

І. В. Гончаренко

Метод “сортуючої” кластеризації (DRSA) для класифікації рослинності

(Представлено членом-кореспондентом НАН України Я. П. Дідухом)

Розглянуто застосування нового методу класифікації фітоценозів “distance-ranked sorting assembling” (DRSA). Проаналізовано аспекти стійкості фітоценотичної класифікації. Показана можливість варіювання масштабу кластеризації та кількості фітоценотичних кластерів, класифікації даних широкого екологічного діапазону та об’єктивізації визначення екотонних угруповань. Якість геоботанічної класифікації запропоновано оцінювати кількістю статистично вірних видів.

Ключові слова: кластерний аналіз, DRSA, класифікація рослинності, фітоценологія.

Класифікація рослинності — фундаментальна задача фітоценології — нерозривно пов’язана із сучасними проблемами збереження біорізноманіття на ценотичному рівні фітосистем. Для пізнання (розвідки) структури (*data mining*) багатоознакових систем у біології, медицині, соціології тощо з метою класифікації об’єктів використовують кластерний аналіз або автоматичну класифікацію. Однак застосування класичних методів кластерного аналізу в фітоценології не набуло популярності. Геоботаніки або винаходять власні методи, наприклад методи Браун–Бланке, або використовують автоматичну класифікацію лише для первинного сортування таблиць, тобто для отримання “стартової конфігурації”, після чого “доводять” її ручним сортуванням.

Одним із “золотих стандартів” автоматичних методів класифікації у фітоценології є алгоритм Twinspan, який імплементований у відомій чеській програмі Juice [1]. Дійсно, у більшості випадків він дає непоганий результат. Але Twinspan ефективний лише за домінування одного (провідного) екологічного фактора [2]. Коли ж градієнт комплексний, а простіше кажучи — широкий, коли фактори, що визначають варіювання видового складу, змінюються в різних частинах градієнта (типах рослинності), ефективність його зменшується. Ситуація покращується, якщо градієнт поділити на відрізки (звужити), і в кожному з них провести ординацію окремо (ДСА, аналіз відповідностей з видаленим трендом). Але і алгоритм Twinspan, і інші методи ординації при багатьох видах з низьким траплянням дають результат, який потребує ручного сортування, а це знову повертає нас до проблеми “мистецтва класифікації” та “некерованого чинника”.

Існують дві основні проблеми, що перешкоджають впровадженню класичних і перевірених часом алгоритмів кластерного аналізу щодо фітоценотичних даних — їх зашумленість та неоднорідність. Шум у фітоценотичних даних зумовлений випадковими або заносними видами. При цьому деякі типові види в фітоценозі, навпаки, можуть бути відсутні, що не свідчить про непридатність для них умов існування (так звана неповночленність фітоценозів). Помилки при ідентифікації видів, підбір описів та визначення їх меж на розсуд, окомірність фітоценометричних показників (покриття, зімкнутості тощо) також зумовлюють шум.

Неоднорідність даних у фітоценотичній таблиці зумовлена тим, що геоботанічні описи відрізняються не лише за екологічними, а й іншими (фенологічними, сукцесійними тощо) характеристиками. Наприклад, описи, що здійснені в різні місяці вегетаційного сезону, виконані в одному локалітеті, але через значний час (понад 10 років); отримані з пробних площ різного розміру, матимуть відмінності у видовому складі, що лише частково відображають екологічну диференціацію та затушовані іншими факторами.

У такому випадку застосування потужних та складних методів математичної статистики не відкриє нам нові залежності чи особливості структури фітоценотичних даних, оскільки вплив на результат якості вихідних даних є істотнішим за вибір того або іншого методу. Першочерговою задачею сучасної геоботанічної науки є розробка “стійких” методів автоматичної класифікації фітоценотичних даних.

У цьому повідомленні представлені результати апробації методу “сортуючої кластеризації, заснованої на рангах відстаней” (DRSA, distance-ranked sorting assembling) [3], який розроблявся для класифікації фітоценозів з урахуванням зазначених “складних особливостей” фітоценотичних даних (багатовимірність, зашумленість, неоднорідність тощо). Але оскільки DRSA працює з матрицею відстаней, отриманою для будь-яких об’єктів у відповідному просторі ознак, то галузь використання методу не обмежується геоботанікою: він може бути застосований у соціології, медицині, психології, і, в першу чергу, там, де досліджують багатоознакові стохастичні системи.

Сутність даного методу полягає в тому, що після отримання матриці відстаней (можуть бути застосовані будь-які метрики або міри подібності) з неї виділяється “активна частина” (і лише остання враховується при групуванні об’єктів). При цьому для кожного об’єкта визначається перелік найближчих сусідів (“глибина сканування” задається через параметр mnp (minimum neighboring points)). Можна уявити, що кожен об’єкт “зв’язаний” зі своїми сусідами, це граф. Важливо, що кількість сусідів однакова у кожного об’єкта, є рівною обраному mnp і не залежить від локальної щільності об’єктів. Далі здійснюється власне групування.

DRSA можна охарактеризувати як ітеративний (неієрархічний) метод кластеризації з використанням матриці суміжності об’єктів та почерговим утворенням і нарощуванням кластерів до досягнення максимуму зв’язності кореляційних плеяд.

1. Цикл утворення нових кластерів, доки є об’єкти з $FI > 0$:

- а) ініціювати $N + 1$ кластер одним об’єктом з максимальним значенням FI ;
- б) інакше — вихід, неклаसифіковані об’єкти помітити як “шум”.

2. Цикл приєднання об’єктів до останнього кластеру, доки є об’єкти з $CI > 0$:

- а) приєднати $m + 1$ об’єкт з максимальним значенням CI , далі — п. 3;
- б) інакше — перевірити кластер на досягнення порогового розміру, далі — п. 1.

3. Перевірка “якості” кластеру (пошук оптимальної конфігурації кластеру для зупинки нарощування):

а) якщо Q -index кластеру збільшився після приєднання об’єкта, то запам’ятати нове досягнуте значення Q -index, перерахувати CI усіх об’єктів, далі — п. 2;

б) інакше — перевірити кластер на досягнення порогового розміру, далі — п. 1.

Тут N — кількість кластерів на даному етапі групування; m — кількість об’єктів у складі останнього кластеру; FI (*freeness index*) — індекс вільності об’єкта, він визначається кількістю серед найближчих сусідів вільних (некласифікованих) об’єктів, досягає максимуму у об’єктів усередині скупчень, мінімуму — поблизу утворених кластерів; CI (*connectedness index*) — індекс зв’язності об’єкта, визначається кількістю серед найближчих сусідів об’єк-

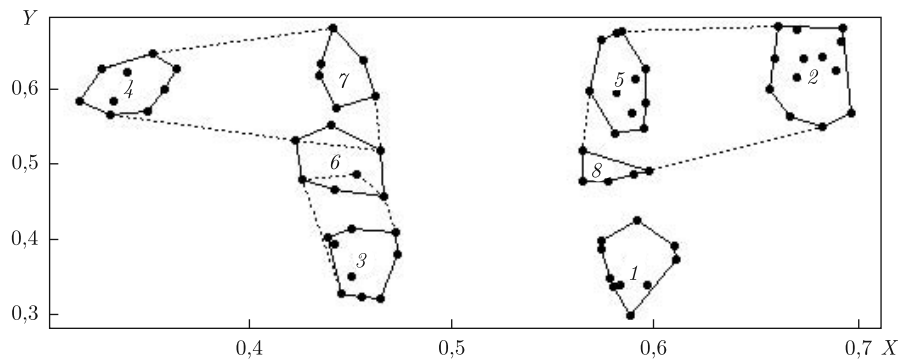


Рис. 1. Кластери DRSA штучних даних (70 точок) при $mnr = 5$ (суцільні лінії) і $mnr = 12$ (пунктирні лінії)

тів “свого” кластеру; максимальний у найближчих до кластеру об’єктів, зменшується далі від нього. Показники FI й CI перераховуються після приєднання об’єктів і описують їх положення в кластерній конфігурації, що змінюється в ході групування; вони залежать від обраного mnr і визначають “сортуючий добір” об’єктів: FI — для вибору об’єкта, що стане “центром” нового кластеру (плеяди), CI — для вибору наступного об’єкта для приєднання до “зростаючого” (останнього) кластеру. Q -index (*quality index*) — індекс “якості” кластеру, визначається кількістю зв’язків (ребер) між об’єктами кластеру та його розміром у даний момент. Це показник зв’язності кореляційної плеяди (кластеру). Максимізація значень Q -index визначає момент припинення нарощування останнього кластеру та перехід до утворення нового. З іншого боку, це дозволяє виділяти щільні кластери. Як буде показано далі, за цим показником DRSA близький до популярного в біологічних дослідженнях методу Варда.

Результат кластеризації методом DRSA штучних даних “географічної моделі” з 70 випадкових точок демонструє рис. 1. Для кластеризації було використано матрицю евклідових відстаней між точками в декартових координатах ($X; Y$).

Як видно з рисунку, ми отримали непересічні кластери. Оскільки DRSA належить до методів неієрархічної кластеризації (*flat clustering*), то звичної дендрограми тут немає. Особливістю DRSA є відсутність необхідності вказувати кількість кластерів до початку групування (як у методі К-середніх). Це важливо, адже геоботаніку апоріорі не відома кількість фітоценонів, і тому вгадати їх кількість заздалегідь було б важко. Оптимальна кількість кластерів визначається у DRSA “автоматично” та після групування. Вона кількість залежить від обраного параметра mnr . Бажаний масштаб кластеризації можна підібрати емпірично, варіюючи mnr . З рисунку також видно, що збільшення mnr дає меншу кількість кластерів, проте більшого розміру. Можна робити зрізи при різних mnr , отримуючи шари і реалізуючи відсутню в методі DRSA ієрархію.

Для апробації на фактичних даних брали масив геоботанічних описів ($N_{obj} = 722$), виконаних автором у північно-східній частині Лівобережного Лісостепу України (Сумська область). Дані мають широкий еколого-фітоценотичний діапазон (ЕФД) і охоплюють 8 класів Браун–Бланке. Розподіл кількості описів такий: *Alnetea glutinosae* (2,2%), *Festuco–Brometea* (26,7%), *Koelerio–Corynephoretea* (1,5%), *Molinio–Arrhenatheretea* (20,3%), *Phragmiti–Magnocaricetea* (8,6 %), *Quercu–Fagetea* (27,7%), *Scheuchzerio–Caricetea* (1,0%), *Vaccinio–Piceetea* (12,1%). Раніше ці описи були класифіковані традиційним методом Браун–Бланке [4]. При апробації DRSA масив описів оброблявся без поділу за типами рослинності

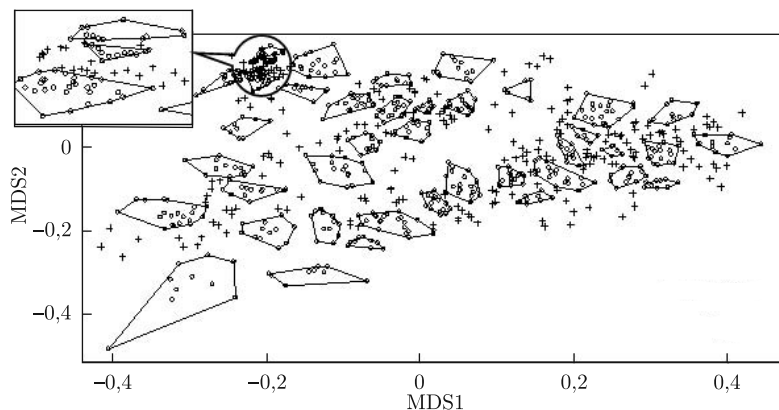


Рис. 2. Диференціація фітоценозів (кластерів DRSA) у ординаційній площині багатовимірного шкалювання

з метою дослідження ефективності методу при широкому ЕФД, коли інші методи ординації (взаємного усереднення (*reciprocal averaging*) тощо) мають обмеження через так званий arch (horseshoe) effect. Належність фітоценозів до раніше виділених асоціацій використали як критерій для оцінки відповідності автоматичної та експертної класифікації (DRSA і Браун–Бланке).

Оцінити якість кластеризації можна в ординаційній площині. Було проведено ординацію методом багатовимірного шкалювання [5]. Оскільки в обох методах використано однакову матрицю відстаней, це є перевіркою нового (DRSA) методу верифікованим (MDS). Кластеризація якісна, якщо кластери не накладаються (рис. 2).

Крім того, що кластери не перетинаються, відзначимо, що DRSA не дає 100% класифікації всіх об'єктів, частина фітоценозів (позначені "+") не увійшла до кластерів. Фактично це — перехідні описи (з екотонним видовим складом). Те що фітоценози, визначені як перехідні, знаходяться між кластерами в ординаційній площині і рідко опиняються всередині кластерів, доводить правильність діагностики екотонних угруповань. Під час класифікації фітоценозів геоботанік теж здійснює бракування перехідних описів, вирізаючи екотони, але робить це на свій розсуд. У DRSA цей процес "автоматизовано", він спирається на інформацію з матриці відстаней, тому бракування об'єктивізується. Крім того, в такий спосіб підвищується ізолюваність кластерів, що важливо з огляду на континуальність рослинності.

Для кожного з фітоценозів, показаних на рис. 2, нам відома еталонна класифікація, що отримана методом Браун–Бланке [4]. Отже, можна зіставити склад "автоматичних" і "традиційних" (класи Браун–Бланке) кластерів. Загальну міру кореляції двох класифікацій оцінимо статистикою Крамера (Cramer's V). Отримано: Cramer's V = 0,81, тобто кореляція значна, проте повної відповідності кластерів DRSA класам Браун–Бланке немає. Частина кластерів об'єднала фітоценози з двох, а іноді з трьох близьких класів: лучно-степові *Molinio-Arrhenatheretea* + *Festuco-Brometea*, лісові *Quercu-Fagetea* + *Vaccinio-Piceetea*, болотно-лучні *Phragmiti-Magnocaricetea* + *Molinio-Arrhenatheretea* тощо (хоча кластер піщаних угруповань *Koelerio-Corynephoretea* залишився цілком самостійним). Це пояснюється тим, що в екотонній природно-кліматичній зоні (Лісостепу) перехідні (щодо оптимумів перерахованих класів) фітоценози зустрічаються не менш часто, ніж типові. Тому асоціації з таких перехідних угруповань у Лісостепу цілком прийнятні: їх не можна бракувати, наприклад, щоб узгодити свою класифікаційну схему з існуючими схемами для класів з оптимумом у іншій природно-кліматичній зоні.

Одночасно було застосовано до згаданого модельного набору даних з 722 описів метод DRSA та інші алгоритми групування (метод Варда (Ward's method), повного та одиничного зв'язку, зваженого середнього (WPGMA), К-середніх), та розраховано кореляцію отриманих класифікацій різними методами. Результат DRSA виявився найбільш схожим з методом Варда (Cramer's $V = 0,86$), який, як відомо, дає щільні кластери і тому часто застосовується для біологічних об'єктів. Найменша схожість була з методом одиничного зв'язку (Cramer's $V = 0,34$). Останній використовують рідко через недолік, відомий як "*chaining phenomenon*".

Існує математичний апарат оцінки якості кластерного рішення [6]. Відомими показниками є статистика силуетів (*avg. silh*), індекс Калінського–Харабаша (*Calinski-Harabasz index, ch*) та ін. Отримані значення "внутрішніх критеріїв" валідації кластерів були близькими в методах DRSA, Варда та повного зв'язку, що свідчить на користь компактності кластерів.

У кластерному аналізі важливо те, якою мірою кластери піддаються інтерпретації. Так, для геоботаніка бажаним є результат, де на кожен кластер (фітоценоз) припадає значна кількість вірних (диференціюючих) видів, і при цьому класифікація має максимально можливий рівень детальності. При подальшому поділі кластерів межі між фітоценозами стають нечіткими, і загальна кількість диференціюючих видів зменшується. Таким чином, чим більшу кількість вірних видів має кожен виділений фітоценоз (або чим більше фітоценозів (кластерів) мають кількість вірних видів більшу за встановлений поріг), тим якіснішою є класифікація фітоценозів. Таким чином можна порівнювати якість експертних класифікацій та ефективність застосування різних автоматичних методів класифікації для одного набору даних.

Спираючись на концепцію вірності видів на статистичній основі (*fidelity index*), чому присвячено чимало робіт європейських фітоценологів [7–9], ми порівняли класифікації нашого модельного набору даних методами DRSA, Ward's, K-means, Браун–Бланке. Для кожного із зазначених методів при порозі *fidelity index* > 0,5 було отримано такі значення кількості вірних видів у середньому на фітоценоз: 4,1, 4,0, 3,8, 3,5 відповідно. Таким чином, метод DRSA показав найкращий (на одному наборі даних порівняно з іншими методами) результат щодо "змістовних" з точки зору геоботаніка фітоценозотичних кластерів. Це доводить, що делімітація кластерів методом DRSA була екологічною: фітоценози мають численні диференціюючі види, що можливо лише за умови, якщо їх межі було проведено там, де закінчуються амплітуди не одного, а групи видів, та починаються амплітуди інших видів.

Кілька слів слід сказати про оцінку відстаней між об'єктами у DRSA: враховуються не абсолютні значення, а ранги, тобто лише порядок розташування об'єктів один відносно іншого. Для класифікації зашумованих фітоценозотичних даних це важливо. І хоча це зменшує потужність запропонованого методу, проте забезпечує його робастність.

По-перше, внаслідок цього зменшується вплив випадкових видів на результат класифікації і ординації. Через їх "домішку" у складі будь-якого фітоценозу значення коефіцієнтів у матриці відстаней не точно відображають екологічну диференціацію і мають "неекологічну похибку". В методах, які спираються на абсолютні значення з матриці відстаней, вона неодмінно вплине на результат — склад і кількість кластерів. Це небажано. Навпаки, у DRSA, що є непараметричним варіантом кластерного аналізу, кластери зміняться менше, а шуми від випадкових видів буде згладжено.

По-друге, зменшується вплив на результат класифікації фітоценозів вибору іншого способу обчислення відстаней (іншого коефіцієнта подібності). В DRSA еквівалентні коефіцієнти

подібності призводять до ідентичних класифікацій. Еквівалентними у фітоценології називають коефіцієнти, які дають однакові ряди фітоценозів, впорядкованих за схожістю [10]. При цьому абсолютні значення коефіцієнтів та формули їх обчислення відрізняються, але порядок розташування фітоценозів (ранги) один відносно іншого залишається сталим. Так, заміна коефіцієнта Жаккара коефіцієнтом Сьоренсена при класифікації фітоценозів, ординації, побудові плеяд тощо не додасть нічого нового, оскільки ці коефіцієнти еквівалентні: $K_j = K_s / (2 - K_s)$. Проте використання матриць цих коефіцієнтів (та отриманих з них відстаней $D = 1 - K$) у кластерному аналізі дає близькі, але різні результати. На нашу думку, це — артефакт: хіба можуть бути різними класифікації фітоценозів одного набору даних, якщо використовуються еквівалентні коефіцієнти?!

Ми брали матриці, розраховані за коефіцієнтами Жаккара та Сьоренсена для згаданого модельного набору даних, і використали їх при класифікації: спочатку за методом DRSA, що спирається на ранжовані відстані, потім за методом Варда, що базується на абсолютних значеннях відстаней. У методі Варда отримано Cramer's $V = 0,84$, тобто кореляція класифікацій з використанням еквівалентних коефіцієнтів була високою, але результат відрізнявся! У DRSA у тій самій ситуації отримано ідентичні класифікації (Cramer's $V = 1,0$), що природніше. Далі ми брали іншу пару коефіцієнтів подібності — Браун–Бланке та Шимкевича–Сімпсона. Як відомо [10], це крайні по ряду ступеневих середніх коефіцієнти. Наскільки зміниться результат класифікації фітоценозів, якщо замінити один коефіцієнт іншим — адже вихідні дані ті самі?! DRSA виявився більш стійким: для модельного набору даних з використанням коефіцієнтів Браун–Бланке та Шимкевича–Сімпсона в DRSA Cramer's $V = 0,76$, а у методі Варда — Cramer's $V = 0,66$.

Оперування рангами (порядком об'єктів) у DRSA має ще одну перевагу для геоботаніка. У різних типах рослинності, наприклад болотній та степовій, середня схожість фітоценозів (бета-різноманітність) істотно відрізняється. Там, де умови жорсткіші, видове багатство (альфа-різноманітність) менше, бо менше видів може пристосуватися до таких умов. Але екологічні амплітуди видів зазвичай ширші, і бета-різноманітність (швидкість зміни видового складу) зменшується. Якщо поєднати болотні, степові тощо фітоценози в одній таблиці, то поділ в одній частині градієнта буде задовільним, в іншій — ні. Заміна абсолютних значень відстаней на ранги (порядок об'єктів) дозволяє проводити класифікацію фітоценозів у DRSA “з нормуванням” щодо середньої щільності синтаксонів у кожному відрізьку градієнта: кластери DRSA — це ніби проекція дендрограми, зрізана на різних рівнях у різних гілках. Тому DRSA залишається ефективним при широкому ЕФД вибірки, тоді як інші методи класифікації та ординації потребують розділення даних хоча б на рівні класів Браун–Бланке та їх окремої обробки.

Важливий аспект кластерного аналізу — стійкість класифікації. Її можна оцінити, змінюючи метод групування об'єктів: кластери, що будуть стійкими і збережуться, природніші. На жаль, геоботанікам добре відома ситуація, коли діагностичні блоки видів “розсипаються” через додавання нових описів. Це пов'язане з проблемою неповноти фітоценотичних даних.

Щоб довести, що запропонований метод доволі стійкий і в цьому аспекті, проведемо обчислювальний експеримент на основі техніки ресамплінгу. Будемо розглядати ситуацію, коли геоботанічні описи до таблиці не додаються, а віднімаються, причому вибираються вони випадково. Генерація усічених вибірок у такий спосіб схожа з підходом, відомим як “метод складного ножа” (*jackknifing*), ідею якого запропонував у 1949 р. М. Кенуй. Наша гіпотеза полягає в тому, що метод, за яким кореляція класифікацій “усічених” вибірок

фітоценозів та “повного” набору буде більшою, є стійкішим не лише до вилучення описів з таблиці, а й до їх додавання. З модельного набору даних ми видаляли по 10% випадкових фітоценозів, отримавши таким чином усічені вибірки з N_{obj} 652, 582, 512, 442 тощо відповідно. Потім провели класифікації усічених вибірок паралельно DRSA та за методом Варда. У випадку з DRSA відповідність класифікацій цих вибірок з вихідним набором ($N_{obj} = 722$) коливалася у межах Сramer’s V 0,92–0,97 (для методу Варда в аналогічній ситуації Сramer’s V був 0,77–0,84). Отже, вилучення даних меншою мірою впливало на результат у методі DRSA.

Таким чином, вказані вище особливості DRSA, як відносна стійкість, визначення оптимальної кількості фітоценотичних кластерів при обраному mpr , можливість обробляти дані широкого ЕФД, діагностика перехідних угруповань, що базується на інформації з матриці відстаней, “щільні” фітоценози зі значною кількістю диференціюючих видів тощо свідчать про перспективність методу для класифікації рослинності.

Цитована література

1. Tichý L. JUICE, software for vegetation classification // J. Veg. Sci. – 2002. – **13**. – P. 451–453.
2. Belbin L., McDonald C. Comparing three classification strategies for use in ecology // J. Veg. Sci. – 1993. – **4**. – P. 341–348.
3. Гончаренко І.В. Свідоцтво про реєстрацію авторського права на твір “Збірка наукових творів “DRSA (distance-ranked sorting assembling) – метод сортуючого кластерного аналізу” № 58837. – від 26.02.2015. – Заяв. № 59398. – І.В. Гончаренко, 31.12.2014. – Бюл. № 36.
4. Гончаренко І.В. Аналіз рослинного покриву північно-східного Лісостепу України // Укр. фітоцен. зб. Сер. А. – 2003. – **1(19)**. – 203 с.
5. Oksanen J., Blanchet F. G., Kindt R., Legendre P., O’Hara R. G., Simpson G. L., Solymos P., Stevens M. H. H., Wagner H. Vegan: Community Ecology Package. – 2010. – URL: <http://cran.r-project.org/web/packages/vegan/>.
6. Halkidi M., Batistakis Y., Vazirgiannis M. On Clustering Validation Techniques // J. Intell., Inform. Systems. – 2001. – **17**. – P. 107–145.
7. Bruelheide H. A new measure of fidelity and its application to defining species groups // J. Veg. Sci. – 2000. – **11**. – P. 167–178.
8. Chytrý M., Tichý L., Holt J., Botta-Dukát Z. Determination of diagnostic species with statistical fidelity measures // J. Veg. Sci. – 2002. – **13**. – P. 79–90.
9. De Cáceres M., Font X., Oliva F. Assessing diagnostic species value in large data sets: A comparison between phi-coefficient and Ochiai index // J. Veg. Sci. – 2008. – **19**. – P. 779–788.
10. Семкин Б.И. Эквивалентность мер близости и иерархическая классификация многомерных данных // Иерархические классификационные построения в географической экологии и систематике. – Владивосток: ДВНЦ АН СССР, 1979. – С. 97–112.

References

1. Tichý L. J. of Veg. Sci., 2002, **13**: 451–453.
2. Belbin L., McDonald C. J. of Veg. Sci., 1993, **4**: 341–348.
3. Goncharenko I. V. Certificate on the Registration of the Copyright for Scientific Work “Collection of scientific works “DRSA (distance-ranked sorting assembling) – a method of sorting clustering” No 58837, 26.02.2015 / № 59398, Goncharenko I.V., 31.12.2014, Bull. No 36 (in Ukrainian).
4. Goncharenko I. V. Analysis of vegetation of the northeastern part of Forest-Steppe of Ukraine, Ukr. Phytosoc. Coll. Ser. A, 2003, **1(19)**: 203 (in Ukrainian).
5. Oksanen J., Blanchet F. G., Kindt R., Legendre P., O’Hara R. G., Simpson G. L., Solymos P., Stevens M. H. H., Wagner H. Vegan: Community Ecology Package, 2010, URL: <http://cran.r-project.org/web/packages/vegan/>.
6. Halkidi M., Batistakis Y., Vazirgiannis M. J. Intel. Inform. Systems, 2001, **17**: 107–145.

7. Bruelheide H. J. Veg. Sci., 2000, **11**: 167–178.
8. Chytrý M., Tichý L., Holt J., Botta-Dukát Z. J. Veg. Sci., 2002, **13**: 79–90.
9. De Cáceres M., Font X., Oliva F. J. Veg. Sci., 2008, **19**: 779–788.
10. Semkin B. I. Ierarhicheskie klassifikatsionnyye postroeniya v geograficheskoy ekologii i sistematike, Vladivostok, DVNC AN SSSR 1979: P. 97–112 (in Russian).

Інститут еволюційної екології НАН України, Київ

Надійшло до редакції 15.05.2015

И. В. Гончаренко

Метод “сортирующей” кластеризации (DRSA) для классификации растительности

Інститут еволюційної екології НАН України, Київ

Рассмотрено применение нового метода классификации фитоценозов “distance-ranked sorting assembling” (DRSA). Проанализированы аспекты устойчивости фитоценотической классификации. Показана возможность варьировать масштаб кластеризации и количество фитоценоотических кластеров, классифицировать данные широкого экологического диапазона и объективизировать определение экотонных сообществ. Качество геоботанической классификации предложено оценивать количеством статистически верных видов.

Ключевые слова: кластерный анализ, DRSA, классификация растительности, фитоценология.

I. V. Goncharenko

A method of “sorting” clustering (DRSA) for the classification of plant communities

Institute for Evolutionary Ecology of the NAS of Ukraine, Kiev

The application of a new clustering technique, named “distance-ranked sorting assembling” (DRSA), for the classification of plant communities is considered. Aspects of the stability of a phytocoenotic classification are analyzed. It is shown that a researcher can vary the scale of clustering and thus change the number of phytocoenons, classify reasonably the data of a wide environmental amplitude, and determine objectively ecotonic releves among the data. We suggest to use the quantity of faithful species per phytocoenon for qualifying clusters in the field of geobotany.

Keywords: cluster analysis, DRSA, vegetation classification, phytocoenology.