

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ, СИСТЕМНИЙ АНАЛІЗ ТА КЕРУВАННЯ

УДК 519.816: 004.9+519.6

DOI: 10.20535/1810-0546.2017.1.87423

П.І. Бідюк^{1*}, О.М. Терентьев¹, Т.І. Просянкіна-Жарова², В.В. Ефендієв²

¹КПІ ім. Ігоря Сікорського, Київ, Україна

²Уманський державний педагогічний університет імені Павла Тичини, Київ, Україна

ПРОГНОЗНЕ МОДЕЛЮВАННЯ НЕЛІНІЙНИХ НЕСТАЦІОНАРНИХ ПРОЦЕСІВ У РОСЛИННИЦТВІ З ВИКОРИСТАННЯМ ІНСТРУМЕНТІВ SAS ENTERPRISE MINER

Background. The issue of providing the increase of production of main agricultural crops in Ukraine under conditions of environmental management requires the use of modern scientific approaches. The complexity of solving this problem lies in the lack of practical experience of applying modern information-analytical systems, where different methods for analysis and modeling of nonlinear non-stationary processes in crop production would be implemented simultaneously. The proposed methodology has the advantage of using the tools of SAS Enterprise Miner – software where a wide range of techniques are implemented, that should be used for predictive modeling of main agricultural crops according to the performed research.

Objective. The goal of the study is in application of the integrated methods of analysis and predictive modeling of non-stationary processes for agricultural crop yield prediction using SAS Enterprise Miner tools.

Methods. To solve the problems stated the following approaches were used: systems analysis, regression analysis, gradient boosting, probabilistic modeling and decision trees. The methodology for developing of crop yield prediction under influence of various groups of factors was offered, and the possibility of their use in decision support systems in agriculture was substantiated.

Results. Based on the analysis of the works of domestic and foreign scientists it was proposed to improve methodology of development of yield prediction of main agricultural crops using integrated analysis methods, which were implemented in the system of SAS Enterprise Miner. The analysis of the obtained results was performed.

Conclusions. Winter wheat and corn yield prediction was performed for the Forest-Steppe Zone using the developed methodic. Different methods of construction of models for prediction of the non-stationary processes were applied; the choice of the worthiest one was reasonably proved. Advanced information technologies, including SAS Enterprise Miner, were used for automatization the process of selecting the optimal model for investigated crop yield prediction.

Keywords: non-stationary processes; regressive model; agricultural crop yield; prediction; decision support systems; SAS Enterprise Miner.

Вступ

Сільське господарство є одним із найбільш ризикованих видів економічної діяльності, оскільки його розвиток триває в умовах мінливого природного середовища та необхідності нарощування виробництва з орієнтацією на світові тенденції. Україна активно використовує свій аграрний потенціал, щороку збільшуючи обсяги експорту продукції рослинництва, зокрема кукурудзи, соняшнику, пшениці. Однак інтенсивне використання ресурсної бази сільського господарства за низького рівня впровадження інноваційних технологій виробництва продукції рослинництва, а також часто і недотримання вимог агротехніки спричиняють значний вплив на ґрунти, знижуючи їх родючість, забруднюючи воду та повітря залишками від розпаду пестицидів і мінеральних добрив тощо, погіршуючи тим самим стан навколишнього середовища. Все це створює загрози по-

дальшому ефективному функціонуванню агробізнесу в Україні та ризики виникнення екологічних катастроф [1, 2].

Саме тому в рамках проекту NUKR.SFPP G4877 “Modeling and Mitigation of Social Disasters Caused by Catastrophes and Terrorism” [3] як одні з пріоритетних розглядаються задачі розробки та впровадження спеціалізованих систем підтримки прийняття рішень (СППР), призначених для використання у сільському господарстві України. Особливе прикладне значення оцінювання стану і прогнозування урожайності сільськогосподарських культур у контексті реалізації СППР для вирішення завдань сталого розвитку та економічної безпеки відзначається і в низці міжнародних та національних програм, зокрема, це завдання є одним із пріоритетних у системі GEOSS [4].

Однак, незважаючи на наявність значних масивів сільськогосподарських даних, які з достатньою повнотою характеризують процеси, що

*corresponding autor: pbidyuke_00@ukr.net

досліджуються, СППР у аграрному секторі України використовуються досить обмежено [5–7]. Значною мірою це пов'язано з відсутністю опрацьованих методик аналізу та прогнозування сільськогосподарських даних на різних рівнях управління, складністю обробки значних обсягів різномірної вхідної інформації, відсутністю у фахівців-аграріїв навичок застосування інтегрованих методів аналізу даних і засобів сучасних інформаційних технологій. Тому, з огляду на практичну значущість проблеми і відсутність опрацьованої та адаптованої до вітчизняних умов методики для розв'язання задач цього класу, завдання стосовно розробки методики моделювання сільськогосподарських даних є актуальним.

Для прогнозування обсягів виробництва сільськогосподарської продукції пропонується розробити множину математичних моделей за допомогою інструментарію системи SAS Enterprise Miner [8], що дає можливість автоматизувати вибір кращих моделей із використанням множини статистичних критеріїв. У дослідженні використано матеріали щодо урожайності основних сільськогосподарських культур підприємств зони лісостепу України [9].

Перевагами запропонованої методики моделювання і прогнозування є можливість опрацьовування широкого спектра процесів, що використовуються у практичній діяльності фахівцями-економістами і агрономами.

Постановка задачі

Метою роботи є удосконалення існуючих методик аналізу, математичного моделювання та прогнозування урожайності сільськогосподарських культур, основаних на використанні комбінацій різних методів, а також виконання обчислювальних експериментів з метою надання прикладів аналізу та прогнозування урожайності окремих сільськогосподарських культур за допомогою математичних моделей, розроблених із використанням сучасних інструментів SAS Enterprise Miner на основі статистичних даних.

Регресійне моделювання та прогнозування урожайності

Оскільки існує тісний зв'язок результативності сільського господарства із станом та динамікою біоресурсів і природокористування, то необхідно удосконалювати існуючу систему йо-

го інформаційно-аналітичного забезпечення, орієнтовану на опрацювання ретроспективних статистичних даних щодо обсягів виробництва продукції та розвитку економічних процесів. Фактично при цьому не використовуються геоінформаційні технології, інформаційно-аналітичні системи, призначені для прогнозування метеорологічних умов, продуктивності тваринництва, формування оптимального раціону відгодівлі худоби, формування технологічних карт вирощування сільськогосподарських культур на основі технологій точного землеробства та раціонального використання природних ресурсів.

Сьогодні для оцінювання урожайності сільськогосподарських культур використовуються різні підходи, серед яких переважають статистичні методи та припущення щодо існування залежності між характеристиками навколишнього середовища, внесенням добрив і урожайністю.

Біофізичні моделі, де як вхідні змінні використовуються дані дистанційного зондування з метою калібрування моделі та її настроювання, застосовуються набагато рідше. Основним недоліком таких моделей є необхідність задавання численних специфічних параметрів для кожної культури: характеристик ґрунту, методів обробки культур, агрометеорологічних даних, строків посіву. Все частіше для побудови моделей, що описують залежність урожайності сільськогосподарських культур від обсягів внесених поживних речовин, стану ґрунтів, впливу факторів навколишнього середовища, застосовують дані, одержані в результаті дистанційного зондування [10]. Ці методи, як правило, прості, а їх реалізація не вимагає задавання значної кількості параметрів. Однак необхідно враховувати, що процеси, які описують урожайність основних сільськогосподарських культур, як правило, є нестационарними, а показники урожайності необхідно опрацьовувати для значної кількості господарств певної ґрунтово-кліматичної зони.

Статистичні методи прогнозування врожайності на основі оцінок фізичних факторів середовища ґрунтуються на припущенні про існування простої залежності між характеристиками навколишнього середовища та врожайністю. Крім того, враховуючи визначальну роль стану формування біомаси рослин у певні фази їх розвитку, для моделювання урожайності сільськогосподарських культур застосовують кількісні показники фотосинтетичної активності біомаси – вегетаційні індекси. Один із

них – NDVI (Normalized Difference Vegetation Index) – нормалізований відносний індекс вегетації – простий показник кількості фотосинтетичної активної біомаси (так званий вегетаційний індекс). У цілому, головною перевагою NDVI є простота його отримання: для обчислення індексу не потрібно додаткових даних і методик, окрім безпосередньо самої космічної зйомки і знання її параметрів. Динаміка сезонної зміни NDVI дає можливість оцінювати стан рослин у різних фазах розвитку. Методи, запропоновані для розв'язання задач цього класу, рідко застосовуються на практиці внаслідок значної трудомісткості необхідних наземних вимірів і складності досліджуваних залежностей [6].

У виконаному дослідженні інформаційною базою є статистичні дані стосовно урожайності основних сільськогосподарських культур за 2007–2013 рр., вирощуваних у зоні лісостепу на полях підприємств, що входять до складу одного з українських холдингів, а також дані щодо впливу фізичних факторів середовища та вегетаційні NDVI-індекси вибраних сільськогосподарських культур.

Побудова моделей за експериментальними даними має істотну теоретичну та практичну основу [1, 6, 7, 10, 11]. Зокрема, існує широкий арсенал методів і засобів структурної та параметричної ідентифікації досліджуваних процесів. Крім того, ці засоби та методи можуть як мати універсальний характер, так і бути певною мірою орієнтованими на конкретні предметні галузі, різний рівень формалізації задачі. Одними з найбільш широко застосовуваних для аналізу експериментальних даних в економіці, фінансах, соціології, біології, психології, медицині, техніці тощо є методи регресійного аналізу [6, 12].

Так, парна лінійна регресія встановлює лінійну залежність між двома змінними [12], наприклад залежність урожайності зернових від кількості внесених добрив, витрат на обробіток ґрунту на полях сівозміни та обсягів реалізованої продукції. У загальному випадку парна вибіркова регресійна модель [13] має вигляд

$$Z = a_0 + a_1 X + E,$$

де Z – вектор значень залежної змінної, $Z = [z_1, z_2, \dots, z_n]$; X – вектор значень незалежної змінної (регресора), $X = [x_1, x_2, \dots, x_n]$; a_0, a_1 – невідомі параметри регресійної моделі; E – век-

тор значень випадкових величин, $E = [e_1, e_2, \dots, e_n]$, поява яких у моделі зумовлена наявністю випадкових збурень, недосконалою структурою моделі, похибками вимірів та обчислень оцінок параметрів моделі.

За великої кількості екзогенних змінних виникає проблема нестійкого оцінювання параметрів моделі, а тому застосовують регресію найменших кутів [14] (least-angle regression (LARS)) – алгоритм відбору ознак у задачах лінійної регресії та метод часткових найменших квадратів (partial least squares (PLS)) [15]. Алгоритм LARS схожий на алгоритм покрокової регресії. Його відмінність полягає в тому, що алгоритм LARS замість послідовного додавання регресорів на кожному кроці змінює їх вагові коефіцієнти.

Метод PLS [15] – статистичний метод, який має деяке відношення до регресії за методом головних компонент, але замість знаходження гіперплощин з мінімальною дисперсією між вихідною і незалежними змінними він знаходить модель лінійної регресії проектуванням прогнозованих і спостережуваних змінних у новому просторі. Оскільки обидві множини даних X і Y переносяться у новий простір, то сімейство PLS-методів відоме як моделі білінійних факторів. Модель регресії, побудована за методом PLS, особливо корисна у випадку, коли матриця предикторів має більше змінних, ніж спостережень, і коли у матриці вимірів X існує мультиколінеарність.

Серед сучасних методів розв'язання задач у багатовимірному просторі також необхідно виділити методи градієнтного бустингу [16]. Задача бустингу у загальному випадку формулюється таким чином [16]. Нехай задано багатовимірний простір X із простором міток Y , з навчальною вибіркою $\{x_i\}_{i=1}^N$, де $x \in X$. Задача зводиться до пошуку пари оптимальних параметрів $\{a_m, b_m\}$ для класифікатора F_m :

$$F_m(x) = F_{m-1}(x) + b_m h(x; a_m), \quad b_m \in R, \quad a_m \in A.$$

Оптимальність розв'язку визначається відповідно до принципу явної максимізації граничних значень, що передбачає введення деякої функції втрат $L(y_i, F_m(x_i))$, $i = \overline{1, N}$, яка показує, наскільки змодельований результат відрізняється від правильної відповіді y_i . Потім мінімізується функціонал похибки [16]:

$$Q_i = \sum_{i=1}^N L(y_i, F_m(x_i)) \rightarrow \min.$$

Для розв'язання задачі мінімізації функціоналу виконують один крок алгоритму градієнтного спуску. У точці, для якої шукають оптимальний приріст, розраховують градієнт функціоналу похибки [16]:

$$\begin{aligned}\nabla Q_i &= \left[\frac{\partial Q}{\partial F_{m-1}}(x_i) \right]_{i=1}^N = \\ &= \left[\frac{\partial \left(\sum_{i=1}^N L(y_i, F_{m-1}) \right)}{\partial F_{m-1}}(x_i) \right]_{i=1}^N = \\ &= \left[\frac{\partial L(y_i, F_{m-1})}{\partial F_{m-1}}(x_i) \right]_{i=1}^N.\end{aligned}$$

Далі, використовуючи відповідні алгоритми навчання, знаходять пари оптимальних параметрів $\{a_m, b_m\}$; для визначення a_m використовується базовий алгоритм навчання, а для b_m – лінійний пошук. Алгоритм градієнтного бустингу застосовують у задачах регресійного моделювання. Так, наприклад, при використанні методу LS-boosting функція втрат має вигляд

$$L(y, F) = \frac{(y - F)^2}{2},$$

а градієнт функціоналу похибки (∇Q) в алгоритмі градієнтного бустингу набуває такого вигляду:

$$\nabla Q_i = y_i - F_{m-1}(x_i),$$

де y_i – значення міток для кожного об'єкта.

Іншим варіантом є алгоритм мінімізації середнього модуля відхилення (LAD-boosting) [16]; при цьому функція втрат має вигляд

$$L(y, F) = |y - F|.$$

Тепер ∇Q набуває вигляду

$$\nabla Q_i = \text{sign}(y_i - F_{m-1}(x_i)).$$

У цьому випадку b_m розраховується через зважене медіанне значення вибірки, тобто виконується задача пошуку порядкової статистики в масиві.

У випадках наявності шуму вимірів та викидів у даних кращі результати можна одержати за допомогою процедури Tree-boosting, оскільки в цьому випадку ∇Q_i набуває значень

в інтервалі $\{+1, -1\}$, а оновлення параметрів відбувається за рахунок медіанних значень. Ще одним методом, що позитивно зарекомендував себе у задачах із сильно зашумленими даними, є метод M-бустингу [16]. Цей метод ґрунтується на функції втрат Хубера, яка має вигляд

$$L(y, F) = \begin{cases} \frac{1}{2}(y - F)^2, & |y - F| \leq \sigma, \\ \sigma \left(|y - F| - \frac{\sigma}{2} \right), & |y - F| > \sigma. \end{cases}$$

Це дає можливість використовувати метод найменших квадратів для незначних похибок та функцію втрат – для великих. Градієнт у цьому випадку має вигляд

$$\begin{aligned}\nabla Q_i &= \\ &= \begin{cases} y_i - F_{m-1}(x_i), & |y_i - F_{m-1}(x_i)| \leq \sigma, \\ \sigma - \text{sign}(y_i - F_{m-1}(x_i)), & |y_i - F_{m-1}(x_i)| > \sigma. \end{cases}\end{aligned}$$

Оскільки розподіл похибок на кожній ітерації бустингу різний, то на кожній ітерації підбирається нове значення σ_m за виразом

$$\sigma_m = \text{quantile}_\alpha \{ |y_i - F_{m-1}(x_i)| \}_{i=1}^N.$$

Унаслідок того, що процеси, які мають місце у сільськогосподарському виробництві, характеризуються множиною сильно взаємодіючих змінних, а також того, що існує потреба формування структур даних для розв'язання інших задач аналізу сільськогосподарських процесів, виникає необхідність використання композицій різних методів.

Сільськогосподарське виробництво має такі особливості: вплив попередника у сівозміні, неоднорідність ґрунтів, особливості технологій вирощування тощо. Тому при оцінюванні впливу різних факторів на зміну урожайності різних сільськогосподарських культур виникає проблема виявлення прихованих закономірностей в оброблюваних даних. Для розв'язання цієї задачі високоякісні прогнози можна одержати, застосовуючи такий різновид дерев рішень, як дерева регресії. Побудова регресійних моделей із застосуванням дерев рішень відбувається у два етапи: спочатку для відбору найбільш значущих змінних будується дерево рішень, а потім – модель лінійної регресії, у якій використано одержані на попередньому етапі множини найбільш значущих змінних. Сьогодні розроблено та реалізовано програмне забезпечення для побудови дерев рішень за алгоритмами

C4.5, C5.0, CART, CHAID, CN2, SPRINT, ID3 тощо. Перевагами дерев рішень є наочність, швидке навчання, в основному прийнятна точність прогнозу, а також можливість розробки правил у випадках, коли знання складно формалізувати.

Формалізм побудови узагальнених графічних моделей поєднує у собі багато методів статистичного моделювання, таких як факторний аналіз, аналіз розподілів, моделі сумішей розподілів, приховані марковські моделі, фільтри Калмана, моделі Айзінга та деякі інші. Перевагою такого підходу є те, що методи дослідження процесів та обробки даних, розроблені в одній області, можуть бути успішно перенесені в інші [10].

Мережа Байєса являє собою пару $\langle G, B \rangle$ [17], у якій перша компонента G – це спрямований нециклічний граф, що відповідає випадковим змінним і записується як множина умов незалежності: кожна змінна незалежна від її батьків на графі. Друга компонента пари B – це множина параметрів у формі таблиць умовних імовірностей.

Компонента B містить параметри $\Theta_{X^{(i)}|\rho\alpha(X^{(i)})} = P(X^{(i)}|\rho\alpha(X^{(i)}))$ для кожного можливого значення $x^{(i)} \in X^{(i)}$ та $\rho\alpha(X^{(i)}) \in P\alpha(X^{(i)})$, де $P\alpha(X^{(i)})$ позначає множину батьків змінної $X^{(i)} \in G$. Кожна змінна $X^{(i)} \in G$ зображується у вигляді вершини. Якщо розглядають більше одного графа, то для визначення батьків змінної $X^{(i)}$ у графі G використовують позначення $P\alpha^G(X^{(i)})$. Повна спільна ймовірність для мережі Байєса обчислюється за формулою [17]

$$P_B(X^{(1)}, \dots, X^{(N)}) = \prod_{i=1}^n P_B(X^{(i)} | P\alpha(X^{(i)})). \quad (1)$$

З математичної точки зору, мережа Байєса – це модель подання існуючих і відсутніх імовірнісних причинно-наслідкових залежностей. При цьому зв'язок $A \rightarrow B$ є причинним, коли подія A є причиною виникнення B , тобто існує механізм, відповідно до якого значення, набуте A , впливає на значення, набуте B .

Теорія побудови мереж Байєса ґрунтується на припущенні, що події є вичерпними і не перетинаються. Якщо ця умова не виконується, то результати застосування мережі будуть неконсистентними (неточними). У випадку, коли події вичерпні і не перетинаються,

ймовірність події E можна обчислити за допомогою умовних ймовірностей [17]:

$$P(E) = \sum_{i=1}^n P(E \cap H_i) = \sum_{i=1}^n P(E|H_i) \cdot P(H_i). \quad (2)$$

Використовуючи рівняння (1), суму перетинів події E з H можна визначити так:

$$P(E \cap H_k) = P(E|H_k) \cdot P(H_k) = P(H_k|E) \cdot P(E).$$

З цієї рівності знайдемо, що

$$P(H_k|E) = \frac{P(E|H_k) \cdot P(H_k)}{P(E)},$$

а з урахуванням (2) отримаємо вираз

$$P(H_k|E) = \frac{P(E|H_k) \cdot P(H_k)}{\sum_{i=1}^n P(E|H_i) \cdot P(H_i)},$$

що представляє собою формулу Байєса. Тут H_k означає будь-яку гіпотезу з n можливих. Ймовірності $P(E|H_k)$ задаються експертами апіорно або розраховуються за навчальними даними. Ймовірності $P(E|H_k)$ є дуже корисними, тому що, як правило, легше знайти ймовірність послідовності подій типу причина-наслідок, ніж навпаки. Значення $P(H_k)$ називають апіорними ймовірностями, вони визначають початкові ймовірності для всіх гіпотез. Переваги байєсівського методу полягають у тому, що апіорні ймовірності можна уточнювати відповідно до реалій перебігу процесу, що досліджується. Це дає можливість уточнювати ймовірності подій при надходженні додаткової інформації та створювати системи адаптивного моделювання.

З урахуванням необхідності опрацювання значних обсягів статистичних даних про урожайність сільськогосподарських культур, показників NDVI та погодних факторів для моделювання і прогнозування використано програмне забезпечення SAS Enterprise Miner [8]. Для виконання дослідження запропоновано методику у вигляді такої послідовності.

1. Завантаження даних.
2. Розділення даних на 2 підвибірки: навчальну (70 %) та перевіірочну (30 %).
3. Побудова моделей.
4. Порівняння моделей за статистичними критеріями якості.
5. Побудова загального звіту за виконаним обчислювальним експериментом.

Типовим прикладом аналізу даних є дослідження урожайності озимої пшениці. Як видно з графіка (рис. 1), значення математичного сподівання показника урожайності врожайності різняться в ці роки у 1,62 разу, а дисперсія – у 6,13 разу, що підтверджує нестационарність цього процесу.

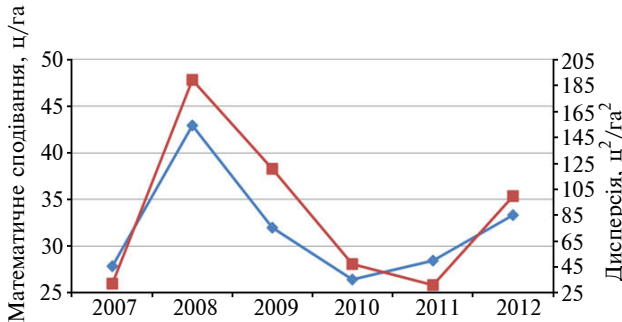


Рис. 1. Графік умовного математичного сподівання та дисперсії урожайності озимої пшениці за 2007–2012 рр.

На рис. 2 зображені графіки автокореляційної та часткової автокореляційної функції часового ряду урожайності кукурудзи (з 1991 по 2013 рр.). Як видно з рис. 2, АКФ і ЧАКФ не є постійно спадаючими. Значення АКФ і ЧАКФ є значними навіть для других різниць для 1, 2, 3, 13 і 18-го лагу (а також для набору лагів для вихідного ряду та перших різниць), що свідчить про наявність тренду другого порядку у значеннях ряду.

На рис. 3 подано графік динаміки урожайності кукурудзи в Україні з 1991 по 2013 рр. з трендом другого порядку. Критерії адекватності побудованого тренду: $R^2 = 0,86$, $R^2_{adj} = 0,84$.

У результаті виконання дослідження отримано моделі лінійної регресії, LARS, PLS, градієнтного бустингу для прогнозування урожайності озимої пшениці та кукурудзи. Моделювання залежності врожайності від внесених матеріалів виконано на прикладі даних про урожайність кукурудзи для двох груп полів (табл. 1). Побудована модель лінійної регресії за критерієм R^2 . Отримані результати для кукурудзи для першої групи полів:

$$\begin{aligned} \text{Productivity} &= \\ &= N \cdot 0,01534 + P \cdot 0,06844 + 1,22805, \end{aligned}$$

для другої групи полів:

$$\begin{aligned} \text{Productivity} &= \\ &= N \cdot 0,00807 + P \cdot 0,0834 + 4,55825. \end{aligned}$$

Для оцінювання статистичних характеристик моделей у дослідженні використані такі статистики [12]: SSE – сума квадратів похибок моделі:

$$SSE = \sum_{i=1}^N (y_i - \hat{y}_i)^2;$$

коефіцієнт детермінації:

$$R^2 = \frac{\text{Var}(\hat{y})}{\text{Var}(y)};$$

скоригований коефіцієнт детермінації R^2 :

$$R^2_{adj} = 1 - (1 - R^2) \frac{n-1}{n-k};$$

$RASE$ (Root Average Square Error) – корінь квадратний із середньої квадратичної похибки:

Вихідний ряд				Перші різниці				Другі різниці			
Autocorrelation	Partial Correlation	AC	PAC	Autocorrelation	Partial Correlation	AC	PAC	Autocorrelation	Partial Correlation	AC	PAC
1	0.646	0.646		1	-0.665	-0.665		1	-0.718	-0.718	
2	0.712	0.506		2	0.330	-0.203		2	0.317	-0.409	
3	0.470	-0.193		3	-0.116	0.023		3	-0.121	-0.232	
4	0.427	-0.113		4	0.144	0.239		4	0.147	0.133	
5	0.273	-0.037		5	-0.195	-0.015		5	-0.247	-0.090	
6	0.221	-0.004		6	0.263	0.133		6	0.256	-0.044	
7	0.086	-0.112		7	-0.181	0.114		7	-0.158	0.026	
8	0.058	-0.005		8	0.078	-0.018		8	0.049	-0.012	
9	-0.109	-0.192		9	0.003	0.006		9	-0.005	0.005	
10	-0.200	-0.253		10	-0.026	-0.063		10	-0.000	-0.066	
11	-0.251	0.093		11	-0.011	-0.065		11	-0.035	-0.090	
12	-0.306	0.052		12	0.042	-0.032		12	0.086	0.045	
13	-0.363	-0.191		13	-0.149	-0.233		13	-0.185	-0.231	
14	-0.332	0.054		14	0.232	0.101		14	0.245	-0.022	
15	-0.347	0.084		15	-0.234	0.011		15	-0.219	-0.041	
16	-0.277	0.003		16	0.149	0.010		16	0.182	0.097	
17	-0.335	-0.196		17	-0.189	-0.205		17	-0.213	-0.129	
18	-0.257	0.045		18	0.190	-0.073		18	0.199	-0.220	
19	-0.278	-0.030		19	-0.205	0.006		19	-0.151	-0.071	
20	-0.168	0.026		20	0.135	-0.089		20	0.071	-0.141	
21	-0.126	0.189		21	-0.095	-0.013					
22	-0.044	-0.072									

Рис. 2. Значення АКФ та ЧАКФ урожайності кукурудзи в Україні з 1991 по 2013 рр.

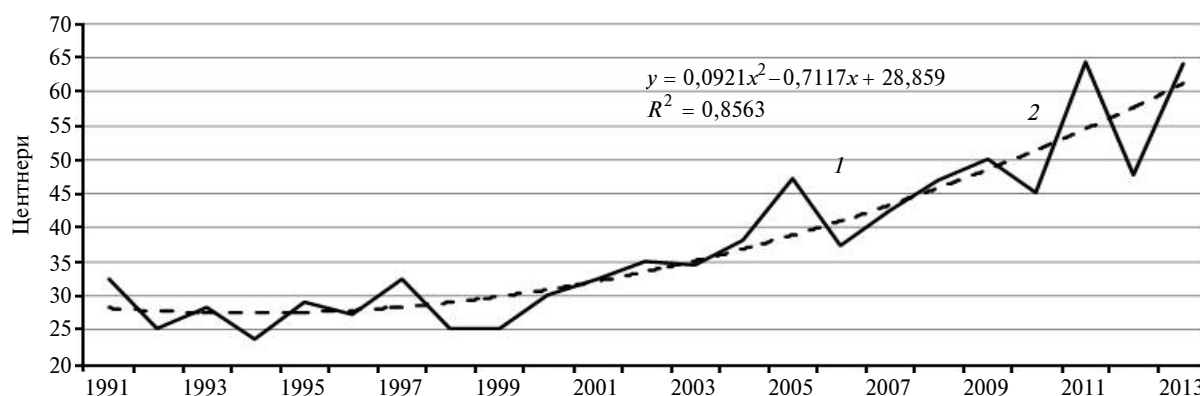


Рис. 3. Урожайність кукурудзи в Україні з 1991 по 2013 рр. із трендом другого порядку: 1 – урожайність; 2 – поліноміальна (врожайність)

$$RASE = \sqrt{\frac{1}{N-k} \sum_{i=1}^N (y_i - \hat{y}_i)^2};$$

MAX E (Maximum Error) – абсолютна величина максимальної похибки:

$$MAX E = \max_{i=1, \dots, N} |y_i - \hat{y}_i|.$$

Таблиця 1. Статистичні характеристики моделі залежності урожайності кукурудзи від обсягів внесених добрив за лінійною регресією

Критерій	<i>RASE</i>	<i>RASE</i> , %	R^2	R^2_{adj}
Значення для 1-ї групи полів	0,22	2,87	0,89	0,87
Значення для 2-ї групи полів	0,05	0,84	0,88	0,86

Моделювання залежності урожайності NDVI, погодних факторів виконано на матеріалах урожайності озимої пшениці. Побудована модель лінійної регресії має такий вигляд:

$$\begin{aligned} \text{Productivity} = & 25,352 \cdot \text{Intercept} + \\ & + 89,646 \cdot \text{ndvi}_{17_18} - 29,044 \cdot \text{ndvi}_{19_20} + \\ & + 74,230 \cdot \text{ndvi}_{23_24} - 89,578 \cdot \text{ndvi}_{25_26} + \\ & + 52,028 \cdot \text{ndvi}_{27_28} + 0,330 \cdot \text{temp}_{17_18} + \\ & + 0,711 \cdot \text{temp}_{19_20} - 4,118 \cdot \text{temp}_{21_22}. \end{aligned}$$

Регресори типу ndvi_{17_18} позначають різницю між значенням нормалізованого відносного індексу вегетації (NDVI) на 18 і 17-му тиж-

нях; відповідно, temp_{17_18} позначає різницю між температурою навколишнього середовища на 18 і 17-му тижнях вегетації. У табл. 2 наведені статистичні характеристики отриманих результатів.

Модель, побудована за методом LARS із вибором моделі за значенням ASE на перівірочній вибірці, має такий вигляд:

$$\begin{aligned} \text{Productivity} = & -32,616 \cdot \text{Intercept} + \\ & + 89,325 \cdot \text{ndvi}_{17_18} - 28,519 \cdot \text{ndvi}_{19_20} + \\ & + 3,361 \cdot \text{ndvi}_{21_22} + \\ & + 70,514 \cdot \text{ndvi}_{23_24} - 89,180 \cdot \text{ndvi}_{25_26} + \\ & + 51,874 \cdot \text{ndvi}_{27_28} - 11,133 \times \\ & \times \text{precip}_{17_18} - 1,462 \cdot \text{precip}_{19_20} + \\ & + 0,572 \cdot \text{temp}_{21_22}. \end{aligned}$$

Регресор precip_{17_18} – це різниця між кількістю опадів на 18 і 17-му тижнях вегетації. У табл. 3 наведено статистичні характеристики результатів прогнозування.

Модель за методом LARS з вибором моделі за значенням SBC (23) має такий вигляд:

$$\begin{aligned} \text{Productivity} = & -2,641 \cdot \text{Intercept} + \\ & + 74,067 \cdot \text{ndvi}_{17_18} + 10,496 \cdot \text{ndvi}_{27_28} - \\ & - 7,323 \cdot \text{precip}_{17_18} - 0,286 \cdot \text{precip}_{19_20} - \\ & - 0,932 \cdot \text{temp}_{21_22}. \end{aligned}$$

Статистичні характеристики результатів наведені в табл. 4.

Таблиця 2. Статистичні характеристики моделювання урожайності озимої пшениці залежно від значень нормалізованого відносного індексу вегетації за методом лінійної регресії

Вибірка	Навчальна			Перевірочна		
	<i>RASE</i>	<i>SSE</i>	<i>MAX E</i>	<i>RASE</i>	<i>SSE</i>	<i>MAX E</i>
Показник Значення	5,21	977,74	13,64	5,54	459,81	11,73

Таблиця 3. Статистичні характеристики моделювання урожайності озимої пшениці залежно від факторів зовнішнього середовища за методом LARS із вибором моделі за значенням ASE на перевірочній вибірці

Вибірка	Навчальна			Перевірочна		
	<i>RASE</i>	<i>SSE</i>	<i>MAX E</i>	<i>RASE</i>	<i>SSE</i>	<i>MAX E</i>
Показник Значення	5,21	977,73	13,64	5,57	465,68	11,74

Таблиця 4. Статистичні характеристики моделювання урожайності озимої пшениці під впливом факторів зовнішнього середовища за методом LARS із вибором моделі за значенням SBC

Вибірка	Навчальна			Перевірочна		
	<i>RASE</i>	<i>SSE</i>	<i>MAX E</i>	<i>RASE</i>	<i>SSE</i>	<i>MAX E</i>
Показник Значення	5,54	1104,23	15,74	6,30	595,91	12,65

Результат моделювання урожайності озимої пшениці за методом PLS (24):

$$\begin{aligned} \text{Productivity} = & -22,657 \cdot \text{Intercept} + \\ & + 89,325 \cdot \text{ndvi}_{17_18} - 28,519 \times \\ & \times \text{ndvi}_{19_20} + 3,361 \cdot \text{ndvi}_{21_22} + \\ & 70,514 \cdot \text{ndvi}_{23_24} - 89180 \cdot \text{ndvi}_{25_26} + \\ & + 51,874 \cdot \text{ndvi}_{27_28} - 2,887 \times \\ & \times \text{precip}_{17_18} - 0,568 \times \\ & \times \text{precip}_{19_20} + 1,240 \cdot \text{precip}_{21_22} + \\ & + 1,340 \cdot \text{temp}_{23_24} - 0,702 \cdot \text{temp}_{25_26}. \end{aligned}$$

У табл. 5 наведені статистичні характеристики результатів.

Метод лінійної регресії з відбором значущих змінних за допомогою дерева рішень:

$$\begin{aligned} \text{Productivity} = & 41,648 \cdot \text{Intercept} + \\ & + 94,555 \cdot \text{ndvi}_{17_18} - 3,955 \cdot \text{temp}_{21_22}. \end{aligned}$$

У табл. 6 наведені статистичні характеристики результатів.

Для методу градієнтного бустингу як значущі були вибрані такі змінні: $\text{avg.air.temp}_{17_18}$, $\text{avg.air.temp}_{21_22}$, ndvi_{17_18} , ndvi_{19_20} , ndvi_{21_22} , ndvi_{23_24} , ndvi_{25_26} , ndvi_{27_28} .

Регресор вигляду $\text{avg.air.temp}_{17_18}$ – це різниця між середньою тижневою температурою повітря на 18 і 17-му тижнях вегетації. У

табл. 7 наведені статистичні характеристики результатів.

Для дослідження урожайності озимої пшениці під впливом факторів зовнішнього середовища побудовано дві моделі у формі мережі Байеса. Для побудови першої моделі вихідні дані були перетворені в дискретну форму за методом Optimal Binning (дискретизація проводилась за допомогою аналітичної системи SAS Enterprise Miner [8]).

Для побудови другої мережі Байеса робились такі перетворення:

- обчислені перші різниці від показників NDVI, температури та кількості опадів;
- дані перетворені в дискретну форму за методом рівних проміжків.

На основі дискретизованих даних мережа Байеса побудована за допомогою програми GeNIe з використанням методу Greedy Thick Thinning. На рис. 5 зображено мережу Байеса для озимої пшениці. Точність класифікації становить 47,06 %. За побудованою мережею врожайність безпосередньо залежить від показників NDVI на 17–18-му тижнях; різниці NDVI на 25–26 і 23–24-му тижнях; різниці NDVI на 27–28 і 25–26-му тижнях; різниці кількості опадів на 19–20 і 17–18-му тижнях; різниці температури на 21–22 і 19–20-му тижнях.

У табл. 8 подано показники якості для усіх побудованих моделей залежності врожайності озимої пшениці від NDVI і погодних факторів.

Таблиця 5. Статистичні характеристики моделювання залежності урожайності озимої пшениці від факторів навколишнього середовища за методом PLS

Вибірка	Навчальна			Перевірочна		
Показник	<i>RASE</i>	<i>SSE</i>	<i>MAX E</i>	<i>RASE</i>	<i>SSE</i>	<i>MAX E</i>
Значення	5,21	977,73	13,64	5,57	465,68	11,74

Таблиця 6. Статистичні характеристики моделювання залежності урожайності озимої пшениці від факторів навколишнього середовища за лінійною регресією з відбором значущих змінних за допомогою дерева рішень

Вибірка	Навчальна			Перевірочна		
Показник	<i>RASE</i>	<i>SSE</i>	<i>MAX E</i>	<i>RASE</i>	<i>SSE</i>	<i>MAX E</i>
Значення	5,54	1105,32	14,27	6,42	618,85	13,16

Таблиця 7. Статистичні характеристики моделювання залежності урожайності озимої пшениці від факторів навколишнього середовища за методом градієнтного бустингу

Вибірка	Навчальна			Перевірочна		
Показник	<i>RASE</i>	<i>SSE</i>	<i>MAX E</i>	<i>RASE</i>	<i>SSE</i>	<i>MAX E</i>
Значення	3,63	475,25	10,58	4,94	366,78	13,04

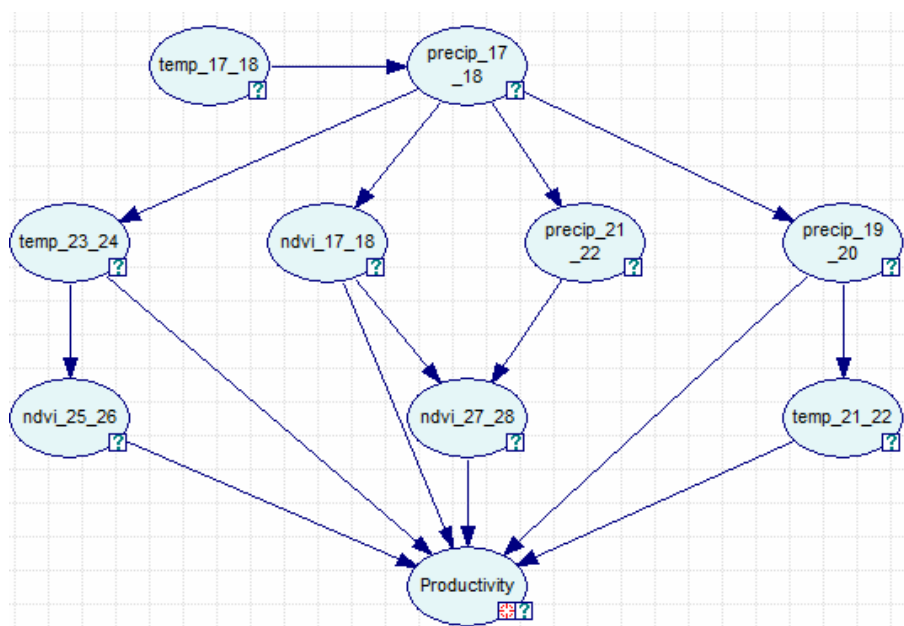


Рис. 5. Мережа Байеса (для озимої пшениці взяті перші різниці вхідних даних)

Як видно з табл. 8, найкращою виявилась модель, побудована за методом градієнтного бустингу. Для порівняння якості побудованих регресійних моделей з мережами Байеса (табл. 9) розраховано такі показники:

- для мереж Байеса розраховано відношення правильно класифікованих значень врожайності (вибиралась та мережа Байеса, у якій це показник вищий);
- для регресійних моделей вхідні дані показника врожайності та показника врожайності, розрахованого за відповідною моделлю,

перетворено в інтервали відповідно до мережі Байеса, з якою виконувалось порівняння.

З отриманих результатів моделювання залежності врожайності озимої пшениці від NDVI та погодних факторів на основі регресійних моделей можна зробити висновок, що найкращою є модель, отримана за методом градієнтного бустингу. Показники якості на навчальній вибірці такі: *RASE* = 3,63; *SSE* = 475,25; *MAX E* = 10,58; на перевірочній вибірці: *RASE* = 4,94; *SSE* = 366,78; *MAX E* = 13,04). Інші моделі показали гірші і схожі результати: *RASE* від 5,21

до 5,54; *SSE* від 977,73 до 1105,32; *MAX E* від 13,64 до 15,74. На перевірочній вибірці отримано такі значення: *RASE* від 5,54 до 6,42; *SSE* від 459,81 до 595,91; *MAX E* від 11,73 до 13,16.

Також побудовані мережі Байеса за вихідними даними, перетвореними в інтервальні значення за методом оптимального розбиття на групи з використанням алгоритму дерева рішень. Для побудови мереж використано спеціальний компонент Optimal Binning у системі SAS Enterprise Miner [8] за вихідними даними,

Таблиця 8. Значення показників якості регресійних моделей залежності врожайності озимої пшениці від NDVI і погодних факторів

Вибірка	Навчальна			Перевірочна		
	<i>RASE</i>	<i>SSE</i>	<i>MAX E</i>	<i>RASE</i>	<i>SSE</i>	<i>MAX E</i>
Regression	5,21	977,74	13,64	5,54	459,81	11,73
LARS	5,21	977,73	13,64	5,57	465,68	11,74
LARS за критерієм мінімального SBC	5,54	1104,23	15,74	6,3	595,91	12,65
PLS	5,21	977,73	13,64	5,57	465,68	11,74
Gradient Boosting	3,63	475,25	10,58	4,94	366,78	13,04

Таблиця 9. Порівняння якості регресійних моделей і мереж Байєса (на прикладі прогнозування урожайності пшениці)

Модель	Відсоток класифікації на тестовому наборі даних
Мережа Байєса	78,43
Gradient Boosting	86,11
Regression	69,44
PLS	69,44
LARS	69,44

вихідними даними, переведеними в різницеву форму і в інтервали за методом рівних інтервалів. Отримані мережі Байєса порівнювались за загальною точністю класифікації, показники якої становили 55,36 і 70,98 % для кукурудзи та 78,43 і 47,06 % для пшениці.

Також виконано порівняння результатів, отриманих за регресійними моделями та байєсівськими мережами. Кращими результатами залежності врожайності для обох досліджуваних культур виявились ті, що отримані за методом градієнтного бустингу.

Висновки

Застосування сучасних технологій прогнозування урожайності сільськогосподарських культур надає змогу вітчизняним аграріям значно підвищити їх конкурентоспроможність на світовому ринку сільськогосподарської продукції. Це стало можливим не лише за рахунок збільшення посівних площ та підвищення урожайності сільськогосподарських культур, внесення значної кількості мінеральних добрив і регуляторів росту, а й завдяки раціональному поєд-

нанню використання поживних речовин, підбору кращих попередників у сівознах, формування біомаси рослин у певні фази їх розвитку на основі збалансованого використання факторів впливу зовнішнього середовища.

Як свідчать результати виконаного дослідження, прийнятних результатів прогнозування такого нестационарного процесу, як урожайність сільськогосподарських культур, можна досягти завдяки застосуванню запропонованої методики, в основу якої покладено композицію регресійних моделей із деревами рішень та байєсівськими мережами.

Серед регресійних моделей урожайності пшениці кращою виявилась модель, отримана за методом градієнтного бустингу. Показники якості на навчальній вибірці такі: *RASE* = 3,63, *SSE* = 475,25, *MAX E* = 10,58; на перевіірочній вибірці: *RASE* = 4,94, *SSE* = 366,78, *MAX E* = 13,04. Серед байєсівських мереж кращими за показником точності класифікації виявились такі: для кукурудзи – модель, побудована за вихідними даними, переведеними в інтервальний вимір за методом рівних інтервалів (точність класифікації – 70,98 %); для озимої пшениці – мережа, побудована за вихідними даними, переведеними в інтервальний вимір за методом Optimal Bining (точність класифікації – 78,43 %).

У подальших дослідженнях для прогнозного моделювання доцільно застосовувати комплекс різнотипних моделей, зокрема регресійні моделі, мережі Байєса, в т.ч. неперервні, гібридні та динамічні; доповнити дослідження моделями іншого типу, наприклад методом групового врахування аргументів, нейронними мережами та іншими методами і моделями інтелектуального аналізу даних.

Список літератури

1. *Ресурсний потенціал аграрної сфери: проблеми та завдання ефективного використання* / О.В. Собкевич, В.М. Русан, А.Д. Юрченко, О.В. Ковальова – К.: НІСД, 2013. – 76 с.
2. *Національна парадигма сталого розвитку України* / За заг. ред. академіка НАН України, д.т.н., проф. Б.Є. Патона. – К.: ДУ “Інститут економіки природокористування та сталого розвитку Національної академії наук України”, 2012. – 72 с.
3. *Офіційний сайт “Україна-НАТО”. Програма НАТО “Наука заради миру та безпеки”*. – Режим доступу: <http://ukraine-nato.mfa.gov.ua/ua/about-nato/science-for-peace>
4. *Офіційний сайт “Group on Earth observations”*. – Режим доступу: <https://www.earthobservations.org/geoss.php>
5. *Програмно-технічний комплекс Control Agro*. – Режим доступу: <http://control.ua/index.php/reseniya/selskoe-khozyajstvo/control-agro>
6. *Регрессионные модели оценки урожайности сельскохозяйственных культур по данным MODIS* / Н.Н. Куссуль, А.Н. Кравченко, С.В. Скакун и др. // *Современные проблемы дистанционного зондирования Земли из космоса*. – 2012. – 9, № 1. – С. 95–107.
7. *Космічний простір на службі аграрія*. – Режим доступу: <http://www.agroprof.com.ua/statti/1115-kosmichnij-prostir-na-sluzhbi-agrarija.html>.
8. *Офіційний сайт корпорації SAS*. – Режим доступу: <http://www.sas.com>
9. *Офіційний сайт державної служби статистики України*. – Режим доступу: <http://www.ukrstat.gov.ua/>
10. *Моніторинг навколишнього середовища з використанням космічних знімків супутника NOAA* / За ред. д-ра фіз.-мат. наук, чл.-кор. НАН України, проф. С.О. Довгого; Ін-т телекомунікацій і глобал. інформ. простору НАН України, Нац. аерокосм. ун-т ім. М.Є. Жуковського “ХАІ”. – К.: Пономаренко Є.В., 2013. – 316 с.
11. *Довгий С.О., Бідюк П.І., Трофимчук О.М.* Системи підтримки прийняття рішень на основі статистично-ймовірнісних методів / НАН України, Ін-т телекомунікацій і глобал. інформ. простору. – К.: Логос, 2014. – 418 с.
12. *Бідюк П.І., Терентьев О.М., Просянкін-Жарова Т.І.* Прикладна статистика. – Вінниця: ПП “ТД “Едельвейс і К”, 2013. – 288 с.
13. *Ймовірно-статистичні невизначеності в системах підтримки прийняття рішень* / О. Трофимчук, П. Бідюк, О. Кожухівська, А. Кожухівський // *Вісник НУ “Львівська політехніка”*. – 2015. – № 826. – С. 237–248.
14. *Least angle regression* / E. Bradly, N. Trevor, I. Johnstone, R. Tibshirani // *The Annals of Statistics*. – 2004. – 32, № 2. – P. 407–499. – Available: <http://projecteuclid.org/euclid.aos/1083178935>
15. *Abdi H.* Partial least square regression (PLS regression) // *Encyclopedia for Research Methods for the Social Sciences*. – 2003. – P. 792–795. – Available: <https://www.utdallas.edu/~herve/Abdi-PLS-pretty.pdf>
16. *Friedman J.H.* Greedy function approximation: a gradient boosting machine // *The Annals of Statistics*. – 2001. – 29, № 5. – P. 1189–1232. – Available: <http://www.jstor.org/stable/2699986>
17. *Бідюк П.І., Кузнецова Н.В.* Основні етапи побудови і приклади застосування мереж Байеса // *Системні дослідження та інформ. технології*. – 2007. – № 4. – С. 26–39.
18. *Classification and Regression Trees* / L. Breiman, J. Friedman, R. Olshen, C. Stone. – Florida: CRC Press, 2000. – 360 p.
19. *Сайт BayesFusion, LLC*. – Режим доступу: <http://www.bayesfusion.com/genie-modeler>

References

- [1] O.V. Sobkevich et al., *Resource Potential of Agrarian Sphere: Problems and Tasks of Effective Use*. Kyiv, Ukraine: NISS, 2013 (in Ukrainian).
- [2] *National Paradigm of Sustainable Development of Ukraine*, B. Paton, Ed. Kyiv, Ukraine: Institute of Environmental Economics and Sustainable Development, 2016 (in Ukrainian).
- [3] *The Official Website “Ukraine-NATO. Program NATO “Science for Peace and Security”* [Online]. Available: <http://ukraine-nato.mfa.gov.ua/ua/about-nato/science-for-peace>
- [4] *The Official Website of the Group on Earth Observations* [Online]. Available: <https://www.earthobservations.org/geoss.php>
- [5] *Program-Technic Complex Control Agro* [Online]. Available: <http://control.ua/index.php/reseniya/selskoe-khozyajstvo/control-agro>
- [6] N.N. Kussul et al., “Regression models for assessing crop yields according to the MODIS”, *Sovremennye Problemy Distantionnogo Zondirovaniya Zemli iz Kosmosa*, vol. 9, no. 1, pp. 95–107, 2012 (in Ukrainian).
- [7] *Outer Space in the Service of Agrarian* [Online]. Available: <http://www.agroprof.com.ua/statti/1115-kosmichnij-prostir-na-sluzhbi-agrarija.html>
- [8] *The Official Website of SAS* [Online]. Available: <http://www.sas.com>

- [9] *The Official Website of State Statistics Service of Ukraine* [Online]. Available: <http://www.ukrstat.gov.ua/>
- [10] S.O. Dovgiy *et al.*, *Environmental Monitoring Using Satellite Images of Satellite NOAA*. Kyiv, Ukraine: Ponomarenko EV, 2013 (in Ukrainian).
- [11] S.O. Dovgiy *et al.*, *Decision Support Systems Based on Statistical and Probabilistic Methods*. Kyiv, Ukraine: Logos, 2014 (in Ukrainian).
- [12] P.I. Bidyuk *et al.*, *Applied Statistics*. Vinnytsya, Ukraine: PP TD Edelweiss & Co, 2013 (in Ukrainian).
- [13] O. Trofymchuk *et al.* "Probabilistic and statistical uncertainties in the Decision Support Systems", *Visnyk NU "L'viv's'ka Politekhnikha"*, no. 826, pp. 237–248, 2015 (in Ukrainian).
- [14] E. Bradley *et al.*, "Least angle regression", *The Annals of Statistics*, vol. 32, no. 2, pp. 407–499, 2004. Available: <http://projecteuclid.org/euclid.aos/1083178935>. doi:10.1214/009053604000000067
- [15] H. Abdi, "Partial least square regression (PLS regression)", *Encyclopedia for Research Methods for the Social Sciences*, pp. 792–795, 2003. Available: <https://www.utdallas.edu/~herve/Abdi-PLS-pretty.pdf>. doi: 10.1002/wics.051
- [16] J.H. Friedman "Greedy function approximation: a gradient boosting machine", *The Annals of Statistics*, vol. 29, no. 5, pp. 1189–1232, 2001. Available: <http://www.jstor.org/stable/2699986>
- [17] P.I. Bidyuk and N.V. Kuznetzova. "The basic stages of building and examples of using Bayesian Networks", *Systemni Doslidzhennya ta Informatsiyni Tekhnolohiyi*, no. 4, pp. 26–39, 2007 (in Ukrainian).
- [18] L. Breiman *et al.*, *Classification and Regression Trees*. Florida: CRC Press, 2000.
- [19] *Website of BayesFusion, LLC* [Online]. Available: <http://www.bayesfusion.com/genie-modeler>

П.І. Бідюк, О.М. Терентьев, Т.І. Просянкина-Жарова, В.В. Ефендієв

ПРОГНОЗНЕ МОДЕЛЮВАННЯ НЕЛІНІЙНИХ НЕСТАЦІОНАРНИХ ПРОЦЕСІВ У РОСЛИННИЦТВІ З ВИКОРИСТАННЯМ ІНСТРУМЕНТІВ SAS ENTERPRISE MINER

Проблематика. Питання забезпечення зростання виробництва основних сільськогосподарських культур України за умов раціонального природокористування потребує застосування сучасних наукових підходів. Складність вирішення цієї проблеми полягає у відсутності практичного досвіду застосування сучасних інформаційно-аналітичних систем, у яких були б реалізовані одночасно різні методики аналізу та моделювання нелінійних нестационарних процесів у рослинництві. Пропонована методика має перевагою використання інструментів SAS Enterprise Miner – програмного забезпечення, у якому реалізовано широкий спектр методів, які, як показало виконане дослідження, доцільно застосовувати для прогнозного моделювання основних сільськогосподарських культур.

Мета дослідження. Метою роботи є дослідження застосування інтегрованих методів аналізу за використання інструментів SAS Enterprise Miner для прогнозного моделювання нестационарних процесів у рослинництві.

Методика реалізації. Для розв'язання поставлених задач використано методи: системного аналізу, регресійного аналізу, градієнтного бустингу, ймовірнісного моделювання, побудови дерев рішень. Запропоновано методику для розробки прогнозів урожайності сільськогосподарських культур за умови впливу різних груп факторів, обґрунтовано можливість їх використання у системах підтримки прийняття рішень аграрного спрямування.

Результати дослідження. На основі аналізу праць вітчизняних і закордонних вчених запропоновано удосконалення методики розробки прогнозів урожайності основних сільськогосподарських культур із використанням інтегрованих методів аналізу, реалізованих у системі SAS Enterprise Miner. Виконано аналіз отриманих результатів.

Висновки. За допомогою розробленої методики виконано прогнозування урожайності озимої пшениці та кукурудзи для зони лісостепу України. Застосовано різні методи побудови моделей для прогнозування цих нестационарних процесів, обґрунтовано вибір найвірогіднішого з них. Для автоматизації процесу вибору оптимальної моделі для прогнозування урожайності досліджуваних культур використано сучасні інформаційні технології, зокрема SAS Enterprise Miner.

Ключові слова: нестационарний процес; регресійна модель; урожайність сільськогосподарських культур; прогнозування; система підтримки прийняття рішень; SAS Enterprise Miner.

П.И. Бидюк, А.Н. Терентьев, Т.И. Просянкина-Жарова, В.В. Ефендиев

ПРОГНОЗНОЕ МОДЕЛИРОВАНИЕ НЕЛИНЕЙНЫХ НЕСТАЦИОНАРНЫХ ПРОЦЕССОВ В РАСТЕНИЕВОДСТВЕ С ИСПОЛЬЗОВАНИЕМ ИНСТРУМЕНТОВ SAS ENTERPRISE MINER

Проблематика. Вопрос обеспечения роста производства основных сельскохозяйственных культур Украины в условиях рационального природопользования требует применения современных научных подходов. Сложность решения данной задачи связана с отсутствием практического опыта применения в сельском хозяйстве современных информационно-аналитических систем, в которых одновременно использовались бы разные методики анализа и моделирования нелинейных нестационарных процессов в растениеводстве. Предлагаемая методика обладает преимуществом использования инструментов SAS Enterprise Miner – программного обеспечения, в котором реализован широкий спектр методов, которые, как показало выполненное исследование, целесообразно применять для прогнозного моделирования основных сельскохозяйственных культур.

Цель исследования. Целью работы является исследование применения интегрированных методов анализа с использованием инструментов SAS Enterprise Miner для прогнозного моделирования нестационарных процессов в растениеводстве.

Методика реалізації. Для рішення поставлених задач використані методи: системного аналізу, регресійного аналізу, градієнтного бустинга, вірогідного моделювання, дерев'яв рішень. Предложена методика для прогнозування урожайності сільськогосподарських культур в залежності від впливу різних груп факторів, обґрунтована можливість її застосування в системах підтримки прийняття рішення, використовуваних в аграрному секторі.

Результати дослідження. В результаті аналізу робіт вітчизняних і зарубіжних авторів пропонується удосконалення методики прогнозування урожайності основних сільськогосподарських культур з використанням інтегрованих методів аналізу, реалізованих в системі SAS Enterprise Miner. Проаналізовані отримані результати.

Висновки. С допомогою розробленої методики виконано прогнозування урожайності озимої пшениці і кукурузи для зони лесостепі України. Застосовані різні методи побудови моделей для прогнозування цих нестационарних процесів, обґрунтовано вибір найбільш вірогідного з них. Для автоматизації процесу вибору оптимальної моделі для прогнозування урожайності досліджуваних культур використані сучасні інформаційні технології, зокрема SAS Enterprise Miner.

Ключові слова: нестационарний процес; регресійна модель; урожайність сільськогосподарських культур; прогнозування; система підтримки прийняття рішень; SAS Enterprise Miner.

Рекомендована Радою
Інституту прикладного системного
аналізу КПІ ім. Ігоря Сікорського

Надійшла до редакції
22 грудня 2016 року