An advanced multivariate statistical approach to study coastal sediment data

Pavlina Simeonova¹, Costel Sarbu², Thomas Spanos³, Vasil Simeonov⁴*, Stefan Tsakovski⁴

¹ Institute of Solid State Physics, Bulgarian Academy of Sciences, Tzarigradsko Chaussee Blvd. 72, 1784 Sofia, Bulgaria
² Faculty of Chemistry and Chemical Engineering, Babes-Bolyai University of Cluj-Napoca, Jarani Janos 11, RO-3400 Cluj-Napoca, Romania
³ General Department, Technological Educational Institution (TEI), St. Lucas, 65403 Kavala, Greece
⁴ Chair of Analytical Chemistry, Faculty of Chemistry, University of Sofia “St. Kl. Okhridski”, 1164 Sofia, J. Bourchier Blvd. 1, Bulgaria

Received 8 July 2005; accepted 20 September

Abstract: The present paper deals with the application of classical and fuzzy principal components analysis to a large data set from coastal sediment analysis. Altogether 126 sampling sites from the Atlantic Coast of the USA are considered and at each site 16 chemical parameters are measured. It is found that four latent factors are responsible for the data structure (“natural”, “anthropogenic”, “bioorganic”, and “organic anthropogenic”). Additionally, estimating the scatter plots for factor scores revealed the similarity between the sampling sites. Geographical and urban factors are found to contribute to the sediment chemical composition. It is shown that the use of fuzzy PCA helps for better data interpretation especially in case of outliers.

Keywords: Coastal sediments, marine pollution, chemometrics, fuzzified principal components analysis

* E-mail: VStimeonov@chem.uni-sofia.bg
1 Introduction

There is serious concern in the modern world about the contamination of the marine environment and especially about the concentration of pollutants in the coastal sediments. Therefore, the careful monitoring of the sediments proved to be an important tool for indication of marine pollution. Most contaminants released into coastal waters rapidly become associated with the particulate matter in the seawater and finally incorporated into sediments. Thus, the coastal sediments become a quite specific bioindicator of the nature of the pollution events. It has to be also kept in mind that natural processes also modify and redistribute anthropogenic contamination between solid and water phases. According to Martin and Whitfield [1] the accumulation of metal contaminants in coastal sediments gives a reliable picture of the spatial and temporal pollution history.

Large-scale monitoring studies in the USA in the last decade delivered important information about major and trace element distribution in wide coastal regions of the country. Scientific groups from the National Oceanic and Atmospheric Administration (NOAA) using programs such as National Status and Trends (NS&T) program for Marine Environment [2, 3] have performed these studies. These studies made it possible to collect substantial amount of information from the eastern and western coasts of the USA over large scales of time and distance. All procedures of choice of sampling locations (distance from the coast), sampling, sample pretreatment, and chemical analysis were preliminary determined in the NS&T program [2, 3].

The idea of specific recognition of the effects of natural and anthropogenic factors in the formation of the marine sediments is of substantial importance. Correlation and regression analysis are usually the main tools for estimation of the contribution of various effects to sediment genesis [4–6]. These approaches are performed after appropriate selection of major components (tracers) related to the morphology of the sediments like iron, total organic carbon (TOC), aluminium or grain size. The reason for the choice is that iron and aluminium are present in constant and high levels in the sediment volume, and are not affected by anthropogenic actions, TOC content reveals important processes of sediment formation with participation of trace metals from anthropogenic activity, and grain size is related to adsorption of heavy metals being also contaminants. In all these cases of statistical data interpretation the approach is strictly univariate (linear regression of tracers as dependent variables). For instance, the baseline model suggested by NS&T teams enables simple estimation of pollution processes with iron as tracer [4]. It is assumed that sampling sites along the coast with population less than 10,000 located 20 km from the coastline are anthropogenically unaffected. The slope of the regression line (iron concentration as function of heavy metal concentration) is compared with the ratio of the element concentrations in the sediment formation rock. If both values are statistically comparable, the model enables calculation of the naturally occurring heavy metal concentration in the sediment sample. Polluted sites are those with statistically significant difference between the heavy metal concentration predicted and the measured one.
Some environmetric studies recently published [7–9] have indicated the restrictions of this model. In addition, the advantages of intelligent data analysis using chemometrics were demonstrated. The separation between anthropogenic and natural contributions to the sediment formation was performed by classical principal components analysis (PCA) and by multiple regression on principal components.

The aim of the present study is to apply an advanced chemometric approach to coastal sediment data in order to reach a better understanding of the role of the sediments as bioindicators of pollution, to apportion the contribution of natural and anthropogenic factors to sediment formation, and check the applicability of the novel chemometric method. Additionally, the ecological situation of the coastal regions could be estimated in a more responsible way.

2 Experimental

2.1 Sampling and sample analysis

The sampling was performed according to the requirements of the program NS&T, which consists of two major projects: The National Benthic Surveillance and the Mussel Watch. The aim of the program is to estimate the environmental pollution along the coastal line of the USA and to localize the pollution sources by analysis of coastal and estuary systems including marine sediments.

The data used in this study are sediment data collected along the coastal line of Gulf of Mexico. Altogether 126 sampling sites are used and from each site, three samples are taken. Each of these samples is a mean from three samplings. In this way for sediment analysis three samples are available to determine organic compounds, three – for inorganic components and three for grain size estimation.

The sediment sampling is performed at depths between 0.1 and 3 m at 1 to 400 m distance from shore. For lesser depths the sample collection is carried out manually by the use of a Teflon spoon but for bigger depths with a special device described in [2, 3].

Sample of 0.10 to 0.45 g dry sediment is located in a Teflon vessel and treated by mixtures like HNO₃-HF, HNO₃ - KCl – HF or HNO₃ – HClO₄ – HF at heating with conventional or microwave heaters. Solution of H₃BO₃ is added to dissolve the insoluble fluorides in the sample. The solution obtained is analyzed by ETAAS, ICP-AES, AAS (cold vapor method) and XRF for Ag, Al, As, Cd, Cr, Cu, Fe, Hg, Mn, Ni, Pb, Se, Sn and Zn. Altogether 16 parameters (chemical variables) were determined in each sample. A full description of sampling, sample preparation and analysis (including TOC and grain size) could be found elsewhere [2, 3].

A summarized statistics of the analytical results is presented in Table 1.
2.2 Chemometrics

Principal component analysis (PCA) is a favorite tool in chemometrics and environmetrics for data compression and information extraction [10, 11]. PCA finds linear combinations of the original measurement variables that describe the significant variations in the data. However, it is well known that PCA, as with any other multivariate statistical method, is sensitive to outliers, missing data, and poor linear correlation between variables due to poorly distributed variables. As a result data transformations have a large impact upon PCA. In this regard one of the most powerful approaches to improve PCA appears to be the fuzzification of the matrix data, thus diminishing the influence of the outliers.

2.3 Fuzzy PCA - optimizing the first component (FPCA-1)

For the data collected on \( p \) variables for \( n \) cases, PCA performs analyses in the \( n \)-dimensional space defined by \( p \) variables and \( p \)-dimensional space defined by \( n \) cases. In PCA straight lines are sought which best fit the clouds of points in the vector spaces (of variables and cases), according to the least squares criterion. This, in turn, yields the principal components (factors) that result in the maximum sums of squares for the orthogonal projections. Consequently, a lower dimensional vector subspace is recovered that best represents the original vector space. Although the first factor is extracted so as to capture the variance to the maximum extent, it can seldom capture the variance in its entirety. What remains should, therefore, be recovered by another (second) factor, a third, etc. However, the number of factors thus extracted will never exceed the number of original variables.

Fuzzy clustering is an important tool for identifying the structure in data [12]. According to the choice of prototypes and the definition of the distance measure, different fuzzy clustering algorithms are obtained. If the prototype of a cluster is a point – the cluster center – it will produce spherical clusters; if the prototype is a line, it will produce tubular clusters, and so on. Also, elements with a high degree of membership in the \( i \)-th cluster (i.e., close to the cluster’s center) will contribute significantly to this weighted average, while elements with a low degree of membership (far from the center) will contribute almost nothing. In what follows we briefly review the Fuzzy (first component) PCA algorithm proposed in reference [13]. We wish to determine the particular membership degrees \( A(x) \) such that the first principal component is best fitted along the points of the data set \( X \). The algorithm is a natural extension of the Fuzzy Regression Algorithm [14, 15]. The fuzzy set in this case may be characterized by a linear prototype, denoted \( L(u,v) \), where \( v \) is the center of the class and \( u \), with \( ||u|| = 1 \), is the main direction. This line is named the first principal component for the set, and its direction is given by the unit eigenvector \( u \) associated with the largest eigenvalue \( \lambda_{max} \) of, for example, the covariance matrix given in relation (1), which is a slight generalization for fuzzy sets of
| Component | Mean  | SD    | SD/mean | Min.  | Max.   |
|-----------|-------|-------|---------|-------|--------|
| TOC       | 19729 | 23477 | 1.19    | 900   | 118773 |
| Al        | 42239 | 25400 | 0.60    | 790   | 108703 |
| As        | 6.82  | 5.17  | 0.76    | 0.35  | 23.54  |
| Cd        | 0.19  | 0.16  | 0.84    | 0.01  | 0.91   |
| Cr        | 47.70 | 26.76 | 0.56    | 3.56  | 170.38 |
| Cu        | 12.97 | 12.59 | 0.97    | 0.92  | 87.67  |
| Fe        | 19708 | 14500 | 0.73    | 823   | 66049  |
| Pb        | 20.50 | 17.65 | 0.86    | 1.26  | 115.13 |
| Mn        | 343   | 324   | 0.95    | 4.0   | 1560   |
| Hg        | 0.06  | 0.07  | 1.17    | 0.001 | 0.24   |
| Ni        | 14.27 | 9.25  | 0.65    | 0.84  | 36.23  |
| Se        | 0.36  | 0.27  | 0.75    | 0.007 | 1.21   |
| Ag        | 0.12  | 0.17  | 1.42    | 0.008 | 1.42   |
| Sn        | 1.69  | 1.18  | 0.70    | 0.20  | 6.65   |
| Zn        | 66.37 | 49.26 | 0.74    | 3.16  | 372.53 |
| Grain size| 0.58  | 0.24  | 0.42    | 0.10  | 0.99   |

Note: The dimension of the chemical components is in ppm; grain size is in mm. SD means standard deviation. The total number of samples is 378.

Table 1 Basic statistics of the input data set.

The classical covariance matrix:

\[
C_{kl} = \frac{\sum_{j=1}^{n} [A_i(x^j)]^2 (x_{jk} - \bar{x}_k)(x_{jl} - \bar{x}_l)}{\sum_{j=1}^{n} [A_i(x^j)]^2} \quad (1)
\]

The algorithm defined in this way permits the determination of the \( A(x^j) \) values that best describe the fuzzy set \( A \) and the relation with its linear prototype (the first principal component).

2.4 Fuzzy PCA – orthogonal (FPCA-O)

Encouraged by the good results obtained with the Fuzzy (first component) PCA, we decided to extend the fuzzy approach one step more. A Fuzzy PCA algorithm is written that would extend the fuzzy clustering scheme with computing each particular principal component, not just the first one. Let us denote \( \lambda_1, \ldots, \lambda_p \) and \( e^1, \ldots, e^p \) the eigenvalues and the eigenvectors, respectively, that will finally be produced by our suggested algorithm. The first fuzzy principal component is computed as with the FPCA-1 algorithm, i.e., by finding the optimal fuzzy membership degrees and the optimal linear prototype for the data set. Let us denote \( \lambda'_1, \ldots, \lambda'_p \) and \( e'^1, \ldots, e'^p \) the eigenvalues and eigenvectors,
respectively, produced in this way. Therefore, we will have

\[ \lambda_1 = \lambda'_1 \]  

(2)

and

\[ e^1 = e'^1 \]  

(3)

The major novelty of this algorithm is in the way the other fuzzy principal components are computed. The original data set is projected onto the hyperplane orthogonal to the first fuzzy principal component, i.e., determined by all the other principal components, as determined by the Fuzzy First Component PCA algorithm. Practically, this may be done by computing the scores and removing the first item from the data vectors. Therefore, we first compute the scores:

\[ x'^j = x^j \cdot (e'^1, \ldots, e'^p) \]  

(4)

and then remove the first component of \( x'^j \), thus producing a subset \( X' \) of \( \mathbb{R}^{p-1} \):

\[ X' = \{ (x'_2, x'_3, \ldots, x'_p) | \exists j : x'^j = (x'^1, x'^2, \ldots, x'^p) \}. \]  

(5)

This produces a data set in a Euclidean space of dimension \( p - 1 \), where \( p \) is the size of the original data set. Denote \( \lambda''_1, \ldots, \lambda''_{p-1} \) and \( e''_1, \ldots, e''_{p-1} \) the eigenvalues and eigenvectors, respectively, produced in this way. The first fuzzy principal component of this projected data set, after being rewritten in terms of the original space, is orthogonal on the first fuzzy principal component, as computed originally. In order to account for the fuzziness in the fuzzy data sets, when rewriting the components in terms of the original space, the eigenvalues computed in the \( p - 1 \) sized space will be multiplied by the fuzzy set fuzziness index \( f_A \), given by

\[ f_A = \frac{1}{n} \sum_{i=1}^{n} A(x^1)^m. \]  

(6)

Thus

\[ \lambda_2 = \lambda''_1 \cdot f_A \]  

(7)

and

\[ e^{2T} = (0, e''_1)^T \cdot (e'^1, \ldots, e'^p)^T, \]  

(8)

where \((0; e''_1)\) denotes a vector having 0 for the first component, and the components of the vector \( e''_1 \) for the other components. In order to determine the third fuzzy principal component we will reason in the same way, but here we start with the projected data set and project it onto the hyperplane orthogonal to the first two fuzzy principal components. This twice-projected data set will be in a Euclidean space of dimension \( p - 2 \). Let us suppose that, after proper transformations have been made at the superior level, these newly produced eigenvectors and eigenvalues (now in the \( \mathbb{R}^{p-1} \) space) and still denoted by \( \lambda''_1, \ldots, \lambda''_{p-1} \) and by \( e''_1, \ldots, e''_{p-1} \). These notations will replace the already computed values. Now we need only a final transformation: to revert these eigenvectors and eigenvalues to the original space. We use relations similar to those used for computing the initial projection, but in reverse [15].

It is worth to mention that the new approach of fuzzy PCA has found recently numerous applications [16–18].
3 Results and Discussion

As already mentioned above the main goals of the study were as follows:

- to detect similarities between the sampling areas from the coastal line where the sampling was performed (data projection using the factor scores from classical or fuzzy PCA);
- to detect latent factors responsible for the data structure and, thus, to identify sources responsible for the sediment formation (by the use of factor loadings from classical and fuzzy PCA);
- to compare the classical and the fuzzy approach of PCA as projection and modeling strategy in environmental chemistry.

In the next figures (Fig. 1 – 6) the scatter plot diagrams for all sampling sites and all chemical variables treated by various PCA approaches (classical and fuzzy) are presented.

Each one of the plots requires a careful interpretation. When one applies classical PCA to the data plot four principal components are found (scree plot check) to describe over 85 % of the total variance of the system (126 sediment samples x 18 variables; 15 of them are chemical components, one is a physical parameter – grain size and the last two are the geographical coordinates of the sites – longitude and latitude). The first latent factors are conditionally named “natural” since it indicates high factor loadings for Al, As, Cr, Fe, Mn, Ni, and grain size. These are major sediment components, which play a substantial role in the sedimentation process – iron, manganese, chromium, nickel, arsenic as typical constituents in iron – containing natural materials. Grain size is included as it is an indication for the significant contribution of the major components to the sediment topology.

The second latent factor contains as significant contributors polluting heavy metals (zinc, mercury, copper, lead, cadmium and tin). This is labeled as “anthropogenic”. The third latent factor represents the participation of the organic matter in the sediment formation (high factor loadings for TOC) and is named “bioorganic”. The bioorganic contribution to the sediment formation is obviously split into two parts: the first one explained by PC3 informs on natural organic matter impact to the sediment mass and the second (explained by PC 4) on the role of the organometallic polluting species to the same impact. Higher factor loadings in PC 4 possess silver and lead, but also TOC and cadmium. The fourth latent factor can thereby be conditionally called “organic anthropogenic”.

The geographical parameters are included in the plots just to indicate their independence on the chemical variables.

It was of substantial interest to show if the application of the fuzzy PCA (first direction and orthogonal) changes the way of variables projection. The 3D plots for both cases are presented in Figs. 2 and 3.

In the case of fuzzy PCA (first direction approach, explained total variance for 4 PCs approximately 80 %) the identified sources for the sediment formation are almost the same as in the case with the classical PCA. One could separate the conditional “natural”
Fig. 1 Scatter plot of loadings corresponding to the first three principal components (classical PCA).

Fig. 2 Scatter plot of loadings corresponding to the first three principal components (Fuzzy PCA, first direction).

factor (with high loadings of Al, Ni, Fe, Mn, Cr, As, and grain size parameter [GS]); further the “anthropogenic” one (significant loading values for Cu, Hg, Zn, Sn, and Cd).
In the next latent factor one detect high contributions of total organic carbon [TOC] and Se, which holds for the conditional “bioorganic” tracer in the sediment formation as well as relation between Pb and Ag to make the latent factor named “organic anthropogenic” due to the relatively high loading for TOC in PC 4. Therefore, no difference between classical PCA and fuzzy PCA (first direction) is found in identification of the sources of sediment formation in the region of interest.

![3D Scatterplot (Vasilodfpca.sta 19v*18c)](image)

**Fig. 3** Scatterplot of loadings corresponding to the first three principal components (Orthogonal Fuzzy PCA).

In the third option of PCA (orthogonal fuzzy PCA, explained total variance approximately 78 %) the results obtained for the latent factors identification do not differ substantially from the other two options. The first two principal components are almost the same and resemble the “natural” and the “anthropogenic” impact in sediment formation. The only difference is that in PC 2 a clear relation between Pb, Cu, Zn, Hg, and Sn is established as Cd changes its belonging with Pb to PC 4, respectively. Thus, the total organic carbon amount is related to Se as in the previous cases but the “organic anthropogenic” factor now is, indeed, another anthropogenic impact (high correlation between Ag and Cd with weaker correlation to TOC). In such a way, the orthogonal fuzzy PCA gives better information on the role of the bioorganic impact to the sediment formation. Probably, the identified PC4 is resulting in specific anthropogenic activity, rather than in formation of organometallic compounds.

In the next three plots the projection of the sampling stations as objects of interest is given. In Fig. 4 the scatter plot of the 126 sites obtained by the application of classical PCA is represented. It may be seen that several sites appear to be typical outliers (sites with code numbers 117, 65, 111, 112, 116, 66, 67, 125, 121) since the rest form a large
Fig. 4 Scatter plot of scores corresponding to the first two principal components (classical PCA).

group of obviously close related in properties sites. The interpretation of the outliers from the scatter plot (Fig. 4) does not seem very simple. For instance, site 117 is located near to a small settlement (less than 1000 inhabitants) and could be accepted as background site with very low level of pollution. On the other hand, site 67 is a typical urban site (above 10000 inhabitants and chemical industry). Considered in such a way, the outliers indicated by the classical approach of PCA indicated either background site with clean environment or urban sites with higher level of pollution impact. The sediments sampled in the neighborhood of similar sites reflect, therefore, the general environmental situation and, maybe, lack information on some important details like type and nature of separate pollution sources.

A further attempt was made to improve the information on the characteristics of the sampling areas by using fuzzy PCA (first direction). The scatter plot for the factor scores is presented in Fig. 5. It is seen that a slight change of the grouping of the sampling sites is at hand. For instance, the former outlier site 117 is in one group of similar objects like 36, 37, 39, 68, and 69. It could mean that the little village with code number 117 resembles sites with different urban profile (e.g. 68, 69 which are bigger towns). The common feature of these sites is that they belong to the same geographical region (Mississippi delta). Similar considerations could be made for other groups of sites to find general features like sites from Texas, from the western part of the Mexican Gulf etc.

The application of the third mode of PCA used in this study, namely the orthogonal fuzzy PCA adds some more interesting aspects of the objects interpretation. The data treatment in this case indicates that no typical outliers are already available. Since the
Fig. 5 Scatter plot of scores corresponding to the first two principal components (Fuzzy PCA, first direction).

Fig. 6 Scatter plot of scores corresponding to the first two principal components (Orthogonal Fuzzy PCA)

classical PCA is very sensitive to outliers and this fact can spoil any classification attempt, the results obtained by the orthogonal fuzzy PCA make it possible to find the following
factors, which regulate the sediment formation in the region of interest (additionally to the identified possible pollution sources):

- the geographical location of the site (with respect to big areas like Mississippi Delta sites, Texas sites, Louisiana sites, Florida sites, sites from the southern or western part of the Mexican Gulf);
- the urban situation within the geographical regions: typical urban sites, small settlements, background sites, industrial sites;
- the population factor.

The software used for all calculations was STATISTICA 6.0, which is a commercially available software product. Additionally, original program for fuzzy PCA calculations was applied. The latter is available from the authors on request.

4 Conclusions

As mentioned, principal component analysis (PCA) is a favorite tool in environmetrics for data compression and information extraction. PCA finds linear combinations of the original measurement variables that describe the significant variations in the data. However, it is well known that PCA, as with any other multivariate statistical method, is sensitive to outliers, missing data, and poor linear correlation between variables due to poorly distributed variables. As a result data transformations have a large impact upon PCA. In this regard one of the most powerful approaches to improve PCA appears to be the fuzzification of the matrix data, thus diminishing the influence of the outliers. In this paper we apply a robust fuzzy PCA algorithm (FPCA). The application of various modes of principal components analysis (classical and fuzzy) to data from sediment monitoring (Atlantic Coast of USA) has indicated that the data interpretation could be positively changed. The traditional latent factor identification, which is normally achieved by estimating of the factor loadings in the classical PCA, could be improved if fuzzy PCA is applied. Very often fuzzy PCA gives better models with higher level of explained total variation. But more important improvement is achieved in the interpretation of the projection scatter plots of the objects (sampling sites). The fuzzy approach eliminates some outliers and makes it possible to determine in a more reliable way the role of “location” factors in sediment formation – the geographical coordinates of the sampling site, the level of population, the total environment of the site.

References

[1] J. Martin and M. Whitfield: The Significance of the River Input of Chemical Elements to the Ocean, Plenum Press, New York, 1983.
[2] National Status and Trends Program: Monitoring Site Descriptions (1984 – 1990) for the National Mussel Watch and Benthic Surveillance Projects, NOAA Office of Oceanography and Marine Assessment, Rockville, MD, 1998.
[3] Second Summary of Data on Chemical Contaminants in Sediments from the NS&TTP, NOAA, Technical Memorandum 59, NOS OMA, Rockville, MD, 1991.

[4] K. Daskalakis and Th. O’Connor: “Distribution of Chemical Concentrations in US Coastal and Estuarine Sediment”, Marine Environ. Res., Vol. 40, (1995), pp. 381–398.

[5] P. Hanson, D. Evans, D. Colby and V. Zdanowicz: “Assessment of Elemental Contamination in Estuarine and Coastal Environments based on Geochemical and Statistical Modeling of Sediments”, Marine Environ. Res., Vol. 36, (1993), pp. 237–266.

[6] A. Cantillo and Th. O’Connor: “Trace Elements Contaminants in Sediments from the NOAA National Status and Trends Programme Compared to Data from Throughout the World”, Chem. Ecol., Vol. 7, (1992), pp. 31–50.

[7] I. Stanimirova, S. Tsakovski and V. Simeonov: “Multivariate Statistical Analysis of Coastal Sediment Data”, Fres. J. Anal. Chem., Vol. 365, (1999), pp. 489–493.

[8] V. Simeonov, S. Tsakovski and D. L. Massart: “Multivariate Statistical Modeling of Coastal Sediment Data”, Toxilogol. Environ. Chem., Vol. 72, (1999), pp. 81–92.

[9] V. Simeonov, I. Stanimirova and S. Tsakovski: “Multivariate Statistical Interpretation of Coastal Sediment Monitoring Data”, Fres. J. Anal. Chem., Vol. 370, (2001), pp. 719–722.

[10] K. Esbensen, S. Schoenkopf and T. Midtgaard: Multivariate Analysis in Practice, CAMO AS, Trondheim, 1994.

[11] B. Vandeginste, D. L. Massart, L. Buydens, S. De Jong, P. Lewi and J. Smeyers-Verbeke: Handbook of Chemometrics and Qualimetrics, Elsevier, Amsterdam, 1998.

[12] C. Sârbu and H.F. Pop: “Fuzzy Soft-Computing Methods and Their Applications in Chemistry”, In: K.B. Lipkowitz, D.B. Boyd and T.R. Cundari (Eds.): Rev. Comput. Chem., Chapter 5, Wiley-VCH, 2004, pp. 249–332.

[13] T. Cundari, C. Sârbu and H.F. Pop: “Robust Fuzzy Principal Component Analysis (FPCA). A Comparative Study Concerning Interaction of Carbon-Hydrogen Bonds with Molybdenum-Oxo Bonds”, J. Chem. Inf. Computer Sci., Vol. 42, (2002), pp. 310–321.

[14] H.F. Pop and C. Sârbu: “A New Fuzzy Regression Algorithm”, Anal. Chem., Vol. 68, (1996), pp. 771–780.

[15] C. Sârbu: “Use of Fuzzy Regression for Calibration in TLC-Densitometry”, J. AOAC Internat., Vol. 83, (2000), pp. 1463–1467.

[16] K.Y. Lee: “Local Fuzzy PCA based GMM with Dimension Reduction on Speaker Identification”, Pattern Recognition Lett., Vol. 25, (2004), pp. 1811–1817.

[17] P. Piraino, E. Parente and P.L.H. McSweeney: “Processing of Chromatographic Data for Chemometric Analysis of Peptide Profiles from Cheese Extracts: A Novel Approach”, J. Agricultur. Food Chem., Vol. 52, (2004), pp. 6904–6911.

[18] C. Sarbu and H. F. Pop: “Principal Components Analysis versus Fuzzy Principal Components Analysis: A Case Study: The Quality of Danube Water (1895-1996)”, Talanta, Vol. 65, (2005), pp. 1215–1220.