

МЕТОДИ ВЕРИФІКАЦІЇ ПРЕДИКАТИВНИХ ҐРУНТОВИХ КАРТ

В. Р. ЧЕРЛІНКА, кандидат біологічних наук, доцент кафедри агротехнологій та ґрунтознавства

Ю. М. ДМИТРУК, доктор біологічних наук, завідувач кафедри агротехнологій та ґрунтознавства

Інститут біології, хімії та біоресурсів

Чернівецький національний університет ім. Ю. Федьковича, Україна

E-mail: v.cherlinka@chnu.edu.ua

Анотація. Аналіз можливих методів верифікації отриманих результатів є необхідним і важливим процесом для винесення обґрунтованих висновків щодо якості проведеного моделювання. На відміну від поширених в літературі рекомендацій щодо процесу верифікації отриманих результатів моделювання ґрунтового покриву суто з математичної сторони, запропоновано, так звану, крос-валідацію, яка дозволяє співставити модельні дані з реальними ґрунтовими обстеженнями та існуючими картами. Це потенційно дає змогу відслідкувати помилки не тільки модельного експерименту, а й «законсервовані» на існуючих ґрунтових картах. Встановлено, що для охоплення всього спектру можливих крос-валідацій польових, картографічних і модельних даних та отримання максимально достовірних і коректних як з математичної, так і ґрунтознавчої сторони результатів, кількість закладених розрізів має відповідати плану великомасштабного ґрунтового обстеження, а схема розташування повинна відповідати репрезентативним точкам місцевості, виділеним на основі геоморфологічного аналізу.

Ключові слова: морфометричні параметри, ЦМР, ґрунтова карта, кореляція, предиктори, прогносні алгоритми, крос-валідація

Актуальність. Для виробництва сільськогосподарської продукції, необхідної для постійно зростаючого населення та задоволення потреб промисловості, виробники у всьому світі зосереджують свої зусилля на максимізації кількості врожаю з одночасним підвищенням його якості та мінімізацією витрат. Єдиний шлях до вирішення цих завдань в 21 столітті — точне землеробство. Проте, застосування точного землеробства традиційно було трудомістким і вартісним для багатьох виробників процесом. У сучасній практиці точного сільського господарства відсутні високоточні та недорогі в реалізації картографічні технології, які б дозволили товаровиробникам керувати варіаціями технологічного характеру відповідно до ґрунтових умов в умовах конкретних ґрунтових ареалів. За допомогою сучасних технологій проксимального і дистанційного зондування та прогнозного моделювання ґрунтових ресурсів такі карти можна створювати

й широко застосовувати для високопродуктивного точного сільського господарства.

Майбутня інтеграцію супутникових даних із даними щодо ґрунтів забезпечить створення точних карт ґрунтів [1–3]. Вони будуть достойною альтернативою поточним методам дослідження земель, а зменшення впливу на навколишнє середовище буде значною мірою досягнуто за рахунок зниження кількості агрохімікатів, що застосовуються. Завершальним продуктом, як це бачиться з позицій сьогодення, буде «прецизійна ґрунтова карта», доступна для товаровиробників у високій роздільній здатності. Вона посилить розуміння потреб рослин в агроєкосистемах з використанням інформації про ґрунти і забезпечить економічний шлях впровадження точного землеробства.

Отже, реалізація пропонованих заходів матиме вплив на економічно та екологічно обґрунтовану інтенсивну сільськогосподарську діяльність, а такий підхід запропонує інноваційний та недосягнутий на сьогодні рівень деталізації для планування поточної діяльності.

Проте, для високопродуктивного агровиробництва, моніторингу та управління якістю довкілля важливо мати всебічну інформацію про ґрунтовий покрив конкретної території. Водночас аналіз літературних джерел показує [4], що великомасштабні ґрунтові обстеження (наприклад, масштабу 1:10 000) проведені не для всіх держав. Так, в Україні, необстеженими залишилося близько 20 з 60 млн. га [5]. Оскільки ретельно досліджувалися тільки землі сільськогосподарського призначення, то території населених пунктів, лісів, гірські регіони в більшості випадків залишилися поза увагою дослідників. Одержані ще в 1957–1961 рр. результати застаріли, в них чимало недоречностей, а корекція цих даних не проводилася понад 25 років [5, 6]. Ситуація в інших державах Європи різна, але аналогічні проблеми зустрічаються нерідко.

Аналіз останніх досліджень та публікацій. Внаслідок помилок за ручного виділення картограм крутизни схилів (в Україні саме на них базуються методи виділення ґрунтів різного ступеня еродованості) наявні похибки, іноді значні, щодо просторового паттерну ґрунтових таксонів. Заповнення прогалів та коригування меж картографічних даних у таких умовах логічне шляхом моделювання. Кількість досліджень, присвячених саме моделюванню хорології таксономічних ґрунтових одиниць наростає [7–13]. При цьому застосовується великий спектр математичних методів: від багатофакторного регресійного аналізу, крігінгу, нейронних мереж до різного роду класифікаційних дерев [14]. Загальна ідея, яка лежить в основі застосування таких методів, полягає у використанні опорних точок ландшафтів з приуроченими до них ґрунтовими таксонами [15], що загалом відповідає концепції великомасштабної ґрунтової зйомки на «ключачах» [16]. Основним джерелом предикторів у такій симуляції є цифрова модель рельєфу (DEM), аналіз якої дозволяє виділити цілий ряд геоморфологічних та пов'язаних з ними параметрів. Оскільки модельовані змінні (різновиди ґрунтів) відносяться не до числового, а категоріального типу даних, а отримані з ЦМР показники, як правило, числові, то використання лише просунутих математичних методів дозволяє встановити неочевидні на перший погляд залежності між усіма цими параметрами [6, 11, 17–20].

Результати моделювання мають, як правило, відхилення від реального стану речей як за рахунок власне моделювання, так і за рахунок ймовірних помилок на самих ґрунтових картах, які беруть за основу для розрахунку вхідних параметрів. Тому аналіз можливих методів верифікації отриманих результатів є необхідним і важливим процесом для винесення обґрунтованих висновків щодо якості проведеного моделювання.

Отже, **метою** нашого **дослідження** в цьому аспекті було дослідження способів верифікації модельованих даних та рекомендації стосовно оптимального поєднання способів верифікації, методики створення навчального набору даних і самого алгоритму побудови моделі. Водночас ми намагалися зробити процес дослідження максимально прозорим та відтворюваним для широкого загалу. Тому ми пропонуємо використовувати великомасштабні ґрунтові та топографічні карти і вільне програмне забезпечення — геоінформаційні системи GRASS [21] і Quantum [22], векторизатор Easy Trace [23] та мову статистичних розрахунків R-statistic [24].

Матеріали і методи досліджень. Згідно мети досліджень нами були визначено такі завдання: а) оцифрування та атрибутування картографічних матеріалів; б) побудова ЦМР з роздільною здатністю 5 м; в) аналіз цифрових моделей рельєфу і виділення з них у GIS GRASS набору карт морфометричних та інших похідних характеристик; г) генерація навчальних вибірок згідно описаних методичних підходів; д) створення в R-statistic симулятивних моделей ґрунтового покриття (карт-версій або карт-моделей) з використанням 14 основних типів предикативних алгоритмів; ж) закладка мережі ґрунтових розрізів і аналіз методів верифікація якості прогнозного моделювання та самої карти ґрунтового покриття.

В якості об'єкта було обрано фрагмент території України в межах Чернівецької області, приурочений до Прут-Дністровського межиріччя. Адміністративно цей ареал відноситься до території міста Чернівці (рис. 1), а при його виборі було вирішено деякі суто українські проблеми, детально описані в ряді праць [6, 25, 26]. Головними з них були проблеми з доступом до топографічних та ґрунтових даних та їх актуальністю, системою координат тощо. Відповідний даній території фрагмент сканованого листа топографічної карти М 1:2 000 був геореєстрований в GIS Quantum [22] з використанням системи координат Pulkovo 1942 CS63 Zone X2 [27].

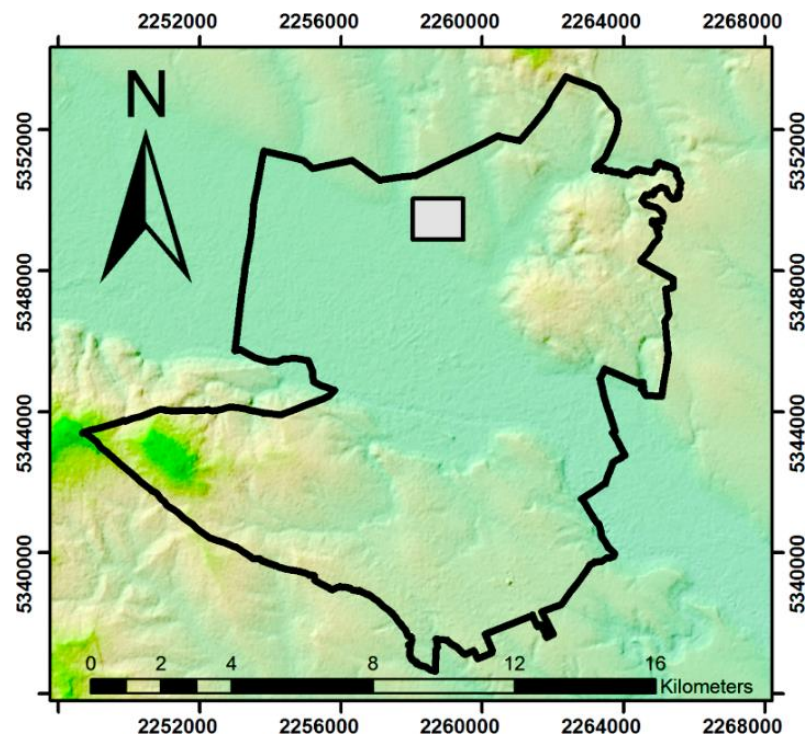


Рис. 1. Ареал досліджень

Інформативні ґрунтові матеріали базувалися на архівній ґрунтовій карті ареалу дослідження (зйомка 1993 р.). Тестовий район було підібрано з можливостями охоплення ґрунтової катени повного складу.

Для опрацювання даних були використані інструментальні можливості вільного програмного забезпечення: оцифрування — Easy Trace [23], підготовка карт морфометричних параметрів — GIS GRASS [21] та побудова предикативної моделі ґрунтового покриття — мова статистичних розрахунків R-statistic [24]. Цифрова модель рельєфу з роздільною здатністю 5 м була побудована за допомогою регуляризованих напружених сплайнів [28] в GIS GRASS, а вибір параметрів сплайну здійснювався згідно рекомендацій [29]. Була обрана саме 5 м роздільна здатність, оскільки при відносно високій точності відтворення топографії вона забезпечує також практичне співпадіння площ векторизованих і растрованих ґрунтів [30].

Цифрова модель рельєфу послужила основою для виділення ряду його морфометричних характеристик: крутизни (slope) та експозиції (aspect) схилів (модуль `r.slope.aspect` [31]), кривизни поверхні (профільної, планової, поздовжньої, мінімальної та максимальної — `profc`, `planc`, `longc`, `minic` та `maxic` відповідно) — модулем `r.param.scale` на основі роботи [32], даних по сонячній радіації (`rad`) — модулем `r.sun` [33], форм рельєфу (`gmf`) в `r.geomorphon` [34]. Також були згенеровані додаткові карти гідрологічних показників: топографічного індексу вологості (`twi`) [35] в `r.topidx`, акумуляції (`flowaccum`) та напрямку водних потоків (`flowdirect`) в `r.terraflow` [36], довжини водних потоків (`flowlength`) [37] в `r.flow` та відстані до них (`diststream`, `r.stream.distance` [38]).

Для створення симулятивних моделей ґрунтового покриття нами був використаний скрипт [30, 39, 40] на мові R-statistic [24], який включає в себе ряд адаптацій для вирішення поставлених завдань та реалізує 14 основних типів предикативних алгоритмів [30, 39].

При цьому створення навчальної вибірки здійснювалося рандомізовано- зваженим методом на основі проведеного детального аналізу [39]). Для оцінки якості моделі використали індекс каппа Когена [41, 42], який в даному випадку показує ступінь відповідності між оригінальними та прогнозними даними.

Результати дослідження та їх обговорення. Аналіз шляхів верифікації реальних, картованих і модельних ґрунтів показує, що існує кілька можливих варіантів цього процесу (рис. 2). Зокрема, їх можна розділити на три групи залежно від виду вхідної інформації: наявна ґрунтова карта (полігональний тип даних), достовірні польові обстеження ґрунтів (точковий тип даних) та їх комбінація. Ці групи даних є базисом для побудови трьох відповідних їм предикативних моделей. Відмітимо, що єдиний вид верифікації, який не потребує моделювання, відповідає першому варіанту (V1) і полягає у порівнянні польових обстежень з картографічним матеріалом. Створені моделі першого ($Model_{map}$), другого ($Model_{profile}$) та третього ($Model_{combined}$) типів будуються на основі вище перелічених алгоритмів і можуть кроссверифікуватися кількома способами, як між собою (V9–V11), так і з джерелами даних (V2–V8). При цьому $Model_{combined}$ можна перевірити 5 варіантами валідації (V6–V10), а дві інші моделі лише 4-ма: V2, V4, V9, V11 та V3, V5, V10, V11 — для $Model_{map}$ і $Model_{profile}$ відповідно. Потреба у тому чи іншому варіанті валідації визначається при плануванні експерименту залежно від його мети і в екстремальних випадках можуть бути залучені всі 11 способів.

Конкретно в нашому випадку, ми планували дослідити точність існуючої карти агрогруп ґрунтів, оскільки існували певні сумніви в її достовірності, зокрема у точності проведення ґрунтових контурів. Саме тому і була запланована та схема закладки ґрунтових розрізів по катені, яка передбачала, в першу чергу, контроль ґрунтових меж та точності виділення самих ґрунтових відмін. Оскільки найпростішим і очевидним є порівняння діагностованих в польових умовах ґрунтів з відображеними на ґрунтовій карті (V1), то аналіз отриманих даних показав, наскільки співпадають реально досліджені ґрунти з картованими. Відмітимо, насамперед те, що підозри щодо якості цієї карти підтвердилися, оскільки ґрунтова карта співпала з закладеними розрізами лише у кількох випадках, а у всіх інших випадках ми спостерігаємо розходження (іноді доволі сильне) між генетичними типами реальних ґрунтів та даними з карти. Це свідчить лише про досить системні помилки даної карти, оскільки не співпадають з картою 15 із 18 закладених розрізів. Враховуючи достатньо простий у геоморфологічному відношенні рельєф, такі помилки картування можна вважати дуже грубими. Окрім того польові дослідження і картовані дані мають лише одну спільну агрогрупу ґрунтів, в ареалі якої якраз і розташовані три розрізи що збігаються. Всі інші агрогрупи мало того, що не співпадають за якістю, не співпадають і за кількістю: за нашими дослідженнями виділено 8 агрогруп проти 6 на карті.

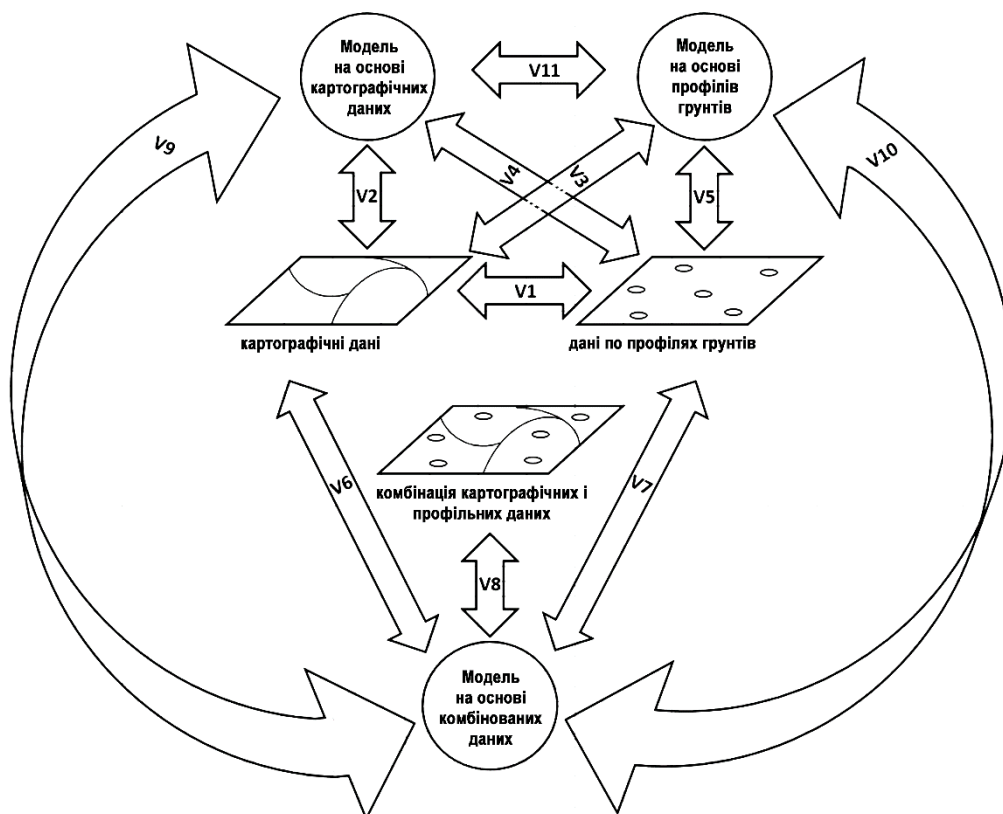


Рис. 2. Типи верифікації реальних, картографічних та модельних даних

Наявність таких грубих розходжень вказує на те, що подальше моделювання на основі картографічних даних ($Model_{map}$) у нашому випадку має сенс лише для оцінки якості предикативних алгоритмів (за рахунок покриття ареалу досліджень нехай і хибними зате повними даними). Відповідно, валідація даних згідно $Model_{combined}$ та відповідних їй варіантів (V6-V10) тут унеможлиблюється, оскільки така модель може бути побудована лише у випадку повного співпадіння польових досліджень з картованими даними. Ще одна модель ($Model_{profile}$) та набір відповідних їй варіантів (V3, V5, V10 та V11) в принципі може бути побудована, оскільки в якості навчального датасету буде використана польова діагностика, яка має вищу достовірність відносно картограми агрогруп за рахунок наших особистих спостережень. Проте, проведене моделювання ґрунтового покриття, апіорі не може бути точним, оскільки мета закладення розрізів у нашому випадку не підпорядкована цілям великомасштабного ґрунтового картографування (яке передбачає зовсім іншу схему закладки ґрунтових профілів), а контролю за точністю ґрунтових меж та виділення самих ґрунтових відмін.

Оскільки результати польових досліджень показали, що продовження роботи можливе лише з $Model_{map}$ і $Model_{profile}$, то ми спробували отримати максимум корисної інформації з цієї ситуації. Точність моделей у випадках оцінки картографічних і реальних даних, як було зауважено вище, визначається за допомогою каппи Когена. Вона, власне, дозволяє встановити, наскільки точно модель може описати умови розміщення ґрунтів, і, відповідно, відтворити ґрунтову карту.

Розглянемо ближче $Model_{map}$. Як нами встановлено, «завдяки» точності існуючої карти, дане моделювання має суто академічний інтерес у зв'язку із

валідацією набору предикативних алгоритмів. Отримані 14 симуляцій ґрунтового покриття показують досить цікаві закономірності щодо якості моделювання навіть при аналізі лише візуальних характеристик отриманих карт. Оцінка їх числових характеристик виявляє, що найбільш точними є 2 класифікаційні моделі (Random forest і Bagged trees) та K-Nearest Neighbors. Такий масив симуляцій ґрунтового покриття цікавий з точки зору його відповідності оригінальній карті і, відповідно до цього, прогностичній «силі» для ареалів з відсутньою інформацією (для досліджень, де це є актуально). Оскільки алгоритми аналізують весь спектр предикативних параметрів, виробляючи класифікаційні правила, то при високому рівні співпадіння модельних і реальних даних можна буде говорити про певний рівень статистичної достовірності результатів і в ареалах «білих плям». Саме тому отримані нами висновки в цьому плані досить обнадійливі з огляду на діапазон значень κ Когена. Якщо проранжувати моделі за зростанням якості предикції по κ основного набору даних, то для алгоритму PLR виявилися найгірші результати серед інших. За ним у порядку зростання розміщуються PLS, NSC, LDA, NDA, MLR, NN, SVM, KNN, RF та BGT. Останні два алгоритми відносяться до класифікаційних, і їх високі результати, очевидно, свідчать про найбільшу придатність такого роду підходів у картографічному відтворенні структури ґрунтового покриття на основі картографічної ж вибірки. Відмітимо, що цілий ряд алгоритмів перетнув межу якості предикції в 90 % (MLR, NN, SVM, KNN, RF та BGT), з яких два останніх (RF та BGT) практично відтворюють оригінальну ґрунтову карту. Це черговий раз підтверджує наші попередні дослідження [39], в яких показано, що рандомізовано-зважена вибірка навчальних даних є оптимальною для задач відтворення та моделювання карт ґрунтового покриття.

Аналіз рівня співставності отриманих результатів за якістю симуляції з подібними дослідженнями показує, що каппа більшості наших моделей перевищує усереднені значення з літературних джерел. Отримані нами дані мають у найгіршій моделі PLR — незначне сходження ($\kappa = 0,01-0,20$), PLS, NSC — істотне сходження ($\kappa = 0,61-0,80$), а у ліпших (LDA, NDA, MLR, NN, SVM, KNN, RF та BGT) — майже повне сходження ($\kappa = 0,81-0,99$). Відповідно, це дозволяє оцінити якість симулятивних карт-версій як таку, що не поступається рівню згаданих нами аналогічних за літературними даними. Окрім того вважаємо, що існує ще певний потенціал щодо підвищення загальної κ , зокрема, шляхом ретельнішого добору предикторів моделі та розширення їх числа включенням даних дистанційного зондування Землі, карт антропогенних відкладів (підстилаючих або материнських порід) тощо. Значним корисним ефектом такого роду моделювання є можливість заповнення прогалин на існуючих картографічних матеріалах даними з предикативних карт-версій та отримання, таким чином, комбінованих ґрунтових карт. Хоча такі результати отримані на деякою мірою «віртуальних» даних, це ніяк не принижує цінність висновків про ранжування предикативних алгоритмів та вибір способу отримання навчальної вибірки. У випадку аналізу картографічних матеріалів з прогалинами в ареалах такий підхід дозволяє отримати хоча б якісь дані на науковій основі з певним рівнем статистичної достовірності та використовувати їх для прикладних

задачах ґрунтознавства, агрономії, землевпорядкування та землеустрою тощо, тобто сферах, де потреба у таких даних стоїть найбільш гостро.

Не менш цікавим є детальний аналіз Model_{profile}. Як ми згадували, структура вихідних даних була спроектована не для цілей великомасштабного ґрунтового картографування, а для зв'язки ґрунтових меж та точності виділення самих ґрунтових відмін. Це на самому старті накладає певні обмеження на якість очікуваних результатів, оскільки точки закладення розрізів не є характерними точками місцевості (ключами у термінології великомасштабного картографування). Використання 14 предикативних алгоритмів дозволило отримати серію прогнозних карт, які загалом дозволяють зробити висновок про недостатність лише 18 розрізів (причому розташованих неоптимально для типового картування на ключах) для повноцінного моделювання ґрунтового покриття такої території. Так, в середньому на 1 ґрунтовий розріз припадає 9,28 га, що навіть по стандарту для цієї категорії складності місцевості є замало [16], хоч і відповідає усередненій площі на 1 розріз [5].

Високі значення κ Когена, отриманих при цьому, не повинні вводити в оману: 6 алгоритмів з 14 показали начебто 100 співпадіння навчальних та повних даних. Насправді це сталося так лише тому, що сама вибірка складається всього з 18 розрізів, і на такому маленькому обсязі даних це співпадіння не є показовим. Звичайно навіть з таких даних можна зробити певні висновки про якість алгоритмів. Так, стабільно високі, і як в попередньому варіанті моделювання 100 % результати показали RF і BGT. Стали 100 % MLR, NN, NDA та LDA. Дещо ліпші параметри каппа стали у PLS і NSC, трохи знизилися у SVM, а у PLR та KNN вони різко впали.

Значно цікавішим є візуальний контроль отриманих симуляцій. З огляду на певний ґрунтознавчий досвід, з отриманих таким способом карт найбільше подібні до реальних лише RF і BGT. Очевидно, що крос-валідації за каппа (у тих випадках де вона становить 100 %) формально будуть достовірними, але такі предикативні карти для використання не рекомендується.

Ці результати чітко окреслюють коло і технічні характеристики майбутніх досліджень:

- кількість закладених розрізів має відповідати плану великомасштабного ґрунтового обстеження;

- схема їх закладки повинна відповідати ключовим точкам місцевості на основі ЦМР і геоморфологічного аналізу.

Лише в такому разі доцільно говорити про задіявання всього спектру можливих крос-валідацій польових, картографічних та модельних даних та отримання максимально достовірних та коректних як з математичної, так і ґрунтознавчої сторони результатів.

Висновки і перспективи. Проведені дослідження показали, що існує 11 можливих шляхів верифікації реальних, картованих і модельних ґрунтів, які можна розділити на три групи залежно від виду вхідної інформації: полігональний тип даних (ґрунтові карти), точковий тип даних (достовірні польові обстеження ґрунтів) та їх комбінація. Таке групування даних передбачає можливості щодо побудови відповідних трьох наборів предикативних моделей, які включають 14 типів основних предикаційних

алгоритмів. Єдиний вид верифікації з описаних, який не потребує безпосереднього моделювання відповідає першому варіанту (V1) і полягає у прямому порівнянні польових обстежень з картографічним матеріалом.

Аналіз отриманих симуляцій ґрунтового покриття з точки зору їх відповідності оригінальній карті дозволяє рекомендувати деякі алгоритми для створення предикативних карт, які містять ареали з відсутньою інформацією про ґрунтовий покрив з результатами певного рівня статистичної достовірності. Показано, що отримані значення k Когена для деяких алгоритмів показують практично 100 % співпадіння модельних карт з оригінальними. Загалом, зростання якості предикції по k показує таку динаміку: PLR, PLS, NSC, LDA, NDA, MLR, NN, SVM, KNN, RF та BGT. Високі результати двох останніх алгоритмів свідчать, очевидно, про їх найбільшу придатність для модельного відтворення структури ґрунтового покриття (на основі картографічної навчальної вибірки). Підтверджено, що рандомізовано-зважена вибірка навчальних даних є оптимальною для задач відтворення та моделювання карт ґрунтового покриття.

References

1. Vykorystannia danykh suputnykovoї ziomky v systemakh tochnoho zemlerobstva / S. R. Truskavetskyi, T. Yu. Byndych, L. P. Koliada ta in. // *Inzheneriia pryrodokorystuvannia*. — 2017. — №1 (7). — S. 29–35.
2. Achasov A. B., Achasova A. O. Metodychni osnovy suchasnoho prostоровoho monitorynhu gruntiv // *Visnyk Kharkivskoho natsionalnoho universytetu imeni V. N. Karazina*. — Seriia: Ekolohiia. — 2011. — № 944. — Vyp. 6. — S. 20–27.
3. Achasov A. B., Titenko H. V., Kurilov V. I. Dani dystantsiinoho zonduvannia yak osnova kartohrafuvannia gruntiv: ekonomichnyi aspekt // *Visnyk Kharkivskoho natsionalnoho universytetu imeni V. N. Karazina*. — Seriia: Ekolohiia. — 2015. — № 1104. — Vyp. 10. — S. 60–66. — Rezhym dostupu: http://journals.urau.ua/visnyk_khnu_ecology/article/download/25458/33191.
4. Soil resources of Europe / Ed. by Robert J. A. Jones, Beata Houskova, Peter Bullock, Luca Montanarella. European Soil Bureau Research Report No.9, EUR 20559 EN. — 2 edition. — Luxembourg : Office for Official Publications of the European Communities, 2005. — 420 p. — Access mode: http://eusoils.jrc.ec.europa.eu/ESDB_Archive/eusoils_docs/esb_rr/n09_EUR20559.pdf.
5. Kanash O. P. Grunty — providna skladova zemelnykh resursiv // *Zemleustrii i kadastr*. — 2013. — № 2. — S. 68–76.
6. Cherlinka V. Using Geostatistics, DEM and Remote Sensing to Clarify Soil Cover Maps of Ukraine // *Soil Science Working for a Living: Applications of soil science to present-day problems* / Ed. by David Dent, Yuriy Dmytruk. — Cham, Switzerland : Springer-Verlag GmbH, 2017. — Jan. — P. 89–100. — Access mode: https://doi.org/10.1007/978-3-319-45417-7_7.
7. MacMillan R. A. Experiences with applied DSM: protocol, availability, quality and capacity building // *Digital Soil Mapping with Limited Data* / Ed. by Alfred E. Hartemink, Alex B. McBratney, Maria de Lourdes Mendonca-Santos. — Amsterdam : Springer Netherlands, 2008. — P. 113–135.
8. Browning D. M., Duniway M. C. Digital soil mapping in the absence of field training data: A case study using terrain attributes and semiautomated soil signature derivation to distinguish ecological potential // *Applied and Environmental Soil Science*. — 2011. — Vol. 2011. — Access mode: <https://doi.org/10.1155/2011/421904>.

9. An appropriate data set size for digital soil mapping in Erechim, Rio Grande do Sul, Brazil / Alexandre ten Caten, Ricardo Simao Diniz Dalmolin, Fabrício de Araujo Pedron et al. // *Revista Brasileira de Ciencia do Solo*. — 2013. — Vol. 37, no. 2. — P. 359–366. — Access mode: <https://doi.org/10.1590/s0100-06832013000200007>.
10. Machine learning for predicting soil classes in three semi-arid landscapes / Colby W. Brungard, Janis L. Boettinger, Michael C. Duniway et al. // *Geoderma*. — 2015. — Vol. 239. — P. 68–83. — Access mode: <https://doi.org/10.1016/j.geoderma.2014.09.019>.
11. Malone B. P., Minasny B., McBratney A. B. Using R for Digital Soil Mapping. Progress in Soil Science. — Springer International Publishing, 2016. — 262 p. — ISBN: 978-3-319-44327-0. — Access mode: <https://doi.org/10.1007/978-3-319-44327-0>.
12. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping / Brandon Heung, Hung Chak Ho, Jin Zhang et al. // *Geoderma*. — 2016. — Vol. 265. — P. 62–77. — Access mode: <https://doi.org/10.1016/j.geoderma.2015.11.014>.
13. Heung B., Hodul M., Schmidt M. G. Comparing the use of training data derived from legacy soil pits and soil survey polygons for mapping soil classes // *Geoderma*. — 2017. — Vol. 290. — P. 51–68. — Access mode: <https://doi.org/10.1016/j.geoderma.2016.12.001>.
14. Florinsky I. V. Digital Terrain Analysis in Soil Science and Geology. — Amsterdam : Academic Press / Elsevier, 2012. — 379 p. — ISBN: 978-0-12-385036-2. — Access mode: <https://doi.org/10.1016/c2010-0-65718-x>.
15. Mapping of reference area representativity using a mathematical soilscape distance / P. Lagacherie, J. M. Robbez-Masson, N. Nguyen-The, J. P. Barthès // *Geoderma*. — 2001. — Apr. — Vol. 101, no. 3-4. — P. 105–118. — Access mode: [https://doi.org/10.1016/s0016-7061\(00\)00101-4](https://doi.org/10.1016/s0016-7061(00)00101-4).
16. Tekhnicheskiye ukazaniya po korrektyrovke materialov krupnomasshtabnoho obsledovaniya pochv kolkhozov y sovkhozov Ukraynskoï SSR. — K. : MSKh USSR, 1977. — 128 s.
17. Digital soil mapping using logistic regression on terrain parameters for several ecological regions in Southern Brazil / E. Giasson, S. R. Figueiredo, C. G. Tornquist, R. T. Clarke // Digital Soil Mapping with Limited Data / Ed. by Alfred E. Hartemink, Alex B. McBratney, Maria de Lourdes Mendonça-Santos. — Amsterdam : Springer Netherlands, 2008. — P. 225–232. — Access mode: https://doi.org/10.1007/978-1-4020-8592-5_19.
18. Updating the 1:50,000 Dutch soil map using legacy soil data: A multinomial logistic regression approach / Bas Kempen, Dick J. Brus, Gerard B. M. Heuvelink, Jetse J. Stoorvogel // *Geoderma*. — 2009. — Jul. — Vol. 151, no. 3. — P. 311–326. — Access mode: <https://doi.org/10.1016/j.geoderma.2009.04.023>.
19. Debella-Gilo M., Etzelmüller B. Spatial prediction of soil classes using digital terrain analysis and multinomial logistic regression modeling integrated in GIS: Examples from Vestfold County, Norway // *Catena*. — 2009. — Apr. — Vol. 77, no. 1. — P. 8–18. — Access mode: <https://doi.org/10.1016/j.catena.2008.12.001>.
20. Hengl T. A practical guide to geostatistical mapping. — 2 edition. — Luxembourg : Office for Official Publications of the European Communities, 2009. — 271 p. — ISBN: 978-90-9024981-0. — Access mode: http://www.academia.edu/download/40396676/A_Practical_Guide_to_Geostatistical_Mapping.pdf.
21. GRASS Development Team. Geographic Resources Analysis Support System (GRASS GIS) Software // *Version 7.2*. — 2017. — Access mode: <http://grass.osgeo.org> (online; accessed: 25.05.2017).

22.QGIS Development Team. QGIS Geographic Information System, Open Source Geospatial Foundation. — 2015. — Access mode: <http://qgis.osgeo.org> (online; accessed: 25.05.2017).

23.EasyTrace group. Easy Trace 7.99. Digitizing software. — 2015. — Access mode: <http://www.easytrace.com> (online; accessed: 25.05.2017).

24.R Development Core Team. R: A language and environment for statistical computing. // *R Foundation for Statistical Computing*. — 2017. — Access mode: <http://www.r-project.org> (online; accessed: 25.05.2017).

25.Cherlinka V. R., Dmytruk Yu. M. Problemy stvorennia, heorektyfikatsii ta vykorystannia krupnomasshtabnykh tsyfrovyykh modelei reliefu // Heopolytyka y ekoeodynamyka rehionov. — 2014. — T. 10, № 1. — S. 239–244. Rezhym dostupu <http://geopolitika.crimea.edu/arhiv/2014/tom10-v-1/040cherlin.pdf>.

26.Cherlinka V. R. Adaptatsiia velykomasshtabnykh kart hruntiv do yikh praktychnoho vykorystannia u HIS // Ahrokhimiia i gruntoznavstvo. Mizhvid. tem. nauk. zb. — Vyp. 84. — Kharkiv : TOV «Smuhasta typohrafiia», 2015. — S. 20–28. — Rezhym dostupu: <http://agrosoil.yolasite.com/resources/2015-AiG-84-pp20-28.pdf>.

27.Evenden G., Warmerdam F. Proj. 4 — Cartographic Projections Library // *Source code and documentation*. — 1990. — Access mode: <https://proj4.org/>.

28.Mitasova H., Mitas L. Interpolation by regularized spline with tension: I. Theory and implementation // *Mathematical Geology*. — 1993. — Vol. 25, no. 6. — P. 641–655. — Access mode: <https://doi.org/10.1007/BF00893171>.

29.Hofierka J., Cebecauer T., Šurí M. Optimisation of interpolation parameters using cross-validation // Digital Terrain Modelling / Ed. by Robert Joseph Peckham, Gyoza Jordan. — Berlin Heidelberg : Springer, 2007. — Lecture Notes in Geoinformation and Cartography. — P. 67–82. — Access mode: http://dx.doi.org/10.1007/978-3-540-36731-4_3.

30.Cherlinka V. R. Vplyv rozdilnoi zdatnosti tsyfrovyykh modelei reliefu na yakist predykatyvnoi symuliacii gruntovoho pokryvu // Gruntoznavstvo. — 2017. — T. 18, № 1-2. — S. 79–95. — Rezhym dostupu : <http://arss.dp.ua/index.php/ass/article/view/11>.

31.Hofierka J., Mitasova H., Neteler M. Geomorphometry in GRASS GIS // Geomorphometry: Concepts, Software, Applications / Ed. by Tomislav Hengl, Hannes I. Reuter. — Amsterdam : Elsevier, 2009. — Vol. 33 of *Developments in Soil Science*. — P. 387–410. — Access mode: [https://doi.org/10.1016/S0166-2481\(08\)00017-2](https://doi.org/10.1016/S0166-2481(08)00017-2).

32.Wood J. D. The geomorphological characterisation of digital elevation models : Phd thesis / Joseph D. Wood ; University of Leicester, UK. — Leicester : UoL, 1996. — 185 p. — Access mode: <https://lra.le.ac.uk/bitstream/2381/34503/1/U077269.pdf>.

33.Hofierka J., Suri M. The solar radiation model for Open source GIS: implementation and applications // Proceedings of the Open source GIS-GRASS users conference / Ed. by Marco Ciolli, Paolo Zatelli ; Dipartimento di Ingegneria Civile e Ambientale. — Trento, Italy : Universita' di Trento, 2002. — P. 1–19. — Access mode: http://www.ing.unitn.it/~Grass/conferences/GRASS2002/proceedings/proceedings/pdfs/Hofierka_Jaroslav.pdf.

34.Jasiewicz J., Stepinski T. F. Geomorphons — a pattern recognition approach to classification and mapping of landforms // *Geomorphology*. — 2013. — Jan. — Vol. 182. — P. 147–156. — Access mode: <https://doi.org/10.1016/j.geomorph.2012.11.005>.

35.Soil attribute prediction using terrain analysis / Ian D. Moore, P. E. Gessler, G. A. Nielsen, G. A. Peterson // *Soil Science Society of America Journal*. — 1993. — Vol. 57, no. 2. — P. 443–452. — Access mode: <https://doi.org/10.2136/sssaj1993.572npb>.

36.Efficient flow computation on massive grid terrain datasets / Lars Arge, Jeffrey S. Chase, Patrick Halpin et al. // *Geoinformatica*. — 2003. — Vol. 7, no. 4. — P. 283–313. — Access mode: <https://doi.org/10.1023/A:102552642>.

37.Mitasova H., Hofierka J. Interpolation by regularized spline with tension: II. Application to terrain modeling and surface geometry analysis // *Mathematical Geology*. — 1993. — Vol. 25, no. 6. — P. 657–669. — Access mode: <https://doi.org/10.1007/BF00893172>.

38.Jasiewicz J., Metz M. A new GRASS GIS toolkit for hortonian analysis of drainage networks // *Computers & Geosciences*. — 2011. — Aug. — Vol. 37, no. 8. — P. 1162–1173. — Access mode: <https://doi.org/10.1016/j.cageo.2011.03.003>.

39.Cherlinka V. R. Variatsii prohnozhnoi efektyvnosti gruntovykh kart zalezchno vid sposobiv pobudovy navchalnykh vybirok predykatyvnykh alhorytmiv // *Ekolohiia ta noosferolohiia*. — 2017. — T. 28, № 3-4. — S. 55–71. — Rezhym dostupu: <http://erae.dp.ua/index.php/erae/article/view/20>.

40.Cherlinka V. R. Morfometrychni parametry reliefu yak bazys dlia predykatyvnoho modeliuvannia prostorovoho poshyrennia gruntovykh vidmin // *Ahrokhimiia i gruntoznavstvo. Mizhvid. tem. nauk. zb.* — Vyp. 86. — Kharkiv : TOV «Smuhasta typohrafiia», 2017. — S. 5–16. — Rezhym dostupu: http://agrosoil.yolasite.com/resources/2017-86/2017_AiG-86_5-16_Chерlinka.pdf.

41.Landis J. R., Koch G. G. The Measurement of Observer Agreement for Categorical Data // *Biometrics*. — 1977. — Mar. — Vol. 33, no. 1. — P. 159–174. — Access mode: <https://doi.org/10.2307/2529310>.

42.Li W., Zhang C. A Random-Path Markov Chain Algorithm for Simulating Categorical Soil Variables from Random Point Samples // *Soil Science Society of America Journal*. — 2007. — Vol. 71, no. 3. — P. 656–668. — Access mode: <https://doi.org/10.2136/sssaj2006.0173>.

МЕТОДЫ ВЕРИФИКАЦИИ ПРЕДИКАТИВНЫХ ПОЧВЕННЫХ КАРТ

В. Р. Черлинка, Ю. М. Дмитрук

Аннотация. Анализ возможных методов верификации полученных результатов является необходимым и важным процессом для получения обоснованных выводов о качестве проведенного моделирования. В отличие от распространенных в литературе рекомендаций относительно процесса верификации полученных результатов моделирования почвенного покрова чисто с математической стороны, предложена, так называемая, кросс-валидация, которая позволяет сопоставить модельные данные с реальными почвенными обследованиями и существующими картами. Это потенциально позволяет отследить ошибки не только модельного эксперимента, но и «законсервированные» на существующих картах. Установлено, что для охвата всего спектра возможных кросс-валидаций полевых, картографических и модельных данных и получения максимально достоверных и корректных как с математической, так и почвоведческих сторон результатов, количество заложенных разрезов должно соответствовать плану крупномасштабного почвенного обследования, а схема их расположения должна соответствовать репрезентативным точкам местности, выделенным на основе геоморфологического анализа.

Ключевые слова: морфометрические параметры, ЦМР, почвенная карта, корреляция, предикторы, прогнозные алгоритмы, кросс-валидация

VERIFICATION METHODS FOR PREDICATIVE SOIL MAPS

V. R. Cherlinka, Y. M. Dmytruk

Abstract. *An analysis of possible verification methods from the results is a necessary and important process for making substantiated conclusions about the quality of the modeling performed. In contrast to the popular literature recommendations on the process of verifying the results of simulation of soil coverings purely from the mathematical side, a so-called cross-validation is proposed, which allows comparing model data with real soil surveys and existing maps. This potentially allows you to track errors not only in the model experiment, but also «preserved» on existing ground maps. It has been established that in order to reach the full spectrum of possible cross-validations of field, cartographic and model data and obtain the most reliable and correct results from both the mathematical and soil sciences, the number of inlaid sections should correspond to the plan of large-scale soil survey, and the layout of the arrangement should correspond to representative points areas allocated on the basis of geomorphological analysis.*

Keywords: *morphometric parameters, DEM, soil map, correlation, predictors, predictive algorithms, crossvalidation*