Agronomy Research **17**(6), 2460–2473, 2019
https://doi.org/10.15159/AR.19.208

Application of geographically weighted principal components analysis based on soybean yield spatial variation for agro-ecological zoning of the territory

A. Zymaroieva[1,*], O. Zhukov[2], T. Fedonyuk[3] and A. Pinkin[4]

[1]Zhytomyr National Agroecological University, Faculty of Forestry, Department of Forest Resources Utilization, Stary Blvd. 7, UA10008 Zhytomyr, Ukraine
[2]Oles Honchar Dnipro National University, Faculty of Biology and Ecology, Department of Zoology and Ecology, pr. Gagarina, 72, UA49010 Dnipro, Ukraine
[3]Zhytomyr National Agroecological University, Faculty of Forestry, Department of Forest Ecology and Life Safety, Stary Blvd. 7, UA10008 Zhytomyr, Ukraine
[4]Zhytomyr National Agroecological University, Faculty of Engineering and Power Engineering, Department of Electrification, Automation of Production and Engineering Ecology, Stary Blvd. 7, UA10008 Zhytomyr, Ukraine
[*]Correspondence: nastya.zymaroeva@gmail.com

Abstract. In this study, the geographically weighted principal components analysis as an alternative method for agro-ecological characterization of the region was provided. The spatial and temporal distribution pattern of soybean yield was analyzed by using spatial statistics technology, which provided a good reference for agricultural development planning. The soybean yield was selected for the present study because it is a comprehensive indicator reflecting the production potential of the regional agroecosystems. The organized data set, which included the average per year yields of soybean in 10 regions (206 administrative districts) of Ukraine, was used for analysis. The regular temporal trend, specific for each district, was previously extracted from the time series data. The principal components analysis of the detrended data allowed to identify four principal components, which altogether can explain 58% of the soybean yield variation. The geographically weighted principal components analysis allowed to reveal that four spatially determined processes were influencing the yield of soybeans and had the oscillatory dynamics of different periodicity. It was hypothesized that the oscillating phenomena were of ecological nature. Geographically weighted principal component analysis revealed spatial units with similar oscillatory component of soybean yield variation. Our study confirmed the hypothesis that within the studied territory there are zones with the specific patterns of the temporal dynamics of soybean yield, which are uniform within each area but qualitatively different between zones. The territorial clusters within which the temporal dynamics of soybean yield is identical can be considered as agro-ecological zones for soybean cultivation.

Key words: cluster analysis, geographically weighted principal components analysis (GWPCA), soybean, spatial variability, productivity, yield.

INTRODUCTION

Sustainable agricultural development requires a systematic effort towards the planning of land use activities in the most appropriate way. Agro-ecological zoning is one of the cornerstones for agricultural planning because survival and failure of particular land use or farming system in a given region heavily relies on careful assessment of agro-climatic resources (Patel, 2003). A framework of agro-ecological zoning describing concepts, methods and procedures was conceptualized for the first time by FAO (1976). Agro-ecological zoning refers to the division of an area of land into land resource mapping units, having a unique combination of landform, soil and climatic characteristics and/or land cover. (FAO, 1996; Patel, 2003). Therefore, each zone has a similar combination, constraints and potential for the use of land, which serves as the focus of recommendations designed to improve the existing land use, either through increased production or by limiting land degradation (Suriadikusumah & Herdiansyah, 2014). The main objectives of agro-ecological zoning are data inventory of environmental resources, identification of homologous environments, determination of agricultural potential of a region, planning for regional development and identification of research priorities. Conventional methods employed are overlaying of maps and various statistical techniques (Aggarwal, 1991).

Principle components analysis (PCA) is a statistical method widely used in exploratory data analysis (Pearson, 1901). This non-parametric method reduces the dimension of a dataset, which simplifies structures hidden in the dataset (Liu et al., 2012). Principal components analysis has been applied by various researchers' area to explore and characterize the relationships between regionalized variables and related environmental factors, and to quantify the spatial variability pattern of these variables (Kumar et al., 2012). In an ecological setting, common applications of PCA are employed to environmental data sets e.g., the soils biogeochemistry data, species abundance data etc. (Legendre & Gallagher, 2001; Kaspari & Yanoviak, 2009).

Geographically weighted principal components analysis (GWPCA) is a localized version of PCA that is an exploratory tool for investigating spatial heterogeneity in the structure of multivariate data (Harris, 2011). Hence, a GWPCA investigates how outputs from a PCA vary spatially (Comber et al., 2016). Spatial changes in data dimensionality and multivariate structure can be explored via maps of the GWPCA outputs. GWPCA can also be used to detect multivariate spatial anomalies (Harris et al., 2015; Comber et al., 2016). In the published literature, GWPCA has been applied for analyzing multivariate population characteristics (Lloyd, 2010), social structure (Harris et al., 2011), soil characteristics (Kumar et al., 2012) and freshwater chemistry data (Harris et al., 2015, Li et al., 2015). However, GWPCA has not been applied to assess spatial variability of crop yields in agricultural landscapes, moreover, it has never been used for agro-ecological zoning of an area.

In this study, we consider the possibility of applying the geographically weighted principal components analysis as an alternative method for agro-ecological characterization of a region. The soybean yield was selected as the basis for agro-ecological zoning, because it is the comprehensive indicator, reflecting the production potential of agroecosystems (Kukal & Irmak, 2018). Crop yield is influenced by both management and environmental factors, but definite quantitative relationships are not easy to obtain because of complicated interactions between these factors (Ruiz-Vega, 1984). However, if the influence of agro-technological and management factors has the general origin and are described by the regression model, the influence of environmental factors leads to yield fluctuations (residuals) that do not fit into the total trend (Zhukov et al., 2018). These residuals also have a complex nature. There is a random noise associated with objective errors in the source data. However, in the regression residuals, we can expect a component that is associated with a regular variation that has an ecological nature (Kunah et al., 2018). Thus, the study of the residuals of the yield regression model allows us to separate the ecological determinants of soybean yield variation. Besides, through GWPCA it is possible to map areas with similar temporary fluctuations in yield, which may be regarded as agro-ecological zones for soybean cultivation.

The objective of this research was to study the spatial variation of the temporal patterns of the soybean yield. We have discussed two alternative hypotheses. The first one is the spatial variation of the soybean yield is per the uniform trend and there is no interruption of the continuous yield dynamics within the studied territory. The second one is that within studied territory there are zones with specific patterns of the temporal dynamics, which are uniform within each area but qualitatively different between zones.

MATERIALS AND METHODS

Time series of the soybean yields for each administrative district was divided into two components: total trend and trend residual. Total trend was determined by the dependence of the yield on time. As an analytical form of the trend, we selected the fourth-degree polynomial. The residuals of the corresponding regression models that describe the trends consist of the random component (noise) and, probably, the regular one that cannot be explained by the selected trend model. These two components are distinguished by their properties: the random component is an independent one for different points of space, and the regular component must be correlated to all or some points in space (administrative districts). We used the principal components analysis (PCA) for the residuals to isolate the regular component of trend models. The presence of the principal components, whose eigenvalues are more than 1, indicates that there exists a correlation in crop yields variation. Data on the yield of soybean were obtained from the State statistics service of Ukraine (http://www.ukrstat.gov.ua/) and its regional offices. Specifically, the organized data set included the average per year yields of the soybean for 10 regions of Ukraine (Vinnytsia, Volyn, Zhytomyr, Kyiv, L'viv, Rivne, Ternopil, Khmelnytsky, Cherkasy, Chernihiv), which include 206 administrative districts (Fig. 1). Information covers a period from 1991 to 2017.

Principal components analysis (PCA) is widely used for dimensionality reduction of the multivariate data set (Liu et al., 2012). Principal component analysis was performed using library stats (R Core Team, 2017). The suitability of yield data for the principal components analysis was evaluated by the Kaiser-Meyer-Olkin (KMO) test (Kaiser, 1974) with the help of the function KMOS from the library REdaS (Maier, 2015) in the environment for statistical computing R (R Core Team, 2017). Horn's (1965) technique for evaluating the components in a principal components analysis was implemented through *paran* function from the library 'paran' (Dinno, 2012). Geographically weighted principal components analysis (GWPCA) may be used to account for spatial heterogeneity in the structure of the multivariate data (Harris et al., 2011). An essential component of the GWPCA modelling is the spatial weighting function that quantifies the spatial relationship or spatial dependency between the observed variables (Fotheringham et al., 2002). A bandwidth for spatial analysis was found optimally using cross-validation with the Gaussian kernel function. Monte Carlo test was performed to examine whether yield data matrix eigenvalues were spatially varying (Iqbal et al., 2005). The GWPCA method is implemented using the GWmodel R package (Gollini et al., 2013). To visualize GWPCA outputs, the spatial distribution of the first four principal components percentage of the total variance was mapped. The locale influence of the variables on principal components 1–4 was visualized by mapping the 'winning variable' with the highest absolute loading. The spatial database was created in ArcGIS 10.0. The spatial autocorrelation, *I*-Moran's statistics (Moran, 1950), was calculated using Geoda095i (Anselin et al., 2005).

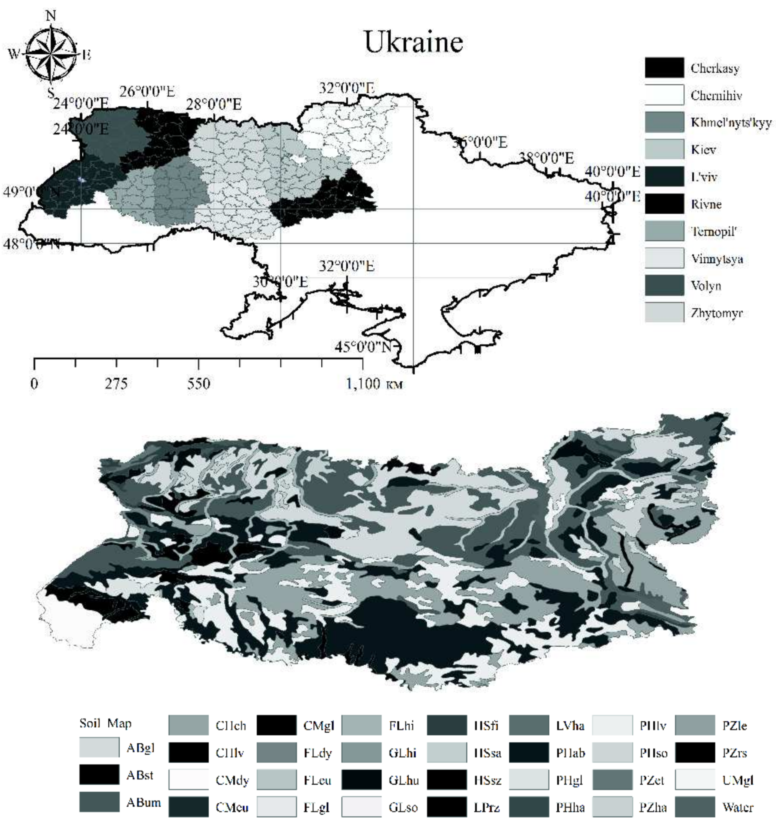

Figure 1. Map of 10 administrative regions in Ukraine, Ecoregions and soil map (Hengl et al., 2017).

Legend: Soil classification according World Reference Base for Soil Resources: ABgl – Albeluvisols Gleyic; ABst – Albeluvisols Stagnic; ABum – Albeluvisols Umbric; CHch – Chernozems Chernic; CHlv – Chernozems Luvic; CMdy – Cambisols Dystric; CMeu – Cambisols Eutric; CMgl – Cambisols Gleyic; FLdy – Fluvisols Dystric; FLeu – Fluvisols Eutric; FLgl – Gleyic Fluvisols; FLhi – Fluvisols Histic; GLhi – Gleysols Histic; GLhu – Gleysols Humic; GLso – Gleysols Sodic; HSfi – Histosols Fibric; HSsa – Histosols Sapric; HSsz – Histosols Salic; LPrz – Leptosols Rendzic; LVha – Haplic Luvisols; PHab – Phaeozems Albic; PHgl – Phaeozems Gleyic; PHha – Phaeozems Haplic; PHlv – Phaeozems Luvic; PHso – Phaeozems Sodic; PZet – Podzols Entic; PZha – Podzols Haplic; PZle – Leptic Podzols; PZrs – Podzols Rustic.

RESULTS AND DISCUSSION

The global principal components analysis

The dissimilar magnitude between regression residuals for administrative areas may lead to biased results from PCA as the variables with the highest sample variances tend to be emphasized in the first few principal components. Hence, all the selected variables need to be standardized by subtracting its mean from that variable and dividing it by its standard deviation. Such data standardization makes each transformed variable have equal importance in the subsequent analysis (Li et al., 2015).

As described before, the total number of 206 units was observed for 27 variables (years). The Kaiser-Meyer-Olkin (KMO) index was run for the overall data set to detect sampling adequacy. As the KMO value is 0.63, according to the Kaiser empirical rule (Kaiser, 1974), the study data should be considered relevant for the principal components analysis.

The PCA of the residuals of the regression model allowed to establish that the number of statistically probable principal components is 4 according to the Horn procedure (Horn, 1965). The four components with eigenvalues larger than 1 explain up to 58% of variation in the regional soybean yield (Table 1).

Table 1. Summary of global PCA

Principal components	Adjusted* eigenvalues	Unadjusted eigenvalues	Estimated bias	Proportion of variance	Standard deviation
1	8.30	9.04	0.73	33.47	3.00
2	2.45	3.08	0.62	11.39	1.75
3	1.33	1.86	0.54	6.90	1.36
4	1.21	1.67	0.46	6.20	1.29

Symbols: *– by Horn's parallel analysis.

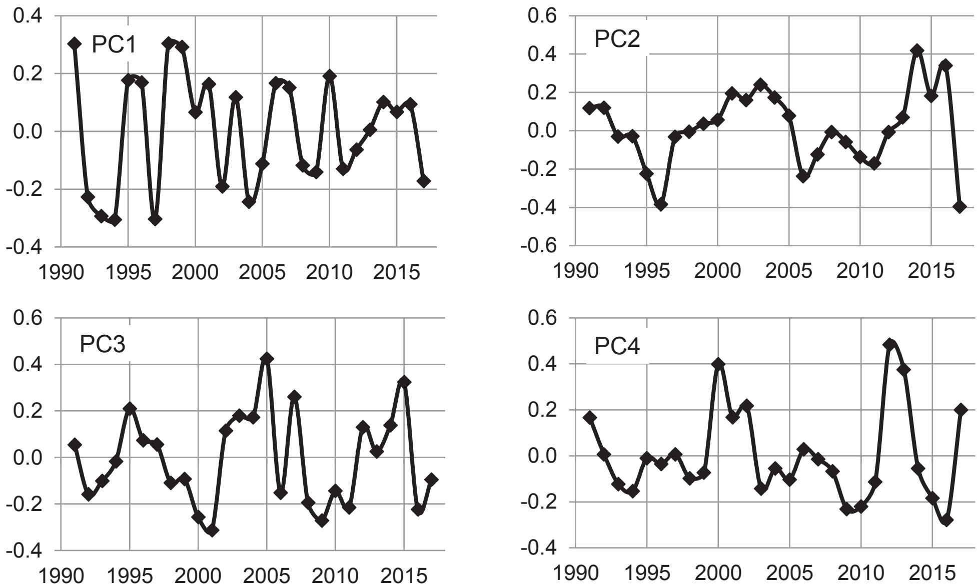

Figure 2. The principal components loadings to the variables.

The variables used in the PCA are the ordinal quantities – the years, so the loadings of the principal components on the variables can be represent ted as dynamic changes in time (Fig. 2). This form of presentation enables us to interpret meaningfully the determined principal components as oscillation processes with different frequencies. Thus, the principal component 1 explains 33.47% of the total variability of the soybean yield. It is characterized by a predominant oscillation process within 5 years. Moreover, this principal component demonstrates a clear trend towards damping of the oscillation process during the study period.

The variation of principal component 1 is spatially determined (*I*-Moran 0.29, $P = 0.001$). The zones with higher values of principal component 1 form clusters in some northern areas of the studied region, as well as in the western ones. The zone with the lower values of principal component 1 forms a cluster in the southeastern direction from the center of the region (Fig. 3).

Principal component 2 explains 11.39% of the total dispersion and as to its fluctuations, most typical is an oscillating process with a lag of ten years. This component demonstrates spatially regular patterns of variation (*I*-Moran 0.48, $P = 0.001$). Clusters with higher values of principal component 2 are located in the south-western and north-eastern regions, and with the lower ones – in the north-west and south-east (Fig. 3).

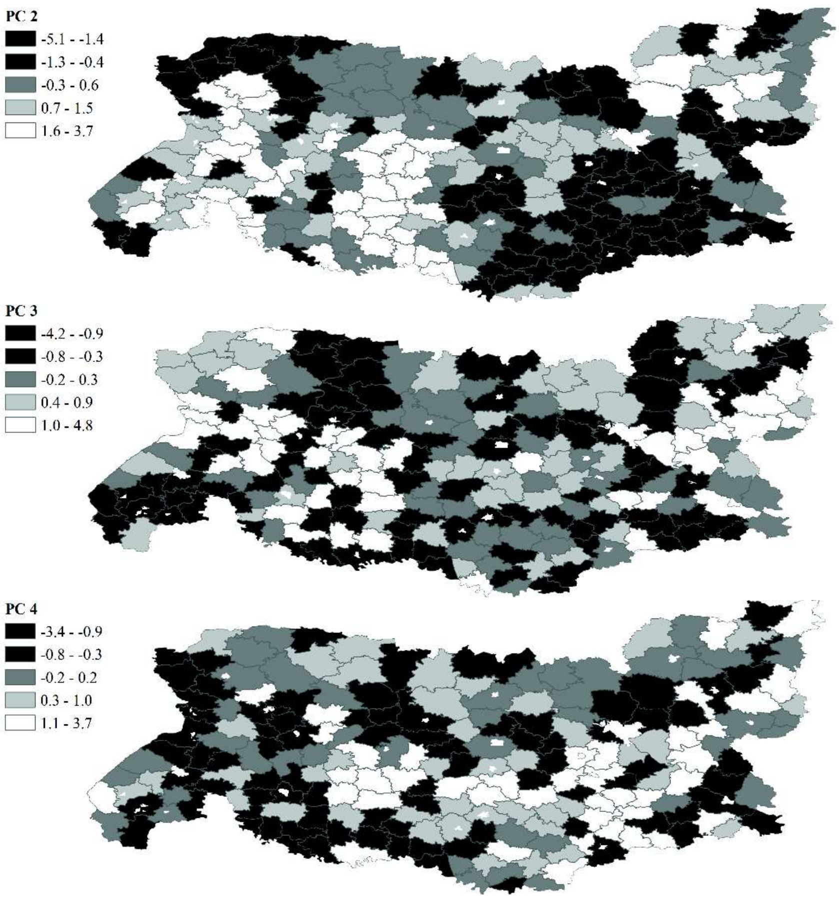

Figure 3. Spatial variation of principal components 1–4.

Principal component 3 explains 6.90% of the soybean yield variability and its characteristic oscillations are repeated every 9-10 years. The high spatial level of the principal component 3 variation is confirmed by *I*-Moran statistics (0.28, $P = 0.001$). Clusters with high values of principal component 3 are typical of the southwest and east, and with low values – of the southeast.

Principal component 4 explains 6.20% of the dispersion of the soybean yield. For its fluctuations in time, the most characteristic period is also a span of 8-9 years (Fig. 2). The spatial patterns of this component are statistically significant (*I*-Moran 0.29, $P = 0.001$). The clusters with the higher values of principal component 4 are characteristic for the center and east of the region, and with lower values – for the west (Fig. 3).

Thus, the global principal components analysis revealed the presence of dynamic processes of soybean yield, which have the oscillatory nature with varying frequencies. We associate oscillatory processes of varying frequency with causes of different nature.

The principal components analysis of the regression model residues of the time trend enables us to prove that within a set of ecological factors four principal components affect the soybean yields to the greatest extent. Specification and detailed research of the origin of these principal components are objective of our subsequent studies. However, at this point we can prove the presence of four spatially determined processes that influence the yield of soybeans and have the oscillatory dynamics of different periodicity.

Geographically weighted principal components analysis

The Monte Carlo test was conducted to examine whether the data matrix eigenvalues are spatially varying ($P = 0.01$). Thus, there is a high degree of spatial non-stationarity present in the data of regional soybean yield.

The previous global PCA results indicate that the first four components can collectively explain 58% of the variance in data structure. Accordingly, it is reasonable to retain the four components for further GWPCA analysis. However, since the paper is limited in scope, only the first two components GWPC 1 and GWPC 2 from GWPCA will be comparatively interpreted in detail.

The results of the procedure GWPCA can be visualized and interpreted by focusing on how the dimensions of the data vary spatially and how the original variables affect the principal components (Li et al., 2015). Percentage of spatial variation of the total variation demonstrates a clearly expressed variability, thus forming spatially homogeneous clusters from north to south of the research region (Fig. 4). Compared with the global analysis of the principal component, GWPCA demonstrates its effectiveness and efficiency in the analysis of spatial patterns of regional placement of soybean yields, using the mapping of spatial variability of the principal components.

It was suggested that the variables with the highest loading values and their impact intensity values can be mapped locally (Lloyd, 2010). Then we can visualize how each of the four variables locally affects the given component, displaying the 'winning variable' with the highest absolute loading. Fig. 5 shows the spatial distribution of variables with the highest absolute loading of GWPC 1–4, respectively.

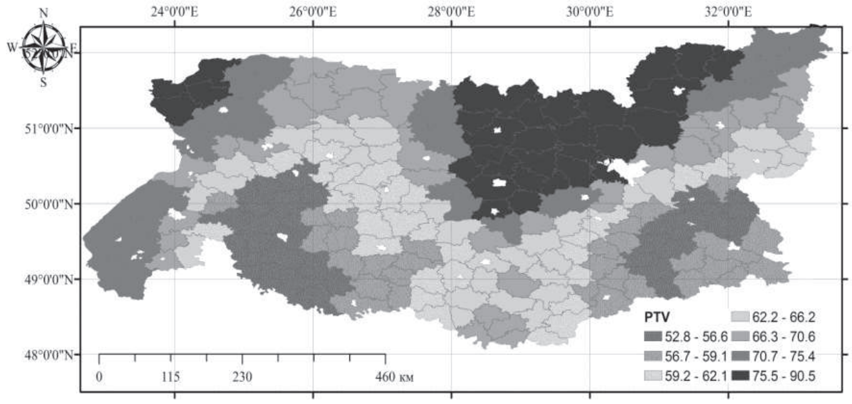

Figure 4. Spatial variability of the percentage of total variance (PTV) of the first four principal components.

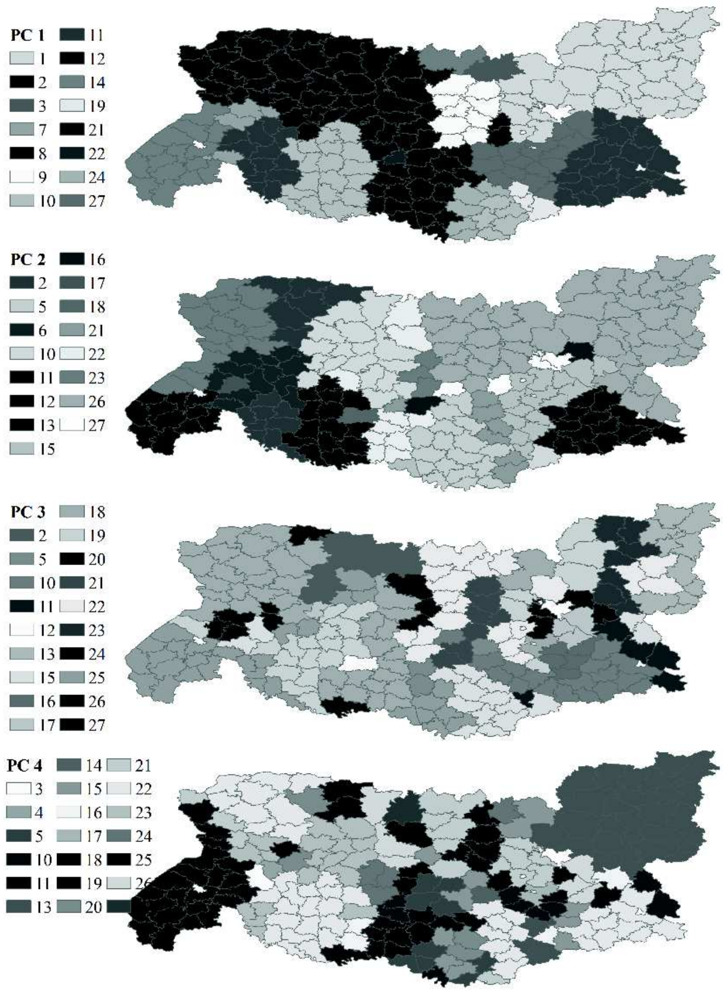

Figure 5. Spatial location of 'winning' variables for principal components 1–4.

The traditional representation of 'winning' variables for the principal components cannot fully reveal the nature of the spatially dependent relationship between the indicators estimated by the principal component analysis. The factor of loading predominance is one of the aspects that reflects the crop yields dynamics. Due to the oscillating nature of such dynamics, predominance is the random outlier of the indicator at a certain moment in comparison with the general recurring dynamics. Therefore, for each of the statistically significant principal components, we conducted the classification of administrative districts by cluster analysis based on distance, which is opposite to the Pearson correlation coefficient. This indicator of distance is sensitive to the form of comparable indicators, and not to their absolute values. This approach allows to identify groups of administrative districts with a similar time dynamic of soybean yields in the aspect of the corresponding principal component. It can be assumed that the aggregates of administrative districts with a similar yield's dynamics are also geographically close and form homogeneous ecological regions.

Cluster analysis of the administrative districts by factor loading values of GWPC 1 revealed three homogeneous clusters (Fig. 6). For each cluster, we calculated the average values of the factor loadings, which helped assess the specificity of the respective clusters (Fig. 7). The general trend of principal component 1 is the damping of the amplitude of the oscillations during the research period and the predominance of higher frequency components of oscillatory dynamics corresponding to the heterogeneity of observations over time or heteroscedasticity. So, the Koenker-Bassett test for clusters 1 and 3 indicates the heteroscedasticity of the time dynamics of factor loadings (1.17, $P = 0.28$ and 1.35, $P = 0.24$, respectively). The heteroscedasticity is established for cluster 2 (5.09, $P = 0.024$). Thus, the qualitative feature of the soybean yield dynamics in the corresponding clusters is the difference in levels of damping of the principal component 1 oscillations over time.

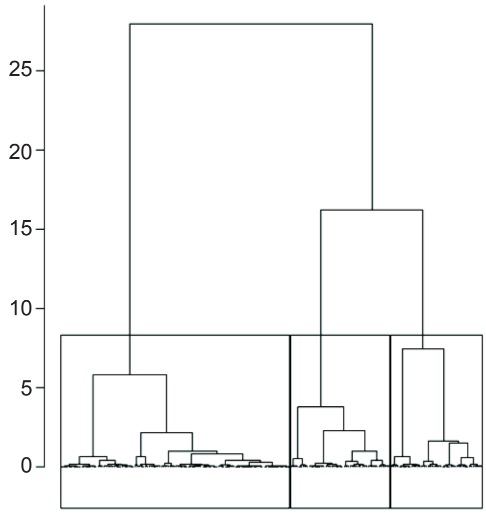

Figure 6. Cluster analysis of administrative districts by factor loadings values GWPC 1.

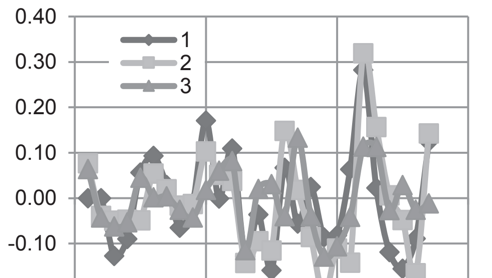

Figure 7. The average values of factor loadings of GWPC 1 for clusters 1–3. Here abscissa is the primary variables (the residuals of the regression models of the trend of yield by years), axis ordinate - factor loadings.

The spatial distribution of administrative districts included in the respective clusters is spatially regular (Fig. 8). Cluster 3 covers the largest part of Ukraine and is located in the north, center and west of the studied territory. Clusters 1 and 2 are located in the south of the research area.

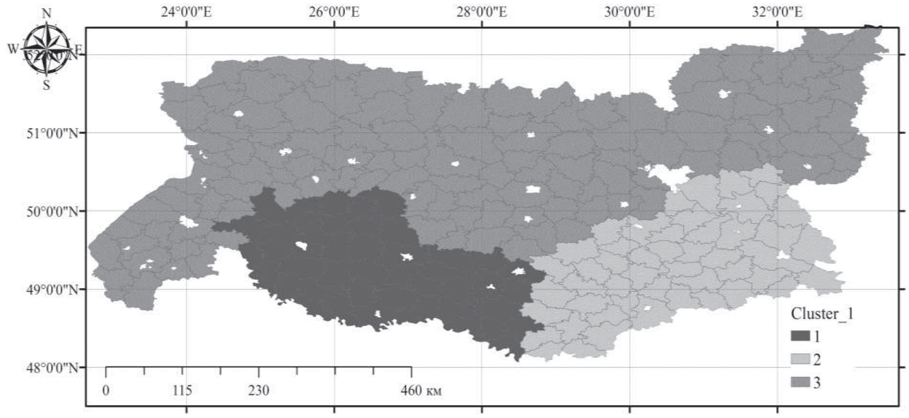

Figure 8. Spatial location of the clusters obtained based on the GWPC 1 loadings.

Cluster analysis of the values of factor loadings GWPC 2 revealed four homogeneous clusters (Fig. 9).

For each cluster, we calculated the average values of factor loadings, which helped assess the specificity of the respective clusters (Fig. 10). For clusters 1 and 3 attenuation during the studied period is characteristic, while for clusters 2 and 4 a fading amplitude was observed in the middle of the research period. In the spatial aspect, cluster 4 occupies the west of the research area. Clusters 1, 2 and 3 are disruptive, so cluster 1 is mainly located in the center, cluster 2 – in the east, and cluster 3 – in the southwest of the research region (Fig. 11).

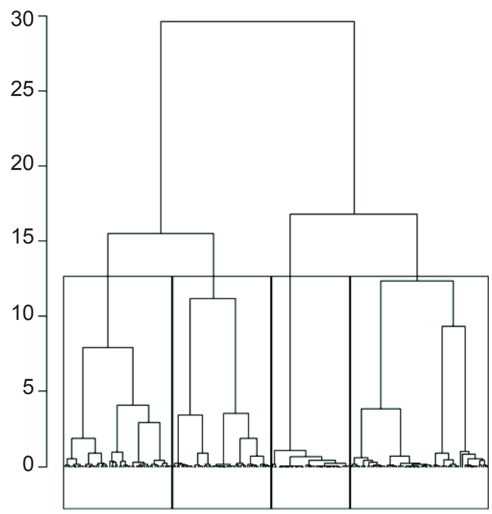

Figure 9. Cluster analysis of administrative districts by factor loadings values GWPC 2.

Principal component 1 (PCA 1) explains the largest part of soybean yield variability (33.5%). It is characterized by oscillatory dynamics with a period of 5 years and has the nature of an irregular component. Principal component 2 (PCA 2) has the amplitude of oscillation of 8–10 years. The principal components are spatially heterogeneous and divide the territory of Ukraine into 4 zones, which are characterized by different sensitivity of soybean yield to environmental factors. Such territorial clusters can be defined as agro-ecological zones, since an agro-ecological zone is an area with a similar course of ecological processes (Sivakumar & Valentin, 1997). Cosequenly, agro-ecological zoning refers to the division of an area (of land) into smaller units, which have similar characteristics related to land suitability potential production and environmental impact (Patel, 2003). The crop yield is a functional indicator of complex relations between plants and their environment (Anderson et al., 2013). Therefore, applying the yield as a basic indicator for agro-ecological zoning is entirely justified.

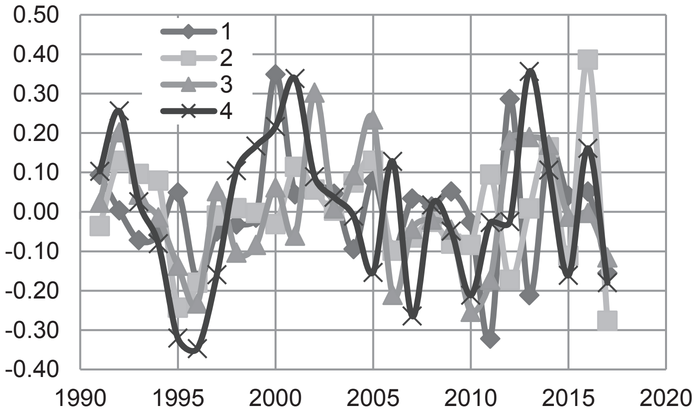

Figure 10. The average values of factor loadings of GWPC 2 for clusters 1–4.

Application of the principal components analysis of the yield dynamics is based on the assumptions that the origin of the relationships within the entire investigated area is homogeneous. Geographically weighted principal components analysis allows us to investigate local patterns of soybean yield dynamics (Patel, 2003). Local models have greater explanatory power than the total model, which is quite natural because the consideration of local specifics allows the more objective reflection of reality (Kumar et al., 2012). Nevertheless, the application of this approach causes certain methodological difficulties for meaningful interpretation. The most common technique of mapping 'winning' variables is not suitable in the case of time series analysis.

Consequently, based on the approximate types of local cycles, clusters were established for each principal component, and instead of displaying 'winning' variables, we applied the mapping of the established clusters. However, this approach has some advantages. Firstly, the ecologically homogeneous zones obtained by our approach (Figs 8, 11) are more compact than the ones that are established using 'winning variables' (Fig. 5). This result was obtained because in the clusters formation, the dominant role is played by the factors of a regular nature, and the random factors are filtered out during the analysis procedure (Zhukov et al., 2018). In fact, 'winning variables' are the result of a predominantly random choice from some lists of important information variables. Therefore, both approaches give a similar picture in general, but the proposed algorithm is less sensitive to random factors. Secondly, the proposed algorithm provides an opportunity to give a meaningful interpretation of the obtained clusters by studying the dynamics characteristics of each cluster in time. In the 'winning variables' approach, the variable itself is a marker of the corresponding spatially homogeneous territory (Kunah et al., 2018). Nevertheless, such an instrument is acceptable when qualitatively diverse variables are used, each of them can be measured in the next period, and thus applied to forecast the phenomenon under study. Among the time series variables, there are no 'more important' or 'less important' years. Besides, all of these variables are in retrospective and could not be re-measured. The patterns based on the cyclic frequency of processes are applied for forecasting. Such features can be set for the selected clusters. Results of the present work reveal that the GWPCA can be used for agro-ecological zoning.

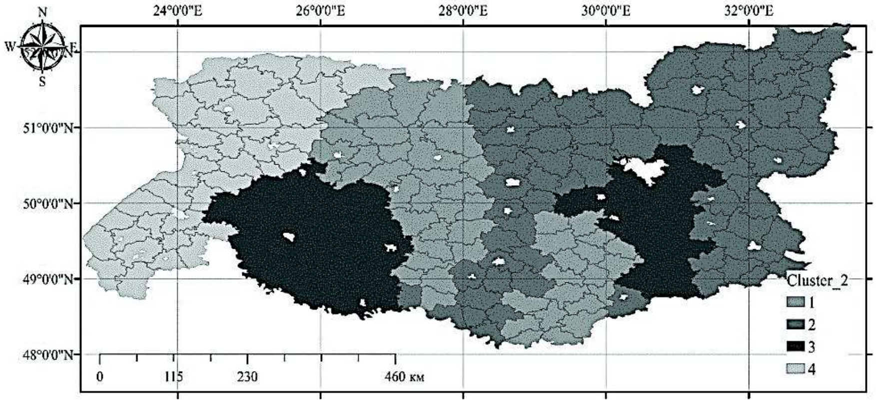

Figure 11. Spatial location of the clusters obtained based on the GWPC 2 loadings.

Consequently, agro-ecological zoning was performed with regard to the uniformity of dynamics of an agricultural area production potential. This approach is fundamentally different from that of zoning based on the total yield of crops (Lazarenko, 1995). A classification based on yields is justified for systems that are in a state close to the steady-state. According to the global climate changes and transformation of the environmental regimes, this approach is unacceptable. The agro-ecological zones proposed in the given research did not differ in the overall level of productivity of soybean during the study period. Features of these zones lie in the values of the principal components and reflect the nature of the relationship between different spatial units. Spatial distribution of the principal components indicates a continual pattern, but their overlapping allowed to determine spatially discrete units, which we identified as agroecological zones. Each zone is characterized by a certain character and dynamics of production capacity and has an invariant pattern of response to varying climatic, environmental, and agroeconomic factors.

CONCLUSIONS

Our study confirmed the hypothesis that within the studied territory there are zones with specific patterns of the temporal dynamics of soybean yield, which are uniform within each area but qualitatively different between zones. The principal components analysis of the regression models' residues of the time trend enabled us to establish 4 principal components, which together explain up to 58% of the variation in the regional soybean yield. Four spatially determined processes influence the yield of soybeans and have the oscillatory dynamics of different periodicity. The oscillating phenomena are of an ecological nature. Geographically weighted principal component analysis revealed spatial units with similar oscillatory component of soybean yield variation. The territorial clusters within which the temporal dynamics of soybean yield is identical can be considered as agro-ecological zones for soybean cultivation. Further study of the nature of the principal components will be the objective of our subsequent studies, as well as the impact of the climate change on the crop yield variability.

REFERENCES

Aggarwal, P.K. 1991. Agro-ecological zoning using crop growth simulation models: characterization of wheat environments of India. In: *Proceedings of the International Symposium on Systems Approaches for Agricultural Development*, Bangkok, Thailand, pp. 2–6.

Anderson, M.C., Cammalleri, C., Hain, C.R., Otkin, J., Zhan, X.W. & Kustas, W. 2013. Using a diagnostic soil-plant-atmosphere model for monitoring drought at field to continental scales. In: *Four decades of progress in monitoring and modeling of processes in the soil-plant-atmosphere system: applications and challenges*, 19, pp. 46–57.

Anselin, L., Ibnu, S. & Youngihn, Kh. 2005. GeoDa: An Introduction to Spatial Data Analysis. *Geographical Analysis* **38** (1), 5–22. doi: 10.1111/j.0016-7363.2005.00671.x

Comber, A.J, Harris, P. & Tsutsumida, N. 2016. Improving land cover classification using input variables derived from a geographically weighted principal components analysis. *ISPRS Journal of Photogrammetry and Remote Sensing* **119**, 347–360. doi: 10.1016/j.isprsjprs.2016.06.014.

Dinno, A. 2012. *paran: Horn's Test of Principal Components/Factors*. R package version 1.5.1. Available at: https://CRAN.R-project.org/package=paran

Gollini, I., Lu, B., Charlton, M., Brunsdon, Ch. & Harris, P. 2013. GWmodel: An R Package for Exploring Spatial Heterogeneity Using Geographically Weighted Models. *Journal of Statistical Software* **63**(17), 1–52.

FAO. 1976. A framework for land evaluation. Food and Agricultural Organisation, Soils Bulletin **32**, Rome, Italy.

FAO. 1996. Guidelines: Agro-ecological zoning. Food and Agricultural Organisation, Soils Bulletin, Rome, Italy.

Fotheringham, A.S., Brunsdon, C. & Charlton, M. 2002. *Geographically Weighted Regression: the Analysis of Spatially Varying Relationships*. John Wiley & Sons, Chichester, 284 pp.

Harris, P., Brunsdon, C. & Charlton, M. 2011. Geographically Weighted Principal Components Analysis. *International Journal of Geographical Information Science* **25**(10), 1717–1736. doi: https://doi.org/10.1080/13658816.2011.554838

Harris, P., Clarke, A, Juggins, S, Brunsdon, C. & Charlton, M. 2015. Enhancements to a Geographically Weighted Principal Component Analysis in the Context of an Application to an Environmental Data Set. *Geographical analysis* **47**(2)*,* 146–172. doi: 10.1111/gean.12048.

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M. & Blagotić, A. 2017. SoilGrids250m: Global gridded soil information based on machine learning. *PLoS ONE* **12**(2), e0169748. doi: https://doi.org/10.1371/journal.pone.0169748

Horn, J.L. 1965. A rationale and a test for the number of factors in factor analysis. *Psychometrika* **30**, 179–185. doi: 10.1007/BF02289447

Iqbal, J., Thomasson, J.A., Jenkins, J.N., Owens, P.R. & Whisler, F.D. 2005. Spatial variability analysis of soil physical properties of alluvial soils. *Soil Science Society America journal* **69**(4), 1338–1350. doi: 10.2136/sssaj2004.0154

Kaiser, H.F. 1974. An Index of Factorial Simplicity. *Psychometrika* **39**(1), 31–36.

Kaspari, M. & Yanoviak, S. 2009. Biogeochemistry and the Structure of Tropical Brown Food Webs. *Ecology* **90**, 3342–3351.

Kukal, M.S. & Irmak, S. 2018. Climate-Driven Crop Yield and Yield Variability and Climate Change Impacts on the U.S. Great Plains Agricultural Production. *Scientific Reports* **8**, 3450. doi:10.1038/s41598-018-21848-2

Kumar, S., Lal, R. & Lloyd, C.D. 2012. Assessing spatial variability in soil characteristics with geographically weighted principal components analysis. *Computational Geosciences* **16**(3), 827–835. doi 10.1007/s10596-012-9290-6

Kunah, O., Pakhomov, O., Zymaroieva, A., Demchuk, N., Skupskyi, R., Bezuhla, L. & Vladyka, Y. 2018. Agroeconomical and agroecological aspects of the rye (*Secale cereale* L.) yelds spatial variation within Polesia and Foreststeppe zones of Ukraine: the useage of the geographically weighted principal components analysis. *Biosystems Diversity* **26**(4), 276–285. doi: https://doi.org/10.15421/011842

Lazarenko, P.I. 1995. Ecological and biological bases of agricultural zoning areas (Dnipropetrovsk region as an example). Kyiv, 476 pp. (in Ukrainian).

Legendre, P. & Gallagher, E. 2001. Ecological Meaningful Transformations for Ordination of Species Data. *Oecologia* **129,** 271–280.

Li, Z., Cheng, J. & Wu, Q. 2015. Analyzing regional economic development patterns in a fast developing province of China through geographically weighted principal components analysis. *Letters in Spatial and Resource Sciences* **9**(3), 233–245.

Liu, J., Pattey E. & Jégo, G. 2012. Assessment of Vegetation Indices for Regional Crop Green LAI Estimation from Landsat Images over Multiple Growing Seasons. *Remote Sensing of Environment* **123**, 347–358. doi:10.1016/j.rse.2012.04.002

Lloyd, C.D. 2010. Analysing population characteristics using geographically weighted principal components analysis: a case study of Northern Ireland in 2001. *Comput. Environ. Urban.* **34**(5), 389–399.

Maier, M.J. 2015. Companion Package to the Book ``R: Einführung durchangewandte Statistik"_. R package version 0.9.3, URL: http://CRAN.R-project.org/package=REdaS

Moran, P.A.P. 1950. Notes on continuous stochastic phenomena. *Biometrika.* **37**(1/2), 17–23. doi: 10.2307/2332142

Patel, N.R. 2003. Remote sensing and GIS application in agro-ecological zoning. Satellite Remote Sensing and GIS Applications in Agricultural Meteorology. In: *Proceedings of a Training Workshop*, (7–11 July), Dehra Dun, India, pp. 213–233.

Pearson, K. 1901. On lines and planes of closest fit to systems of points in space. *Phil. Mag.* **2**(7–12), 559–572.

R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: https://www.R-project.org/.

Ruiz-Vega, J. 1984. Soybean phenology and yield as influenced by environmental and management factors. *Retrospective Theses and Dissertations*, 8213. Available at: https://lib.dr.iastate.edu/rtd/8213

Sivakumar, M.V.K. & Valentin, C. 1997. Agroecological zones and the assessment of crop production potential. *Phil. Trans. R. Soc. Lond*. **352**(1356), 907–916.

Suriadikusumah, A. & Herdiansyah, D.G. 2014. Study on land resources based on agro-ecological zones in Bandung district, West Java -Indonesia. *International Journal of Applied Science and Technology* **4**(4), 212–220.

Zhukov, O.V., Pelina, T.O., Demchuk, O.M., Demchuk, N.I. & Koberniuk, S.O. 2018. Agroecological and agroeconomic aspects of the grain and grain legumes (pulses) yield dynamic within the Dnipropetrovsk region (period 1966–2016). *Biosystems Diversity* **26**(2), 3–10.