Clustering of functional objects in energy load prediction issues

Daniel Kosiorowski[1], Dominik Mielczarek[2], Ewa Szlachtowska[3]

Abstract

Many economic phenomena directly lead to functional data: yield curves, income densities, development trajectories, price trajectories, lives of products, and electricity, heat or water consumption within a day. The Functional Data Analysis (FDA) over the last two decades proved its usefulness in a context of a decomposition of income densities or yield curves, an analyses of huge, sparse economic datasets or in analysing of ultra-high frequency financial time series. The FDA enables us for conducting an effective statistical analysis when number of variables exceeds number of observations. Using FDA we can effectively analyse economic data streams i.e., e.g., perform an analysis of non-equally spaced observed time series and predict of a whole future trajectory of a stream rather than iteratively predict single observations.

In this paper we present and carefully study several economic applications of recently proposed clustering algorithms for functional data. We show their usefulness in a context of functional time series prediction. Theoretical considerations are illustrated by means of empirical examples related to energy consumption prediction.

Keywords: *electricity demand, functional k-means, functional median, functional data analysis*

JEL Classification: C14, C53, C55

1. Introduction

Forecasting electricity demand is becoming more and more important, because there has been a significant increase in the cost of energy production and a market competition has intensified over time. Studies on the prediction of electricity demand usually consider three main issues: short-term forecasts – used to schedule the generation and transmission of electricity, medium-term forecasts – used to schedule the fuel purchases and maintenance, and long-term forecasts – used to develop the power supply and delivery system (generation units, transmission system, and distribution system).

In many countries a transition to a low carbon economy is promoted because of suggestions that a low carbon transition offers challenges and might yield economic benefits comparable to those of the previous industrial revolutions. Decarbonizing transport and heating (for example, through the implementation of electric vehicles and heat pumps) may

[1] Corresponding author: Cracow University of Economics, Rakowicka 27, 31-510 Kraków, Poland, e-mail: daniel.kosiorowski@uek.krakow.pl.

[2] AGH University of Science and Technology, Mickiewicza 30, 30-059 Krakow, Poland, e-mail: dmielcza@wms.mat.agh.edu.pl.

[3] AGH University of Science and Technology, Mickiewicza 30, 30-059 Krakow, Poland, e-mail: szlachto@agh.edu.pl.

increase the demand network. In short, the electricity demand could increase, and the process of transmission may stop working at low voltage levels, or become unstable. Furthermore, with the accelerated spread of low-carbon energy technologies, distribution network operators via mass storage equipment began to collect data on changes in low voltage in order to determine expected changes in demand, followed by the modernization or network development. Precise forecasts of demand at the level of individual households can reduce peak demand, helping to plan the most appropriate charging and discharging cycles of electrical equipment. However before planning a method of transmission, cycles or the entire transmission network, measures are needed to help evaluate the accuracy and usefulness of forecasts of the electricity demand for households.

2. Functional Data Analysis

There is actually an increasing number of situations coming from different fields of applied sciences in which the collected data are curves. The progress of the computing tools, both in terms of memory and computational capacities, allows us to deal with large sets of data. In particular, for a single phenomenon, we can observe a very large set of variables.

Functional data analysis (FDA) extends the classical multivariate methods when data are functions or curves. Functional data analysis is about the analysis of information on curves or functions. The first contributions to functional data analysis concern the factorial analysis and are mainly based on the Karhunen-Loeve expansion of a second order L_2-continuous stochastic process (Loève, 1945). The contributions of Besse (Ferraty and Vieu, 2006) and of Saporta (Ferraty and Vieu, 2006) extends to functional data the principal component analysis, the canonical analysis of two functional variables, the multiple correspondence analysis for functional categorical data and the linear regression on functional data.

According to Ferraty and Vieu (2006), a functional random variable X is a random variable with values in an infinite dimensional space. Then, functional data represents a set of observations $\{X_1,\ldots,X_n\}$ of X . The underlying model for X_i's is generally an i.i.d. sample of random variables drawn from the same distribution as X . A well accepted model for this type of data is to consider it as paths of a stochastic process $X=\{X_t\}_{t\in T}$ taking values in a Hilbert space H of functions defined on some set T . Generally, T represents an interval of time, of wavelengths or any other continuous subset of $\Re$. The main source of difficulty when dealing with functional data, consists in the fact that the observations are supposed to belong to an infinite dimensional space, whereas in practice one only has sampled curves observed into a finite set of time-points. It is usual that we only have discrete observations X_{ij} of each sample path $X_i(t)$ at a finite set of knots $\{t_{ij} : j=1,\dots,m_i\}$. Because of this, the first step in FDA is often the reconstruction of the functional form of data from discrete observations. The most common solution to this problem is to consider that sample paths belong to a finite dimensional space spanned by some basis of functions (see Ramsay and Silverman, 2005). An alternative way of solving this problem is based on nonparametric smoothing of functions (see Ferraty and Vieu, 2006).

2.1 Cluster analysis

Cluster analysis is an important statistical methodology used in a wide variety of fields including business, biology, psychology and medicine. Outlying data can heavily influence standard clustering methods. Many clustering methods are non-robust and these statistical procedures may be heavily influenced by even a small fraction of outlying data. Often the results obtained by cluster analysis may cause an economist to make mistakes – this arising as a result of small errors in the data. This can lead to incorrect business decisions such as wrong description of the types of customers or incorrect household electricity demand forecasting. In recent years, the literature proposes robust clustering algorithms that cope well with the imperfections of economic data. Furthermore, there are modifications of the clustering algorithms for functional data. These algorithms can be used in many fields of science, including marketing research, accounting and forecasting of economic and business phenomena.

The article concerns issues of proper segmentation of electricity consumers and extracting homogeneous periods of daily electricity demand. The results will help to improve the quality of electricity demand forecasting.

2.2 The model for functional k-means

Let $\{x_1(t), x_2(t), ..., x_I(t)\}$ be a set of functional observations with $t \in \Gamma$ and Γ is an interval of $\Re$. Functions lie in a separable Hilbert space H with inner product:

$$\langle f, g \rangle = \int f(t) g(t) dt. \tag{1}$$

Assume that functional observations $x_i(t)$ are given by:

$$x_i(t) = \sum_{g=1}^{G} u_{ig} m_g(t) + \varepsilon_i(t),\ i = 1,\dots,I \tag{2}$$

where m_g are smooth unknown centroid functions, $\varepsilon_i(t)$ denotes an unobservable zero-mean error term, $u_{ig} \in \{0,1\}$, $\sum_g u_{ig} = 1$ for every i and $u_{ig} = 1$ if x_i belongs to the g-th cluster.

The parameters u_{ig} and m_g are estimated by minimizing:

$$\sum_{i,g} u_{ig} \int_{\Gamma} \left[x_i(t) - m_g(t) \right]^2 dt = \sum_{i,g} u_{ig} \, \| x_i(t) - m_g(t) \|^2 \to min. \tag{3}$$

Let $x_i(t_z)$ be an observed curve, $i = 1,\ldots,I$, $z = 1,\ldots,Z$. Assume the centroids admit the basis expansion:

$$m_g(t) = \sum_{j=1}^{J} c_{gj} \varphi_j(t). \tag{4}$$

Then the model in matrix form becomes:

$$X = UC\Phi' + E \tag{5}$$

where X is the $I \times Z$ matrix with elements $\{x_i(t_z)\}$, U is the $I \times G$ matrix with elements $\{u_{ig}\}$, C is the $G \times J$ matrix with elements $\{c_{gj}\}$, Φ is the $Z \times J$ matrix with generic j-th column $\{\varphi(t)\}$ and E is the $I \times Z$ matrix with elements $\{\varepsilon_{iz}\}$.

Then objective function to minimize becomes:

$$h(C,U) = \left\| UC\Phi' - X \right\|^2 = tr\left\{ \left[UC\Phi' - X \right]' \left[UC\Phi' - X \right] \right\}. \tag{6}$$

Given U, Φ and X fixed, it is possible to write:

$$h(C,U) = \left\| (U'U)^{\frac{1}{2}} CW^{\frac{1}{2}} - (U'U)^{-1/2} U'X\Phi W^{-1/2} \right\|^2 + b \tag{7}$$

where b is independent of C and matrix W has components $w_{jr} = \int \varphi_j(t) \varphi_r(t) dt.$

The minimizing C is given by:

$$C = (U'U)^{-1} U'Y\Phi W^{-1}. \tag{8}$$

2.3 The algorithm

By means of the standard FDA tools we compute Φ and W. Next we choose initial value for U and we update C by means of:

$$C = (U'U)^{-1} U'Y\Phi W^{-1}. \tag{9}$$

Given the current estimate of C, update U by means of the k-means algorithm as follows: we set $u_{ig} = 1$ if $\left\|x_i - m_g\right\|^2 = \min\left\{\left\|x_i - m_l\right\|^2 : g = 1,\ldots,G; l \neq g\right\}$ and $u_{ig} = 0$ otherwise. Then we repeat until convergence.

FDA is improved and the results are more interpretable when smoothing is included in the model estimation. The problem is how to impose the smoothness condition. There are two approaches. The first approach is filtering approach. The second approach is regularization – contextual smoothing (better than the first one).

The smoothing step is incorporated by adding a roughness penalty to criterion (3) and minimizing:

$$\sum_{i,g} u_{ig} \| x_i(t) - m_g \|^2 + \lambda \sum_g \left\| m_g^{"} \right\|^2 \to min, \tag{10}$$

where $m_g^{"}$ is the second derivative of m_g. Given the u_{ig}, minimizing (10) is equivalent to minimize:

$$\sum_g \| \overline{x}_g - m_g \|^2 + \lambda \sum_g \left\| m_g^{"} \right\|^2 \to min. \tag{11}$$

3. Electricity market in Australia

We applied the functional k-means algorithm for functional data showing the electricity demand in Australia. All the considered data was taken from the AEMO Australian Energy Market Operator (http://www.aemo.com.au/). The series is for New South Wales (abbreviated as NSW), which is one of state of Australia. Daily electricity consumption is measured every half hour. Fig. 1 depicts the time series for the energy consumption in 2014. The electricity demand in 2014 is characterized by high symmetry. In fact, this reflects the seasonality of the seasons. Therefore, we compared the electricity demand in each quarter.

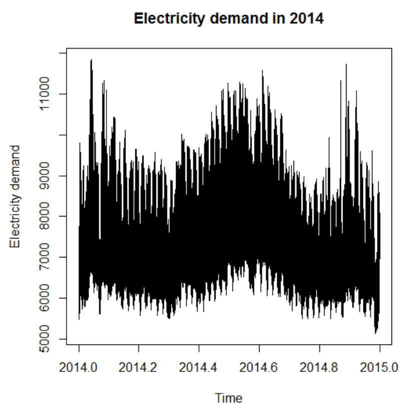


Fig. 1. Electricity demand in Australia in 2014.

The following graphs in Fig. 2 illustrate the time series describing the electricity demand in Australia electricity in each quarter of 2014.

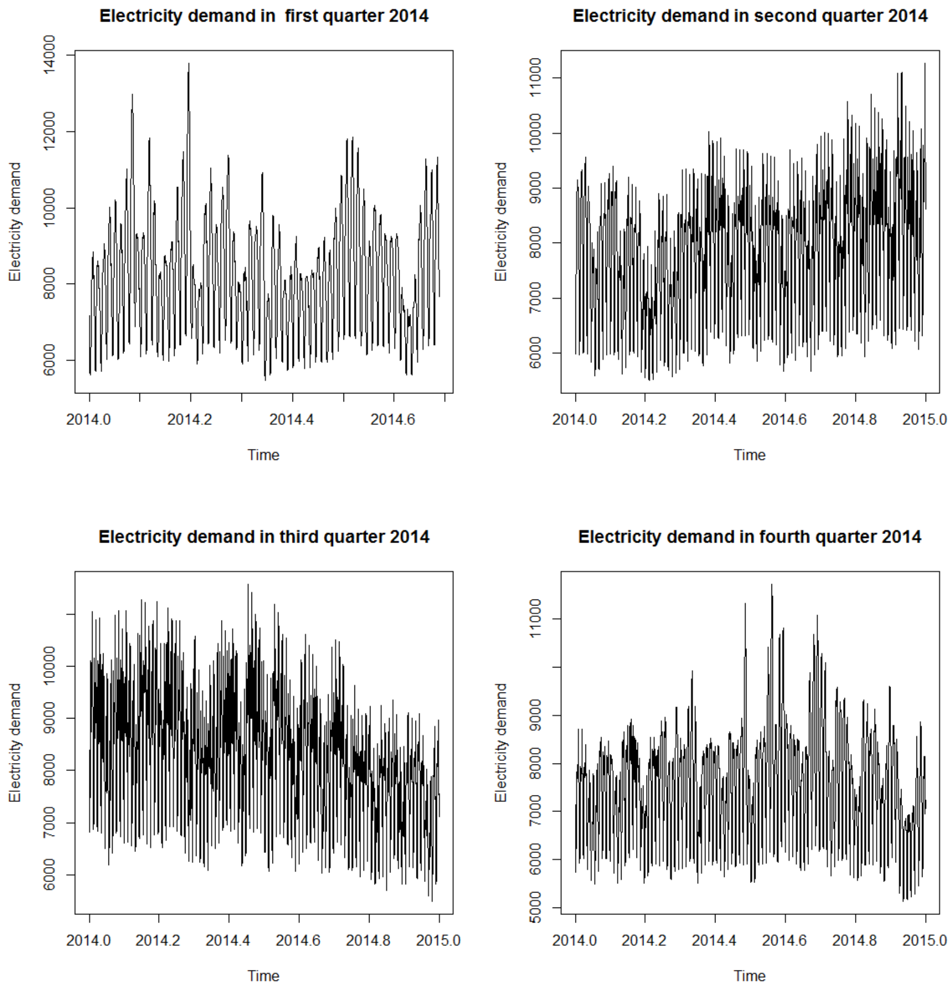
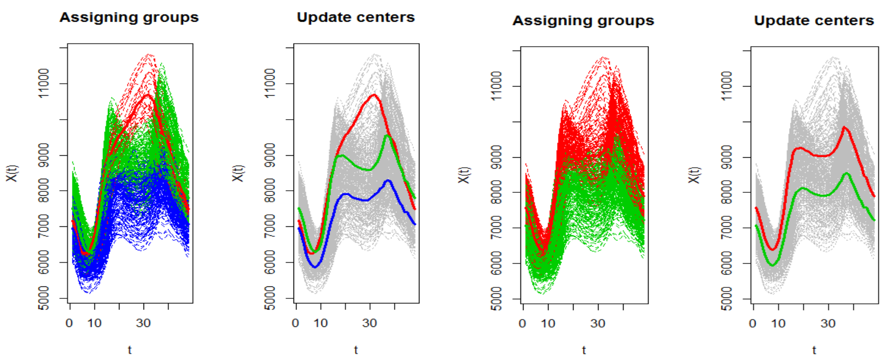


Fig. 2. Electricity demand in each quarter of 2014.

The result is that demand in the first and fourth quarter are comparable, as well as demand in the second and third quarter are comparable.

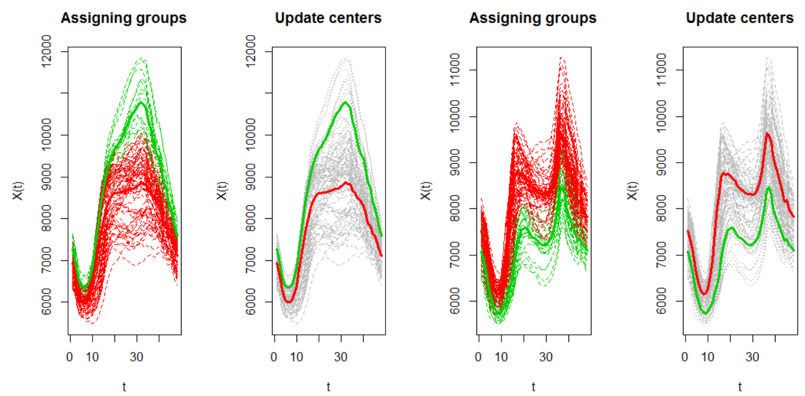
We used the functional k-means algorithm for functional data in each quarter of 2014. We search the optimal number of clusters for daily demand profiles in each quarter. The choice of the number of clusters is very important in the functional k-means algorithm. When we consider k = 3, the third cluster partially overlaps with the other components. Moreover, this clustering procedure adequately handles the severe overlap of two clusters (see Fig. 3). Thus we can expect that the division into two clusters corresponding to the increased and reduced need for electricity.



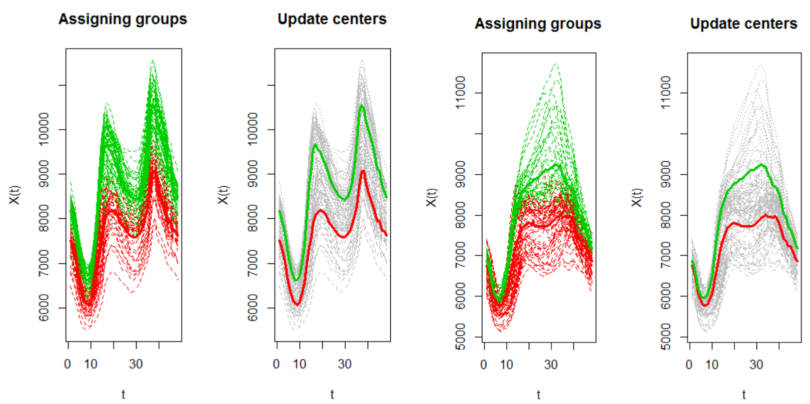
a) k = 3 **b)** k = 2

Fig. 3. Applying the algorithm functional k-means a) for k = 3, b) for k = 2.

The figures below show the assignment of the two clusters, obtained by the use of functional k-means algorithm for each of the quarters of 2014.



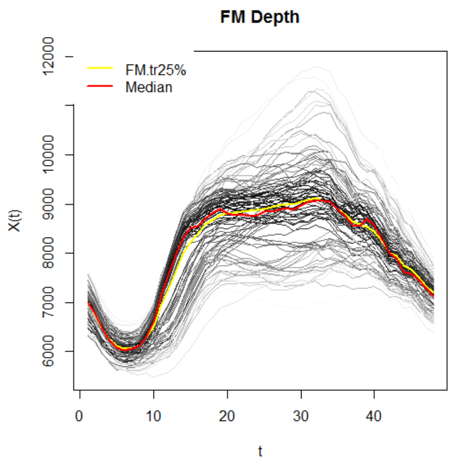
a) The first quarter of 2014 **b)** The second quarter of 2014



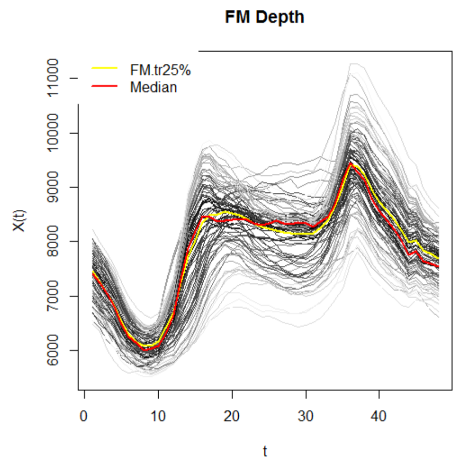
c) The third quarter of 2014 **d)** The fourth quarter of 2014

Fig. 4. Clustering results for each quarter of 2014 when applying the functional k-means algorithm with k = 2.

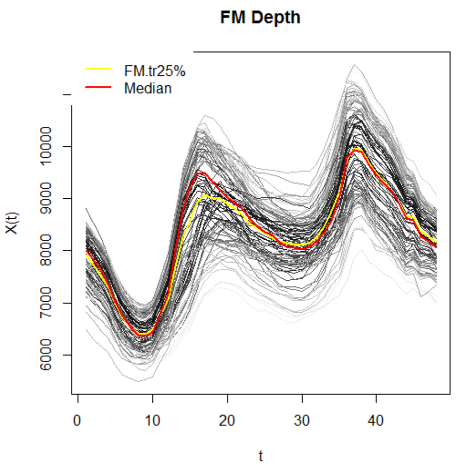
The figures below show a comparison of the functional medians for each quarter of 2014 (see Fig. 5). The graphs show that the functional medians are a “little” different from the functional 25% trimmed means.



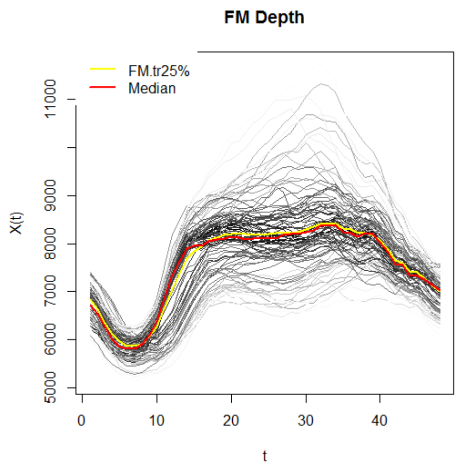
a) The first quarter of 2014



b) The second quarter of 2014



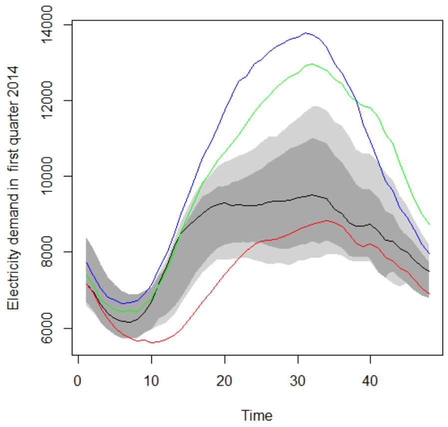
c) The third quarter of 2014



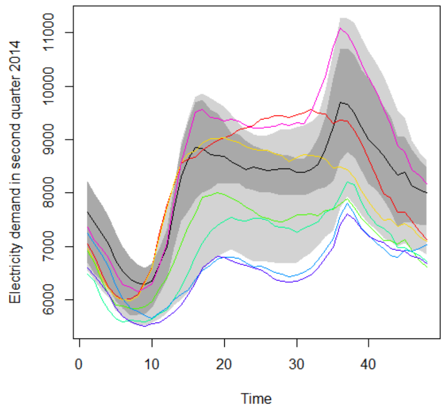
d) The fourth quarter of 2014

Fig. 5. The comparison of the functional medians and the 25% trimmed functional means for each quarter of 2014.

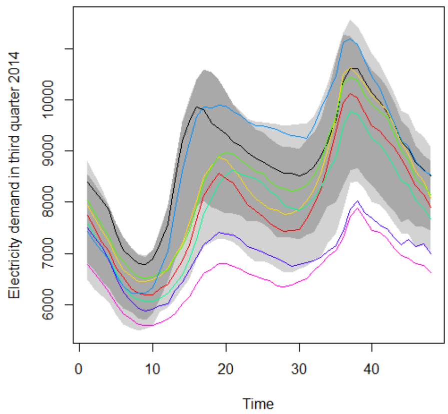
From the below figures, the electricity demand is significantly different in the second and third quarter in comparison to the first and fourth quarter. We can observe similar trend in the first and fourth quarters, and a similar trend in the second and third quarter (see Fig. 6).



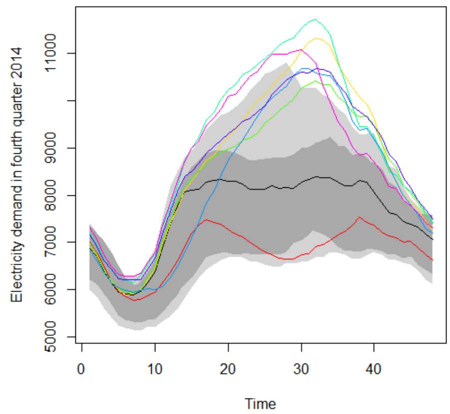
a) The first quarter of 2014



b) The second quarter of 2014



c) The third quarter of 2014



d) The fourth quarter of 2014

Fig. 6. Functional box plots for electricity demand in Australia in 2014.

Conclusion

As seen in Fig. 4 the observations have been divided into two clusters in each of the quarters. This suggests that there are two groups of customers. Households can be divided due to the electricity demand, i.e. high and low energy consumption. This is important in planning the modernization or network development, as well as in determining the subscriptions for different customer groups.

Moreover one of the most complex problems when applying the functional k-means algorithm is the choice of the number of clusters, k. In some cases one might have an idea of the number of clusters in advance, but usually k is completely unknown. Future plans are to propose an algorithm that would help in determining the number of clusters.

Acknowledgements

Authors thank for the financial support related to CUE Faculty of Management fund for the research potential preservation and the Polish National Science Center for financial support from grant UMO-2011/03/B/HS4/01138.

References

Ferraty, F., & Vieu, P. (2006). *Nonparametric functional data analysis: theory and practice*. Springer Science & Business Media.

Gattone, S. A., & Rocci, R. (2012). Clustering curves on a reduced subspace. *Journal of Computational and Graphical Statistics*, *21*(2), 361-379.

Górecki, T., & Krzyśko, M. (2012). Functional Principal Components Analysis. In: Pociecha, J., & Decker, R. (eds.), *Data analysis methods and its applications*. Warszawa: C.H. Beck, 71-87.

Kosiorowski, D. (2014). Functional Regression in Short Term Prediction of Economic Times Series. *Statistics in Transition new series*, *15*(4), 611-626.

Kosiorowski, D., Mielczarek, D., Rydlewski, J., & Snarska, M. (2014). Sparse Methods for Analysis of Sparse Multivariate Data from Big Economic Databases. *Sampling methods and estimation*, *Statistics in Transition new series*, *15*(1), 111-132.

Loève, M. (1945). *Fonctions aléatoires de second ordre*. C. R. Acad. Sci. Paris, 220-469.

Febrero-Bande, M., & Oviedo de la Fuente, M. (2012). Statistical computing in functional data analysis: the R package fda. usc. *Journal of Statistical Software*, *51*(4), 1-28.

Pociecha, J. (1986). *Statystyczne metody segmentacji rynku.* Kraków: Wydawnictwo Akademii Ekonomicznej w Krakowie.

Silverman, B. W., & Ramsay, J. O. (2005). *Functional Data Analysis*. Springer.