Analyzing cereal and grain legumes (pulses) yields patterns in the forest and forest-steppe zones of Ukraine using geographically weighted principal components analysis

Anastasiia ZYMAROIEVA [1, 2], Oleksandr ZHUKOV [3]

Analyzing cereal and grain legumes (pulses) yields patterns in the forest and forest-steppe zones of Ukraine using geographically weighted principal components analysis

Abstract: This paper aims to explore spatial heterogeneity present in the crop yields data collected from 170 administrative districts in the forest and forest-steppe zones of Ukraine for 27 years using the PCA and GWPCA methods. As a result of the principal component analysis of cereal and grain legumes (pulses) yields variability seven principal components were determined which together explain 66.8 % of the overall yields variability. The global PCA revealed the presence of dynamic processes of the cereal and grain legumes yields variation which have the oscillatory nature with different frequencies. We associate oscillatory processes of the varying frequency with causes of a different nature. The oscillating processes with a period of ten years or more may be of climatic origin. The oscillatory process with the longest period (13 years) is characteristic for the principal component 1, which explains the largest part of cereal and grain legumes yields variability (22.6 %). It is possible to assume that among agroecological factors climate change mostly affects crop productivity. The cluster analysis of administrative districts was conducted based on the cereal and leguminous yield dynamics. The clusters are geographically defined administrative districts that together forming spatially connected areas, which we identified as agroecological zones.

Key words: yield; cereals; leguminous crops; spatial and temporal variability; geographically weighted principal components analysis

Analiza vzorcev pridelkov žit in zrnatih stročnic na območju gozda in lesostepe Ukrajine z geografsko tehtano analizo glavnih komponent

Izvleček: Namen prispevka je bil preučiti prostorsko heterogenost pridelkov poljščin iz podatkov zbranih iz 170 administrativnih okrožij na območju gozda in lesostepe Ukrajine v obdobju zadnjih 27 let z uporabo PCA in GWPCA metod. Rezultat analize spremenljivosti pridelkov žit in zrnatih stročnic z analizo glavnih komponent je bila določitev sedmih glavnih component, ki so skupno razložile 66,8 % celokupne variabilnosti pridelkov. Globalna analiza glavnih component je odkrila prisotnost dinamičnih procesov v spremenljivosti pridelkov žit in zrnatih stročnic, ki nihajo z različnimi frekvencami. Oscilatorne procese z različnimi frekvencami povezujemo z različnimi vzroki. Nihajoči procesi s periodo desetih ali več let so lahko povezanimi s podnebjem. Oscilatorni proces z najdalšo period (13 let) je značilen za prvo glavno komponento, ki razloži največji delež nihanja pridelkov žit in zrnatih stročnic (22,6 %). Mogoče je zaključiti, da med agroekološkimi dejavniki sprememba podnebja najbolj vpliva na pridelek poljščin. Klasterska analiza administrativnih območij je bila izvedena na osnovi dinamike spremeljivosti pridelkov žit in zrnatih stročnic. Grozdi so zemljepisno omejena administrativna območja, ki tvorijo skupaj prostorsko povezana območja, ki so označena kot agroekološke cone.

Ključne besede: pridelek; žita; zrnate stročnice; prostorska in časovna spremenljivost; analiza geografsko tehtanih glavnih komponent

---

1   Polissia National University, Faculty of Foresty and Ecology, Department of forest resources utilization, Zhytomyr, Ukraine
2   Corresponding author, e-mail: nastya.zymaroeva@gmail.com
3   Bogdan Khmelnitsky Melitopol State Pedagogical University, Faculty of Chemistry and Biology, Botany and Horticulture Department, Melitopol, Ukraine

1   INTRODUCTION

Agricultural producers have known for many years that temporal and spatial variations in crop yields are a reality of farming. Year-to-year fluctuations in crop performance can be caused by seasonal differences in factors such as growing season conditions, differences in weed, insect and disease pressures and possibly the appropriateness of management decisions (Lauzon et al., 2005; Frieler et al., 2017).

Crop yields are frequently heterogeneous across space and time. Spatiotemporal variation can be broken down into its spatial and temporal components (Hammond & Kolasa, 2014). Synchrony and persistence are important components of spatiotemporal variability. When the same crop rises or declines in the same year in each of the two regions of the country, they are in synchrony. Persistence on the other hand refers to consistent differences in mean yield between two regions or other spatial units. Spatial patterns are diagnostic when they are used to uncover hidden mechanisms in the landscape, and predictive when they indicate the likely future behavior of a process (Hammond & Kolasa, 2014; Kong et al., 2018).

There have been several approaches to explore long-term trends in historical crop yield data based on process-based or statistical models (Lobell et al., 2013; Frieler et al., 2017). In this study, we use a collection of local (non-stationary) statistical models, termed geographically weighted principal components analysis (GWPCA).

GWPCA is a localized version of PCA that is an exploratory tool for investigating spatial heterogeneity in the structure of multivariate data. It not only provides a useful investigative tool, but also lends itself to many uses of PCA at a localised scale (Harris et al., 2011).

Principle component analysis (PCA) is a statistical method widely used in exploratory data analysis (Pearson, 1901). This non-parametric method compresses the dimension of a dataset and thus can reveal some simplified structures hidden in the dataset (Liu et al., 2012). Principal component analysis has been applied by various research area with the aim of exploring and characterizing the relationships between regionalized variables and related environmental factors, and to quantify the spatial variability pattern of these variables (Kumar et al., 2012; Li & Huang, 2008). In an ecological setting, common applications of PCA are to environmental data sets e.g., the soils biogeochemistry data (Kaspari & Yanoviak, 2009), although via a suitable transform, PCA can also be applied to species abundance data (Legendre & Gallagher, 2001; Harris et al., 2015).

PCA analysis can be easily expanded using geographically weighted correlation coefficients where the weights are determined with respect to each observation location allowing a separate PCA (geographically weighted PCA, GWPCA) to be conducted for each sampling location (Kumar et al., 2012). 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 (Fotheringham et al., 2002). GWPCA can also be used to detect multivariate spatial anomalies (Harris et al., 2015).

In the published literature, GWPCA has been extensively 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 the spatial variability of crop yield in agricultural landscapes inherently with spatially heterogeneous. To fill in the gap, this paper aims to explore such spatial heterogeneity present in the crop yields data collected from 170 administrative districts in forest and forest-prairie zone of Ukraine for 27 years using the GWPCA method.

The maps produced from GWPCA provide quantitative evidence and spatial details for supporting spatial land management and regional development strategy and help identify the spatial differentiation status of the regional agricultural development.

2   MATERIALS AND METHODS

Crop data were obtained from the State Statistics Service of Ukraine. Specifically, the organized data set included the average per year yields of the cereal and grain legumes (pulses) for 7 regions of Ukraine, which include 170 administrative districts over 27 years (1991– 2017). The cereal crops includes wheat (winter and spring), rye (winter and spring), barley (winter and spring), spring oats, buckwheat and millet. Grain legumes are beans and peas. The State Statistics Service of Ukraine provides information on the yield of cereals and grain legumes in one category. We considered the yield of cereal and grain legumes crops as a marker of the productive potential of the agrolandscape.

The time series of crop yields for each administrative district was divided into two components: total trend and trend residual. The total trend was explained by the dependence of the yield from time. As an analytic form of the trend we chose the fourth-degree polynomial (Zymaroieva et al., 2019b; Zymaroieva et al., 2020a). 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

Table 1: Summary of global PCA

Principal components	Adjusted eigenvalues*	Unadjusted eigenvalues	Estimated bias	Proportion of variance	Standard deviation
1	5.28	6.10	0.82	22.60	2.46
2	2.46	3.15	0.69	11.68	1.77
3	2.07	2.67	0.59	9.88	1.63
4	1.23	1.74	0.51	6.45	1.32
5	1.14	1.58	0.44	5.84	1.25
6	1.06	1.43	0.37	5.31	1.19
7	1.04	1.35	0.31	4.99	1.16

Legend: * – by Horn's parallel analysis

is independent 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 for the residuals in order to isolate the regular component of trend models. The presence of the principal components, whose eigenvalues more than one unit, indicates the existence of correlation in crop yields variation (Zymaroieva et al., 2019a).

How data is suited for the principal component analysis was estimated by Kaiser-Meyer-Olkin (KMO) test (Kaiser, 1974). Calculations were performed using library REdaS (Hatzinger et al., 2014) in the environment for statistical computing R (R Core Team, 2017). PCA was performed using library stats (R Core Team, 2017). The GWPCA method is implemented using the GWmodel R package (http://cran.rstudio.com/) (Zhukov & Ponomarenko, 2018). The spatial database was created in ArcGIS 10.0. The spatial autocorrelation, *I*-Moran's statistics (Moran, 1950), was used to calculate the global coefficient. *I*-Moran's is a measure of autocorrelation similar to the Pearson's correlation statistics, and both statistics range from +1.0 meaning strong positive spatial autocorrelation, to 0 meaning a random pattern, to –1.0 indicating strong negative spatial autocorrelation (Iqbal et al., 2005). Heteroscedastic testing becomes very challenging for high-dimensional regressions. Heteroscedasticity implies that the variance of the disturbance term is not constant over the data range. Koenker-Bassett test evaluates heteroscedasticity by comparing different quantile or expectile estimates (Koenker & Bassett, 1978). The global Moran's statistics were calculated using Geoda095i (http://www.geoda.uiuc.edu/) (Anselin et al., 2005).

3   RESULTS

The residuals of the corresponding regression models have a complex nature. Definitely, there is a random noise associated with objective errors in the source data. Also, in the regression residuals, we can expect a component that is associated with a regular variation that may have an ecological nature (Zymaroieva et al., 2019a).

The dissimilar magnitude between these 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).

There is another question to be answered before implementing a PCA analysis: is the sample size large enough for the statistical analysis? Is there a certain redundancy between the variables? We studied crop yields in 170 administrative districts for 27 variables (years). The Kaiser-Meyer-Olkin (KMO) index is run for the overall data set to detect sampling adequacy. As the KMO value is 0.65, according to the Kaiser empirical rule (Kaiser, 1974), the study data should be considered mediocre for the PCA.

The PCA of the residuals of the regression model allowed establishing that the number of statistically probable principal components is 7 according to the Horn procedure (Horn, 1965). The first seven components with eigenvalues larger than unity totally explain up to 66.8 % of variation in the regional cereal yield.

The variables used in the PCA are the ordinal quantities – the years, so the loadings of the principal components on the variables can be represented as dynamic changes in time (Fig. 1). This form of presentation allows interpreting meaningfully the installed principal components as oscillation processes with the different frequency. Thus, principal component 1 explains 22,60 % of the total variability of the grain and grain legumes yields. It is characterized by a predominant oscillation process within the period of 13 years. The variation of

the principal component 1 is clearly spatially determined (*I*-Morana 0.29, *p* = 0.001). The zones with higher values of the principal component 1 form clusters in some areas of the east of the studied region, as well as in the south, southwest and west. The zone with the lower values of the principal component 1 forms a clear cluster in the southwestern direction from the center of the region (Fig. 2).

Principal component 2 explains 11.68 % of the variability of the space of signs and its fluctuation has characteristic period of 6-7 years. This component demonstrates spatially regular patterns of variation (*I*-Moran 0.48, *p* = 0.001). Clusters with higher values of the principal component 2 are located on the north and northeast, and with the lower ones – the east and south (Fig. 2).

Principal component 3 explains 9.88 % of the total variability of the cereal and grain legumes yields and it has characteristic fluctuations with a period of 8-9 years. This component has a high level of spatial variability (*I*-Moran 0.51, *p* = 0.001). Clusters with higher values of the principal component 3 are common for the northwest and southeast, and with lower values for the east and southwest (Fig.2).

Principal component 4 explains 6.45 % of the variability of the space of signs. For its fluctuations in time, the period of 8-9 years is also the most characteristic (phase shift between components 3 and 4 is 4 years) (Fig. 1). The spatial patterns of this component are statistically significant (*I*-Morana 0.29, *p* = 0.001). The cluster with the higher values of the principal component 4 is characteristic for the center and north of the region, and with lowered values – for the east (Fig.2).

Principal component 5 describes 5.84 % of the variability of the feature space and is characterized by fluctuations with a period of 4 years. The *I*-Moran's index value for the PC5 is 0.39 (p = 0.001), which reveals a statistically positive spatial autocorrelation and as such demonstrates a highly clustering spatial pattern. Clusters with higher values of the principal component 5 are concentrated in the west, and with the lower ones in the central part of the region (Fig.2).

Principal component 6 explains 5.31 % of variability and represents the most high-frequency component of the grain yield dynamics with the most typical fluctuation period of 3 years. The variability of this component is characterized by the spatial component (*I*-Moran 0.19, *p* = 0.001). The zone with the higher scores of the main component 6 forms a clearly defined cluster in the south of the region (Fig. 2).
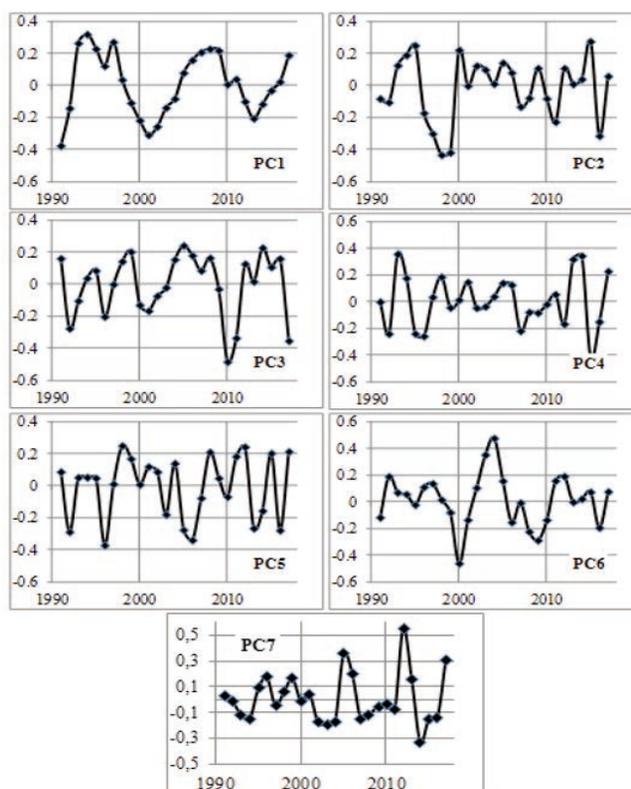

Figure 1: The principal components loadings to the variables 1 – 7
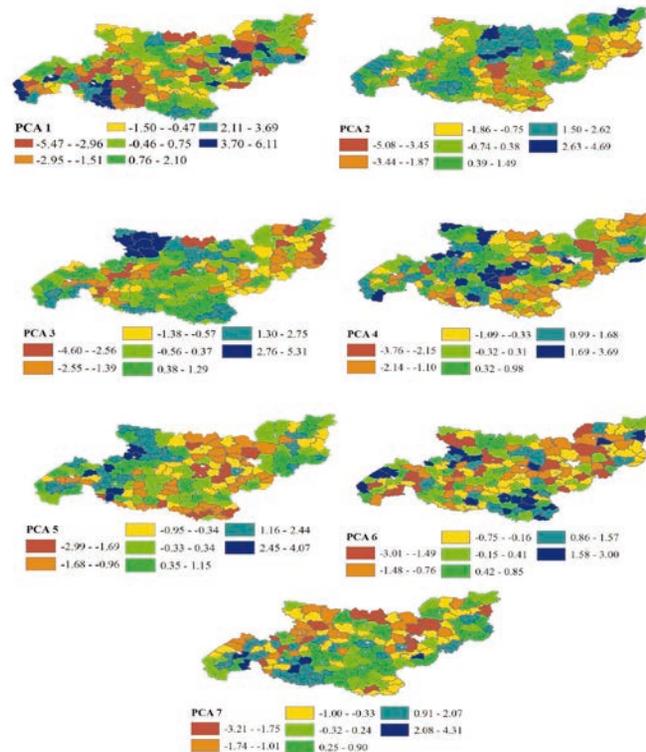
Figure 2: Spatial variation of the principal components 1-7

Principal component 7 describes 4.99 % of the variability of the feature space and varies in time within a period of about 5 years. The *I*-Moran statistic value (0.27, *p* = 0.001) indicates statistically positive autocorrelation and, thus, demonstrates the presence of a significantly clustered spatial pattern. Clusters with higher scores of the principal component 7 are concentrated in the southwest, and with the lower ones, in two clusters in the north of the region (Fig.2).

3.1 GEOGRAPHICALLY WEIGHTED PRINCIPAL COMPONENT ANALYSIS

Monte Carlo test was conducted to examine whether data matrix eigenvalues are spatially varying. As shown in Figure 3, the *p*-value for testing the local eigenvalues of standard deviations from GWPCA is 0.05. This value demonstrates that the spatial invariant hypothesis of local eigenvalues is significantly rejected at the 95 % level; or rather, there is a certain degree of spatial non-stationarity present in the data of regional cereal yield.

A key feature in GWPCA is finding the scale at which each localized PCA should operate; that is, choosing the kernel bandwidth. Before searching for an

optimal bandwidth, it is necessary to decide a prior upon the number of components to retain (Harris et al., 2015; Gollini et al., 2015; Li et al., 2015). The previous global PCA results indicate that the first seven components can collectively explain 66.8 % of the variance in data structure. Consequently, it is reasonable to retain seven components for further GWPCA analysis.

For this study, we are guided by an automatic routine for bandwidth selection. Through an adaptive bandwidth selection procedure, an optimal bandwidth of 441 km has been reached, which is chosen to run the GWPCA analysis.

The GWPCA outputs can be visualized and interpreted, focusing on: (1) how the data dimensionality varies spatially and (2) how the original variables influence the components (Li et al., 2015). The spatial distribution of local PTV of first seven components can be mapped. The percentage of the spatial variation of the general variation shows a clearly pronounced variability, with the formation of spatially homogeneous clusters in the meridional direction. Compared with the outputs from global PCA, the GWPCA has exhibited its power and strength in analyzing spatial patterns of regional cereal yields by mapping spatial variations of local principal components. Further, the local variance at each admin-
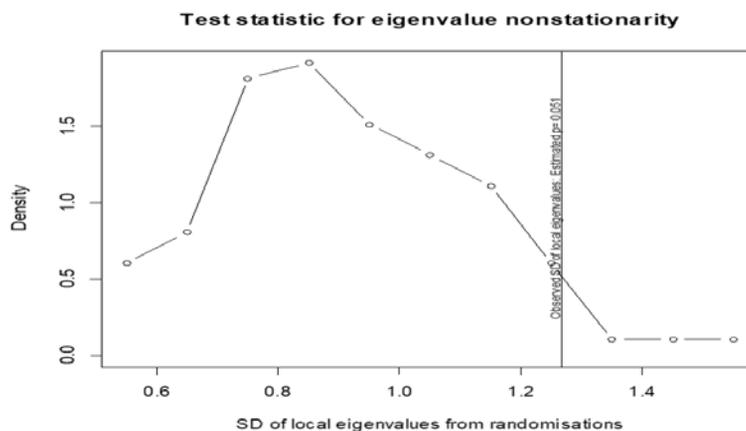
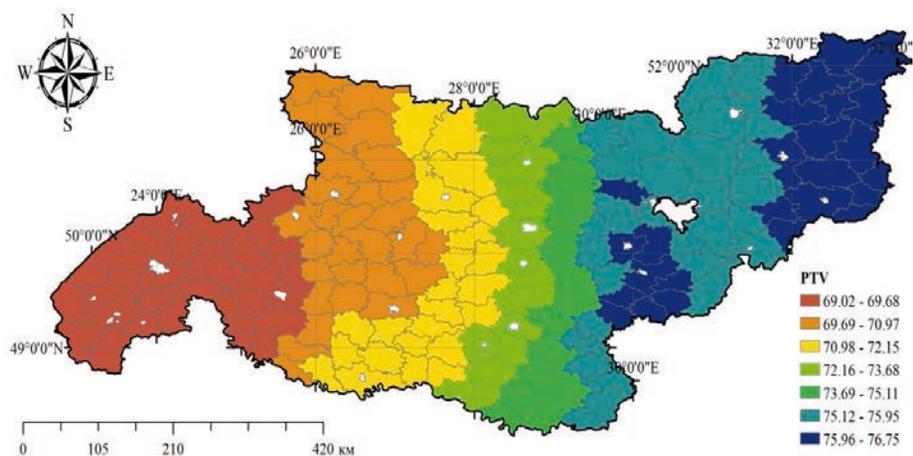**Figure 3:** Monte Carlo test for GWPCA

Figure 4: Spatial variation of the percentage of the total variation of the first seven principal components (percentage of total variance – PTV)

istrative district explained by the calculated GWPCA 1–7 can be visualized by mapping as well (fig. 4), which shows a clear east-west trend with the highest percentage variances distributed in the east, intermediate level in the central areas and the lowest values in the west. The obvious spatial clustering trend identified from the variance values in Figure 4 suggests that the interactions among these variables converge spatially.

It was suggested that the variables with the highest loading values and their impact intensity values can be mapped locally (Lloyd, 2010). We can next visualize how each of the seven variables locally influences a given component, by mapping the 'winning variable' with the highest absolute loading. Figure 5 shows the spatial distribution of variables with the absolute highest loading from GWPC 1–7 respectively.

The largest absolute loadings of a variable, which are the peculiarities of the spatial distribution of the grain

yields in a given year, can be interpreted as a marker of the greatest sensitivity to oscillatory dynamics over time, as was shown for the global PCA. The local solutions can either largely correspond to the global outcome or vary in the significance of the oscillatory processes at the regional level, which may cause a change in the order of the components, or may be manifested in gaining statistical significance for processes that are not statistically significant at the global level.

The traditional representation of the "winning" variables for the principal components cannot fully reveal the nature of the spatially dependent relationship between the indicators estimated by the PCA. The overweigh of the factor loading is one of the aspects that reflects in the crop yields dynamics. Due to the oscillating nature of this dynamics, overweighs are the random outlier of the indicator at a certain moment of time in comparison with the general repetitive dynamics. Therefore, for each
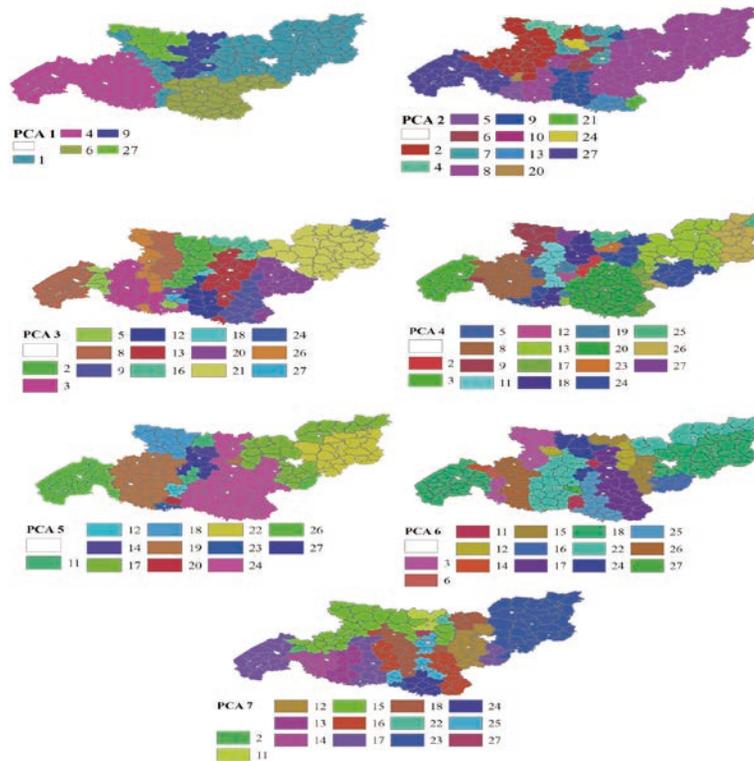
Figure 5: Spatial location of variables with the largest loading for the principal components 1-7
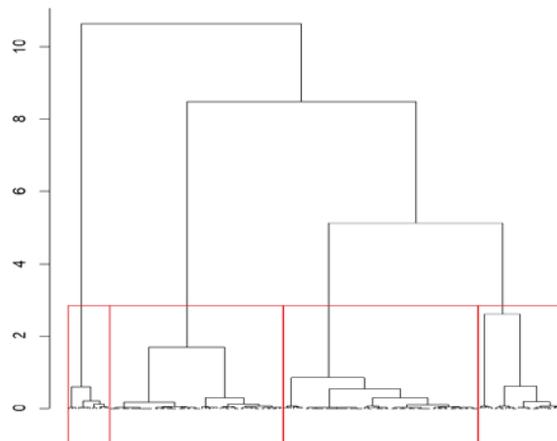

Figure 6: Cluster analysis of administrative districts by factor loadings values GWPC

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. Such a distance is sensitive to the form of comparable indicators, and not to their absolute values. This approach allows identifying groups of administrative districts with the similar time dynamics of cereal and grain legumes yields in the aspect

of the corresponding principal component. It can be assumed that the aggregate of the administrative districts with similar yields dynamics are also geographically close and form the homogeneous ecological regions.

To be consistent with the global PCA analysis, only the first component GWPC 1 from GWPCA will be interpreted in details, because it explains 22.60 % of the total variability.
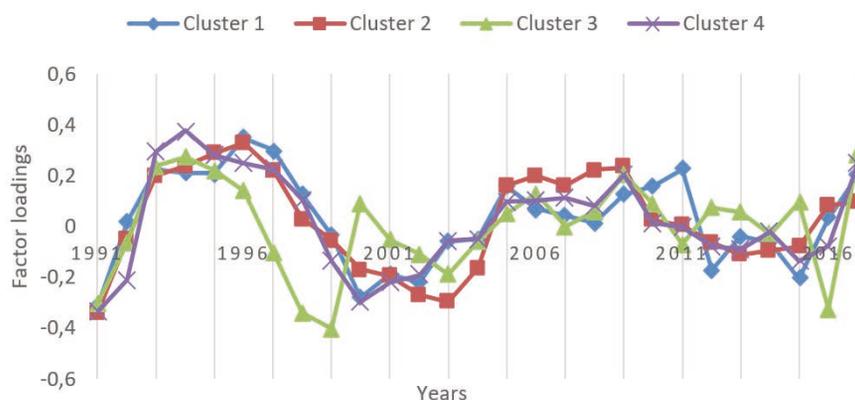
Figure 7: The average values of factor loadings of GWPC 1 for clusters 1-4
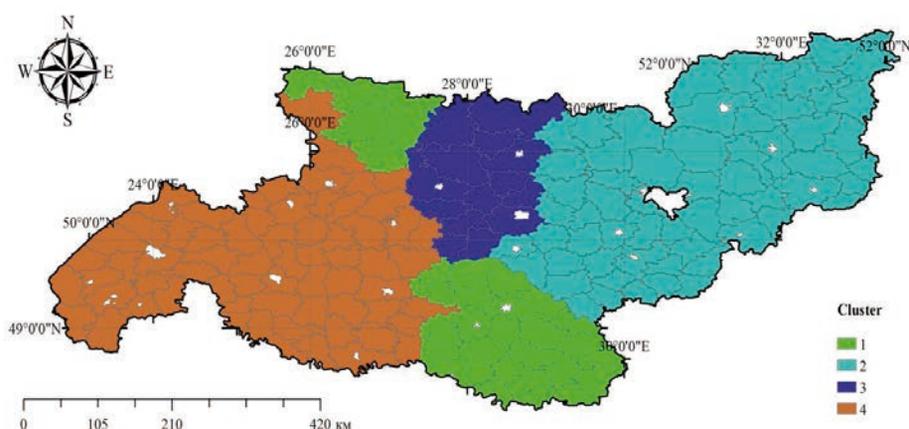

Figure 8: Spatial location of clusters obtained on the basis of factor loadings of GWPC 1

Cluster analysis of administrative districts by factor loading values GWPC 1 allowed the establishment of four homogeneous clusters (Fig. 6).

The average factor loadings for each cluster were calculated and the specifics of the corresponding clusters can be appreciated (Fig. 5, 7). For cluster 4, the greatest loading is characteristic for the variable 4 (yield in 1994), which corresponds to the so-called "winning" variable. But the information of the "winning variable" does not indicate anything about the characteristics of the overall yield dynamics within the corresponding cluster. For the cluster 3, the "winning" variable is the smallest (the largest module) variable 9, which is obviously an outlier, if the factor loading by this variable would be smaller, then the bidder for the "winning" would be the variable 27, which is also by its origin most likely an outlier.

Established clusters and their factor loadings distributions explaining the process progress that are characteristic for the corresponding cluster. Appar-

ently, the general trend of the principal component 1 is the damping of the amplitude of oscillations during the research period and the predominance of higher frequency components of oscillatory dynamics corresponding to the heterogeneity of observations overtime or the heteroscedasticity. So, the Koenker-Bassett test for cluster 1 indicates the heteroscedasticity of the time dynamics of factor loadings (3.54, $p = 0.06$) (Koenker & Bassett, 1982). The heteroscedasticity level is even higher for cluster 2 (6.89, $p = 0.008$) and for cluster 4 (10.71, $p = 0.001$). The variation of factor loadings in cluster 3 is homogeneous over time (Koenker -Bassett test 1.62, $p = 0.20$). Thus, the qualitative feature of the grain and grain legumes yields dynamics in the corresponding clusters is the different level of damping of the oscillations of the principal component 1 over time.

Spatial arrangement of administrative areas included in the corresponding clusters is spatially regular (fig. 8).

Cluster 2 occupies the east and the center of the studied area. Cluster 3 is located northwest of cluster 2. Cluster 1 is disruptive and is located in the north and south of the region. Cluster 4 covers the west of the region. The central part of the region is characterized by a steady-state oscillation regime during the research period (cluster 3) or a relatively low level of heteroscedasticity (cluster 1). For the east and the west, the damping amplitude of oscillation of factor 1 is characteristic.

4 DISCUSSION

One of the most recent approaches to the quantifying spatial variations for specific land management is based on the division of the field into land management zones according to yield level (Khosla et al., 2002). This analysis of yield maps is a fundamental tool in the investigation and understanding of the causes of yield and crop quality variations and may become the decision procedure for land management (Filho et al., 2010).

In the course of our research, we made emphasis on the evaluation of the correlative relationship between time series of cereal and grain legumes (pulses) yield within the administrative districts of the forest and forest-prairie zone of Ukraine (1991-2017). The obtained results indicate that productivity as a result of agroecosystems functioning has a complex nature and is affected by the influence of different factors. The impact of these factors can be identified through the research of synchronous dynamics characteristics. The synchronous dynamics expresses itself through the forming of the correlation relationship. The correlation matrix is the basis for the PCA and cluster analysis. PCA allows to discover the main variability trends of agricultural crops productivity. Cluster analysis led to the establishment of the homogeneous ecological area (Zhukov et al., 2018; Zymaroieva et al., 2020b).

As such, PCA enables to identify the main statistical characteristics of the regional agricultural development and reveal the intrinsic complicate interactions among the selected variables (Li et al., 2015). Thus, the global PCA revealed the presence of dynamic processes of cereal and grain legumes yields which have the oscillatory nature with different frequencies. We associate oscillatory processes of the varying frequency with causes of different nature. The oscillating processes within a period of 10 years or more may be of climatic origin. So, the oscillatory process within the longest period (13 years – larger periods are hypothetical due to relatively limited time series) is characteristic for the principal component 1 (PCA 1), which explains the largest part of grain and grain legumes yields variability (22.6 %). It

is possible to assume that among agro-ecological factors climate changes mostly influence crop productivity. For all other principal components, oscillation processes are more frequent (from 3 to 9 years). High frequency yield components may have the character of noise and may have an environmental origin as a consequence of such phenomena as the impact of diseases and pests, or the impact of weather anomalies.

Agroecological zoning was made according to the principle of uniformity of character dynamics of the production potential of agricultural areas. This approach is fundamentally different from that of zoning based on the total yield of crops (Lazarenko, 1995). Classification on the basis of absolute yield value is justified for systems that are close to the steady-state. In the face of global climate change and the transformation of the environmental regimes, this approach is unacceptable. The agroecological zones proposed by us do not differ in the overall level of productivity of grain and leguminous during the study period. Features of these zones are due to the values of principal components and reflect the nature of the relationship between different spatial units. Spatial distribution of principal components indicates a continual pattern, but their overlapping allows us 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.

5 CONCLUSIONS

The global principal components analysis revealed the presence of dynamic processes of cereal and grain legumes yields which have the oscillatory nature with different frequencies. We associate oscillatory processes of the varying frequency with causes of different nature.

The oscillating processes with a period of 10 years or more may be of climatic origin. So, the oscillatory process with the longest period (13 years – longer periods are hypothetical due to relatively limited time series) is characteristic for the principal component 1 (PCA 1), which explains the largest part of cereal and grain legumes yields variability (22.6 %). It is possible to assume that among agro-ecological factors climate changes mostly influence crop productivity. For all other principal components, oscillation processes are more frequent (from 3 to 9 years). High frequency yield components may have the character of noise or may have environmental origins as a consequence of such phenomena as the impact of diseases and pests, or the impact of weather anomalies.

Due to geographically weighted principal components analysis, spatial units with a similar oscillating component of the cereal and grain legumes yields variation were revealed. Since we consider only the environmental component of yield variation, territorial clusters, within which the yields dynamics are the same, can be considered as agroecological zones for crops cultivation.

6   REFERENCES

Anselin, L., Syabri, I., Kho, Y. (2005). GeoDa: An introduction to spatial data analysis. *Geographical Analysis, 38* (1), 5–22. https://doi.org/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. ISSN0924-2716. https://doi.org/10.1016/j.isprsjprs.2016.06.014

Filho, O. G., Vieira, S. R., Chiba, M. K., Nagumo, C. H., Dechen, S. C. F. (2010). Spatial and temporal variability of crop yield and some Rhodic Hapludox properties under no-tillage. *Revista Brasileira de Ciência do Solo, 34*(1). http://dx.doi.org/10.1590/S0100-06832010000100001

Fotheringham, A.S., Brunsdon, C., Charlton, M. (2002). *Geographically weighted regression: The analysis of spatially varying relationships.* Wiley, Chichester, 284 p. ISBN: 978-0-471-49616-8.

Frieler, K., Schauberger, B., Arneth, A., Balkovi, J., Chryssanthacopoulos, J., Deryng, D., … Leverman, A. (2017). Understanding the weather signal in national crop-yield variability. *Earth's Future, 5,* 605-616. http://dx.doi.org/10.1002/2016EF000525

Gollini, I., Lu, B., Charlton, M., Brunsdon, Ch., Harris, P. (2015). GWmodel: An R package for exploring spatial heterogeneity using geographically weighted models. *Journal of Statistical Software, 63*(17), 1–52. https://doi.org/10.18637/jss.v063.i17

Hammond, M. P. & Kolasa, J. (2014). Spatial variation as a tool for inferring temporal variation and diagnosing types of mechanisms in ecosystems. *PloS one, 9*(2), e89245. https://doi.org/10.1371/journal.pone.0089245

Harris, P., Brunsdon, C. & Charlton, M. (2011). Geographically Weighted Principal Components Analysis. *International Journal of Geographical Information Science, 25*(10), 1717–1736. http://dx.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. http://dx.doi.org/10.1111/gean.12048

Hatzinger, R., Hornik, K., Nagel, H., Maier, M. J. (2014). *R: Einführung durch angewandte Statistik* (2nd ed.). München: Pearson Studium.

Horn, J.L. (1965). A rationale and a test for the number of factors in factor analysis. *Psychometrika, 30,* 179–185. http://dx.doi.org/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. http://dx.doi.org/10.2136/sssaj2004.0154

Kaiser, H. F. (1974). An Index of Factorial Simplicity. *Psychometrika, 39* (1), 31–36. *https://doi.org/10.1007/BF02291575*

Kaspari, M. & Yanoviak, S. (2009). Biogeochemistry and the Structure of Tropical Brown Food Webs. *Ecology, 90,* 3342–51. *https://doi.org/10.1890/08-1795.1*

Khosla, R., Fleming, K., Delgado, J.A., Shaver, T., Westfall, D.G. (2002). Use site-specific management zones to improve nitrogen management for precision agriculture. *Journal of Soil and Water Conservation, 57,* 513-518.

Koenker, R. & Bassett, G.Jr. (1978). Regression Quantiles. *Econometrica, 46* (1), 33-49.

Koenker, R. & Bassett, G.Jr. (1982). Robust tests for heteroscedasticity based on regression quantiles. *Econometrica, 50*(1), 43-61. *https://doi.org/10.2307/1913643*

Kong, L.Q., Zheng, H., Rao, E.M., Xiao, Y., Ouyang, Z.Y., Li, C. (2018). Evaluating indirect and direct effects of eco-restoration policy on soil conservation service in Yangtze River Basin. *Science of the total environment, 631-632,* 887-894. http://dx.doi.org/10.1016/j.scitotenv.2018.03.117.

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. http://dx.doi.org/10.1007/s10596-012-9290-6

Lauzon, J. D., Fallow, D. J., O'Halloran, O. P., Gregory, S. D. L., Bertoldi, A. P. (2005). Assessing the temporal stability of spatial patterns in crop yields using combine yield monitor data. *Canadian journal of soil science, 85*(3), 439-451. https://doi.org/10.4141/S04-067

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

Legendre, P. & Gallagher, E. (2001). Ecological Meaningful Transformations for Ordination of Species Data. *Oecologia, 129,* 271–80. *https://doi.org/10.1007/s004420100716*

Li, Y.S. & Huang, M.B. (2008). Pasture yield and soil water depletion of continuous growing alfalfa in the Loess Plateau of China. *Agriculture, Ecosystem & Environment, 124*(1–2), 24–32. https://doi.org/10.1016/j.agee.2007.08.007

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 Science, 9*(3), 233-245. https://doi.org/10.1007/s12076-015-0154-2

Liu, X., Zhu, X. H., Qiu, P., Chen, W. (2012). Correlation-Matrix-Based Hierarchical Clustering Method for Functional Connectivity Analysis. *Journal of Neuroscience Methods, 211*(1), 94-102. http://dx.doi.org/10.1016/j.jneumeth.2012.08.016

Lloyd, C.D. (2010). Analysing population characteristics using geographically weighted principal components analysis: a case study of Northern Ireland in 2001. *Computers, Environment and Urban System, 34*(5), 389–399. *https://doi.org/10.1016/j.compenvurbsys.2010.02.005*

Lobell, D. B., Hammer, G. L., McLean, C., Roberts, M. J., Schlenker, W. (2013). The critical role of extreme heat for maize

production in the United States. *Nature Climate Change, 3*(5), 497–501. https://doi.org/10.1038/nclimate1832

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

Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine, 2*(7–12), 559–572. https://doi.org/10.1080/14786440109462720

R Core Team (2017). R: A language and environment for statistical computing. R foundation for statistical computing, Vienna, Austria. Retrived from URL https://www.R-project.org/

State Statistics Service of Ukraine. Retrieved from http://www.ukrstat.gov.ua/

Zhukov, O. V. & Ponomarenko, S. V. (2018). Spatial-time dynamics of cereals of grain and grain crops in Poltava region. *Bulletin of Poltava State Agrarian Academy, 1,* 55–62.

Zhukov, O.V., Pelina, T.O., Demchuk, O. M., Demchuk, N. I., Koberniuk, S.O. (2018). Agroecological and agro-economic aspects of the grain and grain legumes (pulses) yield dynamic within the Dnipropetrovsk region (period 1966-2016). *Biosystems Diversity, 26*(2), 3–10. *https://doi.org/10.15421/011826*

Zymaroieva, A., Zhukov, O., Fedonyuk, T., Pinkin, A. (2019a). Application of geographically weighted principal components analysis based on soybean yield spatial variation for agro-ecological zoning of the territory. *Agronomy Research, 17*(6), 2460–2473. https://doi.org/10.15159/AR.19.208

Zymaroieva, A., Zhukov O., Romanchuck, L., Pinkin A. (2019b). Spatiotemporal dynamics of cereals grains and grain legumes yield in Ukraine. *Bulgarian Journal of Agricultural Science, 25*(6), 1107–1113.

Zymaroieva, A., Zhukov, O., Romanchuck, L. (2020a). The spatial patterns of long-term temporal trends in yields of soybean *(Glycine max (L.) Merril)* in the Central European mixed forests (Polissya) and East European forest steppe ecoregions within Ukraine. *Journal of Central European Agriculture*, *21*(2), 320-332. https://doi.org/10.5513/JCEA01/21.2.2402

Zymaroieva A., Zhukov O., Fedonyuk T., Pinkina T. (2020b). The spatio-temporal trend of rapeseed yields in Ukraine as a marker of agro-economic factors influence. *Agronomy Research,* 18*(S2)*, 1584–1596. https://doi.org/10.15159/AR.20.119