Colloquium Biometricum 41
2011, 5–17

PRINCIPAL COMPONENTS ANALYSIS FOR FUNCTIONAL DATA

Karol Deręgowski[1], Mirosław Krzyśko[1,2]

[1]President Stanislaw Wojciechowski Higher Vocational State School in Kalisz
Institute of Management, Nowy Świat 4, 62–800 Kalisz, Poland
[2]Adam Mickiewicz University, Faculty of Mathematics and Computer Science
Umultowska 87, 61–614 Poznań, Poland
e–mail: kadere@o2.pl; mkrzysko@amu.edu.pl

This paper is dedicated to the memory of Professor Wiktor Oktaba.

Summary

In this paper we present the construction of functional principal components and show that the problem of FPCA is reduced to multivariate PCA performed on some covariance matrix. An example concerning with data on how children walk is also presented.

Key words and phrases: functional principal components, orthonormal basis, Bayesian Information Criterion, least squares method

Classification AMS 2010: 62H25

1. Introduction

When data are observed as a function of time (such as financial time series, temperature recorded by some central source, etc.), we refer to them as functional data (see Ramsay and Dalzell, 1991). Since in many statistical applications realizations of continuous time series are available as observations of a process recorded in discrete time intervals, one crucial step is to convert discrete data to continuous functions, that is from vectors to curves. Functional data are a very convenient approach to dealing with data depending on time, providing theoretical tools which are indispensable for analyzing observations of the process recorded in discrete time intervals and converting them to continuous functions.

Typically, the sample data contain a number of *N* independent replications of form

$$\{x_i(t), \quad i = 1, 2, \ldots, N, \quad t \in [0, T]\} \tag{1.1}$$

and the record of replication $x_i(t), \quad i = 1,2,\ldots,N$ might consist of J_i pairs $\{t_{ij}, y_{ij}\}$, $j = 1,2,\ldots J_i$, where t_{ij} denotes the argument, and y_{ij} the observed functional value. The choice of t_{ij} is very nonrestrictive, e.g. the argument values may vary between the records and need not be equally spaced. Furthermore, the number of observations J_i can differ between the records. But nevertheless, the argument should lie in the range of values of interest, that means $t_{ij} \in [0, T]$ for all *i*, *j*.

Normally the construction of the functional observations $x_i(t)$ using the discrete data y_{ij} takes place separately or independently for each record *i*. Therefore we will simplify notation by assuming that a simple function $x(t)$ is being estimated.

The first task is to convert the values $y_1, y_2, \ldots, y_J$ to a function x with values $x(t)$ computable for any *t*, called functional objects. The conversion from discrete data to functions may involve smoothing (Ramsay and Silverman, 2005). One smoothing procedure often used involves obtaining a representative function $x(t)$ as the linear combination of *K*+1 base orthonormal functions φ_k:

$$\left\{ x(t) = \sum_{k=0}^{K} c_k \varphi_k(t), \quad t \in [0, T] \right\}. \tag{1.2}$$

A simple linear smoother is obtained if we determine the coefficients of the expansion c_k by minimizing the least squares criterion

$$S(c_0, c_1, ..., c_k) = \sum_{j=1}^{J} \left(y_j - \sum_{k=0}^{K} c_k \varphi_k (t_j) \right)^2 . \tag{1.3}$$

The criterion is expressed more clearly in matrix terms as

$$S(\mathbf{c}) = (\mathbf{y} - \mathbf{\Phi c})^{\mathrm{T}} (\mathbf{y} - \mathbf{\Phi c}) ,$$

where vector $\mathbf{c} = (c_0, c_1, \ldots, c_K)^{\mathrm{T}}$ contains the coefficients c_k and $\mathbf{\Phi}$ is a J by K+1 matrix containing the values $\varphi_k(t_j)$.

Taking the derivative of criterion $S(\mathbf{c})$ with respect to $\mathbf{c}$ yields the equation

$$2\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi c} - 2\mathbf{\Phi}^{\mathrm{T}}\mathbf{y} = \mathbf{0}$$

and solving this for $\mathbf{c}$ provides the estimate $\hat{\mathbf{c}}$ that minimizes the least squares solution,

$$\hat{\mathbf{c}} = \left(\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}\right)^{-1} \mathbf{\Phi}^{\mathrm{T}}\mathbf{y} . \tag{1.4}$$

The smoothness degree depends on *K*, since small (large) values of *K* induce more (less) smoothed curves. The optimal number *K* of basic elements was selected by applying the Bayesian Information Criterion (BIC) to each function $x(t)$ separately, and then taking the most frequently occurring value (the modal value) over all functions. The BIC measures goodness of fit (see Shmueli, 2010).

Principal component analysis (PCA) is a standard approach to the exploration of variability in multivariable data. PCA uses an eigenvalue decomposition of the covariance matrix of the data to find directions in the observations space along which the data have the highest variability. For each principal component, the analysis yields a loading vector or weight vector which gives the direction of variability corresponding to that component.

In the functional context, each principal component is specified by a principal component weight function or eigenfunction $u(t)$ defined on the same range of *t* as the functional data.

The construction of functional principal components is described in Section 2. Section 3 contains an example concerning with data on how children walk.

2. Functional principal components

Suppose we observe a sample of the process $X(t) \in L_2([0,T])$, where $L_2([0,T])$ is the Hilbert space of square integrable functions on the interval $[0,T]$, equipped with the scalar product

$$\langle u,v\rangle = \int u(s)\,v(s)\,ds\,. \tag{2.1}$$

Remark. All integrals are taken over the interval $[0,T]$.

Moreover, suppose that $EX(t) = 0$ and assume the existence of the variance $Var_X(t) = E[X^2(t)]$ and covariance $Cov_X(s,t) = E[X(s)X(t)]$ functions of $X(t)$; $s,t \in [0,T]$.

In functional principal components analysis (FPCA) we want to find orthonormal weight functions $u_1, u_2, \ldots$, such that the variance of the linear transformation is maximal.

The weight functions satisfy:

$$\|u_m\|^2 = \int u_m^2(t)\,dt = 1\,,$$

$$\langle u_l, u_m\rangle = \int u_l(t)\,u_m(t)\,dt = 0\,,\ l \neq m\,.$$

The linear combination is:

$$U_m = \langle u_m, X\rangle = \int u_m(t)\,X(t)\,dt\,, \tag{2.2}$$

and the desired weight functions solve:

$$\underset{\langle u_l,\, u_m\rangle = \delta_{lm},\, l \le m}{\arg\max} \; Var\langle u_m, X\rangle , \tag{2.3}$$

or equivalently:

$$\underset{\langle u_l,\, u_m\rangle = \delta_{lm},\, l \le m}{\arg\max} \int\int u_m(s)\, Cov_X(s,t)\, u_m(t)\, ds\, dt \;, \tag{2.4}$$

where δ_{lm} is the Kronecker delta.
The solution is obtained by solving the Fredholm functional eigenequation

$$\int Cov_X(s,t)\, u(t)\, dt = \lambda u(s) . \tag{2.5}$$

The eigenfunctions $u_1, u_2, \ldots$, sorted with respect to the corresponding eigenvalues $\lambda_1 \ge \lambda_2 \ge \ldots$ solve the FPCA problem (2.3). The following relationship holds between eigenvalues and eigenfunctions:

$$\lambda_m = Var(U_m) = Var\left[\int u_m(t)\, X(t)\, dt\right] = \int\int u_m(s)\, Cov_X(s,t)\, u_m(t)\, ds\, dt .$$

In practice, the covariance function $Cov_X(s,t)$ is unknown and must be estimated from the functional dataset (1.1). For the functional sample (1.1), the estimator of the covariance function $Cov_X(s,t)$ has the form:

$$C\hat{o}v_X(s,t) = \vartheta(s,t) = \frac{1}{N}\sum_{i=1}^{N} x_i(s)\, x_i(t) = \frac{1}{N}\boldsymbol{\varphi}^T(s)\mathbf{C}^{\mathrm{T}}\mathbf{C}\boldsymbol{\varphi}(t), \tag{2.6}$$

where

$$\boldsymbol{\varphi}^{\mathrm{T}}(t) = (\varphi_0(t), \varphi_1(t), \ldots, \varphi_K(t)) \quad s,t \in [0,T] \tag{2.7}$$

and $\mathbf{C} = (c_{ik}) \quad i = 1,2,\ldots,N;\; k = 0,1,\ldots,K$.

Dauxois, Pousse and Romain (1982) show that $\left\| Cov_X(s,t) - \hat{C}ov_X(s,t) \right\| \to 0$ with probability one.

Suppose that the weight function $u(t)$ has the expansion

$$u(t) = \sum_{k=0}^{K} d_k \varphi_k(t) = \boldsymbol{\varphi}^{\mathrm{T}}(t)\mathbf{d}, \tag{2.8}$$

where $\boldsymbol{\varphi}^{\mathrm{T}}(t)$ is given by (2.7) and $\mathbf{d} = (d_0, d_1, \ldots, d_K)^{\mathrm{T}}$.
Using this notation we have:

$$\int C\hat{o}v_X(s,t)\,u(t)\,dt = \frac{1}{N}\boldsymbol{\varphi}^{\mathrm{T}}(\mathbf{s})\mathbf{C}^{\mathrm{T}}\mathbf{C}\left[\int \boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)dt\right]\mathbf{d} = \frac{1}{N}\boldsymbol{\varphi}^{\mathrm{T}}(s)\mathbf{C}^{\mathrm{T}}\mathbf{C}\mathbf{d},$$

since

$$\int \boldsymbol{\varphi}(t)\boldsymbol{\varphi}^{\mathrm{T}}(t)dt = \mathbf{I}_{K+1}.$$

Hence we have

$$\frac{1}{N}\boldsymbol{\varphi}^{\mathrm{T}}(s)\mathbf{C}^{\mathrm{T}}\mathbf{C}\mathbf{d} = \lambda\boldsymbol{\varphi}^{\mathrm{T}}(s)\mathbf{d}. \tag{2.9}$$

Multiplying both sides of equation (2.9) by $\boldsymbol{\varphi}(s)$, and then integrating, we obtain

$$\frac{1}{N}\mathbf{C}^{\mathrm{T}}\mathbf{C}\,\mathbf{d} = \lambda\mathbf{d}. \tag{2.10}$$

We see that the problem of FPCA is reduced to multivariate PCA performed on the matrix $\frac{1}{N}\mathbf{C}^{\mathrm{T}}\mathbf{C}$. In practice, the matrix $\mathbf{C}$ is unknown and must be estimated by $\hat{\mathbf{C}} = (\hat{\mathbf{c}}_{ik})$, $i = 1,2,\ldots,N;\quad k = 0,1,\ldots,K$ (see formula (1.4)).
Next, we find the nonzero eigenvalues λ_j and corresponding eigenvectors $\mathbf{d}_j$ of the matrix $\frac{1}{N}\hat{\mathbf{C}}^{\mathrm{T}}\hat{\mathbf{C}}$.
Having determined the eigenvectors $\mathbf{d}_j$ we determine its weight functions

$$u_j(t) = \mathbf{d}_j^{\mathrm{T}} \boldsymbol{\varphi}(t), \quad t \in [0, T]. \tag{2.11}$$

Hence the *j*th functional principal component is equal to

$$\begin{aligned} U_{ij} &= \langle u_j(t), x_i(t) \rangle = \int u_j(t)\, x_i(t)\, dt = \sum_{k=0}^{K} \sum_{l=0}^{K} c_{il} d_{jk} \int \varphi_k(t) \varphi_l(t) dt = \\ &= \sum_{k=0}^{K} c_{ik} d_{jk} = \mathbf{c}_i^{\mathrm{T}} \mathbf{d}_j, \quad i = 1, 2, ..., N, \ j = 1, 2, ... \end{aligned} \tag{2.12}$$

3. Analysis of the gait data

Here we consider gait data described by Olshen et al. (1989) consisting of measurements of the angles made by the hip and by the knee of each of 39 children at twenty time points in a single gait cycle. Thus corresponding to each children is a time series of 20 elements.

The discrete time series for the N = 39 children were centered and then transformed into continuous function in the form (1.2) on the range [0, 1].

The base functions $\varphi_k(t),\ k = 0, 1, \ldots, K$, form a Fourier orthonormal basis in the space $L^2([0,1])$:

$$\varphi_0(t) = 1, \ \ \varphi_{2k-1}(t) = \sqrt{2} \sin 2k\pi t, \ \ \varphi_{2k}(t) = \sqrt{2} \cos 2k\pi t, \ k = 1, 2, \ldots .$$

The values of the coefficients c_k in the expansion (1.2) were estimated by the least squares method. These coefficients form the matrix $\hat{\mathbf{C}} = (\hat{c}_{rs})$.

The optimum number K of base functions $\varphi_k(t)$ in the expansion (1.2) was selected using the Bayesian Information Criterion (BIC). The optimal values of K for each of the 39 functions $x(t)$ are contained in Table 1.
The frequency distribution of the values of K is shown in Table 2.

Hence the joint K for expansions of all functions $x_i(t),\ i = 1, 2, \ldots, 39$, is equal to 4 (for hip) and is equal to 6 (for knee).

Table 1. Optimal *K* for each of the children.

No.	*K*		No.	*K*		No.	*K*	
	hip	knee		hip	knee		hip	knee
1	18	6	14	4	6	27	8	18
2	8	4	15	4	4	28	18	14
3	6	6	16	6	10	29	12	4
4	4	6	17	2	4	30	8	14
5	2	8	18	4	4	31	6	6
6	4	6	19	4	6	32	4	6
7	6	2	20	4	6	33	6	10
8	6	18	21	8	2	34	4	10
9	10	6	22	4	8	35	8	8
10	6	4	23	4	4	36	4	4
11	4	6	24	2	2	37	8	18
12	8	6	25	4	10	38	8	10
13	6	6	26	4	6	39	6	8

Table 2. Frequency distribution of values of *K*.

K		2	4	6	8	10	12	18
Frequency	hip	3	15	9	8	1	1	2
	knee	3	8	14	4	5	2	3

The most frequently considered objects are presented on a plot of the first two functional principal components. In this case the criterion for explanation of variability by the first two functional principal components is the expression $\frac{\lambda_1 + \lambda_2}{\sum \lambda_i} 100\%$, where $\lambda_1 \geq \lambda_2 \geq \ldots$ are the non–zero eigenvalues of the matrix $\frac{1}{N}\hat{\mathbf{C}}^T\hat{\mathbf{C}}$. In our case $\lambda_1 = 31.63$, $\lambda_2 = 5.36$ for hip and $\lambda_1 = 14.79$, $\lambda_2 = 8.72$ for knee. All the non–zero eigenvalues are shown in Figures 1 and 2. The percentage of variability explained by the first two functional principal components is equal to 86.0% for hip and 68.85% for knee.

The eigenvalues λ_1 and λ_2 correspond to the eigenvectors

$\mathbf{d}_{1\,\text{hip}} = (0.978; 0.042; 0.201; -0.031; -0.007)^T$;
$\mathbf{d}_{2\,\text{hip}} = (-0.120; 0.885; 0.373; -0.180; 0.175)^T$;
$\mathbf{d}_{1\,\text{knee}} = (-0.511; -0.118; -0.696; 0.379; 0.238; 0.043; 0.193)^T$;
$\mathbf{d}_{2\,\text{knee}} = (0.795; 0.061; -0.244; 0.536; 0.108; 0.023; 0.073)^T$.

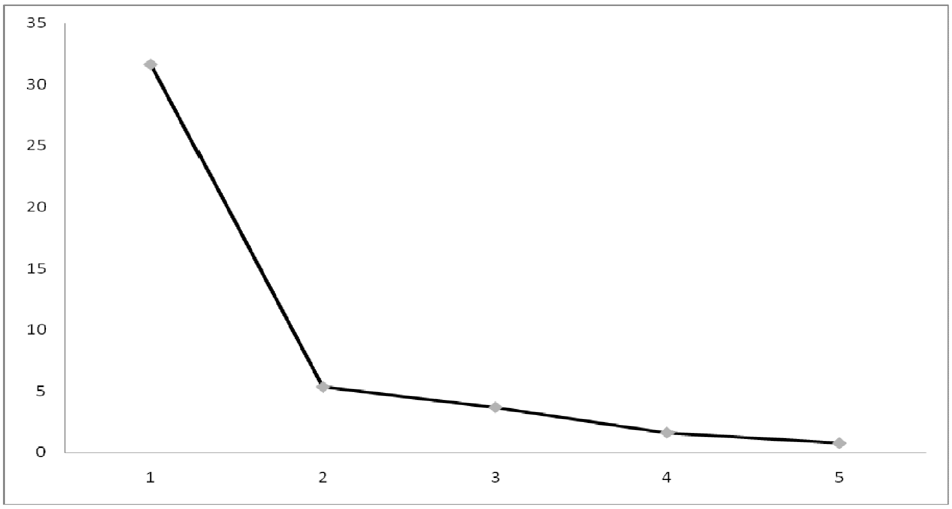

Fig. 1. Scree graph (hip)

Fig. 2. Scree graph (knee)

Figures 3 – 6 show the first two weight functions of the form (2.11) constructed from the vectors $\mathbf{d}_1$ and $\mathbf{d}_2$ and the base functions $\varphi_k(t)$.

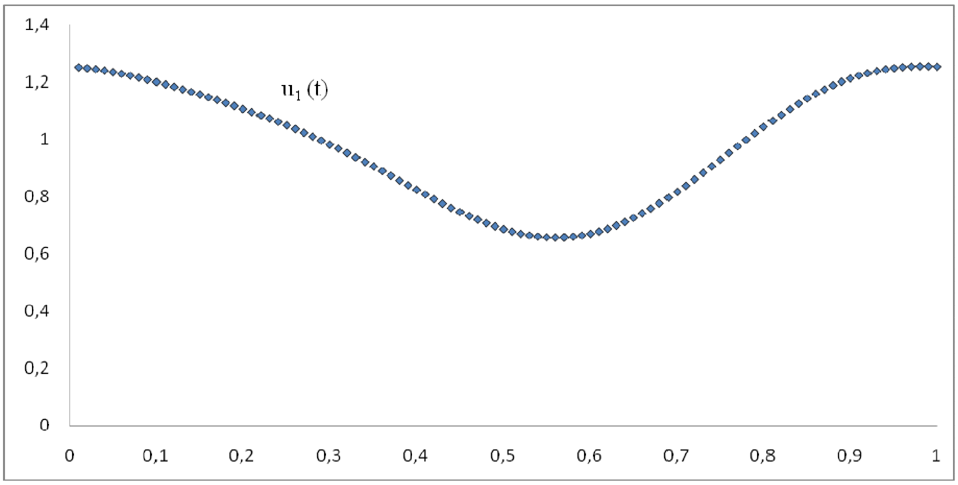

Fig. 3. The weight function $u_{1\,\mathrm{hip}}(t)$ (73.59%)

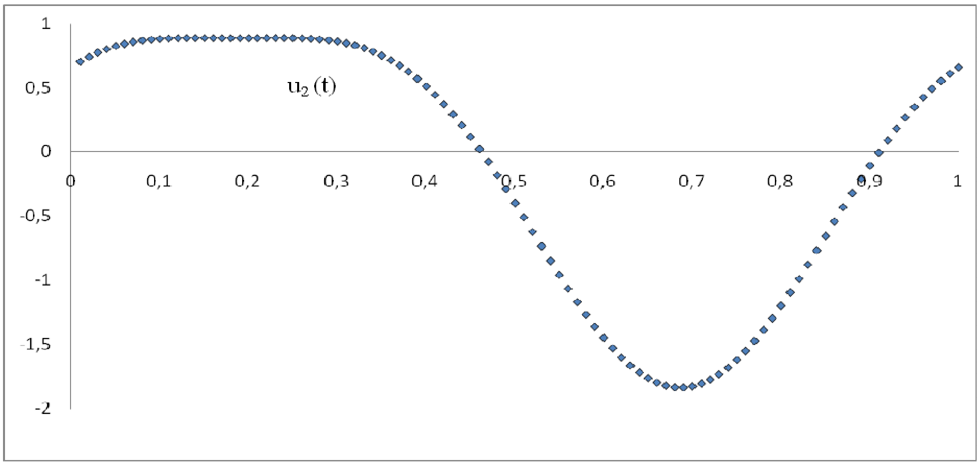

Fig. 4. The weight function $u_{2\,\mathrm{hip}}(t)$ (12.42%)

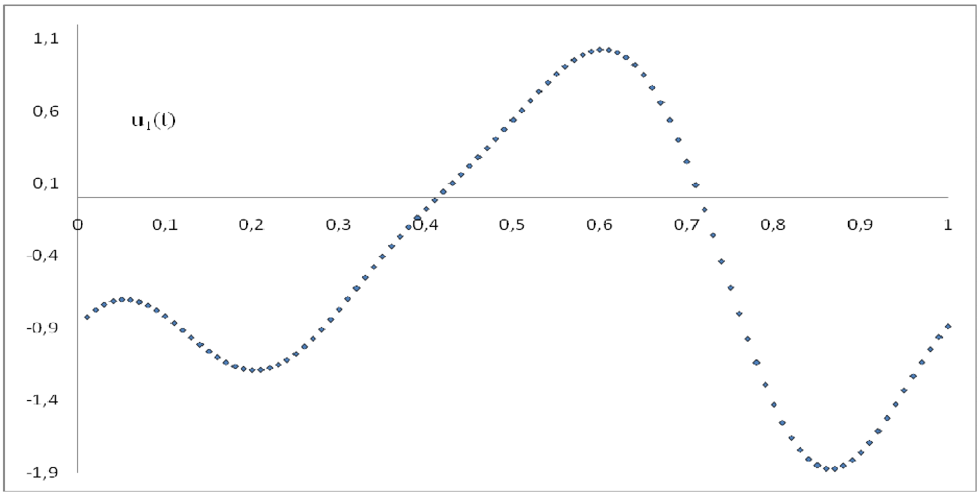

Fig. 5. The weight function $u_{1\,\text{knee}}(t)$ (43.31%)

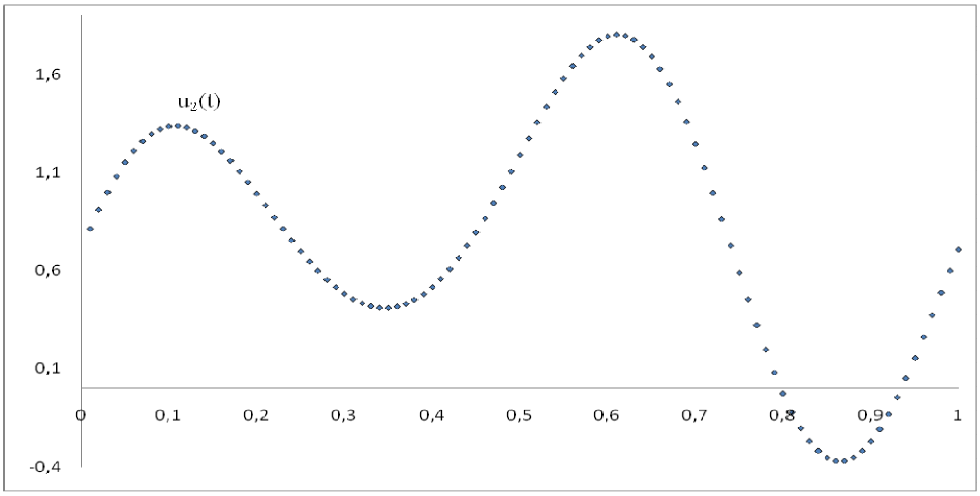

Fig. 6. The weight function $u_{2\,\text{knee}}(t)$ (25.53%)

Figures 7 – 8 present the 39 children on a plot of the first two functional principal components; the children's coordinates were computed using formula (2.12).

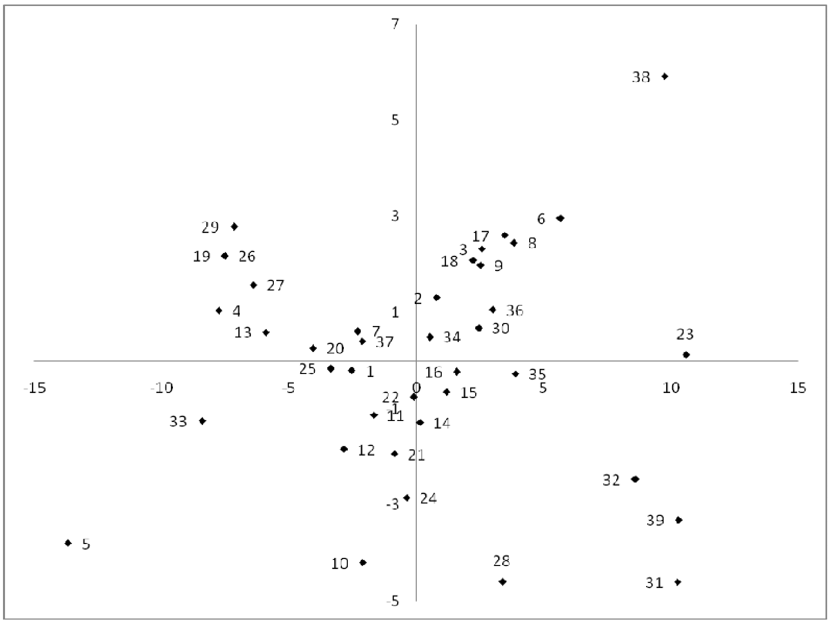

Fig. 7. Plotted values of the first two functional principal components for individual child (hip)

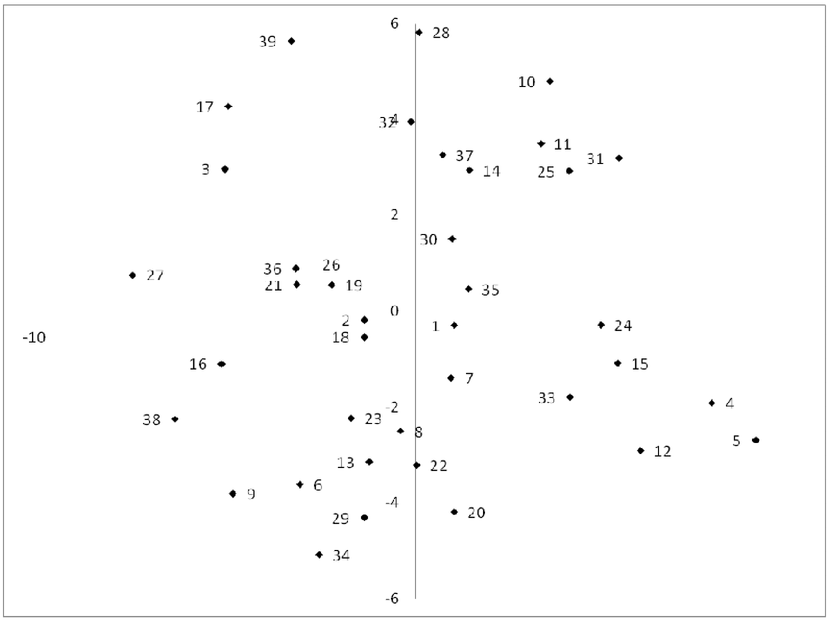

Fig. 8. Plotted values of the first two functional principal components for individual child (knee)

References

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

Olshen R.A., Binden E.N., Wyatt M.P., Sutherland D.H. (1989). Gait analysis and the bootstrap. *Annals of Statistics* 17, 1419 – 1440.

Ramsay J.O., Danzell C.J. (1991). Some tools for functional data analysis. *J.R. Statist. Soc*. B 53, 539–572.

Ramsay J.O., Silverman B.W. (2002). *Applied Functional Data Analysis*. Springer–Verlag, New York.

Ramsay J.O., Silverman B.W. (2005). *Functional Data Analysis*. Second Edition, Springer–Verlag, New York.

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