



Генетика, селекція, біотехнологія

УДК 631.522/524

© 2024

ЗАСТОСУВАННЯ ІННОВАЦІЙНИХ МЕТОДІВ МАТЕМАТИКО- СТАТИСТИЧНОГО АНАЛІЗУ В СУЧАСНИХ ТЕХНОЛОГІЯХ СЕЛЕКЦІЇ РОСЛИН В УКРАЇНІ

В.В. Чернуський¹, С.Д. Орлов², С.С. Климчук³

¹кандидат сільськогосподарських наук

²доктор сільськогосподарських наук

^{1,2}Інститут біоенергетичних культур і цукрових буряків НААН
вул. Клінічна, 25, м. Київ, 03141, Україна

³Уладово-Люлинецька дослідно-селекційна станція

Інституту біоенергетичних культур і цукрових буряків НААН
вул. Семполовського, 15, с. Уладівське Калинівського р-ну Вінницької обл.

e-mail: ¹vadimchernuskiy_58@ukr.net, ²orlov.stanislav48@gmail.com, ³uldss1888@ukr.net

ORCID: ¹0000-0002-8477-1050, ²0000-0001-5759-862X

Надійшла 24.04.2024

Мета. Розробити принципи застосування інноваційних статистичних методів аналізу експериментальних даних для оптимізації системи добору селекційних зразків на продуктивність і адаптивність в умовах зміни клімату. **Методи.** Польовий (для формування методично верифікованих експериментальних популяцій селекційних зразків), інструментально-камеральний (для автоматизації структурного аналізу рослин), матрично-цифровий (для формування метабази даних), математико-статистичний. **Результати.** Взаємодія генотип – середовище (ВГС) є важливим елементом селекційних технологій, спрямованих на пришвидшення темпів створення сортів в умовах зміни клімату шляхом оптимізації системи добору. За результатами досліджень зернової продуктивності селекційних зразків гороху посівного в конкурсному сортовипробуванні 2021 – 2023 рр. на Уладово-Люлинецькій дослідно-селекційній станції Інституту біоенергетичних культур і цукрових буряків НААН виявлено значну мінливість статистичних характеристик варіаційних рядів за роками на достовірному рівні. Найвищі показники середньої міжпопуляційної мінливості (26,3 ц/га) зафіксовано в 2021 р., найнижчі (16,5 ц/га) – 2023 р. За однакового набору генотипів на різницю рівнів зернової продуктивності значною мірою вплинула форма епігенотипової мінливості – ВГС. За результатами аналізу регресійних рівнянь, установлених за формами поверхонь відгуку продуктивності селекційних зразків у 2021 – 2023 рр. доведено, що диференціальним і аналізуювальним

темпоральним середовищем є аналітична геометрична поверхня 2022 р. Висновки. На міжпопуляційному рівні за роками досліджень встановлено сумарну дію і модульний взаємозв'язок між елементами генотипової та епігенотипової мінливості, яка проявляється під впливом стресових факторів зовнішнього середовища. В умовах стресорної дії погодних факторів сумарний розмах фенотипової мінливості мінімізується, епігенотипової мінливості — максимізується відповідно до закону ентальпійного (продуктивного) мінімуму та ентропійного (адаптивного) максимуму на фоні стабільної генотипової мінливості порівняно з роками з оптимальними умовами для росту і розвитку рослин. Отже, добір на внутрішньопопуляційному рівні на високу генотипову складову продуктивності та виявлення високого коефіцієнта успадковування доцільно проводити в роки з несприятливими умовами для розвитку ознаки (за аналізу поверхні відгуку як селекційне плато), оскільки її абсолютний «дельта – прояв» є мінімальним і визначається переважно генотиповою складовою.

Ключові слова: горох, конкурсне сортовипробування, регресійні рівняння, аналітичні геометричні поверхні відгуку.

DOI: <https://doi.org/10.31073/agrovisnyk202406-07>

Взаємодія генотип — середовище (ВГС) є важливим елементом селекційних технологій з урахуванням фенотипової мінливості і пришвидшення темпів створення сортів в умовах зміни клімату.

У програмах селекції рослин використано дані випробувань у кількох середовищах (МЕТ) для вибору кращих ліній з метою збільшення генетичного прибутку. Точність вибору можна підвищити за використання передових методів статистичного аналізу із застосуванням інформативних моделей для аналізу взаємодії генотипу з навколишнім середовищем [1].

Аналіз кореляційних шляхів показав, що зміни кліматичних факторів спричинили 26% варіації врожайності [2].

На думку дослідників [3], інтеграція біологічних ідей у моделі прогнозування зазвичай ґрунтується на створенні чітких формулювань для GxE. Ці стратегії моделювання мають інтегрувати фенотипову інформацію, що належить до різних масштабів біологічної організації, зокрема і в різних умовах середовища. Практичною реалізацією може бути класична схема дисперсій і коваріацій, доповнена введенням в оцінку біологічної інформації у формі альтернативних

статистичних інструментів, таких, як попередні байєсівські методи або штрафні багатовимірні регресії.

Дослідженнями [4] встановлено, що індекси екстремальної погоди мають статистично значущу інформацію на рівні оцінки лінійних і нелінійних методів попарної залежності. Розглянуто додаткову цінність методів нелінійної регресії та висунуто припущення, що лінійні підходи самі по собі можуть бути недостатніми. Розроблено методологію та актуалізовано шкалу індексів і наскрізних систем оцінки ризику щодо впливу клімату на врожайність.

Кліматичні фактори, такі як кількість опадів і температура, відіграють важливу роль у характеристиках росту рослинності. Зв'язок між кліматом і ростом рослинності можна точно передбачити. Авторегресійну модель, перетворену за Койком, використовували для побудови регресійних моделей [5].

У роботі [6] проаналізовано, як екстремальні погодні умови впливають на врожайність. Застосовано оцінку впливу зміни клімату через регресійну модель погоди та днів вегетації (GDD) на врожайність сільськогосподарських культур із використанням багаторічних даних.

Автори [7] дослідили вплив зміни клімату на врожайність основних сільськогосподарських культур у гірських районах Непалу та визначили їх зв'язки на основі регресійної моделі між історичними кліматичними даними та даними про врожайність продовольчих культур. Методи нахилу Манна-Кендалла та Сена було використано для аналізу тенденцій і кількісного визначення взаємозв'язків. Подібну методика застосовано і виявлено подібні тенденції в умовах гірського Пакистану [8].

У роботі [9] проаналізовано параметри опадів та дані про врожайність з наступним аналізом урожайності за допомогою регресії з 3 компонентами опадів: річною кількістю, кумулятивною кількістю упродовж культурного циклу (листопад–червень) та кумулятивною кількістю опадів протягом весни [9].

Автори [10], використовуючи методи кластеризації, оцінили взаємозв'язок між кліматичними факторами та мінливістю врожайності в локальному масштабі для конкретних сортів і стадій росту. Показано чіткі взаємозв'язки в різних комбінаціях місця розташування та сорту, причому 6 — 46% просторово-часової мінливості врожаю залежали від впливу кліматичних факторів, а реакція культури на погоду була нелінійною і залежною від сорту.

Мета досліджень — розробити принципи застосування інноваційних статистичних методів аналізу експериментальних даних для оптимізації системи добору селекційних зразків на продуктивність і адаптивність в умовах зміни клімату.

Методика досліджень. Лабораторні та польові досліді, проведені на Уладово-Люлинецькій дослідно-селекційній станції Інституту біоенергетичних культур і цукрових буряків НААН, закладали в полі селекційно-насінницької сівозміни, де попередником була гречка.

Ґрунти — чорноземи типові глибокі малогумусні крупнопилувато-середньосуглинкові з умістом гумусу 3,72%. Реакція ґрунтового розчину близька до нейтральної, уміст легкогідролізованого азоту — 12,02, рухомого фосфору (за Чиріковим) — 19,4, рухомого калію — 10,4 мг/100 г ґрунту. Погодні умови вегетаційного періоду 2021–2023 рр. показано в табл. 1.

Конкурсне (станційне) сортовипробування закладалося в 4-разовій повторності. Розміщення номерів у повтореннях — системне. Облікова площа ділянки — 25 м². У 2021–2023 рр. у конкурсному сортовипробуванні висівали 35–40 селекційних номерів, зокрема вивчали 26 селекційних ліній, що належать до 22 гібридних комбінацій. Стандартами були сорти Алекс УЛ, Улус, Улюбленець та Юлій. Упродовж вегетації проводили фенологічні та біометричні спостереження, відзначали інтенсивність росту, стійкість рослин до полягання та хвороб, оцінювали смакові якості у фазі технічної стиглості.

Як зазначають автори [11], оцінка сортів за допомогою випробувань у різних середовищах (МЕТ) є важливим кроком у програмах селекції рослин. Одна з цілей цих оцінок — визначити взаємодію генотипу з навколишнім середовищем (GEI). Методика визначення впливу GEI

1. Погодні умови вегетаційних періодів (2021–2023 рр.)

Місяць	2021		2022		2023	
	Температура, С°	Опади, мм	Температура, С°	Опади, мм	Температура, С°	Опади, мм
Квітень	6,6	22,8	7,0	56,1	7,7	84,2
Травень	13,2	97,0	13,6	53,6	15,2	4,5
Червень	19,4	78,2	19,4	57,6	18,6	96,9
Липень	22,3	73,4	19,0	56,1	20,2	93,4

на продуктивність сортів ґрунтується на аналізі адаптивності та стабільності. Дослідження, що базуються на лінійній регресії, мають обмеження, тому що переважно в схемах із статистичною незбалансованістю спостерігається неоднорідність залишкових дисперсій і генетичної коваріації. Альтернативою є використання моделей випадкової регресії (RRM), в яких параметри генотипів характеризуються як норма реакції з використанням багаторічних даних або повторних вимірювань з урахуванням інформації щодо коваріаційної функції. Застосовано використання MET на основі поліномів Лежандра та генотип ідеотипних відстаней. Подібні принципи виділення та аналізу ВГС запропоновано авторами [12–14].

Згідно з методикою [15], у регресії середня квадратична помилка (MSE), середня квадратична помилка (RMSE), середня абсолютна помилка (MAE) і коефіцієнт детермінації (R^2) є найпоширенішими оцінками помилок. MSE розраховується обчисленням середнього квадрата різниць між прогнозованими оцінками та справжніми значеннями. Нижча MSE свідчить про кращу відповідність між моделлю та даними. Рівняння для розрахунку MSE:

$$MSE = \sum_{i=1}^n (Y' - Y)^2 / n, \quad (1)$$

де Y' — прогнозоване значення, Y — спостережуване значення, n — загальна кількість спостережень. RMSE визначає, наскільки регресійна модель відповідає даним. Вона розраховується обчисленням середнього квадрата похибок між прогнозованими та фактичними значеннями, використовується як міра точності регресійної моделі й визначається за допомогою рівняння:

$$RMSE = \sqrt{\sum (Y' - Y)^2 / n}. \quad (2)$$

R^2 — вимірює, яка частина дисперсії в даних експерименту пояснюється моделлю. Він змінюється від 0 до 1, причому вищі значення свідчать про кращу відповідність, і розраховується за допомогою рівняння:

$$R^2 = SST - SSE / SST, \quad (3)$$

де SST — сума квадратів для загальної суми, SSE — сума квадратів для помилки. MAE є мірою середньої величини помилок у наборі прогнозів без урахування їхнього напрямку. Це середнє значення абсолютних відмінностей між прогнозом і фактичним спостереженням за тестовою вибіркою, де всі індивідуальні відмінності мають однакову нейромережеву вагу. MAE — найпоширеніший показник для регресійних моделей і хороший показник помилки загальної розраховування, який обчислюється за рівнянням:

$$MAE = \sum_{i=1}^n |Y'_i - Y_i| / n [15]. \quad (4)$$

Відповідно до нашого напрямку досліджень, помилки регресії в статистичному значенні стосовно біологічних явищ слід розглядати як аналог епігенетичного елемента (ВГС) у структурі фенотипової мінливості.

Результати досліджень. За результатами досліджень зернової продуктивності селекційних зразків гороху посівного в конкурсному сортовипробуванні 2021–2023 рр. виявлено значну мінливість статистичних характеристик варіаційних рядів за роками. Так, найвищі показники середньої міжпопуляційної мінливості (26,3 ц/га) зафіксовано у 2021 р., найнижчі (16,5 ц/га) — у 2023 (табл. 2).

2. Статистична характеристика варіаційного ряду зернової продуктивності селекційних зразків гороху посівного в конкурсному сортовипробуванні (2021–2023 рр.), ц/га

Розсадник. Статистика варіаційного ряду	Середнє	Мінімум	Максимум	Дисперсія	Стандартне відхилення	Стандартна похибка	Асиметрія	Стандартна похибка
Кс 2021	26,3	14,4	34,5	25,4	5,04	0,92	-0,71	0,43
Кс 2022	21,5	12,2	30,7	17,2	4,15	0,64	-0,21	0,37
Кс 2023	16,5	10,1	25,5	16,7	4,09	0,83	0,42	0,47

Відповідно до методики [15] використано набір із 3 випробувань, класифікованих як несприятливі (Кс 2023) або

сприятливі (Кс 2021, 2022) темпоральні середовища. Результати показали, що RRM дає змогу прогнозувати генотипові

3. Статистична характеристика варіаційного ряду зернової продуктивності селекційних зразків гороху посівного в конкурсному сортовипробуванні за множинною регресією R^2 (2021 – 2023 рр.), ц/га

Розсадник. Статистика варіаційного ряду	Множинна регресія R^2	SS	сс	MS	SS	сс	MS	F	p
Кс 2021	0,978	209,1	1,0	209,1	4,594	22,0	0,209	1001,2	0,0
Кс 2022	0,730	148,1	1	148,1	51,44	22,0	2,338	63,33	0,0
Кс 2023	0,980	376,3	1,0	376	8,27	22,0	0,38	1001,2	0,0

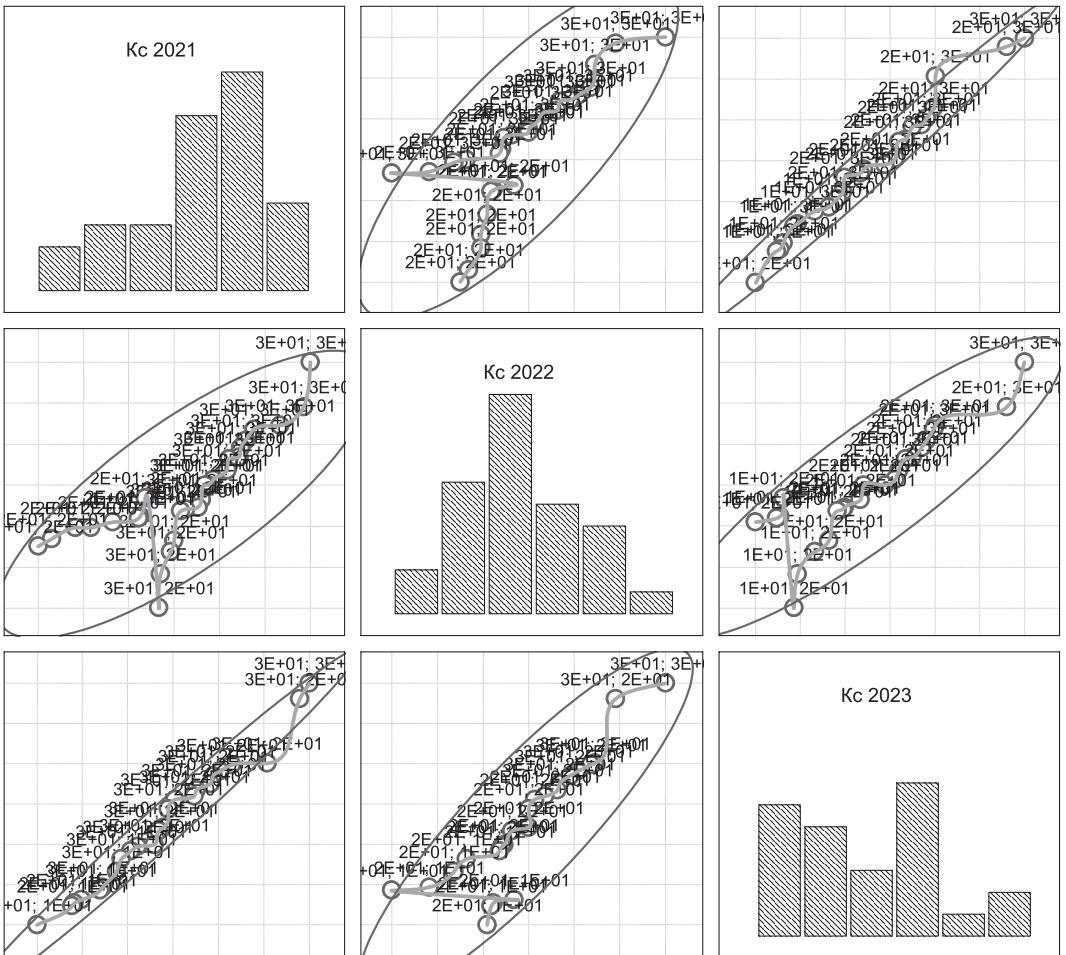


Рис. 1. Матриця розподілу варіаційного ряду зернової продуктивності селекційних зразків гороху посівного в конкурсному сортовипробуванні (2021 – 2023 рр.), ц/га

значення сортів на міжпопуляційному рівні в середовищах, де вони не були оцінені з високою точністю. За цими значеннями можна виміряти генотипову адаптованість за ідіотипами та їх нормами реакції [11]. Зокрема, за результатами множинної регресії у 2021–2023 рр. відзначено високий рівень R^2 , що свідчить про

відповідність експериментальних даних статистичній моделі (табл. 3).

Відповідно до розподілу селекційних номерів за продуктивністю в конкурсному сортовипробуванні, відзначено відсутність дискретності в поліноміальному ряді нормального розподілу 2022 р. та наявність біфуркацій у сплайнах 2021 і 2023 рр. Попри

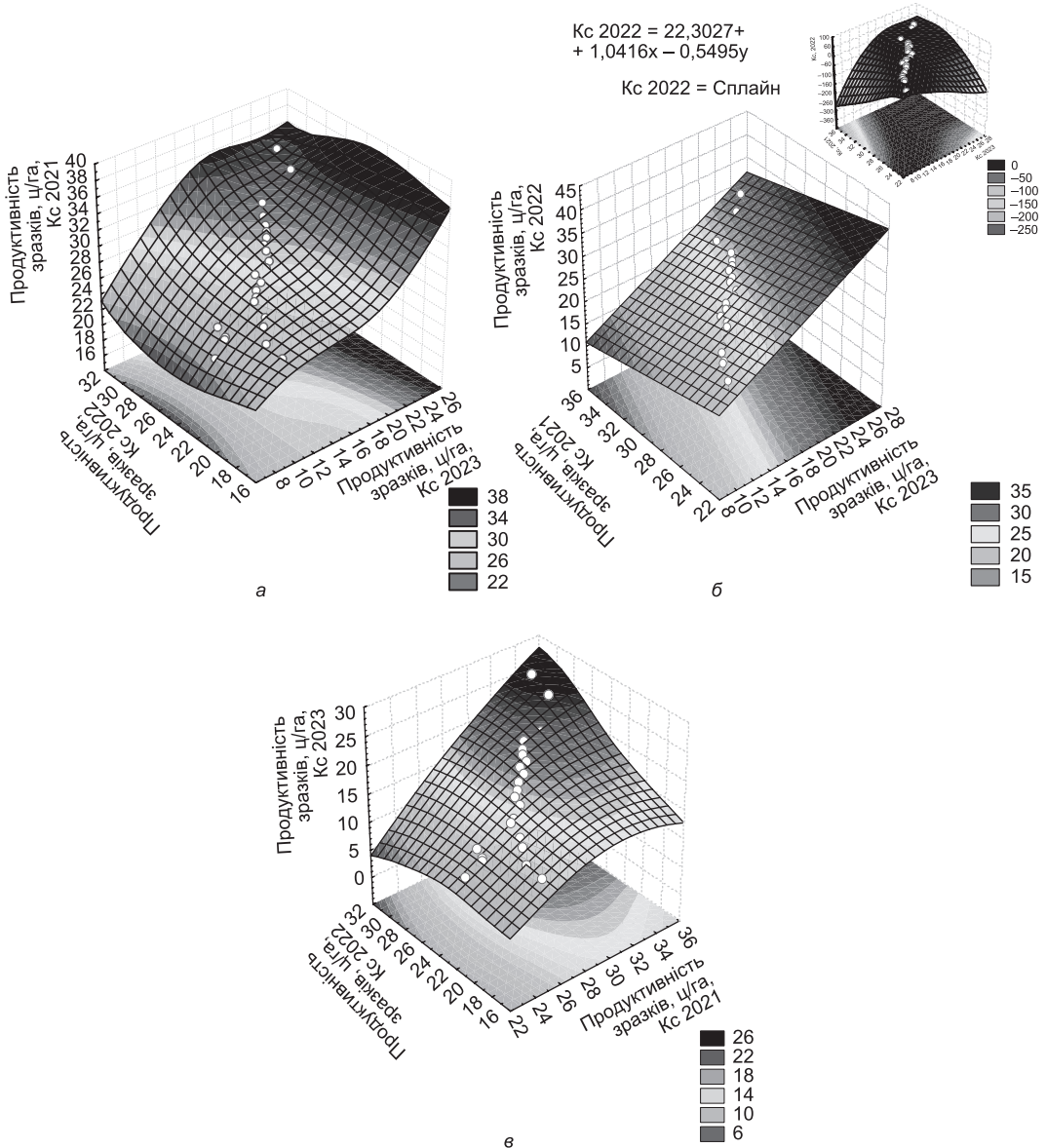


Рис. 2. Аналітична геометрична поверхня відгуку варіаційного ряду зернової продуктивності селекційних зразків гороху посівного в конкурсному сортовипробуванні, (2021–2023 рр.), ц/га

різниці в розподілі зразків у різні роки досліджень спостерігаються постійні робастні викиди 2 зразків із високими продуктивними адаптивними властивостями (рис. 1).

За продуктивними характеристиками на внутрішньо-популяційному рівні набору зразків у 2021–2023 рр. параметричні показники умов вегетації темпорального середовища забезпечили візуальну диференціацію зразків на комплексній поверхні відгуку. Зокрема, спостерігається біфуркація в зоні низької врожайності (10–14 ц/га) і робастні викиди продуктивності 2 зразків у зоні максимальних показників (26 ц/га). Тобто відзначається диференціальна здатність середовища на паралельну про-

дуктивність і адаптивність. За однакового набору генотипів на різницю рівнів зернової продуктивності значною мірою вплинула форма епігенотипової мінливості — ВГС. За результатами аналізу регресійних рівнянь, визначених за формами поверхонь відгуку продуктивності селекційних зразків у 2021–2023 рр., встановлено, що диференціальним і аналізуючим темпоральним середовищем є аналітична геометрична поверхня 2022 року. Плюс — мінус добір за комплексною і компонентними ознаками доцільно проводити на поверхнях 2021 і 2023 рр. відповідно до регресійного рівняння $K_s 2022 = 22,30 + 1,04x - 0, 54y$ (рис. 2, а, б, в).

Висновки

На міжпопуляційному рівні за роками досліджень встановлено сумарну дію і модульний взаємозв'язок між елементами генотипової та епігенотипової мінливості, яка проявляється під тиском стресових факторів зовнішнього середовища.

В умовах стресорної дії погодних факторів сумарний розмах фенотипової мінливості значно мінімізується, епігенотипової мінливості — максимізується відповідно до закону ентальпійного (продуктивного) мінімуму та ентропійного (адаптивного) максимуму на фоні стабільної генотипової мінливості порівняно з роками з оптимальними умовами

для росту і розвитку рослин. Причому тригерний вплив несприятливих факторів продовжується, навіть в умовах оптимального подальшого розвитку інших компенсаторних ознак.

Отже, добір на високу генотипову складову продуктивності та виявлення високого коефіцієнта успадкування доцільно проводити в роки з несприятливими умовами для розвитку ознаки (за аналізу поверхні відгуку як селекційне плато), оскільки її абсолютний фенотиповий «дельта-прояв» є мінімальним і визначається переважно генотиповою складовою.

Chernuskiy V.¹, Orlov S.², Klymchuk S.³
Institute of Bioenergy Crops and Sugar Beet;
e-mail: ¹vadimchernuskiy_58@ukr.net, ²orlov.stanislav48@gmail.com, ³uldss1888@ukr.net; ORCID:
¹0000-0002-8477-1050, ²0000-0001-5759-862X

Application of innovative methods of mathematical and statistical analysis in modern plant breeding technologies in Ukraine

Goal. To develop the principles of applying innovative statistical methods of experimental data analysis to optimize the system of selection of breeding samples for productivity and adaptability in conditions of climate change. **Methods.** Field (for the formation of methodologically verified experimental populations of breeding samples), instrumental-camera (for automating the structural analysis of plants), matrix-digital (for the formation

of a metadatabase), mathematical-statistical. **Results.** Genotype-environment interaction (GEI) is an important element of breeding technologies aimed at accelerating the pace of variety creation in climate change by optimizing the selection system. According to the results of studies of the grain productivity of selection samples of seed peas in the competitive variety test of 2021–2023, significant variability of the statistical characteristics of the variation series by year at a reliable level was revealed. The highest indicators of the average interpopulation variability (26.3 c/ha) were recorded in 2021, and the lowest (16.5 c/ha) in 2023. With the same set of genotypes, the difference in the levels of grain productivity was largely influenced by the form of epigenotypic variability — GEI. According to the results of the analysis of the

regression equations established by the shapes of the response surfaces of the productivity of the breeding samples in 2021–2023, it was proved that the differentiating and analyzing temporal environment was the analytical geometric surface of 2022. **Conclusions.** At the interpopulation level, over the years of research, the cumulative effect and modular relationship between the elements of genotypic and epigenotypic variability, which manifests itself under the influence of environmental stressors, was established. In the conditions of the stressful effect of weather factors, the total range of phenotypic variability was minimized, and epigenotypic variability was maximized under the law of enthalpy (productive) minimum and entropic (adaptive) maximum

against the background of stable genotypic variability compared to years with optimal conditions for plant growth and development. Therefore, selection at the intra-population level for a high genotypic component of productivity and detection of a high heritability coefficient is expedient to be carried out in years with unfavorable conditions for the development of the trait (according to the analysis of the response surface as a selection plateau), since its absolute “delta manifestation” is minimal and is determined mainly by the genotypic component.

Key words: peas, competitive variety testing, regression equations, analytical geometric response surfaces.

DOI: <https://doi.org/10.31073/agrovisnyk202406-07>

Бібліографія

1. Smith A., Ganesalingam A., Lisle C. et al. Use of Contemporary Groups in the Construction of Multi-Environment Trial Datasets for Selection in Plant Breeding Programs. *Front. Plant Sci.* 2021. 11:623586. doi: 10.3389/fpls.2020.623586
2. Ren H., Liu M., Zhang J. et al. Effects of agronomic traits and climatic factors on yield and yield stability of summer maize (*Zea mays* L.) in the Huang-Huai-Hai Plain in China. *Front Plant Sci.* 2022. № 10; 13:1050064. doi: 10.3389/fpls.2022.1050064
3. Bustos-Korts D., Romagosa I., Borrás-Gelonch G. et al. Genotype by Environment Interaction and Adaptation. In: Meyers, R. (eds) *Encyclopedia of Sustainability Science and Technology*. 2018. Springer. New York, NY. doi: 10.1007/978-1-4939-2493-6_199-3
4. Konduri V.S., Vandal T.J., Ganguly S., Ganguly A.R. Data Science for Weather Impacts on Crop Yield. *Front. Sustain. Food Syst.* 2020. 4:52. doi: 10.3389/fsufs.2020.00052
5. Phiri M., Shiferaw Y.A., Tesfamichael S.G. Biome-level relationships between vegetation indices and climate variables using time-series analysis of remotely-sensed data. *GIScience & Remote Sensing*. 2020. 57(4). P. 464–482. doi: 10.1080/15481603.2020.1733325
6. Ajetomobi, Joshua. Effects of weather extremes on crop yields in Nigeria. *African J. of Food, Agriculture, Nutrition and Development*. 2016/12/06. VL 16. P. 11168–11184. doi: 10.18697/ajfand.76.15685
7. Poudel S., Shaw R. The Relationships between Climate Variability and Crop Yield in a Mountainous Environment: A Case Study in Lamjung District, Nepal. *Climate*. 2016. 4. 13. doi: 10.3390/cli4010013
8. Zahid K.B., Ali G., Sabir S., Fayaz M. Relationship between climate and year to year variability in crop yield: selected districts of Pakistan. *Sarhad J. of Agriculture*. 2019. 35(1): 36–42. doi: 10.17582/journal.sja/2019/35.1.36.42
9. Chourghal N., Belguerri H., Maamri K. Diagnosis of the rainfall-wheat yield relationship in the current and future climate change conditions in Eastern Algeria BIOSYSTEMS DIVERSITY. 2023. V. 31. N 2. doi: 10.15421/012316
10. Delerce S., Dorado H., Grillon A. et al. Assessing Weather-Yield Relationships in Rice at Local Scale Using Data Mining Approaches. *PLoS ONE*. 2016 11(8): e0161620. doi: 10.1371/journal.pone.0161620
11. Michel Henriques de Souza, Jos Domingos Pereira Júnior, Skarlet De Marco Steckling et al. Adaptability and stability analyses of plants using random regression models. *Plos one*. 2020 Dec 2; 15 (12): e 0233200. doi: 10.1371/journal.pone.0233200. eCollection 2020.
12. Corrêa A.M., Teodoro P.E., Gonçalves M.C. et al. Artificial intelligence in the selection of common bean genotypes with high phenotypic stability. *Genet Mol Res*. 2016. Apr. 28; 15(2). doi: 10.4238/gmr.15028230
13. Dominic L. Waters, Julius H.J. van der Werf et al. Partitioning the forms of genotype-by-environment interaction in the reaction norm analysis of stability. *Theor Appl Genet*. 2023. Apr. 7; 136(5):99. doi: 10.1007/s00122-023-04319-9
14. Cooper M., Voss-Fels K.P., Messina C.D. et al. Tackling G × E × M interactions to close on-farm yield-gaps: creating novel pathways for crop improvement by predicting contributions of genetics and management to crop productivity. *Theor Appl Genet*. 2021; 134(6):1625–1644. doi: 10.1007/s00122-021-03812-3
15. Yoosefzadeh Najafabadi M., Hesami M., Eskandari M. Machine Learning-Assisted Approaches in Modernized Plant Breeding Programs. *Genes*. 2023. 14. 777. doi: 10.3390/genes14040777