

УДК 911.2

НЕЧІТКИЙ КЛАСТЕРНИЙ АНАЛІЗ В ЛАНДШАФТОЗНАВЧОМУ КАРТОГРАФУВАННІ

Свідзінська Д.В.

Київський національний університет імені Тараса Шевченка

Дискретність-континуальність ландшафту обумовлює високий потенціал використання в ландшафтознавчому картографуванні слабкоформалізованих методів. На цій основі розглянуто методичні засади і практичний досвід застосування алгоритму нечітких k -середніх для відображення та аналізу морфодинамічної конфігурації ландшафту. **Ключові слова:** нечіткий кластерний аналіз, ландшафтознавче картографування, нечіткі k -середні, морфодинамічні мікрогеохори, SRTM ЦМВ

Постановка проблеми. Дискретність-континуальність є однією з найбільш важливих загальних властивостей ландшафту [2]. Неможливість її представлення в межах єдиної картографічної моделі завжди пов'язувалась з протилежністю прояву дискретності через ареали, а континуальності – через поле. На думку [2,6] потенціал для подолання цих складнощів закладено в теорії нечітких множин, що актуалізує питання картографування на її основі. Це дає можливість по-новому поглянути на градієнтні та математико-статистичні критерії, які є основою сучасного автоматизованого ландшафтознавчого картографування.

Аналіз останніх досліджень. На відміну від традиційної жорсткої класифікації теорія нечітких множин дозволяє використовувати показник, що контролює міру перекриття між класами, тобто один об'єкт може одночасно належати до двох або більше класів [1,2,5]. Завдяки цій властивості нечітке групування є більш доцільним для аналізу континуальних явищ, параметри яких змінюються поступово – його активно застосовують в класифікуванні даних про форми рельєфу [8], ґрунти [11], ландшафти [7], розпізнаванні образів і обробці даних дистанційного зондування [5]. Теоретико-методологічні засади нечіткого кластерного аналізу детально висвітлено в роботах [1,2,5,7], тому ми зупинимось лише на принципових методичних моментах застосування алгоритму нечітких k -середніх (також відомого як c -середніх або FCM від англ. Fuzzy C-Means), що є одним з найбільш популярних методів кластеризації [7,8,11,16 та ін.].

Ціль та завдання дослідження. Ціллю дослідження є розробка та апробація методу картографування і аналізу морфодинамічної конфігурації ландшафту на основі нечіткого кластерного аналізу методом k -середніх. При цьому послідовно вирішуються наступні завдання: 1) ознайомлення з методичною сутністю алгоритму нечітких k -середніх; 2) інтерпретація кількісних показників, що використовуються для оцінювання результатів класифікації; 3) відо-

браження та аналіз морфодинамічної конфігурації ландшафту через нечітку кластеризацію.

Точність картографування відповідає масштабу 1:100 000, а виділені на основі геоморфометричного ГІС-аналізу ЦМР [16] елементарні ділянки ландшафту – рангу мікрогеохор (урочищ). Таким чином, одиницею картографування виступають морфодинамічні мікрогеохори, як елементарні ділянки ландшафту однорідні за особливостями форми поверхні (спадистість, опуклість-увігнутість) та динамічними параметрами (напрямок, інтенсивність) латеральних абіотичних потоків.

Територія дослідження. Тестова ділянка площею 100 км² розташована в межах Орільсько-Самарської низовинної області Північностепової підзони [3].

В геоморфологічному відношенні територія являє собою акумулятивно-денудаційну, увалисту, сильно-розчленовану рівнину [4] з діапазоном абсолютних висот 70-168 м та ухилом поверхні від 0 до 8°.

Матеріали та програмне забезпечення. Вихідним матеріалом дослідження стала відкрита ЦМВ SRTM [9] приведена до роздільної здатності в 50 м та підготована для подальшого аналізу за допомогою процедур фільтрації [15,17] і гідрологічної корекції [14]. На її основі було розраховано ряд геоморфометричних показників, що описують форму поверхні та динамічні властивості латеральних абіотичних потоків [16]. Поверхні розподілу значень цих показників послуговували емпіричною основою для проведення процедури нечіткого групування даних.

Кластерний аналіз реалізовано в середовищі програмного забезпечення FuzMe v. 3.5с, яке є універсальним засобом виконання некерованої нечіткої класифікації даних з можливостями оцінки її загальної невизначеності та обґрунтування оптимального числа класів [10].

Нечіткий кластерний аналіз методом k -середніх. Процедура кластеризації ґрунтувалась на використанні Евклідової дистанції, яка являє собою просту відстань в багатовимірному ознаковому просторі. Застосування цієї метрики не включає попереднього масштабування даних,

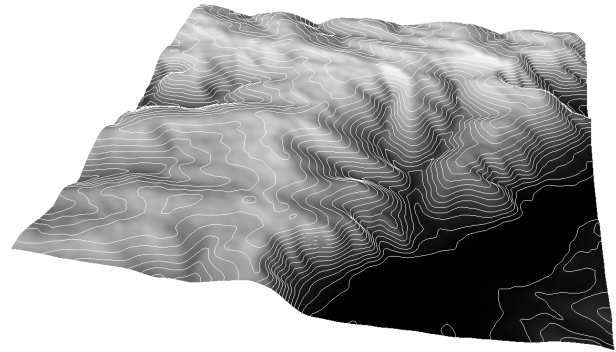
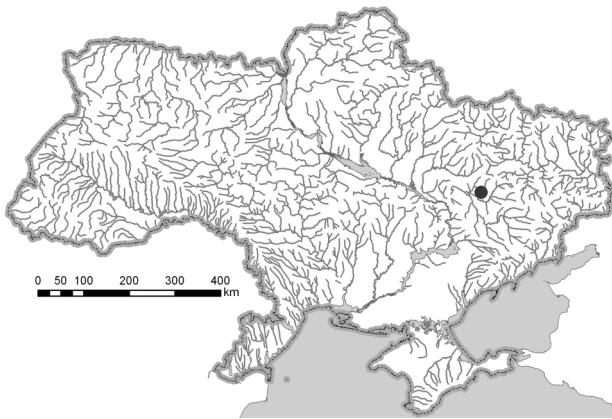


Рис. 1 Місцерозташування та зовнішній вигляд тестової ділянки

тому з метою усунення похибки всі змінні попередньо було стандартизовано. Це дозволило звести значення за усіма параметрами до уніфікованого діапазону $[0;1]$. Після вибору метрики та підготовки даних можливо приступити до виконання ітераційної процедури кластеризації на основі алгоритму FCM (рис. 2).

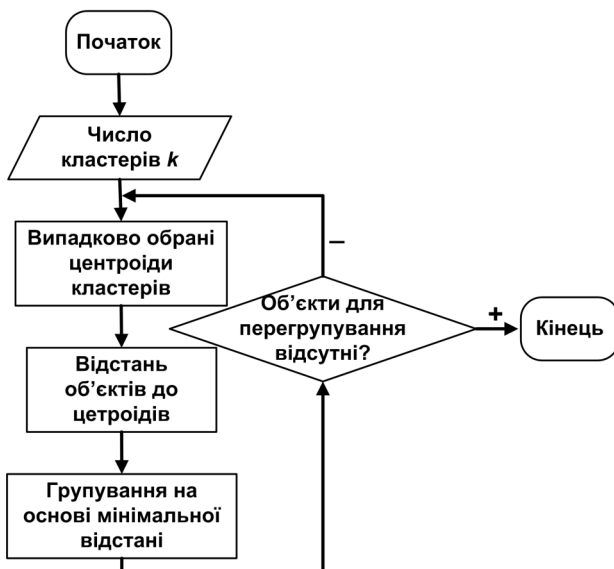


Рис. 2 Алгоритм кластерного аналізу методом k -середніх

Основними вихідними параметрами для здійснення нечіткого групування даних є кількість класів та показник нечіткості класифікації. В нашому випадку аналіз здійснювався для кількості кластерів від 2 до 10. Міра нечіткості класифікації кількісно контролює ступінь перекриття класів. Її значення можуть коливатись в діапазоні від 1 (жорстке групування, коли кожен об'єкт належить тільки до одного класу) до ∞ (відсутність окремих класів в результаті їх повного накладання). Питання визначення оптимального значення міри нечіткості детально викладені в багатьох публікаціях (зокрема, [12]), де її діапазон коливається в середньому від 1,1 до 3. Оскільки робота не розглядає вплив змін параметрів кластерного аналізу на результат класифікації (це потребувало

б окремого дослідження), значення міри нечіткості було встановлене на компромісному рівні 1,3, що добре працює з обраною метрикою [13].

Процедура кластеризації була завершена згідно встановлених вихідних параметрів, але FCM не завжди знаходить оптимальну конфігурацію, що співвідноситься з мінімумом функції відстані між об'єктом та центром кластера. Також цей метод є чутливим до випадково обраних початкових центрів кластерів на основі яких відбувається наступний перерозподіл об'єктів. З метою усунення цих ефектів для визначення центрів кластерів було обрано опцію довільного початку процедури, тобто стартові кластери обирались випадковим шляхом. Ця опція може суттєво змінити результати розрахунків для одного і того ж самого набору даних. Після того, як число кластерів і значення індексу нечіткості було визначено, кластеризація набору даних повторювалась десять разів. В усіх випадках тестування отримані результати були майже ідентичними, що ілюструє якісне та стабільне відображення груп, утворених вихідними даними.

Оцінювання результатів кластеризації. Для об'єктивного відбору результатів використовувався показник лямбди Уїлкса (Wilk's lambda), який являє собою загальну міру відмінності між кластерами [5]. Значення показника близькі до 0 ідентифікують відмінності між середніми значеннями групи, а близькі до 1 – їх подібність. Таким чином, було відібрано результати з найнижчими значеннями лямбди Уїлкса серед 10 повторень.

Обґрунтування оптимального числа кластерів спиралось на показники якості нечіткої класифікації [5,7,11]:

- індекс нечіткості виконання кластеризації (Fuzziness Performance Index – FPI) – міра змішаності класів, значення якої коливаються в діапазоні від 0 до 1, де 0 відповідає жорсткому розподілу з чіткими класами.

- модифікована невизначеність розділення (Modified Partition Entropy – MPE) – оцінює рівень

дезорганізації, що створюється певним числом класів, де значення близькі до 1 означають високий рівень дезорганізації, а 0 – виразу організації.

- дистанція відокремлення (separate distance – S) – відношення загальної варіативності розділення, центроїдів кластерів та відокремленості їх векторів.

Мінімальні значення вищенаведених індексів ідентифікують найкраще розділення і допомагають визначити оптимальну кількість класів. Як видно з рис. 3, показники демонструють часткову узгодженість, що пояснюється різними підходами до їх розрахунку. Тим не менше, діагностуються два альтернативні варіанти розподілу – 3 та 6 класів відповідно, що дає можливість співставити між собою результати класифікації.

Картографічне відображення результатів класифікації ґрунтувалось на відтворенні геопросторового розподілу значень міри належності та індексу сплутаності. Міра нечіткої належності (fuzzy membership value) покликана кількісно оцінити наскільки певний елемент множини відноситься до даного класу і коливається в діапазоні від 0 (не належить) до 1 (належить) [5,7,11]. Перехід від нечіткої до «жорсткої» класифікації можливий на основі її максимальних значень. Індекс сплутаності класифікації (confusion index – CI) інформує наскільки добре був класифікований кожний окремий об'єкт [5,7,11]. Якщо CI близький до 0, елемент демонструє чітку належність до домінуючого класу; коли CI наближається до 1, відмінності у значеннях міри належності до двох домінуючих класів незначні, що породжує сплутаність між ними.

Представлення морфодинамічної конфігу-

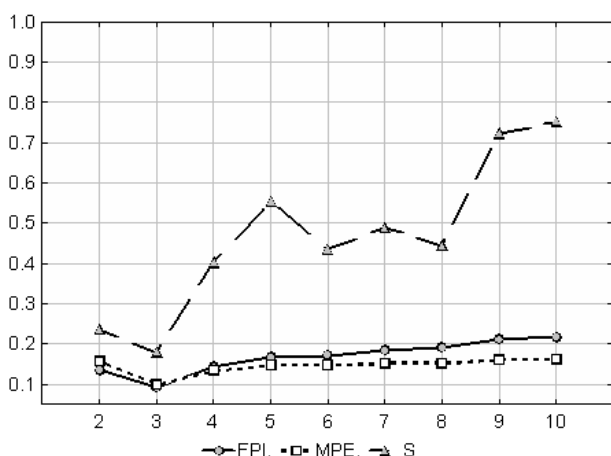


Рис. 3 Визначення оптимального числа класів

рації ландшафту через нечітку кластеризацію. Успішність процедури у виявленні класів, придатних до картографічного відображення, оцінювалась на основі відтворення геопросторового

розподілу значень міри належності, що були розраховані для кожного пікселя в кожному класі (рис. 4). Карти демонструють, що більша частина території була безпомилково віднесена до одного класу. Така однозначність є результатом значущої геопросторової кореляції у вихідних даних [6] і робить доцільним зведення всього набору карт до вигляду компактних «жорстких» класів (рис. 5).

Картографічне відображення геопросторового розподілу значень CI (рис. 5 а,с) відтворює ділянки де класифікація була неуспішною. Лінії максимальних значень CI ідентифікують межі між класами, а оконтурені ними ділянки можуть бути віднесені до єдиного «жорсткого» класу (рис. 5 b,d). З детальною семантичною інтерпретацією морфодинамічних мікрогеохор у випадку розподілу на 3 та 6 класів можна ознайомитись в [16]. Картографічне ж їх відображення демонструє геопросторову значущість розділення типів 1 та 2 із збільшенням числа класів від 3 до 6.

Обговорення та висновки. Некерований нечіткий кластерний аналіз методом *k*-середніх – об'єктивна і достовірна процедура групування даних, яка дозволяє уникнути більшості складнощів, що властиві загальноприйнятим підходам класифікування. По-перше, FCM надає можливості визначення оптимального (при даному наборі параметрів) числа класів та аналізу особливостей поширення даних всередині них. По-друге, завдяки об'єктивності вихідних даних та статистичних параметрів класифікації, виключається суб'єктивізм дослідника. По-третє, за рахунок використання методів нечіткого кластерного аналізу усувається жорстка детермінованість, невласлива природним утворенням в реальності. При цьому створюються широкі можливості статистичної інтерпретації результатів (класів, їх достовірності, міри належності), які є більш об'єктивними і незалежними від дисциплінарного поля, ніж традиційне семантичне наповнення.

Для подальшого удосконалення методики важливо розглянути декілька питань, які стосуються як відпрацювання самої класифікаційної процедури, так і підходів до її інтерпретації. Для покращення самого класифікаційного алгоритму вартими аналізу є питання пов'язані з впливом значення міри належності на результати кластеризації, потенціалом застосування інших метрик дистанції (наприклад, Махаланобіса). Також заслуговує на увагу розширення набору вихідних даних, наприклад, розрахованими на основі даних дистанційного зондування спектральними індексами (зокрема, NDVI), що зробить класифікацію більш комплексною і надасть їй ландшафтно-екологічного змісту.

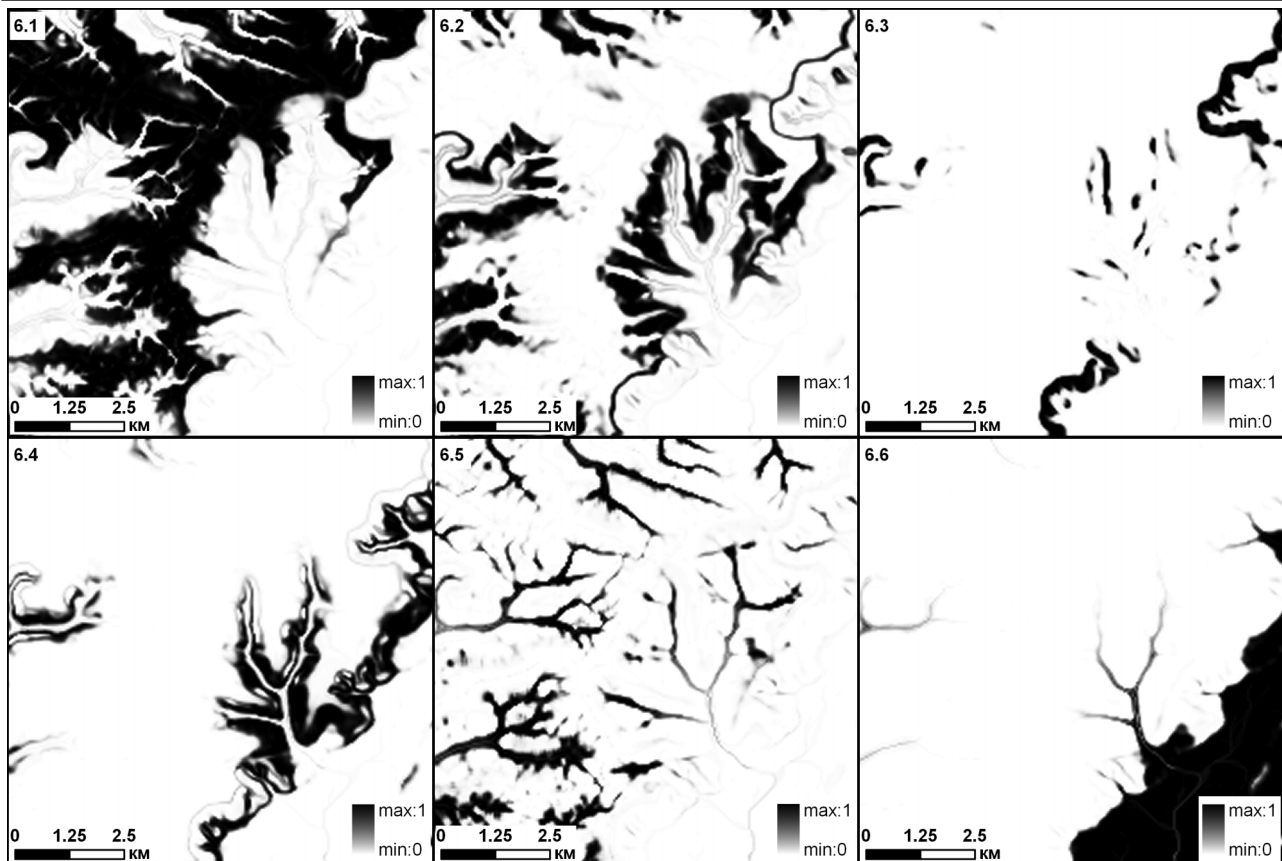


Рис. 4 Геопросторовий розподіл значень міри належності для 6 класів

Інтерпретаційний потенціал алгоритму пов'язаний з можливістю визначення і картографування каркасних та структурних ліній рельєфу, ядер типовості та екотонних смуг ландшафту на основі значень СІ. Масиви емпіричних даних, що генеруються в процесі аналізу, варто покласти в основу кількісного оцінювання мір подібності між місцями ландшафту та їх нечіткої ординації.

Список літератури

1. Андреев В.Л. Анализ эколого-географических данных с использованием теории нечетких множеств. – Л.: Наука, 1987. – 154 с.
2. Гродзинський М.Д. Пізнання ландшафту: місце і простір. Монографія у 2-х т. – К: ВПЦ «Київський університет», 2005. – Т.1. – С. 60-68, 416-427
3. Маринич О.М., Пархоменко Г.О., Петренко О.М., Шищенко П.Г. Удосконалена схема фізико-географічного районування України // Український географічний журнал. – 2003. – №3. – С. 16-21
4. Паленко В.П., Барщевський М.Є., Бортник С.Ю. та ін. Загальне геоморфологічне районування території України // Український географічний журнал. – 2004. – №1. – С. 3-11
5. Bezdek J., Keller J., Krisnapuram R., Pal N. Fuzzy Models and Algorithms for Pattern Recognition and Image Processing (the Handbooks of Fuzzy Sets). – New York: Springer, 2005. – 785 p.
6. Burrough P.A., van Gaans P.F.M, Hootsmans R.J. Continuous classification in soil survey: spatial correlation, confusion and boundaries // Geoderma. – 1997. – vol. 77. – P. 115–135
7. Burrough P., Wilson J., van Gaans P., Hansen A. Fuzzy k-means classification of topo-climatic data as an aid to forest mapping in the Greater Yellowstone Area, USA // Landscape Ecology. – 2001. – vol. 16. – P. 523-546
8. Burrough P., van Gaans P., MacMillan R. High-resolution landform classification using fuzzy k-means // Fuzzy Sets and Systems. – 2000. – vol. 113. – P. 37-52
9. Jarvis A., Reuter H., Nelson A., Guevara E. Hole-filled seamless SRTM data V4. International Centre for Tropical Agriculture (CIAT). – 2008. – <http://srtm.csi.cgiar.org> – (23.02.2012)
10. Minasny B., McBratney A. FuzME v. 3.5c. Australian Centre for Precision Agriculture. The University of Sydney, Australia. – 2002. – <http://sydney.edu.au/agriculture/pal/software/fuzme.shtml> – (23.02.2012)
11. McBratney A., Odeh I. Application of fuzzy sets in soil science: fuzzy logic, fuzzy measurements and fuzzy decisions // Geoderma. – 1997. – vol. 77(2-4). – P. 85-113
12. Okeke F., Karnieli A. Linear mixture model approach for selecting fuzzy exponent value in fuzzy c-means algorithm // Ecological Informatics. – 2006. – No. 1. – P. 117-124
13. Panchal V.K., Kundra H., Kaur J. Comparative Study of Particle Swarm Optimization based Unsupervised Clustering Techniques // International Journal of Computer Science and Network Security. – 2009. – Vol. 9, No.10. – P. 132-140
14. Planchon O., Darboux F. A fast, simple and versatile algorithm to fill the depressions of digital elevation models // Catena. – 2002. – vol. 46(2-3). – P. 159-176
15. Selige T., Buhner J., Ringeler A. Processing of SRTM X-SAR data to correct interferometric elevation models

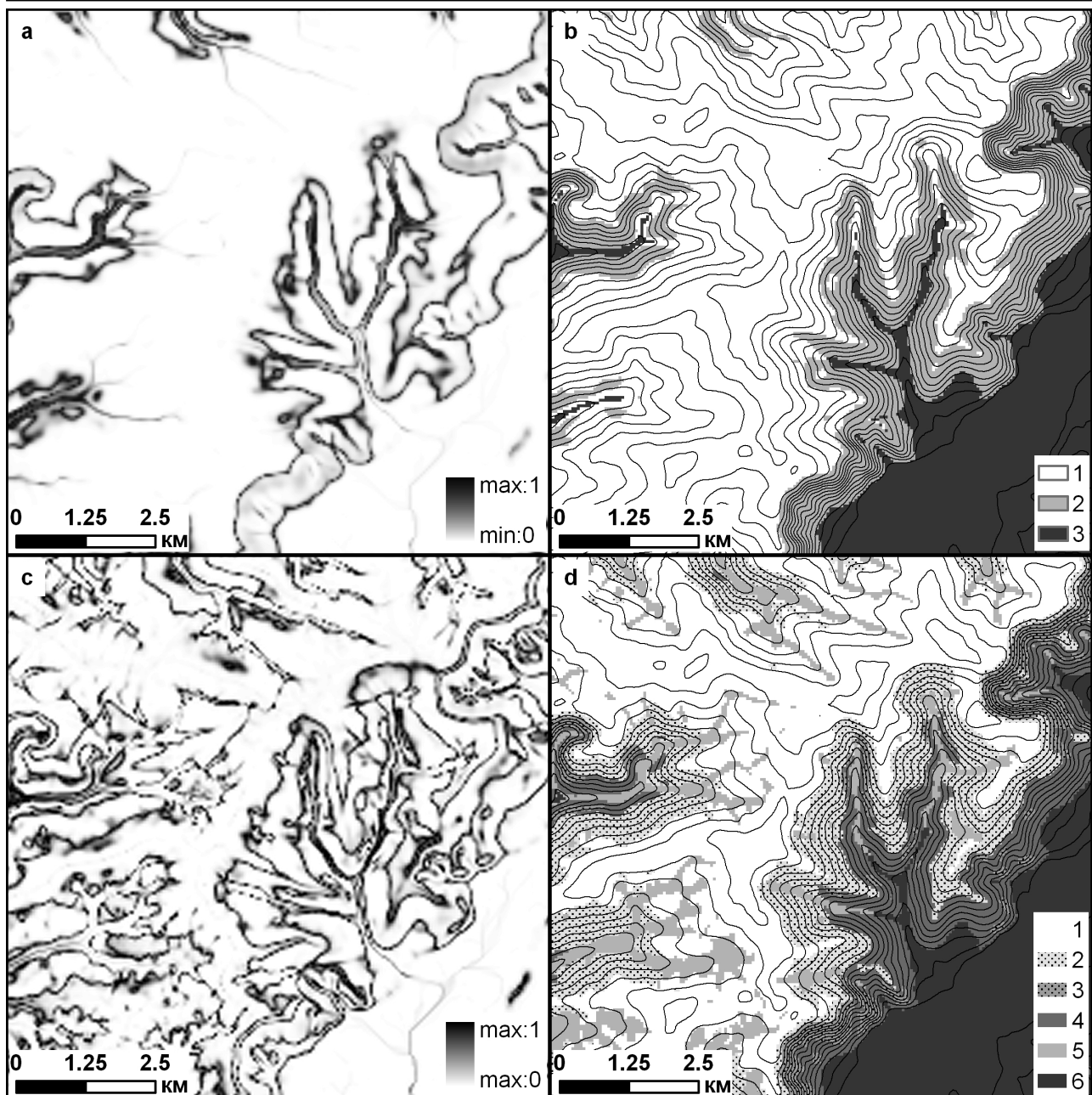


Рис. 5 Геопросторовий розподіл значень СІ (а, с) та представлення результатів класифікації у вигляді «жорстких» класів (b, d)

for land surface process applications // SAGA – Analysis and Modelling Applications. Göttinger Geographische Abhandlungen. – 2006. – vol. 115. – P. 97-104

16. Svidzinska D. Mapping of landscape spatial dynamics patterns by the fuzzy clustering analysis // The Problems

of Landscape Ecology. – 2011. – vol. 30 (у друці)

17. Vosselman G. Slope based filtering of laser altimetry data // IAPRS. – 2000. – Vol. XXXIII, Part B3. – Amsterdam, The Netherlands. – P. 935-942

Свидзинская Д.В. Нечеткий кластерный анализ в ландшафтном картографировании. Дискретность-континуальность ландшафта обуславливает высокий потенциал использования в ландшафтном картографировании слабоформализованных методов. Исходя из этого рассмотрены методические положения и практический опыт применения алгоритма нечетких k-средних для отображения и анализа морфодинамической конфигурации ландшафта.

Ключевые слова: нечеткий кластерный анализ, ландшафтное картографирование, нечеткие k-средние, морфодинамические микрогеохоры, SRTM ЦМБ

Svidzinska D.V. Fuzzy clustering analysis for landscape mapping. Landscape discretion-continuity determines the high potential of weakly-formalized methods application in landscape mapping. The methodical principles and practical application experience of the fuzzy k-means algorithm for the reflection and analysis of morphodynamic landscape configuration is discussed on this basis.

Key words: fuzzy clustering analysis, landscape mapping, fuzzy k-means, morphodynamic microgeochores, SRTM DEM