Statistical Analysis of Aircraft Trajectories: a Functional Data Analysis Approach

Florence NICOL

ENAC-DEVI

ALLDATA 2017
April 26, 2017

Outline

Outline

What is the best representation?

An old problem

Figure: Fragment from the Tomb of Nebamun, Thebes, *British Museum*.

What is the best representation?

An old problem

Figure: Mug shot of American gangster Al Capone, 1931.

What is the best representation?

An old problem

Figure: Marie-Thérèse portrait, Picasso.

Main ideas

Context

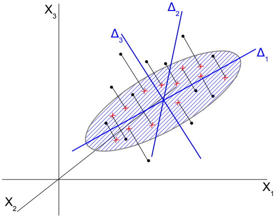
- a large number of numeric variables possibly correlated,
- analyze data variability by studying the covariance structure.

What do we want to do?

- create a small number of new descriptors,
- capture the maximum amount of variation in the data.

How can we do that?

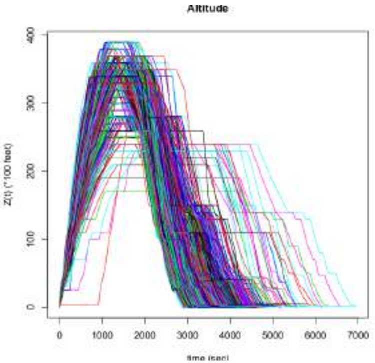
- Looking for new orthogonal directions such that the variance of the projected data is maximal.



Outline

What are Functional Data?

Rather than a sample of points x_i, $i = 1, \ldots, n$, we observe a sample of entangled curves $x_i(t)$, images or functions.



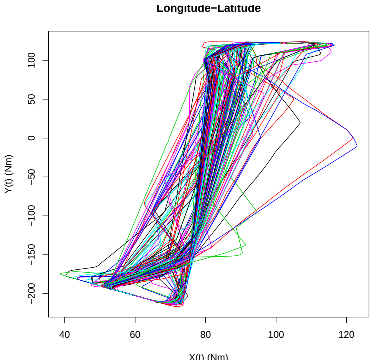


Figure: Sample of aircraft trajectories (Paris-Toulouse).

What are functional data?

Functional variable (f.v.)

$X = \{X(t), t \in J\}$ is a functional space $\mathcal{H}$-valued random variable

- a continuous stochastic process on a compact interval J,
- $\mathcal{H}$ is the separable Hilbert space $L^2(J)$.

Observed data

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

Discretized observed data

- Functional data $x_i(t)$ are observed discretely

$$x_i(t_{ij}), i = 1, \ldots, n, j = 1, \ldots, N_i.$$

- Often given at the same time arguments $t_1, \ldots, t_N$.

Functional Data Analysis

Inference about functional data

Multivariate statistical techniques are inadequate !

- They don't take into account to the functional nature of data.
- A hard drawback: the curse of dimensionality $N \gg n$.

Extend multivariate methods to the functional case

- Functional principal component analysis (FPCA), Clustering.
- Functional linear models, Functional analysis of variance... etc

Generalization is not trivial!

- Data live in infinite dimensional spaces.
- Two types of errors:
  - sampling error in random functions drawn from an underlying process,
  - measurement error when functions are discrete noisy sample paths.

Outline

Generalization to the functional case

Multivariate PCA	**Functional PCA**
individual vector $x_i \in \mathbb{R}^p$	function $x_i(t) \in \mathcal{H}$
eigenvector u	eigenfunction γ
mean vector $\overline{x} \in \mathbb{R}^p$	mean function $\mu(t) \in \mathcal{H}$
covariance matrix	covariance operator
inner product $$\langle u, x_i \rangle = u^T x_i$$	inner product $$\langle \gamma, x_i \rangle = \int_J \gamma(t) x_i(t) dt$$

Generalization to the functional case

Maximization of variance

The weight function γ_1 maximizes the variance of the projected data

$$\gamma_1 = \underset{\|\gamma_1\|=1}{\operatorname{argmax}} \operatorname{Var}\left(\langle \gamma_1, X \rangle \right).$$

The subsequent weight functions γ_k can be found analogously subject to the additional constraint (orthogonality)

$$\langle \gamma_i, \gamma_k \rangle = \int_J \gamma_i(t)\gamma_k(t) = 0,\ i < k.$$

- $\gamma_1, \gamma_2, \ldots$ are called *functional principal components*,
- orthogonality constraints ensure that γ_i indicates something new,
- the amount of variation $\lambda_i = \operatorname{Var}\left(\langle \gamma_i, X \rangle \right)$ will decline stepwise.

Estimation

Let $X_1, \ldots, X_n$ be a sample of independent functional variables.

Karhunen-Loève representation

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

Interpretation

- Principal components $\widehat{\gamma}_i$ are modes of variation of individual trajectories.
- Random scores $A_{ij} = \langle \widehat{\gamma}_i, X_j \rangle$ are proportionality factors: measure the influence of the principal component $\widehat{\gamma}_i$ on the shape of X_j.

Estimation

Reduction dimension tool: a small number $L \ll n$ is needed

$$X_j(t) \simeq \sum_{i=1}^{L} A_{ij}\widehat{\gamma}_i(t) + \mu(t).$$

▶ A small number L of components is often sufficient to account for a large part of variation.
▶ High values of L are associated with high frequency components which represent the sampling noise.

Quality of representation: % of total variation (Scree Plot)

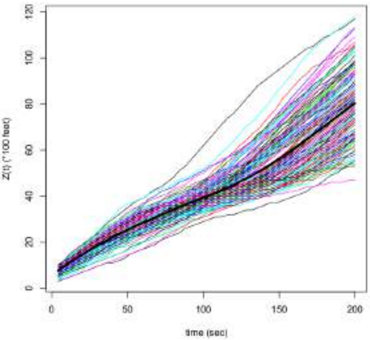
$$\tau_i = \frac{\widehat{\lambda}_i}{\sum_{i=1}^{n}\widehat{\lambda}_i}, \qquad \tau_L^C = \frac{\sum_{k=1}^{L}\widehat{\lambda}_k}{\sum_{i=1}^{n}\widehat{\lambda}_i}.$$

Outline

Univariate FPCA: Flight Level

Route: Paris Orly airport ⟶ Toulouse airport

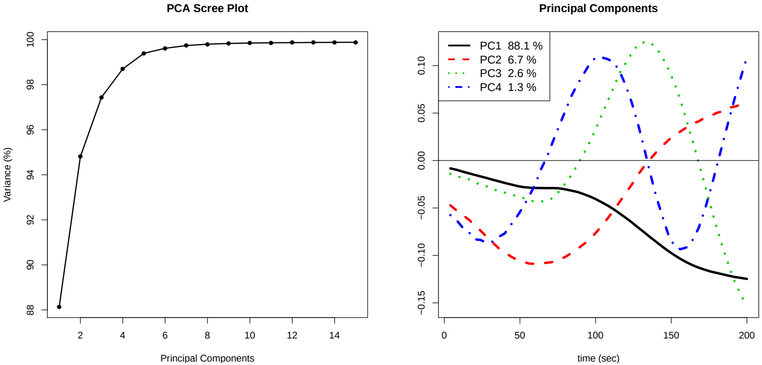
Flight level



- Aircraft type: A319(25%), A320(41%), A321(24%), B733 (4%), B463 (2%) AT type (4%).
- Aircraft trajectories measured at 4 seconds intervals.

Univariate FPCA: Flight Level

Effects on the mean trajectory of adding (+) or substracting (-) PC



- 4 components = $98,7\%$ of total variation.

Univariate FPCA: Flight Level

Effects on the mean trajectory of adding (+) or substracting (-) PC

Overall effect

88.1%

Takeoff effect

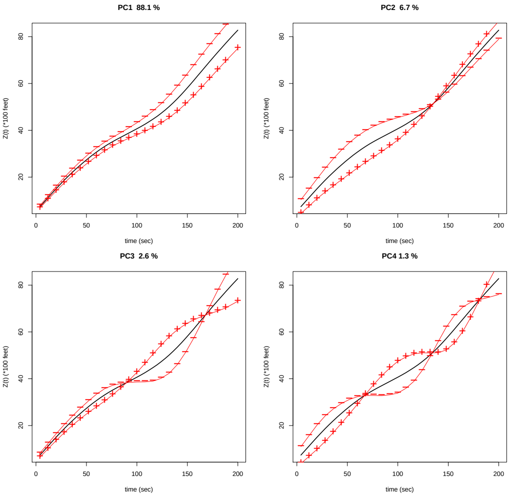
6.7%

First step effect

2.6%

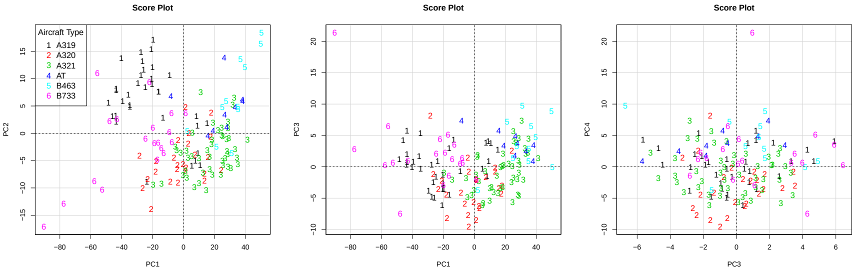
Time shift effect

1.3%



Univariate FPCA: Flight Level

Score scatterplots by aircraft types



- detect outliers and clusters in the data,
- interpret clusters,
- explain individual behaviour relatively to modes of variation.

Univariate FPCA: Flight Level

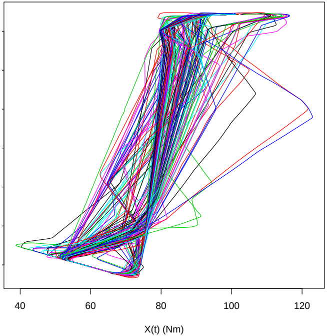
Table: Individual scores by aircraft type

Aircraft type	PC1 Overall	PC2 Take-off	PC3 First step	PC4 Time shift	Outlier
AT, E120, B463	+	+	+	+	
A320	0	-	-	-	
B733	-	-	+	+	*
A319	-	+	0	0	
A321	+	-	-	0	

Multivariate FPCA

Route: Paris Charles de Gaulle airport $\longrightarrow$ Toulouse airport

Longitude-Latitude trajectories



Route

Paris Charles de Gaulle

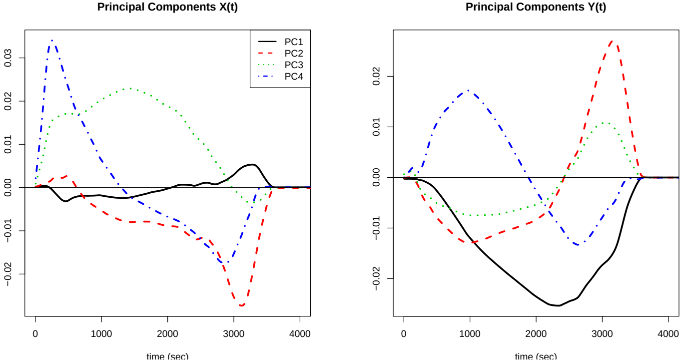
$\Downarrow$

Toulouse

- Aircraft type: A319(25%), A320(41%), A321(24%), B733 (4%), B463 (2%) AT type (4%).
- Aircraft trajectories measured at 4 seconds intervals.

Multivariate FPCA

Principal Components (PC): Principal components in X and Y-coordinates



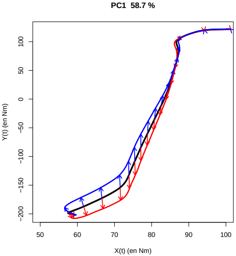
Principal Component	Total Var	X-coord	Y-coord
PC1 Overall effect	58%	2%	98%
PC2 Landing effect	14.7%	48%	52%
PC3 Separation effect	12.9%	86%	14%
PC4 Change procedure effect	6%	66%	34%

Multivariate FPCA

Effects on the mean trajectory of adding (+) or substracting (-) PC

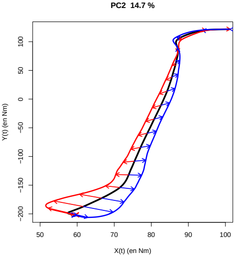
Overall effect

58%



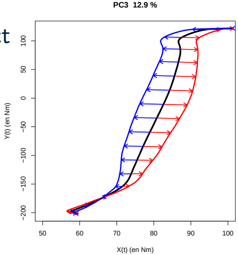
Landing effect

14.7%



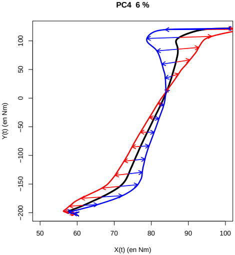
Separation effect

12.9%



Procedure effect

6%



Multivariate FPCA

Mean cluster Trajectories and the overall mean (black curve)

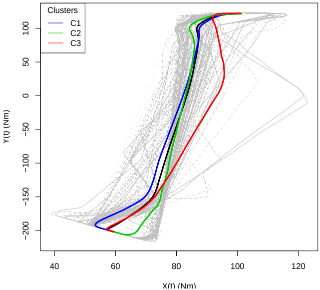


Table: k-means on scores

Aircraft type	*Cluster 1*	*Cluster 2*	*Cluster 3*
A319	15	18	0
A320	14	14	1
A321	25	28	0
AT	2	0	8
B463	10	0	2
B733	22	1	2

Multivariate FPCA

Mean cluster Trajectories (Route: Paris Orly airport ⇆ Toulouse airport)

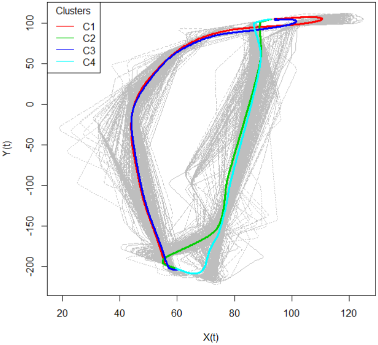
Longitude-Latitude trajectories

Route

Paris Orly

⇑

Toulouse



Route

Paris Orly

⇓

Toulouse

▶ FPCA is able to separate the two clusters located at the right side: a standard approach procedure and a short one at Toulouse airport.

Outline

Conclusion and future works

A dimension reduction tool

- An empirical basis function expansion.
- Dimension reduction: use score vectors instead of functions $x_i(t)$.

A powerful visualization tool

- Explore the ways in which trajectories vary.
- Reveal clusters and atypical trajectories

Other applications

- Generalization to 3D trajectories.
- Generate samples of trajectories.
- Reduce the dimension of simulated models.

Short bibliography

- P. Besse, J.O. Ramsay, Principal component analysis of sampled curves, *Psychometrika*, **51** (1986), 285311.
- Bookstein F.L., *Morphometric tools for landmark data:geometry and biology*, Cambridge: Cambridge University Press, 1991.
- 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), 136154.
- F. Nicol, Functional Principal Component Analysis of Aircraft Trajectories, ISIATM, July 810, 2013, Toulouse, France.
- J.O. Ramsay, G. Hooker and S. Graves, *Functional data analysis with R and Matlab*, Springer-Verlag, New York, 2009.