Asian Journal of Agriculture and Food Science (ISSN: 2321 – 1571)
Volume 02 – Issue 02, April 2014

Using of Multidimensional Statistical Method to Analyse of Heavy Metals Contaminated Soil

Ružena Králiková[1], Miriam Andrejiová[2]

[1] Department of Environmentalistics, Faculty of Mechanical Engineering
Technical university of Kosice, Letna 9, Kosice, Slovak Republic
Email: ruzena.kralikova {at} tuke.sk

[2] Department of Applied mathematics and Informatics, Faculty of Mechanical Engineering
Technical university of Kosice, Letna 9, Kosice, Slovak Republic
Email: miriam.andrejiova {at} tuke.sk

ABSTRACT— ***Environmental pollution has recently become one of the key social issues. Deterioration of the environment with varying degrees of devastation in individual regions negatively contributes to public health and the quality of the entire ecosystem. Due to its use of agrochemicals, the agriculture itself along with individual industries cause the soil pollution with heavy metals. The paper investigates the degree of heavy metal contamination of agricultural soil in selected Slovakian districts over a monitored period of time. Selected multidimensional statistical methods were applied to analyze and evaluate the degree of soil contamination.***

Keywords— soil c1ontamination, principal of component analysis, cluster analysis

---

1. INTRODUCTION

Slovakia is situated in the centre of Europe and covers an area of about fifty thousand km^2, and in so doing agricultural land accounts for almost half the territory, while chernozems (black soils) represent 60 percent of the utilized agricultural soil. Soil surveys aimed at soil identification and assessment are a pre-requisite for rational utilization and protection of both production and non-production soil functions in Slovak Republic, including the modelling and development of optimization programs for protection and effective utilization of soil as one of the basic components of the environment. Soil parameters as well as its production and environmental functions in space and time are determined according to soil monitoring in order to identify and forecast its evolutionary changes. Under a bill that amends and supplements Act No. 220/2004 Coll. on protection and utilization of agricultural soil and on amendment to Act No. 245/2003 Coll. on integrated pollution prevention and control and on amendment to, and supplementation of, certain acts, as amended, soil is defined as an irreplaceable natural resource with several innate functions worth utmost protection. The bill emphasizes the purpose of agricultural soil as a non-recoverable natural resource and a basic component of the environment [8]. Characteristic features of the agricultural soil fund of Slovakia are a long-term, gradual decrease in its acreage, reclassification of productive plough-land to other soil categories, a decrease in the acreage of all categories of farmland, and a significant increase in surplus unused agricultural land.

Agricultural ecosystems differ from natural systems in that they are disturbed strongly by human practices such as harvesting, tillage and sowing. The basic function of agricultural production lies in food production to feed a population, performing thus a production function [17]. The European Union enforces the principle of multifunctionality in agriculture, which is not solely oriented at food production and the efficiency of production. The multifunctionality is defined by three basic activities: the production of industrial raw materials, forage and, to a certain degree, also foodstuff, the preservation of rural environment and countryside, and the development of social infrastructure and the structure of rural communities. From this perspective, an areal survey of agricultural soil contamination is of great importance and requires special attention. Any industrial influences and possible negative effects of applied herbicides are mainly due to chemical and textural parameters of treated soil. Organic and inorganic contaminants are determined in soil samples based on monitoring activities [5], [6]. Since the representativeness and objectivity of a monitoring network is statistically conditioned by its density and spatial arrangement, in terms of statistics it is a selection limited by the allotted funding as well as the organizational possibilities of the monitoring system [3].

2. MATERIALS AND METHOD

2.1 Soil monitoring

The in situ of agricultural soil contamination in the Slovak Republic is conducted in 5-year cycles by the Department of Environment and Ecological Agriculture of the Central Supervisory and Testing Institute of Agriculture in Bratislava. The limits for selected monitored soil parameters ($mg.kg^{-1}$ of dry matter) are stipulated in a decision issued by the Ministry of Agriculture and Rural Development of the Slovak Republic to Act. No.531/1994-540. *Cycle 1* of the Area Survey of Soil Contamination (ASSC) was conducted from 1991 to 1995. Soil samples of 19 257 plots, which corresponds to 782 905 ha, were analyzed in this period. With respect to analyses, 21 322 soil samples were analyzed in total within the ASSC program, and 107 314 analyses were performed while monitoring 40 different parameters (10 inorganic, 30 organic) [9]. *Cycle 2* was conducted from 1996 to 2000. During this cycle were analyzed 8 921 soil samples, the total count of analyses performed being 105 351 (monitoring 56 parameters, 13 inorganic, 43 organic). In *Cycle 3* (from 2001 to 2005) were analyzed 7 605 soil samples, the total count of analyses performed being 96 665 (monitoring 56 different parameters, 13 inorganic, 43 organic). The main goal of the second and third cycle was to monitor soil contaminants in selected cadastral areas. These were shortlisted by the Department of Environment and Ecological Agriculture on the basis of increased levels of contaminants in the soil analyzed in Cycle 1 and 2. Those cadastral areas were selected where the contamination limit had been exceeded at least in one of the defined parameters [1], [2]. To get a full picture of soil contamination in Slovakia, the data files include also the results of analyses made in the cadastral areas that are a part of the Coordinated Targeted Monitoring System. *Cycle 4* has been in progress since 2006. The task is to keep monitoring levels of contaminants in the soil in selected cadastral areas. The selection was made according to higher contamination levels identified in the soil analyses conducted in previous cycles [13], [14].

2.2 Statistical analysis

Multidimensional statistical methods were applied in the analysis of the degree of soil contamination with heavy metals in Slovakia. Using these statistical methods in the monitoring of soil contamination can be found in many works [11], [12], [18].

Principal Component Analysis (PCA) is one of the basic multidimensional methods. Its aim is to reduce the number of interrelated variables down to principal components with as little information loss as possible [7], [16]. The first principle component corresponds to the greatest input variability; each next component contributes to the variability less and less. PCA can also be applied to detect outliers, which sometimes need to be eliminated from further analyses, and it is a useful method of classification of objects into clusters [10], [16].

Cluster analysis belongs to multidimensional statistical methods. Its goal is to group the monitored objects into certain similar, homogeneous groups, known as clusters. Its graphic representation is a dendrogram. There is a distinction between hierarchical and non-hierarchical methods according to the cluster formation mode [4], [10], [15]. The most common hierarchical methods are the nearest neighbour method (single linkage method), the furthest neighbour method (complete linkage method), average linkage method, centroid method, median method and Ward's method. One of the most common non-hierarchical methods is the k-means clustering.

The goal of a *multidimensional comparison* is to replace several indicators (variables) with a single and integrating indicator according to which objects can be arranged. Basic multidimensional comparison techniques include the simple ordering method, the scoring method, the standardized variable method, and the fictitious point distance method. Before their application, it is important to determine the type of variables, i.e. whether they are stimulating, de-stimulating or nominal [16].

3. RESULT AND DISCUSSION

The degree of soil contamination with heavy metals in selected Slovak regions are characterized by six variables, they represent the levels of a particular heavy metals in a soil sample ($mg.kg^{-1}$): Chrome (Cr), Nickel (Ni), Cadmium (Cd), Arsenic (As), Mercury (Hg), and Lead (Pb). 63 districts were included in the areal soil contamination survey, the measured values of individual variables representing the mean values of the first years of Cycle 4. Zinc and Copper were not included in the analysis since these parameters had only been monitored in certain districts in the given period of time. The districts in which no soil samples had been taken during the period under review were not included in the survey either.

It was primarily determined in the data analysis and evaluation whether there was any correlation between the input variables, i.e. there is a dependency between them. Since the surveyed variables were measured in various units, a correlation matrix was used to identify the dependencies (Table 1). It follows from the correlation matrix that there is a strong correlation between Pb-Cd ($r=0.76$). A significant correlation also exists between Ni-Cr ($r=0.60$), and Cd-Cr ($r=0.56$).

Table 1: Correlation Matrix

	Cr	**Ni**	**Cd**	**As**	**Hg**	**Pb**
Cr	1.00					
Ni	**0.60****	1.00				
Cd	**0.56*****	0.28	1.00			
As	-0.03	-0.09	0.29	1.00		
Hg	-0.04	-0.05	0.11	0.21	1.00	
Pb	0.31	-0.02	**0.76*****	0.47	0.26	1.00

Note: *p<0.05, **p<0.01, ***p<0.001

Because of the relatively strong correlation between these pairs, PCA was applied to identify new, independent variables. The original variables were replaced with new, independent ones, known as *principal components.* To extract the principal components we use the eigenvalues which are indicated in the first row of Table 2. The second row presents the proportion of variability atributable to the individual principal components. In the third row are cumulative proportions of variability (in %). The first principal component accounts for approximately 40.8 percent of the total data variance, the second for 26.5, the third for 14.4, and the fourth one for 10.8 percent of the total variance. The last two principal components account for approximately 7.5 percent of the total variance (Table 2).

Table 2: Summary of Principal components analysis

	Comp.1	**Comp.2**	**Comp.3**	**Comp.4**	**Comp.5**	**Comp.6**
Eigenvalues	2.446	1.591	0.864	0.646	0.288	0.165
Proportion (%)	40.8	26.5	14.4	10.8	4.8	2.7
Cumulative (%)	40.8	67.3	81.7	92.5	97.3	100

However, there are several techniques how to determine an appropriate number of principal components. A commonly used criterion is the Kaiser-Guttman rule, according to which only those eigenvalues that are larger than 1 are incorporated in the model [7], [10], [16]. According to another rule, there should only be so many principle components that account for 70 to 90 percent of the total variance. It means that the first three principal components, which cumulatively account for more than 81.7 percent of the total data variance, were used to define the new variables. Eigenvectors represent weights, which make combination of standardized variables possible. Eigenvector coefficients, the so-called component score coefficient matrix for the first three principal components are shown in Table 3.

Table 3: Component score for the first three principal components

Parameters	**Comp.1**	**Comp.2**	**Comp.3**
Cr	0.456	**-0.447**	-0.050
Ni	0.268	**-0.574**	-0.298
Cd	**0.576**	0.003	0.166
As	0.282	**0.484**	0.206
Hg	0.157	0.383	**-0.900**
Pb	**0.533**	0.300	0.169

The first component accounts for 40.8 percent of the total variance with important positive loadings of the variables Cd and Pb. The second component has strong loadings of the variables Cr, Ni and As. The third component is mainly composed of the variable Hg.

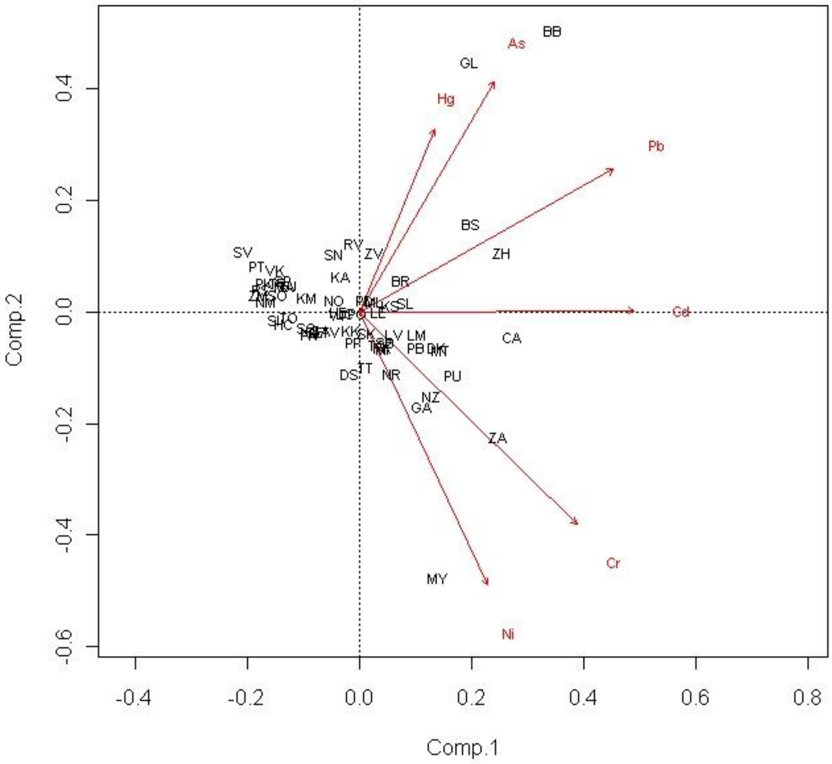

Figure 1: Biplot (R package)

Figure 1 is a graphic representation of the original variables in the coordinate system of one principle component in relation to another (biplot). For instance, it follows from the graph that there is an abundance of the variable Pb in Ziar nad Hronom (ZH) and Banska Stiavnica (BS) districts. There is a high incidence of the variable Cd in Cadca (CA) district, the variable As in Banska Bystrica (BB) district, the variable Cr in Zilina (ZA) district and the variable Ni in Myjava (MY) district. Biplot sorted the counties into several clusters according to a similar degree of contamination. There is a large cluster composed mostly of the counties with a lower degree of soil contamination with heavy metals. Gelnica (GL) and Banská Bystrica (BB) Counties seem to be significantly out of line.

The *cluster analysis* was applied to sort and group the districts with similar characteristics based on soil contamination with heavy metals. Hierarchical agglomerative methods were chosen subsequently. The Euclidean distance, being the most common technique, was used as the distance metric, and the three principle variables generated via PCA were selected as the input variables. The Cophenetic correlation coefficient (CCC) was used to verify the results and determine the best clustering method (Table 4). The highest coefficient value defines the best clustering method. The closer its value is to 1, the more appropriate the hierarchical agglomerative clustering method is for the analysed data structure definition [10], [15]. The results indicate that average-link clustering and the centroid method seem to be the most suitable clustering techniques.

Table 4: The Cophenetic correlation coefficient (CCC)

Methods	CCC	Methods	CCC
Average linkage method	**0.928**	Single linkage method	0.899
Ward's method	0.567	Complete linkage method	0.788
Centroid method	**0.911**	Median method	0.895

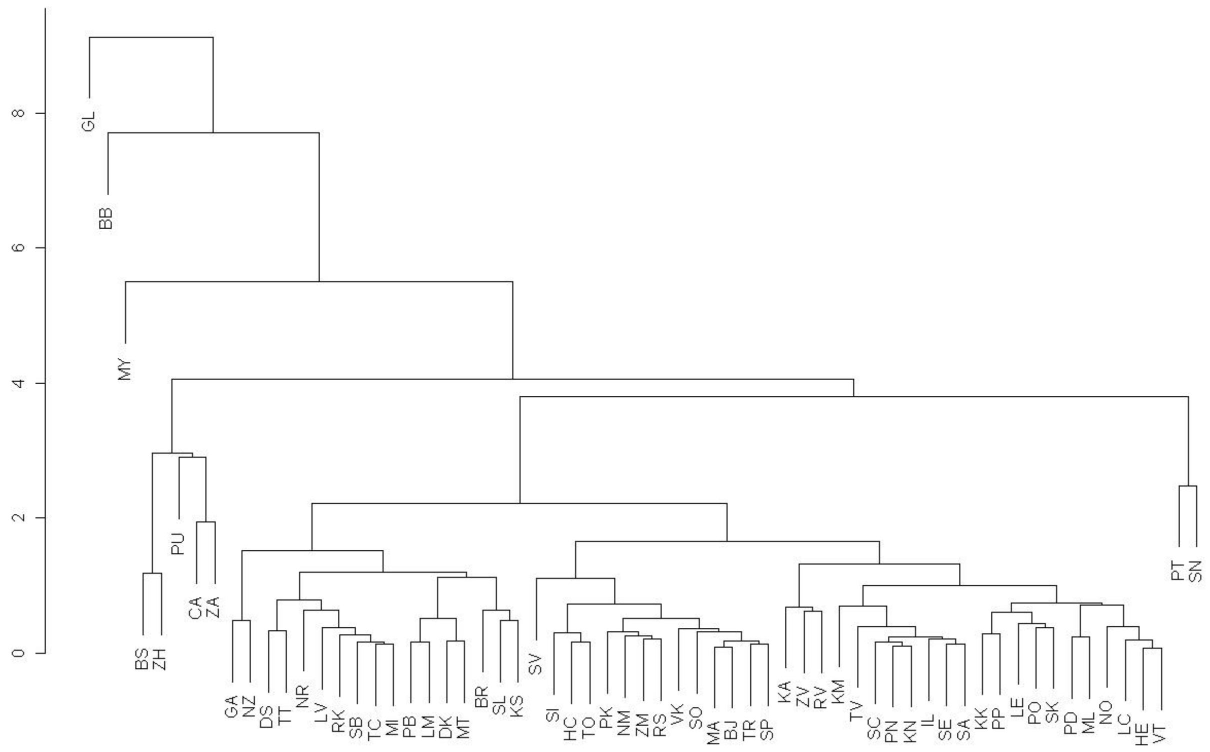

Figure 2: Cluster dendrogram – Average linkage method (R package)

The district similarity dendrogram obtained through average linking clustering represents the ultimate cluster analysis result (Figure 2). One of the dimensions stands for the Counties and the other for the distance between them. According to the dendrogram, Gelnica (GL), Myjava (MY), and Banska Bystrica (BB) Counties are very different from the others. These three are clearly separated and form independent groups. The Banska Bystrica district has been recently marked by mining activities, which results into exceeded limits of heavy metals values in surrounding nowadays. There is a galvanic sediment junkyard in Myjava district, which comes from production of metals and metal products, and has been affected by environmental inappropriate technological applications. A soil of Gelnica district is burdened by natural resources mining, mechanical engineering and non-ferrous metals metallurgy.

According to [8], they can be considered distant objects that may affect clustering. Poltár (PT) and Spisska Nova Ves (SN) districts form another independent group. Districts Cadca (CA), Zilina (ZA), Banska Stiavnica (BS) and Ziar nad Hronom (ZH) constitute a larger cluster. Just like the previous districts, these also belong to the ones with higher parameter values, which characterize higher degree of soil contamination with heavy metals. The next three large clusters are composed of the counties with average, below-the-limit values of the parameters under review.

The districts under consideration were arranged according to the degree of heavy-metal contamination, using multidimensional comparison. The degree of contamination was ascertained by means of two such techniques, namely the standardized variable method and the fictitious object method [16]. All indicators were considered de-stimulating variables (their decrease being a positive manifestation). The results of both applied methods of multidimensional comparison are comparable and show, that Gelnica (GL), Banska Bystrica (BB), Puchov (PU), Zilina (ZA), Cadca (CA) and Ziar nad Hronom (ZH) belong to those with the highest degree of soil contamination. On the contrary, Svidnik (SV), Zlate Moravce (ZM), Pezinok (PK), Rimavska Sobota (RS) and Nove Mesto nad Vahom (NM) belong to the five counties with the lowest degree of soil contamination with heavy metals.

The mentioned areas that are affected by mining activities and metals production, engineering, energy industry, transport, and which are nowadays marked out by heterogeneousness, and exceeded limit values of monitored parameters (heavy metals) in soils, represents environmental burdens. Ministry of Environment of Slovak Republic recommends the areas for their detailed research, risk analysis, possibly some kind of reconstruction in the near future.

4. CONCLUSION

Multidimensional statistical methods were applied in the survey of areal pollution of agricultural soil in Slovakia. The survey was oriented at monitoring of soil contamination with the six most significant heavy metals

(As, Cd, Cr, Hg, Ni, Pb), taking into consideration the percentage of contaminated soil in the analyzed samples in 63 districts. The original dependent variables, whose values represented the mean indicator values in individual districts over the monitored period of time, were replaced with three independent input variables. These were determined according to PCA and cumulatively accounted for almost 80 percent of the total variance. The results of the hierarchical agglomerative method of average-link clustering pointed to the fact that Gelnica (GL), Banska Bystrica (BB), and Myjava (MY) districts are outlying (in compared to other districts).

For instance, the average mercury levels in Gelnica district were 2.69 $mg.kg^{-1}$, whereas the limit stipulated by the Ministry of Agriculture and Rural Development of the Slovak Republic is 0.3 $mg.kg^{-1}$. The average arsenic levels measured in Banska Bystrica district over the given period were 16.59 $mg.kg^{-1}$, which is 4.15 times the admissible limit (5 $mg.kg^{-1}$). The average nickel levels in Myjava district (8.85 $mg.kg^{-1}$) were close to the limit (10 $mg.kg^{-1}$). Parameter values exceeded the limits in 48 percent of the surveyed plots in Spisska Nova Ves district and 65.4 percent of plots in Gelnica district. The results of the areal soil-contamination survey, published in the 2008 and 2009 Report, indicate that the highest chrome levels (90 $mg.kg^{-1}$, 2007 ASSC), which were nine times the limit, were measured in Zilina district over the period under consideration. Nickel levels that were 3.8 times the limit were measured in Myjava district (38 $mg.kg^{-1}$, 2007 ASSC). The highest copper levels over the given period of time were found in Gelnica district (83.77 $mg.kg^{-1}$, 2006 ASSC), although this parameter was not evaluated (limit 20 $mg.kg^{-1}$). Greatly over-the-limit levels of arsenic (185.90 $mg.kg^{-1}$) were established in Banska Bystrica district during the 2006 ASSC. This value was 37.2 times the limit, which is incredible (limit 5 $mg.kg^{-1}$). Excessive levels of lead were found in Banska Stiavnica district in the 2006 ASSC (159 $mg.kg^{-1}$, limit 30 $mg.kg^{-1}$). The highest mercury level (Gelnica, 8.07 $mg.kg^{-1}$, 2007 ASSC) was almost 27 times the admissible limit (0.3 $mg.kg^{-1}$).

According to our survey results, the districts with a higher degree of agricultural soil contamination with heavy metals are located in different regions of Slovakia, and the degree of contamination depends, among other factors, on the type of soil, mining activities, local industries, traffic, or usage of artificial fertilizers and pesticides. The overall soil condition is significantly affected not only by current public attitude and approach to this topic, but especially by careless behaviour in the past.

Industrial activities had released to the environment many toxic chemicals, that is heavy metals and persistent organic pollutants due to accidental spills or improper management. It resulted in many contaminated sites all over world. Soil, sediment, and groundwater contamination has been a major problem at these polluted sites, which need urgent decontamination to protect public health and environment. At present many decontamination technologies are found to be ineffective or expensive to remediate sites with heterogeneous subsurface conditions and contaminant mixtures. There is an urgent need to develop new technologies that can overcome these challenges and that will cost-effective.

5. ACKNOWLEDGEMENT

This paper was prepared with the support of grant project KEGA 032/TUKE/012 (50%) and project OPVaV-2008/2.2/01-SORO (50%).

6. REFERENCES

[1] Barancikova, G., et al., Soil monitoring of Slovak Republic. Present state and development of monitored soils as the base to their protection and next land use (1997-2001 years), VUPOP, Bratislava, 2002.
[2] Barancikova, G., et al., Soil monitoring of Slovak Republic. Present state and development of monitored soils as the base to their protection and next land use (2002-2006 years), VÚPOP, Bratislava, 2009.
[3] Duris, M., "Environmental functions of soils in agricultural land White Carpathians and Myjavská pahorkatina", Acta Environmentalica Universitatis Comenianae, vol. 14, no.2, pp. 31-40, 2006.
[4] Everitt, B. S., Landau, S., Leese, M., Cluster Analysis, Wiley, London, 2001.
[5] Hiller E., Cernansky S., Krascsenits Z., Milicka J., "Effect of soil and sediment composition on acetochlor sorption and desorption", Environmental Science and Pollution Research, vol. 16, no. 5, pp. 546-554, 2009.
[6] Hofman J., Dusek L., "Biochemical analysis of soil organic matter and microbial biomass composition - a pilot study", European Journal of Soil Biology, vol. 39, no. 4, pp. 217-224, 2003.
[7] Jollifte, I. T., Principal Component Analysis, Springer– Verlag, New York, 2002.
[8] Kralikova, R., Badida, M., Halasz, J., Technique for the environmental protection, Elfa, Košice, 2007.
[9] Linkes, V. et al, Soil monitoring of Slovak Republic. Present state of monitored soil properties, VÚPÚ, Bratislava, 1997.

[10] Meloun, M., Militky, J., Compendium of statistical data processing, Academia, Praha, 2006.
[11] Morrison, J. M. et al., "Cluster analysis of a regional-scale soil geochemical dataset in northern California", Applied Geochemistry, vol. 26, pp. 105-107, 2011.
[12] Muhammad, S., Shah, M. T., Khan, S., "Heavy metal concentrations in soil and wild plants growing around Pb-Zn sulphide terrain in the Kohistan region, northern Pakistan", Microchemical Journal, vol. 99, no. 1, pp. 67-75, 2001.
[13] Novakova, K., Report on Surface soil contamination survey – 2008, Department of Environment and Ecological Agriculture, Bratislava, 2008.
[14] Novakova, K., Heroldova, T., Report on Surface soil contamination survey – 2009, Department of Environment and Ecological Agriculture, Bratislava, 2009.
[15] Rezankova, H., Husek, D., Snasej, V., Clusters analysis of data, PB, Praha, 2009.
[16] Stankovicova, I., Vojtkova, M., Multidimensional statistical methods with application, IURA Edition, Bratislava, 2007.
[17] Zoborsky, I. M., Agricultural Economy, SPU, Nitra, 2006.
[18] Wong, S. C. et al., "Heavy metals in agricultural soils of Pearl River Delta south China", Environmental Pollution, vol.119, no. 6, pp. 33-44, 2002.