
SOIL CARTOGRAPHY




V. R. Cherlinka  Cand. Sci. (Biol.), Assoc. Prof.

UDK 631.4:551.4:004.942

*Yuriy Fedkovych Chernivtsi National University,
str. Lesia Ukrainka, 25, Chernivtsi, Ukraine, 58012*

INFLUENCE OF RESOLUTION OF DIGITAL ELEVATION MODELS ON THE QUALITY OF PREDICATIVE SIMULATION OF SOIL COVER

Abstract. The main objective was to investigate the effect of the digital model of relief (DEM) on the qualitative characteristics of simulative soil maps, which are obtained through simulation using a typical set of materials that can be potentially available for soil scientist in modern Ukrainian realities. Achievement of this goal has been accomplished by solving a number of the following tasks: a) digitization and attribution of cartographic materials; b) construction of a DEM with a resolution of 2.5; 5; 10; 15; 20, 25, 30 and 50 m; c) analysis of digital elevation models and generation on this base a set of maps of morphometric and other derivatives characteristics in the GIS GRASS; d) the creation simulation models of soil cover (maps-versions or maps-models) using 14 basic types of predictive algorithms for areas with available soil information and for those where it is absent. As an object was selected a fragment of the territory of Ukraine (4200×4200 m) within the limits of Glybotsky district of the Chernivtsi region, confined to the Prut-Siret interfluvium (North Bukovyna) with contrast geomorphological conditions. This area has different administrative subordination and economic use but is covered with soil cartographic materials only by 49.43%. For data processing were used instrumental possibilities of free software: geo-rectifications of maps material – GIS Quantum, digitalization – Easy Trace, preparation of maps morphometric parameters – GRASS GIS and building simulative soil maps – R, a language and environment for statistical computing. The series of digital elevation models was constructed using by regularized splines with tension in GIS GRASS. It served as the basis for the selection of a number of morphometric characteristics of the relief: slope and aspect, curvature of the surface (profile, plan, longitudinal, minimal and maximal), calculations of data on the amount of solar radiation and forms of relief. Additional maps of hydrological indicators were also generated: the topographic wetness index, flow accumulation and direction, length of water flows and distance to them. To create simulation models of soil cover, a R-statistic script was written that includes a number of adaptations for solving set tasks and implements the different types of predicative algorithms such as: Multinomial Logistic Regression, Decision Trees, Neural Networks, Random Forests, Naive Bayes, K-Nearest Neighbors, Flexible Discriminant Analysis, Nonlinear Discriminant Analysis, Support Vector Machines, Linear Discriminant Analysis, Partial Least Squares Discriminant Analysis, Penalized Logistic Regression, Nearest Shrunken Centroids and Bagged Trees. To assess the quality of the obtained models, the Cohen's Kappa Index (κ) was used which best represents the degree of compliance between the original and the simulated data. The final result of the conducted model experiment was a set of simulative soil map maps for the whole territory of the research, which was built on the basis of the educational selection formed on the median lines of the soil types. It has been established that there is a general trend where the values of κ for training data are higher than κ of simulation models based on them (except for the

 Tel.: +38050-158-25-85, e-mail: v.cherlinka@chnu.edu.ua

DOI: 10.15421/041709

ISSN 1684-9094. Gruntoznavstvo. 2017. Vol. 18, no. 1-2

79

Bugget Trees model). It is shown that the Decision Trees model is the leader in the quality of the resulting map-version, the Random Forests algorithm is second in top, and Bagged Trees – it is slightly inferior to them on all the steps grid resolution DEM. Of the promising models that are worthy of attention in the simulation of the soil cover, it should be noted Neural Networks, K-Neighbors and Support Vector Machines. In particular, the Neural Network, showing not very high results on domestic cartographic material, is a leader in the analysis of foreign cartographic sources, which can indirectly indicate the difference in the approaches and methodology of mapping. Also was investigated an important parameter such as the time of computation and it was found that with a decrease in the resolution of up to 2,5 m, it dramatically increases. For the approximate calculation of time costs, the equation of the power function is proposed, which allows them to obtain at different values of the grid cell. The visual analysis of the best of possible simulations of the soil cover for each of the grid sizes in comparison with the original map showed that at large sampling steps of the model there is a loss of the soil details of the soil situation, although κ is at the same time is rather high. Therefore, in choosing the optimal resolution of the DEM in separate investigations, the class of the topographic complexity of the region should be taken into account, while adhering to the recommendation to choose the highest resolution when increasing the saturation of the relief with small details. Simulation variants of the obtained maps-versions have a generally higher spatial resolution of mapping, which is manifested in the deeper development of the contours of the soil situation and the higher detail of the small areas of soils than the manually constructed archival map. This is true for maps-versions with a resolution of up to 30 m, since higher grid cell size values already show a significant generalization of both the original map and its predicative variants. Taking into account the fact that from as 60.5 million hectares of land of Ukraine are currently covered by large-scale soil surveys of about 45 million ha only, that is, about a quarter territory of the country still presents a various white spots, process of creating simulations maps based on a broad analysis of existing predicative algorithms has some scientific novelty. A significant beneficial effect of this kind of modeling is the ability to fill gaps on existing cartographic materials with data from predicative map-versions and, thus, obtaining composite soil maps. This certainly does not exclude the need for a large-scale soil survey of such areas, but in the absence possibility of its carrying out, it allows to obtain certain data on a scientific basis with a high level of statistical reliability. It also allows it to be used in applied problems of soil science, agronomy and land management, that is, in areas where the need for such data is most acute.

Key words: simulation, morphometric parameters, DEM, soil map, predictive algorithms.

УДК 631.4:551.4:004.942 **В. Р. Черлинка** канд. биол. наук, доц.

*Черновицкий национальный университет им. Ю. Федьковича,
ул. Леси Украинки, 25, г. Черновцы, Украина, 58012,
тел.: +38050-158-25-85, e-mail: v.cherlinka@chnu.edu.ua*

ВЛИЯНИЕ РАЗРЕШАЮЩЕЙ СПОСОБНОСТИ ЦИФРОВЫХ МОДЕЛЕЙ РЕЛЬЕФА НА КАЧЕСТВО ПРЕДИКАТИВНОЙ СИМУЛЯЦИИ ПОЧВЕННОГО ПОКРОВА

Аннотация. Основной целью математического эксперимента было исследование влияния разрешения ЦМР на качественные характеристики симулятивных почвенных карт, получаемых путем моделирования при использовании стандартного набора материалов, которые могут быть потенциально доступными рядовому почвоведу или ученому в современных украинских реалиях. При этом показано, что морфометрические параметры рельефа и его производные являются надежным базисом предикативного моделирования пространственного распространения почвенных таксономических единиц с достаточно высокой надежностью, а представленная методика имеет значительную перспективу в научно-производственных задачах. На основе корреляционного анализа была осуществлена оценка тесноты связи и роли упомянутых параметров в изменчивости почвенного покрова, что с привлечением анализа главных компонент позволило выбрать 9 базовых предикторов модели: абсолютные высоты, топографический индекс влажности, количество солнечной радиации на единицу площади, крутизну склонов, продольную и максимальную кривизну топографической поверхности, аккумуляцию, длину и расстояние до водных потоков. Сделана расширенная оценка качества симулятивных почвенных карт при различных значениях шага дискретизации

ЦМР. Установлені различия в качестве прогнозных почвенных карт при использовании 14 основных типов предикативных алгоритмов и рекомендованы наиболее пригодные для такого рода задач модели, в частности Decision Trees, Random Forests, при этом выделен ряд моделей, которые потенциально могут показывать высокие результаты, в частности Bugget Trees, K-Nearest Neighbors, Support Vector Machines и Neural Networks.

Ключевые слова: симуляция, морфометрические параметры, ЦМР, почвенная карта, предикационные алгоритмы.

УДК 631.4:551.4:004.942

В. Р. Черлінка

канд. біол. наук, доц.

*Чернівецький національний університет ім. Ю. Федьковича,
вул. Лесі Українки, 25, м. Чернівці, Україна, 58012,
тел.: +380501582585, e-mail: v.cherlinka@chnu.edu.ua*

ВПЛИВ РОЗДІЛЬНОЇ ЗДАТНОСТІ ЦИФРОВИХ МОДЕЛЕЙ РЕЛЬЄФУ НА ЯКІСТЬ ПРЕДИКАТИВНОЇ СИМУЛЯЦІЇ ҐРУНТОВОГО ПОКРИВУ

Анотація. Основною метою математичного експерименту було дослідження впливу роздільної здатності ЦМР на якісні характеристики симулятивних ґрунтових карт, які отримуються шляхом моделювання при використанні типового набору матеріалів, які можуть бути потенційно доступними пересічному ґрунтознавцю чи науковцю в сучасних українських реаліях. При цьому показано, що морфометричні параметри рельєфу та його деривати є надійним базисом предикативного моделювання просторового поширення ґрунтових відмін з достатньо високою точністю, а представлена методика має значну перспективу в науково-виробничих задачах. На основі кореляційного аналізу була здійснена оцінка тісноти зв'язку та ролі згаданих параметрів у мінливості ґрунтового покриття, що з залученням аналізу головних компонент дозволило обрати 9 базових предикторів моделі: абсолютні висоти, топографічний індекс вологості, кількість сонячної радіації на одиницю площі, крутизну схилів, поздовжню та максимальну кривизну топографічної поверхні, акумуляцію, довжину та відстань до водних потоків. Зроблено розширену оцінку якості симулятивних ґрунтових карт при різних значеннях роздільної здатності ЦМР. Встановлено відмінності у якості прогнозних ґрунтових карт при використанні 14 основних типів предикативних алгоритмів та рекомендовано найбільш придатні для такого роду задач моделі, зокрема Decision Trees, Random Forests, та виокремлено деякі з них, які потенційно можуть показувати високі результати, зокрема Bugget Trees, K-Nearest Neighbors, Support Vector Machines та Neural Networks.

Ключові слова: симуляція, морфометричні параметри, ЦМР, ґрунтова карта, предикаційні алгоритми.

ВСТУП

Ретельний аналіз ситуації з ґрунтовими картографічними матеріалами в Україні показує, що наразі існує ряд важливих і складно розв'язуваних проблем, що стосуються питань їх наявності, доступності, актуальності та точності. Зокрема, для значних за площею територій досі не створено великомасштабних карт ґрунтів, особливо у гірській частині Карпат та Криму, для покритих лісовою рослинністю рівнинних територій, ареалів більшості населених пунктів тощо. У локаціях, де карти є в наявності, доводиться констатувати факт їх значної застарілості (Polchyna et al., 2004; Achasov et al., 2015; Cherlinka, 2017). Окрім цього, під час сучасних локальних досліджень ґрунтового покриття часто викриваються грубі помилки у контурах ґрунтової ситуації, наявність яких спричинена як недоліками застосовуваних на час обстежень методик картографування, так і помилками технічного характеру, низькою кваліфікацією виконавців, високими вимогами щодо термінів виконання робіт тощо. З відомих і доступних широкому загалу фахівців матеріалів відмітимо карту ґрунтів України М 1:2 500 000 та інші її варіанти в дрібно- та середньомасштабному представленні, карти ґрунтів областей М 1:200 000 та електронний варіант

Публічної кадастрової карти України. З існуючих карт більш крупного масштабу згадаємо картограми ґрунтів та їх агровиробничих груп М 1:10 000/25 000, створені для господарств за часів СРСР. Проте останні є недоступними для пересічних користувачів, а юридичні питання доступу до них та їх використання вимагають окремого дослідження. В основному весь цей масив інформації базується на матеріалах великомасштабного обстеження ґрунтів 1957–1961 рр. та їх часткового коректування, яке проводилося до початку 1990 рр. 20 століття. Отже, з 60,5 млн га земель України наразі охоплено великомасштабними обстеженнями близько 45 млн, тобто близько четвертини території країни досі являє собою різновеликі білі плями. Очевидний шлях подолання такої кризи – новітнє великомасштабне обстеження ґрунтів на основі інноваційних науково-методичних підходів з використанням наземних обстежень, безпілотних літальних апаратів (БПЛА) та дистанційного зондування Землі (ДЗЗ) має певні недоліки, зокрема: високі часові та трудові витрати, відповідні економічні затрати (які не на часі для бюджету держави) та відсутність у достатній кількості кадрового забезпечення. Тому, враховуючи сучасний політико-економічний стан України, логічним виглядає шлях моделювання, а саме аналіз наявних великомасштабних ґрунтових карт та відповідних їм цифрових моделей рельєфу і симуляція на цій основі предикативних моделей ґрунтового покриття для локацій з відсутніми даними, що є на даний момент актуальним і важливим у світлі оптимізації нормативної грошової оцінки земель та інших науково-практичних задач.

В останні кілька декад спостерігаються якісні та кількісні зміни в підходах до математичного моделювання в ґрунтознавстві, які візуалізуються в таких основних напрямках: моделювання ґрунтового профілю та базових його характеристик, або цифрова ґрунтова морфометрія (Hartemink and Minasny, 2014), створення карт ґрунтових параметрів (Pennock et al., 1987; Pachepsky et al., 2001; Park et al., 2001; Florinsky et al., 2002; Park and Van De Giesen, 2004; Florinsky et al., 2009; Malone et al., 2016), моделювання таксономічних ґрунтових одиниць (Merot et al., 1995; Zhu et al., 2001; McBratney et al., 2003; Walter et al., 2006; MacMillan, 2008; Mourier et al., 2008; Penizek and Boruvka, 2008; Barringer et al., 2008; Smith et al., 2010; Malone et al., 2016). Це вимагає застосування великого розмаїття математичних методів: багатofакторного регресійного аналізу, гібридних геостатистичних методів, регресійного крігінгу, методів нечіткої логіки, логістичної регресії, нейронних мереж тощо (Florinsky, 2012). Загальна ідея, яка лежить в основі застосування таких методів, полягає у використанні опорних точок місцевості та приурочених до них класифікаційних ґрунтових одиниць (Lagacherie et al., 2001). Аналіз цифрової моделі рельєфу (ЦМР) дозволяє виділити цілий ряд геоморфологічних параметрів, які пов'язані з даним ґрунтовим таксоном, і застосувати їх в якості предикторів. Оскільки самі ґрунтові таксони мають не числовий, а описовий характер, тобто відносяться до категоріального типу даних, а отримані з ЦМР показники, навпаки, – числові, то використання просунутого та витонченого математичного апарату є доволі логічним і широко використовуваним кроком (Giasson et al., 2008; Kempen et al., 2009; Debella-Gilo and Eitzelmüller, 2009; Hengl, 2009; Malone et al., 2016; Cherlinka, 2017)

Якщо цифрова модель рельєфу є основою, з якої видобувається переважна більшість предикторів, то важливого значення набуває знання про базові її характеристики, зокрема таку з найголовніших і практично недосліджених в даному контексті, як роздільна здатність ρ , або крок дискретизації моделі, питання вибору якого досі залишається дискусійним та обговорюваним в геоінформації та геоматиці (Hengl, 2006; Dmytruk and Cherlinka, 2012). Відповідно до цього метою нашої роботи стало дослідження впливу роздільної здатності на якісні характеристики симулятивних ґрунтових карт шляхом проведення математичного експерименту при використанні типового набору матеріалів, які можуть бути потенційно доступними пересічному ґрунтознавцю чи науковцю в сучасних

українських реаліях. До них ми відносимо насамперед власне великомасштабні ґрунтові та топографічні карти (М 1:10 000) й вільне програмне забезпечення – геоінформаційні системи GRASS (GRASS Development Team, 2017) і Quantum (QGIS Development Team, 2015), векторизатор Easy Trace (EasyTrace group, 2015) та мову статистичних розрахунків R-statistic (R Development Core Team, 2017).

МАТЕРІАЛИ ТА МЕТОДИ ДОСЛІДЖЕННЯ

Відповідно до поставленої мети нами були визначені такі завдання: а) оцифрування та атрибутування картографічних матеріалів; б) побудова серії ЦМР з роздільною здатністю ρ , рівною 2,5; 5; 10; 15; 20, 25, 30 та 50 м; в) аналіз цифрових моделей рельєфу і генерація з них у ГІС GRASS набору карт морфометричних та інших похідних характеристик; г) створення в R-statistic симулятивних моделей ґрунтового покриву (карт-версій або карт-моделей) з використанням 14 основних типів предикативних алгоритмів як для ареалів з наявною ґрунтовою інформацією, так і для тих, де вона не представлена; д) аналіз отриманих результатів і висновки щодо величини оптимальної роздільної здатності для прогнозного моделювання ґрунтового покриву та його тривалості.

В якості об'єкта було обрано фрагмент території України (рис. 1, а) в межах Чернівецької області (рис. 1, б), приурочений до Прут-Сіретського межиріччя (Північна Буковина) з контрастними геоморфологічними умовами і адміністративно належний Глибоцькому району (рис. 1, в). Даний ареал (4200×4200 м) має різне підпорядкування та господарське використання, а при його виборі були вирішені типові проблеми, що часто виникають при роботах такого характеру (Cherlinka and Dmytruk, 2014; Cherlinka, 2015; Cherlinka, 2017). Система координат проекту була приведена до СК 1963 (зона X2), сканований лист топографічної карти М 1:10 000 – М-35-136-Г-б-2 (рис. 1, г) георектифікувався з використанням створеної векторної математичної основи, а геоприв'язка ґрунтових карт здійснювалася до характерних точок місцевості та адміністративних меж існуючих на даний момент сільських рад (Михайлівської (н.п. Червона Діброва та Михайлівка) – А, Волоцької – Б, Валя-Кузьминської – В та Тарашанської – Г Глибоцького району Чернівецької області. Інформативні ґрунтові матеріали базувалися на серії архівних ґрунтових карт колгоспу «Україна» с. Тарашани (територія Тарашанської сільської ради (зйомка 1993 р.), колгоспу ім. Т. Г. Шевченка с. Михайлівка і с. Червона Діброва (територія Михайлівської сільської ради (обстеження 1992 р.) та колгоспу ім. Фрунзе с. Волока, с. Грушівка і с. Валя Кузьмина (територія сучасних Волоцької та Валя-Кузьминської сільських рад (обстеження 1991 р.). Після зведення номенклатурного списку ґрунтів до єдиної системи та узгодження контурів і типів ґрунтів на межах господарств інформація була дигіталізована та отримано попередні дані щодо відсотка покриття території ґрунтовими обстеженнями: з 1764 га загальної площі для 872 га, або 49,43 %, дані повністю відсутні (рис. 2, табл. 1). «Білі» плями приурочені до території населених пунктів та лісових масивів, які входять до меж згаданих сільських рад.

Для опрацювання даних були використані інструментальні можливості вільного програмного забезпечення: георектифікація картографічного матеріалу – ГІС Quantum (QGIS Development Team, 2015), оцифрування – Easy Trace (EasyTrace group, 2015), підготовка карт морфометричних параметрів – ГІС GRASS (GRASS Development Team, 2017) та симуляція карт ґрунтового покриву – мова статистичних розрахунків R- statistic (R Development Core Team, 2017). Серія цифрових моделей рельєфу з роздільною здатністю $\rho = \{2,5; 5; 10; 15; 20; 25; 30; 50\}$ м будувалася за допомогою регуляризованих напружених сплайнів (Mitášová and Mitáš, 1993) в ГІС GRASS. Вона послужила основою для виділення ряду морфометричних характеристик рельєфу: крутизни та експозиції схилів (модуль *r.slope.aspect* (Hofierka et al., 2009)), кривизни поверхні (профільної, планової, поздовжньої, мінімальної та максимальної) – модулем *r.param.scale* на основі (Wood, 1996), даних по сонячній

Таблиця 1

Площі ґрунтових відмін території симуляції за різних значень роздільної здатності моделі рельєфу, га/%

Шифр	Номенклатурний перелік ґрунтів	Архів-на карта									
		3	4	5	6	7	8	9	10	11	
0	Обстеження не проводилось	872,02/ 49,43	872,05/ 49,44	872,00/ 49,43	871,99/ 49,43	870,57/ 49,00	872,32/ 49,45	871,63/ 49,41	874,71/ 48,89	874,75/ 49,59	
1	Сірі лісові поверхнево-глеюваті крупнопилувато-легкосуглинкові на лесовидних суглинках	30,18/ 1,71	30,18/ 1,71	30,16/ 1,71	30,32/ 1,72	30,26/ 1,70	30,40/ 1,72	30,00/ 1,70	30,24/ 1,69	29,00/ 1,64	
2	Сірі лісові поверхнево-глеюваті слабозмиті крупнопилувато-легкосуглинкові на лесовидних суглинках	105,45/ 5,98	105,45/ 5,98	105,45/ 5,98	105,37/ 5,97	105,57/ 5,94	105,28/ 5,97	104,88/ 5,95	105,75/ 5,91	105,75/ 5,99	
3	Сірі лісові поверхнево-глеюваті середньозмиті крупнопилувато-середньосуглинкові на лесовидних суглинках	34,09/ 1,93	34,10/ 1,93	34,06/ 1,93	34,17/ 1,94	33,93/ 1,91	34,28/ 1,94	34,81/ 1,97	33,84/ 1,89	33,50/ 1,90	
4	Темно-сірі лісові поверхнево-глеюваті крупнопилувато-легкосуглинкові на лесовидних суглинках	49,40/ 2,80	49,39/ 2,80	49,46/ 2,80	49,42/ 2,80	49,95/ 2,81	49,36/ 2,80	49,19/ 2,79	48,15/ 2,69	49,25/ 2,79	
5	Темно-сірі лісові поверхнево-глеюваті слабозмиті крупнопилувато-легкосуглинкові на лесовидних суглинках	81,54/ 4,62	81,54/ 4,62	81,57/ 4,62	81,32/ 4,61	81,81/ 4,60	81,32/ 4,61	82,06/ 4,65	80,82/ 4,52	82,75/ 4,69	
6	Дерново-середньопідзолисті поверхнево-глеюваті крупнопилувато-легкосуглинкові на лесовидних суглинках	122,84/ 6,96	122,84/ 6,96	122,81/ 6,96	122,82/ 6,96	123,08/ 6,93	122,64/ 6,95	122,88/ 6,97	121,86/ 6,81	121,75/ 6,90	
7	Дерново-середньопідзолисті поверхнево-глеюваті слабозмиті крупнопилувато-легкосуглинкові на лесовидних суглинках	145,17/ 8,23	145,17/ 8,23	145,18/ 8,23	145,13/ 8,23	145,22/ 8,17	145,28/ 8,24	145,00/ 8,22	145,53/ 8,13	145,75/ 8,26	
8	Дерново-середньопідзолисті поверхнево-глеюваті середньозмиті крупнопилувато-легкосуглинкові на лесовидних суглинках	21,28/ 1,21	21,28/ 1,21	21,30/ 1,21	21,23/ 1,20	21,31/ 1,20	21,16/ 1,20	21,25/ 1,20	21,24/ 1,19	21,25/ 1,20	
9	Дерново-підзолисті поверхнево-глеюваті сильнозмиті на давніх зсувах крупнопилувато-важкосуглинкові на підстелених глинами лесовидних суглинках	178,74/ 10,13	178,74/ 10,13	178,70/ 10,13	178,58/ 10,12	178,99/ 10,07	178,04/ 10,09	179,19/ 10,16	177,93/ 9,94	176,00/ 9,98	

Закінчення табл. 1

1	2	3	4	5	6	7	8	9	10	11
10	Дерново-підзолисті поверхнево-глеюваті середньозмиті на давніх зсувах крупнопилувато-середньосуглинкові на підстелених глинами лесовидних суглинках	17,99/ 1,02	17,99/ 1,02	17,98/ 1,02	18,20/ 1,03	17,89/ 1,01	18,24/ 1,03	18,44/ 1,05	18,27/ 1,02	18,25/ 1,03
11	Лучно-болотні крупнопилувато-середньосуглинкові на давніх алювіально-делювіальних відкладах	70,91/ 4,02	70,89/ 4,02	70,91/ 4,02	71,05/ 4,03	70,99/ 4,00	71,40/ 4,05	70,75/ 4,01	70,65/ 3,95	71,00/ 4,02
12	Лучні піщанисто-середньосуглинкові на сучасних делювіальних відкладах	5,26/ 0,30	5,26/ 0,30	5,24/ 0,30	5,26/ 0,30	5,27/ 0,30	5,12/ 0,29	5,06/ 0,29	5,13/ 0,29	5,50/ 0,31
13	Виходи лесовидних суглинків крупнопилувато-важкосуглинкові	29,14/ 1,65	29,13/ 1,65	29,19/ 1,65	29,14/ 1,65	29,18/ 1,64	29,16/ 1,65	28,88/ 1,64	29,88/ 1,67	29,50/ 1,67
	* Дані відсутні (доплаткова площа, яка виникає при деяких роздільних здатностях для забезпечення кратності кроку моделі)					12,62/ 0,71			25,29/ 1,41	
	Сума площ обстежених та необстежених територій	1764,0/ 100,00	1764,0/ 100,00	1764,0/ 99,99	1764,0/ 99,99	1764,0/ 99,99	1764,0/ 99,99	1764,0/ 100,01	1764,0/ 98,59	1764,0/ 99,97
	Всього	1764,0/ 100,00	1764,0/ 100,00	1764,0/ 99,99	1764,0/ 99,99	1776,62/ 99,99	1764,0/ 99,99	1764,0/ 100,01	1789,29/ 100,00	1764,0/ 99,97

радіації – модулем r.sun (Hofierka and Šúri, 2002), форм рельєфу в r.geomorphon (Jasiewicz and Metz, 2011).

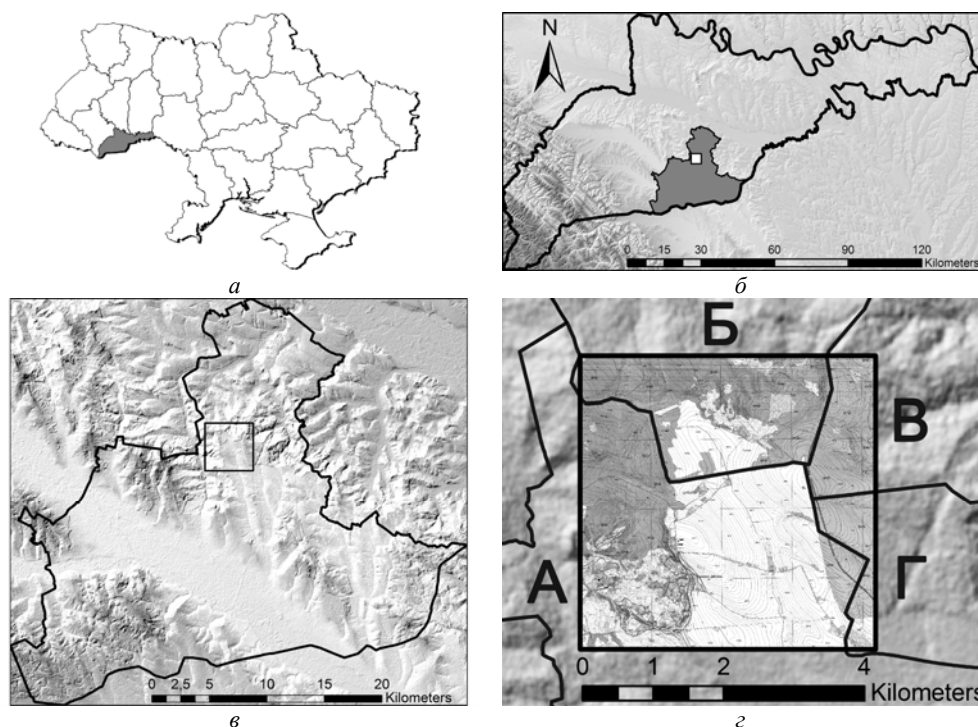


Рис. 1. Географічне розташування регіону досліджень у межах України (а), Чернівецької області (б), Глибоцького району (в) та схема тестового полігону (г). Для фону використано дані SRTM – NASA’s Shuttle Radar Topography Mission

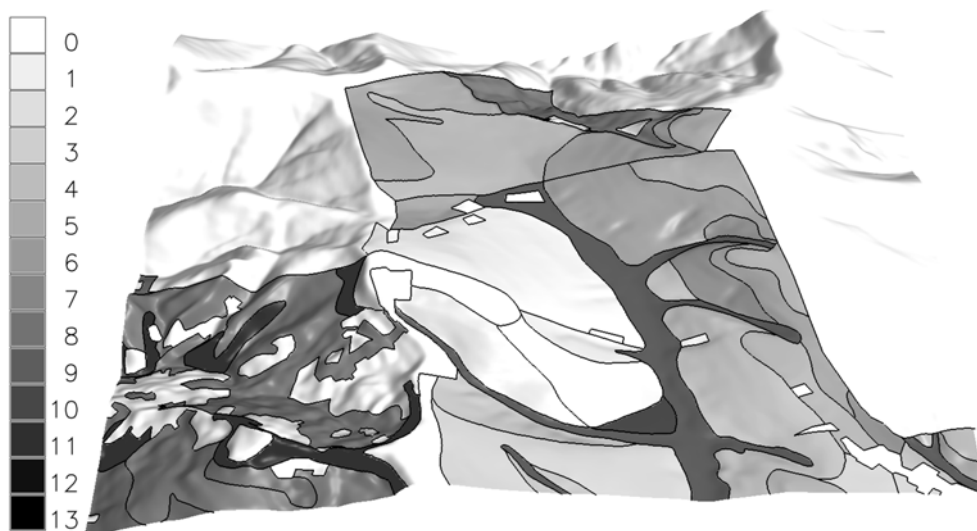


Рис. 2. Драпована оригінальною ґрунтовою картою цифрова модель рельєфу ($\rho = 2,5$ м) території досліджень (номери ґрунтів відповідають їх порядковим номерам у номенклатурному переліку в табл. 1)

Для створення симулятивних моделей ґрунтового покриття нами був написаний скрипт на мові R-statistic (R Development Core Team, 2017), який включає в себе ряд

адаптацій для вирішення поставлених завдань та реалізує 14 основних типів предикативних алгоритмів, зокрема: 1. Multinomial Logistic Regression – MLR (Giasson et al., 2008; Kempen et al., 2009; Debella-Gilo and Etzelmüller, 2009; Hengl, 2009; Malone et al., 2016; Cherlinka, 2017). 2. Neural Networks – NN (Venables and Ripley, 2002; Ripley and Venables, 2016). 3. Decision Trees – DT (Venables and Ripley, 2002). 4. Random Forests – RF (Breiman, 2001; Cutler et al., 2012). 5. Naive Bayes – NBS (Lowd and Domingos, 2005). K-Nearest Neighbors – KNN (Liu, 2011). 7. Flexible Discriminant Analysis – FDA (Liu, 2011). 8. Nonlinear Discriminant Analysis – NDA (Hastie et al., 2009). 9. Support Vector Machines – SVM (Venables and Ripley, 2002; Hastie et al., 2009). 10. Linear Discriminant Analysis – LDA (Hastie et al., 2009). 11. Partial Least Squares Discriminant Analysis – PLS (Kuhn and Johnson, 2013). 12. Penalized Logistic Regression – PLR (Kuhn and Johnson, 2013). 13. Nearest Shrunken Centroids – NSC (Hastie et al., 2009). 14. Bagged Trees – BGT (Hastie et al., 2009; Kuhn and Johnson, 2013).

Моделювання передбачало створення навчальної вибірки за запропонованою (Dobos and Hengl, 2009) та апробованою (Cherlinka, 2017) методикою і перенесення отриманих параметрів на повний набір даних. Для оцінки якості отриманих моделей використали індекс каппа Когеа κ (Landis and Koch, 1977; Li and Zhang, 2007; Kuhn, 2008; Grinand et al., 2008; Hengl, 2009; Malone et al., 2016), який в даному випадку показує ступінь відповідності між оригінальними та симулятивними даними.

РЕЗУЛЬТАТИ ТА ЇХ ОБГОВОРЕННЯ

Набір предикторів, ЦМР та ґрунтових таксонів був підданий кореляційному аналізу. При цьому, оскільки кореляційна матриця є гетерогенною, дослідили кореляцію Пірсона між числовими даними – висоти, експозиції, похили, кривизни тощо; полісеріальну між числовими та категоріальними змінними – ґрунтовими відмінами і геоморфологічними класами рельєфу (Fox, 2016) та V-коефіцієнт Крамера між категоріальними параметрами (Bildler and Loughin, 2014).

Значення даних коефіцієнтів між ґрунтовими таксонами та формами рельєфу варіює залежно від кроку дискретизації моделей рельєфу. Слід відмітити, що, незважаючи на досить низькі значення багатьох коефіцієнтів кореляції, вони в більшості своїй є статистично значущими при $p=0,05$, $0,01$ і навіть $0,001$. Деякі параметри (наприклад, мінімальна кривизна та абсолютні висоти, планова та профільна кривизна) показали тотожний характер варіювання та зв'язків з іншими ознаками, що вже на цьому етапі дозволило деякі з них (мінімальну та профільну кривизни) виключити з переліку предикторів моделі. Виходячи з того, що коефіцієнт кореляції Пірсона є мірою лінійної залежності і не виключає існування іншого її виду, а також наявний розкид варіювання та тісноти зв'язків, очевидно, що вибір предикторів лише на основі кореляційного аналізу є не зовсім коректним. Тому нами були залучені можливості методу головних компонент (Venables and Ripley, 2002), де шляхом виключення було встановлено набір з 9 відносно незалежних параметрів (абсолютні висоти, топографічний індекс вологості, кількість сонячної радіації на одиницю площі, крутизну схилів, поздовжню та максимальну кривизну топографічної поверхні, акумуляцію, довжину та відстань до водних потоків), які в подальшому використовувалися в якості базових предикторів при симуляціях.

Також зауважимо, що розмір ґрунтових ареалів на векторизованій та растрових картах дещо різняться залежно від величини роздільної здатності (табл. 1). Всупереч загальній сходимості результатів по площі (1764 га), аналіз кумулятивних відмінностей в площах по модулю показує наростання відмінностей залежно від кроку дискретизації моделі і в розглянутому їх інтервалі становить 0,10; 0,32; 1,19; 3,42; 3,17; 4,36; 8,87 та 11,54 га відповідно. При цьому спостерігається також аналогічне значне спотворення в контурах ґрунтової ситуації (рис. 3), що в крайніх випадках може призводити до «випадання» дрібних ареалів. За рахунок такого

рандомного нівелювання площ пряма та зворотна дискретизація/векторизація ґрунтових карт призводить до якісно відмінних результатів при роздільних здатностях вище 30 м, і тому орієнтування лише на табличні дані площ при оцінці точності отриманих моделей не буде коректним.

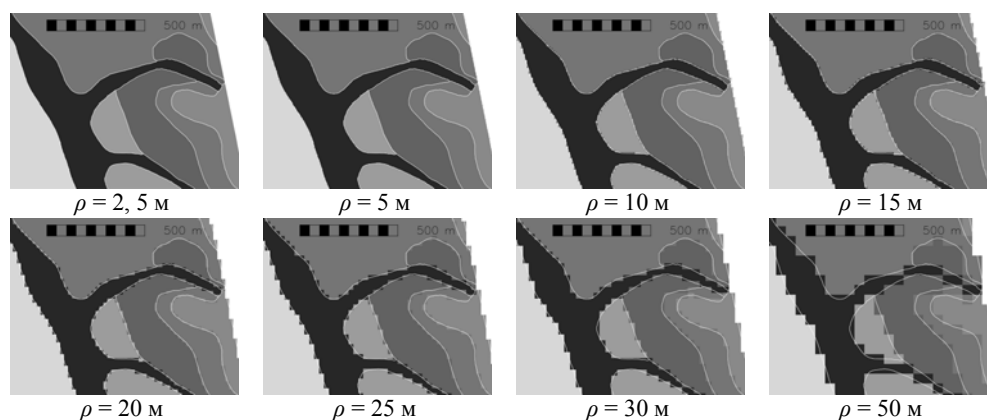


Рис. 3. Зміни в якості відображення контурів ґрунтів при різних кроках дискретизації моделі рельєфу

Кінцевим результатом проведеного модельного експерименту став набір симулятивних ґрунтових карт-версій для всієї території дослідження. Як було зауважено, карти-версії будувалися на основі навчальної вибірки, яка формувалася по медіанних осях ґрунтових таксонів. При цьому кількість навчальних пікселів значно різнилася для різних кроків дискретизації моделей і варіювала від 25919 для кроку 2,5 м (рис. 4, а), 2000 при $\rho = 25$ м (рис. 4, б) та 690 для частоти дискретизації 50 м (рис. 4, в).

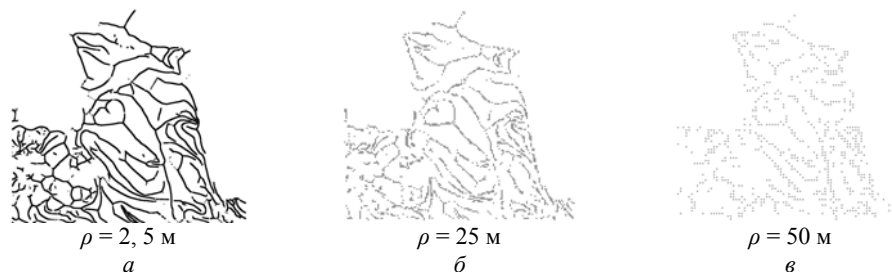


Рис. 4. Медіанні осі ґрунтових таксонів навчальної вибірки

Після отримання всього спектру розв'язків в R-statistic для навчальних вибірок провели перерахунок моделей на весь масив даних. Для оцінки якісних їх характеристик дані представили у вигляді матриці помилок з використанням крос-табуляції для показу співвідношення значень таксономічних класів, отриманих з архівної та предиктивних карт. При інтерпретації результатів було припущено, що прогнозна карта потенційно неточна, а архівна – навпаки, добре відображає реальну ґрунтову ситуацію. Для побудови матриць були використані всі комірки отриманих растрових карт, які містили інформацію про ґрунтові відміни. Головна діагональ цих матриць показує випадки, де розрахункові класи та дані архівної карти збігаються (правильна класифікація). Сума значень діагональних елементів показує загальну кількість правильно класифікованих пікселів растру, а відношення цієї кількості до загальної кількості елементів матриці вказує на загальну точність класифікації. Позадіагональні елементи даних матриць показують випадки

розбіжностей між розрахунковими і архівними таксонами, або так звані помилки класифікації. Загальним при прямому порівнянні площ між даними архівної та прогнозних карт в межах наявних даних є ситуація, коли деякі ґрунти з номенклатурного переліку або взагалі не попадають до прогнозу, або різко зменшують свою присутність. Лише кілька категорій нарощують свої площі, а у решти варіації відносно архівної карти коливаються від незначних до суттєвих. Ці коливання сильно залежать від класу використаної моделі і, на жаль, не дають чіткої картини точності симулятивних карт, оскільки не враховують просторову точність співпадіння таксонів ґрунтів. Навіть використання показника «чутливості», або Sensitivity (Kuhn, 2008), який показує, наскільки добре результат класифікації для класу збігається з перевіряльними даними, не є найбільш надійним параметром, оскільки він не враховує ймовірностей випадкової класифікації комірок растру (Grinand et al., 2008). Саме тому для оцінки точності предиктових карт ґрунтів нами також використано більш коректний індекс, а саме *kanma* Когена κ (Landis and Koch, 1977; Li and Zhang, 2007; Kuhn, 2008; Malone et al., 2016), яка за рахунок врахування ймовірностей хибної класифікації завжди мала значення трохи нижче, ніж загальна точність. Водночас це дозволяє з більшою коректністю оцінювати точність предикції в ареалах, де дані по ґрунтах були відсутні.

Отримані результати для наших даних по величині індексу κ (табл. 2) показують цікаві результати. По-перше, існує загальний тренд, коли значення κ для навчальних вибірок є вищими за κ побудованих на їх основі симулятивних моделей (окрім одного винятку в моделі Bugget Trees). По-друге, існують моделі, які явним чином за своїми неординарними результатами виділяються на загальному фоні, зокрема модель DT і RF, а також дуже нестандартний алгоритм BT, який при одних з найнижчих показників по навчальній вибірці стабільно має 3-тє місце (а при роздільній здатності 2,5 та 5 м і друге) по якості результатних моделей. Модель DT в даному дослідженні залишається поза конкуренцією, маючи найвищу прогностичну здатність (лише для $\rho = 10$ м має 2-ге місце). Модель RF є другою за якістю отримуваних результатів, зокрема займає 1-ше місце при $\rho = 10$ м і 2-ге за всіма іншими кроками дискретизації. Єдиним негативним моментом є те, що при $\rho = 2, 5$ м розв'язки не були отримані, що може бути зумовлено надзвичайною строкатістю вихідних умов при такій точності ЦМР та високій чутливості самої моделі. З перспективних моделей, які заслуговують на увагу при симуляціях ґрунтового покриття, відмітимо також моделі KNN, SVM та NN. Зокрема, Neural Network, показавши не найбільш високі результати на вітчизняному картографічному матеріалі, є лідером при аналізі закордонних ґрунтових карт, що опосередковано може свідчити про різницю в самих підходах та методології картографування.

Характеристика модельного експерименту буде неповною без інформації про такий важливий параметр, як час компутації. При циклічному переборі варіантів моделі виявилось, що при зменшенні роздільної здатності до 2,5 м він драматично зростає (рис. 5). Для орієнтовного розрахунку часових витрат можна скористатися рівнянням регресії у вигляді степеневі функції (наведене на рис. 5), де x – бажане значення ρ , а y – орієнтовний час обчислень. Слід зауважити, що попри високу точність даного рівняння в даному дослідженні ($R^2=0,9887$) співставими часові результати можуть бути отримані лише в умовах, подібних до експериментальних: операційна система Debian GNU/Linux 9 (stretch з ядром 4.9.0-3-amd64 x86_64) та процесор Intel Core I7-5700HQ CPU@3.50GHz з 16 Gb RAM. Спостереження за ходом дослідження та його подальший аналіз показали, що оптимізація компутаційного часу можлива шляхом виключення малоефективних предикативних алгоритмів, або вибором достатньої для користувачьких потреб роздільної здатності моделі.

Також викликає зацікавленість оцінка рівня співставимості отриманих нами результатів по якості симуляції з подібними дослідженнями. Так, у роботі (Hengl, 2009) 51–67 % вважається непоганим показником. У праці (Grinand et al., 2008) $\kappa = 67$ –87 % для навчальної вибірки і становить близько 30 % – для основного набору даних. Для дрібномасштабних ґрунтових карт Giasson et al. (2008) отримали значення

Таблиця 2

Розподіл величини індексу k залежно від кроку дискретизації ЦМР та типу симуляційної моделі

ρ, m	Тип симуляційної моделі													
	MLR		NN		DT		RF		NBS		KNN		FDA	
	$k_{\text{н}}$	k	$k_{\text{н}}$	k	$k_{\text{н}}$	k	$k_{\text{н}}$	k	$k_{\text{н}}$	k	$k_{\text{н}}$	k	$k_{\text{н}}$	k
2,5	63,38	60,09	73,48	68,78	100,00	77,56	–	–	73,22	56,54	98,48	69,27	62,80	47,24
5	66,75	61,27	71,59	68,52	100,00	77,74	100,00	77,40	73,59	58,39	98,27	70,67	67,14	50,97
10	64,85	61,01	73,94	67,96	100,00	75,89	100,00	76,25	72,69	58,34	96,15	70,05	66,60	51,47
15	62,83	61,93	69,35	68,06	100,00	76,00	100,00	75,04	68,78	55,10	99,43	68,66	64,51	51,22
20	67,61	64,23	71,04	66,67	100,00	78,12	100,00	76,94	68,54	56,01	93,10	70,43	64,27	50,19
25	71,02	66,55	76,66	71,44	100,00	77,56	100,00	77,27	70,27	51,18	91,82	72,43	67,77	54,16
30	74,00	66,98	73,91	72,01	100,00	78,54	100,00	77,93	73,86	56,81	92,23	72,61	63,02	47,85
50	77,96	65,63	71,34	76,87	100,00	79,98	100,00	79,93	79,66	63,78	90,35	70,30	70,72	55,66

Тип симуляційної моделі

ρ, m	NDA		SVM		LDA		PLS		PLR		NCS		BGT	
	$k_{\text{н}}$	k	$k_{\text{н}}$	k	$k_{\text{н}}$	k	$k_{\text{н}}$	k	$k_{\text{н}}$	k	$k_{\text{н}}$	k	$k_{\text{н}}$	k
2,5	58,21	42,44	88,16	68,25	51,47	36,76	25,09	24,89	0,16	1,16	37,05	34,38	14,67	74,98
5	63,40	46,32	88,04	67,97	52,76	37,12	25,75	25,20	0,32	1,79	38,73	28,59	14,78	72,84
10	67,06	49,23	86,54	66,83	50,88	35,15	26,15	25,46	0,25	2,01	36,59	27,63	15,62	74,00
15	67,96	47,30	84,10	66,63	46,21	32,16	22,69	25,93	0,21	1,15	35,75	28,89	17,11	73,28
20	67,14	49,19	83,93	68,28	44,62	33,77	24,00	27,75	0,84	3,15	32,64	28,45	13,54	74,97
25	73,06	55,18	84,40	70,28	51,71	39,06	25,96	28,95	0,36	1,87	38,71	31,50	14,48	75,91
30	78,00	60,87	81,48	68,87	56,59	45,60	30,57	31,18	1,03	3,67	37,76	30,61	10,55	76,91
50	88,68	69,51	83,64	68,87	55,76	44,94	48,01	39,45	0,96	3,30	44,70	38,58	5,73	78,80

* $k_{\text{н}}$ – капша навчальної вибірки, k – капша основного набору даних.

κ 37–54 %. У Malone et al. (2016) її величина варіює в межах 35–40 %. Тому для чіткої описової характеристики рівня співпадіння карт пропонуємо використовувати шкалу, розроблену Landis and Koch (1977), а саме: для $\kappa < 0,01$ – сходження відсутнє; 0,01–0,20 – незначне сходження; 0,21–0,40 – задовільне сходження; 0,41–0,60 – помірне сходження; 0,61–0,80 – істотне сходження; 0,81–0,99 – майже повне сходження. Відповідно до цього отримані нами дані свідчать про помірне в більшості та істотне сходження у кращих з представлених моделей, і це дозволяє оцінити якість симулятивних карт-версій як таку, що не поступається рівню аналогічних за літературними даними. Окрім того, вважаємо, що існує ще значний потенціал щодо підвищення загальної κ , зокрема шляхом більш ретельного добору предикторів моделі та розширення їх числа включенням даних дистанційного зондування Землі, карт антропогенних відкладів тощо.

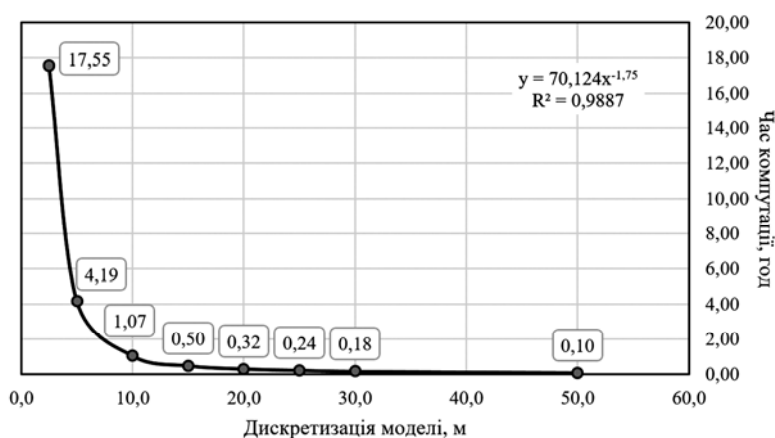


Рис. 5. Час компютації при різних кроках дискретизації ЦМР

Візуальний аналіз найліпших для кожної з роздільних здатностей симулятивних карт ґрунтового покриття в порівнянні з оригінальною картою (рис. 6) показує, що при великих кроках дискретизації моделі має місце втрата дрібних деталей ґрунтової ситуації, хоч κ при цьому досить висока. Тому при виборі оптимальної роздільної здатності ЦМР при окремих дослідженнях слід враховувати клас топографічної складності місцевості, дотримуючись рекомендації обирати вищу роздільну здатність при збільшенні насиченості рельєфу дрібними деталями. Варто зауважити, що точність симуляції залежить не тільки від типу моделі, набору предикторів, складності території в геоморфологічному відношенні, а найголовніше – від якості архівної карти, яка є основою для моделювання. Вона залежить як від методики картування, так і від складу виконавців, добротності польових та камеральних робіт. З останнім часто трапляються проблеми, хоча слід зауважити, що існують чудові картографічні матеріали ґрунтознавців «старої» школи. З негативних прикладів згадаємо вкрай неякісно побудовані картограми крутості схилів, яка лежить в основі виділення ґрунтових відмін різного ступеня змитості. Зрозуміло, що який би повний набір предикторів не використовувався, чи найбільш вишукані техніки створення навчальних вибірок/алгоритмів – прогнозна модель буде давати похибки відносно архівної карти. Врахуємо ще тотальну відсутність в наших умовах великомасштабних карт антропогенних відкладів – а це один з найголовніших предикторних параметрів. Незважаючи на згадані проблемні питання, вважаємо, що обраний нами спосіб моделювання непогано описує просторовий розподіл ґрунтів. Характеризуючи симуляційні варіанти отриманих нами карт-версій (рис. 6), відмітимо в загальному більш високу просторову роздільну здатність картографування, яка

виявляється в глибшому пропрацюванні контурів ґрунтової ситуації та вищій деталізації невеликих ареалів ґрунтів, ніж на вручну побудованій архівній карті. Це стосується, звичайно, карт-версій з роздільною здатністю до 30 м, оскільки при вищих значеннях ρ вже помітна значна генералізація як оригінальної карти, так і її предикативних варіантів (рис. 3, 6).

Значним корисним ефектом такого роду моделювання є можливість заповнення прогалів на існуючих картографічних матеріалах даними з предикативних карт-версій та отримання, таким чином, комбінованих ґрунтових карт. Звісно, це не виключає потреби у великомасштабному ґрунтовому обстеженні таких ареалів, але за відсутності можливості його проведення дозволяє отримати хоча б якісь дані на науковій основі з певним рівнем статистичної достовірності. Це також дозволяє її використовувати у прикладних задачах ґрунтознавства, агрономії, землевпорядкування та землеустрою, тобто сферах, де потреба в таких даних стоїть найбільш гостро.

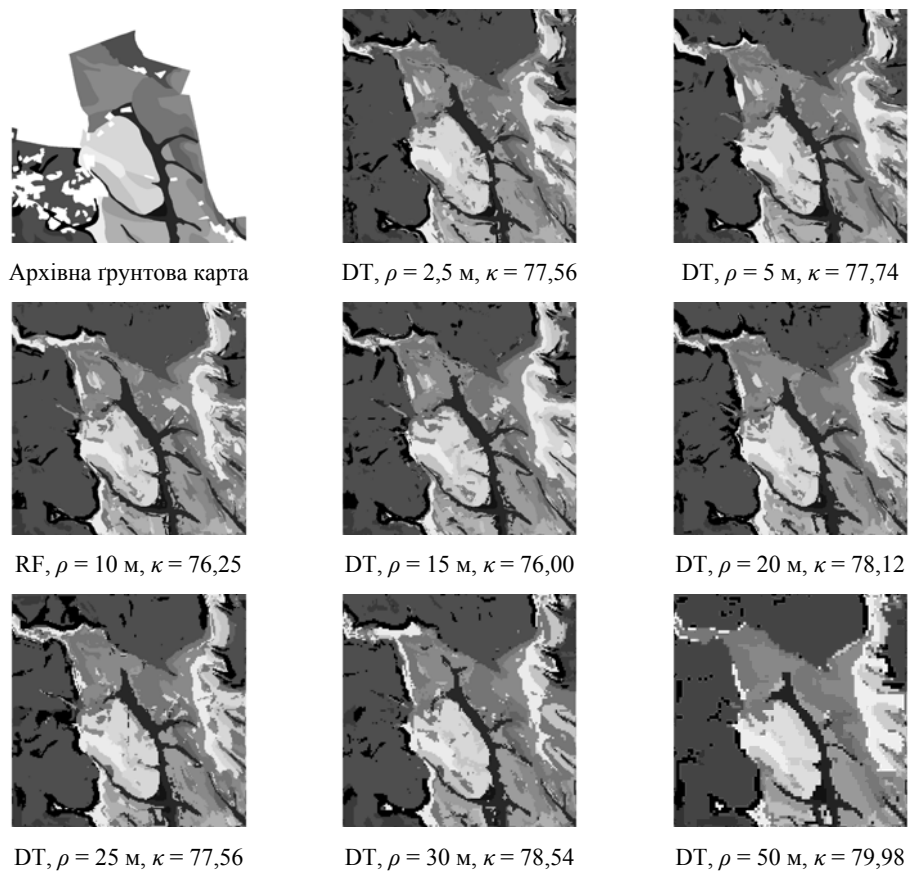


Рис. 6. Основні результати симуляції карт ґрунтового покриття (показано найточніші моделі для кожного значення роздільної здатності)

ВИСНОВКИ

Проведений математичний експеримент виявив, що існує значний вплив роздільної здатності ЦМР на якісні характеристики симулятивних ґрунтових карт, які отримуються шляхом моделювання при використанні типового набору матеріалів, які можуть бути потенційно доступними пересічному ґрунтознавцю чи науковцю в сучасних українських реаліях. При цьому показано, що морфометричні параметри

рельєфу та його деривати є надійним базисом предикативного моделювання просторового поширення ґрунтових відмін з достатньо високою точністю, а представлена методика має значну перспективу в науково-виробничих задачах. На основі кореляційного аналізу була здійснена оцінка тісноти зв'язку та ролі згаданих параметрів у мінливості ґрунтового покриву, що з залученням аналізу головних компонент дозволило обрати 9 базових предикторів моделі: абсолютні висоти, топографічний індекс вологості, кількість сонячної радіації на одиницю площі, крутизну схилів, поздовжню та максимальну кривизну топографічної поверхні, акумуляцію, довжину та відстань до водних потоків. Зроблено розширену оцінку якості симулятивних ґрунтових карт при різних значеннях роздільної здатності ЦМР. Встановлено відмінності у якості прогнозних ґрунтових карт при використанні 14 основних типів предикативних алгоритмів та встановлено найбільш придатні моделі для такого роду задач, зокрема Decision Trees, Random Forests, та виокремлено деякі з них, які потенційно за відповідного тюнінгу можуть показувати більш високі результати, зокрема Bugget Trees, K-Nearest Neighbors, Support Vector Machines та Neural Networks.

ПОДЯКИ

Дослідження виконане за фінансової підтримки Національної стипендіальної програми по обміну студентів, аспірантів, викладачів вузів і науковців Уряду Словацької Республіки, керованої SAIA – Словацькою академічною інформаційною агенцією. Грант №2016/2017: id17680.

Автор вдячний колегам з кафедри ґрунтознавства Чернівецького національного університету ім. Ю. Федьковича та Інституту географії Університету ім. П. Й. Шафарика в Кошицях (Словаччина) за корисні поради та критичні зауваження в ході виконання роботи.

СПИСОК ЛІТЕРАТУРИ / REFERENCES

- Achasov, A. B., Titenko, H. V., Kurilov, V. I., 2015. Dani dystantsiinoho zonduvannia yak osnova kartohrafuvannia gruntiv: ekonomichnyi aspekt [Data of remote sensing as the basis of soil mapping: the economic aspect]. Visnyk Kharkivskoho natsionalnoho universytetu imeni V. N. Karazina. Seriya: Ekolohiia (1104. – Vyp. 10), 60–66 (in Ukrainian).
- Arge, L., Chase, J. S., Halpin, P., Toma, L., Vitter, J. S., Urban, D., Wickremesinghe, R., 2003. Efficient flow computation on massive grid terrain datasets. *GeoInformatica* 7 (4), 283–313.
- Barringer, J. R. F., Hewitt, A. E., Schmidt, J., et al., 2008. National Mapping of Landform Elements in Support of S-Map, a New Zealand Soils Database. In: Zhou, Q., Lees, B., Tang, G.-a. (Eds.), *Advances in Digital Terrain Analysis*. Springer, Berlin Heidelberg, 443–458.
- Bilder, C. R., Loughin, T. M., 2014. *Analysis of Categorical Data* with R. Taylor & Francis, Boca Raton.
- Breiman, L., 2001. Random forests. *Machine learning* 45 (1), 5–32.
- Cherlinka, V. R., 2015. Adaptatsiia velykomasshtabnykh kart gruntiv do yikh praktychnoho vykorystannia u HIS [Adaptation of large-scale soil maps to their practical use in GIS.]. In: Ahrokhimiia i gruntoznavstvo. Mizhvidomchyi tematychnyi naukovyi zbirnyk. Vyp. 84. TOV «Smuhasta typohrafiia», Kharkiv, 20–28 (in Ukrainian).
- Cherlinka, V., Jan. 2017. Using Geostatistics, DEM and Remote Sensing to Clarify Soil Cover Maps of Ukraine. In: Dent, D., Dmytruk, Y. (Eds.), *Soil Science Working for a Living: Applications of soil science to present-day problems*. Springer-Verlag GmbH, Cham, Switzerland, Ch. 7, 89–100.
- Cherlinka, V. R., Dmytruk, Yu. M., 2014. Problemy stvorennia, heorektyfikatsii ta vykorystannia krupnomasshtabnykh tsyfrovnykh modelei reliefu [Problems of creation, georectification and use of large-scale digital elevation models]. *Heopolytyka y ekoheodynamyka rehyonov* 10 (1), 239–244 (in Ukrainian).
- Cutler, A., Cutler, D. R., Stevens, J. R., 2012. *Random Forests*. Springer US, Boston, MA, 157–175.
- Debella-Gilo, M., Eitzelmüller, B., Apr. 2009. Spatial prediction of soil classes using digital terrain analysis and multinomial logistic regression modeling integrated in

- GIS: Examples from Vestfold County, Norway. *Catena* 77 (1), 8–18.
- Dmytruk, Yu. M., Cherlinka, V. R., 2012. Vplyv rozdilnoi zdatnosti TsMR na tochnist vykhidnykh danykh pry modeliuванні protsesiv vodnoi erozii [Influence of DEM resolution on accuracy of output data in the modeling of water erosion processes]. In: *Fizychna heohrafiia ta heomorfologhiia. – Mizhvidomchyi naukovyi zbirnyk. – Vyp. 2 (66). VHL «Obrii», Kyiv, pp. 95–102 (in Ukrainian).*
- Dobos, E., Hengl, T., 2009. Soil mapping applications. In: Hengl, T., Reuter, H. I. (Eds.), *Geomorphometry: Concepts, Software, Applications. Vol. 33 of Developments in Soil Science.* Elsevier, Amsterdam, Ch. 20, 461–479.
- EasyTrace group, 2015. Easy Trace 7.99. Digitizing software.
- Florinsky, I. V., 2012. *Digital Terrain Analysis in Soil Science and Geology.* Academic Press. Elsevier, Amsterdam.
- Florinsky, I. V., Eilers, R. G., Manning, G. R., Fuller, L. G., Jan. 2002. Prediction of soil properties by digital terrain modelling. *Environmental Modelling & Software* 17 (3), 295–311.
- Florinsky, I. V., Eilers, R. G., Wiebe, B. H., Fitzgerald, M. M., Oct. 2009. Dynamics of soil salinity in the Canadian prairies: Application of singular spectrum analysis. *Environmental Modelling & Software* 24 (10), 1182–1195.
- Fox, J., 2016. Polycor: Polychoric and polyserial correlations. R package version 0.7-9, 1–10.
- Giasson, E., Figueiredo, S. R., Tornquist, C. G., Clarke, R. T., 2008. Digital soil mapping using logistic regression on terrain parameters for several ecological regions in Southern Brazil. In: Hartemink, E., McBratney, A. B., de Lourdes Mendonça-Santos, M. (Eds.), *Digital Soil Mapping with Limited Data.* Springer Netherlands, Amsterdam, Ch. 19, 225–232.
- GRASS Development Team, 2017. *Geographic Resources Analysis Support System (GRASS GIS) Software. Version 7.2.*
- Grinand, C., Arrouays, D., Laroche, B., Martin, M. P., Jan. 2008. Extrapolating regional soil landscapes from an existing soil map: Sampling intensity, validation procedures, and integration of spatial context. *Geoderma* 143 (1), 180–190.
- Hartemink, A. E., Minasny, B., Oct. 2014. Towards digital soil morphometrics. *Geoderma* 230-231, 305–317.
- Hastie, T., Tibshirani, R., Friedman, J., 2009. *The elements of statistical learning: Data mining, inference, and prediction, 2nd Edition.* Springer Series in Statistics. Springer, New York.
- Hengl, T., 2006. Finding the right pixel size. *Computers & Geosciences* (32), 1283–1298.
- Hengl, T., 2009. *A practical guide to geostatistical mapping, 2nd Edition.* University of Amsterdam, Amsterdam. URL <http://spatial-analyst.net/book/>
- Hofierka, J., Mitášová, H., Neteler, M., 2009. Geomorphometry in GRASS GIS. In: Hengl, T., Reuter, H. I. (Eds.), *Geomorphometry: Concepts, Software, Applications. Vol. 33 of Developments in Soil Science.* Elsevier, Amsterdam, Ch. 17, 387–410.
- Hofierka, J., Šúri, M., 11-13 September 2002. The solar radiation model for Open source GIS: implementation and applications. In: Ciolli, M., Zatelli, P. (Eds.), *Proceedings of the Open source GIS-GRASS users conference.* Dipartimento di Ingegneria Civile e Ambientale, Università di Trento, Trento, Italy, 1–19.
- Jasiewicz, J., Metz, M., Aug. 2011. A new GRASS GIS toolkit for hortonian analysis of drainage networks. *Computers & Geosciences* 37 (8), 1162–1173.
- Jasiewicz, J., Stepinski, T. F., Jan. 2013. Geomorphons – a pattern recognition approach to classification and mapping of landforms. *Geomorphology* 182, 147–156.
- Kempen, B., Brus, D. J., Heuvelink, G. B. M., Stoorvogel, J. J., Jul. 2009. Updating the 1:50,000 Dutch soil map using legacy soil data: A multinomial logistic regression approach. *Geoderma* 151 (3), 311–326.
- Kuhn, M., 2008. *Building Predictive Models in R Using the caret Package.* Journal of Statistical Software 28 (5), 1–26.
- Kuhn, M., Johnson, K., 2013. *Applied Predictive Modeling.* Springer, New York.
- Lagacherie, P., Robbez-Masson, J. M., Nguyen-The, N., Barthès, J. P., Apr. 2001. Mapping of reference area representativity using a mathematical soilscape distance. *Geoderma* 101 (3-4), 105–118.
- Landis, J. R., Koch, G. G., Mar. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33 (1), 159–174.
- Li, W., Zhang, C., 2007. A Random-Path Markov Chain Algorithm for Simulating Categorical Soil Variables from Random Point Samples. *Soil Science Society of America Journal* 71 (3), 656–668.
- Liu, B., 2011. *Web Data Mining: Exploring Hyperlinks, Contents and Usage Data, 2nd*

- Edition. Springer-Verlag GmbH, London New York Dordrecht.
- Lowd, D., Domingos, P., 2005. Naive bayes models for probability estimation. In: Proceedings of the 22nd international conference on Machine learning – ICML 05. Bonn, Germany. ACM Press, New York, NY, USA, 529–536.
- MacMillan, R. A., 2008. Experiences with applied DSM: protocol, availability, quality and capacity building. In: Hartemink, A. E., McBratney, A. B., de Lourdes Mendonça-Santos, M. (Eds.), Digital Soil Mapping with Limited Data. Springer Netherlands, Amsterdam, 113–135.
- Malone, B. P., Minasny, B., McBratney, A. B., 2016. Using R for Digital Soil Mapping. Progress in Soil Science. Springer International Publishing.
- McBratney, A. B., Santos, M. L. M., Minasny, B., Nov. 2003. On digital soil mapping. Geoderma 117 (1-2), 3–52.
- Merot, P., Ezzahar, B., Walter, C., Arousseau, P., Jan. 1995. Mapping waterlogging of soils using digital terrain models. Hydrological Processes 9 (1), 27–34.
- Mitášová, H., Hofierka, J., 1993. Interpolation by regularized spline with tension: II. Application to terrain modeling and surface geometry analysis. Mathematical Geology 25 (6), 657–669.
- Mitášová, H., Mitáš, L., 1993. Interpolation by regularized spline with tension: I. Theory and implementation. Mathematical Geology 25 (6), 641–655.
- Moore, I. D., Gessler, P. E., Nielsen, G. A., Peterson, G. A., 1993. Soil attribute prediction using terrain analysis. Soil Science Society of America Journal 57 (2), 443–452.
- Mourier, B., Walter, C., Merot, P., Jan. 2008. Soil distribution in valleys according to stream order. Catena 72 (3), 395–404.
- Pachepsky, Y. A., Timlin, D. J., Rawls, W. J., 2001. Soil water retention as related to topographic variables. Soil Science Society of America Journal 65 (6), 1787–1795.
- Park, S. J., McSweeney, K., Lowery, B., Oct. 2001. Identification of the spatial distribution of soils using a process-based terrain characterization. Geoderma 103 (3-4), 249–272.
- Park, S. J., Van De Giesen, N., Aug. 2004. Soil-landscape delineation to define spatial sampling domains for hillslope hydrology. Journal of Hydrology 295 (1-4), 28–46.
- Penizek, V., Boruvka, L., 2008. The digital terrain model as a tool for improved delineation of alluvial soils. In: Hartemink, A. E., McBratney, A. B., de Lourdes Mendonça-Santos, M. (Eds.), Digital Soil Mapping with Limited Data. Springer Netherlands, Amsterdam, 319–326.
- Pennock, D. J., Zebarth, B. J., De Jong, E., Oct. 1987. Landform classification and soil distribution in hummocky terrain, Saskatchewan, Canada. Geoderma 40 (3-4), 297–315.
- Polchyna, S. M., Nikorych, V. A., Danchu, O. A., 2004. Zastosuvannya suchasnoi systemy klasyfikatsii gruntiv FAO/WRB do karty gruntovoho pokryvu Chernivetskoï oblasti [Application of the modern FAO / WRB soil classification system to the map of soil cover of Chernivtsi region]. Gruntoznavstvo 5 (1–2), 27–33 (in Ukrainian).
- QGIS Development Team, 2015. QGIS Geographic Information System. URL <http://qgis.osgeo.org>
- R Development Core Team, 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing. URL <http://www.r-project.org>
- Ripley, B., Venables, W., 2016. R-package nnet: Feed-forward neural networks and multinomial log-linear models. v.7.3-12.
- Smith, S., Bulmer, C., Flager, E., Frank, G., Filatow, D., 2010. Digital soil mapping at multiple scales in British Columbia, Canada. In: Program and Abstracts, 4th Global Workshop on Digital Soil Mapping, 24–26 May 2010, Rome, Italy. International Union of Soil Sciences, IUSS, Rome, 1–9.
- Venables, W. N., Ripley, B. D., Feb. 2002. Modern Applied Statistics with S, 4th Edition. Vol. 53 (1) of Statistics and Computing. Springer-Verlag, New York.
- Walter, C., Lagacherie, P., Follain, S., 2006. Integrating pedological knowledge into digital soil mapping. In: Lagacherie, P., McBratney, A. B., Voltz, M. (Eds.), Digital Soil Mapping: An Introductory Perspective. Vol. 31 of Developments in Soil Science. Elsevier, Amsterdam, Ch. 22, 281–301.
- Wood, J. D., 1996. The geomorphological characterisation of digital elevation models. Phd thesis, University of Leicester, UK, Leicester.
- Zhu, A.-X., Hudson, B., Burt, J., Lubich, K., Simonson, D., 2001. Soil mapping using GIS, expert knowledge, and fuzzy logic. Soil Science Society of America Journal 65 (5), 1463–1472.

Стаття надійшла в редакцію: 25.05.2017