



# Механізація, електрифікація

УДК 519.254

© 2024

## ОБҐРУНТУВАННЯ МЕТОДУ АПРОКСИМАЦІЇ ЕКСПЕРИМЕНТАЛЬНИХ ДАНИХ

*В.Т. Надикто*

*доктор технічних наук, професор, член-кореспондент НААН  
Таврійський державний агротехнологічний університет імені Дмитра Моторного  
вул. Жуковського, 66, м. Запоріжжя, 69063, Україна  
Національна академія аграрних наук України  
вул. Михайла Омеляновича-Павленка, 9, м. Київ, 01010, Україна  
e-mail: nadyktonvt@meta.ua  
ORCID: 0000-0002-1770-8297  
Надійшла 17.07.2024*

**Мета.** Провести порівняння лінійного і криволінійного методів апроксимації експериментальних даних з наданням рекомендацій щодо їхнього використання для забезпечення достовірності та адекватності отриманих результатів. **Методи.** Аналіз лінійного і криволінійного методів апроксимації експериментальних даних здійснено з використанням основ вищої математики, теорії математичної статистики випадкових процесів, регресійного та кореляційного аналізів. Розрахунки виконано у програмному середовищі Microsoft Excel версії 2016. **Результати.** Досягнення поставленої мети здійснено на прикладі аналізу залежності твердості ґрунту від його щільності. Для використаних експериментальних даних тісноту їх лінійного зв'язку оцінено коефіцієнтом кореляції, а криволінійного — кореляційним відношенням. Подальше порівняння цих статистичних показників через F-критерій показало, що фактичне значення цього критерію (1,52) є значно меншим за табличне (19,40). З урахуванням цього, попри менше значення коефіцієнта кореляції (0,77) у порівнянні з кореляційним відношенням (0,84), функціональна залежність між щільністю і твердістю ґрунту на статистичному рівні значущості 0,05 виявилася лінійною. **Висновки.** Отримання адекватного наукового результату при обробленні експериментальних даних залежить від правильного вибору кореляційного зв'язку між аргументом і функцією досліджуваного стохастичного процесу. Вид такого зв'язку (лінійний чи криволінійний) на заданому рівні ймовірності має базуватися на порівнянні коефіцієнта кореляції і кореляційного відношення дослідних даних через такий статистичний показник, як F-критерій. За меншого значення останнього у порівнянні з табличним кореляційний зв'язок між аргументом і функцією

**випадкового процесу приймається лінійним, у протилежному випадку — криволінійним.**

**Ключові слова:** лінійна залежність, криволінійна залежність, коефіцієнт кореляції, кореляційне відношення, F-критерій.

**DOI:** <https://doi.org/10.31073/agrovisnyk202409-05>

У сучасній науці широко використовують методи апроксимації для аналізу експериментальних даних і вивчення поведінки різноманітних динамічних і технологічних систем [1–5]. Серед них виділяють два основних підходи: лінійна та криволінійна апроксимації, кожна з яких має свої переваги та обмеження.

Лінійна апроксимація часто використовується для спрощення аналізу і базується на припущенні пропорційної залежності між змінними [6–8]. Така функція характеризується прямою лінією, для якої будь-які зміни аргументу ( $X$ ) за однакових його прирощень зумовлюють однакові кількісні зміни функції ( $Y$ ). Лінійна апроксимація дає можливість швидко отримати загальний вид системи та оцінити її основні тенденції. Однак, враховуючи, що багато фізичних явищ мають нелінійну природу, лінійна інтерпретація результату може призвести до втрати його значущості та точності.

Кількісною мірою тісноти лінійного зв'язку між аргументом  $X$  і функцією  $Y$  є безрозмірний коефіцієнт кореляції  $r$ , який змінюється у діапазоні від  $-1$  до  $+1$ . Частина науковців вважають, що значення цього параметра вже на рівні  $0,5$  є досить високим і таким, за якого  $50\%$  зміни функції зумовлені відповідною зміною аргументу. У дійсності це не так. Більш точно і адекватно ступінь зв'язку між величинами  $X$  і  $Y$  відтворює квадрат коефіцієнта кореляції  $r^2$ , який називають коефіцієнтом детермінації (Доспехов Б.А., 1985). З огляду на це, за значення коефіцієнта кореляції  $r = 0,5$  величина коефіцієнта детермінації  $r^2 = 0,25$  означає, що лише на  $25\%$  зміни функції зумовлені змінами аргументу. Решта ( $75\%$ ) варіації  $Y$  зумовлена зміною інших, не врахованих факторів.

Прийнято вважати, що за умови  $r < 0,3$  лінійна кореляція слабка, коли  $r = 0,3 - 0,7$  — середня, а при  $r > 0,7$  — сильна. Коли ж  $r = 0$ , то вважають, що між  $X$  і  $Y$  немає лінійного зв'язку, у той час, як криволінійний може цілком існувати.

Криволінійна апроксимація [9, 10] дозволяє враховувати більш складні залежності між змінними, надаючи більш точне відображення реальності. Цей метод зокрема ефективний у випадках, коли система проявляє нелінійну динаміку або взаємодіє з різними факторами.

Кількісним показником рівня тісноти нелінійного зв'язку між аргументом  $X$  і функцією  $Y$  є кореляційне відношення  $\eta$ . На відміну від коефіцієнта кореляції, показник  $\eta$  змінюється у діапазоні від  $0$  до  $1$ . При цьому, за умови чистого функціонального зв'язку між  $X$  і  $Y$ , кореляційне відношення  $\eta = 1$ . Якщо  $\eta = 0$ , то тоді кореляція між аргументом і функцією відсутня.

Проблема полягає в тому, що у багатьох практичних випадках дослідникові невідомий характер зв'язку між досліджуваними параметрами: лінійний чи криволінійний. Особливо це стосується тих процесів, де значення функції  $Y$  характеризуються коефіцієнтом варіації, більшим за  $20\%$ . За законами математичної статистики [11] саме таке значення цього показника характеризує випадковий процес як високо варіабельний.

Аналізуючи такі процеси за відсутності відповідного оцінювального механізму, науковці досить часто приймають необґрунтоване суб'єктивне рішення. Як показує практика, імовірність його неадекватності при цьому може бути досить високою. Особливо це стосується процесів із вираженим екстремумом (максимумом і/або мінімумом). Як правило,

його наявність потребує предметного пояснення. В іншому випадку помилковий вибір методу апроксимації експериментальних даних може призвести до недостовірних наукових висновків.

**Мета досліджень** — порівняння лінійного і криволінійного методів апроксимації експериментальних даних з наданням рекомендацій щодо їхнього використання для забезпечення достовірності та адекватності отриманих результатів.

**Методи досліджень.** Аналіз лінійного і криволінійного методів апроксимації експериментальних даних здійснено з використанням основ вищої математики, теорії математичної статистики випадкових процесів, регресійного та кореляційного аналізів. Розрахунки здійснено у програмному середовищі Microsoft Excel версії 2016.

**Результати досліджень.** Для прикладу розглянемо такий важливий у аграрних технологіях процес, як залежність твердості ґрунту від його щільності. Однозначно визначеного тлумачення з цього питання нині не існує. Одні науковці відстоюють лінійний характер зв'язку між цими параметрами [12, 13]. За незмінної вологості ґрунту такий результат, на наш погляд, є цілком логічним. Адже від більш щільного середовища природно очікувати і більшу твердість. Інші дослідники стверджують, що цей зв'язок може і навіть має бути криволінійним [14].

Для подальшого аналізу розглянемо наступну графічну інтерпретацію зв'язку між щільністю і твердістю ґрунту (рис. 1). На перший погляд, криволінійна апроксимація залежності між цими параметрами ґрунтового середовища є цілком задовільною.

Спробуємо оцінити достовірність такого результату. Для цього розрахуємо для даних рис. 1 коефіцієнт кореляції ( $r$ ) і кореляційне відношення ( $\eta$ ). Аргументом ( $X$ ) у цьому процесі є щільність ґрунту, а функцією ( $Y$ ) — його твердість.

На визначенні показника  $r$  зупинятися не будемо, тому що у програмному середовищі Microsoft Excel ця процедура є досить формалізованою, а тому не становить будь-яких труднощів.

Для розрахунку кореляційного відношення  $\eta$  масив  $Y$  (рис. 1) із 12 експериментальних даних згідно із рекомендаціями (Доспехов Б.А., 1985) розіб'ємо на 4 групи ( $k = 4$ ) по 3 значення у кожній. Потім знайдемо середнє значення ( $\bar{y}$ ) усього масиву і середні значення ( $\bar{y}_x$ ) для кожної із 4-х груп функції  $Y$ . Величину кореляційного відношення розрахуємо із використанням такого виразу:

$$\eta = \sqrt{\frac{\sum(Y - \bar{y})^2 - \sum(Y - \bar{y}_x)^2}{\sum(Y - \bar{y})^2}}. \quad (1)$$

Після визначення значень статистичних показників  $r$  і  $\eta$  проведемо їх

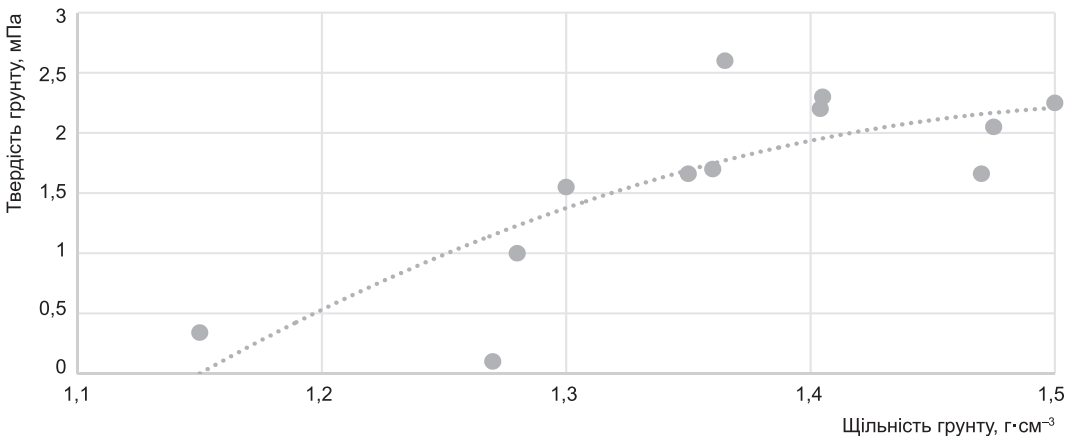


Рис. 1. Прогнозований характер залежності твердості ґрунту від його щільності

|    | A        | B    | C | D      | E   | F   | G    | H            | I     | J                 |
|----|----------|------|---|--------|-----|-----|------|--------------|-------|-------------------|
| 1  | 1,15     | 0,34 |   |        |     | 1,6 |      |              |       |                   |
| 2  | 1,27     | 0,1  |   | Ср_1 = | 0,5 | 2,3 | 1,29 |              |       |                   |
| 3  | 1,28     | 1    |   |        |     | 0,4 |      |              |       |                   |
| 4  | 1,3      | 1,55 |   |        |     | 0,0 |      |              |       |                   |
| 5  | 1,35     | 1,66 |   | Ср_2 = | 1,6 | 0,0 | 0,00 |              |       |                   |
| 6  | 1,36     | 1,7  |   |        |     | 0,0 |      |              |       |                   |
| 7  | 1,365    | 2,6  |   |        |     | 1,0 |      |              |       |                   |
| 8  | 1,404    | 2,2  |   | Ср_3 = | 2,4 | 0,3 | 0,56 |              |       |                   |
| 9  | 1,405    | 2,3  |   |        |     | 0,5 |      |              |       |                   |
| 10 | 1,47     | 1,66 |   |        |     | 0,0 |      |              |       |                   |
| 11 | 1,475    | 2,05 |   | Ср_4 = | 2,0 | 0,2 | 0,14 |              |       |                   |
| 12 | 1,5      | 2,25 |   |        |     | 0,4 |      |              |       |                   |
| 13 |          |      |   |        |     |     |      |              |       |                   |
| 14 | Середнє: | 1,6  |   |        |     | 6,7 | 1,99 | Кор. відн. = | 0,84  |                   |
| 15 |          |      |   |        |     |     |      | Коеф. кор. = | 0,77  |                   |
| 16 |          |      |   |        |     |     |      | Ффакт =      | 1,52  |                   |
| 17 |          |      |   |        |     |     |      | Фтабл =      | 19,40 | Кореляція лінійна |

Рис. 2. Фрагмент Excel-таблиці при визначенні характеру кореляційного зв'язку між дослідними даними

порівняння за допомогою спеціального F-критерію:

$$F_{\text{факт}} = \frac{(\eta^2 - r^2) \cdot (n - k)}{(1 - \eta^2) \cdot (k - 2)}, \quad (2)$$

де  $n = 12$  — масив значень твердості ґрунту (функція  $Y$ , рис. 1);  $k = 4$  — кількість груп значень функції  $Y$ .

Отримане із виразу (2) розрахункове (фактичне) значення цього статистичного критерію порівнюємо із табличним ( $F_{\text{табл}}$ ). Методику визначення

останнього з урахуванням прийнятого рівня статистичної значущості додано докладно викладено у джерелі (Доспехов Б.А., 1985). За умови  $F_{\text{факт}} < F_{\text{табл}}$  зв'язок між  $X$  і  $Y$  вважається лінійним, в іншому випадку — криволінійним.

Розрахункова частина цього процесу, здійснена у програмному середовищі Microsoft Excel, представлена на рис. 2.

Аналіз показує, що у кількісному вираженні кореляційне відношення  $\eta = 0,84$  більше за значення коефіцієнта кореляції  $r = 0,77$ . Водночас за F-критерієм,

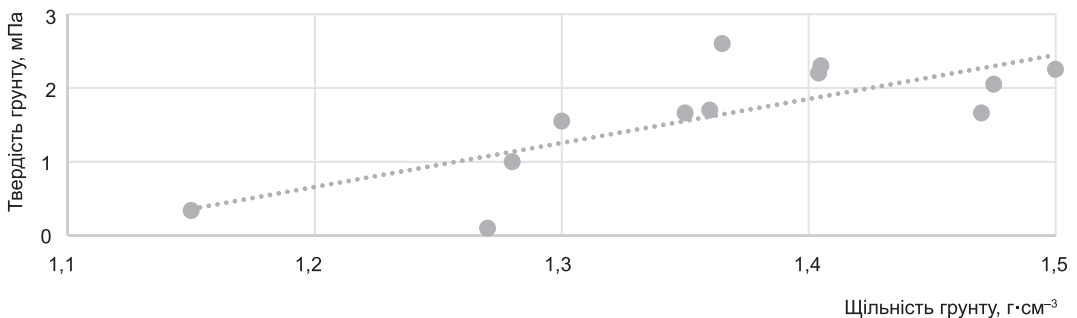


Рис. 3. Реальний характер залежності твердості ґрунту від його щільності

|    | A        | B    | C | D     | E   | F   | G    | H            | I     | J                      |
|----|----------|------|---|-------|-----|-----|------|--------------|-------|------------------------|
| 1  | 1,15     | 0,34 |   |       |     | 2,4 |      |              |       |                        |
| 2  | 1,27     | 2,1  |   | Ср_1= | 1,6 | 0,0 | 0,10 |              |       |                        |
| 3  | 1,28     | 2,3  |   |       |     | 0,2 |      |              |       |                        |
| 4  | 1,3      | 1,55 |   |       |     | 0,1 |      |              |       |                        |
| 5  | 1,35     | 1,66 |   | Ср_2= | 1,6 | 0,1 | 0,07 |              |       |                        |
| 6  | 1,36     | 1,7  |   |       |     | 0,0 |      |              |       |                        |
| 7  | 1,365    | 2,6  |   |       |     | 0,5 |      |              |       |                        |
| 8  | 1,404    | 2,2  |   | Ср_3= | 2,4 | 0,1 | 0,22 |              |       |                        |
| 9  | 1,405    | 2,3  |   |       |     | 0,2 |      |              |       |                        |
| 10 | 1,47     | 1,66 |   |       |     | 0,1 |      |              |       |                        |
| 11 | 1,475    | 2,05 |   | Ср_4= | 2,0 | 0,0 | 0,01 |              |       |                        |
| 12 | 1,5      | 2,25 |   |       |     | 0,1 |      |              |       |                        |
| 13 |          |      |   |       |     |     |      |              |       |                        |
| 14 | Середнє: | 1,9  |   |       |     | 3,8 | 0,40 | Кор. відн. = | 0,95  |                        |
| 15 |          |      |   |       |     |     |      | Коеф.кор.=   | 0,58  |                        |
| 16 |          |      |   |       |     |     |      | Fфакт =      | 21,47 |                        |
| 17 |          |      |   |       |     |     |      | Fтабл=       | 19,40 | Кореляція криволінійна |

Рис. 4. Фрагмент Excel-таблиці при визначенні характеру кореляційного зв'язку між скоорованими дослідними даними

розрахованим з довірчою імовірністю 95%, більш значимим є саме лінійний зв'язок між аргументом X і функцією Y. Це випливає із того, що

$$F_{\text{факт}} = 1,52 < F_{\text{табл}} = 19,40.$$

З огляду на це, представлена на рис. 1 залежність між щільністю ґрунту і його твердістю повинна мати лінійну інтерпретацію (рис. 3).

Для кращого розуміння питання зміні значення експериментальних даних функції Y на такі, які відображено у клітинках B2 і B3 Excel-таблиці (рис. 4).

У цьому випадку фактичне ( $F_{\text{факт}}$ ) значення F-критерію (21,47) більше за табличне ( $F_{\text{табл}}$ ), яке дорівнює 19,40. З огляду на це залежність між щільністю і твердістю ґрунту слід вважати криволінійною. У подальшому завдання полягає у виборі

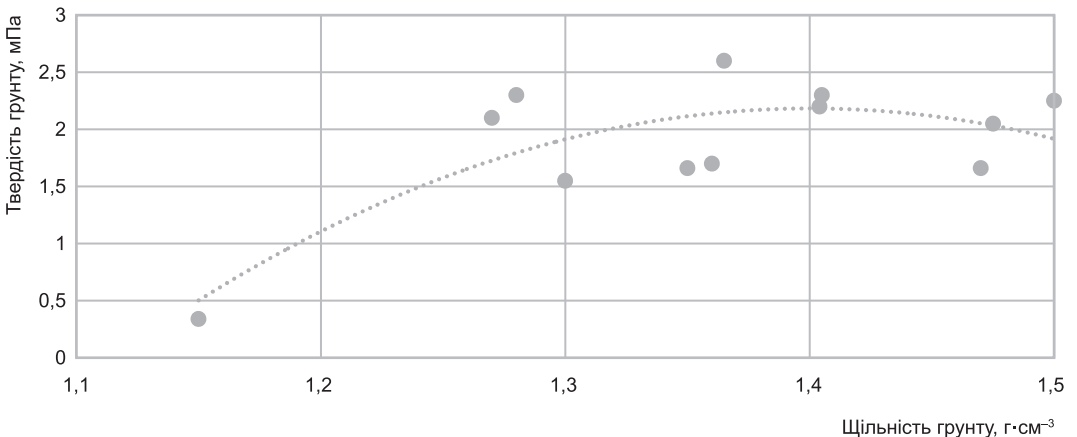


Рис. 5. Характер залежності твердості ґрунту від його щільності для скоригованих дослідних даних

такого математичного рівняння, яке б найбільш адекватно описувало природу цієї залежності.

Спробуємо, для прикладу, зв'язок між експериментальними даними, представленними на рис. 4, виразити рівнянням полінома другого порядку. У цьому випадку отримуємо (рис. 5):

$$Y = -26,80 \cdot X^2 + 75,07 \cdot X - 50,38. \quad (3)$$

При цьому слід зазначити, що рівняння (3), як результат процесу інтерполяції експериментальних даних, справедливе тільки в межах зміни щільності ґрунту від 1,1 до 1,5 г·см<sup>-3</sup>, тобто в межах отриманого масиву даних. Ця обставина має завжди доводитися до читача, у той час як автори далеко не завжди це роблять.

Вибравши криволінійну функцію для екстраполяції результатів, отримали таку цікаву обставину: за позначки 1,4 г·см<sup>-3</sup>

функція (3) має екстремум (див. рис. 5). Згідно з основами фізики ґрунту зростання його щільності зумовлює відповідне зростання твердості. Цей факт, як уже зазначалося вище, є цілком логічним, а тому будь-яких сумнівів не викликає. Але як і чим пояснити, що за щільності того самого ґрунту, більшої за 1,4 г·см<sup>-3</sup>, його твердість має тенденцію до зменшення?

Справа в тому, що нелогічність такого результату зумовлена саме некоректним вибором вихідних даних, і як наслідок — виду функціональної залежності (рівняння регресії), яка характеризує криволінійний зв'язок між щільністю і твердістю ґрунту. Тому у підсумку важливо: 1) правильно визначити характер зв'язку між аргументом і функцією; 2) у разі криволінійної залежності між ними вибрати коректне рівняння регресії.

## Висновки

*Отримання адекватного наукового результату при обробленні експериментальних даних залежить від правильного вибору кореляційного зв'язку між аргументом і функцією досліджуваного стохастичного процесу. Вид такого зв'язку (лінійний чи криволінійний) на заданому рівні імовірності має базуватися на порівнянні коефіцієнта кореляції і кореляційного відношення дослідних даних через F-критерій. За меншого значення*

*останнього у порівнянні з табличним кореляційний зв'язок між аргументом і функцією випадкового процесу приймається лінійним, у протилежному випадку — криволінійним.*

*Вибір коректного типу криволінійної регресії дослідних даних має здійснюватися шляхом їх інтерполяції таким математичним рівнянням, яке не суперечить внутрішній природі досліджуваного явища і є логічно обґрунтованим.*

### Nadykto V.

*Dmytro Motorny Tavria State Agrotechnological University, 66 Zhukovskiy Str., Zaporizhzhia, Ukraine, 69063; e-mail: nadyktonvt@meta.ua; ORCID: 0000-0002-1770-8297*

### **Justification of the method of approximation of experimental data**

**Goal.** To carry out a comparison of linear and curvilinear methods of approximation of experimental data with the provision of recommendations for their use to ensure the reliability and adequacy of the obtained results. **Methods.** The analysis of linear and curvilinear methods of approximation of experimental data was carried out using the foundations of higher mathematics,

the theory of mathematical statistics of random processes, regression and correlation analyses. Calculations were performed in the Microsoft Excel software environment version 2016.

**Results.** The goal was achieved using the example of the analysis of the dependence of soil hardness on its density. For the used experimental data, the tightness of their linear connection was estimated by the correlation coefficient, and the curvilinear one by the correlation ratio. Further comparison of these statistical indicators using the F-criterion showed that the actual value of this criterion (1.52) was significantly lower than the table value (19.40). With this in mind, despite the lower value of the correlation coefficient (0.77) compared to the correlation ratio (0.84),

the functional relationship between density and soil hardness at the statistical significance level of 0.05 was linear. **Conclusions.** Obtaining an adequate scientific result when processing experimental data depends on the correct choice of the correlation between the argument and the function of the studied stochastic process. The type of such a relationship (linear or curvilinear) at a given level of probability should be based on a comparison of the correlation coefficient and

the correlation ratio of experimental data through such a statistical indicator as the F-criterion. For a smaller value of the latter compared to the table, the correlation between the argument and the function of the random process is assumed to be linear, in the opposite case - curvilinear.

**Key words:** linear dependence, curvilinear dependence, correlation coefficient, correlation relation, F-criterion.

DOI: <https://doi.org/10.31073/agrovisnyk202409-05>

## Бібліографія

1. Cohen A., D'Ales J.P. Nonlinear approximation of random functions. *SIAM J. of Applied Mathematics*. 1997. V. 57. Is. 2. P. 518–540.
2. Temlyakov V.N. Nonlinear Methods of Approximation. *Foundations Computational Mathematics*. 2003. V. 3. P. 33–107.
3. Мойсишин В.М., Швасютин А.П., Процюк В.Р. та ін. Удосконалена методика обробки результатів факторного експерименту. *Прикарпатський вісник НТШ*. 2022. № 17(64). С. 75–95.
4. Busby D., Farmer C.L., Iske A. Hierarchical nonlinear approximation for experimental design and statistical data fitting. *SIAM J. of Applied Mathematics*. 2007. V. 29. Is. 1. P. 49–69.
5. Чернуський В.В., Чернуська Т.А. Принципи і методологія обробки експериментальних даних відповідно до концепції комплексної селекції традиційних культур Полісся на продуктивність і адаптивність. *Агропромислове виробництво Полісся*. 2015. № 8. С. 28–35.
6. Reinert D.J., Suzuki L.E.A.S., Reichert J.M. Interrelation of mechanical and physical soil properties of six South Brazilian soils under no-tillage. *International Soil Tillage Research Organisation 17th Triennial Conference*. Kiel, Germany. 2006. P. 1598–1602.
7. Al-Janobi A., Al-Hamed S., Aboukarima A., Almajhadi Y. Modeling of Draft and Energy Requirements of a Moldboard Plow Using Artificial Neural Networks Based on Two Novel Variables. *Engenharia Agricola*. 2020. V. 40. Is. 3. P. 363–373.
8. Tatiček M., Bauer F., Sedlak P. et al. The effect of setup of three point linkage on energetic and performance parameters of tractor aggregate. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*. 2011. V. 59. Is. 5. P. 253–262.
9. Adee E., Hansen F.D., Ruiz Diaz D.A. et al. Corn response as affected by planting distance from the center of strip-till fertilized rows. *Frontiers in Plant Sciences*. 2016. V. 7. P. 1232.
10. Xu Q., Liu H., Li M. et al. The Presence of the Biochar Interlayer Effectively Inhibits Soil Water Evaporation and Salt Migration to the Soil Surface. *Agriculture*. 2023. V. 13. Is. 3. P. 638.
11. Witte R.S., Witte J.S. *Statistics*. John Wiley & Sons, 2017.
12. Kumar A., Chen Y., Sadek M.A.A. et al. Soil cone index in relation to soil texture, moisture content, and bulk density for no-tillage and conventional tillage. *Research in Agricultural Engineering: CIGR J*. 2012. V. 14. Is. 1. P. 26–37.
13. Sayedahmed A. Modeling and Correlation of Soil Cone Index for Bulk Density, Moisture Content and Penetration Depth Levels in a Sandy Loam Soil. *J. of Soil Sciences and Agricultural Engineering*. 2015. V. 6. Is. 2. P. 259–273.
14. Медведев В.В. Твердость почв. Харьков: КП «Городская типография», 2009. 152 с.