CONTEMPORARY AGRICULTURE / *SAVREMENA POLJOPRIVREDA* 60 (1-2) 122-129, 2011.
UDC: 63(497.1)(051)-"540.2" ISSN 0350-1205

UDC: 658.2:007.2

APPLICATION OF PRINCIPAL COMPONENT ANALYSIS TO REALIZE VARIABILITY OF CHEMICAL PROPERTIES OF CHERNOZEM*

SLOBODAN NIĆIN, LJILJANA NEŠIĆ, VLADIMIR ĆIRIĆ[1]

SUMMARY: This paper presents a practical application of the principal components analysis (PCA), as one of the methods of multivariate analysis used for reducing the dimensionality of the problem (reducing a set of a large number of variables down to a smaller number of mutually uncorrelated linear combinations). Seventeen soil properties from 95 samples were analyzed in order to single out the soil properties which have significant impact on soil variability. The paper also presents a comparison of certain results obtained by using this method and by the application of factor analysis. This research showed that heavy metals contribute most to the first principal component, organic matter contributes most to the second, while inorganic compounds (oxides) contribute most to the third component. Accordingly, it is these variables that contribute most also to accountability of the total variability of the observed phenomenon. The determined accountability of the total variability by the first principal component was around 35%, by the second around 20%, while with the third it was 14%, which means that the first three dimensions together account for around 70% of the total variability of the observed set of variables.

Key words: *multivariate analysis, principal component analysis, variability, eigenvalues, soil, soil properties.*

INTRODUCTION

One of the main objectives in soil analysis is an integrated overview of its properties as indicators of soil fertility. The methods used in multivariate statistics are considered to be powerful tools for obtaining these integrated approaches helping researchers to obtain as much information as possible from available data. Sena et al. (2002) argue that soil scientists should be better acquainted with multivariate methods, which, in

Original scientific paper / *Originalni naučni rad*
[1]Slobodan Nićin Msci, research assistant, Ljiljana Nešić PhD, associate professor, Vladimir Ćirić MSci, teaching assistant, Faculty of Agriculture, Novi Sad.
Corresponding author: Slobodan Nićin, Faculty of Agriculture, Trg D. Obradovića 8, 21000 Novi Sad, R. Serbia. Tel.: +381 21 485-3380, e-mail: bobanicin@yahoo.com
*The paper is a part of the research work on the project "The quality of agricultural land, safe use of pesticides and removal of arsenic from water in order to protect the environment", financed by the Provincial Secretariat for Science and Technological Development, from 2007 to 2008.

comparison to univariate and bivariate methods, greatly increase the capacity of extraction and interpretation of soil-related data.

Multivariate analysis is an analysis of a large number of variables, which enables their examination and quantification, as well as identification of their dependence, *i.e.* links between a larger number of variables.

Principal components analysis (PCA) is one of many methods of multivariate analysis. This method enables transformation of a large number of variables into a smaller number of latent variables (principal components, PCs) which are not inter-correlated. These transformed variables represent linear functions of input variables.

Kovačić (1994) stresses the importance of application of the PCA in socio-economic research where it is necessary to reduce an exceptionally large number of indicators of this type to a small number of representative indicators. Many studies (Chang et al., 2001; Visconti et al., 2009; Kummer et al., 2010) showed purposefulness of applying the PCA in examining interrelation of soil properties and their impact on soil variability.

According to Koutsoyiannis (1977), the PCA is just a special case of a more general factor analysis (FA) model. This author argues that formation of a series of variables from the existing variables is the objective of the PCA. The first linear combination called the principal component PC takes the maximum possible share of the total variability in the series consisting of all input variables, the second PC takes the maximal possible remaining variability, and the third takes then the maximal possible remaining variability, etc.

The PCA has a wide application in sociological, economic, biological research as well as the research connected to crop and animal production. The advantages of applying the PCA in such studies are especially important when a large number of variables are to be analyzed. The benefits of PCA analysis and biplot are even greater when applied to larger sets of soil-related data (Sena et al., 2002).

One of the most important applications of the PCA is in regression analysis, where, apart from reducing dimensionality in multiple regression analysis, it also reduces multicollinearity. This stems from the fact that multiple regression analysis is performed using a small number of transformed mutually uncorrelated variables.

The aim of this research is to apply the PCA on a practical example related to soil research and to single out soil proprieties which have statistically significant impact on soil variability.

MATERIAL AND METHODS

Research site and sampling methods

Sampling was conducted on the territory of Srem while the localities on which the samples were taken were determined with a net of squares of 4 x 4 km, which was drawn on the existing pedologic map of Vojvodina R 1: 50,000. In this way, every sample is representative for 1 600 ha of soil. In this research the properties of different soil types were examined, and for the purposes of this paper the data for the soil type chernozem were used. Chernozem belongs to the class of humus-accumulative soils (Miljković, 1996) with the average humus accumulation in Vojvodina about 165 t C ha^{-1} up to 100 cm of depth (Manojlović et al., 2010). Chernozem is naturally fertile soil but it requires significant amount of fertilizers in order to obtain high yield (Bogdanović et al., 2010).

The research included a total of 95 samples. Considering high diversity of the parent material of soil and the climate in Srem, the samples collected in this way are significantly different regarding their physical and chemical properties, thus representing a broad spectrum of different properties for the analysis even within the same soil type. Soil samples were collected in disturbed state in accordance with the propositions of the System of Soil Fertility Control, *i.e.* with an soil probe, up to the depth of 30 cm, in accordance with the system of circular control plots. Prior to chemical analysis, the samples were sieved through a 2 mm sieve and homogenized. Each sample is also accompanied by the data on its precise position with the coordinates and altitude, since GPS technology was used for location of the sampling sites.

Soil analyses

In each analyzed sample, 17 soil properties, *i.e.* variables were analyzed: pH in KCl, pH in water, $CaCO_3$, humus, total nitrogen, P_2O_5, K_2O, copper, cobalt, manganese, arsenic, chromium, nickel, cadmium, lead, zinc and iron. The pH values in KCl and in water were determined potentiometrically, while the values of $CaCO_3$ content were determined by volumetric method on Scheibler's calcimeter. The humus content was determined by wet oxidation of organic carbon with $K_2Cr_2O_7$ and multiplying the carbon with a coefficient (1.724), while the content of the total nitrogen was obtained by calculation from the humus content. The amount of easily soluble P_2O_5 and K_2O in soil was determined by Al-method. The total content of microelements and heavy metals (Cu, Zn, Mn, Pb, Co, Cr, Ni, Cd, As and Hg) extracted with concentrated HNO_3, was determined using inductively coupled plasma on ICP-OES VistaPro Varian, while the Fe content was determined using AAS.

Statistical analysis

Detailed descriptive statistics of the results of examined samples was conducted. The principal components analysis (PCA) was performed in the following way: a) correlation matrix of soil variables was developed; b) varimax rotation was applied on the resulting matrix; c) component defining variables were selected (variables that saturate most the PC). Descriptive analysis of the data and the PCA were performed using the software package STATISTICA 10.0.

RESULTS AND DISCUSSION

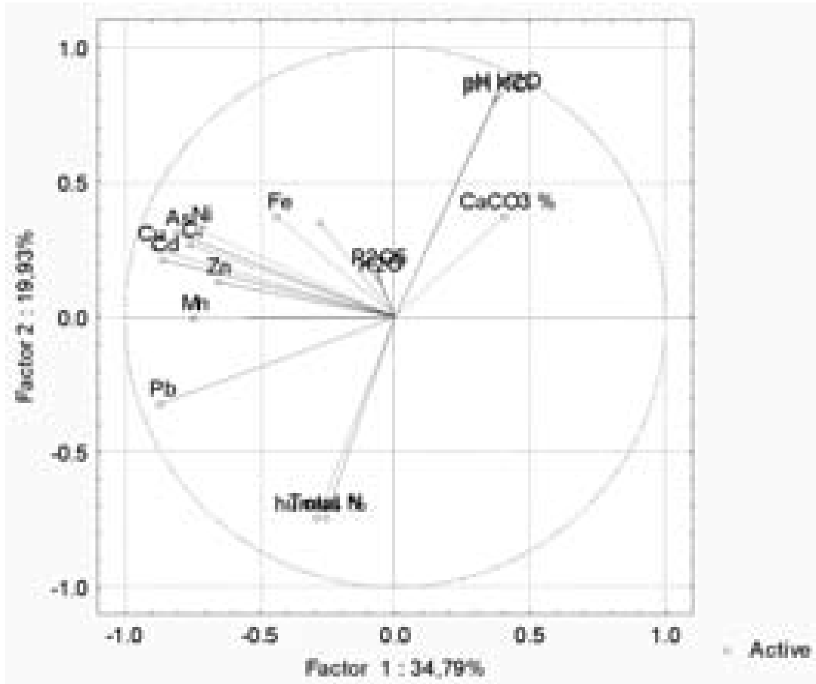
Descriptive statistics of analyzed soil properties (variables) are given in Table 1. Projection of the variables in the factorial plane (Graph. 1) indicates that the variables cobalt, lead and cadmium contribute most to the first PC (which accounts for 34.79% of variability), and thus to the total variability of the basic set. The second PC (which accounts for 19.93% of variability) is contributed most by the variables pH in KCl, pH in H_2O, humus and the total nitrogen. In total, the first and second PCs account for 54.72% of the total variability of the observed set. The third PC accounts for 14% of the total variability of the observed set. Accordingly, three PCs together account for around 70% of the total variability of the initial set of variables. With the first three components (Giuffré et al. 2006) explained 90% of the overall variation in Argentinian Argiudolls. The greatest contribution to the third PC is given by the variables P_2O_5 and K_2O. Onweremadu et al. (2007) argue that by applying PCA on 12 variables in the soils of Nigeria, 28.5% of the total variance was accounted for with pH value (first PC), and 27.3% with organic carbon (second PC), which means that pH values and organic carbon are highly important for sustainability of these soils.

Table 1. Descriptive statistics of soil properties
Tabela 1. Deskriptivna statistika osobina zemljišta

Soil properties / *Osobine zemljišta*	Mean *Srednja vrednost*	Standard deviation *Standardna devijacija*
pH (KCl)	6.60	1.117
pH (H_2O)	7.79	1.003
$CaCO_3$ (%)	5.88	8.060
Humus (%)	2.90	1.060
Total / *Ukupni* N(%)	0.21	0.058
P_2O_5 (mg/100g)	18.62	33.644
K_2O (mg/100g)	21.25	15.905
Cu (mg/kg)	24.34	13.792
Co (mg/kg)	12.50	3.609
Mn (mg/kg)	627.10	367.605
As (mg/kg)	7.74	2.949
Cr (mg/kg)	31.92	14.309
Ni (mg/kg)	46.55	32.330
Cd (mg/kg)	0.34	0.094
Pb (mg/kg)	22.13	5.714
Zn (mg/kg)	48.40	18.086
Fe (mg/kg)	25258.42	3349.903

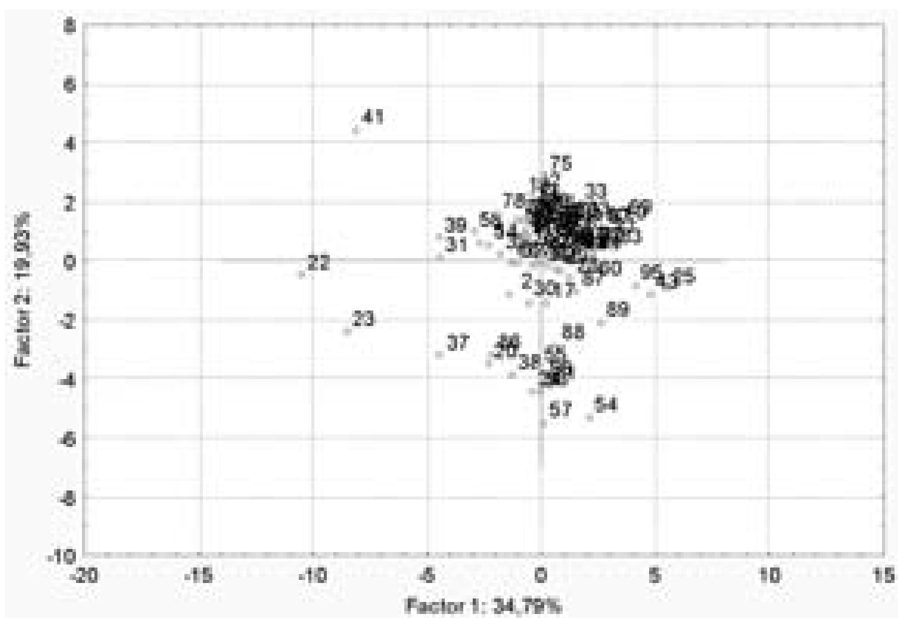
Factor coordinates of variables based on the coefficients of correlation indicate that the highest values of these coefficients are for the variables cobalt (-0.90), lead (-0.87) and cadmium (-0.86) (compared with the factor no. 1) and variables pH in KCl (0.82), pH in H_2O (0.81) and humus and the total nitrogen (-0.74 each) (compared with the factor 2).

Factor coordinates of individual observations (Graph. 2) indicate that the total variability is influenced most by the samples numbered 22 (-10.5), 23 (-8.5), 41 (-8.1), 25 (5.5) and 43 (4.8) (for factor no. 1) and the samples numbered 57 (5.51), 54 (5.32), 41 (4.44), 24 and 53 (4.43) and 21 (4.23) (for factor no. 2).



Graph. 1. Projection of variables in factorial plane

Grafikon 1. Prikaz promenljivih u faktorskoj ravni

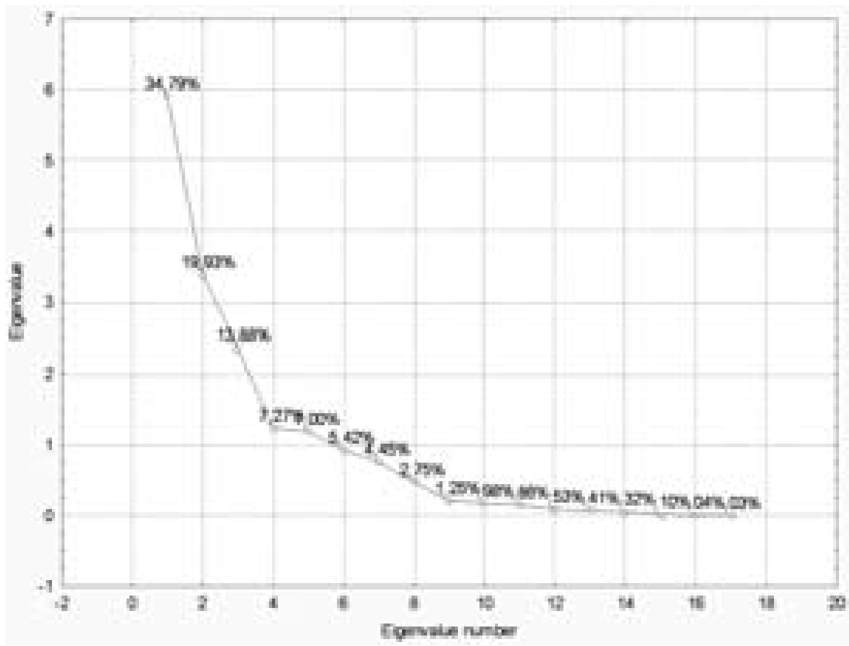


Graph. 2. Projection of cases in factorial plane

Grafikon 2. Prikaz pojedinačnih posmatranja u faktorskoj ravni

Eigenvalues (Graph. 3) indicate that the first factor accounts for slightly over one third of the total variability of the observed set (34.78 %). The second factor accounts for 19.93 % variability, the third for 13.88 %, the forth 7.27 %, the fifth 7.00 %, the sixth 5.42 %, the seventh 4.45 %, the eighth 2.74 %, the ninth 2.03 %, the tenth 0.98 %, the eleventh 0.86 %, the twelfth 0.53 %, the thirteenth 0.41 %, the fourteenth 0.32 %, the fifteenth 0.10 %, the sixteenth 0.04 % and the seventeenth 0.03 %. The first three factors account for slightly over two third of the total variability (68.60 %) (Table 2).

Input values (unrotated) for the factors in application of FA show that variability is accounted for by the same variables as with the PCA.



Graph. 3. Scree- plot of eigenvalues
Grafikon. 3. Scree dijagram svojstvenih vrednosti

Rotated input values (varimax rotation) for the factors in FA indicate that the first factor is largely accounted for by the variables cobalt (0.93), cadmium (0.88) and arsenic (0.86). The second factor is most accounted for by the variables pH in H_2O (-0.91), pH in KCl (-0.89) and humus (0.80). The factor in FA represents linear combination of a larger number of variables.

Table 2. Eigenvalues on the basis of correlation correlation matrix
Tabela 2. Svojstvene vrednosti na bazi korelacione matrice

Eigenvalue *Svojstvena vrednost*	% Total variance *% Ukupna varijacija*	Cumulative eigenvalue *Kumulativna sv. vrednost*	Cumulative % *Kumulativ %*
5.913948	34.78793	5.91395	34.7879
3.388275	19.93103	9.30222	54.7190
2.359530	13.87959	11.66175	68.5985
1.235977	7.27045	12.89773	75.8690
1.190190	7.00112	14.08792	82.8701
0.921563	5.42096	15.00948	88.2911
0.756603	4.45060	15.76609	92.7417
0.467134	2.74784	16.23322	95.4895
0.211650	1.24500	16.44487	96.7345
0.166863	0.98155	16.61173	97.7161
0.146395	0.86114	16.75813	98.5772
0.089806	0.52827	16.84793	99.1055

0.068879	0.40517	16.91681	99.5107
0.054654	0.32150	16.97147	99.8322
0.017010	0.10006	16.98848	99.9322
0.007134	0.04197	16.99561	99.9742
0.004389	0.02581	17.00000	100.0000

CONCLUSION

On the basis of the performed statistical analysis using PCA it was determined that the variables cobalt, lead and cadmium (making the smallest angle with the axis of the first component), contribute most to the first PC (which accounts for 34.79% of variability), and thus also the total variability of the basic set. The second PC (which accounts for 19.93% of variability) is contributed most by the variables pH in KCl, pH in H_2O, humus and the total nitrogen. The third PC is most contributed by the variables P_2O_5 and K_2O.

In total, the first, second and third PCs account for around 70% of the total variability of the observed set of variables.

This research showed that heavy metals contribute most to the first PC, organic matter contributes most to the second PC, while inorganic compounds (oxides) contribute most to the third PC. Accordingly, these components contribute most also to accountability of the total variability of the observed phenomenon.

Application of the PCA in this research points to the conclusion that the observed and quantified variables can be used in scientific and experimental work for examining the impact of a larger number of variables in a specific case, *i.e.* in examples related to soil research.

The practical significance of this paper is that the results obtained by application of the PCA can provide better and more detailed understanding of the impact of certain factors on the observed phenomenon.

REFERENCES

BOGDANOVIĆ, D., MILOŠEV, D., ŠEREMEŠIĆ, S., JUG, I., ĐALOVIĆ, I.: Mineral nitrogen dynamic in soil of different fertility as affected by agronomic practices. Contemporary Agriculture, 59(3-4)278-288, 2010.

CHANG, C. W., LAIRD, D. A., MAUSBACH, M. J., HURBURGH, C. R.: Near-Infrared Reflectance Spectroscopy–Principal Components Regression Analyses of Soil Properties. Soil Sci. Soc. Am. J., 65:480–490, 2001.

GIUFFRÉ, L., ROMANIUK, R., CONTI, M.E., BARTOLONI N.: Multivariate evaluation by quality indicators of no-tillage system in Argiudolls of rolling pampa (Argentina). Biol Fertil Soils, 42:556–560, 2006.

KOVAČIĆ Z.: Multivarijaciona analiza. Ekonomski fakultet, Beograd, 1994.

KOUTSOYIANNIS,A.: Theory of Econometrics. The Macmillan Press ltd., London and Basingstoke, 1977.

KUMMER, L., MELO, V. F., BARROS, Y. J., AZEVEDO, J. C. R.: Scientia Agraria, Curitiba, 11(6)469-480, 2010.

MANOJLOVIĆ, M., ĆIRIĆ, V., NEŠIĆ, LJ., BELIĆ, M.: The importance of soil organic carbon for soil, agriculture and global climate change, Contemporary Agriculture, 59(1-2)194-205, 2010.
MILJKOVIĆ, N.: Osnovi pedologije. Univerzitet u Novom Sadu, Prirodno-matematički fakultet, Institut za geografiju, Novi Sad, 1996.
NIĆIN S., Primena metoda glavnih komponenata u kvantitativnoj analizi. Magistarska teza, Poljoprivredni fakultet, Novi Sad, 2008.
ONWEREMADU, E. U., OSUJI, G. E.,. ESHETT, E. T., OPARA C. C.,. IBEAWUCHI, I. I.: Characterization of Soil Properties of Owner Managed Farms of Abia and Imo States, for Sustainable Crop Production in Southeastern Nigeria. Journal of American Science, 3(1)28-37, 2007.
SENA, M.M., FRIGHETTO, R.T.S., VALARINI, P.J., TOKESHI, H., POPPI, R.J.: Discrimination of management effects on soil parameters by using principal component analysis: a multivariate analysis case study. Soil & Tillage Research, 67:171–180, 2002.
VISCONTI, F., MIGUEL DE PAZ, J., RUBIO, J. L.: Principal component analysis of chemical properties of soil saturation extracts from an irrigated Mediterranean area: Implications for calcite equilibrium in soil solutions. Geoderma, 151:407–416, 2009.

ANALIZA HEMIJSKIH SVOJSTAVA ČERNOZEMA PRIMENOM METODE GLAVNIH KOMPONENATA

SLOBODAN NIĆIN, LJILJANA NEŠIĆ, VLADIMIR ĆIRIĆ

Izvod

U ovom radu prikazana je praktična primena metoda glavnih komponenata (PCA) kao jednog od metoda multivarijacione analize koji se koristi kod smanjenja dimenzionalnosti problema (svođenja skupa velikog broja promenljivih na manji broj međusobno nekoreliranih linearnih kombinacija). Sedamnaest svojstava iz 95 uzoraka zemljišta je analizirano da bi se izdvojila svojstva zemljišta koja imaju značajan uticaj na njegovu varijabilnost. Takođe, u radu je izvršeno i upoređenje nekih od rezultata dobijenih primenom ovog metoda i dobijenih primenom faktorske analize.

U ovom proučavanju pokazalo se da teški metali najviše doprinose prvoj glavnoj komponenti, organska materija najviše doprinosi drugoj, dok neorganska jedinjenja (oksidi) najviše doprinose trećoj glavnoj komponenti. Samim tim, ove promenljive najviše doprinose i objašnjenosti ukupnog varijabiliteta posmatrane pojave. Konstatovana objašnjenost ukupne varijabilnosti pomoću prve glavne komponente iznosila je oko 35 %, kod druge oko 20 %, dok je kod treće iznosila oko 14 %, što znači da je ukupno pomoću prve tri dimenzije objašnjeno oko 70 % ukupne varijabilnosti posmatranog skupa promenljivih.

Ključne reči: multivarijaciona analiza, metod glavnih komponenata, varijabilitet, svojstvene vrednosti, zemljište, osobine zemljišta.

Received / *Primljen*: 20.04.2011.
Accepted / *Prihvaćen*: 30.05.2011.