Functional principal component analysis of aircraft trajectories

Florence Nicol

Functional Principal Component Analysis of Aircraft Trajectories

Florence NICOL [a,1]
[a] *ENAC-MAIAA, Toulouse, FRANCE*

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 method is applied to aircraft trajectories and the problem of registration is discussed when phase and amplitude variations are mixed.

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

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. 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, the general framework for Functional Data Analysis (FDA) is presented and the FPCA approach is formalized in Section 2. In Section 3, 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. 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

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

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 [7] as follows.

Definition 1.1 *A random variable $X = \{X(t), t \in J\}$ is called functional variable (f.v.) if it takes values in an infinite dimensional space (a functional space $\mathscr{H}$). An observation x of X is called a functional datum.*

Definition 1.2 *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 $\mathscr{H}$ 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 $\|x\|^2 = \langle x,x\rangle$. 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 [7].

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

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

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 functional data analysis, particularly in Functional Principal Component Analysis, as will be seen in the next section.

Definition 1.3 *The covariance operator $\Gamma : \mathscr{H} \longrightarrow \mathscr{H}$ is defined by*

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

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 [4].

Note that X is a functional space $\mathscr{H}$-valued random function and its observation x is a non-random function of $\mathscr{H}$. 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, 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 [14] 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 error. 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 trajectories are

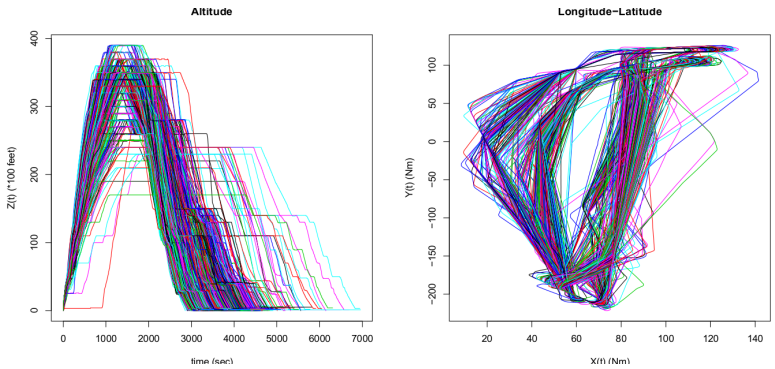

Figure 1. Sample of aircraft trajectories (Paris-Toulouse).

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 1. 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.

2. 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 [12,9]. 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 [1,13,19]. 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 functional data in a understandable manner. Functional Principal Component Analysis (FPCA) is a 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 random score vectors, 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.

2.1. 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 $\mathscr{H}$.

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 random function of $\mathscr{H}$. 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

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

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 than 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 $\theta_i = \langle\gamma_i, X\rangle = \int_J \gamma_i(t)X(t)dt$ are called *principal component scores* of X into the γ_i-direction [14].

2.2. Estimation

When the covariance function is unknown, we can replace it by its sample version

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

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

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

The estimators $\widehat{\gamma}_1,\ldots,\widehat{\gamma}_n$ are then 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 $\theta_{ij} = \langle\widehat{\gamma}_i, X_j\rangle$ into the γ_i-direction, $j = 1,\ldots,n$, are centered and uncorrelated random variables such that $\frac{1}{n}\sum_{j=1}^{n}\theta_{ij}^2 = \widehat{\lambda}_i$. Dauxois et al. [5] showed consistency and asymptotic properties of $\widehat{\Gamma}_n$, $\widehat{\gamma}_i$ and $\widehat{\lambda}_i$ under mild assumptions.

Several estimation methods of scores and principal component functions were developed for FPCA and asymptotic results was studied in [5]. Rao [13] and Tucker [19] first introduced the earliest approach of PCA that linked factor analysis methods with growth curve models. This method is based on numerical integration or quadrature rules when functional data are 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. Assume that func-

tions are observed (or estimated by using interpolation or smoothing techniques) at the same time arguments, no necessarily equally spaced. This yields an $n \times N$ data matrix

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

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*. Numerical quadrature schemes can be used to involve a discrete approximation of the functional eigenequation Eq. (2)

$$\Sigma_n W \widetilde{\gamma}_m = \widetilde{\lambda}_m \widetilde{\gamma}_m, \tag{3}$$

where $\Sigma_n = (\widehat{\sigma}_n(t_i,t_j))_{i,j=1,\dots,N}$ 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 $\widetilde{\gamma}_m = (\widetilde{\gamma}_m(t_1),\dots,\widetilde{\gamma}_m(t_N))$ are the eigenvectors associated with the eigenvalues $\widetilde{\lambda}_m$ of the matrix $\Sigma_n W$. The orthonormality constraints are now

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

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. 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. In Eckstein [6], a standard PCA procedure was applied on a discretized ground speed dataset. Data are not considered as functions but the dataset is defined by a serie of variables that are the ground speed at each given time. This dataset consists of 180 ground speed measurements at 1 second intervals for a large number of departures. In this particular case, the solutions correspond with those obtained by a FPCA procedure because the weight matrix W is the identity matrix. Otherwise an orthonormalization correction is needed using Gramm-Schmidt procedure.

A more sophisticated method is based on expansion of functional data on known basis functions such as a Fourier basis or spline functions. Functional data are estimated by their projections $\widetilde{x}_i$ onto a linear functional space spanned by K known basis functions.

This method will better take into account the functional nature of the data and implies to reduce the eigenequation to discrete or matrix form. Furthermore, in many applications, functional data are assumed to be smooth, and yet, the estimated principal components functions may be rough and present important variability because of sampling error, observation noise and the choice of basis functions. Rather than first smoothing functional data before proceeding with FPCA [14], it makes sense to incorporate this smoothness assumption into the estimation procedure [16,18]. The smoothed FPCA approaches, also called *regularized* FPCA in [14], are based on the well-known roughness penalty approaches. Detailed algorithms of these methods are available in [14].

2.3. 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 $\mathscr{H}$ 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 $\mathrm{Var}(\theta_i) = \lambda_i \geq 0$. Another important property for FPCA involves the best L-term approximation property.

Proposition 2.1 *For any further orthogonal basis $\psi_1, \psi_2, \ldots$ of $\mathscr{H}$ 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 components. Then, the maximization problem in Eq. (1) is equivalent to the minimization problem of the mean integrated square error

$$\min_{\gamma_1, \ldots, \gamma_L} \mathbf{E}\left[\|X - \sum_{i=1}^{L} \theta_i \gamma_i\|^2 \right] \tag{4}$$

that is solved by the first L eigenfunctions $\gamma_1, \gamma_2, \ldots, \gamma_L$ of the covariance operator Γ ordered by the corresponding eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_L$. The two approaches are then connected by the Karhunen-Loève decomposition for which the integrated mean squared error in Eq. (4) is minimum if $\gamma_1, \ldots, \gamma_L$ are the first L eigenfunctions of Γ and $\theta_i = \langle \gamma_i, X \rangle$. Because each functional variable X_j admits the Karhunen-Loève decomposition,

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

we can interpret the random scores $\theta_{ij} = \langle \widehat{\gamma}_i, X_j \rangle$ as proportionality factors that represent strengths of the representation of each individual trajectory by the ith principal compo-

nent 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 directions 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 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 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

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

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. Determining a reasonable number L of components is often a crucial issue in functional analysis. Indeed, 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 [11] and [2].

3. THE REGISTRATION PROBLEM

The process of registration, well known in the field of functional data analysis [17,8, 14], 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 2. In addition, more complex analysis such as 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.

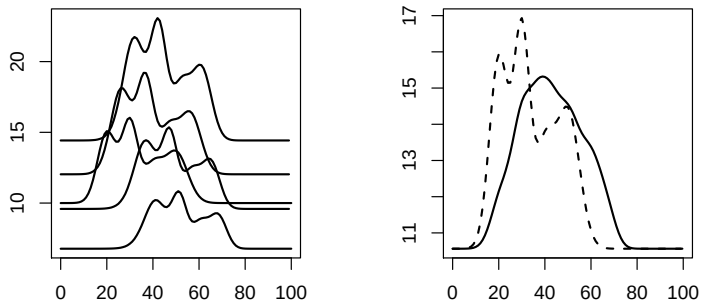

Figure 2. The left panel gives plot of simulated sample functions and the right panel displays mean functions of unregistered curves (solid line) and registered curves (dashed line).

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 [3,10,8]. 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.

4. APPLICATION TO AIRCRAFT TRAJECTORIES

4.1. 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

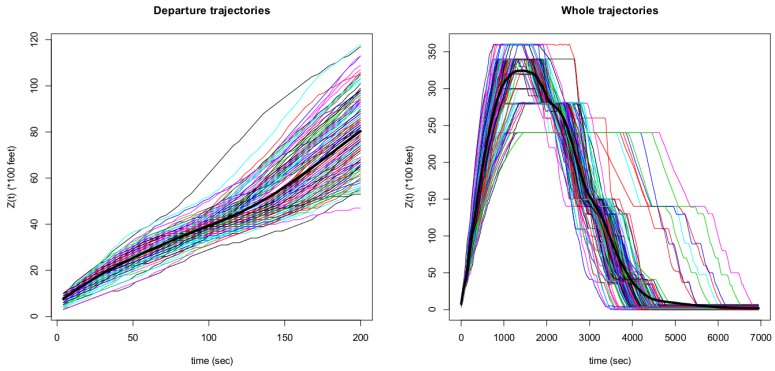

Figure 3. Sample of aircraft trajectories on the range of 4-200 seconds (left panel) and whole trajectories between Toulouse and Paris Charles de Gaulle airports. The heavy solid line is the mean of trajectories.

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 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 3 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.

4.2. 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 4 displays the first 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

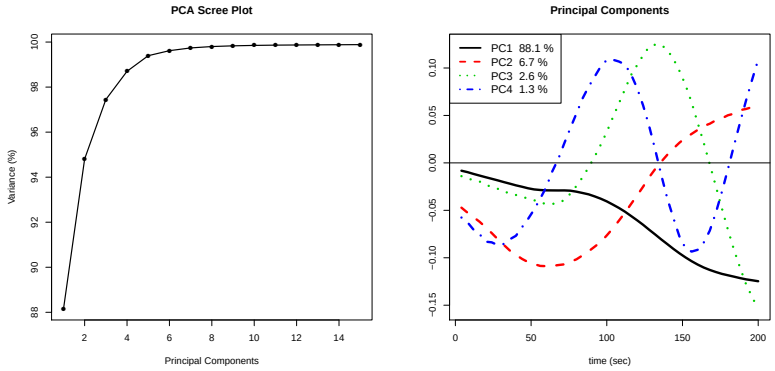

Figure 4. The left panel gives the scree plot of the cumulated variance explained by principal components and the right panel displays the first four principal component curves of aircraft trajectories.

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 [14] 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 5 displays the *overall effect* increasing with time 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 6 displays aircraft scores by aircraft type of the *overall*

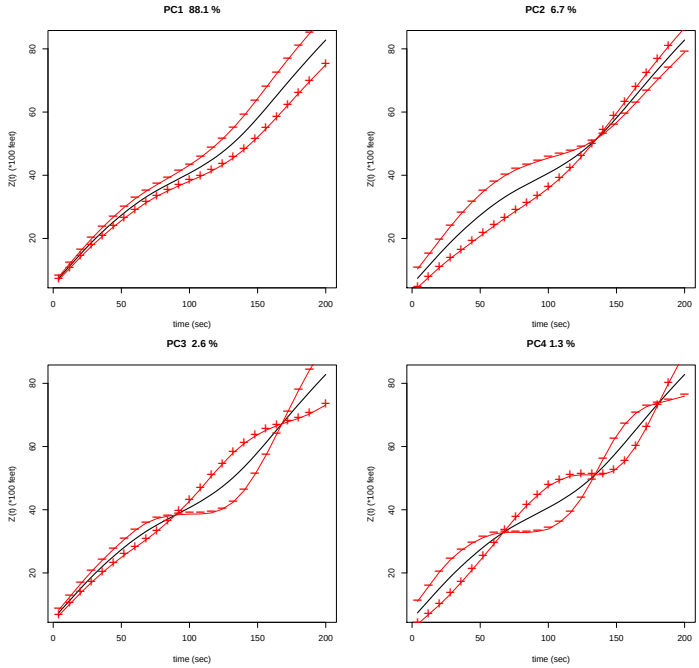

Figure 5. The effects on the mean aircraft trajectory (solid curve) of adding (+) and substracting (-) a multiple of each of the first four functional principal components.

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 see that smallest aircraft types such as AT and E120 should

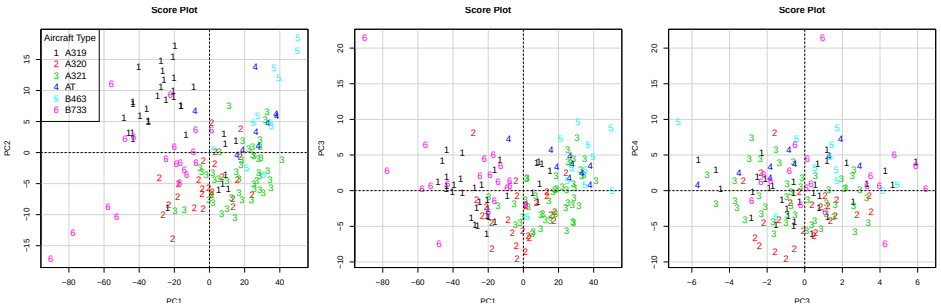

Figure 6. Scatterplots of the individual trajectory scores by aircraft type.

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 6 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 6 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 pro-

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	

duce 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.

4.3. 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 7 that unregistered trajectories exhibit high phase variation. These

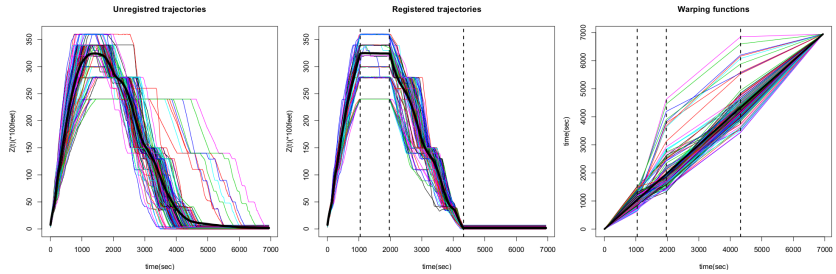

Figure 7. Whole trajectories between Toulouse and Paris Charles de Gaulle airports: unregistered in the left panel and registered in the middle panel. The heavy solid line is the mean of trajectories. The right panel displays warping functions estimated by landmark registration.

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.

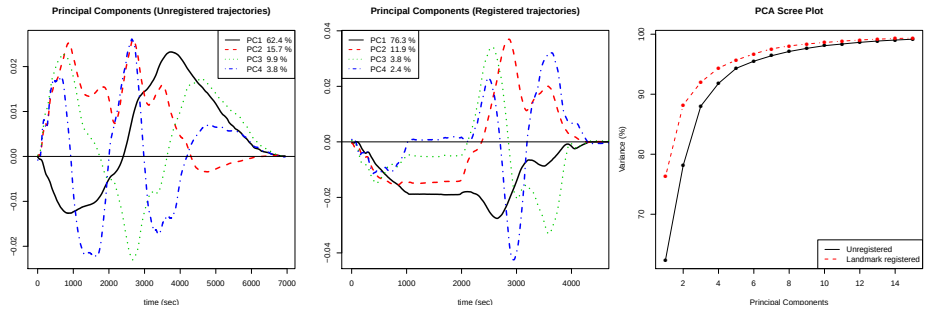

Figure 8. The first four principal components of aircraft trajectories: unregistred in the left panel and registered in the middle panel. The right panel displays the scree plots of cumulated variance.

Figure 8 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 9 displays the effects of this prin-

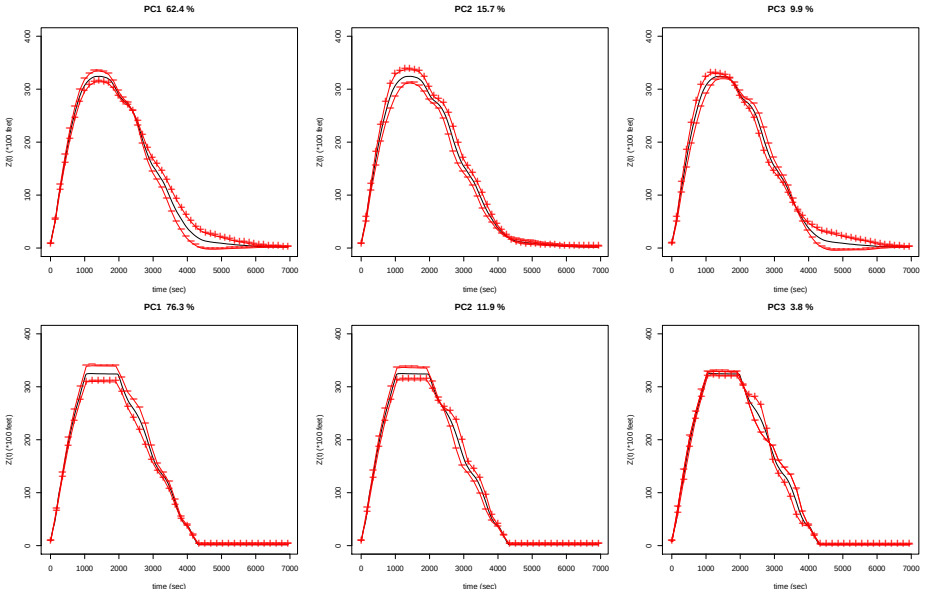

Figure 9. The effects on the mean aircraft trajectory (solid curve) of adding (+) and substracting (-) a multiple of each of the first four principal components for unregistered trajectories in the three top panels and registered trajectories in the three bottom panels.

cipal 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 8 and Figure 9 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.

Conclusion

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.

References

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

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

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

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

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

[6] A. Eckstein, Data driven modeling for the simulation of converging runway operations. *4th International Conference on Research in Air Transportation*, Budapest, 1982.

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

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

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

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

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

[12] K. Pearson, 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** (1901), 559-572.

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

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

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

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

[17] 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.

[18] B.W. Silverman, Smoothed functional principal components analysis by choice of norm. *Annals of Statistics*, **24** (1996), 1-24.

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