

doi: 10.33249/2663-2144-2019-84-11-51-60

UDC [634.23:551.58](477.7)

REGRESSION ANALYSIS OF THE DEPENDENCE OF THE CHERRY YIELD FROM HYDROTHERMAL FACTORS IN THE CONDITIONS OF MULTICOLLINEARITY**V. Malkina, I. Ivanova, M. Serdiuk, I. Kryvonos, E. Bilous***e-mail: irynaivanova2017@gmasl.com*Dmytro Motorny Tavria State Agrotechnological University
18, B. Khmelnytsky Ave, Melitopol, Zaporizhzhia obl., 72312, Ukraine

In the conditions of the Southern Steppe zone of Ukraine the influence of weather factors of the region on the formation of cherry yield within the limits of 2007–2019 years of researches is revealed. The correlation analysis allowed us to identify ten weather factors that have a notable (significant) and strong, both direct and inverse linear correlation with cherry yield ($r = 0.68... -0.86$).

The model describing the influence of hydrothermal factors on cherry yield is proposed. Multicollinearity effects were detected when analyzing data using paired correlation coefficients and VIF. In this case, using the least squares method to construct and analyze the regression model is inefficient. A regression model based on the LASSO method is proposed. The LASSO method allows reliable estimation of regression parameters. On the basis of the constructed model the factors influencing the cherry yield index are analyzed and it is shown that the most significant factor is the average monthly rainfall for August, mm, then the factor is the absolute minimum relative humidity in May, %, the amount of rainfall during the flowering period, mm, the amount of the effective flowering temperatures, °C, the difference between average maximum and minimum flowering temperatures, °C, the total number of days with precipitation during flowering period, day, the average of maximum air temperatures during flowering period, the amount of active temperatures during flowering period, °C, the amount of active temperatures during the growing season (before the fruit ripening phase), °C, the hydrothermal coefficient (HTC) during flowering period. The indices of the share of influence of each factor on the total variance of the cherry yield index are determined. The share of influence of the factor is $x_1 - 9.90\%$, factor $x_2 - 5.56\%$, factor $x_3 - 1.95\%$, factor $x_4 - 8.22\%$, factor $x_5 - 14.83\%$, factor $x_6 - 10.12\%$, factor $x_7 - 12.12\%$, factor $x_8 - -3.81\%$, factor $x_9 - 27.96\%$, factor $x_{10} - 5.54\%$.

Key words: *cherry yield, weather factors, regression analysis, multicollinearity, LASSO method.*

РЕГРЕСІЙНИЙ АНАЛІЗ ЗАЛЕЖНОСТІ УРОЖАЙНОСТІ ВИШНІ ВІД ГІДРОТЕРМІЧНИХ ФАКТОРІВ В УМОВАХ МУЛЬТИКОЛІНЕАРНОСТІ**В. М. Малкіна, І. Є. Іванова, М. Є. Сердюк, І. А. Кривонос, Е. С. Білоус***e-mail: irynaivanova2017@gmasl.com*Таврійський державний агротехнологічний університет імені Дмитра Моторного
пр. Б. Хмельницького, 18, м. Мелітополь, Запорізька обл., 72312, Україна

В умовах Південного Степу України виявлено вплив погодних факторів регіону на формування показників врожайності вишні в межах 2007–2019 років досліджень. Проведення кореляційного аналізу дозволило визначити десять погодних факторів, що мають помітний (значний) та сильний як прямий, так і обернений лінійний кореляційний зв'язки з урожайністю вишні ($r = 0,68...0,86$).

Запропоновано модель, яка описує вплив гідротермічних факторів на врожайність вишні. При аналізі даних з допомогою парних коефіцієнтів кореляції і показника VIF був виявлений ефект мультиколінеарності. У цьому випадку використовувати метод найменших квадратів для побудови і аналізу регресійної моделі є неефективним. Запропонована регресійна модель, яка побудована на основі методу LASSO. Метод LASSO дозволяє побудувати достовірні оцінки параметрів регресії. На основі побудованої моделі проаналізовано чинники, що впливають на показник врожайності вишні і показано, що найбільш значущими є такі фактори: середньомісячна сума опадів за серпень, мм; абсолютна мінімальна відносна вологість повітря в травні, %; сума опадів у період цвітіння, мм; сума ефективних температур у період цвітіння, °C; різниця між середньою максимальною та

мінімальною температурами повітря в період цвітіння, °С; загальна кількість днів з опадами в період цвітіння, доба; середня з максимальних значень температур повітря в період цвітіння; сума активних температур в період цвітіння, °С; сума активних температур за вегетаційний період (до фази досягання плодів), °С; гідротермічний коефіцієнт (ГТК) в період цвітіння. Визначено показники частки впливу кожного фактора на загальну дисперсію показника врожайності вишні. Доля впливу фактору складає $x_1 - 9,90\%$, фактору $x_2 - 5,56\%$, фактору $x_3 - 1,95\%$, фактору $x_4 - 8,22\%$, фактору $x_5 - 14,83\%$, фактору $x_6 - 10,12\%$, фактору $x_7 - 12,12\%$, фактору $x_8 - 3,81\%$, фактору $x_9 - 27,96\%$, фактору $x_{10} - 5,54\%$.

Ключові слова: врожайність вишні, погодні фактори, регресійний аналіз, мультиколінеарність, метод LASSO.

Вступ

Вишня – одна з найбільш популярних культур у населення, завдяки своїй скоростиглості та невибагливості до умов навколишнього середовища. Найморозостійкіша кісточкова плодова порода, що за зазначеним показником поступається тільки яблуні. Культура не вимоглива до ґрунтових умов і тому набула поширення по всій території України (Bublyk, 2005; Vasylyshyna, 2019).

Найбільші площі вишневих садів у Донецькій, Вінницькій і Дніпропетровській областях – близько четвертої частини усіх площ, значна їх кількість знаходиться у Луганській, Київській, Черкаській, Запорізькій та Полтавській областях – більше третини площ (Bohdaniuk, 2016).

В останнє десятиліття до Реєстру включено низку нових високопродуктивних сортів вишні селекції станції зрощуваного садівництва УААН – Шалунья, Примітна, Ожиданіє, Ігрушка, Встреча, селекції Мліївського інституту садівництва УААН – Альфа, іноземної селекції – Норд Стар. Сорти вишні та вишнево-черешневих гібридів, що створені селекціонерами півдня Степової зони України М.І. Туровцевим і В.О. Туровцевою, відрізняються високою продуктивністю та стійкістю до несприятливих факторів. Однак найповніше ці сорти можуть реалізувати свої кращі ознаки тільки при розміщенні їх в найбільш сприятливих зонах вирощування (Turovceva et al., 2013).

Зміни клімату можуть мати істотні наслідки, які вплинуть на врожайність та якість плодів культур. Тому, актуальним є питання дослідження впливу метеорологічних умов на врожайність провідних плодів культур Півдня Степової зони України. Завдання науки в цих умовах зводиться до того, щоб підготувати відповідну інформацію для прийняття рішень (Semenov et al., 2000; Vasylyshyna & Postolenko, 2019).

Останнім часом в світовій аграрній науці, у зв'язку зі значним зниженням врожаю сільськогосподарських (в тому числі і плодівих) культур, під дією погодних умов у провідних країнах склалася чітка думка про те, що в найближчий час прогрес в аграрному виробництві буде відбуватися не стільки від розвитку біології та техніки, а завдяки удосконаленню методів отримання погодних даних, вивчення впливу погодних факторів на продуктивність порід, технологій розміщення цих порід в регіонах, де вони здатні максимально реалізувати свій біологічний потенціал (Ivashenko, 2019). Тому адаптація плодівих культур, зокрема вишні, до умов довкілля також вимагає розробки нових методик та створення моделей продуктивності цієї культури на основі погодних факторів.

Ефективним інструментом для виявлення впливу на урожайність вишні різних природних факторів є аналіз регресійної моделі, побудованої на основі зібраних експериментальних спостережень.

Як відомо, для отримання незміщених ефективних оцінок параметрів регресійної моделі повинні виконуватися умови теореми Гауса-Маркова (Damodar, 2004).

Одна з умов теореми – вектори-фактори повинні бути незалежні, тобто фактори не повинні корелювати між собою. Порушення цієї умови призводить до прояву ефекту мультиколінеарності.

Ефект мультиколінеарності означає, що принаймні два чинники (незалежні змінні), які беруть участь в моделі, зв'язані тісною кореляційною залежністю між собою. Питанням складнощів і негативному впливу мультиколінеарності на весь процес дослідження присвячено багато публікацій (Aiken & West, 1991; Graham, 2003). Як описано в літературних джерелах, основна проблема при проявленні мультиколінеарності – нестабільні і зміщені

похибки, що означає велике значення показника p -values для оцінки статистичної значущості параметрів регресійної моделі. Як результат, це призводить до неефективних оцінок параметрів моделі.

Мультиколінеарність серед факторів робить практично неможливо адекватну інтерпретацію виявлення і оцінки впливу кожного регресора (фактора) на результуючу ознаку. Це відбувається через ефект перекриття інформації. Як відомо, коефіцієнти регресійної моделі використовують для інтерпретації показника зміни очікуваного значення результуючої ознаки через посилення ролі фактора на одиницю при незмінних значеннях інших змінних. При прояві ефекту мультиколінеарності така інтерпретація не дає адекватного результату.

Тому слід розглядати дві проблеми - як виявити ефект мультиколінеарності і як побудувати ефективну регресійну модель в умовах мультиколінеарності.

Відомо багато способів виявлення мультиколінеарності. Ознакою мультиколінеарності також є велике значення коефіцієнта детермінації, адекватність моделі (яка підтверджується критерієм Фішера) і, при цьому, велика кількість незначних параметрів моделі (що перевіряється за критерієм Стюдента). Одним з широко використовуваних методів виявлення мультиколінеарності є метод на основі показника variance inflation factor (VIF) (Yu-Kang et al., 2004; Kutner et al., 2004), який вимірює збільшення дисперсії в порівнянні з тим, яка б вона була в разі однофакторної регресії:

$$VIF = \frac{1}{1-R_j^2},$$

де R_j^2 – коефіцієнт детермінації регресії фактора X_j від інших факторів.

Вважається (Aiken & West, 1991; Graham, 2003), що якщо $VIF > 10$, то присутній ефект мультиколінеарності.

При побудові адекватної регресійної моделі в умовах мультиколінеарності найчастіше пропонується видалити з аналізу один з факторів, які корелюють. Однак, це призведе до втрати важливої інформації. Ми не зможемо на підставі такої моделі оцінити ступінь впливу на результуючу ознаку тих чинників, які ми видалили з розгляду. До того ж, в нашому випадку, корелюють між собою практично всі фактори, і видалення їх з моделі призведе до неефективності всього дослідження.

Для вирішення другої проблеми при проведенні досліджень в умовах виявленої мультиколінеарності пропонується використовувати методи регуляризації, які коригують відхилення від нормального розподілу залишків. Таким методом, наприклад, є метод LASSO (Hastie et al., 2015; James et al., 2013), який пропонується у роботі.

Як відомо, при визначенні параметрів регресійної моделі на основі МНК мінімізують функцію:

$$RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2,$$

де:

y_i – експериментальні значення регресанта,

\hat{y}_i – теоретичні значення регресанта, які розраховуються на основі побудованого рівняння регресії.

Згідно з методом LASSO, для визначення параметрів моделі знаходять мінімум функції:

$$L = RSS + \lambda \sum_{i=1}^n |\beta_i|,$$

де: λ – заданий параметр (штраф);

β_i – коефіцієнти регресійної моделі.

Іншими словами, вводиться штраф за невивпавдано великі значення параметрів β_i . Причому величина цього штрафу пропорційна величині параметра λ , за допомогою якого можна підібрати більш стійке рішення.

Суть алгоритму визначення параметра λ полягає в наступному. Дані випадковим чином розбивають на Q підвибірок. Одна з підвбірок є контрольною, інші $Q - 1$ в сукупності утворюють загальну вибірку. Визначається вектор значень $\lambda = \{\lambda_s\}$ з деяким кроком. Для λ_s за навчальною вибіркою будується регресійна модель. Для кожної моделі обчислюється помилка прогнозу, тобто сума квадратів залишків регресії:

$$RSS_{\lambda_s}^q = \sum_{i=1}^n (y_i - \sum_{j=0}^{k-1} \hat{b}_j(q, \lambda_s) x_{ij})^2,$$

де: $q = 1, Q$ – номер блока, обраного в якості вибірки;

y_i – експериментальні значення регресанта;

x_{ij} – експериментальні значення регресорів;

λ_s – параметр.

Потім обчислюють середнє значення помилки за всіма блоками:

$$MSE_{\lambda_s} = \frac{1}{Q} \sum_{q=1}^Q RSS_{\lambda_s}^q.$$

В якості оптимального λ обирається таке λ_s , за якого MSE_{λ_s} буде мінімальним.

Для оцінки ступеня впливу кожного фактора окремо на показник врожайності вишні знаходять для кожної із запропонованих моделей

(побудованих за значень параметра λ_{min} і λ_{1se}) коефіцієнт детермінації R^2 , коефіцієнти еластичності E_i , коефіцієнти β_i .

Коефіцієнт детермінації визначається за формулою:

$$R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2},$$

де:

\hat{y}_i – теоретичні значення врожайності, розраховані на підставі регресійної моделі,

y_i – експериментальні значення врожайності,

\bar{y} – середнє значення показника урожайності.

Коефіцієнт характеризує ступінь впливу виділених чинників на врожайність вишні в порівнянні з іншими неврахованими і випадковими факторами.

На підставі коефіцієнтів

$$\beta_i = b_i \frac{\bar{S}_{x_i}}{\bar{S}_y}$$

де b_i – розраховані коефіцієнти регресійної моделі,

\bar{S}_{x_i} – середні квадратичні відхилення факторів,

\bar{S}_y – середнє квадратичне відхилення показника y ,

впорядкуємо фактори за ступенем їх впливу на показник «Урожайність вишні».

Таким чином, встановлення пріоритетності основних параметрів погодних умов та побудова ефективної регресійної моделі в умовах мультиколінеарності з метою визначення найповнішого прояву біологічного потенціалу вишні, що вирощена в умовах Півдня степової зони України, є дуже актуальним питанням.

Матеріали та методи

Дослідження проводилися в 2007–2019 роках в межах Мелітопольського району Запорізької області. З метою дослідження впливу погодних чинників на врожайність вишні використано данні, що були надані Головним управлінням статистики в Запорізькій області та метеорологічні данні за період з 2007 по 2019 роки Мелітопольської метеостанції.

Метою досліджень було встановити пріоритетність об'єктивних агрокліматичних показників, що мають вплив на врожайність вишні в умовах Півдня Степової зони України та створити ефективну регресійну математичну модель врожайності культури на підставі виявлених стресових чинників.

Для досягнення мети нами було проведено кореляційний та регресійний аналізи: розрахована міцність кореляційних зв'язків між агрокліматичними показниками та врожайністю культури; визначено комплекс погодних факторів, що суттєво впливають на врожайність вишні; встановлено моделі залежності врожайності вишні від погодних умов, що склалися в Запорізькій області у 2007–2019 рр., виходячи з практичної доцільності визначена найбільш ефективна математична модель; отримано рівняння регресії, яке можна використовувати для побудови статистичних прогнозів.

Розрахунок моделей врожайності вишні у зв'язку з абіотичними погодними факторами проводили за наступною схемою (Ivanova et al., 2019; Bublyk, 2005):

1.Збір даних та створення комп'ютерної бази про середню врожайність вишні в умовах Мелітопольського району. Середню врожайність породи визначали в виробничих насадженнях, узагальнюючі врожайність сортименту регіону.

2.Створення комп'ютерної бази погодних умов у роки досліджень з відбором показників: температура (мінімальна, середня, максимальна), сума опадів, кількість днів з опадами більш одного міліметра, середня відносна вологість повітря.

3. Розрахунок показників: тривалість безморозного періоду, сума опадів, гідрометеорологічні коефіцієнти та перепад температур за певні періоди, суми активних температур та інші показники.

4. Визначення погодних факторів, що суттєво впливають на врожайність вишні.

5. Аналіз визначених в пункті 4 факторів методами регресійного аналізу з метою виявлення ефекту кореляції між факторами.

6. Побудова і аналіз регресійної моделі залежності врожайності вишні від погодних умов.

У період проведення досліджень було використано представлені нижче методи варіаційної статистики, що дозволили проаналізувати, обробити експериментальні данні та зробити прогноз остаточних результатів: проводили математичну обробку, парний та множинний кореляційний і регресійний аналізи – за загально-прийнятими методиками (Bublyk, 2005). Розраховано показник наявності ефекту мультиколінеарності – фактор інфляції дисперсії VIF. Також були використані комп'ютерні

програми Msoffice Excel, програмне середовище RStudio.

Результати досліджень та обговорення

За допомогою методів регресійно-кореляційного аналізу було отримано сільськогосподарську оцінку впливу погодних умов на врожайність вишні в період 2007–2019 рр. Врожайність – це складна ознака, яка залежить від багатьох чинників, тому проводили пошук кореляційних зв'язків між врожайністю вишні (Y) та комплексом гідротермічних умов (факторів) - x_i , де $i = 1..10$. Обрано 20 чинників (факторів), які можуть мати вплив на зміни врожайності вишні. Для визначення погодних факторів, що суттєво впливають на врожайність вишні в розрізі кожного року, проаналізовано 20 парних кореляційних залежностей на етапах: вегетаційний період, цвітіння, досягання плодів та в період знімання врожаю (Ivanova & Herasko, 2019; Ivanova & Malkina, 2019).

На рисунках 1–3 показані графіки зміни показників факторів зовнішнього середовища, які наведені за нормованою шкалою, де x_1 – сума активних температур за вегетаційний період (до фази досягання плодів), °C, x_2 – різниця між середньою максимальною та мінімальною температурами повітря в період цвітіння, °C, x_3 –

сума активних температур в період цвітіння, °C, x_4 – сума ефективних температур в період цвітіння, °C, x_5 – гідротермічний коефіцієнт (ГТК) в період цвітіння», x_6 – середньомісячна сума опадів за серпень, мм, x_7 – абсолютна мінімальна відносна вологість повітря в травні, %, x_8 – сума опадів у період цвітіння, мм, x_9 – загальна кількість днів з опадами в період цвітіння, доба, x_{10} – середня з максимальних значень температур повітря у період цвітіння.

Згідно з даними рисунку 1 максимальна врожайність вишні за роки досліджень в діапазоні 9,9...11,4 т/га спостерігалася на фоні значень суми активних температур вище 10°C (САТ 10) за вегетаційний період (до фази досягання плодів); 1391,6 °C ... 1499,1 °C ($r_{YX1} = 0,77$). Різниця між середньою максимальною та мінімальною температурами повітря впродовж 13 років становить від 10,7 °C до 18,6°C; значення r у досліджуваного гідротермічного параметру по відношенню до врожайності вишні становить $r_{YX2} = 0,68$. Для забезпечення максимальної врожайності черешні на рівні 2010, 2013 та 2018 років, показник – сума активних температур вище 10°C (САТ 10) у період цвітіння, повинен мати діапазон значень 209,3–261,4 °C, тенденція підтверджується значенням $r_{YX3} = 0,76$.



Рис. 1. Динаміка зміни показників, наведених у нормованій шкалі, де x_1 – сума активних температур за вегетаційний період (до фази досягання плодів), °C, x_2 – різниця між середньою максимальною та мінімальною температурами повітря в період цвітіння, °C, x_3 – сума активних температур у період цвітіння, °C, y – урожайність вишні

Аналіз даних рисунку 2 свідчить, що діапазон значень досліджуваного показника – суми ефективних температур вище 10°C у період цвітіння (СЕТ10) для отримання максимальної врожайності культури на рівні 2013 та 2018 років повинен становити – 91,4...92,6 °C ($r_{YX4} = 0,75$). Значення ГТК для 2010, 2013, 2018 років, в які відмічено найвищу врожайність вишні, дорівнює 0,0–0,1 ($r_{YX5} = -0,75$). Коливання середньомісячної суми опадів в серпні по роках досліджень

відбувається в діапазоні від 0 мм до 42,4 мм. Оптимальна врожайність досліджуваної культури визначена в роки з максимальним значенням суми опадів у серпні при значеннях показника 36,9–42,4 мм ($r_{YX6} = 0,84$). Абсолютна з мінімальних значень відносної вологості повітря в травні склала 13...14% у 2013 та 2018 роках, коли спостерігалася найвища врожайність 9,9...11,4 т/га ($r_{YX7} = -0,73$).

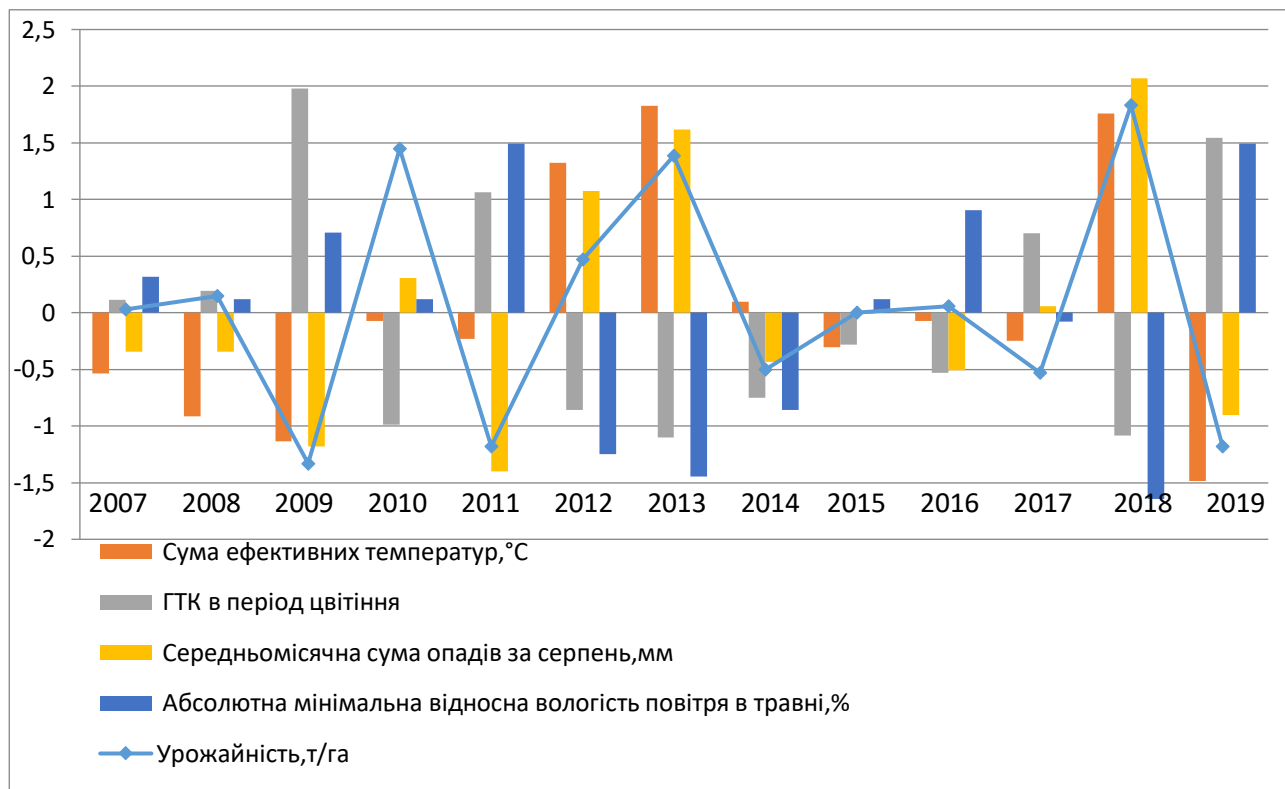


Рис. 2. Динаміка зміни показників, наведених у нормованій шкалі, де x_4 – сума ефективних температур у період цвітіння, °C, x_5 – гідротермічний коефіцієнт (ГТК) в період цвітіння, x_6 – середньомісячна сума опадів за серпень, мм, x_7 – абсолютна мінімальна відносна вологість повітря в травні, %, y – урожайність вишні

Аналіз динаміки гідротермічних факторів по роках досліджень показує, що максимальна врожайність досліджуваної культури (9,9...11,4 т/га) визначена в роки з мінімальним значенням суми опадів у період цвітіння на рівні значень 0,7 мм до 3,3 мм ($r_{YX8} = -0,77$). Загальна кількість днів з опадами у період цвітіння в інтервалі досліджуваних років зафіксована з мінімальною кількістю днів що склали 2 доби. Між

урожайністю та досліджуваним параметром визначено міцну зворотну кореляційну залежність при $r_{YX9} = -0,86$. Середня з максимальних значень температур повітря у період цвітіння в досліджуваному інтервалі років коливається в діапазоні 17,8...25,4 °C. У погодного фактора по відношенню до врожайності вишні визначений прямий міцний кореляційний зв'язок – $r_{YX10} = 0,68$;



Рис. 3. Динаміка зміни показників, наведених у нормованій шкалі, де x_8 – сума опадів в період цвітіння, мм, x_9 – загальна кількість днів з опадами в період цвітіння, доба, x_{10} – середня з максимальних значень температур повітря в період цвітіння, y – урожайність вишні

Процес нормування значень проводиться для приведення різнопланових показників до єдиної шкали вимірювань за формулою:

$$z_{ij} = \frac{x_{ij} - \bar{x}_i}{\bar{\sigma}_i},$$

де: x_{ij} – реальне значення фактору x_i в j -му спостереженні,

$$\bar{x}_i = \frac{\sum_{j=1}^n x_{ij}}{n} - \text{середнє значення фактору } x_i,$$

$$\bar{\sigma}_i - \text{дисперсія фактору } x_i.$$

В таблиці 1 наведені розраховані значення середніх \bar{x}_i і середньоквадратичних відхилень $\bar{\sigma}_i$ для кожного фактору.

Таблиця 1. Значення середніх \bar{x}_i і середньоквадратичних відхилень $\bar{\sigma}_i$

i	1	2	3	4	5	6	7	8	9	10
\bar{x}_i	1260,100	12,631	189,846	60,615	1,283	17,123	21,385	20,177	4,231	20,354
$\bar{\sigma}_i$	152,703	1,568	36,404	17,533	1,139	12,228	5,107	16,154	2,006	2,321

Для проведення кореляційного аналізу побудована матриця парних коефіцієнтів кореляції, яка наведена в таблиці 2. Аналіз тісноти кореляційного зв'язку для десяти погодних факторів (x_i) по відношенню до урожайності вишні (Y) проведено згідно зі шкалою англійського статистика Чеддока (Ivanova, 2019). Згідно з даними таблиці 1, для десяти погодних факторів (x_i , $i=1..10$) встановлено середню та сильну лінійну кореляційну залежність з урожайністю вишні ($r_{yx_i} = 0,68..-0,86$). Це такі фактори (x_i): сума активних температур за

вегетаційний період, до етапу досягання плодів (x_1), середньомісячна сума опадів за серпень (x_6), абсолютна мінімальна відносна вологість повітря в травні (x_7); у період цвітіння розглядали наступні гідротермічні параметри: різниця між середньою максимальною та мінімальною температурами повітря (x_2), сума активних температур (x_3), сума ефективних температур (x_4), гідротермічний коефіцієнт, ГТК (x_5), сума та загальна кількість днів з опадами (x_8 , x_9) та середня з максимальних значень температур повітря у період цвітіння (x_{10}).

Таблиця 2. Кореляційна матриця

	Y	x ₁	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₈	x ₉	x ₁₀
Y	1	0,769	0,679	0,759	0,757	-0,753	0,839	-0,728	-0,772	-0,864	0,677
x ₁	0,769	1	0,545	0,469	0,468	-0,518	0,607	-0,388	-0,618	-0,463	0,530
x ₂	0,679	0,545	1	0,425	0,706	-0,584	0,808	-0,676	-0,724	-0,477	0,719
x ₃	0,759	0,469	0,425	1	0,779	-0,794	0,590	-0,596	-0,635	-0,798	0,688
x ₄	0,757	0,468	0,706	0,779	1	-0,672	0,841	-0,845	-0,652	-0,765	0,907
x ₅	-0,753	-0,518	-0,584	-0,794	-0,672	1	-0,562	0,674	0,945	0,816	-0,742
x ₆	0,839	0,607	0,808	0,590	0,841	-0,562	1	-0,897	-0,644	-0,700	0,746
x ₇	-0,728	-0,388	-0,676	-0,596	-0,845	0,674	-0,897	1	0,722	0,787	-0,752
x ₈	-0,772	-0,618	-0,724	-0,635	-0,652	0,945	-0,644	0,722	1	0,742	-0,728
x ₉	-0,864	-0,463	-0,477	-0,798	-0,765	0,816	-0,700	0,787	0,742	1	-0,665
x ₁₀	0,677	0,530	0,719	0,688	0,907	-0,742	0,746	-0,752	-0,728	-0,665	1

Перед тим як будувати регресійну модель (Ivanova, 2019; Ivanova, 2019) проведено аналіз даних на наявність ефекту мультиколінеарності.

Практично всі парні коефіцієнти кореляції між факторами мають значення, близькі по модулю до одиниці, що вже говорить про високий рівень кореляційної зв'язку між ними.

Лінійна регресійна модель, побудована на основі МНК, має вид:

$$y = 10,22 - 0,003x_1 - 1,151x_2 - 0,017x_3 + 0,105x_4 + 0,049x_5 + 0,431x_6 + 0,945x_7 - 0,187x_8 - 0,758x_9 - 0,479x_{10}.$$

Аналіз моделі показав, що практично всі фактори при рівні значущості 0,05 незначимі, а значення коефіцієнта детермінації дорівнює 0,9958. При цьому, модель адекватна, згідно з критерієм Фішера при рівні значущості 0,05 (p-value: 0,02072). Це свідчить про наявність сильної кореляції факторів, тобто про невиконання однієї з умов теореми Гауса-Маркова (Graham, 2003). Значення показників VIF для кожного фактора наведені в таблиці 3.

Таблиця 3. Значення показника VIF

Фактори	x ₁	x ₂	x ₃	x ₄	x ₅	x ₆	x ₇	x ₈	x ₉	x ₁₀
VIF	15,14	8,66	10,67	46,32	45,99	54,48	17,41	10,78	42,67	11,45

Як видно, практично всі значення показника VIF перевищують значення 10, що говорить про сильну кореляцію факторів між собою. Іншими словами, присутній ефект мультиколінеарності.

Для побудови регресійної моделі в умовах мультиколінеарності було використано метод LASSO, який здійснює регуляризацію параметрів і дозволяє подолати недоліки методу найменших квадратів.

Після побудови регресійної моделі методом LASSO був виконаний аналіз регресії за різних значень параметра λ .

Після проведення кросвалідації були визначені оптимальні значення параметра λ : $\lambda_{min} = 0,000224834$ і $\lambda_{1se} = 0,03417277$.

Коефіцієнти регресії при значеннях λ наведені в таблиці 4.

Таблиця 4. Коефіцієнти регресії, розраховані методом LASSO

Коефіцієнти регресії при $\lambda_{min} = 0,000224834$	b_0	b_1	b_2	b_3	b_4	b_5
	-8,8561	0,0008	0,4392	-0,0066	0,0579	1,6099
	b_6	b_7	b_8	b_9	b_{10}	
	0,1021	0,0293	-0,0293	-1,7263	-0,2945	

Наведемо оцінки дисперсій похибок для лінійної регресії, побудованої методом найменших квадратів і методом LASSO.

В першому випадку (для МНК) оцінка дисперсії похибок становить 2,5124. Для регресії, яка побудована методом LASSO при λ_{min} оцінка

дисперсії похибок дорівнює 2,5156. Для регресії, побудованої методом LASSO при λ_{1se} дисперсії похибок дорівнює 4,3341.

Виконаємо аналіз побудованої моделі шляхом дослідження коефіцієнту детермінації R^2 , коефіцієнтів еластичності та параметрів β_i .

У нашому випадку значення коефіцієнта детермінації $R^2 = 0,9928$ (для першої моделі) і $R^2 = 0,9864$ (для другої моделі), що говорить про

суттєвий вплив факторів на показник врожайності у порівнянні з випадковими похибками.

Значення коефіцієнтів еластичності для кожного коефіцієнта регресійної моделі наведені в таблиці 5. На основі коефіцієнтів еластичності робимо наступний висновок.

Розраховані коефіцієнти β_i наведені у таблиці 5.

Таблиця 5. Коефіцієнти еластичності та β_i

β_i	x_1	x_2	x_3	x_4	x_5
	0,03609	0,20342	-0,07099	0,299925	0,541895
	x_6	x_7	x_8	x_9	x_{10}
	0,368873	0,04421	-0,13984	-1,02309	-0,201984

На основі першої моделі фактори розташовуються у наступному порядку (від найвпливовішого фактору, до того, що має найменший вплив): x_9 – загальна кількість днів з опадами у період цвітіння; x_5 – гідротермічний коефіцієнт (ГТК) в період цвітіння; x_7 – абсолютна мінімальна відносна вологість повітря в травні; x_6 – середньомісячна сума опадів за серпень; x_1 – сума активних температур за вегетаційний період (до фази досягання плодів); x_4 – сума ефективних температур у період цвітіння; x_2 – різниця між середньою максимальною та мінімальною температурами повітря у період цвітіння; x_{10} – середня з максимальних значень температур повітря в період цвітіння; x_8 – сума опадів у період цвітіння; x_3 – сума активних температур у період цвітіння.

На основі показників цих же коефіцієнтів оцінимо долю впливу кожного фактору в дисперсії показника урожайності вишні по абсолютному значенню (результати наведено на основі першої моделі).

Доля впливу фактору складає: x_1 – 9,90%, фактору x_2 – 5,56%, фактору x_3 – 1,95%, фактору x_4 – 8,22%, фактору x_5 – 14,83%, фактору x_6 – 10,12%, фактору x_7 – 12,12%, фактору x_8 – 3,81%, фактору x_9 – 27,96%, фактору x_{10} – 5,54%.

Висновки

У підсумку, в умовах Південного Степу України виявлено вплив погодних факторів регіону на формування показників врожайності

вишні в межах 2007–2019 років досліджень. Проведення кореляційного аналізу дозволило визначити десять погодних факторів, що мають помітний (значний) та сильний як прямий, так і обернений лінійний кореляційний зв'язки з урожайністю вишні ($r = 0,68$ – $0,86$).

При аналізі впливу факторів погодних умов на урожайність вишні методами регресійного аналізу вкрай важливо враховувати ефект кореляції факторів. Несприятливий ефект мультиколінеарності негативно позначається на інтерпретації побудованої моделі, а саме, за аналізу ступеня впливу кожного фактора на досліджуваний показник. Запропоновано модель, яка описує вплив гідротермічних факторів на врожайність вишні. За аналізу даних з допомогою парних коефіцієнтів кореляції і показника VIF був виявлений ефект мультиколінеарності. У цьому випадку використовувати метод найменших квадратів для побудови і аналізу регресійної моделі є неефективним. Запропонована регресійна модель, яка побудована на основі методу LASSO. Метод LASSO дозволяє побудувати достовірні оцінки параметрів регресії. На основі побудованої моделі проаналізовано чинники, що впливають на показник врожайності вишні і показано, що найбільш значущим є фактор): загальна кількість днів з опадами у період цвітіння, доба; гідротермічний коефіцієнт (ГТК) у період цвітіння; абсолютна мінімальна відносна вологість повітря в травні, %; середньомісячна сума опадів за серпень, мм; сума активних температур за вегетаційний період (до фази досягання плодів); сума ефективних температур в період цвітіння, °C; різниця між

середньою, максимальною та мінімальною температурами повітря в період цвітіння, °С; середня з максимальних значень температур повітря у період цвітіння; сума опадів у період цвітіння, мм; сума активних температур у період цвітіння, °С.

Визначено показники частки впливу кожного фактору на загальну дисперсію показника врожайності вишні. Доля впливу фактору складає: x_1 – 9,90%, фактору x_2 – 5,56%, фактору x_3 – 1,95%, фактору x_4 – 8,22%, фактору x_5 – 14,83%, фактору x_6 – 10,12%, фактору x_7 – 12,12%, фактору x_8 – 3,81%, фактору x_9 – 27,96%, фактору x_{10} – 5,54%.

References

- Aiken, L. S. & West, S. G. (1991). *Multiple Regression: Testing and Interpreting Interactions*. Newbury Park : Sage Publications.
- Bohdaniuk, O. V. (2016). Otsinka vplyvu chynnykiv na urozhainist plodovo-iahidnykh kultur v konteksti efektyvnoho upravlinnia sadivnytstvom [Assessing the impact of factors on the yield of fruit and berry crops in the context of effective horticulture management]. *Molodyi vchenyi*, 11, 555–558 [in Ukrainian].
- Bublyk, M. O. (2005). *Metodolohichni ta tekhnolohichni osnovy pidvyshchennia produktyvnosti suchasnoho sadivnytstva* [Methodological and technological foundations for improving the productivity of modern gardening]. Kyiv : Nora-print [in Ukrainian].
- Damodar, N. G. (2004). *Basic Econometrics* (4-th ed.). New York : McGraw-Hill.
- Graham, M. H. (2003). Confronting Multicollinearity in Ecological Multiple Regression. *Ecology*, 84 (11), 289–281. doi.org/10.1890/02-3114.
- Hastie, T., Tibshirani, R. & Wainwright, M. (2015). *Statistical Learning with Sparsity. The Lasso and Generalization*. London, New York : Chapman & Hall.
- Ivanova, I. Ye., Serdiuk, M. Ye., Herasko, T. V., Bilous, E. S. & Kryvonos, I. A. (2019). Urozhainist chereszni zalezno vid klimatychnykh umov rokiv vyroshchuvannia [The productivity of cherry is depending on the climatic terms of years of growing]. *Visnyk Ahrarnoi nauky Prychornomia*, 3, 32–48 [in Ukrainian].
- Ivanova, I. Ye., Serdiuk, M. Ye., Malkina, V. M., Shkinder-Barmina, A. M. & Kryvonos, I. A. (2019). Urozhainist vyshni zalezno vid klimatychnykh umov rokiv vyroshchuvannia [Cherry yields depending on the climatic conditions of the growing years]. *Visnyk Ahrarnoi nauky Prychornomia*, 4, 29–50 [in Ukrainian].
- Ivashenko, A. (2012). Budushee Ukrainy – v globalnoj sisteme proizvoditelej prodovolstviya [The future of Ukraine lies in the global system of food producers]. *Gazeta. zn. ua*. Retrieved from http://gazeta.zn.ua/LAW/budushee_ukrainy_v_globalnoy_sisteme_proizvoditelej_prodovolstviya.html [in Russian].
- James, G., Witten, D., Hastie, T. & Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. [S.I.]: Springer.
- Kutner, M. H., Nachtsheim, C. J. & Neter, J. (2004). *Applied Learning Regression Models* (4-th ed.). New York : McGraw-Hill.
- Semenov, V. A., Minin, V. B. & Jakushev, M. P. (2000). Climate changing and field experimental practice. *Aspects of Applied Biology*, 61, 113–120.
- Serdyuk, M. E. & Rastorguev, A. B. (2013). Ocenka vliyaniya pogodnykh faktorov na urozhajnost yabloni v usloviyah Yuzhnoj stepnoj zony Ukrainy [Assessment of the influence of weather factors on the yield of apple trees in the conditions of the southern steppe zone of Ukraine]. *Plodovodstvo*, 25, 132–140 [in Russian].
- Turovceva, V. A., Turovcev, N. I., Shkinder-Barmina, A. N. & Turovceva, N. N. (2013). Sozdanie sortov vishni i dyukov na yuge Ukrainy [Creation of varieties of cherries and dykes in the south of Ukraine]. *Metodologicheskoe obospechenie selekcii sadovykh kultur i vinograda na sovremennom etape*, 1, 135–142 [in Russian].
- Vasylyshyna, O. V. & Postolenko, Ye. P. (2019). Vplyv osoblyvostei sortu na yakist zamorozhenykh plodiv vyshni [Influence of variety characteristics on the quality of frozen cherry fruit]. *Naukovi horyzonty. «Scientific horizons»*, 2 (75), 44–49. doi: <https://doi.org/10.33249/2663-2144-2019-75-2-44-49> [in Ukrainian].
- Vasylyshyna, O. V. (2019). Vplyv obrobky rozchynom alginatu natriyu na plody vyshni pid chas zberigannya [The influence of sodium alginate processing on fruits of cherry of the storage]. *Naukovi horyzonty. «Scientific horizons»*, 10 (83), 35–40. doi: <https://doi.org/10.33249/2663-2144-2019-83-10-35-40> [in Ukrainian].
- Yu-Kang, Tu, Clerehugh, V., Gilthorpe, M. S. (2004). Collinearity in linear regression is a serious problem in oral health research. *Eur J Oral Sci*, 112 (5), 389–397. doi:10.1111/j.1600-0722.2004.00160.x.