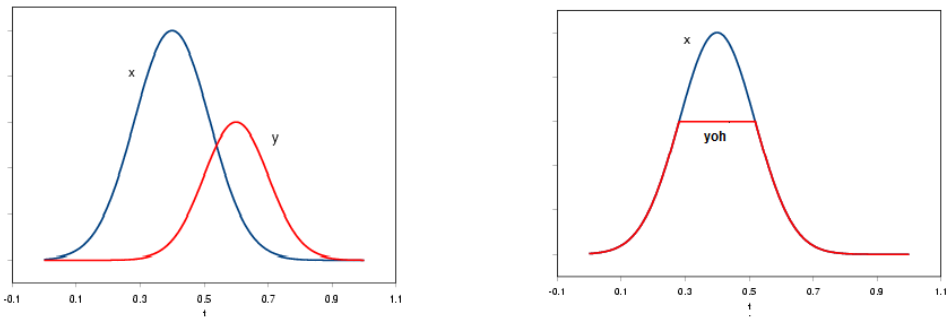

Figure 5: The left panel gives plots of two shifted curves and the right panel displays the registered curve $y \circ h$ in presence of amplitude variation.

tion of suitable warping functions h_i in presence of amplitude variation. By adding an amplitude effect for explicitly taking into account to amplitude variation, any trajectory can be represented in the form $a_i(t)x(t) = y_i \circ h_i(t)$. The unknown amplitude functions $a_i(t)$ should be sufficiently smooth positive functions so that they retain the basic structure of trajectories. Note that estimation of amplitude functions is not of interest but it is just a tool to improve registration procedures as intensity differences may be studied by a FPCA procedure. In parametric methods [29, 1], we consider a parametric family of transformations, the simplest case being the scale-shift registration with constant amplitude, when h_i is restricted to be affine, giving $h_i(t) = \beta_i t + \delta_i$, $\beta_i > 0$ such that $\alpha_i x(t) = y_i(\beta_i t + \delta_i)$, $\alpha_i > 0$. In the literature, a few semi-parametric approaches have been also considered in which functions are obtained from a linear combination of common shape functions by using some parametric transformations, but no particular specification about shape functions is made [19, 10, 16, 18]. In many applications, these methods may appear too restrictive because more complex possibly non-linear time warping functions are necessary. Non-parametric approaches are often more adapted and have received increasing attention in the literature. For example, a monotonization operator can be used to characterize a flexible family of smooth monotone warping functions that can be described by a homogeneous linear differentiable equation [22]. Then warping functions are estimated by minimizing a global fitting criterion in which a penalization term yields both smoothness and monotonicity of warping functions. Note that this latter registration method as well as landmark reg-

istration are implemented in softwares R and Matlab and can be downloaded through the internet site `www.functionaldata.org` [26].

In registration, there is an intrinsic problem of unidentifiability that may lead to badly identify time warping transformations. The challenge in presence of amplitude variation is to correctly separate amplitude and phase variations in estimation procedure. In semi-parametric approaches, a possible solution is to require some normalization conditions on parameters. In a purely non-parametric context, this problem turns out to be insuperably difficult to solve. Indeed, there is no way of correctly identifying warping functions in the sense that we could always find another time transformation $\tilde{h}_i \neq h_i$ which compensates a different amplitude function $\tilde{a}_i \neq a_i$ such that $a_i.x \circ h_i^{-1} = \tilde{a}_i.x \circ \tilde{h}_i^{-1}$ except if amplitude function is assumed to be constant. In this latter case, any trajectory can be represented in the form $\alpha_i.x \circ h_i^{-1}$ for some unique warping function h_i and constant amplitude α_i can be easily identified as being a proportionality factor between the registered trajectory and the target at a certain landmark point. Unfortunatly, even in this case, there is no mean to correctly identify warping functions and amplitude if trajectories are linear. Then, differences between trajectories may be wrongly explained either by amplitude variation or by phase variation.

For illustrating this problem, speed discrepancy related to aircraft type may be adequately handled by affine registration, provided trajectories are full ones (i.e. include synchronization of both departure and arrival point). However, in order to well perform a FPCA procedure, we need to estimate possibly non-linear time warping transformations. Unfortunatly, non-parametric smooth monotone warping methods may not be relevant for aircraft trajectory registration in some cases and may lead to inadequate results for the following reasons. Firstly, as all trajectories have not the same origin-destination pair, the assumption that they are sample paths from a single stochastic process is not satisfied. Then, it may not be relevant to register functions that don't share similar structure (i.e. similar sequence of shape features such as peaks, valleys etc ...). In addition, for segments of flight paths, things are more complicated since the observations may match only a small portion of a larger reference trajectory. Finally, latitude and longitude trajectories are essentially linear when time argument grids at which trajectories are observed are not enough fine to highlight their detailed structures. Only altitude trajectories appear to be piecewise linear and consist of a sequence of flight levels connected by climb or descent phases. Latitude, longitude and altitude exhibit high local variability both in amplitude and in phase that can be difficult to identify because of the problem of unidentifiability due to linearity. Note that the first problem is less crucial that the two latters if the main goal is less to correctly align trajectories than compute distance between trajectories. For instance, we may expect that this problem will not too much disturb a clustering procedure. Indeed, after registration, trajectories sampled from the same stochastic process should be correctly registered while trajectories sampled from different stochastic processes will badly be registered and remain far away. The more problematic unidentifiability problem is illustrated in Figure 6 for both longitude and latitude trajectories. We can clearly see the effect of linearity on registration procedure: the estimated warping functions mainly explain amplitude-part variation. This completely destroys structure of the registered trajectory $(x(t), y(t))$ and will lead to compute a mistaken distance between two trajectories. As an alternative, we might perform a landmark registration method. However, this may be unworkable because of possibly missing marker events in truncated observations. Indeed, it is particularly irrelevant to match the first and last time locations of truncated trajectories to departure and arrival points of a full template.

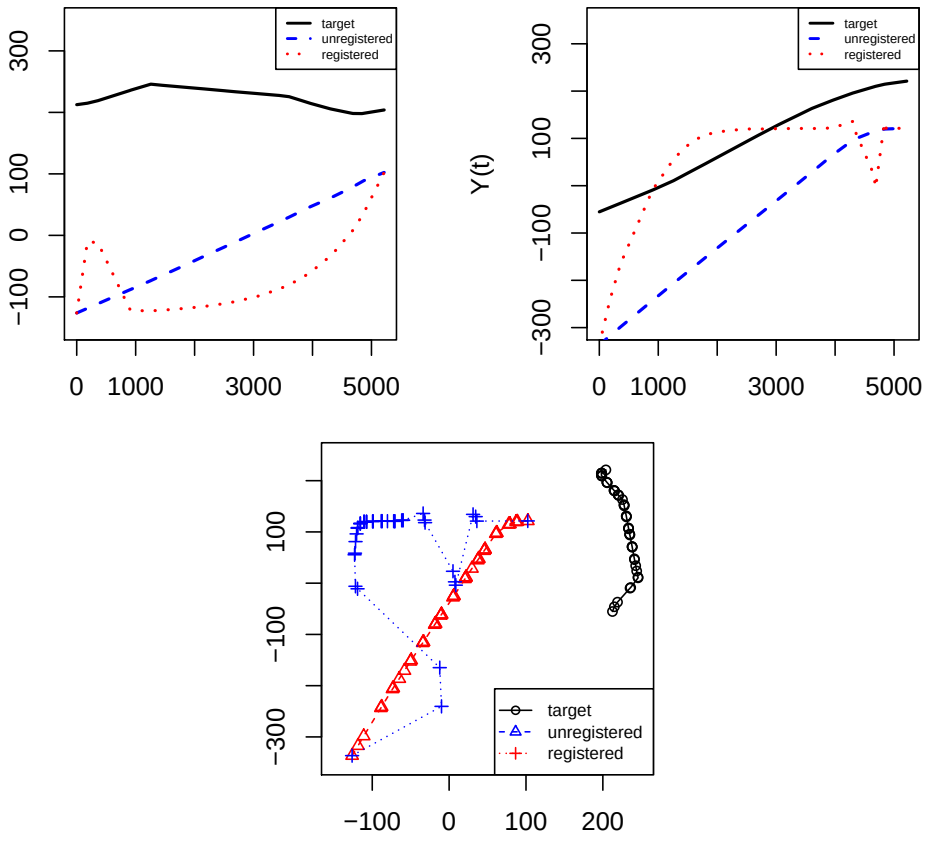

Figure 6: The two upper panels give plots of latitude and longitude trajectories and the bottom panel displays $X - Y$ trajectories.

To avoid this problem as well as the unidentifiability problem, we may perform a purely parametric approach applied to altitude trajectories by simply using the above described scale-shift registration with constant amplitude. Altitude trajectories can be described as piecewise linear functions composed by one single flight level eventually connected by climb or descent phases. We will then estimate warping functions by registering only segments of flight paths that match this flight level rather than full trajectories (work in progress).

To illustrate the influence of phase variation on a FPCA procedure, we generated 200 sinus curves over the interval $[0, 1]$ of the form

$$y_i(t) = 30a_i \sin \pi t, \tag{9}$$

where the coefficients a_i were randomly generated from the gaussian distribution $\mathcal{N}(1; 0.05^2)$. The associated warping functions h_i were $h_i(t) = t^5$. The unregistered curves $y_i(t) = x_i^*(h_i(t))$ are shown in the left panel of Figure 7.

A FPCA procedure applied to the unregistered curves will produce too many principal components than it is needed rather than just only one when curves are synchroniezd. In particular, the second principal component can be interpreted as a purely time shift effect as shown in Figure 8 and is not of interest for the analysis of the variability of the curves. In addition, phase variation influences the shape of the first principal component which is composed of two bumps rather than just one. This component may not be representative of the structure of the curves. Finally, the score histogram of the first component has not a symmetric unimodal distribution. There is a bimodal variation in the data and scores of the two first principal components have a kind of correlation directly related to the bimodal grouping. This bimodal distribution may be problematic if we approximate the score distribution of the first

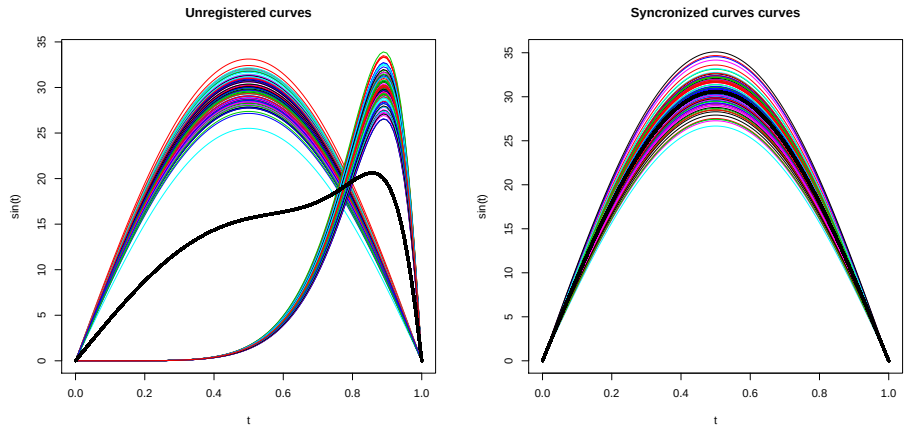

Figure 7: The left panel displays 200 random unregistered curves generated from (9) and warping functions $h_i(t) = t^5$. The right panel gives these curves after registration. The heavy solid lines are the respective pointwise means of these curves.

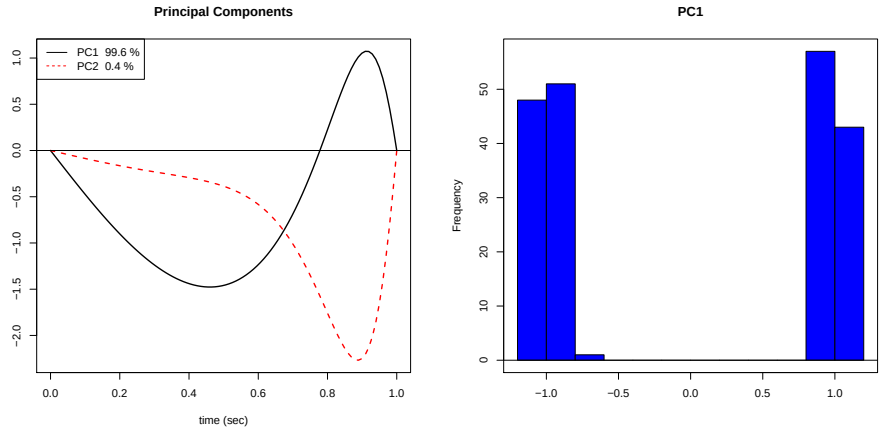

Figure 8: The left panel gives the two first principal components curves and the right panel gives the score histogram on the first pricipal component.

principal component by an unimodal distribution, for instance a gaussian distribution. This problem remains important even if only one outlier curve is out of the group of synchronized curves.

6 Application to Aircraft Trajectories

6.1 The Aircraft Trajectory Dataset

We now apply the previously described FPCA technique to a 1077 aircraft trajectory dataset. These data consist of radar tracks between Paris Orly, Charles de Gaulle (CDG) and Toulouse Blagnac airports recorded during two weeks. Most of the aircrafts are Airbus A319 (25%), A320 (41%) and A321 (24%), followed by Boeing B733 (4%) and B463 (2%) a member of British Aerospace BAe 146 family. Other aircraft types (A318, A333, B738, E120, AT43, AT45 and AT72) account for a smaller amount of aircrafts. Radar measurements are observed in the range of 4-6960 seconds at 4 seconds intervals. As noted in Section 1, the assumption that all trajectories are sample paths from a single stochastic process defined on a time interval is clearly not satisfied in the case of aircrafts: departure times are different, even on the same origin-destination pair and the time to destination is related to the aircraft type and the wind experienced

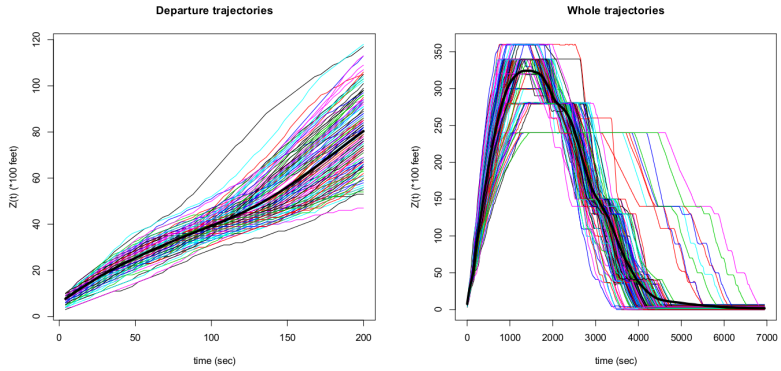

Figure 9: Sample of aircraft trajectories on the range of 4-200 seconds (left panel) and whole trajectories between Toulouse and Paris Charles de Gaulle airports. The heavy solid line is the mean of trajectories.

along the flight. Without loss of generality, we will assign a common starting time 0 to the first radar measurement of the flights. Trajectory altitudes in Figure 9 consist of a sequence of flight levels (FL) measured in hundreds of feet and connected by climb or descent phases. These data exhibit high local variability in amplitude and in phase but our goal is to analyze the amplitude variability by means of a FPCA technique. As observed raw data were passed through pre-processing filters, we get radar measurements at a fine grid of time arguments with few noise. We have then used the discretization method described in Section 2. We will first focus on departure trajectories to avoid the registration problem. Next, we will analyze whole trajectories and compare FPCA results for unregistered and registered trajectories.

6.2 Departure Data

As phase variation may badly influence FPCA, each track was reduced to the range of 4-200 seconds for which phase variations seem negligible. Figure 10 displays the first

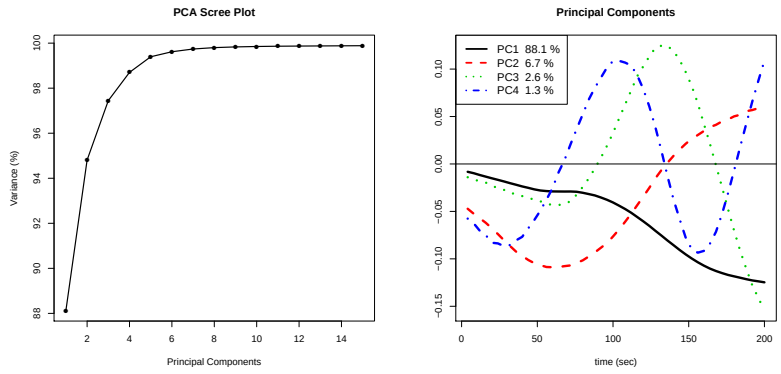

Figure 10: The left panel gives the scree plot of the cumulated variance explained by principal components and the right panel displays the first four principal component curves of aircraft trajectories.

four principal component functions for these track data after the overall mean has been removed from each track. Note that principal component functions are defined only to within a sign change. The percentage 88.1% of total variation explained by the first principal component indicates that this type of variation strongly dominates all other types of variation. The first principal component is a negative function, decreasing with time. It quantifies an overall decrease in altitude that we can call *overall effect* (PC1).

This effect begins to be important around 100 seconds after takeoff and is growing with time. Aircrafts with high negative scores would show especially above-average tracks displaying more important climb rates increasing with time. As the second principal component must be orthogonal to the first one, it will define a less important mode of variation. It accounts for 6.7% of total variation and consists of a high negative contribution for the 0-140 seconds climb phase with minimum at around 60 seconds followed by a much less important positive contribution. As the third and fourth components are required to be orthogonal to the first two components as well as to each other, they account for small proportions of total variation. The third component accounts for only 2.6% of total variation and consists of negative contributions for the two 0-90 seconds and 170-200 seconds phases. The fourth principal component is difficult to interpret and accounts for a very small percent of total variation. Nevertheless, we can see that it looks like the third principal component except for a time shift.

A helpful graphical representation proposed in [25] facilitates the interpretation of each principal component. It consists in visualizing effects of each functional principal component on the overall mean function by adding and substracting a suitable multiple of each principal component. Figure 11 displays the *overall effect* increasing with time

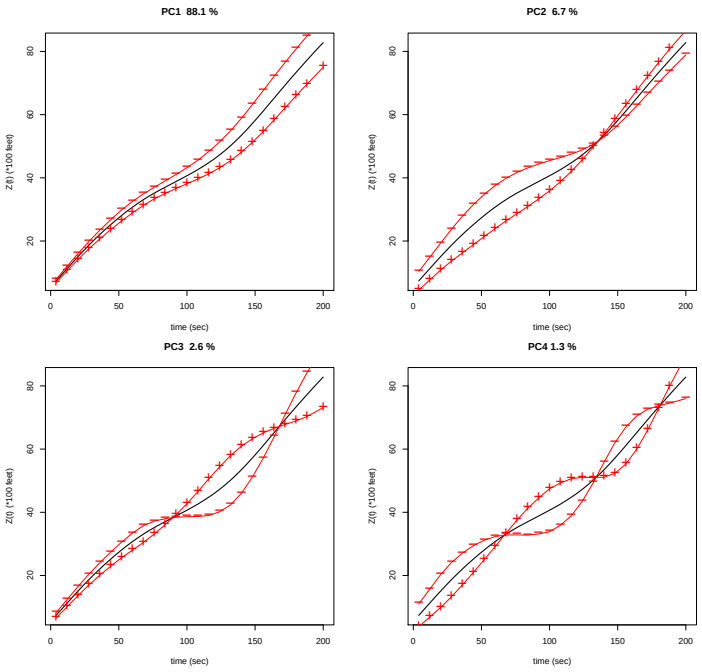

Figure 11: The effects on the mean aircraft trajectory (solid curve) of adding (+) and substracting (-) a multiple of each of the first four functional principal components.

due to the first principal component. The second principal component indicates a mode of variation corresponding to early climb rates. Aircrafts with high negative scores would display higher climb rates up to 140 seconds and later slightly reverting to the mean path. On the other hand, those with high positive scores would display smaller climb rates and trajectories seem to be linear. We call this effect the *takeoff effect* (PC2). We can also easily see the effect of the third component on the overall mean. Aircrafts with high negative scores would display an overall trajectory up to 70 seconds followed by a constant flight level during 60 seconds (4000 feet), later reverting to higher climb rates to compensate it. We call this effect the *first level effect* (PC3). Furthermore, we can visualize the effect due to the fourth principal component that we call *time shift effect* (PC4). High negative scores would display earlier first flight level (3000 feet) at 120 seconds.

Finally, pairwise scatterplots of aircraft scores may reveal patterns of interest and clusters in aircraft trajectories by route and aircraft type. In addition, these plots may also be used to detect outliers. For simplifying scatterplots, FPCA was applied to a 145 aircraft trajectory dataset between Toulouse Blagnac and Paris Charles de Gaulle airports and we have grouped together AT43, AT45, AT72 and E120 aircraft types, now labeled AT type. We have found similar components to those observed previously. The scatterplot in the left panel of Figure 12 displays aircraft scores by aircraft type of the *overall effect* (PC1) against the *takeoff effect* (PC2). Clearly, the first component divides aircraft trajectories in two groups: AT, B463 and most of A321 with positive PC1 scores (under-average trajectories with overall lower climb rates) and A319, B733 with negative scores (above-average trajectories with overall higher climb rates). The second component corresponding to the *takeoff effect* (PC2) divides trajectories in a different manner: AT, B463 and A319 with positive scores (slower takeoff) and A320, A321 with negative scores (faster takeoff). Then, we can

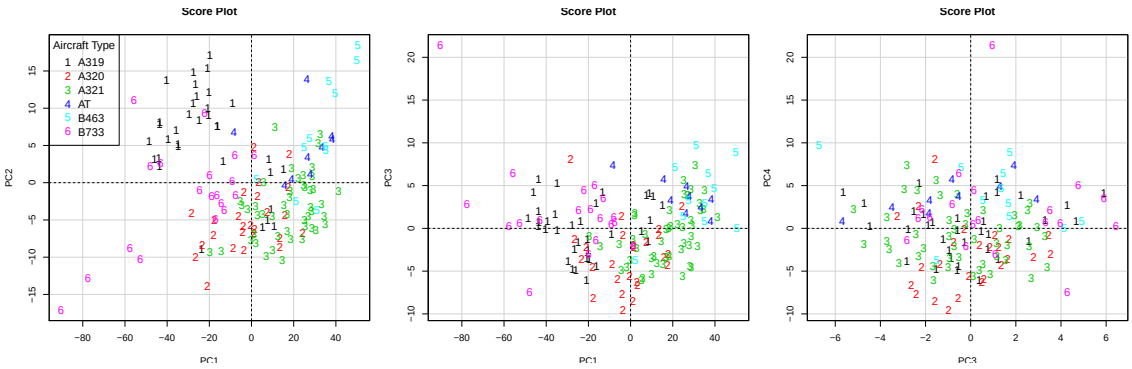

Figure 12: Scatterplots of the individual trajectory scores by aircraft type.

see that smallest aircraft types such as AT and E120 should display overall lower climb rates associated with slower takeoff. In addition, A319 and A321 aircrafts trajectories are completely different: A319 aircrafts have negative PC1 scores (overall higher climb rates) associated with positive PC2 scores (slow takeoff) while A321 aircrafts have positive PC1 scores (overall lower climb rates) associated with negative PC3 scores (fast takeoff).

The second scatterplot in the middle panel of Figure 12 shows aircraft scores by aircraft type of the *overall effect* (PC1) against the *first level effect* (PC3). Firstly, we clearly detect one outlier with very high negative PC1 score and very high positive PC3 score due to a B733 aircraft. This aircraft displays a very atypical trajectory with a global high climb rate and no first level effect to compensate it. Moreover, the third component divides trajectories in two groups: AT, B463, B733 with positive scores (no first level effect) and A320, most of A321 with negative scores associated with a first level effect.

The third scatterplot in the right panel of Figure 12 gives aircraft scores of the *first level effect* (PC3) against the *time shift effect* (PC4). The fourth component divides trajectories in two groups: AT, B733 with positive scores (later first level) and A320 with negative scores (ealier first level).We can find again the same outlier than the previous one with one very high positive PC4 score. This B733 aircraft has an overall above-average trajectory with fast takeoff and no early first level effect to compensate it. We can easily summaryze the previous results in Table 1. Note that phase variation may produce too many components and trajectories may be characterized by only three principal components rather than four components. To improve results, phase variation should be removed by using a registration procedure.

Table 1: Individual scores by aircraft type

Aircraft type	PC1	PC2	PC3	PC4	Outlier
AT, E120, B463	+	+	+	+	
A320	0	-	-	-	
B733	-	-	+	+	*
A319	-	+	0	0	
A321	+	-	-	0	

6.3 Whole Trajectory Data

We now consider whole trajectories between Toulouse and Paris Charles de Gaulle airports and compare FPCA results obtained from unregistered and registered trajectories. We can see in Figure 13 that unregistered trajectories exhibit high phase

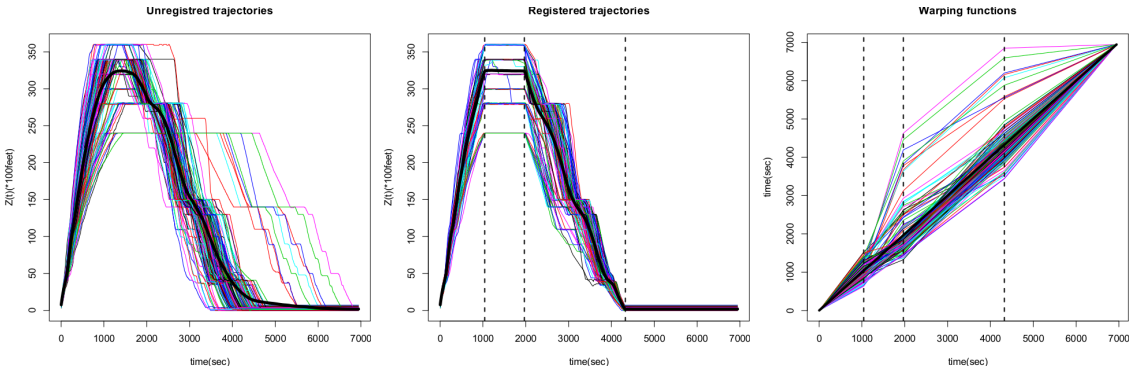

Figure 13: Whole trajectories between Toulouse and Paris CDG airports: unregistered (left panel) and registered (middle panel). The heavy solid line is the mean of trajectories. The right panel displays warping functions estimated by landmark registration.

variation. These differences in dynamics may disturb the sample mean function and consequently a FPCA procedure. Altitude trajectories can be described as piecewise linear functions composed by one maximum flight level connected by climb or descent phases. The registered mean function is more representative of such structure of trajectories. Trajectories have been registered by using a landmark registration procedure with three markers: the time to destination and the two time locations of segments that match the maximum flight level.

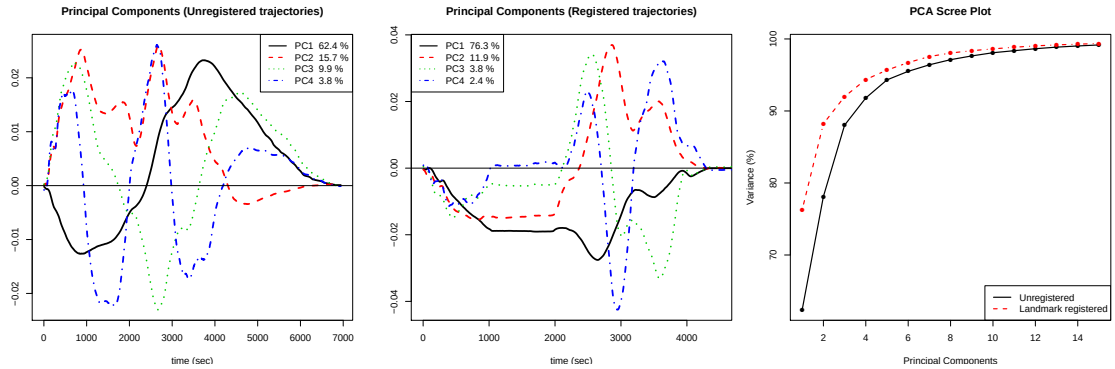

Figure 14: The first four principal components of aircraft trajectories: unregistred in the left panel and registered in the middle panel. The right panel displays the scree plots of cumulated variance.

Figure 14 displays the first four principal components for unregistered trajectories in the left panel and registered trajectories in the middle panel. For unregistered trajectories, the fourth principal component looks like the third one except for a time shift. The first principal component consists of a negative contribution for the 0-2500 seconds phase followed by an important positive contribution.

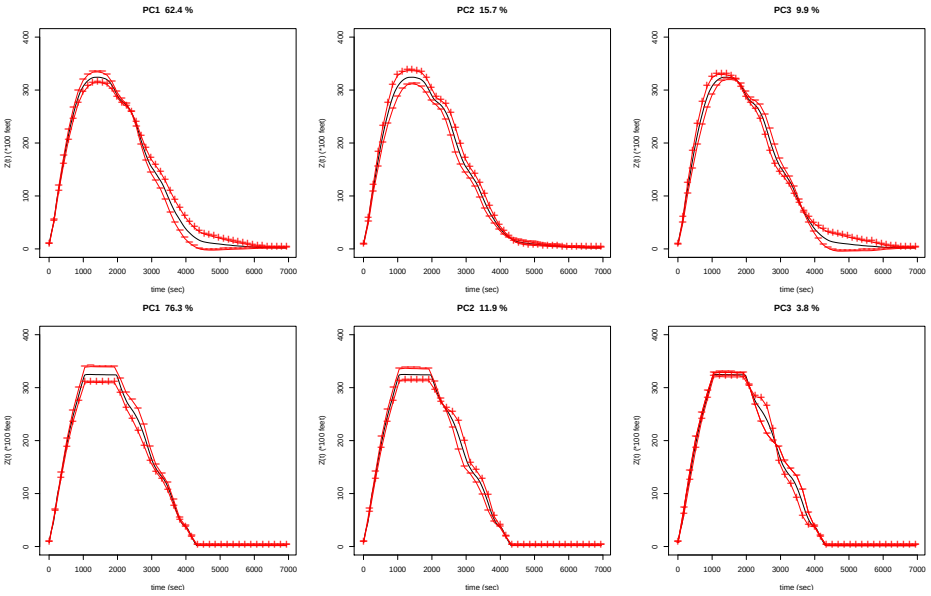

Figure 15: The effects on the mean aircraft trajectory (solid curve) of adding (+) and substracting (-) a multiple of each of the first four principal components for unregistered trajectories in the three top panels and registered trajectories in the three bottom panels.

Figure 15 displays the effects of this principal component on the mean function. We clearly visualize that this effect corresponds to an increase in the differences between the maximum level and the descent phase: trajectories with higher flight level have a faster descent phase and trajectories with lower flight level have a slower descent phase. The second principal component corresponds to an overall increase in altitudes and the third component displays a time shift in the arrival phase and in the maximum flight level followed by two different descent phases. Phase variation has probably disturbed the estimation of the principal components because amplitude and phase variation are mixed.

For registred trajectories, we can see in Figure 14 and Figure 15 that the first principal component now corresponds to an overall increase in altitude and now accounts for the main percentage of total variation with 76.3% rather than 15.7%. The second principal component displays the differences between the maximum flight level and the descent phase with a less important variation (11.9% of total variation rather than 62.4%). The time shift effect is removed from the third component which corresponds to the two different descent phases and represents only 3.8% of total variation rather than 9.9%. These differences are due to phase variation that are mixed to amplitude variation when trajectories are shifted. Finally, principal components of registered trajectories capture a more important proportion of total variation than principal components of unregistered trajectories. We only need three components to capture 92% of total variation instead of four principal components in the case of unregistered trajectories. Then, a preliminary registration procedure leads to reduce the number of principal components. Modes of variation are more representative and will better explain the main directions in which aircraft trajectories vary.

7 Conclusion

Functional Data Analysis (FDA) consists in studying a sample of random trajectories called *functional data*, generated from an underlying stochastic process. Functional Principal Component Analysis (FPCA) generalizes the standard multivariate PCA to the infinite-dimensional case by analyzing the covariance structure of functional data. It can be seen from two different points of view, these two approaches being connected by the Karhunen-Loève decomposition: in a non-parametric point of view, variability of functional data is characterized by spectral decomposition of the sample covariance operator and in a semi-parametric model, the individual trajectories are characterized through a linear combination of empirical basis functions called *functional principal components* with random coefficients called *principal component scores*. These components are estimated by the eigenfunctions corresponding to the largest eigenvalues of the sample covariance operator and they can be interpreted as modes of variation.

FPCA has many advantages. By characterizing individual trajectories through an empirical Karhunen-Loève decomposition, FPCA can be used as a dimension reduction technique. Moreover, rather than studying infinite-dimensional functional data, we can focus on a finite-dimensional vector of random scores that can be used into further statistical analysis such as cluster analysis. In addition, the estimated coefficients are uncorrelated and may be more convenient for subsequent applications. Finally, FPCA may be better than alternative representation of functional data by fixed basis functions such as Fourier series, wavelets or B-splines that may require a larger number of fixed basis functions to correctly represent a given sample of trajectories. This idea is used in principal component regression in which the regression function is expanding in the basis of the empirical eigenfunctions.

FPCA is a powerful tool to analyze and visualize the main directions in which trajectories vary. As in the multivariate case, pairwise scatterplots of scores may reveal patterns of interest, clusters in the data and atypical trajectories. We have successfully applied this technique to analyze aircraft trajectories and it can be easily extended to the three dimensional case. Moreover, this technique may be useful to generate aircraft trajectories by sampling the principal component scores. However, a FPCA procedure should not be directly applied to whole trajectories because phase and amplitude variations may be mixed. This registration problem remains crucial because the assumption that all trajectories are sample paths from a single stochastic process is not satisfied and may be complex in the case of three dimensional aircraft trajectories. For this reason, a preliminary step may consist of the registration of trajectories by suitable transformations of time called *warping functions* before proceeding to sophisticated statistical analysis such as FPCA or clustering. This problem is difficult to solve particularly when trajectories are linear. Rather than register trajectories in a first step, an alternative consists in combining registration with FPCA as is done in [18, 29]. A first approach of this problem was proposed in [29]: phase variation is assumed to consist of individual time shifts $h(t) = t + \tau$. More recently, in [18], warping functions are represented as linear combinations of known basis functions such as spline basis and are incorporated into a FPCA procedure. As a generalization, a more complex model consisting of a double FPCA approach will be studied in which warping functions are modeled as a linear combination of unknown basis functions (work in progress). This approach can be useful to generate realistic sample of aircraft trajectories taking into account possible phase variation. Another future work consists in using a Common FPCA procedure that can be useful to provide common components explaining the structure of trajectories sampled from different routes.

References

[1] Ashburner J., Neelin P., Evans D.I. and Friston K., Incorporating prior knowledge into image registration, *NeuroImage*,6(4),344-352, 1997.

[2] Besse P., Ramsay J.O., Principal component analysis of sampled curves, *Psychometrika*, 51, 285–311, 1986.

[3] Besse P., PCA stability and choice of dimensionality, *Statistics and Probability Letters*, 13, 405–410, 1992.

[4] Bookstein F.L., *Morphometric tools for landmark data:geometry and biology*, Cambridge: Cambridge University Press, 1991.

[5] Conway J. B., *A Course in Functional Analysis*, New York: Springer-Verlag, 1985.

[6] Dauxois J., Pousse A., *Les analyses factorielles en calcul des Probabilités et en Statistique: Essai d'étude synthétique*, PhD thesis, University of Toulouse, 1976.

[7] Dauxois J., Pousse A., and Romain Y., Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *Journal of Multivariate Analysis*, 12, 136–154, 1982.

[8] Deville J., Méthodes statistiques et numériques de l'analyse harmonique, *Ann. Insee*, 15, 1974.

[9] Ferraty F. and Vieu P. *Nonparametric functional data analysis*, New York: Springer-Verlag, 2006.

[10] Gasser Th. and Kneip A., Convergence and consistency results to self-modeling nonlinear regression, *Annals of Statistics*, 16, 82-112, 1988.

[11] Gasser Th. and Kneip A., Searching for structure in curve samples, *Journal of the Amererican Statistical Association*, 90, 1179-1188, 1995.

[12] Härdle W. and Marron J., Semi-parametric comparison of regression curves, *Annals of Statistics*, 18, 63-89, 1990.

[13] Hötelling H., Analysis of a complex of statistical variables into principal components. *J*ournal of Educational Psychology, 24 (6-7), 417-441 and 498-520, 1933.

[14] Kneip A. and Gasser Th., Statistical tools to analyze data representing a sample of curves, *Annals of Statistics*, 20, 1266-1305, 1992.

[15] Kneip A., Nonparametric estimation of common regressors for similar curve data. *Annals of Statistics*, 22, 1386-1428, 1994.

[16] Kneip A. and Engel J., Model estimation in nonlinear regression under shape invariance, *Annals of Statistics*, 23, 551-570, 1995.

[17] Kneip A. and Utikal K., Inference for density families using functional principal component analysis. *Journal of the American Statistical Association*, 96, 519-532, 2001.

[18] Kneip A. and Ramsey J.O., Combining registration and fitting for functional models, *Journal of Amer. Statist. Assoc.*, 103, 483-484, 2008.

[19] Lawton W.H. et al, Self-modeling nonlinear regression, *Technometrica*, 14, 513-532, 1972.

[20] Pearson K., On lines and planes of closest fit to systems of points in space. *The London, Edinburgh and Dublin Philosophical Magazine and Journal in Science*, 2, 559-572, 1901.

[21] Rao C. R., Some statistical methods for the comparison of growth curves. *Biometrics*, 14, 1–17, 1958.

[22] Ramsay J.O., Estimating smooth monotone functions, *Journal of the Royal Statistical Society, Series B*, 60, 365-375, 1996.

[23] Ramsay J.O. and Li X., Curve registration, *Journal of the Royal Statistical Society, Series B*, 60, 351-363, 1998.

[24] Ramsay J.O and Silverman B.W., *Applied Functional Data Analysis*, New York: Springer, 2002.

[25] Ramsay J.O. and Silverman B.W., *Functional data analysis*, New York: Springer-Verlag, 2005.

[26] Ramsay J.O., Hooker G. and Graves S., *Functional data analysis with R and Matlab*, New York: Springer-Verlag, 2009.

[27] Rice J. and Silverman B., Estimating the mean and covariance structure nonparametrically when the data are curves. *Journal of the Royal Statistical Society, Series B*, 53, 233–243, 1991.

[28] Sakoe H. and Chiba S., Dynamic programming algorithm optimization for spoken word recognition, *IEEE Trans. on Acoustics Signal and Speech Process.*,26, 43-49, 1978.

[29] Silverman B.W., Incorporating parametric effects into functional principal components analysis, *Journal of the Royal Statistical Society, Series B*, 57, 673-689, 1995.

[30] Silverman B.W., Smoothed functional principal components analysis by choice of norm. *Annals of Statistics*, 24, 1–24, 1996.

[31] Tucker L.R., Determination of parameters of a functional relation by factor analysis. *Psychometrika*, 23, 19-23,1958.

École Nationale de l'Aviation
7, avenue Édouard Belin
BP 54005
31055 Toulouse cedex 4
Tèl. +33 (0)5 62 17 40 00
Fax +33 (0)5 62 17 40 23

ENAC

La référence aéronautique

www.enac.fr