

МЕТОДИКА ВИБОРУ МАТЕМАТИЧНОЇ МОДЕЛІ ЕКОЛОГІЧНОГО ПРОЦЕСУ

В.В. МІКУЛІН

Анотація. У XXI ст. пошук підходів до вирішення екологічних проблем зумовлений загрозами існування навколишнього середовища внаслідок провадження різноманітних видів діяльності та бездіяльності людини. Більшу половину земної поверхні змінено людиною. Ця модифікація називається зміною в землекористуванні. Нелінійності у змінах у землекористуванні можна вивчати за допомогою інструментів інтелектуального аналізу даних. Запропоновано розглядати більш детально три моделі для зміни землекористування: штучні нейронні мережі, методи класифікації і регресії за допомогою побудови дерева рішень і багатовимірні адаптивні регресійні сплайни. Порівнюються результати трьох інструментів інтелектуального аналізу даних.

Ключові слова: множинна класифікація, штучні нейронні мережі, багатовимірні адаптивні регресійні сплайни, множинна детермінація, оцінка правдоподібності, комірка, зміни у землекористуванні.

ВСТУП

В Україні дедалі більш актуальною проблемою стає охорона навколишнього середовища. Показники екологічного стану в Україні наближаються до критичних значень за результатами аудита ООН. Для України екологічна безпека стає одним із перших пріоритетів формування національної політики. Однією з необхідних умов сучасності в Україні є реконструкція людської і господарської діяльності з урахуванням екологічних факторів. Тільки за допомогою екологічного аналізу ситуації діючих об'єктів діяльності та об'єктів, які перебувають на стадії проектування, можна досягти позитивних змін. Актуальність цієї теми зумовлено загрозами існування навколишнього середовища внаслідок провадження різноманітних видів діяльності та бездіяльності людини.

Натепер у галузі промисловості екологія розглядається як важливий аспект майже всіх нових проектів розвитку виробництва. Розвиток екологічного менеджменту як наукового напрямку спонукає до збільшення кількості методів і прийомів, які можуть бути застосовані як інструменти для поліпшення екологічних параметрів кожного виробничого процесу.

У сучасному підході до удосконалення технологічних процесів для контролю екологічних параметрів дедалі частіше використовують матема-

тичні методи прогнозування. Актуальність розроблення таких методів полягає у передбаченні поведінки екологічних параметрів технологічних процесів з метою прогнозування потенційної небезпеки впровадження технологій на виробництві.

З огляду на потенційні екологічні наслідки зміни клімату основне завдання сьогодення — оцінювання динамічних ефектів зовнішнього впливу клімату на флуктуацію. Це не є тривіальним завданням, оскільки клімат може взаємодіяти з іншими факторами, такими, як густота населення, ступінь міжвидової конкуренції, вплив випасу і людського ставлення до фізичного забруднення навколишнього середовища. Хоча відносна значущість внутрішніх факторів і змін зовнішнього навколишнього середовища у визначенні динаміки популяцій тварин уже давно є центральною темою в екології, інтерактивний вплив клімату і густота населення, як правило, ігноруються. Взаємодія між кліматом та іншими факторами може призвести до появи складної динаміки, що не належним чином моделюється лінійними системами. Виконано огляд відомих робіт із цієї тематики. Сьогодні існує декілька типів моделей, які успішно використовуються в прогнозуванні екологічних процесів.

Одним з інструментів, які можуть бути використані для цієї мети, є штучні нейронні мережі (ШНМ) [1]. Як об'єкт моделювання за допомогою ШНМ був викид газоподібного діоксиду сірки, постійний моніторинг якого проводився у місті Бор (Сербія). Крім того, для процедури моделювання використовувалися метеорологічні параметри: сезон, швидкість вітру, напрямки вітру, температура повітря, вологість і атмосферний тиск.

Ці дослідження дали змогу сформувати математичну модель, яка дозволить прогнозувати надмірну емісію діоксиду сірки з достатньою точністю, на підставі записаних метеорологічних і технологічних параметрів, отриманих з виробництва, зокрема мідеплавильного комплексу. Ця модель не може бути використана для поліпшення технологічного процесу, однак вона придатна для непрямого захисту від впливу на громадян підвищеного вмісту діоксиду сірки в повітрі.

Описана математична модель ґрунтується на прогнозуванні динаміки метеорологічних і технологічних параметрів, впливає з плану виробництва мідеплавильного комплексу і може бути використана для прогнозування потенційного надмірного викиду діоксиду сірки в повітрі. Таким чином, за ті дні року, коли передбачається надлишкова емісія, можна скоригувати план виробництва, у тому числі щодо зменшення кількості концентрату, що підлягає обробленню та збільшенню виробництва сірчаної кислоти. Ці поправки можуть бути виконані ітеративно до передбаченого вмісту діоксиду сірки, що не знижується менше від його граничного значення.

Користувачі таких моделей прагнуть отримати якомога точніших прогнозів і дедалі частіше потребують формальних оцінок невизначеності в модельних прогнозах для обґрунтованого прийняття рішень та управління ризиками. Одним із нетривіальних інструментів нелінійного моделювання є байєсівські методи формування логічного висновку. Такий підхід запропоновано до розроблення моделі стохастичного процесу, у якому екофізіологічні властивості планктонних груповань виражені авторегресійними випадко-

вими процесами [2]. Цей підхід відображає вплив змін у планктонних групуваннях з плином часу і застосовується до тематичного дослідження в океані на станції Папа з використанням методу Монте-Карло для ланцюгів Маркова. Отримані результати свідчать про те, що, спираючись на об'єктивну попередню ретроспективу, можна виокремити корисну інформацію про стан моделі та підмножини параметрів, і навіть зробити корисні довгострокові прогнози на підставі рідкісних і зашумлених спостережень.

Формулювання байєсівської моделі дасть змогу інтегрувати вимірювання за допомогою стандартних спостережень змінних стану в межах послідовної і строгої структури логічного висновку.

Для множинної класифікації використовують методи класифікації і регресії за допомогою регресійного дерева рішень (РДР) і багатовимірних адаптивних регресійних сплайнів (БАРС), які свідчать про їх широке застосування для моделювання складних екологічних процесів. Насамперед ці моделі потрібні для того, щоб переконатися, що правила зупинення не заважають моделі вилучати правильні дані в ході тренувального прогону. Отже, модель спрощується через зменшення складності моделі і видалення розривів моделі, які не роблять її більш точною. Також для множинної класифікації використовується модель трансформації землекористування, яка має дві особливості на відміну від початкової структури. По-перше, змінено початкову структуру моделі ШНМ, яка використовується у підходах до змін у землекористуванні для бінарної класифікації, де кількість вузлів у вихідному шарі відповідає кількості бажаних результатів. По-друге, використано стратегію кодування «один за класом» для вихідного шару, який містить комбінацію k -двійкових чисел для подання атрибутів до категорії, кожна з яких пов'язана з одним із переходів. Щоб визначити робочий стан переходу для кожного класу землекористування, тільки один з k чисел у вихідному шарі має бути кодований як один, а решта k чисел дорівнюють нулю. Усі вузли у вихідному шарі кодуються як нуль, якщо використання землі зберігається між двома кроками. Модель множинної класифікації у змінах у землекористуванні дозволяє користувачу визначити входи, приховані блоки і вихідні блоки.

Мета роботи — вибір із трьох математичних моделей (ШНМ, РДР, БАРС) кращої. Одна із запропонованих моделей ґрунтується на формальному і неформальному виборі моделі у випадку неповноти вхідних даних. Друга обрана модель прийнятна до оцінювання моделей і ґрунтується на байєсовських критеріях вибору моделі, що показують задовільні результати за відомого розподілу помилок унаслідок моделювання процесів. Третя модель оцінювання екологічних процесів — дедуктивне моделювання (від загального до часткового) математичного сподівання, дисперсії та густонаселеності, яке використовувалося в автоматизованому оцінюванні фінансових процесів.

ПОСТАНОВКА ЗАВДАННЯ

Розробити методику побудови математичної моделі та її апробації на фактичних даних стосовно екологічних процесів. За допомогою обраних трьох

математичних моделей (ШНМ, РДР, БАРС) проаналізувати та дослідити розглянутий підхід до моделювання екологічних процесів.

МЕТОДИКА ПОБУДОВИ МОДЕЛІ ЕКОЛОГІЧНОГО ПРОЦЕСУ

Виділимо етапи побудови моделі екологічного процесу: 1) аналіз досліджуваних процесів; 2) підготовку даних для побудови моделі (нормування, заповнення пропусків, екстремальні значення, фільтрацію); 3) оцінювання структури і параметрів моделі; 4) аналіз адекватності моделі.

Аналіз досліджуваного процесу — це фундаментальний етап побудови моделей різних процесів. Для його виконання дослідник повинен мати реальний досвід практичної побудови процесів різної природи. Етап аналізу процесу полягає у виконанні таких завдань [6, 7, 8]:

- а) визначення розмірності (кількості входів і виходів) моделі процесу;
- б) визначення логічних зв'язків між змінними досліджуваного процесу та аналіз можливості їх математичного опису;
- в) визначення кількості зовнішніх збурень та їх типу (детерміноване чи стохастичне);
- г) установлення потенційної можливості виділення окремих підпроцесів (декомпозиція), які є простішими як у функціонуванні, так і у математичному описі. Однак такий процес розподілу є досить складним, виконання якого потребує застосування спеціальних математичних методів;
- д) визначення функції рівнів і типів зв'язків між ними у випадку ієрархічної структури досліджуваного процесу. У технологічних процесах раціонально будувати дерево ієрархії, що дозволить поділити процес на два і більше рівнів функціонування і керування;
- е) використання бази знань, нагромадженої за рахунок наукових досліджень та їх результатів щодо особливостей життєвого циклу процесу, а також застосування напрацьованих законів і закономірностей цього процесу, виявлення досвіду теоретичного чи експериментального дослідження;
- ж) визначення переваг, недоліків, можливостей подальшого удосконалення досліджуваного процесу за наявності розроблених моделей. Використання робочих моделей дозволить оптимізувати процес побудови і застосування моделі.

Підготовка даних для побудови моделі

Метою підготовки експериментальних даних для побудови моделі є зведення даних до зручної форми для визначення закономірностей у досліджуваних процесах. У прикладному сенсі оброблення даних спрямовано на виявлення відмінності між відповідними показниками та оцінювання достовірності такої відмінності. Оскільки дані фактично є випадковими величинами, у процесі дослідження може виникнути недостовірний результат моделювання процесу, тобто різний результат за однакових вхідних умов.

Процес попереднього оброблення експериментальних даних включає: а) нормування — логарифмування даних або зведення до зручного діапазону їх зміни; б) коригування даних — заповнення пропусків та зменшення екстремальних значень, яких явно немає в основному інтервалі значень

змінних; в) формування перших або різниць вищих порядків, необхідних для аналізу відповідних складових часового ряду.

Оцінювання структури і параметрів моделі

На цьому етапі оцінюються коефіцієнти (параметри) та структури моделі [7, 8]. Параметри узагальненої лінійної моделі в основному оцінюються за допомогою методу найменших квадратів (МНК) з використанням принципу економії або збереження. Цей принцип обмежує використання кількості коефіцієнтів для їх оцінювання, що не перевищує їх необхідної кількості. Така кількість формується на основі потреби у збереженні статистичних характеристик процесу.

Найпоширенішими методами оцінювання параметрів моделі є такі: МНК; метод максимальної правдоподібності (ММП); метод допоміжної (інструментальної) змінної (МДП); нелінійний метод найменших квадратів (НМНК) та їх рекурсивні версії (РМНК, РММП, РМДП). Для отримання незміщених, консистентних та ефективних оцінок вектора параметрів θ лінійної регресійної математичної моделі, наприклад, моделі змішаної регресії

$$y(k) = a_0 + a_1 y(k-1) + a_2 y(k-2) + b_1 x(k) + b_2 z(k) + \varepsilon(k),$$

за допомогою МНК необхідно задовольнити такі умови: $\varepsilon(k)$ — некорельована послідовність випадкових чисел з нульовим середнім; $E[\varepsilon(k)] = 0$,

$$\text{cov}[\varepsilon(k)] = E[\varepsilon(k)\varepsilon(j)] = \begin{cases} \sigma_\varepsilon^2, & k = j \\ 0, & k \neq j \end{cases}, \text{ тобто послідовності } \varepsilon(k) \text{ і } y(k) \text{ не}$$

повинні бути корельовані між собою. Зазначимо, що перевірити виконання наведених умов можна тільки після оцінювання коефіцієнтів моделі, а до оцінювання можна лише постулювати їх виконання [9].

Діагностика моделі

На цьому етапі оцінюється адекватність процесу моделювання шляхом перевірки оцінених кандидатів. Аналіз ґрунтується на оцінюванні таких параметрів:

1. Коефіцієнт множинної детермінації R^2 , який обчислюється таким чином:

$$R^2 = \frac{\text{var}(\tilde{y})}{\text{var}(y)} = 1 - \frac{SSE}{SST},$$

де $\text{var}(\tilde{y})$ — дисперсія залежної змінної, оціненої за допомогою побудованої моделі; $\text{var}(y)$ — дисперсія вимірів залежної змінної; $SSE =$

$$= \sum_{k=1}^N [y(k) - \tilde{y}(k)]^2 \text{ — сума квадратів похибок (залишків) моделі (sum of}$$

$$\text{squared errors); } SST = \sum_{k=1}^N [y(k) - \bar{y}]^2 \text{ — загальна сума квадратів (total sum of}$$

$$\text{squares); } \bar{y} \text{ — середнє значення; } SST = SSE + SSR, \text{ де } SSR = \sum_{k=1}^N [\tilde{y}(k) - \bar{y}]^2$$

— загальна сума квадратів для регресії (sum of squares for regression). Найкращим значенням є $R^2 = 1$, тобто коли дисперсії вимірів змінної та цієї ж змінної, оціненої за рівнянням, збігаються [9].

2. Сума квадратів похибок для вибраної моделі повинна бути мінімальною, тобто $\sum_{k=1}^N e^2(k) = \sum_{k=1}^N (\tilde{y}(k) - y(k))^2 \rightarrow \min(\tilde{y})$ порівняно з усіма іншими моделями [9].

3. Похибки моделі не повинні бути корельовані між собою. Корельованість похибок визначають за статистикою Дарбіна–Уотсона DW , яка розраховується за формулою $DW = 2 - 2\rho$, де $\rho = E[e(k)e(k-1)] / \sigma_e^2$ — коефіцієнт кореляції між сусідніми в часі значеннями похибки; σ_e^2 — дисперсія послідовності похибок $\{e(k)\}$. Таким чином, за повної відсутності кореляції між похибками $DW = 2$ — це ідеальне значення. Граничними значеннями для DW є 0 (якщо $\rho = 1$) і 4 (якщо $\rho = -1$) [9].

Приклад побудови математичних моделей екологічного процесу

Зміни у землекористуванні являють собою складний процес і моделювання цих систем є теж досить складним. Добре відомо, що чинники впливу на зміни у землекористуванні працюють у різних просторово-часових масштабах нелінійним чином і, отже, для моделювання цієї динаміки потрібні нелінійні інструменти. Нелінійні методи часто застосовуються у моделюванні навколишнього середовища, такі як ШНМ, що набули поширення протягом останніх двох десятиліть. Багатовимірні адаптивні регресійні сплайни і РДР теж відомі моделі, зокрема РДР обчислює ймовірність результатів з використанням декількох просторових збудників, щоб показати монотонні результати; БАРС долає обмеження кусково-постійних функцій у РДР, генеруючи кусково-лінійні моделі з використанням базисних функцій.

Однією з проблем моделювання є те, що в межах обраного регіону відбувається кілька змін у землекористуванні. Це характерно для деяких регіонів, де міські ліси будуть перетворені із земель сільського господарства, у той час, як сусідні перетворюються в сільськогосподарські. Проте деякі дослідники розглядали кілька переходів у землекористуванні в одній і тій же моделі і таким чином спрощували процес зміни у землекористуванні. Під час моделювання або імітації більш ніж один результат часто створює проблему МК. Кілька підходів були розроблені протягом останніх трьох десятиліть для імітації змін у землекористуванні, використовуючи численні змінні навколишнього середовища. Однак більшість з них обмежує їх застосування тільки для одного переходу в землекористуванні. Наприклад, модель трансформації землі (МТЗ), яка являє собою ШНМ, використовувалась у різних місцях по всьому світу для моделювання тільки однієї зміни. У роботі пропонуються різні стратегічні схеми кодування, які можна застосувати для зміни своїх моделей математичної класифікації (МК). В аналізі екологічних процесів використовуються три інструменти інтелектуального аналізу даних (ШНМ, РДР і БАРС) для моделювання МК.

Моделі МК повинні фіксувати кожен комірку на карті до унікального класу землекористування; проте часто комірки можна віднести до більш ніж одного класу землекористування у МК (неоднозначне прогнозування). Деякі моделі можуть комунікувати з декількома класами землекористування, таких як комірчасті автоматичні моделі, моделі логістичної регресії, агентсько-орієнтовані моделі і ШНМ. У деяких з них використовується простий ієрархічний підхід до МК, у той час, як в інших — конкурентний підхід до розподілу землі. Дослідження, запропоновані вирішити різними способами МК, не змогли виявити ефективного правила для усунення проблем у моделюванні МК. Інша мета дослідження — запропонувати простий метод для вирішення проблем шляхом усунення неоднозначних прогнозів.

Таким чином, існує необхідність порівняти і протиставити нелінійні інструменти для моделювання МК, визначити методи, які працюють краще ніж МК, і дослідити їх виконання для умов, коли кілька переходів у землекористуванні відбуваються в одному місці. Порівнюються три відомі підходи інтелектуального аналізу даних (ШНМ, РДР і БАРС) для моделювання МК з метою виявлення закономірностей використання земель на Середньому Заході США.

У дослідженні поєднуються три нелінійні інструменти інтелектуального аналізу даних для імітації більш як одного класу для двох різних регіонів Середнього Заходу США. Метою цієї роботи є: 1) запропонувати різні стратегії кодування схеми трьох інструментів для кількох переходів з одного стану в інший стан землекористування, розроблених для однієї зміни; 2) розробити ефективні правила для подолання неоднозначних прогнозів МК; 3) порівняти кілька переходів у землекористуванні трьох методів інтелектуального аналізу даних один з одним щодо їх потенціалу для сільського господарства, лісів та моделювання міського землекористування у басейні річки Маскегон (*MRV*) та у Південно-східному Вісконсіні (*CEVI*), використовуючи відносну робочу характеристику (*ROC*) і відсоток правильних порівнянь (*PCM*); 4) зробити висновки про моделі змін у галузі землекористування на Середньому Заході США.

Моделювання РДР. Вузли А, Б, В у дереві містять комірки: міську територію, ліс, землю сільськогосподарського призначення відповідно.

Комірка, яка задовольняє умову у вузлах, переходить по лівому плечу, або перехід відбувається по правому плечу. Однак вузли Г (рис. 1) містять комірки — інші переходи у землекористуванні, або які залишаються у тому самому класі землекористування. Відстані до лісу і сільського господарства в *CEVI* та відстані до сільського господарства, доріг і лісів у *MRV* є найбільш значущими для міських умов, що імітують приріст лісу і сільського господарства одночасно (рис. 1, а, б). За розрахунком приросту визначається якість розмежування (чим більший бал, тим вища якість). Відношення дисперсії до розмежування у міру спадання наведено в табл. 1.

За відстані до сільськогосподарських розмежувань 75 м приріст становить 0,065, що набагато нижчий, ніж за основної відстані розмежувань 63 м (за відстані до лісу приріст — 0,094) у *CEVI*. Відстань до чагарника є найкращою змінною (51 м), приріст — 0,064 як головний показник у *MRV* (табл. 1).

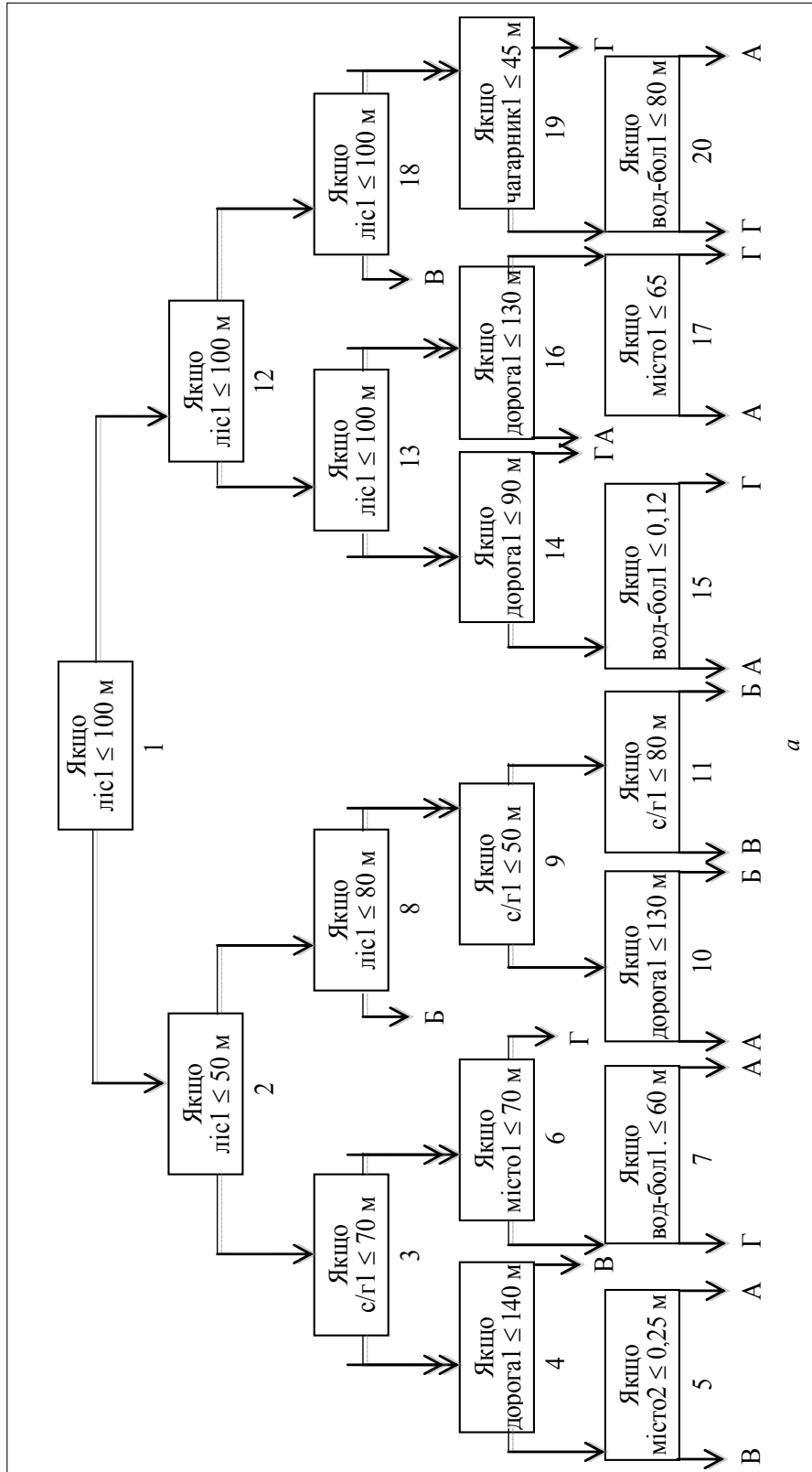
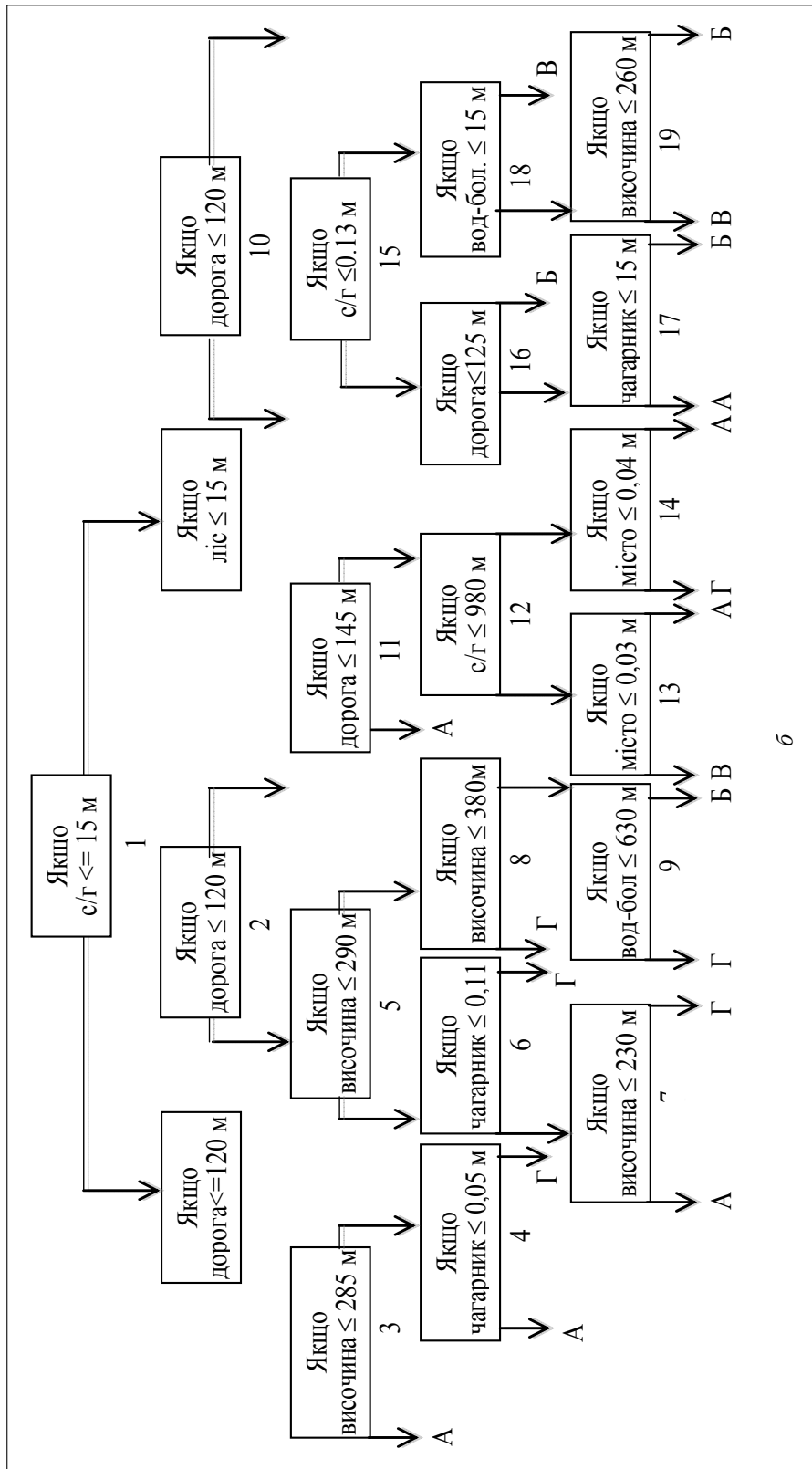


Рис. 1. Візуалізація основних розгалужувачів BARC: дерево рішень у SEVI (a) і MRV (б) і MRV (б); вод-бол — водянисто-бологиста місцевість; с/г — сільськогосподарська земля. (Див. також с. 123)



б

Рис. 1. Закієчення

Комірки, розміщені в межах А, Б і В вузлів з великими значеннями придатності мають найбільшу ймовірність існування міських, лісових і приросту сільськогосподарських земель відповідно (рис. 1, а, б). У *CEVI* комірки на вузлі 8 (відстань до лісу менша за 80 м), вузол 20 (відстань до заболоченого місця більша ніж 80 м) і вузол 18 (відстань до сільського господарства менша ніж 75 м) найбільш придатні для лісів, міського і сільського господарства відповідно (рис. 1). У *MRV* комірки у вузлі 17 (відстань до чагарника понад 40 м), вузол 14 (щільність міського господарювання більша ніж 0,04) і вузла 18 (відстань до заболоченого місця більша ніж 60 м) мають найбільшу придатність і вигоду для лісового, міського і сільського господарства відповідно (рис. 1, б). Ця процедура триває від вузла зі значеннями більшої придатності до інших вузлів з більш низьким значенням придатності до РДР, де не задовольняється загальна кількість еталонних змін для кожного класу землекористування (табл. 2).

Таблиця 1. Змінна, розмежування і приріст

Крок	Змінні	Розмежування	Приріст
<i>CEVI</i>			
0	Ліс 1	63	0,09451
1	С/г 1	75	0,06582
2	Місто 1	51	0,03752
3	Ліс 2	0,11346	0,02820
4	Дорога 1	121	0,01776
5	Місто 2	0,16735	0,01632
6	Вод-бол 1	51	0,01451
7	С/г 2	0,43037	0,01371
8	Схил 1	3,14572	0,01317
9	Чагарник 1	142	0,00712
10	Парк 1	1499	0,00648
11	Чагарник 2	0,05948	0,00610
12	Вод-бол 2	0,09972	0,00473
13	Височина 1	265	0,00458
14	Вода 1	101	0,00456
15	Потік 1	121	0,00418
16	Аспект 1	0,50	0,00349
<i>MRV</i>			
0	С/г 1	345	0,06587
1	Чагарник 1	51	0,06486
2	Дорога 1	114	0,03823
3	Ліс 1	190	0,03040
4	Місто 2	0,03543	0,01517
5	Місто 1	270	0,01333
6	С/г 2	0,14085	0,00980
7	Височина 1	266	0,00939
8	Вод-бол 1	875	0,00910
9	Чагарник 2	0,13359	0,00829
10	Вод-бол 2	0,04490	0,00374
11	Ліс 2	0,37794	0,00346
12	Потік 1	915	0,00227
13	Парк 1	2772	0,00110
14	Схил 1	1,14232	0,00089
15	Аспект 1	5,58117	0,00056

Таблиця 2. Приріст, часовий інтервал і розмір комірок для *CEBI* і *MRV*

Місцевість, яка вивчається	Часовий інтервал, роки	Приріст с/г земель	Приріст лісу	Приріст міських земель	Інші зміни	Всього	Розмір комірок, м
<i>CEBI</i>	1990–2000	135,709	79,467	491,031	314,265	1020,472	30×30
<i>MRV</i>	1978–1998	120,013	641,237	412,302	693,598	1967,150	30×30

Зміни у землекористуванні мають однакові значення придатності комірок у кожному вузлі. Дослідник здатен знайти вузли через їх сортування від високих до низьких концентрацій для різних класів землекористування. Наприклад, фермери, містобудівники та управляючі природними ресурсами цікавляться розширенням сільського господарства, урбанізацією і збільшенням лісів у певному регіоні (наприклад, вузли 3 і 18 у *CEBI* для збільшення сільського господарства; вузли 10 і 14 у *MRV* для урбанізації і вузли 7 і 11 у *CEBI* для збільшення лісів). Показник придатності відображає кожен вид землі, тобто кожен клас з власною відміткою (інші види землекористування, сільськогосподарські землі, ліси і міські землі — Г, В, Б і А відповідно). Вузли 3, 5, 7 і 14 мають найвищу концентрацію сільськогосподарських, міських, лісових земель та інших переходів землекористування в *CEBI*. Вузли 11, 10, 19 і 13 мають також найвищу концентрацію сільського, міського господарства та збільшення лісового господарства й інших видів переходів у землекористуванні у *MRV* відповідно.

Моделювання БАРС. Змінна з великим стандартним відхиленням (*ANOVA*) краще описує взаємозв'язок між входами і виходами. Відстань до сільського господарства зі стандартними відхиленнями 0,74 і 0,29 показують більший внесок для моделювання переходів у землекористуванні в *CEBI* і *MRV* відповідно (рис. 2).

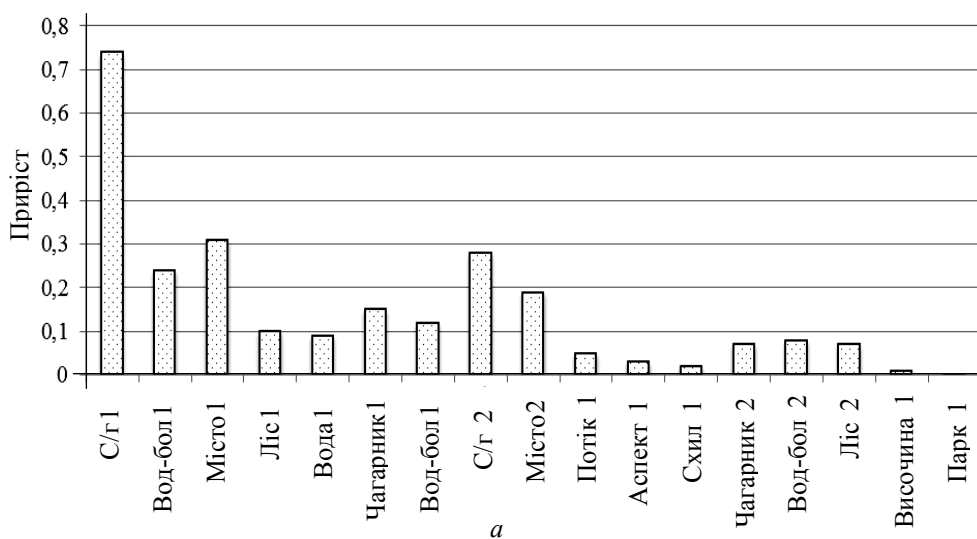


Рис. 2. Дисперсія у БАРС: а — *CEBI*; б — *MRV*. (Див. також с.126)

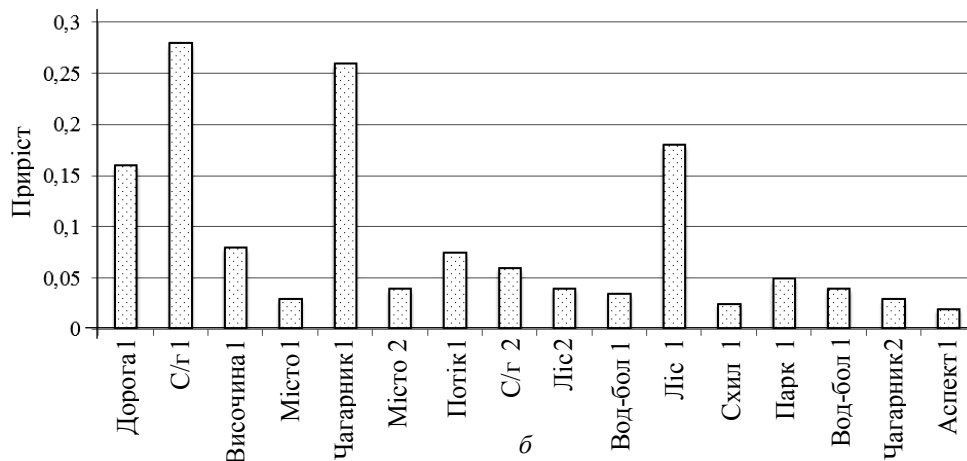


Рис. 2. Закінчення

Відстань до міських земель зі стандартним відхиленням 0,31 і відстані до чагарників зі стандартним відхиленням 0,25 вказують на більший внесок для імітації переходів у землекористуванні в *CEBI* і *MRV* (рис. 2), відстань до чагарників із чотирьох базисних функцій в *CEBI* і підвищенні до 8 базисних функцій у *MRV* містять найбільшу кількість базисних функцій в БАРС (табл. 3).

Таблиця 3. Коефіцієнти, змінні та вузли в БАРС для *CEBI*

<i>N</i>	Коефіцієнт	Змінна	Стан	Вузол
0	-0,984680			
1	-0,061248	С/г 1	-	67
2	0,035816	С/г 1	+	67
3	-0,031057	Місто 1	+	42
4	-0,015252	Вод-бол 1	-	161
5	0,014962	Вод-бол 1	+	161
6	-0,001308	Дорога 1	-	161
7	0,003081	Дорога 1	+	161
8	-0,022361	Вод-бол 1	+	60
9	-0,000101	Вода 1	-	90
10	-0,006427	Вода 1	+	90
11	-0,004736	Чагарник 1	-	366
12	0,004286	Чагарник 1	+	366
13	0,000015	Ліс 1	-	67
14	-0,004473	Ліс 1	+	67
15	0,012819	Місто 1	-	42
16	-3,674577	Місто 2	+	0,7726
17	-1,079093	Місто 2	-	0,7726
18	-3,326859	С/г 2	+	0,0790
19	-2,891896	С/г 2	-	0,0790
20	-0,007839	Чагарник 1	+	134
21	0,000535	Аспект 1	+	145
22	0,000145	Аспект 1	-	145
23	-0,722788	Чагарник 2	+	0,3728
24	-1,050544	Чагарник 2	-	0,3728
25	0,999261	Вод-бол 2	+	0,1020
26	0,056512	Вод-бол 2	-	0,1020

Продовження табл. 3

<i>N</i>	Коефіцієнт	Змінна	Стан	Вузол
27	0.792730	Ліс 2	+	-0.0000
28	-0.006119	Схил 1	-	2.8624
29	-0.027395	Схил 1	+	2.8624
30	0.003083	Чагарник 1	-	174
31	-0.000693	Височина 1	+	267
32	-0.001118	Височина 1	-	267
33	-1.464719	Вод-бол 2	+	0.5492
34	0.002106	Місто 1	+	84
35	-0.000475	Потік 1	+	240
36	0.000135	Потік 1	-	1218
37	0.000110	Вода 1	+	684
38	4.434641	С/г 2	-	0.1045
39	0.000002	Парк 1	+	0.0001

Порівняння моделей інтелектуального аналізу даних

Відповідно до РСМ (рис. 3, а) МТЗ і РДР мали однакову точність і більшу точність, ніж БАРС для імітації міського, сільського господарства і збільшення лісів в обох регіонах.

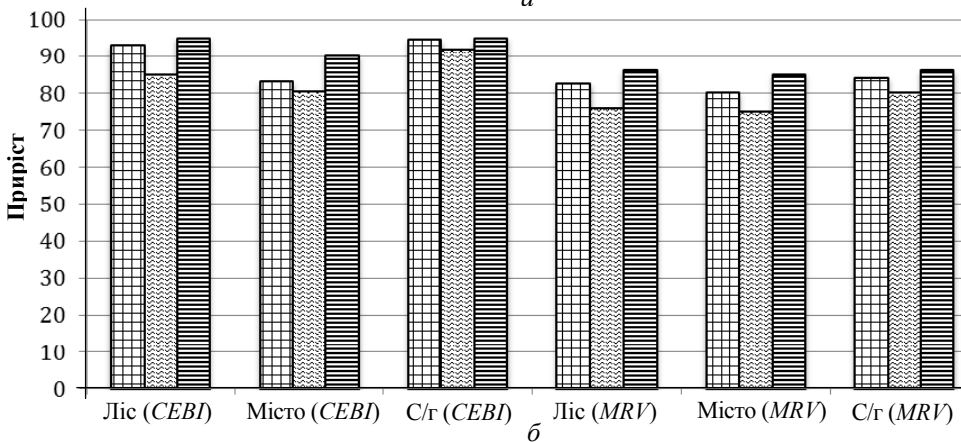
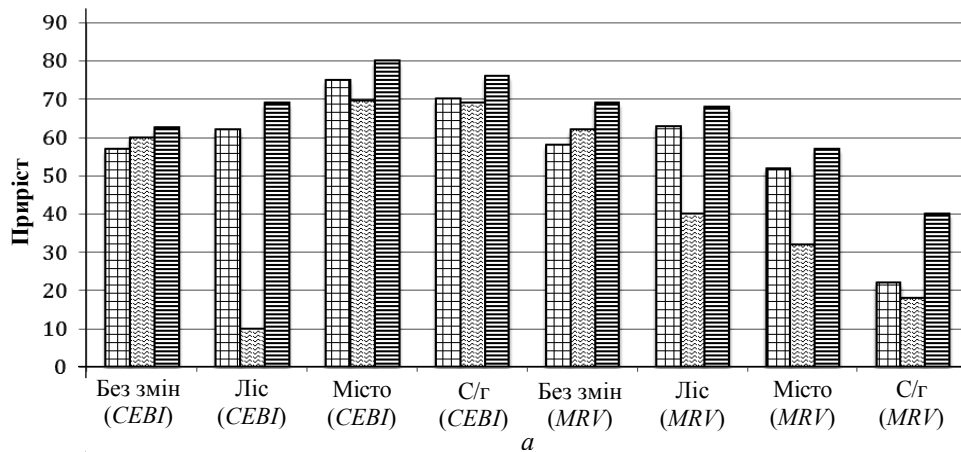


Рис. 3. Розрахований РСМ для незмінних сільськогосподарського, міського та лісового приростів у CEBI і MRV; ROC для моделей (ШНМ, РДР, БАРС) для приростів сільського, міського та лісового господарств у CEBI і MRV

Тільки для моделювання збільшення лісу в *CEBI* відмінність між МТЗ і РДР з БАРС була великою і зумовлена тим, що кілька комірок показали приріст лісів протягом 10 років (близько 7,7%). Відповідно до *ROC* (рис. 3, б) МТЗ і РДР приріст значно перевищив БАРС за даними перевірки. Характеристики *ROC* для трьох моделей були подібні в *CEBI*, проте *ROC* для МТЗ і РДР більші, ніж *ROC* для БАРС у *MRV*.

ВИСНОВКИ

Таким чином, алгоритми класифікації допомагають зрозуміти існуючу структуру даних і можуть бути використані для прогнозування класу землекористування з використанням відповідних методів інтелектуального аналізу даних. Більшість моделей змін у землекористуванні розроблено для бінарного прогнозування. У цій роботі реформуються три нелінійні інструменти інтелектуального аналізу даних для моделювання декількох змін для двох різних регіонів Середнього Заходу США (сільського господарства, лісових і міських земель). Модифіковано методи МТЗ, РДР і БАРС для МК, які спочатку розроблялись для бінарного моделювання змін. Виявлено, що стратегія кодування один-у-класі (наприклад, МТЗ) є більш придатною, ніж стратегія кодування розподіленого висновку (наприклад, РДР і БАРС) для моделювання задач МК. Результати дослідження підтверджують висновок про можливість використання МТЗ, РДР і БАРС для моделювання декількох змін у землекористуванні.

Оцінка моделі для бінарної класифікації проста. Використано таблицю на випадок непередбачених обставин для порівняння модельованих процесів і вибору карт земель у двох категоріях — змінних і незмінних. Проте далеко не всі дослідження були зосереджені на побудові моделі змін землекористування для МК або порівняння різних моделей цих змін для багаторазового їх моделювання. Це дослідження дає змогу імітувати кілька переходів землекористування з використанням нелінійних інструментів. Результати показують, що моделі інтелектуального аналізу даних виконуються досить добре з використанням жорсткої класифікації *PCM* і м'якої класифікацію (*ROC*). Виявлено, що МТЗ дала послідовно вищий ступінь узгодженості ніж РДР і БАРС для кожного класу землекористування. Метод ШНМ має більш високу точність, але мало що пояснює (див. рис. 3). Методи РДР і БАРС передбачають, що найбільше їх значення залежать від чинників, пов'язаних з кожним видом переходу у землекористуванні (див. табл. 1 і 2).

Показано, що зміни у землекористуванні є послідовними у центрально-західній частині США (*CEBI* і *MRV*). Ділянки біля лісу перетворюються в сільське господарство і сільськогосподарські землі, згодом — у міські. Перший період землекористування, як правило, швидший. Неоднозначні передбачення в МК відбуваються через складність моделей змін, коли підходи до інтелектуального аналізу даних не можуть визначати чіткі межі між класами землекористування. Кількість комірок з неоднозначними прогнозами залежить від здатності процедур інтелектуального аналізу даних розмежувати класи землекористування, різні чинники, кількість еталонних змін для кожного класу землекористування і вихідних класів землекористування.

Тут розроблено простий метод для вирішення конфліктних проблем у МК. Це досягається усуненням неоднозначних прогнозів для даних комірок у моделюванні землекористування, яке відображає модельовані карти землекористування з довільними класами землекористування.

Необхідно більш поглиблено вивчати проблематику моделювання кількох переходів землекористування. Моделі землекористування, особливо для МК, генерують як помилковий позитивний результат, так і помилковий негативний результат класу помилок. Отже, використання тільки однієї моделі не може бути ідеальним, вони можуть доповнювати одна одну. Застосування двох або більше потужних інструментів, таких як РДР і БАРС у поєднанні з моделлю ШНМ допомагає зрозуміти базовий складний процес у землекористуванні.

ЛІТЕРАТУРА

1. *Mihajlović I.* Statistical modelling in ecological management using the artificial neural networks (ANNs) / I. Mihajlović, Đ. Nikolić, N. Strbac and Z. Zivković // *Serbian Journal of Management*. — 2010, No 5 (1). — P. 39–50.
2. *Parslow J.* Bayesian learning and predictability in a stochastic nonlinear dynamical model / J. Parslow, N. Cressie, P. Campbell E., E. Jones and L. Murray // *Ecological Applications*. — 2013. — No 23 (4). — P. 679–698.
3. *Verbeke G.* Formal and Informal Model Selection with Incomplete Data / G. Verbeke, G. Molenberghs and C. Beunckens // *Statistical Science*. — 2008. — Vol. 23, No. 2. — P. 201–218.
4. *Bubna K.* Model selection and surface merging in reconstruction algorithms / K. Bubna, C.V. Stewart // *Journal Computer Vision and Image Understanding archive*. — 2000. — Vol. 80, Issue 2. — P. 215–245.
5. *Sucarrat G.* Automated Model Selection in Finance: General-to-Specific Modelling of the Mean, Variance and Density. *Oxford bulletin of economics and statistics* / G. Sucarrat, A. Escribano // *Oxford: Wiley-Blackwell*. — 2012. — Vol. 74. — P. 716–735.
6. *Бідюк П.І.* Аналіз часових рядів : навч. посіб. / П.І. Бідюк, В.Д. Романенко, О.Л. Тимошук. — К: НТУУ «КПІ», 2013. — С. 115–158.
7. *McCullagh P.* *Generalized Linear Models* [Text] / P. McCullagh, J.A. Nelder. — New York: Chapman & Hall, 1990. — 526 p. Doi: 10.1007/978-1-4899-3242-6.
8. *Трухан С.В.* Прогнозування актуальних процесів за допомогою узагальнених лінійних моделей / С.В. Трухан, П.І. Бідюк // *Наукові вісті НТУУ «КПІ»*. — 2014. — № 2. — С. 14–20.
9. *Довгий С.О.* Системи підтримки прийняття рішень на основі статистично-ймовірнісних методів / С.О. Довгий, П.І. Бідюк, О.М. Трофимчук. — К.: Логос, 2014. — 419 с.
10. *Tayyebi A.* Modeling multiple land use changes using ANN, CART and MARS: Comparing tradeoffs in goodness of fit and explanatory power of data mining tools / A. Tayyebi, Bryan C. Pijanowski // *International Journal of Applied Earth Observation and Geoinformation* 28 (2014). — P. 102–116.

Надійшла 27. 10. 2016