Principal differential analysis of the Aneurisk65 data set

Matilde Dalla Rosa · Laura M. Sangalli · Simone Vantini

Received: 2 August 2013 / Revised: 29 April 2014 / Accepted: 2 May 2014 /
Published online: 22 May 2014

M. Dalla Rosa · L. M. Sangalli · S. Vantini (✉)
MOX - Department of Mathematics, Politecnico di Milano,
Piazza Leonardo da Vinci, 32, 20133 Milano, Italy
e-mail: simone.vantini@polimi.it

M. Dalla Rosa
e-mail: matilde.dallarosa@mail.polimi.it

L. M. Sangalli
e-mail: laura.sangalli@polimi.it

1 Introduction

Principal differential analysis (PDA) (Ramsay 1996; Ramsay and Silverman 2005) is a technique that enables the estimation of a differential operator from a functional data set. This technique has already been used to analyze various types of applications such as the study of free handwriting (Ramsay 2000), the analysis of the movement of the lips during the release of the syllable Bob (Ramsay et al. 1996), economic models (Wang et al. 2008), weather (Ramsay and Silverman 2005), chemical models (Poyton et al. 2006), articulatory motion for speech analysis (Reimer and Rudzicz 2010), and sounds (Winsberg and Depalle 1999). In these papers, the main goal of PDA was the estimation of the the unknown parameters of a usually well known underlying differential operator. In contrast, in this work, we consider situations were there is not a known differential operator governing the phenomenon behavior. PDA is used here to better understand and possibly shade some new lights on the phenomenon under investigation. In particular, the estimated differential operator is used to obtain a convenient representation of the data. It provides a finite-dimensional space onto which the data can be projected and where the variability related to linear relations among derivatives can be explored (Ramsay 1996). In this perspective PDA will be applied here to analyze the AneuRisk65 data set, in order to explore the geometry of cerebral vessels.

The AneuRisk project[1] is a scientific project that aimed at investigating the role of vessel morphology, blood fluid dynamics and biomechanical properties of the vascular wall, on the pathogenesis of cerebral aneurysms. The project has gathered together researchers of different scientific fields, ranging from neurosurgery and neuroradiology to statistics, numerical analysis and bio-engineering.

Cerebral aneurysms are deformations of cerebral vessels characterized by a bulge of the vessel wall. This is a common pathology in the adult population, usually asymptomatic and not disrupting. On the other hand, the rupture of a cerebral aneurysm, even if quite uncommon, is usually a tragic event, with very high mortality. Unfortunately, rupture-preventing therapies, both endovascular and surgical treatments, are not without risks; this adds to the fact that in clinical practice general indications about rupture risk are still missing. Even the origin of the aneurysmal pathology is still unclear. Possible explanations that have been discussed in the medical literature focus on interactions between the biomechanical properties of artery walls and hemodynamic factors, such as wall shear stress and pressure; the hemodynamics is in turn strictly dependent on vascular geometry. In particular, it has been conjectured that the

[1] The project involved MOX Laboratory for Modeling and Scientific Computing (Dip. di Matematica, Politecnico di Milano), Laboratory of Biological Structure Mechanics (Dip. di Ingegneria Strutturale, Politecnico di Milano), Istituto Mario Negri (Ranica), Ospedale Niguarda Ca? Granda (Milano) and Ospedale Maggiore Policlinico (Milano), and has been supported by Fondazione Politecnico di Milano and Siemens Medical Solutions Italia. Detailed descriptions of the project's aims can be found at AneuRisk webpage http://mox.polimi.it/it/progetti/aneurisk/, where AneuRisk65 data can be downloaded. These data include the image reconstructions of one of the main cerebral vessels, the Inner Carotid Artery (ICA), described in terms of the vessel centreline and of the vessel radius profile. An increasing data warehouse concerning aneurysm pathology can be accessed from the AneuRisk Web Repository http://ecm2.mathcs.emory.edu/aneurisk managed by Emory University and Orobix.

pathogenesis of these deformations is influenced by the morphological shape of cerebral arteries, through the effect that the morphology has on the hemodynamics. For this reason, the main goal of the AneuRisk project has been the study of relationships between vessel morphology and aneurysm presence and location.

The AneuRisk65 data set is based on a set of three-dimensional angiographic images taken from 65 subjects, hospitalized at Niguarda Ca Granda Hospital (Milan), who were suspected of being affected by cerebral aneurysms. These data include the image reconstructions of one of the main cerebral vessels, the Inner Carotid Artery, described in terms of the vessel centreline and of the vessel radius profile. A thorough description of the elicitation of the AneuRisk65 data set can be found in Antiga et al. (2008), Sangalli et al. (2009) and Passerini et al. (2012). In Sangalli et al. (2009), functional principal component analysis (FPCA) was used to explore variability of the 65 ICA radius profiles. This pointed out important features amenable of a biological interpretation and able to discriminate aneurysms presence and location. In this manuscript we study the variability of the 65 radius profiles by means of PDA, thus exploring interesting possible relations among the radius profile derivatives, in order to further validate the findings enlightened in Sangalli et al. (2009) and possibly shade new insights on the data structure.

The paper is organized as follows. In Sect. 2, the PDA methodology is summarized and commented. In Sect. 3, PDA and FPCA are applied to the analysis of three simulated data sets to point out similarity and differences between the two approaches. In Sect. 4, the Aneurisk65 data set is analyzed by means of FPCA and PDA. Finally, in Sect. 5, some concluding remarks are reported.

All analyses have been performed using R (R Core Team 2012).

2 PDA as a dimensional reduction tool

Let $\{x_i\}_{i=1,\ldots,N}$ be a set of N real-valued functional data $x_i = x_i(\cdot)$, defined on the common domain (a, b). To estimate a linear differential operator of order m assume that $x_i, i = 1, \ldots, N$ belong to the Sobolev space $H^m(a, b)$. We consider linear differential operators L, with constant coefficients, of the form

$$Lx(s) = D^m x(s) + \beta_{m-1} D^{m-1} x(s) + \cdots + \beta_1 Dx(s) + \beta_0 x(s) \qquad \text{for } s \in (a, b) \tag{1}$$

where $D^m x$ denotes the m-order derivative of the function $x \in H^m(a, b)$. PDA estimates the operator L, defined by the coefficient vector $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{m-1})'$, by minimizing the sum of squared differential residuals

$$RSS(L) = \sum_{i=1}^{N} ||Lx_i||^2, \tag{2}$$

with $||\cdot||$ being the $L^2(a, b)$ norm induced by the usual inner product in $L^2(a, b)$, which is indicated in the following with $< \cdot, \cdot >$. The driving idea behind the minimization

problem (2) is the search for a differential operator able to describe a large part of the variability shown by the functional data set $\{x_i\}_{i=1,\ldots,N}$.

The minimization problem (2) is analytically solved for the value $\hat{\boldsymbol{\beta}} = \mathbf{A}^{-1}\mathbf{b}$, where $(\mathbf{A})_{j_1 j_2} = \sum_{i=1}^{N} < D^{j_1}x_i, D^{j_2}x_i >$ and $(\mathbf{b})_{j_1} = \sum_{i=1}^{N} < D^{j_1}x_i, D^{m}x_i >$ with $j_1, j_2, \ldots \in \{0, 1, \ldots, m-1\}$. The corresponding estimate $\hat{L}$ of the unknown operator L is:

$$\hat{L}x(s) = D^m x(s) + \hat{\beta}_{m-1}D^{m-1}x(s) + \cdots + \hat{\beta}_1 Dx(s) + \hat{\beta}_0 x(s) \quad \text{for} \quad x \in H^m(a,b).$$

Notice that the minimization problem (2) is formally identical to the minimization problem encountered within the ordinary least squares estimation of the unknown parameters of a functional–functional concurrent regression model with constant parameters,

$$D^m x_i(s) = -\beta_{m-1}D^{m-1}x_i(s) - \cdots - \beta_1 Dx_i(s) - \beta_0 x_i(s) + Lx_i(s) \quad \text{for} \quad i = 1, \ldots, N,$$

being in this case $D^m x_i$ the functional response, $-D^{m-1}x_i, \ldots, -Dx_i, -x_i$ the functional regressors, and Lx_i the residual term.

The estimation of the linear operator L and the Partitioning Principle of Hilbert spaces (see, e.g., Rudin 1991) jointly enable the orthogonal decomposition of the functional space the data belong to, into two different components $ker(\hat{L})$ and $ker(\hat{L})^{\perp}$, where $ker(\hat{L})$ is the m-dimensional linear space of all functions $\hat{x}$ satisfying the linear differential relation $\hat{L}\hat{x} = 0$, and $ker(\hat{L})^{\perp}$ is its orthogonal counter part. This means that for $i = 1, \ldots, N$, the function x_i can be univocally decomposed in the sum of two orthogonal components $\hat{x}_i$ and $\hat{e}_i$ such that $x_i = \hat{x}_i + \hat{e}_i$ with $\hat{L}\hat{x}_i = 0$ and $\hat{L}\hat{e}_i = \hat{L}x_i$. The functions $\hat{x}_i$ and $\hat{e}_i$ are named structural and residual component, respectively. Indeed, the function $\hat{x}_i$ gathers those differential features of x_i, common to the entire functional data set, that can be described by the linear differential equation $\hat{L}\hat{x}_i = 0$. On the contrary, the function $\hat{e}_i$ gathers those differential features of x_i that the linear differential operator $\hat{L}$ cannot handle, i.e., $\hat{L}x_i = \hat{L}\hat{e}_i$. Due to the orthogonality of $\hat{x}_i$ and $\hat{e}_i$, for $i = 1, \ldots, N$ the structural component $\hat{x}_i$ can be identified with the solution of the following minimization problem:

$$\min_{x \in ker(\hat{L})} ||x - x_i||^2 \quad \text{for} \quad i = 1, \ldots, N. \tag{3}$$

It is well known that any solution of a non degenerated linear differential equation of order m can be expressed as a linear combination of m linearly independent complex-valued exponential functions $\psi_1, \ldots, \psi_m$. In particular we have that $ker(\hat{L}) = \text{span}\{\hat{\psi}_1, \ldots, \hat{\psi}_m\}$ with $\hat{\psi}_1(s) = e^{\hat{\lambda}_1 s}, \ldots, \hat{\psi}_m(s) = e^{\hat{\lambda}_m s}$ where $\hat{\lambda}_j$ are the m different complex roots of the characteristic polynomial $\lambda^m + \hat{\beta}_{m-1}\lambda^{m-1} + \cdots + \hat{\beta}_1\lambda + \hat{\beta}_0$. This analytical characterization of the space $ker(\hat{L})$ provides an explicit solution to the problem (3). Indeed, for $i = 1, \ldots, N$, $\hat{x}_i(s) = \sum_{j=1}^{m} c_{ij}\hat{\psi}_j(s)$ with scores $(\hat{c}_{i1}, \hat{c}_{i2}, \ldots, \hat{c}_{im})' = -\mathbf{C}^{-1}\mathbf{d}_i$, where $(\mathbf{C})_{j_1 j_2} = < e^{\hat{\lambda}_{j_1} s}, e^{\hat{\lambda}_{j_2} s} >$ and

$(\mathbf{d}_i)_{j_1} = < e^{\hat{\lambda}_{j_1} s}, x_i(s) >$. Finally, the differential residual term is simply computed as $\hat{e}_i = x_i - \hat{x}_i$.

The final result of the overall procedure is a new functional data set $\{\hat{x}_1, \ldots, \hat{x}_N\}$ derived from the original data set $\{x_1, \ldots, x_N\}$ belonging to an m dimensional functional space provided with an analytical representation and interpretable in terms of linear relations among different order derivatives.

A useful tool to measure the effectiveness of the obtained dimensional reduction is the quantity

$$RSQ = \frac{\sum_{i=1}^{N} ||D^m x_i||^2 - \sum_{i=1}^{N} ||\hat{L} x_i||^2}{\sum_{i=1}^{N} ||D^m x_i||^2}$$

introduced in Ramsay (1996). For linear differential operators L with constant coefficients $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{m-1})'$, as in (1), RSQ is a ratio between the structural variability and the total variability. Indeed $RSQ = 1$ when there is only structural variability (i.e., $x_i = \hat{x}_i$ for $i = 1, 2, \ldots, N$) and $RSQ = 0$ when there is no structural variability (i.e., $x_i = \hat{e}_i$ for $i = 1, 2, \ldots, N$).

A couple of theoretical remarks about the dimensional reduction obtained by the projection of functional data onto $ker(\hat{L})$ need to be mentioned. The first remark concerns to the link between PDA and functional regression. Note that, while in the minimization problem (3), N actual functional regression analyses are performed (the only unusual feature with respect to a traditional functional regression being that the regressors are not known but estimated in a previous stage of the analysis), in the minimization problem (2), the link is instead just formal and does not concern the modeling aspect. Indeed, in (2), the functional "regressors" $-D^j x_i$, $j = 0, \ldots, m-1$, are random and the random functional "error" Lx_i is not independent from the functional "regressors".

The second remark is related to the link between PDA and FPCA. FPCA is probably the most used approach to perform a dimensional reduction of a functional data set $\{x_i\}_{i=1,\ldots,N}$, with $x_i \in L^2(a, b)$. This is due to the fact that for any dimension $k = 1, 2, \ldots, N-1$ FPCA provides an affine k-dimensional subspace of $L^2(a, b)$ which is statistically optimal in an $L^2(a, b)$ perspective. Indeed, it can be proven that the sum of the squared $L^2(a, b)$-norms of the residuals of the orthogonal projections of the functional data onto this k-dimensional subspace is minimal over all possible k-dimensional subspaces of $L^2(a, b)$. On the whole, FPCA provides a sequence of nested best approximating affine subspaces of increasing dimension. In detail, these subspaces are all centered in the sample mean function and generated by the first k eigenfunctions of the sample covariance operator, which are real-valued and orthogonal being the covariance operator positive semi-definite. The goodness of fit to the original data of the projections of the data on the k-dimensional FPCA affine subspace is commonly measured by means of the fraction of explained total variance (namely, the ratio between the sum of the first k eigenvalues of the sample covariance operator and the sum of first $N-1$ eigenvalues of the sample covariance operator) which represents the ratio between the average squared $L^2(a, b)$ distance of the projections to the sample mean and the average squared $L^2(a, b)$ distance of the original functions

to the sample mean. For further details about FPCA refer for instance to Ramsay and Silverman (2005).

It is worth noticing that, differently from FPCA, where the dimensional reduction is driven just by the point-wise values of the functional data along the domain (a, b), the dimensional reduction obtained by PDA is driven by the values of linear combinations of the first m derivatives of the functional data along the domain (a, b). PDA is therefore able to capture hidden features of the data that are purely functional in their nature. By definition, FPCA is expected to provide an effective dimensional reduction in any situation where most of the functional variability is expressed within some finite dimensional subspace; at the same time a simple analytical expression of the principal components is often missing and the interpretation of these components is often non-trivial. On the contrary, PDA is expected to provide an effective dimensional reduction only in those situations where most of the functional variability is expressed within specific finite dimensional subspaces, those generated by some functions of the form $e^{(\alpha \pm \omega i)s}$ with α and $\omega \in \mathbb{R}$. On the other hand, PDA always provides interpretable results in terms of constant, exponential, sinusoidal, or damped-sinusoidal functions. Since the effectiveness of PDA is related to particular finite dimensional subspaces, it is obvious that the dimensional reduction provided by FPCA is always more effective than the one provided by PDA. The aim of FPCA is indeed to achieve the most effective dimensional reduction from a purely geometric point of view. The aim of PDA is instead to provide a compact interpretable model in term of relationship between derivatives. It is thus interesting to determine if PDA can be a useful tool to have a different insight of the functional variability, at least in cases where both PDA and FPCA provide an effective dimensional reduction. A more thorough discussion about similarities and differences between PDA and FPCA is reported in Ramsay (1996).

3 Comparing FPCA and PDA of three simulated data sets

In this section we compare PDA and FPCA in the dimensional reduction of three simulated data sets, of $N = 200$ functions each, generated according three different scenarios. In Case A we consider two clusters of functional data generated from the same orthogonal and sinusoidal basis. In Case B data are generated from a polynomial basis. Finally, in Case C there is a clear underlying differential model, namely we consider the harmonic oscillator in a viscous fluid (Feynman et al. 2013). All data are generated over the interval [0, 10], on a evenly-spaced grid of 101 abscissa points $\{s_k = k/10\}$ with $k = 0, 1, \ldots, 100$.

In Case A the 200 functional data are generated as

$$x_{ik} = a_i + b_i \sin\left(\frac{2\pi}{10} s_k\right) + c_i \cos\left(\frac{2\pi}{10} s_k\right) + \varepsilon_{ik},$$
$$k = 0, 1, \ldots, 100, \; i = 1, \ldots, 200,$$

with random coefficients $(a_i, b_i, c_i)' \sim iid \; \mathcal{N}_3((1, 1, 1)', \mathbf{I})$ for $i = 1, \ldots, 100$, $(a_i, b_i, c_i)' \sim iid \; \mathcal{N}_3((-1, -1, -1)', \mathbf{I})$ for $i = 101, \ldots, 200$, and measurement errors $\varepsilon_{ik} \sim iid \; \mathcal{N}_1(0, 10^2)$. The two different distributions assumed for the random

coefficients, respectively of the first 100 and last 100 functional data, generate two clusters of functions.

In Case B the 200 functional data are generated as

$$x_{ik} = a_i \frac{s_k}{10} + b_i \left(\frac{s_k}{10}\right)^2 + c_i \left(\frac{s_k}{10}\right)^3 + \varepsilon_{ik}, \quad k = 0, 1, \ldots, 100, \quad i = 1, \ldots, 200,$$

with random coefficients $(a_i, b_i, c_i)' \sim iid\ \mathcal{N}_3((2, -2, 2)', 16 \cdot \mathbf{I})$, and measurement errors $\varepsilon_{ik} \sim iid\ \mathcal{N}_1(0, 10^2)$.

In Case C the 200 functional data are generated as

$$x_{ik} = a_i e^{\lambda_1 s_k} + b_i e^{\lambda_2 s_k} + \varepsilon_{ik}, \quad k = 0, 1, \ldots, 100, \quad i = 1, \ldots, 200,$$

with random coefficients $(a_i, b_i)' \sim iid\ \mathcal{N}_2((2, -2)', \mathbf{I})$ for $i = 1, \ldots, 100$, $(a_i, b_i)' \sim iid\ \mathcal{N}_2((2, -2)', \mathbf{I})$ for $i = 101, \ldots, 200$, and measurement errors $\varepsilon_{ik} \sim iid\ \mathcal{N}_1(0, 10^2)$. Parameters λ_1 and λ_2 are the real roots of the characteristic polynomial of the differential equation $m \cdot D^2 x = -6\pi\eta R \cdot Dx - k \cdot x$ describing the dynamics of a sphere of mass $m = 1$ and radius $R = 1/2$ attached to a coil with elasticity constant $k = 1$ in a viscous fluid with drag constant $\eta = 1/4$.

The original data are discrete and noisy observations of curves. PDA requires the estimation of the function derivatives up to the considered order of the differential operator. Here in particular we shall explore differential operators of maximum order $m = 4$. To obtain the estimates of the curves and correspondingly of their derivatives of order $m \leq 4$ from noisy data we use smoothing splines of degree 5, with penalization of fourth order derivative. These curve estimates (zero derivative) are also used for FPCA. The curve estimates, for the three simulation cases, are displayed in the left column panels of Fig. 1 (Case A top, Case B center, Case C bottom).

In the left panels of Table 1 we report, for the three cases, the RSQ and the estimated coefficients of the linear differential operators the order $m = 1, 2, 3, 4$, provided by PDA. On the right panels of the same table we report the fraction of explained total variance in FPCA. Figure 1, center and right column panels, displays the estimated basis functions detected in the three simulation cases by PDA (center panels) for the chosen order of the differential operator, and by FPCA (right panels) for the chosen number of principal components.

Before focussing on the comparison between FPCA and PDA results, it is important to recall the different meanings of the fraction of explained total variance and of the RSQ index, used to evaluate the goodness of the dimensional reduction achieved by FPCA and by PDA, respectively. Indeed, even though these indexes are both bounded between 0 and 1, with lower values associated to less effective dimension reductions and higher values associated to more effective ones, these indices should not be directly compared. Indeed, while the fraction of explained total variance used in FPCA exactly measures to what extent the variability of the functional data set is captured by its reduced representation, the RSQ index used in PDA is more similar to a regression-R^2 index, and measures to what extent the mth derivatives can be described by means of a linear combination of lower-order derivatives and thus provides a goodness of fit for a particular linear differential model. The different nature of the two indices is

Table 1 PDA for the three simulated cases A, B, and C, the estimated coefficients of the linear differential operator of order $m = 1, 2, 3, 4$ provided by PDA and the corresponding RSQ

PDA							FPCA	
	$D^4 f$	$D^3 f$	$D^2 f$	$D^1 f$	$D^0 f$	RSQ		Exp. var.
Case A								
$m = 1$				1	−0.0034	0.0001	$k = 1$	0.6783
$m = 2$			1	0.0003	0.2102	0.5342	$k = 2$	0.8844
$m = 3$		1	0.0000	0.3965	0.0004	0.9999	$k = 3$	1.0000
$m = 4$	1	−0.0243	0.3972	−0.0095	−0.0002	0.9997	$k = 4$	1.0000
Case B								
$m = 1$				1	−0.2182	0.8155	$k = 1$	0.9742
$m = 2$			1	−0.2603	0.0288	0.6955	$k = 2$	0.9997
$m = 3$		1	−0.1111	−0.0059	0.0013	0.4763	$k = 3$	1.0000
$m = 4$	1	−0.0001	0.0001	0.0000	0.0000	0.0194	$k = 4$	1.0000
Case C								
$m = 1$				1	0.6246	0.1942	$k = 1$	0.6650
$m = 2$			1	1.3226	0.7803	0.9658	$k = 2$	0.9994
$m = 3$		1	1.5952	1.2905	0.3870	0.9867	$k = 3$	0.9996
$m = 4$	1	1.7676	1.8584	1.1052	0.2698	0.9667	$k = 4$	0.9999

FPCA for the three simulated cases A, B, and C, the fraction of total variance explained by the first $k = 1, 2, 3, 4$ functional principal components

also enlightened by the fact that the fraction of explained total variance is monotonic with respect to the dimension of the functional subspace considered, while RSQ is not.

In case A, note that the set $\{1, \sin((2\pi/10)s), \cos((2\pi/10)s)\}$ is an orthogonal basis for the three-dimensional space which all functions (when not affected by measurement error) would belong to, and that this space also coincides with $ker(D^3+((2\pi/10))^2 D)$. It is thus obvious that both the projection on the first three functional principal components estimated by FPCA, and the projection on the kernel of a third-order operator estimated by PDA, are expected to be effective dimensional reduction tools. Indeed they are, as it is shown by the fraction of explained total variance by the first three principal components and by the RSQ when a third-order linear differential operator is used (see Table 1). In particular, the obtained RSQ strongly support the choice of $m = 3$ order differential operator. This choice is also supported by the comparison between the estimated coefficients of the third-order operator and the fourth-order operator (see Table 1), as the latter coefficients resemble the former ones, but shifted by an order of derivative. Indeed, the estimated third-order operator correctly targets the operator $D^3 + ((2\pi/10))^2 D$, while the estimated fourth-order operator targets the operator $D^4 + ((2\pi/10))^2 D^2$ whose kernel includes the kernel of the previous one. Considering FPCA results, it emerges that the presence of two clusters of functional data drives in this case the determination of the first principal component, that detects the direction connecting the two clusters, while the second and third components

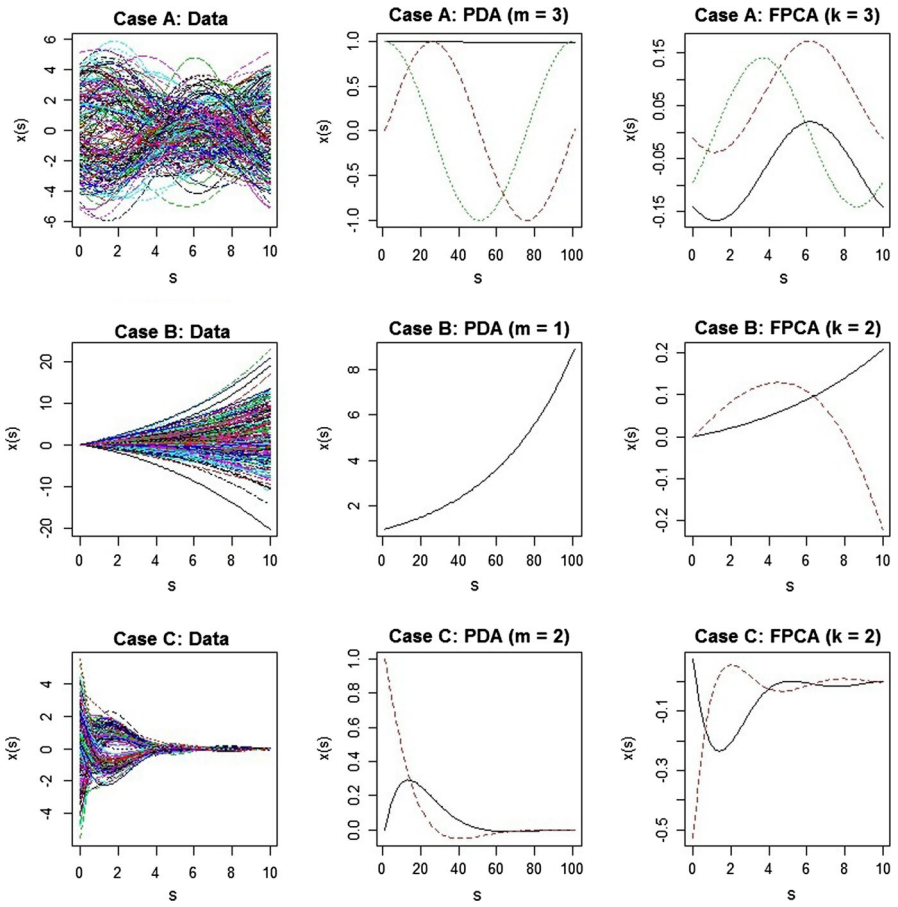

Fig. 1 *First column* the three simulated functional data sets as estimated by smoothing splines of degree 5; *Case A top, Case B center, Case C bottom*. *Second column* corresponding basis functions detected by PDA. *Third column* corresponding basis functions (principal components) detected by FPCA

describe the residual variability. The simple description of functional data as sum of a constant term and a sinusoidal term thus remains unveiled by FPCA. PDA instead describes these clusters as made of functions characterized by the same differential model, but with different boundary conditions. Note moreover that PDA provides explicit estimates of the frequency of the sinusoidal component while in FPCA the latter can only be estimated heuristically.

Case B represents a scenario where PDA is not in principle expected to be effective, since data are generated from a polynomial basis, without any underlying differential model. On the other hand, an analysis of RSQ values highlights that PDA, for some order of the differential operator, provides in fact an effective representation of the functional data, that is easily interpretable in terms of exponential, sinusoidal and dumped sinusoidal functions. This representation is alternative to the orthogonal representation provided by FPCA.

Finally, in Case C there is a clear underlying differential operator, modeling the physics of the observed phenomenon. PDA is in this case expected to yield a more meaningful description of the phenomenon than FPCA. Indeed, in this last example, the PDA representation is easily interpretable in terms of physical quantities, while the eigenfunctions and eigenvalues of the sample covariance operator have no simple physical interpretation.

4 Analysis of the AneuRisk65 data set

In this section we compare PDA and FPCA in the analysis of the AneuRisk65 data. The aim of the analysis is to to explore a conjecture grounded on practical experience of neuroradiologists at Niguarda Ca' Granda Hospital (E. Boccardi, personal communication): cerebral arteries of patients with an aneurysm at the terminal bifurcation of the internal carotid artery (ICA), or after it, show peculiar geometric features. It should be mentioned that aneurysm at or after the terminal bifurcation of the ICA are the most life-threatening; the possible rupture of one such aneurysm is fatal in most cases.

We recall that AnueRisk65 data contains the image reconstructions of the ICA of 65 subjects. To explore this conjecture, the 65 patients are divided in two groups:

Upper group formed by 33 subjects having an aneurysm at or after the terminal bifurcation of the ICA;

Lower group formed by 32 subjects, including 25 subjects having an aneurysm along the ICA, before its terminal bifurcation, and seven subjects not displaying any visible aneurysm during the angiography.

In particular, we shall here focus on the analysis of the radius profiles of the ICA. For each subject i, AneuRisk65 dataset includes the measurements of the radius $\{R_{ij}\,;\, j = 1, \ldots, n_i\}$ of the vessel lumen section, along a fine grid of points $\{s_{ij}\,;\, j = 1, \ldots, n_i\}$, starting from the terminal bifurcation of the ICA moving upstream towards the heart. Likewise in Sect. 3, a functional representation of the radius profiles and of their derivatives of order $m \leq 4$, is obtained by means of smoothing splines of degree five, with penalization of the fourth derivative. Moreover, in order to enable meaningful comparisons across patients, the 65 ICA radius profiles are registered as described in Sangalli et al. (2009), using the `fdakma` R package Patriarca et al. (2013). Since the 65 curves are observed on different abscissa intervals, the following analyses focus on the segment of the ICA where the radius measurements are available for all subjects, i.e., for values of the abscissa between -3.29 and -0.74 cm. The conventional negative sign of the abscissa parameter highlights that we are moving upstream, i.e., in opposite direction with respect to blood flow, the origin of the abscissa corresponding to the terminal bifurcation of the ICA. The registered ICA radius profiles over this abscissa interval are displayed in Fig. 3, panel E.

We first summarize the results obtained by means of FPCA, along the lines described in detail in Sangalli et al. (2009). Figure 3, right panels, show the projections of the 65 radius profiles on the subspace generated by the first $k = 1, 2, 3, 4$ principal components of the radius profile, colored in blue for subjects in the Upper group and in red for subjects in the Lower group. The right panels of Table 2 report the percentage of

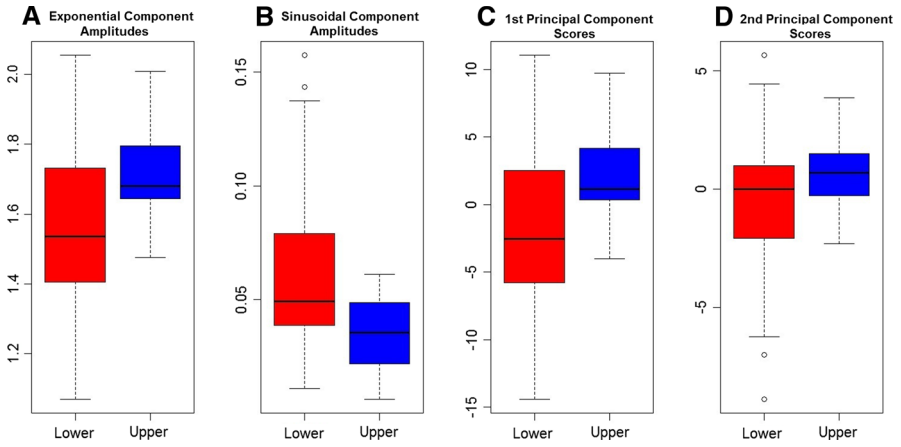

Fig. 2 *Panels A and B*: *boxplots* of the amplitudes of the exponential component $\hat{\psi}_1$ (*panel A*) and of the amplitudes of the sinusoidal component generated by $\hat{\psi}_2$ and $\hat{\psi}_3$ (*panel B*), estimated by PDA, for the Lower and Upper groups in the AneuRisk65 data set. *Panels C and D*: *boxplots* of the scores of the first (*panel C*) and second functional principal components (*panel D*)

total variance explained by the first k principal components. A visual inspection of the projections on the first principal component (Fig. 3, panel F) and on the first and second principal component (Fig. 3, panel G) highlights that these principal components are amenable of a clear biological interpretation and are also useful for the discrimination of the two groups of subjects. The first principal component is interpretable as an average size of the carotid, distinguishing between narrow and wide ICAs. Notice that the radius profiles are all characterized by a slight progressive narrowing of the vessel moving from the heart towards the terminal bifurcation of the ICA; this is the so-called tapering effect. The second principal components emphasizes this tapering effect in the terminal tract of the ICA, distinguishing between ICA presenting a more or less marked tapering in their terminal part.

Figure 2, panels C and D, reports the distributions of the scores corresponding to the first and second principal component for subjects in the Upper group and subjects in the Lower group. These scores may be used to discriminate the two groups of patients. In fact, the distribution of FPCA scores have significantly different means and variances for the two groups, as confirmed by appropriate t tests and F tests for equality of means and variances. According to these differences, Upper group patients tend to have wider and more tapered ICAs, compared to patients belonging to the Lower group. Moreover the variance of these geometrical features is significantly smaller in the Upper group than in the Lower group. The Upper group is indeed very well characterized in terms of the geometrical features represented by the first two principal components of radius. A quadratic discriminant analysis based on the scores of the first two principal components misclassifies 20.0 % of subjects (23.1 % using leave-one-out crossvalidation), supporting a strong association between ICA radius and aneurysm presence and location, and highlighting strong statistical evidence in favor of the conjecture explored within the project.

Table 2 PDA For the AneuRisk65 dataset, the estimated coefficients of the linear differential operators of order $m = 1, 2, 3, 4$, provided by PDA, and the corresponding *RSQ*

PDA							FPCA	
Radius	$D^4 f$	$D^3 f$	$D^2 f$	$D^1 f$	$D^0 f$	RSQ	Radius	Exp. var.
$m = 1$				1	0.0082	0.0261	$k = 1$	0.6654
$m = 2$			1	0.0128	0.0027	0.0028	$k = 2$	0.7910
$m = 3$		1	0.0129	0.9845	0.0076	0.4145	$k = 3$	0.8557
$m = 4$	1	0.0145	2.3601	0.0407	0.0038	0.6185	$k = 4$	0.9061

FPCA the fraction of total variance explained by the first $k = 1, 2, 3, 4$ functional principal components

We now present the results obtained by means of PDA. In the framework of PDA, the first problem to face is finding the order m of the linear operator $\hat{L}$, that provides the best compromise between a satisfactory goodness of fit and an easy interpretability of the results. Figure 3, left panels, show the projections of the 65 radius profiles on the kernels of the linear differential operators of order $m = 1, 2, 3, 4$ estimated by PDA. Table 2 reports the corresponding RSQ values. The RSQ values highlight that the linear differential operators of order one and two are not able to describe the dynamics of the radius along the ICA while linear differential operators of higher order do. For instance, considering the order three operator, the 65 radius functions and their first two derivatives are able to explain more than 40 % of the variability shown by their respective third derivatives.

Favoring parsimony and easy interpretability of the results, and considering the fact that the estimates of high order derivatives might not be very accurate, we consider here the $m = 3$ order operator. This is then estimated as follows:

$$\hat{L}x = D^3x + 0.0129D^2x + 0.9845D^1x + 0.0076x. \tag{4}$$

The three roots of the associated characteristic polynomial are $\hat{\lambda}_1 = -0.0077$, $\hat{\lambda}_2 = -0.0026 + 0.9922i$, and $\hat{\lambda}_3 = -0.0026 - 0.9922i$. The corresponding basis $\{\hat{\psi}_1, \hat{\psi}_2, \hat{\psi}_3\}$, spanning $ker(\hat{L})$, is thus given by:

$$\hat{\psi}_1(s) = e^{-0.0077s}, \hat{\psi}_2(s) = e^{-0.0026s}\cos(0.9922s), \hat{\psi}_3(s) = e^{-0.0026s}\sin(0.9922s).$$

The first basis function $\hat{\psi}_1$ defines a decreasing function that describes jointly the overall size and the tapering of the 65 ICAs. The two remaining basis functions $\hat{\psi}_2$ and $\hat{\psi}_3$ jointly define a slightly dumped sinusoidal function (being the real parts of the adjoint roots very small), of period 6.33 mm and with arbitrary phase, and describe variations of the radius along the ICA.

Figure 3 panel C reports the projections of the 65 ICA radius profiles in $Kern(\hat{L})$, colored in blue for subjects in the Upper group and in red for subjects in the Lower group. This representation can be heuristically compared with the representation in panel G, where the radius profiles are projected in the subspace generated by the first two functional principal components. Both the FPCA-based and the PDA-based representation highlight important, but also different, features of the functional data

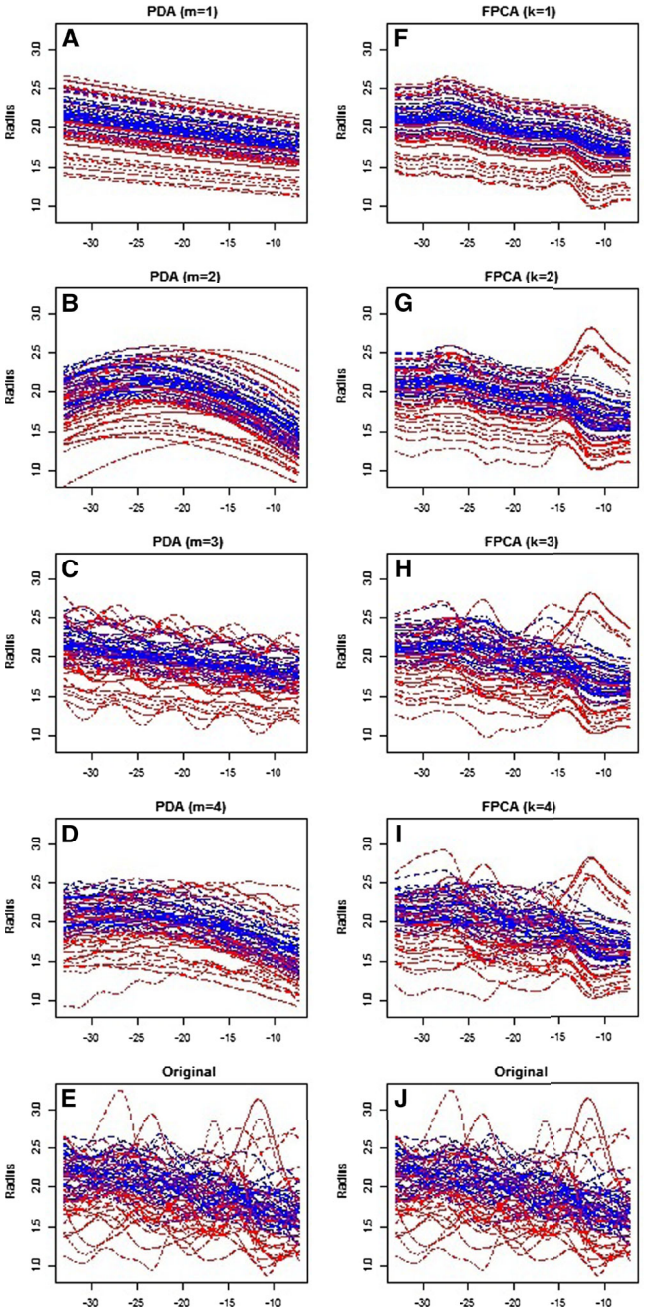

Fig. 3 *Panels A to D*: for the AneuRisk65 data set, the projections of the 65 radius functions in the kernels of the linear differential operators of order $m = 1, 2, 3, 4$ estimated via PDA. *Panels F to I*: the projections on the affine subspaces generated by the first $k = 1, 2, 3, 4$ functional principal components. *Lower panels* the 65 original radius functions. *Curves are colored in blue* for subjects in the Upper group and in *red* for subjects in the Lower group (color figure online)

set. The FPCA-based dimensional reduction obtained through the first two principal components points out the existence of an average tapering of the vessels (represented by the functional mean), the presence of wider versus narrower vessels (represented by the first principal component), and the presence of an extra source of variability in the radius at the very end of the ICA, controlling the tapering in the terminal part of the ICA (represented by the second principal component). See Sangalli et al. (2009) for details. The PDA-based dimensional reduction obtained through the third order differential operator in (4) points out the presence of wider and more tapered ICAs versus narrower and less tapered ICAs (represented by the exponential function $\hat{\psi}_1$, that jointly describes the size and tapering effects) and the presence of ICAs with smoothed profiles, characterized by regularly a decreasing radius, versus ICAs with corrugated profiles, characterized by stronger fluctuations in the radius (represented by the sinusoidal function generated by $\hat{\psi}_2$ and $\hat{\psi}_3$, that describes this rippling effect).

Likewise FPCA, also PDA highlights differences in the geometrical features of the ICA of Upper and Lower group subjects. Focussing on the discrimination of the two groups, panels A and B of Fig. 2 report the boxplots of the amplitudes of the exponential component $\hat{\psi}_1$ and the boxplots of the amplitudes of the sinusoidal component generated by $\hat{\psi}_2$ and $\hat{\psi}_3$, for the two groups of subjects. As shown by Fig. 3 panel C and by these boxplots, and confirmed by suitable tests, Upper group subjects have more homogeneous ICA radius profiles with respect to Lower group subjects, which display more varying behaviors. In particular, Upper group subjects typically present wider, more tapered, and less rippled ICAs than Lower group subjects. The phase of the sinusoidal component turns out to be uninformative in terms of Upper and Lower group characterization. Moreover, it is interesting to note the presence in the Lower group of subjects whose ICAs have strong fluctuations in the radius. Finally, a bivariate quadratic discriminant analysis based on the amplitudes of the exponential and sinusoidal components misclassifies 8.4 % of the subjects (21.5 % using leave-one-out crossvalidation), a slightly better misclassification error than the one achieved relying on the first two principal component.

It is interesting to note that the first principal component scores and the amplitudes of the exponential component provided by PDA are almost linearly dependent, while the second principal component scores and the amplitudes of the sinusoidal component provided by PDA are almost uncorrelated. This supports that the rippling effect detected by PDA is not associated to the variability described by the second functional principal component. Looking at panel H of Fig. 3, one could claim that the rippling effect is gathered also by the third principal component. This hypothesis is though rejected since the third principal component scores are not discriminant of the Lower and Upper groups, differently from the amplitudes of the sinusoidal component provided by PDA that describes this rippling effect.

Summarizing, the conclusions drawn from PDA about the size and tapering effects of the ICA are in complete agreement with the results obtained by FPCA. The conclusions about the rippling effect of the ICA are instead specific of the PDA approach, as a similar feature is not pointed out by FPCA. This can be explained by the fact that these fluctuations in the radius, even if large and present in many subjects, do not

occur at the same abscissa values; thus FPCA, that differently from PDA focuses on pointwise values of the functions, is not able to recognize them as a unique variability feature.

As a final remark, the different meaning of RSQ value in PDA and of the fraction of explained total variance in FPCA, pointed out in the last paragraph of Sect. 3, is clearly enlightened in the comparison between the dimensional reduction obtained by using the first differential operator provided by PDA (Fig. 3, panel A) and by using the first principal component (Fig. 3, panel F). The two representations are qualitatively similar, both gathering the size and tapering effect, and are identical in terms of discrimination between the Upper and the Lower group subjects, being the amplitudes of the exponential components almost proportional to the first principal component scores. Despite of that, the RSQ used in PDA is very low (2.61 %) while the fraction of total variance explained by the first principal component is quite high (66.54 %).

5 Conclusions

The analysis of the AneuRisk65 data set and of the synthetic data sets show that PDA can be a useful tool, as an alternative to FPCA, to perform a dimensional reduction of a functional data set, also when the phenomenon under investigation is not commonly described via a differential model governing the phenomenon behavior. In particular, when effective as a dimensional reduction tool, PDA can provide a representation of functional data (easily interpretable in terms of exponential, sinusoidal, or dumped-sinusoidal components) that can detect important features of the data that FPCA is not able to reveal, providing a different insight into the analysis of a functional data set.

References

Antiga L, Piccinelli M, Botti L, Ene-Iordache B, Remuzzi A, Steinman DA (2008) An image-based modeling framework for patient-specific computational hemodynamics. Med Biol Eng Comput 46:1097–1120

Feynman R, Leighton R, Sands M (2013) The Feynman lectures on physics, vol 1, online edition. In: Gottlieb MA, Pfeiffer R (eds) http://www.feynmanlectures.caltech.edu. Basic Books

Passerini T, Sangalli LM, Vantini S, Piccinelli M, Bacigaluppi S, Antiga L, Boccardi E, Secchi P, Veneziani A (2012) An integrated CFD-statistical investigation of parent vasculature of cerebral aneurysms. Cardiol Eng Tech 3(1):26–40

Patriarca M, Sangalli LM, Secchi P, Vantini S, Vitelli V (2013) fdakma: Clustering and alignment of of a functional dataset. R package version 1.0. http://CRAN.R-project.org/package=fdakma

Poyton AA, Varziri MS, McAuley KB, McLellan P-J, Ramsay JO (2006) Parameter estimation in continuous-time dynamic models using principal differential analysis. Comput Chem Eng 30:698–708

R Core Team (2012) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/

Ramsay JO (1996) Principal differential analysis: data reduction by differential operators. J R Stat Soc Ser B 58(3):495–508

Ramsay JO, Munhall KG, Gracco VL, Ostry DJ (1996) Functional data analysis of lip motion. J Acoust Soc Am 99:3718–3727

Ramsay JO (2000) Functional components of variation in handwriting. J Am Stat Assoc 95:9–16

Ramsay JO, Silverman BW (2005) Functional data analysis, IInd edn. Springer, NY

Reimer M, Rudzicz F (2010) Identifying articulatory goals from kinematic data using principal differential analysis. Proc Interspeech 2010:1608–1611

Rudin W (1991) Functional analysis, 2nd edn. McGraw-Hill, New York
Sangalli LM, Vantini S, Secchi P, Veneziani A (2009) A case study in exploratory functional data analysis: geometrical features of the internal carotid artery. J Am Stat Assoc 104(485):37–48
Wang S, Jank W, Shmueli G, Smith P (2008) Modeling price dynamics in eBay auctions using differential equations. J Am Stat Assoc 103(483):1100–1118
Winsberg S, Depalle P (1999) Applications of principal differential analysis to data reduction and extraction of musical features of sound. In: Proceedings of international computer music conference (ICMC'99)