

УДК 631.4:551.4:004.942

ПРЕДИКАТИВНІ ҐРУНТОВІ КАРТОГРАФІЧНІ МАТЕРІАЛИ ЯК ЕЛЕМЕНТ СУЧАСНИХ ВЕЛИКОМАСШТАБНИХ ОБСТЕЖЕНЬ

Ю. Дмитрук, д. б. н.

ORCID ID: 0000-0002-3157-0503

В. Черлінка, д. б. н.

ORCID ID: 0000-0001-5354-4851

І. Демид, здобувач

ORCID ID: 0000-0001-5819-6390

Чернівецький національний університет імені Юрія Федьковича<https://doi.org/10.31734/agronomy2019.01.202>

Дмитрук Ю., Черлінка В., Демид І. Предикативні ґрунтові картографічні матеріали як елемент сучасних великомасштабних обстежень

Розглянуто поточний стан із ґрунтовими картографічними даними в Україні та показано, що значні території (до 33 %) не покрито великомасштабними ґрунтовими обстеженнями. Показано шлях отримання карт-версій ґрунтового покриття чи картограм агровиробничих груп, суть якого полягає в морфометричному аналізі цифрових моделей рельєфу, виділенні з них низки предикторів та пов'язанні наявних картографічних ґрунтових матеріалів з цими даними через створення математичної предикативної моделі з використанням опорних точок ландшафтів та приурочених до них ґрунтових таксонів, що дає змогу заповнювати необстежені ареали прогнозними даними.

Порівняння картографічних даних, отриманих за різними алгоритмічними методиками, показує, що Random Forest володіє найбільшою прогностичною силою з усіх представлених у нашому дослідженні – 89,3 % проти 54,8 % нейронної мережі та 76,8 % у *k*-найближчих середніх. Високий ступінь збігання прогнозних і картографічних даних алгоритму рандомних лісів у межах картованих ділянок дає підстави до використання отриманих матеріалів в ареалах «білих плям». За відсутності будь-яких варіантів картографічної інформації предикативні варіанти можуть мати застосування в практиці сільськогосподарського виробництва, плануванні системи удобрення, розробки системи протиерозійних заходів тощо. У ймовірному випадку проведення чергових великомасштабних ґрунтових обстежень отримані карти-версії є елементом, який допоможе значною мірою скоротити та мінімізувати фінансові, часові й трудові витрати при польових обстеженнях за рахунок закладання напіврозрізів та прикопок замість повнопрофільних їхніх варіантів у разі збігання прогнозних і реально спостережуваних ґрунтових відмін.

Ключові слова: ґрунтова карта, картограма агровиробничих груп ґрунтів, предикативні алгоритми, симуляція, морфометричні параметри, цифрова модель рельєфу.

Dmytruk Yu., Cherlinka V., Demyd I. Predictive soil cartographic materials as elements of modern large-scale surveys

The current state of soil map data in Ukraine is considered and it is shown that significant territories (up to 33 %) are not covered by large-scale soil surveys. A way is shown for obtaining maps of soil cover or cartograms of agro-industrial groups, the essence of which is the morphometric analysis of digital relief models, the selection of a number of predictors from them and the association of existing cartographic soil materials with these data by creating a mathematical predicative model using the reference points of the landscapes and soil taxons linked to them, which allows you to fill in unexplored areas with forecast data.

Comparison of the map data obtained by different algorithmic methods shows that Random Forest has the highest predictive power from all present in this study – 89,3 % vs. 54,8 % of the neural network and 76,8 % in the *k*-nearest neighbor. The high degree of coincidence of forecast and cartographic data of the random forests' algorithm within the research boundaries of the plotted areas gives rise to the use of the obtained materials in the areas of «white spots». In the absence of any variants of cartographic information, predicative variants can be used in the practice of agricultural production, the planning of fertilizer system, the development of a system of anti-erosion measures, etc. In the probable case of upcoming regular large-scale soil surveys, the maps-version obtained is an element that will allow a significant reduction and minimization of financial, time and labor costs for field surveys by introducing half-profile and cuts instead of full-profile variants in cases of coincidence of predicted and actually observed soil.

Key words: soil map, cartogram of agro-industrial groups of soils, predictive algorithms, simulation, morphometric parameters, digital relief model.

Постановка проблеми. Наявні в системі Держгеокадастру України великомасштабні ґрунтові картографічні матеріали, зокрема картограми агровиробничих груп ґрунтів, створені ще в 60–70-ті роки минулого століття і не відображають сучасний стан ґрунтового покриття.

Значна частина карт має незадовільну вихідну якість, а проведена векторизація у 2012 р. додала численних проблем із геоприв'язкою та атрибуванням даних. До того ж, близько третини території країни ніколи не було покрито суцільною ґрунтовою зйомкою. Оскільки сучасні економічні умови не створюють передумов для повноформатної актуалізації існуючих матеріалів та дообстеження білих плям, а розвиток інтенсивних агродеградацій (водної ерозії, забруднення ґрунтового покриву тощо) нагально вимагає робіт такого роду, то необхідно знаходити малозатратні шляхи здобування детальної інформації про якісний та кількісний стан ґрунтів. Одним із них є залучення прогресивних способів великомасштабних ґрунтових обстежень, складовими елементами яких є предикативне, або прогнозне, моделювання просторового поширення ґрунтових відмін. Дослідження в такому напрямі мають, безумовно, високу затребуваність та актуальність в умовах сучасної України.

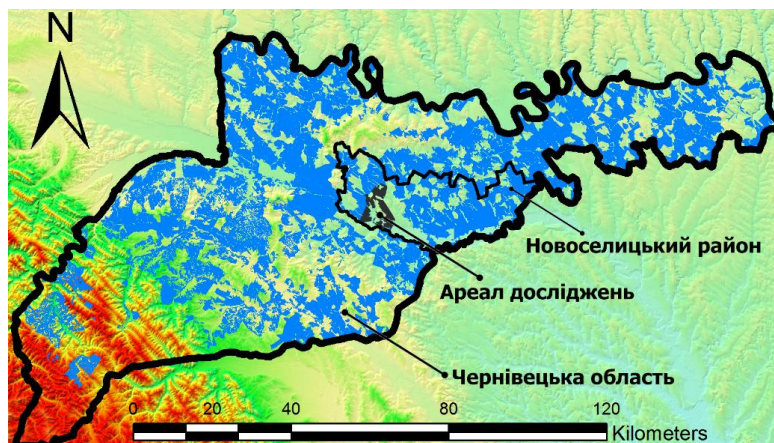
Аналіз останніх досліджень і публікацій.

Великомасштабні ґрунтові картографічні матеріали, до яких ми відносимо як карти ґрунтів, так і їхніх агрогруп, нагально необхідні для сучасного високоінтенсивного сільськогосподарського виробництва. Проте згідно з [1] картографічним даним, отриманим ще під час першого туру великомасштабного ґрунтового обстеження 1957–1961 рр., не вистачало точності, а проведення другого туру ґрунтових зйомок для удосконалення цих матеріалів (коректування матеріалів ґрунтових обстежень) у масштабі 1:10 000 в 1969–1991 рр., хоч і охопило 84,6 % території сільськогосподарських підприємств, викликає чимало запитань щодо якості [5]. Навіть такі обстеження після 1991 р. практично не проводили [1]. Тому доводиться констатувати факт, що отримані дані досі є базовим джерелом інформації про ґрунтовий покрив, тобто про 40 млн га, або 66 % території України. Інформацію про ґрунти решту 33 % площі держави за відсутності належного фінансування здобути без використання моделювання практично неможливо.

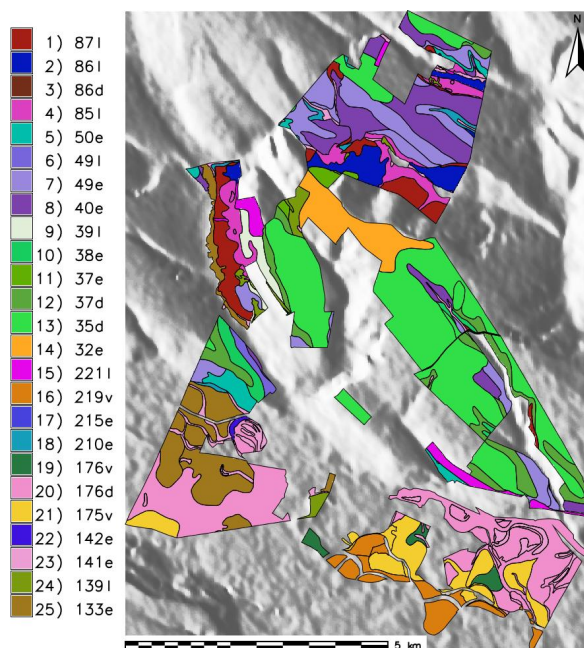
Постановка завдання. Основним завданням нашого дослідження був аналіз ситуації з недостатнім покриттям території України великомасштабними ґрунтовими картографічними матеріалами та показ методичних підходів до вирішення вказаної проблеми, зокрема через створення предикативних карт ґрунтів чи їхніх агрогруп для

Виклад основного матеріалу. Як об'єкт було обрано фрагмент території України (рис. 1, а) в межах Чернівецької області (рис. 1, б), приурочений до Прут-Дністровського межиріччя (Північна Буковина) з контрастними геоморфологічними умовами та адміністративно належний Новоселицькому району. Цей ареал, вписаний в квадрат розмірами 12150×9000 м, має різне адміністративне підпорядкування та господарське використання. Система координат проекту була приведена до СК 1963 (зона X2). Як цифрову модель рельєфу використали дані SRTM [15], передискретизовані до роздільної здатності 25 м. Відповідно до висновків [2] геоприв'язка ґрунтових карт здійснювалася до характерних точок місцевості та адміністративних меж існуючих на теперішній момент сільських рад: Боянівської (н. п. Бояни) та Припрутьської (н. п. Припруття та Боянівка) Новоселицького району Чернівецької області. Інформативні ґрунтові матеріали базувалися на серії архівних картограм агрогосподарських груп ґрунтів колгоспу «Батьківщина» (с. Бояни, с. Припруття та с. Боянівка), зокрема, коректування матеріалів крупномасштабного обстеження 1980 року. Ця інформація була дигіталізована та отримано чітку картину щодо покриття території ґрунтовими обстеженнями (рис. 1, б): «білі» плями приурочені до території населених пунктів та лісових масивів, які входять до меж згаданих сільських рад. Досі необстеженою залишається також територія Української науково-дослідної станції карантину рослин Інституту захисту рослин НААН, яка знаходиться в межах сучасної Боянівської сільської ради.

Оскільки вихідний варіант цифрової моделі рельєфу містить значний відсоток «зашумлених» даних (рис. 2, а), який унеможливорює якісний геоморфологічний аналіз, проведено згладжування моделі рельєфу алгоритмом *mdnoise* згідно із запропонованою методикою [16] з числом ітерацій, рівним 20. На основі знешумленої цифрової моделі рельєфу (ЦМР) виділено низку геоморфологічних характеристик (крутизна та експозиція схилів, кривизна поверхні (поздовжня та максимальна), кількість сонячної радіації, топографічний індекс вологості, акумуляція, напрям та довжина водних потоків і відстань до них) та проведено аналіз ступеня диференціації ґрунтового покриву залежно від форм рельєфу, що детальніше описано в [3; 4; 6]. Описані в цих працях результати стали підґрунтям для розробленої методики моделювання ґрунтового покриву, у т. ч. відновлення великомасштабних карт ґрунтів чи їхніх агрогруп для необстежених досі ареалів.



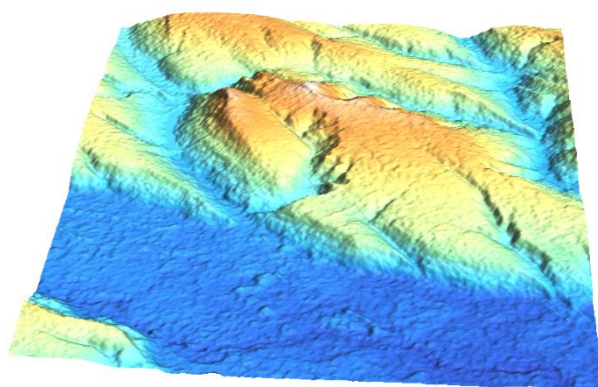
а)



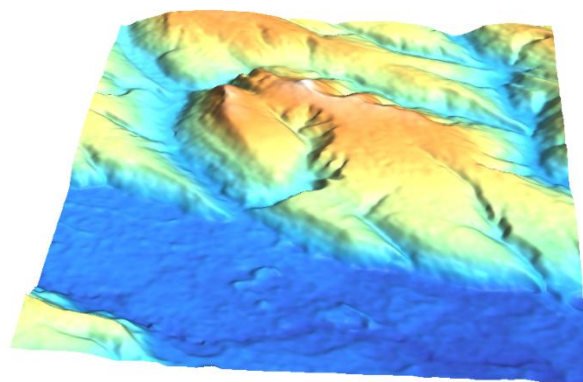
б)

Рис. 1. Географічне розташування регіону досліджень в межах Чернівецької області (а) та картограма агро виробничих груп ґрунтів досліджуваного ареалу (б).*

*Для фону використано дані SRTM – NASA's Shuttle Radar Topography Mission.



а)



б)

Рис. 2. Цифрова модель рельєфу SRTM досліджуваного ареалу до (а) та після (б) знешумлення.

Для опрацювання даних були використані інструментальні можливості вільного програмного забезпечення: георектифікація картографічного матеріалу – ГІС Quantum [12], оцифрування – Easy Trace [8], підготовка карт морфометричних параметрів і генерація ЦМР – ГІС GRASS [9] та симуляція карт ґрунтового покриття – мова статистичних розрахунків R-statistic [13].

Для створення предикативної картограми агропромислових груп ґрунтів використали розроблений нами скрипт на мові R-statistic, який охоплює в себе ряд адаптацій для виконання поставлених зав-

дань та реалізує 14 основних типів предикативних алгоритмів (зокрема, використані алгоритми Neural Network – NN [14], K-Nearest Neighbors – KNN [11] та Random Forest – RF [16]), а для оцінки якості отриманих моделей використали індекс каппа Когена κ [10], який показує ступінь відповідності між оригінальними та предикативними даними.

У процесі побудови прогнозу картограми агропромислових груп отримано предикативні її варіанти, для яких показник каппа навчальної вибірки становить 100 % та 54–89 % для всього досліджуваного ареалу (рис. 3, а-в).

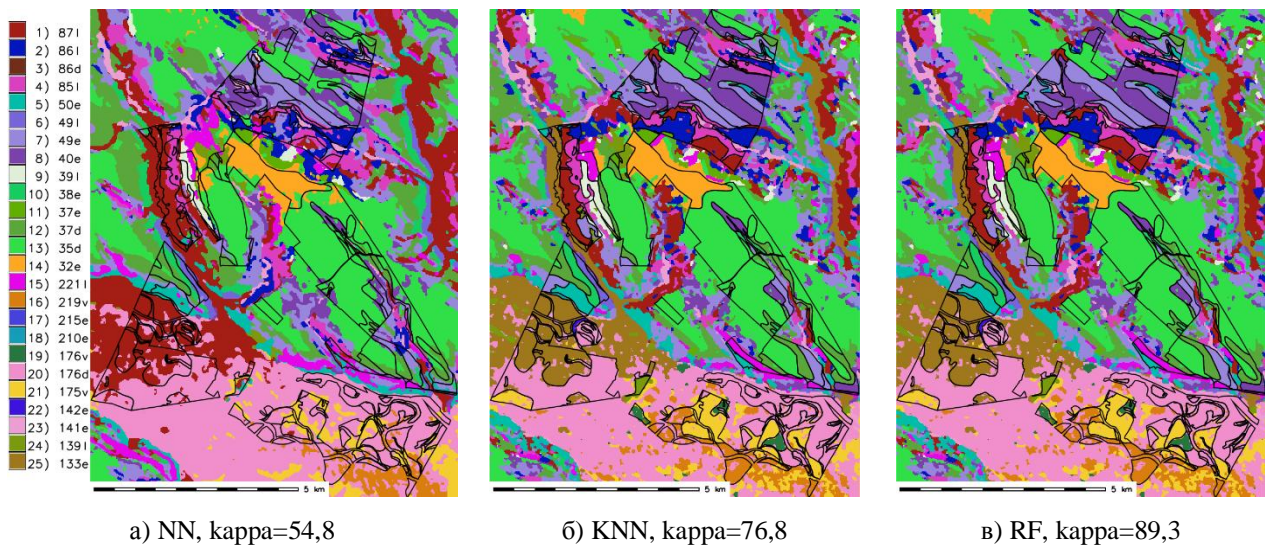


Рис. 3. Варіанти прогнозних картограм агропромислових груп ґрунтів на основі різних предикативних алгоритмів.

Порівняння картографічних даних, отриманих за різними алгоритмічними методиками, показує, що Random Forest володіє найбільшою прогностичною силою з усіх представлених у нашому дослідженні – 89,3 % проти 54,8 % нейронної мережі та 76,8 % у κ -найближчих середніх. Високий ступінь збігання прогнозних і картографічних даних алгоритму випадкових лісів у межах картованих ділянок дає підстави до використання отриманих матеріалів в ареалах «білих плям». За відсутності будь-яких варіантів картографічної інформації предикативні варіанти можуть мати застосування в практиці сільськогосподарського виробництва, плануванні системи удобрення, розробці системи протиерозійних заходів тощо.

У ймовірному проведенні чергових великомасштабних ґрунтових обстежень отримані карти-версії є елементом, який дасть змогу значною мірою скоротити й здешевити польові обстеження за рахунок закладання напіврозрізів та прикопок замість повнопрофільних їхніх варіантів у разі

збігання прогнозних і реально спостережуваних ґрунтових відмін.

Висновки. Поточна ситуація з ґрунтовими картографічними матеріалами вимагає нагального їхнього оновлення відповідно до сучасних вимог держави. Показано шлях отримання карт – версій ґрунтового покриття чи картограм агропромислових груп, суть якого полягає в аналізі цифрових моделей рельєфу та наявних картографічних ґрунтових матеріалів і на цій основі заповнення необстежених ареалів прогнозними даними. Останні можуть бути використані для мінімізації фінансових, часових і трудових витрат у разі проведення чергових великомасштабних ґрунтових обстежень. Окрім того, слід вказати на наявність резервів щодо підвищення якості предикції, зокрема за використання прогресивніших алгоритмів, модифікації кількості і розширення діапазону предикторів тощо.

Бібліографічний список

1. Канаш О. П. Ґрунти – провідна складова земельних ресурсів. *Землеустрій і кадастр*. 2013. № 2. С. 68–76.
2. Черлінка В. Р. Адаптація великомасштабних карт ґрунтів до їх практичного використання у ГІС. *Агророхімія і ґрунтознавство*. Харків, 2015. Вип. 84. С. 20–28.
3. Черлінка В. Р. Варіації прогнозованої ефективності ґрунтових карт залежно від способів побудови навчальних вибірок предикативних алгоритмів. *Екологія та ноосферологія*. 2017. Т. 28, № 3–4. С. 55–71.
4. Черлінка В. Р. Вплив роздільної здатності цифрових моделей рельєфу на якість предикативної симуляції ґрунтового покриву. *Ґрунтознавство*. 2017. Т. 18, № 1–2. С. 79–95.
5. Черлінка В. Р., Лобова О. В. Методичні підходи до узгодження ґрунтових картографічних матеріалів на межах адміністративно-територіальних одиниць України. *Наукові доповіді НУБіП України*. 2018. №. 6 (76). С. 1–15.
6. Черлінка В. Р. Морфометричні параметри рельєфу як базис для предикативного моделювання просторового поширення ґрунтових відмін. *Агророхімія і ґрунтознавство*. Харків, 2017. Вип. 86. С. 5–16.
7. 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. Part 2, chapt. 7. P. 89–100.
8. EasyTrace group. *Easy Trace 7.99*. 2015. Digitizing software. URL: <http://www.easytrace.com> (Last accessed: 01.05.2019).
9. GRASS Development Team. *Geographic Resources Analysis Support System (GRASS GIS) Software*. Version 7.2. 2017. URL: <http://grass.osgeo.org> (Last accessed: 01.05.2019).
10. Landis J. R., Koch G. G. The Measurement of Observer Agreement for Categorical Data. *Biometrics*. 1977. Vol. 33, N 1. P. 159–174.
11. Liu B. *Web Data Mining: Exploring Hyperlinks, Contents and Usage Data*. London; New York; Dordrecht: Springer-Verlag GmbH, 2011. 622 p.
12. QGIS Development Team. *QGIS Geographic Information System*. 2015. URL: <http://qgis.osgeo.org> (Last accessed: 01.05.2019).
13. R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. 2017. URL: <http://www.r-project.org> (Last accessed: 01.05.2019).
14. Ripley B., Venables W. *R-package nnet: Feed-forward neural networks and multinomial log-linear models*. v.7.3-12. 2016. URL: <https://cran.r-project.org/package=nnet> (Last accessed: 01.05.2019).
15. The Shuttle Radar Topography Mission (SRTM), NASA. *U.S. Releases Enhanced Shuttle Land Elevation Data*. 2015. URL: <http://www2.jpl.nasa.gov/srtm> (Last accessed: 01.05.2019).
16. Wright M. N., Ziegler A. Ranger: A fast implementation of random forests for high dimensional data in C++ and R. *Journal of Statistical Software*. 2017. Vol. 77. № 1. P. 1–17.

Стаття надійшла 12.05.2019.