STATISTICS IN TRANSITION new series, Spring 2014
Vol. 15, No. 2, pp. 231–242

METHODS OF REDUCING DIMENSION FOR FUNCTIONAL DATA

Tomasz Górecki[1], **Mirosław Krzyśko**[2], **Łukasz Waszak**[3], **Waldemar Wołyński**[4]

ABSTRACT

In classical data analysis, objects are characterized by many features observed at one point of time. We would like to present them graphically, to see their configuration, eliminate outlying observations, observe relationships between them or to classify them. In recent years methods for representing data by functions have received much attention. In this paper we discuss a new method of constructing principal components for multivariate functional data. We illustrate our method with data from environmental studies.

Key words: multivariate functional data, functional data analysis, principal component analysis, multivariate principal component analysis.

1. Introduction

The idea of principal component analysis (PCA) is to reduce the dimensionality of a data set consisting of a large number of correlated variables, while retaining as much as possible of the variation present in the data set. This is achieved by transforming them to a new set of variables, the principal components, which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all of the original variables.

In recent years methods for representing data by functions or curves have received much attention. Such data are known in the literature as functional data (Ramsay and Silverman, 2005). Examples of functional data can be found in various application domains, such as medicine, economics, meteorology and

[1] Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Poland. E-mail: tomasz.gorecki@amu.edu.pl.
[2] Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Poland. E-mail: mkrzysko@amu.edu.pl.
[3] Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Poland. E-mail: lwaszak@amu.edu.pl.
[4] Faculty of Mathematics and Computer Science, Adam Mickiewicz University, Poland. E-mail: wolynski@amu.edu.pl.

many others. In previous papers on functional data analysis, objects are characterized by only one feature observed at many time points (see Ramsay and Silverman (2005)). In many applications there is a need to use statistical methods for objects characterized by many features observed at many time points (double multivariate data). In this case, such data are called multivariate functional data. A pioneering theoretical work was that of Besse (1979), where random variables take values in a general Hilbert space. Saporta (1981) presents an analysis of multivariate functional data from the point of view of factorial methods (principal components and canonical analysis). Finally, Jacques and Preda (2014) proposed principal component analysis for multivariate functional data (MFPCA) applied to the methods of cluster analysis. In this paper we propose another method of construction of principal components for multivariate functional data, along with an in-depth interpretation of these variables.

2. Classical principal component analysis (PCA)

Suppose we observe a p-dimensional random vector $\boldsymbol{X} = (X_1, X_2, ..., X_p)' \in \mathbb{R}^p$. We further assume that $\mathrm{E}(\boldsymbol{X}) = \mathbf{0}$ and $\mathrm{Var}(\boldsymbol{X}) = \boldsymbol{\Sigma}$.

In the first step we seek a variable U_1 in the form

$$U_1 =< \boldsymbol{u}_1, \boldsymbol{X} >= \boldsymbol{u}_1' \boldsymbol{X} = \sum_{i=1}^{p} u_{1i} X_i,$$

having maximum variance for all $\boldsymbol{u} \in \mathbb{R}^p$ such that $< \boldsymbol{u}, \boldsymbol{u} >= 1$.

Let

$$\lambda_1 = \sup_{\boldsymbol{u} \in \mathbb{R}^p} \mathrm{Var}(< \boldsymbol{u}, \boldsymbol{X} >) = \mathrm{Var}(< \boldsymbol{u}_1, \boldsymbol{X} >) = \boldsymbol{u}_1' \Sigma \boldsymbol{u}_1,$$

where $< \boldsymbol{u}_1, \boldsymbol{u}_1 >= \boldsymbol{u}_1' \boldsymbol{u}_1 = 1$.

The random variable U_1 will be called the first principal component, and the vector $\boldsymbol{u}_1$ will be called the vector of weights of the first principal component.

In the next step we seek a variable $U_2 =< \boldsymbol{u}_2, \boldsymbol{X} >= \boldsymbol{u}_2' \boldsymbol{X}$ which is not correlated with the first principal component U_1 and which has maximum variance. We continue this process until we obtain p new variables $U_1, U_2, \ldots, U_p$ (principal components).

In general, the kth principal component $U_k =< \boldsymbol{u}_k, \boldsymbol{X} >= \boldsymbol{u}_k' \boldsymbol{X}$ satisfies the conditions:

$$\lambda_k = \sup_{\boldsymbol{u} \in \mathbb{R}^p} \mathrm{Var}(< \boldsymbol{u}, \boldsymbol{X} >) = \mathrm{Var}(< \boldsymbol{u}_k, \boldsymbol{X} >) = \boldsymbol{u}_k' \Sigma \boldsymbol{u}_k,$$

$$< \boldsymbol{u}_{\kappa_1}, \boldsymbol{u}_{\kappa_2} >= \delta_{\kappa_1 \kappa_2}, \qquad \kappa_1, \kappa_2 = 1, ..., k.$$

The expression $(\lambda_k, \boldsymbol{u}_k)$ will be called the kth principal system of the variable $\boldsymbol{X}$ (Jolliffe (2002)).

It can be shown that $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_p$ and $\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_p$ are the eigenvalues and corresponding eigenvectors of the covariance matrix $\boldsymbol{\Sigma}$.

In practice this matrix is unknown, and must be estimated from the sample. Let $\boldsymbol{x} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n)$ be realizations of the vector $\boldsymbol{X}$.

Then

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{n}\boldsymbol{x}\boldsymbol{x}'.$$

Moreover, let $\hat{\lambda}_1 \geqslant \hat{\lambda}_2 \geqslant \cdots \geqslant \hat{\lambda}_p$ and $\hat{\boldsymbol{u}}_1, \hat{\boldsymbol{u}}_2, \ldots, \hat{\boldsymbol{u}}_p$ be eigenvalues and corresponding eigenvectors of the matrix $\hat{\boldsymbol{\Sigma}}$.

Then $(\hat{\lambda}_k, \hat{\boldsymbol{u}}_k)$ is called the kth principal system of the sample of the vector $\boldsymbol{X}$.

The coordinates of the projection of the ith realization $\boldsymbol{x}_i$ of the vector $\boldsymbol{X}$ on the kth principal component are equal to:

$$\hat{U}_{ik} = < \hat{\boldsymbol{u}}_k, \boldsymbol{x}_i > = \hat{\boldsymbol{u}}_k' \boldsymbol{x}_i,$$

for $i = 1, 2, ..., n, k = 1, 2, ..., p$. Finally, the coordinates of the projection of the i th realization $\boldsymbol{x}_i$ of the vector $\boldsymbol{X}$ on the plane of the first two principal components from the sample are equal to $(\hat{\boldsymbol{u}}_1' \boldsymbol{x}_i, \hat{\boldsymbol{u}}_2' \boldsymbol{x}_i), i = 1, 2, ..., n$.

3. Multivariate functional principal component analysis (MFPCA)

The functional case of PCA (FPCA) is a more informative way of looking at the variability structure in the variance-covariance function for one-dimensional functional data (Górecki and Krzyśko (2012)). In this section we present PCA for multivariate functional data (MFPCA) (Jacques and Preda (2014)).

Suppose that we are observing a p-dimensional stochastic process $\boldsymbol{X}(t) = (X_1(t), X_2(t), ..., X_p(t))'$, with continuous parameter $t \in I$. We will further assume that $\mathrm{E}(\boldsymbol{X}(t)) = \boldsymbol{0}$ and $\boldsymbol{X}(t) \in L_2^p(I)$, where $L_2(I)$ is a Hilbert space of square integrable functions on the interval I equipped with the following inner product:

$$< \boldsymbol{u}(t), \boldsymbol{v}(t) > = \int_I \boldsymbol{u}'(t)\boldsymbol{v}(t)dt.$$

Moreover, assume that the kth component of the process $\boldsymbol{X}(t)$ can be represented by a finite number of orthonormal basis functions $\{\varphi_b\}$

$$X_k(t) = \sum_{b=0}^{B_k} c_{kb}\varphi_b(t), t \in I, k = 1, 2, ..., p,$$

where c_{kb} are random variables such that $\mathrm{E}(c_{kb}) = 0$, $\mathrm{Var}(c_{kb}) < \infty$, $k = 1, 2, ..., p, b = 0, ..., B_k$.

Let

$$\boldsymbol{c} = (c_{10}, ..., c_{1B_1}, ..., c_{p0}, ..., c_{pB_p})',$$

$$\boldsymbol{\Phi}(t) = \begin{bmatrix} \boldsymbol{\varphi}_1'(t) & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\varphi}_2'(t) & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \boldsymbol{\varphi}_p'(t) \end{bmatrix}, \tag{1}$$

where $\boldsymbol{\varphi}_k(t) = (\varphi_0(t), ..., \varphi_{B_k}(t))'$, $k = 1, ..., p$.

Then, the process $\boldsymbol{X}(t)$ can be represented as

$$\boldsymbol{X}(t) = \boldsymbol{\Phi}(t)\boldsymbol{c},\ t \in I,\ \mathrm{E}(\boldsymbol{c}) = \mathbf{0},\ \mathrm{Var}(\boldsymbol{c}) = \Sigma_{\boldsymbol{c}}.$$

We are interested to find the inner product

$$U = < \boldsymbol{u}(t), \boldsymbol{X}(t) > = \int_I \boldsymbol{u}'(t)\boldsymbol{X}(t)dt$$

having maximal variance for all $\boldsymbol{u}(t) \in L_2^p(I)$ such that $< \boldsymbol{u}(t), \boldsymbol{u}(t) > = 1$. It may be assumed that the vector weight function $\boldsymbol{u}(t)$ and the process $\boldsymbol{X}(t)$ are in the same space, i.e. the function $\boldsymbol{u}(t)$ can be written in the form:

$$\boldsymbol{u}(t) = \boldsymbol{\Phi}(t)\boldsymbol{u},$$

where $\boldsymbol{u} \in \mathbb{R}^{K+p}$, $K = B_1 + ... + B_p$. Then

$$< \boldsymbol{u}(t), \boldsymbol{X}(t) > = < \boldsymbol{\Phi}(t)\boldsymbol{u}, \boldsymbol{\Phi}(t)\boldsymbol{c} > = \boldsymbol{u}' < \boldsymbol{\Phi}(t), \boldsymbol{\Phi}(t) > \boldsymbol{c} = \boldsymbol{u}'\boldsymbol{c}$$

and

$$\mathrm{E}(< \boldsymbol{u}(t), \boldsymbol{X}(t) >) = \boldsymbol{u}'E(\boldsymbol{c}) = \boldsymbol{u}'\mathbf{0} = 0,$$

$$\mathrm{Var}(< \boldsymbol{u}(t), \boldsymbol{X}(t) >) = \boldsymbol{u}'E(\boldsymbol{c}\boldsymbol{c}')\boldsymbol{u} = \boldsymbol{u}'\boldsymbol{\Sigma}_{\boldsymbol{c}}\boldsymbol{u}.$$

Let

$$\lambda_1 = \sup_{\boldsymbol{u}(t) \in L_2^p(I)} \mathrm{Var}(< \boldsymbol{u}(t), \boldsymbol{X}(t) >) = \mathrm{Var}(< \boldsymbol{u}_1(t), \boldsymbol{X}(t) >) = \boldsymbol{u}_1'\boldsymbol{\Sigma}_{\boldsymbol{c}}\boldsymbol{u}_1,$$

where $< \boldsymbol{u}_1(t), \boldsymbol{u}_1(t) > = \boldsymbol{u}_1'\boldsymbol{u}_1 = 1$.
The inner product $U_1 = < \boldsymbol{u}_1(t), \boldsymbol{X}(t) > = \boldsymbol{u}_1'\boldsymbol{c}$ will be called the first principal component, and the vector function $\boldsymbol{u}_1(t)$ will be called the first vector weight function. Subsequently we look for the second principal component $U_2 = < \boldsymbol{u}_2(t), \boldsymbol{X}(t) > = \boldsymbol{u}_2'\boldsymbol{c}$, maximizing $\mathrm{Var}(< \boldsymbol{u}(t), \boldsymbol{X}(t) >) = \boldsymbol{u}'\boldsymbol{\Sigma}_{\boldsymbol{c}}\boldsymbol{u}$, such that $< \boldsymbol{u}_2(t), \boldsymbol{u}_2(t) > = \boldsymbol{u}_2'\boldsymbol{u}_2 = 1$, and not correlated with the first functional principal component U_1,i.e. subject to the restriction $< \boldsymbol{u}_1(t), \boldsymbol{u}_2(t) > = \boldsymbol{u}_1'\boldsymbol{u}_2 = 0$.

In general, the kth functional principal component $U_k = < \boldsymbol{u}_k(t), \boldsymbol{X}(t) > = \boldsymbol{u}_k'\boldsymbol{c}$ satisfies the conditions:

$$\lambda_k = \sup_{\boldsymbol{u}(t) \in L_2^p(I)} \mathrm{Var}(< \boldsymbol{u}(t), \boldsymbol{X}(t) >) = \mathrm{Var}(< \boldsymbol{u}_k(t), \boldsymbol{X}(t) >) = \boldsymbol{u}_k'\boldsymbol{\Sigma}_{\boldsymbol{c}}\boldsymbol{u}_k,$$

$$< \boldsymbol{u}_{\kappa_1}(t), \boldsymbol{u}_{\kappa_2}(t) > = \delta_{\kappa_1\kappa_2}, \qquad \kappa_1, \kappa_2 = 1, ..., k.$$

The expression $(\lambda_k, \boldsymbol{u}_k(t))$ will be called the *k*th principal system of the process $\boldsymbol{X}(t)$.

Now, let us consider the principal components of the random vector $\boldsymbol{c}$. The k th principal component $U_k^* =< \boldsymbol{u}_k, \boldsymbol{c} >$ of this vector satisfies the conditions:

$$\gamma_k = \sup_{\boldsymbol{u} \in \mathbb{R}^{K+p}} Var(< \boldsymbol{u}, \boldsymbol{c} >) = \sup_{\boldsymbol{u} \in \mathbb{R}^{K+p}} \boldsymbol{u}' Var(\boldsymbol{c}) \boldsymbol{u} = \sup_{\boldsymbol{u} \in \mathbb{R}^{K+p}} \boldsymbol{u}' \boldsymbol{\Sigma_c} \boldsymbol{u},$$

$$\boldsymbol{u}'_{\kappa_1} \boldsymbol{u}_{\kappa_2} = \delta_{\kappa_1 \kappa_2},$$

where $\kappa_1, \kappa_2 = 1, ..., k$, $K = B_1 + ... + B_p$. The expression $(\gamma_k, \boldsymbol{u}_k)$ will be called the *k*th principal system of the vector $\boldsymbol{c}$.

Determining the *k*th principal system of the vector $\boldsymbol{c}$ is equivalent to solving for the eigenvalue and corresponding eigenvectors of the covariance matrix $\boldsymbol{\Sigma_c}$ of that vector, standardized so that $\boldsymbol{u}'_{\kappa_1} \boldsymbol{u}_{\kappa_2} = \delta_{\kappa_1 \kappa_2}$.

From the above considerations, we have the following theorem:

Theorem. The *k*th principal system $(\lambda_k, \boldsymbol{u}_k(t))$ of the stochastic process $\boldsymbol{X}(t)$ is related to the *k*th principal system $(\gamma_k, \boldsymbol{u}_k)$ of the random vector $\boldsymbol{c}$ by the equations:

$$\lambda_k = \gamma_k, \qquad \boldsymbol{u}_k(t) = \boldsymbol{\Phi}(t) \boldsymbol{u}_k, \qquad t \in I,$$

where $k = 1, ..., K + p$, $K = B_1 + B_2 + \cdots + B_p$.

Principal component analysis for random vectors $\boldsymbol{c}$ is based on the matrix $\boldsymbol{\Sigma_c}$. In practice this matrix is unknown. We estimate it on the basis of n independent realizations $\boldsymbol{x}_1(t), \boldsymbol{x}_2(t), ..., \boldsymbol{x}_n(t)$ of the random process $\boldsymbol{X}(t)$.

Typically data are recorded at discrete moments in time. The process of transformation of discrete data to functional data is performed for each variable $X_1, X_2, \ldots X_p$ separately.

Let x_{kj} denote an observed value of feature X_k, $k = 1, 2, \ldots p$ at the *j*th time point t_j, where $j = 1, 2, ..., J$. Then our data consist of pJ pairs of (t_j, x_{kj}). This discrete data can be smoothed by continuous functions $x_k(t)$, where $t \in I$ (Ramsay and Silverman (2005)). Let I be a compact set such that $t_j \in I$, for $j = 1, ..., J$. Let us assume that the function $x_k(t)$ has the following representation

$$x_k(t) = \sum_{b=0}^{B_k} c_{kb} \varphi_b(t), \; t \in I, \; k = 1, \ldots, p, \tag{2}$$

where $\{\varphi_b\}$ are orthonormal basis functions, and $c_{k0}, c_{k1}, \ldots, c_{kB_k}$ are the coefficients.

Let $\boldsymbol{x}_k = (x_{k1}, x_{k2}, \ldots, x_{kJ})'$, $\boldsymbol{c}_k = (c_{k0}, c_{k1}, \ldots, c_{kB_k})'$ and $\boldsymbol{\Phi}_k(t)$ be a matrix of dimension $J \times (B_k + 1)$ containing the values $\varphi_b(t_j)$, $b = 0, 1, ..., B$, $j = 1, 2, ..., J$, $k = 1, \ldots, p$. The coefficient $\boldsymbol{c}_k$ in (2) is estimated by the least squares method, that is, so as to minimize the function:

$$S(\boldsymbol{c}_k) = (\boldsymbol{x}_k - \boldsymbol{\Phi}_k(t) \boldsymbol{c}_k)' (\boldsymbol{x}_k - \boldsymbol{\Phi}_k(t) \boldsymbol{c}_k), \; k = 1, \ldots, p.$$

Differentiating $S(\boldsymbol{c}_k)$ with respect to the vector $\boldsymbol{c}_k$, we obtain the least squares method estimator

$$\hat{\boldsymbol{c}_k} = \left(\boldsymbol{\Phi}_k'(t)\boldsymbol{\Phi}_k(t)\right)^{-1}\boldsymbol{\Phi}_k'(t)\boldsymbol{x}_k,\ k = 1,\dots,p.$$

The degree of smoothness of the function $x_k(t)$ depends on the value B_k (a small value of B_k causes more smoothing of the curves). The optimum value for B_k may be selected using the Bayesian information criterion BIC (see Shmueli (2010)).

Let us assume that there are n independent pairs of values (t_j, x_{kij}), $k = 1,\dots,p,\ i = 1,...,n,\ j = 1,...,J$. These discrete data are smoothed to continuous functions in the following form:

$$x_{ki}(t) = \sum_{b=0}^{B_{ki}} \hat{c}_{kib}\varphi_b(t),\ k = 1,\dots,p,\ i = 1,...,n,\ t \in I.$$

Among all the $B_{k1}, B_{k2}, ..., B_{kn}$ one common value of B_k is chosen, as the modal value of the numbers $B_{k1}, B_{k2}, ..., B_{kn}$, and we assume that each function $x_{ki}(t)$ has the form

$$x_{ki}(t) = \sum_{b=0}^{B_{k}} \hat{c}_{kib}\varphi_b(t),\ k = 1,\dots,p,\ i = 1,...,n,\ t \in I.$$

The data $\{x_{k1}(t), ..., x_{kn}(t)\}$ are called functional data (see Ramsay and Silverman (2005)).

Finally, each of n independent realizations $\boldsymbol{x}_1(t), \boldsymbol{x}_2(t), ..., \boldsymbol{x}_n(t)$ has the form $\boldsymbol{x}_i(t) = \boldsymbol{\Phi}(t)\hat{\boldsymbol{c}}_i$ where $\boldsymbol{\Phi}(t)$ is given by (1) and the vectors $\hat{\boldsymbol{c}}_i = (\hat{c}_{10}, ..., \hat{c}_{1B_1}, ..., \hat{c}_{p0}, ..., \hat{c}_{pB_p})'$ are centred, $i = 1, 2, ..., n$.

Let $\hat{\boldsymbol{C}} = (\hat{\boldsymbol{c}}_1, \hat{\boldsymbol{c}}_2, ..., \hat{\boldsymbol{c}}_n)$. Then

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{c}} = \frac{1}{n}\hat{\boldsymbol{C}}\hat{\boldsymbol{C}}'.$$

Let $\hat{\gamma}_1 \geqslant \hat{\gamma}_2 \geqslant ... \geqslant \hat{\gamma}_s$ be non-zero eigenvalues of the matrix $\hat{\boldsymbol{\Sigma}}_{\boldsymbol{c}}$, and $\hat{\boldsymbol{u}}_1, \hat{\boldsymbol{u}}_2, ..., \hat{\boldsymbol{u}}_s$ the corresponding eigenvectors, where $s =$rank$(\hat{\boldsymbol{\Sigma}}_{\boldsymbol{c}})$.

Moreover, the kth principal system of the random process $\boldsymbol{X}(t)$ determined from the sample has the following form:

$$(\hat{\lambda}_k = \hat{\gamma}_k, \hat{\boldsymbol{u}}_k(t) = \boldsymbol{\Phi}(t)\hat{\boldsymbol{u}}_k), \qquad k = 1, ..., s.$$

The coordinates of the projection of the ith realization $\boldsymbol{x}_i(t)$ of the process $\boldsymbol{X}(t)$ on the kth functional principal component are equal to:

$$\hat{U}_{ik} =< \hat{\boldsymbol{u}}_k(t), \boldsymbol{x}_i(t) >=< \boldsymbol{\Phi}(t)\hat{\boldsymbol{u}}_k, \boldsymbol{\Phi}(t)\hat{\boldsymbol{c}}_i >= \hat{\boldsymbol{u}}_k' < \boldsymbol{\Phi}(t), \boldsymbol{\Phi}(t) > \hat{\boldsymbol{c}}_i = \hat{\boldsymbol{u}}_k'\hat{\boldsymbol{c}}_i,$$

for $i = 1, 2, ..., n, k = 1, 2, ..., s$. Finally, the coordinates of the projection of the i th realization $\boldsymbol{x}_i(t)$ of the process $\boldsymbol{X}(t)$ on the plane of the first two functional principal components from the sample are equal to $(\hat{\boldsymbol{u}}_1'\hat{\boldsymbol{c}}_i, \hat{\boldsymbol{u}}_2'\hat{\boldsymbol{c}}_i), i = 1, 2, ..., n$.

4. Example

Data relating to environmental protection were obtained from Professor W. Ratajczak of the Spatial Econometry Group at the Geographical and Geological Sciences Faculty of Adam Mickiewicz University, Poznań. The analysis relates to the 16 Polish provinces ($n = 16$). On the graphs, the provinces are denoted by numbers as given in Table 1.

Table 1. Designations of provinces

1	ŁÓDZKIE
2	MAZOWIECKIE
3	MAŁOPOLSKIE
4	ŚLĄSKIE
5	LUBELSKIE
6	PODKARPACKIE
7	PODLASKIE
8	ŚWIĘTOKRZYSKIE
9	LUBUSKIE
10	WIELKOPOLSKIE
11	ZACHODNIOPOMORSKIE
12	DOLNOŚLĄSKIE
13	OPOLSKIE
14	KUJAWSKO-POMORSKIE
15	POMORSKIE
16	WARMIŃSKO-MAZURSKIE

The analyzed data cover a period of 10 years, from 2002 to 2011 ($J = 10$). Each province was characterized by a group of 6 features ($p = 6$):

1. Gaseous pollutant emissions [t/km^2]
2. Dust pollutant emissions [kg/km^2]
3. Solid waste produced [t/km^2]
4. Total liquid waste [dam^3/1000 residents]
5. Industrial liquid waste [dam^3/1000 residents]
6. Household and industrial water consumption [dam^3/1000 residents]

The classical method of principal component analysis (PCA) permits only separate analysis for each year of observation. Tables 2–5 contain the weights and the percentage contributions for the first and second principal component.

Table 2. Weights (eigenvectors) of the first principal component (analysis for a fixed time)

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
1	0.6181	0.6480	0.6404	0.6535	0.6985	0.7076	0.7564	0.7514	0.7591	0.7748
2	0.4230	0.3964	0.3612	0.3140	0.2773	0.3042	0.2200	0.1962	0.2017	0.1884
3	0.6609	0.6486	0.6762	0.6869	0.6579	0.6363	0.6144	0.6283	0.6180	0.6022
4	0.0013	0.0010	0.0007	0.0008	0.0008	0.0006	0.0008	0.0009	0.0006	0.0005
5	-0.0333	-0.0326	-0.0316	-0.0331	-0.0320	-0.0289	-0.0297	-0.0303	-0.0215	-0.0257
6	-0.0347	-0.0338	-0.0344	-0.0370	-0.0372	-0.0327	-0.0339	-0.0358	-0.0274	-0.0300

Table 3. Percentage contribution of the original variables in the structure of the first principal component (analysis for a fixed time)

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
1	38.2048	41.9904	41.0112	42.7062	48.7902	50.0698	57.2141	56.4602	57.6233	60.0315
2	17.8929	15.7133	13.0465	9.8596	7.6895	9.2538	4.8400	3.8494	4.0683	3.5495
3	43.6789	42.0682	45.7246	47.1832	43.2832	40.4878	37.7487	39.4761	38.1924	36.2645
4	0.0002	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0000
5	0.1109	0.1063	0.0999	0.1096	0.1024	0.0835	0.0882	0.0918	0.0462	0.0660
6	0.1204	0.1142	0.1183	0.1369	0.1384	0.1069	0.1149	0.1282	0.0751	0.0900

Table 4. Weights (eigenvectors) of the second principal component (analysis for a fixed time)

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
1	0.7697	0.7405	0.6493	0.7145	0.3636	0.4864	0.4985	0.4124	0.3529	0.3770
2	-0.2033	-0.1625	0.0214	0.0060	0.0548	0.0398	0.0368	0.0161	-0.0100	-0.0103
3	-0.5764	-0.6277	-0.5937	-0.6673	-0.3448	-0.5116	-0.5798	-0.4388	-0.3824	-0.4280
4	0.0013	0.0022	0.0024	0.0023	0.0018	0.0008	0.0010	0.0005	0.0022	0.0018
5	0.1235	0.1194	0.3322	0.1459	0.5975	0.4978	0.4560	0.5652	0.6065	0.5830
6	0.1368	0.1304	0.3393	0.1511	0.6237	0.5022	0.4539	0.5636	0.6010	0.5785

Table 5. Percentage contribution of the original variables in the structure of the second principal component (analysis for a fixed time)

	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011
1	59.2438	54.8340	42.1590	51.0510	13.2205	23.6585	24.8502	17.0074	12.4538	14.2129
2	4.1331	2.6406	0.0458	0.0036	0.3003	0.1584	0.1354	0.0259	0.0100	0.0106
3	33.2237	39.4007	35.2480	44.5289	11.8887	26.1735	33.6168	19.2545	14.6230	18.3184
4	0.0002	0.0005	0.0006	0.0005	0.0003	0.0001	0.0001	0.0000	0.0005	0.0003
5	1.5252	1.4256	11.0357	2.1287	35.7006	24.7805	20.7936	31.9451	36.7842	33.9889
6	1.8714	1.7004	11.5124	2.2831	38.9002	25.2205	20.6025	31.7645	36.1201	33.4662

The relative position of the 16 provinces (in 2002 and 2011) in the system of the first two principal components is shown in Figure 1.

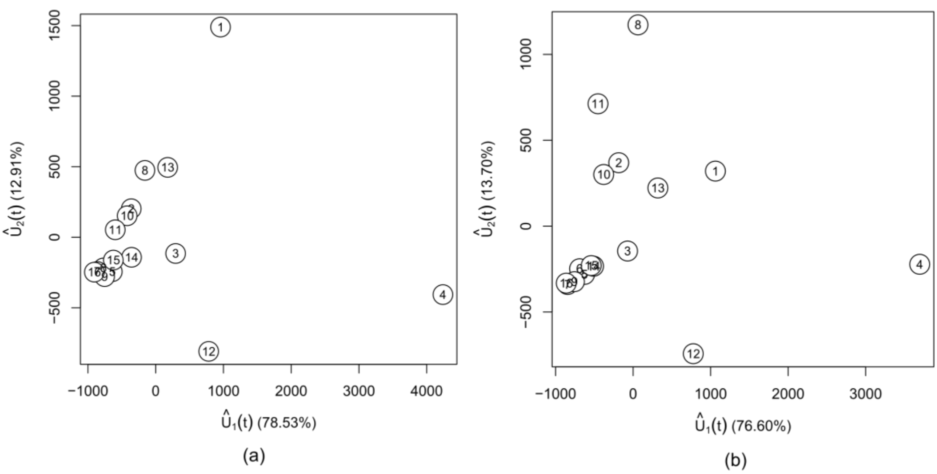

Figure 1. Projection of the six-dimensional vectors representing the 16 provinces on the plane of the first two principal components, (a) year 2002, (b) year 2011

The functional principal components method enables combined analysis of the data for the whole of the studied period of time. The data were transformed to functional data by the method described in Section 3. The calculations were performed using the Fourier basis. The time interval $[0,T]=[0,10]$ was divided into moments of time in the following way: $t_1=0.5(2002)$, $t_2=1.5(2003),\ldots,$ $t_{10}=9.5(2011)$. Moreover, in view of the small number of time periods ($J=10$), for each variable the maximum number of basis components was taken, equal to $B_1 = \cdots = B_{10} = 9$.

Tables 6–7 show the coefficients of the weight functions for the first and second functional principal components.

Table 6. Coefficients of weight functions for the first functional principal component

	$\hat{u}_0$	$\hat{u}_1$	$\hat{u}_2$	$\hat{u}_3$	$\hat{u}_4$	$\hat{u}_5$	$\hat{u}_6$	$\hat{u}_7$	$\hat{u}_8$	Area
1	0.6947	-0.0007	-0.0179	-0.0056	0.0189	-0.0220	-0.0104	-0.0064	-0.0046	2.1968
2	0.2927	0.0794	0.0094	0.0401	0.0138	0.0114	-0.0101	0.0299	0.0011	0.9256
3	0.6443	0.0551	-0.0088	0.0086	0.0129	-0.0010	-0.0074	0.0147	-0.0002	2.0375
4	0.0008	0.0001	0.0000	0.0001	0.0000	0.0001	0.0000	0.0001	0.0000	0.0025
5	-0.0317	0.0001	0.0009	-0.0005	-0.0009	0.0003	0.0003	0.0009	-0.0004	0.1002
6	-0.0355	0.0013	0.0011	0.0006	-0.0008	0.0003	0.0004	0.0013	-0.0004	0.1123

Table 7. Coefficients of weight functions for the second functional principal component

	$\hat{u}_0$	$\hat{u}_1$	$\hat{u}_2$	$\hat{u}_3$	$\hat{u}_4$	$\hat{u}_5$	$\hat{u}_6$	$\hat{u}_7$	$\hat{u}_8$	Area
1	0.5771	-0.0005	0.0680	0.0392	0.0193	0.0165	-0.0105	-0.0091	-0.0114	1.8250
2	-0.0122	-0.0227	-0.0289	-0.0087	-0.0092	-0.0140	0.0010	-0.0157	0.0005	0.1021
3	-0.5636	-0.0314	0.0210	-0.0042	-0.0070	0.0018	0.0017	-0.0078	0.0034	1.7823
4	0.0017	0.0004	0.0003	-0.0001	-0.0002	-0.0002	-0.0003	-0.0001	-0.0002	0.0054
5	0.4080	-0.0210	0.0120	-0.0081	0.0096	-0.0065	-0.0155	-0.0117	0.0055	1.2902
6	0.4120	-0.0164	0.0081	-0.0069	0.0122	-0.0041	-0.0163	-0.0115	0.0072	1.3029

At a given time point *t*, the greater is the absolute value of a component of the vector weight, the greater is the contribution in the structure of the given functional principal component, from the process $X(t)$ corresponding to that component. The total contribution of a particular primary process $X_i(t)$ in the structure of a particular functional principal component is equal to the area under the module weighting function corresponding to this process. These contributions for the six components of the vector process $\boldsymbol{X}(t)$, and the first and second functional principal components are given in Tables 6–7.

Figure 2 shows the six weight functions for the first and second functional principal components.

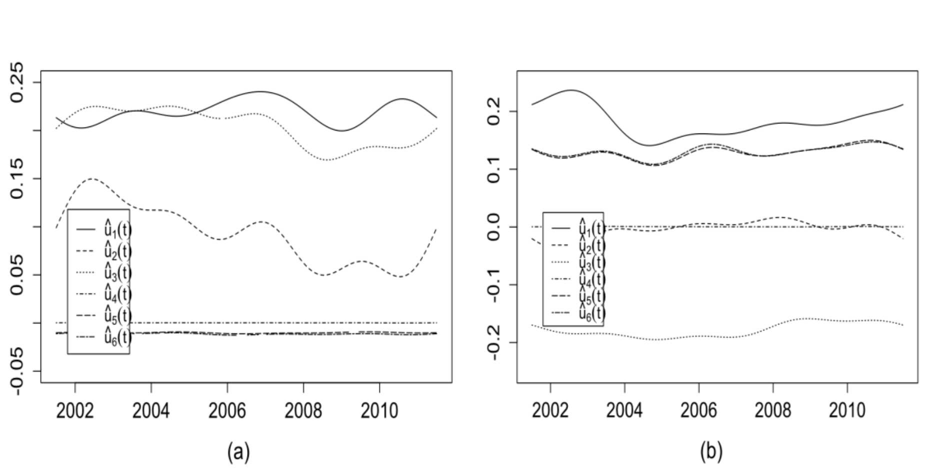

Figure 2. Weight functions for the first (a) and second (b) functional principal component (MFPCA)

The relative positions of the 16 provinces in the system of the first two functional principal components are shown in Figure 3. The system of the first two functional principal components retains 90.33% of the total variation.

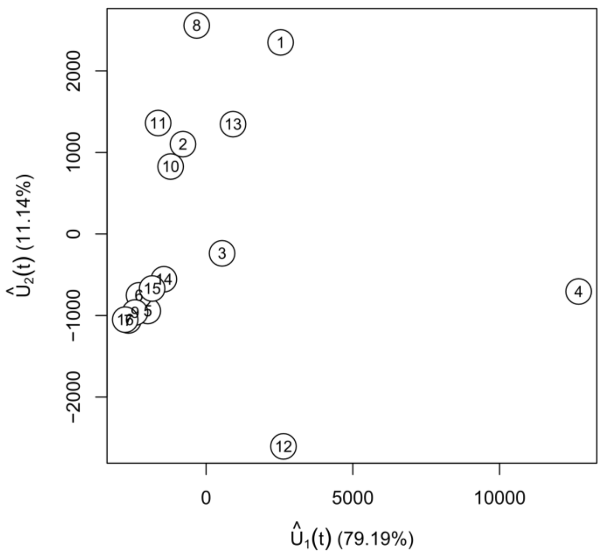

Figure 3. Projection of multidimensional functional data representing the 16 provinces on the plane of the first two functional principal components

5. Conclusions

This paper introduces and analyzes a new method of constructing principal components for multivariate functional data. This method was applied to environmental multivariate time series concerning the Polish provinces. Our research has shown, on this example, that the use of a multivariate functional principal components analysis gives good results. Of course, the performance of the algorithm needs to be further evaluated on additional real and artificial data sets. In a similar way, we can extend similar methods like functional discriminant coordinates (Górecki et al. (2014)) and canonical correlation analysis (Krzyśko, Waszak (2013)) to multivariate case. This is the direction of our future research.

REFERENCES

BESSE, P., (1979). Etude descriptive d'un processus. Ph.D. thesis, Université Paul Sabatier.

FISHER, R. A., (1936). The use of multiple measurements in taxonomic problem, Annals of Eugenics 7, 179–188.

GÓRECKI, T., KRZYŚKO, M., (2012). Functional Principal Components Analysis, Data analysis methods and its applications, C.H. Beck 71–87.

GÓRECKI, T., KRZYŚKO, M., WASZAK, Ł., (2014). Functional Discriminant Coordinates, Communication in Statistics – Theory and Methods 43(5), 1013–1025.

JACQUES, J., PREDA, C., (2014). Model-based clustering for multivariate functional data, Computational Statistics & Data Analysis 71, 92–106.

JOLLIFFE, I. T., (2002). Principal Component Analysis, Second Edition, Springer.

KRZYŚKO, M., WASZAK, Ł., (2013). Canonical correlation analysis for functional data, Biometrical Letters 50(2), 95–105.

RAMSAY, J. O., SILVERMAN, B. W., (2005). Functional Data Analysis, Second Edition, Springer.

SAPORTA, G., (1981). Methodes exploratoires d'analyse de données temporelles. Cahiers du Buro. Ph.D. thesis.

SHMUELI, G., (2010). To explain or to predict? Statistical Science 25(3), 289–310.