

doi: 10.33249/2663-2144-2019-83-10-20-27

UDC 631.559

**THE PERSPECTIVES OF THE APPLICATION OF GEOGRAPHICALLY WEIGHTED PRINCIPAL COMPONENTS ANALYSIS FOR ESTIMATION OF MAIZE YIELDS SPATIAL VARIABILITY****A. Zymaroieva***e-mail: nastya.zymaroieva@gmail.com*Zhytomyr National Agroecological University,  
7, Stary Blvd, Zhytomyr, 10002, Ukraine

World population is projected to increase 35% by 2050, which will require a 70–100% rise in food production given projected trends in diets, consumption, and income. Therefore, the study of the reasons of the stagnation of grain crop yields and the opportunities for its increase in the countries of the world, and in Ukraine, in particular, is a critical issue at the present time. This paper aims to explore spatial heterogeneity present in the maize yields data collected from 206 administrative districts in Polissya and Forest-Steppe zone of Ukraine for 27 years using the GWPCA method. The principal components analysis of the residues of the time trend regression models allowed establishing 2 principal components, which together explain 34.1% of variation in the maize yield. There are two spatial determined processes that influence the yield of maize and have the oscillatory dynamics of different periodicity. The oscillating phenomena are of an ecological nature. Geographically weighted principal component analysis showed spatial non-stationary environmental regimes, which determine the oscillatory component of variation of maize yield over time. Geographically weighted principal component analysis allows us to investigate local patterns of maize yield dynamics. Based on the approximate types of local dynamics, we have conducted a cluster analysis for each principal component. Due to cluster mapping the ecologically homogeneous territories where the dynamics of environmental processes are the same can be established. Consequently, the application of geographically weighted principal components analysis with regard to yield data of any crop enables to conduct agroecological zoning of the territory and to identify dynamic aspects of yield determinants.

**Key words:** yield, maize, variation, dynamics, trend, geographically weighted principal components analysis (GWPCA).

**ПЕРСПЕКТИВИ ВИКОРИСТАННЯ ГЕОГРАФІЧНО ЗВАЖЕНОГО АНАЛІЗУ ГОЛОВНИХ КОМПОНЕНТ ДЛЯ ОЦІНКИ ПРОСТОРОВОЇ ВАРІАБЕЛЬНОСТІ ВРОЖАЙНОСТІ КУКУРУДЗИ****А. А. Зимароєва***e-mail: nastya.zymaroieva@gmail.com*Житомирський національний агроекологічний університет  
бульвар Старий, 7, м. Житомир, 10008, Україна

Прогнозують, що до 2050 року чисельність населення світу зросте на 35% і вимагатиме збільшення обсягів виробництва продовольства на 70–100%. Тому визначення причин стагнації врожайності сільськогосподарських культур та можливостей її нарощення в країнах світу, і в Україні зокрема, є критично важливим питанням наразі. Мета дослідження: за допомогою географічно зваженого аналізу головних компонент (GWPCA) дослідити, на прикладі кукурудзи, просторову неоднорідність, яка присутня в статистичних даних урожайності цієї культури, зібраних з 206 адміністративних районів Поліської та Лісостепової зон України впродовж 27 років і на її основі провести агроекологічне зонування території. Глобальний аналіз головних компонент залишків регресійної моделі дозволив встановити 2 головні компоненти, які разом пояснюють 34,1% загальної варіабельності урожайності кукурудзи. Географічно зважений аналіз головних компонент показав просторову нестационарність екологічних режимів, які визначають коливальну складову варіювання

урожайності кукурудзи у часі. А також, GWPCA дозволив дослідити локальні патерни у динаміці врожайності кукурудзи. На основі наближених типів локальної динаміки нами проведено кластерний аналіз для кожної головної компоненти. За допомогою картування кластерів можна встановити екологічно гомогенні території, де динаміка екологічних процесів однакова. Отже, застосування географічно зваженого аналізу головних компонент щодо даних врожайності будь-якої сільськогосподарської культури дозволяє провести агроекологічне зонування території та виявити динамічні аспекти детермінантів врожайності.

**Ключові слова:** урожайність, кукурудза, варіювання, динаміка, тренд, географічно зважений аналіз головних компонент.

### Вступ

За прогнозами вчених до 2050 року чисельність населення світу зросте на 35% і вимагатиме збільшення обсягів виробництва продовольства на 70–100% враховуючи сучасні тенденції в дієтах, споживанні та доходах (Rosegrant et al, 2009). Збільшення виробництва продуктів харчування може досягатися за рахунок підвищення врожайності сільськогосподарських угідь на існуючих сільськогосподарських землях, розширення площі виробництва рослин або обома цими шляхами. Проте, розширення площі посівів викликає значне збільшення викидів парникових газів, що, відповідно, призводить до зміни клімату (Wart et al, 2013).

Потенціал зростання врожайності також має ліміт, зокрема, за останніми дослідженнями, урожайність рису є стагнаційною в Японії та Китаї; урожайність кукурудзи застійна у Китаї, Італії та Франції; а врожайність пшениці не зростає у Північній Європі та Індії. Така стагнація врожайності в основних районах вирощування зернових має вплив на інші регіони, які змушені прискорювати темпи приросту врожаю або розширювати площі під культурами, аби скоротити різницю між глобальними пропозицією і попитом (Rosegrant et al, 2009, Wart et al, 2013). Тому, дослідження причин стагнації врожайності та можливостей щодо її нарощення в інших країнах світу, зокрема і в Україні, є критично важливим питанням наразі.

Агроекологічна зона – це територія з аналогічним перебігом екологічних процесів (Sivakumar & Valentin, 1997). Отже, агроекологічне районування означає поділ площі (земельної ділянки) на менші одиниці, які мають подібні характеристики, пов'язані з потенціалом придатності земель до виробництва та впливом на довкілля (Patel, 2003). Врожайність культури є функціональним показником, який повною мірою відбиває складні взаємовідносини між рослиною

та довкіллям (Anderson et al, 2013). Тому застосування врожайності як основного показника для агроекологічного районування цілком виправдано.

Існує декілька підходів до вивчення довгострокових трендів у даних варіювання врожайності на основі технічних або статистичних моделей (Patel, 2003, Lloyd, 2010). У цьому дослідженні ми пропонуємо використати низку локальних (нестационарних) статистичних моделей, що називається географічно зваженим аналізом головних компонент (geographically weighted principal components analysis – GWPCA).

GWPCA – локалізована версія аналізу головних компонент (PCA), є дослідницьким інструментом для визначення просторової неоднорідності в структурі багатовимірних даних. Даний метод не тільки надає корисний інструмент для досліджень, але також розвиває результати PCA в локалізованому масштабі. Отже, GWPCA визначає, як результати отримані за допомогою PCA, варіюються просторово. Просторові зміни в розмірності даних і багатовимірній структурі можуть бути досліджені за допомогою карт результатів географічно зваженого аналізу головних компонент. GWPCA може також використовуватися для виявлення багатовимірних просторових аномалій (Harris, 2015).

В різноманітних літературних джерелах GWPCA широко застосовується для аналізу багатовимірних характеристик популяції (Lloyd, 2010), соціальної структури (Harris, 2015), характеристик ґрунтів та даних хімії прісної води (Li et al, 2015). Проте, GWPCA не застосовувався для оцінки просторової мінливості врожайності сільськогосподарських культур, які за своєю сутністю є просторово гетерогенними. З метою заповнення цієї прогалини були проведені ці дослідження.

Метою даної роботи є вивчення за допомогою методу GWPCA просторової

неоднорідності, яка присутня в даних урожайності, на прикладі кукурудзи, що зібрані з 206 адміністративних районів Поліської та Лісостепової зон України протягом 27 років.

Відповідно до поставленої мети передбачалося вирішення наступних завдань:

- виявити коливальну складову варіювання урожайності кукурудзи у часі;
- дослідити локальні тренди у динаміці врожайності кукурудзи;
- провести агроекологічне зонування території щодо вирощування кукурудзи.

### Матеріали та методи

Дані з урожайності кукурудзи у Поліській та Лісостеповій зонах України представлені Державною службою статистики України (<http://www.ukrstat.gov.ua/>). Відомості охоплюють часовий період з 1991 по 2017 рр. Дані мають характер середньої врожайності культури по адміністративному району. Територія охоплює 206 адміністративних районів з десяти областей України (Вінницька, Волинська, Житомирська, Київська, Львівська, Рівненська, Тернопільська, Хмельницька, Черкаська, Чернігівська).

Часові ряди врожайності для кожного адміністративного району були розділені на дві складові: загальний тренд і залишковий тренд. Загальний тренд відображав залежність врожайності від часу. Як аналітичну форму тренду ми обрали поліном 4-го ступеня. Залишки відповідних регресійних моделей, що описують загальний тренд, складаються з випадкової складової (шум) і, ймовірно, регулярної, яку неможливо пояснити вибраною моделлю тренду. Ці дві компоненти відрізняються своїми властивостями: випадкова складова є незалежною для різних точок простору, а регулярна складова має корелювати з усіма або з деякими точками простору (адміністративними районами). Ми використовували аналіз основних компонент (PCA) для залишків (викидів) регресійної моделі, щоб виділити регулярну складову моделей тренду. Наявність головних компонент, чий власні значення більше ніж одна одиниця, свідчать про наявність кореляції в варіаціях урожайності сільськогосподарських культур. Обчислення виконані за допомогою бібліотеки REdaS для середовища статистичних розрахунків R. Метод GWPCA реалізований за допомогою пакета GWmodel R (<http://cran.rstudio.com/>) (Kunah, 2018).

Просторова база даних була створена в ArcGIS 10.0. Для обчислення глобального коефіцієнта просторової автокореляції застосована статистика I-Морана, яка подібна до коефіцієнта кореляції Пірсона. Обидві статистики варіюють у межах від +1,0 (указує на сильну позитивну кореляцію) до 0 (указує на випадковий патерн) та до -1,0 (указує на сильну негативну автокореляцію). Глобальна статистика Морана обчислена із застосуванням програми Geoda095i (<http://www.geoda.uiuc.edu/>) (Kunah, 2018; Zymarioieva, 2019).

### Результати досліджень та обговорення

У наших попередніх дослідженнях (Zymarioieva, 2019) було встановлено, що остаточний тренд врожайності кукурудзи в межах досліджуваної області найкраще описується поліномом четвертого порядку. Регресія за допомогою поліному четвертого порядку пояснює значну частину дисперсії урожайності кукурудзи (від 62 до 83% в залежності від району досліджень). Наявність тренду, математична форма якого є незмінною, вказує на наявність постійно діючих зовнішніх факторів на динаміку процесу. В якості такого фактору ми розглядаємо агроекономічні та агротехнологічні чинники. Але на кінцевий результат вплив здійснюють і фактори іншої природи, що мають локальний характер. У подальшому аналізі будуть застосовані залишки регресійної моделі. Ці залишки, на нашу думку, мають саме екологічну природу.

Для того, щоб встановити чи мають наші дані просторову складову варіювання, було проведено тест Монте-Карло. Як показано на рис. 1,  $p$ -рівень для тестування стандартного відхилення локальних власних чисел за результатами GWPCA становить 0,03. Це значення вказує, що існує високий рівень просторової нестационарності, який представлений у даних по врожайності кукурудзи. Отже, ці дані є придатними для проведення географічно-зваженого аналізу головних компонент.

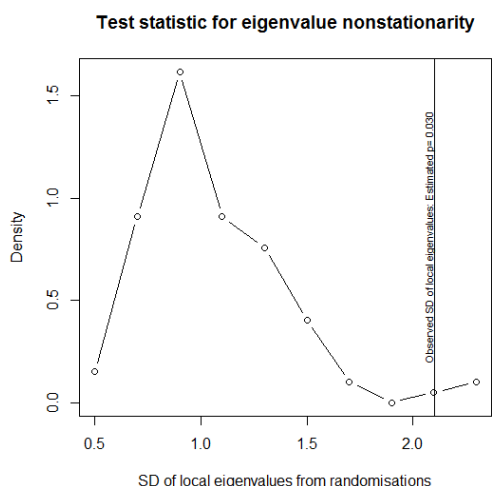


Рис. 1. Тест Монте-Карло для GWPCA

Друге питання, яке необхідно вирішити перед проведенням GWPCA : яку кількість головних компонент залишити (Harris, 2015). Аналіз головних компонент залишків регресійної моделі дозволив встановити, що за процедурою Горна (Horn, 1965) кількість статистично вірогідних головних компонент становить 5 (табл. 1). Разом перші п'ять головних компонент пояснюють 51,9% загальної варіабельності простору ознак. Для подальшого аналізу за критерієм «осипу» ми залишили перші 2 головні компоненти, які разом пояснюють 34,1% загальної варіабельності простору ознак. Таким чином, рішення залишити дві головні компоненти для подальшої процедури GWPCA є цілком обґрунтованим.

Таблиця 1. Результати глобального аналізу головних компонент

| Головна компонента | Налаштоване* власне значення | Власне значення | Зсув | Пояснена варіація | Стандартне відхилення |
|--------------------|------------------------------|-----------------|------|-------------------|-----------------------|
| 1                  | 4,82                         | 5,55            | 0,73 | 20,57             | 2,35                  |
| 2                  | 3,02                         | 3,64            | 0,62 | 13,48             | 1,90                  |
| 3                  | 1,23                         | 1,76            | 0,53 | 6,52              | 1,32                  |
| 4                  | 1,17                         | 1,63            | 0,46 | 6,03              | 1,27                  |
| 5                  | 1,03                         | 1,43            | 0,40 | 5,29              | 1,19                  |

Позначки: \* – за процедурою Горна.

У процесі процедури адаптивної селекції вікна пропускання було встановлене оптимальне вікно пропускання з 54 найближчими сусідами, яке було обрано для виконання GWPCA процедури. Для одержання відповідних до результатів глобального аналізу головних компонент, з метою порівняння були інтерпретовані тільки дві перші головні компоненти GWPC 1, GWPC 2.

візуалізовані та інтерпретовані фокусууючись на тому, як розмірність даних просторово варіює та на тому, як вихідні змінні впливають на головні компоненти. Процент просторового варіювання загальної варіації демонструє чітко виражену мінливість, при цьому, райони із низьким відсотком варіювання розміщені переважно на сході та півдні, а з високим, відповідно, на заході, півночі та центрі досліджуваного регіону (рис. 2).

Результати процедури GWPCA можуть бути

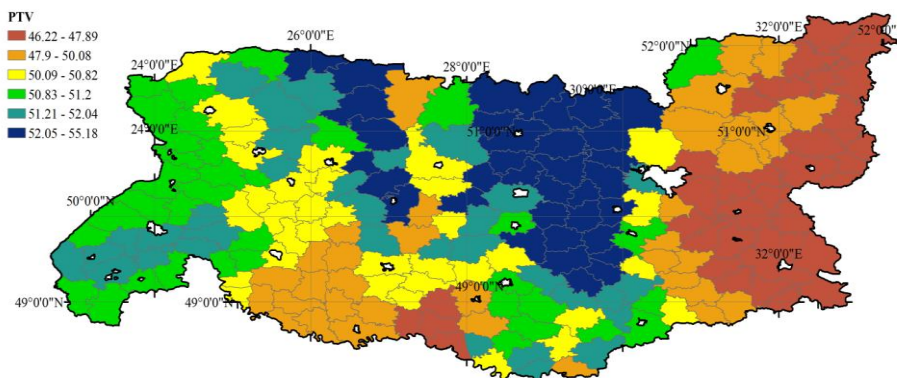


Рис. 2. Просторове варіювання проценту загальної варіації перших трьох головних компонент (percentage of total variance – PTV)

Порівняно з глобальним аналізом головних компонент, GWPCA демонструє свою ефективність та результативність в аналізі просторових патернів регіонального розміщення урожайності кукурудзи за допомогою картування просторової варіабельності головних компонент. Змінні з найбільшими навантаженнями та інтенсивність їх впливу можуть бути локально відображені (Lloyd, 2010). Традиційне представлення «виграшних» змінних для головних компонент не може повною мірою показати природу просторово залежного взаємозв'язку між показниками, який оцінений за допомогою аналізу головних компонент, оскільки у ролі «виграшних» змінних виступають дискретні величини – роки. Тому, для кожної зі статистично достовірних головних компонент нами проведена процедура класифікації адміністративних районів за допомогою кластерного аналізу на основі відстані, яка є зворотною до коефіцієнта кореляції Пірсона. Такий показник відстані чутливий до форми порівнюваних показників, а не до їх абсолютних значень. Вказаний підхід дозволяє виділити два адміністративні райони, які характеризуються подібною часовою динамікою врожайності кукурудзи у аспекті відповідної головної компоненти. Можна припустити, що сукупності адміністративних районів, які характеризуються подібною динамікою врожайності, також географічно наближені та формують однорідні екологічні регіони.

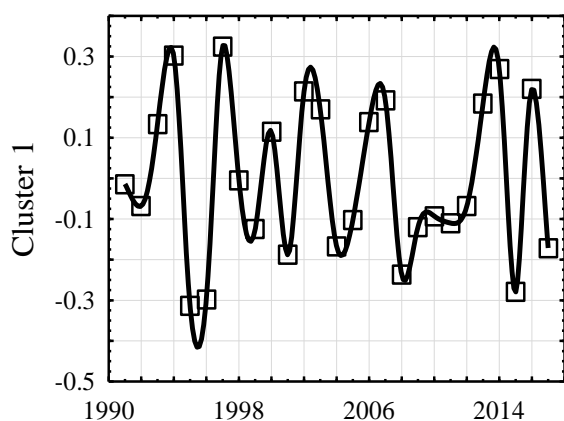


Рис. 4. Середні значення факторних навантажень GWPCA 1 для кластерів 1–2. Вісь абсцис – роки, вісь ординат – факторні навантаження

Просторове розміщення адміністративних районів, які включені у відповідні кластери, є

Кластерний аналіз адміністративних районів за значеннями факторних навантажень GWPCA 1 дозволив встановити три гомогенних кластера (рис. 3). Для кожного кластера були розраховані середні значення факторних навантажень, що дало змогу оцінити специфіку відповідних кластерів (рис. 4).

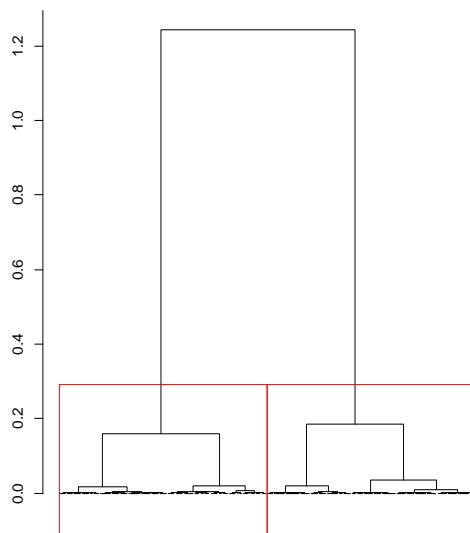
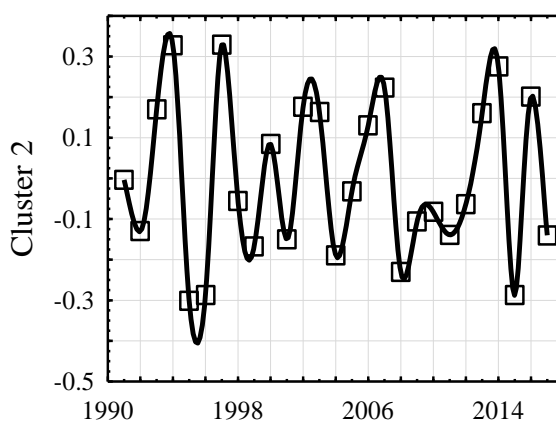
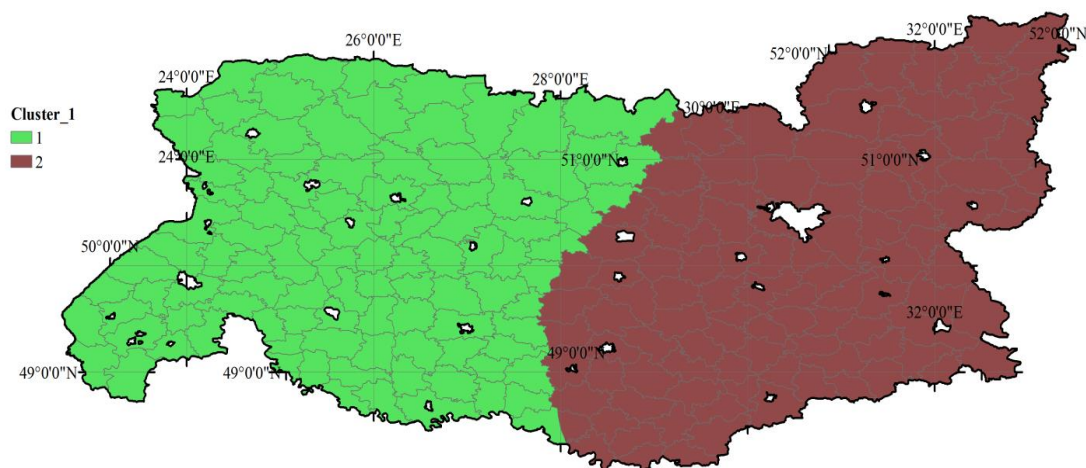


Рис. 3. Кластерний аналіз адміністративних районів за значеннями факторних навантажень GWPCA 1

Встановлені кластери та їх профільні розподіли факторних навантажень дають уявлення про перебіг процесів, які характерні для відповідного кластеру. Часові відмінності у перебігу процесів незначні (рис. 4). Ці два процеси сильно скорельовані ( $r = 0,98$ ,  $p < 0,001$ ).

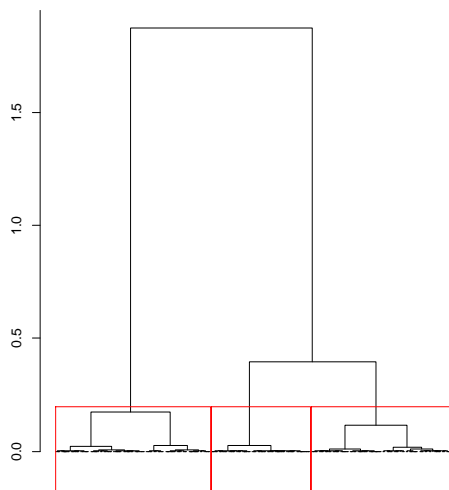


просторово регулярним і ділить територію практично навпіл (рис. 5).



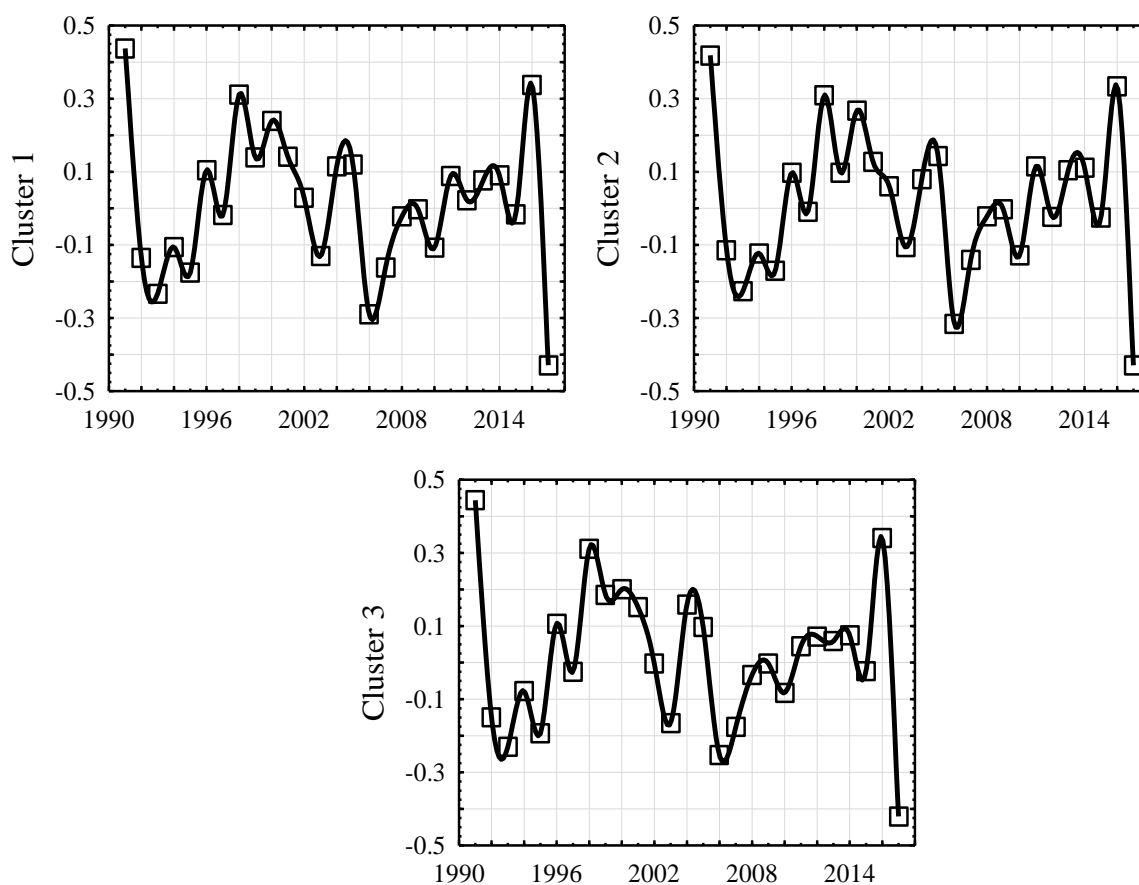
**Рис. 5. Просторове розміщення кластерів, одержаних на основі факторних навантажень GWPC 1**

Кластерний аналіз адміністративних районів за значеннями факторних навантажень GWPC 2 дозволив встановити три гомогенних кластера (рис. 6). Для кожного кластера були розраховані середні значення факторних навантажень, що дало змогу оцінити специфіку відповідних кластерів (рис. 6). Усі три кластера мають подібну динаміку коливань, що говорить про схожість та взаємообумовленість процесів, які їх визначають. У просторовому аспекті ці кластери розділяють досліджену територію майже на три рівні ділянки (рис. 8): кластер 1 займає центральну частину, кластер 2 – займає східну, а кластер 3 – західну частину території.

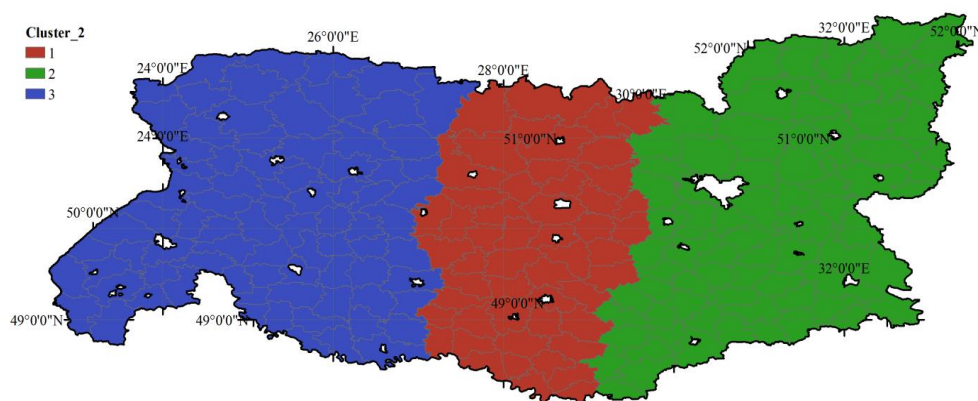


**Рис. 6. Кластерний аналіз адміністративних районів за значеннями факторних навантажень GWPC 2**

Застосування методу головних компонент для аналізу динаміки врожайності базується на припущенні одноманітності характеру взаємозв'язків у межах усїєї дослідженої території. Географічно зважений аналіз головних компонент дозволяє дослідити локальні патерни у динаміці врожайності сільськогосподарських культур, і, зокрема, кукурудзи. Локальні моделі характеризуються більшою пояснювальною здатністю, ніж тотальна модель, що цілком закономірно, оскільки урахування локальної специфіки дозволяє більш предметно віддзеркалити реальність. Але застосування такого підходу викликає певні методичні труднощі для змістовної інтерпретації (Kunah, 2018). Найбільш поширений прийом – картографування «виграшних» змінних – не є придатним у випадку аналізу часових рядів (Li et al., 2015).



**Рис. 7. Середні значення факторних навантажень GWPC 2 для кластерів 1–3. Вісь абсцис – первинні змінні (залишки регресійних моделей тренду урожайності по роках), вісь ординат – факторні навантаження**



**Рис. 8. Просторове розміщення кластерів, одержаних на основі факторних навантажень GWPC 2**

Тому, на основі наближених типів локальної динаміки, нами проведено кластерний аналіз для кожної головної компоненти і застосовано саме

картографування цих кластерів. Такий підхід має певні переваги. Так, за допомогою картування кластерів, можна встановити екологічно

гомогенні території, де динаміка екологічних процесів однакова (Sivakumar & Valentin, 1997). Ці території за своєю суттю є агроекологічними зонами щодо вирощування кукурудзи. Також, запропонований нами методичний прийом надає можливості змістовно інтерпретувати одержані кластери за допомогою дослідження динаміки кожного кластера у часі. Якщо паралельно накладати карти кліматичних змін, родючості ґрунтів, ландшафтного біорізноманіття тощо, то можна виявити екологічні детермінанти врожайності культури, у нашому випадку – кукурудзи. Отже, застосування географічно зваженого аналізу головних компонент щодо даних врожайності кукурудзи дозволяє провести агроекологічне зонування території та виявити динамічні аспекти детермінантів врожайності.

### Висновки

Наше дослідження підтвердило гіпотезу, що в межах досліджуваної території є зони із специфічними закономірностями часової динаміки врожайності кукурудзи, які є однаковими в межах кожної зони, але якісно відрізняються між зонами. Аналіз головних компонент залишків регресійної моделі часового тренду врожайності дозволив нам встановити 2 головні компоненти, які разом пояснюють до 34,1% дисперсії врожайності кукурудзи. Це свідчить про те, що два просторово обумовлені процеси впливають на врожайність кукурудзи та мають коливальну динаміку різної періодичності. Ці коливальні явища мають екологічну природу. Географічно зважений аналіз головних компонент дозволив виявити просторові одиниці з подібною коливальною складовою коливання врожайності кукурудзи. Територіальні кластери, в межах яких динаміка врожайності кукурудзи однакова, можна вважати агроекологічними зонами щодо вирощування кукурудзи. Вивчення природи головних компонент буде метою наших наступних досліджень, зокрема, планується вивчення впливу змін клімату на варіювання врожайності.

### References

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. *Four decades of progress in monitoring and modeling of processes in the soil-plant-atmosphere system: applications and challenges*, 19, 46–57.

Harris, P., Clarke, A., Juggins, S., Brunson, 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, 146–172. doi: 10.1111/gean.12048.

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

Kunah, O. M., Pakhomov, O. Y., Zymarioieva, A. A., Demchuk, N. I., Skupskiy, R. M., Bezuhla, L. S. & Vladyka, Y. P. (2018). Agroecological and agroecological aspects of the rye (*Secale cereale* L.) yields spatial variation within Polesia and Foreststeppe zones of Ukraine: the useage of the geographically weighted principal components analysis. *Biosystem Diversity*, 26 (4), 276–285. doi: <https://doi.org/10.15421/011842>.

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.

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.

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

Rosegrant, M. W., Ringler, C., & Zhu, T. J. (2009). Water for agriculture: maintaining food security under growing scarcity. *Ann. Rev. Env. Resour.*, 34, 205–222.

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.

Wart, J. K., Kersebaum, K. C., Peng, S., Milner, M. & Cassmana, K. G. (2013). Estimating crop yield potential at regional to national scales. *Field Crops Research*, 143, 34–43. doi: <https://doi.org/10.1016/j.fcr.2012.11.018>.

Zymarioieva, A. A. (2019). Prostorovo-chasovi zakonomirnosti variiuvannia urozhainosti kukurudzy v Ukraini [Spatial-temporal patterns of maize yield variation within Ukraine]. *Naukovi horyzonty*, 2, 58–66 [in Ukrainian].