

УДК 574:+502.7

О. Гетманець, канд. фіз.-мат. наук, доц.
E-mail: getmanets54@gmail.com;
А. Некос, д-р геогр. наук, проф.
E-mail: nekos@karazin.ua;
М. Пеліхатий, д-р фіз.-мат. наук, проф.
E-mail: nikolay42@gmail.com
Харківський національний університет ім. В. Каразіна,
пл. Свободи, 6, м. Харків, 61022, Україна

КЛАСТЕРНИЙ АНАЛІЗ І РАДІАЦІЙНИЙ МОНІТОРИНГ ДОВКІЛЛЯ

(Рекомендовано членом редакційної колегії д-ром геол. наук, ст. наук співроб. О.Л. Шевченком)

Побудова поля радіаційного фону на місцевості на підставі даних вимірювань, що були проведені в кінцевому числі точок, – одне з найважливіших завдань радіаційного моніторингу. Мета роботи: дослідження можливості застосування кластерного аналізу для завдань радіаційного моніторингу довкілля. Кластерний аналіз є багатовимірним статистичним аналізом. Його головне призначення – розбиття множини досліджуваних об'єктів (точок спостережень) на однорідні групи або кластери, тобто вирішується завдання класифікації даних і виявлення в них певної структури. Методи досліджень: вимірювання потужності амбієнтної дози безперервного рентгенівського і гамма-випромінювання на місцевості за допомогою дозиметра МКС-05 "ТЕРРА-0", обробка отриманих даних методами кластерного аналізу із застосуванням комп'ютерної програми "Statistics-10", при цьому кожна точка кластера характеризувалася трьома координатами: двома координатами на місцевості й потужністю амбієнтної дози випромінювання в цій точці; відстанню між двома точками була обрана евклідова відстань. Результати: після обробки даних із застосуванням різних способів кластеризації: методу ближнього сусіда, методу зваженого попарного середнього і методу Варда було встановлено, що результати аналізу якісно практично збігаються між собою, це доводить достовірність застосування кластерного аналізу для задач радіаційного моніторингу місцевості й побудови карт радіаційного забруднення. Висновки: у роботі вперше сформульовано поняття "радіаційного кластера", що об'єднує координати на площині з потужністю амбієнтної дози; доведена можливість застосування кластерного аналізу для побудови карти радіаційного забруднення довкілля шляхом послідовного проектування від більш пов'язаних до менш пов'язаних радіаційних кластерів на площину контрольованої зони. У цьому сенсі кластерний аналіз аналогічний операторному підходу до побудови поля радіації. Для подальших досліджень становить певний інтерес вивчення питань інтеграції кластерного аналізу з геоінформаційними системами.

Ключові слова: радіаційний моніторинг, рентгенівське та гамма-випромінювання, кластерний аналіз.

Вступ. Побудова поля радіаційного фону на місцевості є одним з найважливіших завдань радіаційного моніторингу. Виникає питання, як на підставі вимірювань, що були проведені в кінцевому числі розпорошених точок, відновити цілісну картину розподілу радіації на всій площі контрольованої зони. У багатьох аспектах завдання відновлення поля радіаційного фону за даними точкових вимірювань є аналогічним до завдання відновлення зображення за його точковими фрагментами (Некос та ін., 2007; 2008).

Найбільш загальний підхід до завдання відновлення зображення дається в операторному вигляді (Василенко, Тараторин, 1987). При цьому підході у випадку двох координат x і y на площині процесом формування зображення є перетворення початкового зображення $f(x, y)$ в отримане $g(x, y)$ за допомогою деякого оператора A :

$$g(x, y) = Af(x, y). \quad (1)$$

Структура оператора A може бути досить складною. Вважається, що в загальному випадку оператор A є еквівалентним добутку декількох послідовних операторів A_i ($i = 1, 2, \dots, n$), кожний з яких відповідає різним поступовим етапам отримання зображення:

$$g(x, y) = A_n \cdot A_{n-1} \dots A_1 \cdot A_0 f(x, y). \quad (2)$$

Важливим є порядок розташування операторів, оскільки наявність нелінійності не дозволяє гарантувати їхню комутативність. Завдання полягає в знаходженні зворотного перетворення A^{-1} :

$$f(x, y) = A^{-1}g(x, y). \quad (3)$$

Різноманіття моделей формування зображень можна поділити на дві великі групи: детерміновані та статистичні (Василенко, Тараторин, 1987). Детерміновані моделі пов'язані з розв'язанням інтегральних рівнянь лінійної фільтрації. У статистичних моделях вважається, що зображення є деяким статистичним об'єктом і для описання його формування застосовують методи статистичної теорії зв'язку.

Метою даної роботи є запровадження альтернативної моделі побудови поля радіаційного фону (зображення),

заснованої на методах кластерного аналізу (Duranand Odell, 1974; Мандель, 1988). Кластерний аналіз є багатовимірним статистичним аналізом. Його головне призначення – розбиття множини об'єктів, що досліджуються (точок спостережень), на однорідні групи або кластери, тобто вирішується завдання класифікації даних і виявлення в них певної структури. У кластерному аналізі схожість або різниця між об'єктами встановлюється залежно від метричної відстані між ними. Якщо кожен об'єкт має k ознак ("координат"), то він може розглядатися як точка у k -вимірному просторі, і схожість з іншими об'єктами буде визначатися як відповідна відстань між ними.

Оберемо координатами кожної з точок спостережень N_i дві координати x_i і y_i , які визначають її положення на площині, і третю координату z_i – значення амбієнтного еквівалента потужності дози неперервного рентгенівського та гамма-випромінювання в цій точці, тобто точка N_i має координати: $N_i(x_i, y_i, z_i)$. Як відстань d_{ik} між двома точками N_i і N_k оберемо найбільш поширену евклідову відстань:

$d_{ik} = \sqrt{(x_i - x_k)^2 + (y_i - y_k)^2 + (z_i - z_k)^2}$. Далі для зв'язування усіх координат до єдиного масштабу необхідно зробити нормалізацію (або стандартизацію) координат. Для цього віднімемо від кожної з координат її середнє значення за всіма точками спостережень і поділимо різницю на відповідне середнє квадратичне відхилення. Кластер, що був отриманий при об'єднанні схожих точок спостережень або схожих кластерів, називатимемо далі "радіаційним кластером".

Слід зазначити, що існує певна кількість робіт, присвячених застосуванню кластерного аналізу для завдань екологічного моніторингу довкілля (Поліщук, 2005; Saksena, 2003; Atushi, 2014). Але їхньою метою головним чином є визначення, виділення і локалізація основних джерел забруднень, а також оптимізація кількості пунктів екологічних спостережень.

Об'єкти та методи досліджень. Матеріалами досліджень були використані результати вимірювань рівня амбієнтного еквівалента потужності дози неперервного

рентгенівського та гамма-випромінювання на частині території Немишлянського району м. Харкова, що були проведені у травні–червні 2017 р. Вимірювання проводилися за допомогою дозиметра МКС-05 "ТЕРРА-0" з абсолютною похибкою приладу 0,01 мкЗв/год, який застосовується для дозиметричного і радіометричного контролю на промислових підприємствах; для екологічних досліджень; для контролю радіаційної чистоти житлових приміщень, будівель і споруд і прилеглих до них територій. Карта території, на якій проводилися дослідження, наведена на рис. 1. На цій карті показано розташування усіх 19 контрольних точок спостережень. Відлік координат здійснювався відносно контрольної точки

№ 1: координати 1(0;0). Обробка результатів вимірювань проводилася методами кластерного аналізу за допомогою програми "Statistics-10" (SPSSInc., US).

Результати досліджень. Як приклад у табл. 1 наведено результати вимірювань амбієнтного еквівалента потужності дози неперервного рентгенівського та гамма-випромінювання в усіх контрольних точках, наведених на рис. 1, станом на 31 травня 2017 р. У цій таблиці наведено як абсолютні значення результатів вимірювань і геометричні координати кожної контрольної точки, так і стандартизовані (зведені до єдиного масштабу) значення усіх трьох "координат".

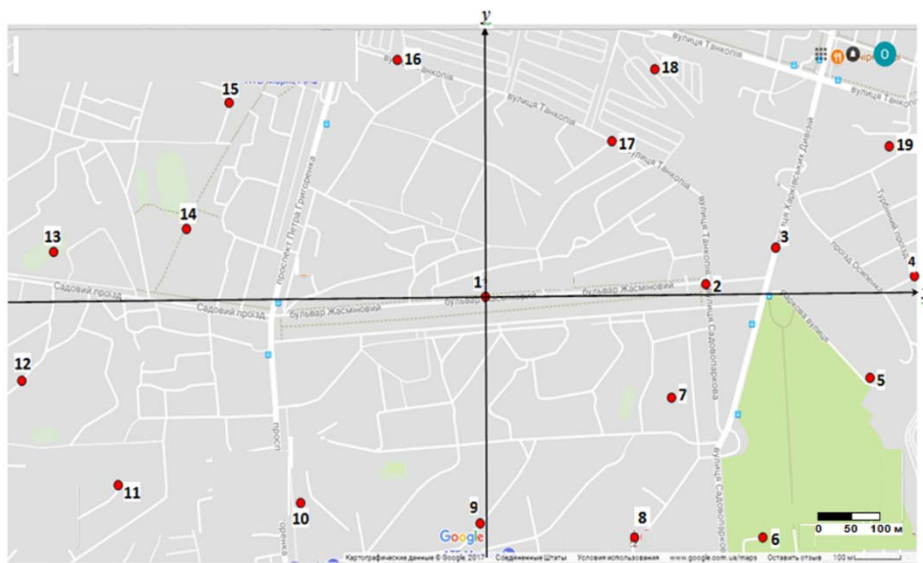


Рис. 1. Карта-схема розташування точок спостережень на контрольованій території

Результати вимірювань амбієнтного еквівалента потужності дози неперервного рентгенівського та гамма-випромінювання станом на 31 травня 2017 року

Таблиця 1

№ точки	Потужність дози d , мкЗв/год	Координати		Стандартизовані координати		
		x , м	y , м	$\tilde{d}$	$\tilde{y}$	$\tilde{x}$
1	0,19	0	0	0,426	-0,097	0,064
2	0,16	493	21	-0,630	0,691	0,131
3	0,13	629	93	-1,687	0,908	0,363
4	0,14	921	36	-1,334	1,374	0,180
5	0,18	829	-164	0,074	1,227	-0,464
6	0,15	593	-471	-0,982	0,850	-1,451
7	0,15	400	-193	-0,982	0,542	-0,557
8	0,18	314	-464	0,074	0,404	-1,429
9	0,19	-14	-436	0,426	-0,120	-1,339
10	0,19	-400	-393	0,426	-0,736	-1,201
11	0,22	-786	-350	1,483	-1,353	-1,062
12	0,16	-1000	-150	-0,630	-1,695	-0,419
13	0,20	-929	100	0,778	-1,582	0,385
14	0,15	-643	136	-0,982	-1,125	0,501
15	0,21	-550	379	1,131	-0,976	1,283
16	0,23	-186	464	1,835	-0,394	1,557
17	0,21	264	300	1,131	0,325	1,029
18	0,18	357	436	0,074	0,473	1,467
19	0,16	864	279	-0,630	1,283	0,961

Далі була безпосередньо застосована кластеризація отриманих даних. Слід зазначити, що існує багато схем кластеризації. Ми обрали для порівняння три найбільш поширені з них: 1) "метод найдальшого сусіда" (Completelinkage) – в алгоритмі якого відстань між двома кластерами визначається як максимальна відстань між усіма можливими парами елементів із цих кластерів; при реалізації цього методу, як правило, формуються відносно компактні кластери, які складаються з об'єктів з великою схожістю; 2) "метод зваженого попарного

середнього" (Weightedpair-groupaverages) – у цьому методі відстань між двома різними кластерами обчислюється як середня відстань між усіма парами об'єктів з них, при цьому кількість об'єктів в кластерах використовується як ваговий коефіцієнт; даний метод застосовують, коли передбачаються нерівні розміри кластерів; 3) "метод Варда" (Ward'smethod) – як цільову функцію використовують внутрішньокластерну суму квадратів відхилень відстаней між кожною точкою кластера і середньою за кластером; на кожному кроці об'єднуються

такі два кластери, які призводять до мінімального збільшення цільової функції, тобто внутрішньої групової суми квадратів; цей метод направлений на об'єднання близько розташованих кластерів.

Для наочності порівняння результати застосування кожної схеми кластеризації щодо даних табл. 1 на рис. 2–4 наведено відповідні "дендрограми". Тут на вертикальній осі вказано номери відповідних точок спостережень, а на горизонтальній – "Linkage Distance" – "відстань зв'язку". Чим вона менша, тим сильніше зв'язані кластери. Вважається, якщо "відстань зв'язку" перевищує 3–4 відносні одиниці, то процес кластеризації завершується (Duran, Odell, 1974; Мандель, 1988).

Порівнюючи між собою ці "дендрограми", можна зробити висновок, що вони якісно збігаються, тобто вказують на утворення кластерів з однакових елементів (точок).

Щодо 3d геометрії радіаційний кластер являє собою "кулю", що висить над контрольованою площиною (x, y). Радіус кулі (радіус кластера) – це відстань від найдаальшої

точки кластера до його центра, який є середнім геометричним місцем точок кластера.

Для побудови карти радіаційного фону на місцевості слід спроектувати ці кулі на площину (x, y), узявши за значення потужності дози випромінювання – середнє геометричне за кластером.

Порядок проектування такий: на першому (передньому) плані розташована проекція кластера з найменшою відстанню зв'язку, на другому – проекція з наступною за значенням відстанню і так далі, доки на задньому плані не з'явиться середнє за всіма точками спостережень значення потужності дози. При цьому можна скористатися таблицями відстаней зв'язку, які програма "Statistics-10" видає для кожної схеми кластеризації. Не вдаючись у деталі обчислень, наведемо на рис. 5–7 результати проектування радіаційних кластерів на площину контрольованої території (рис. 1) для різних схем кластеризації. На цих картах для прив'язки до карти місцевості (рис. 1) як реперні наведено всі контрольні точки вимірювань.

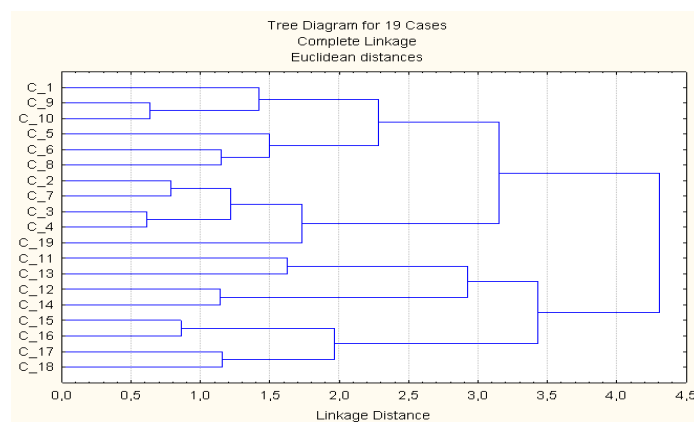


Рис. 2. Схема кластеризації за методом найдаальшого сусіда

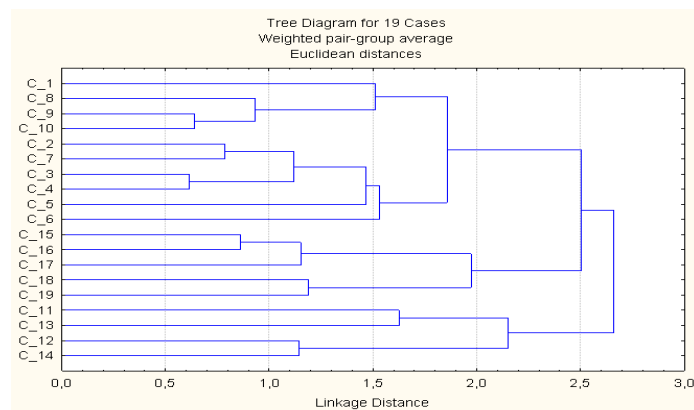


Рис. 3. Схема кластеризації за методом зваженого попарного середнього

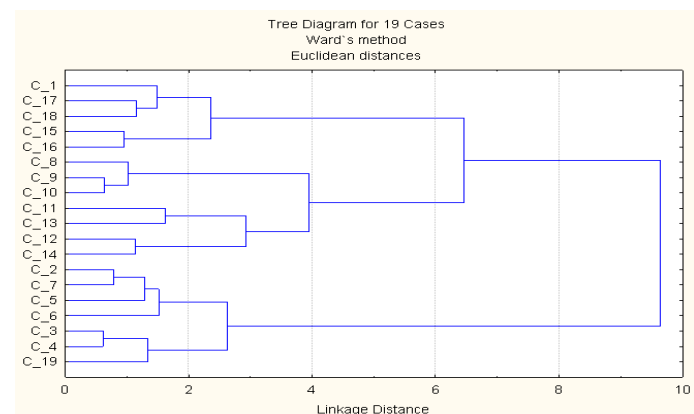


Рис. 4. Схема кластеризації за методом Варда

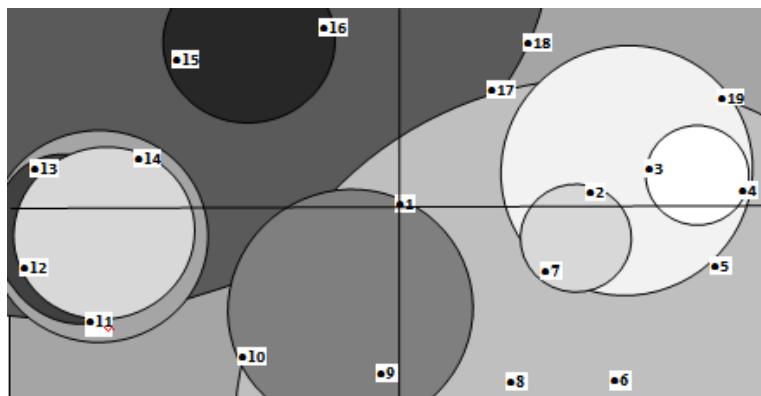


Рис. 5. Карта радіаційного фону станом на 31.05.17, побудована за методом найдаляшого сусіда: білий колір – 0,14 мкЗв/год, чорний – 0,22 мкЗв/год через 0,01 мкЗв/год за тоном

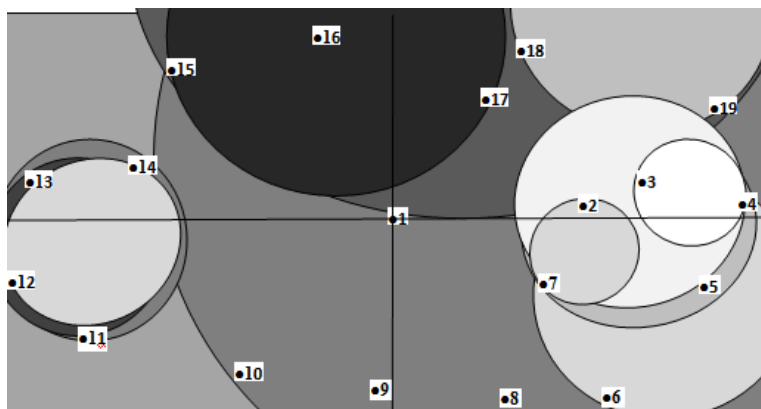


Рис. 6. Карта радіаційного фону станом на 31.05.17, побудована за методом зваженого попарного середнього. Умовні позначення див. рис. 5

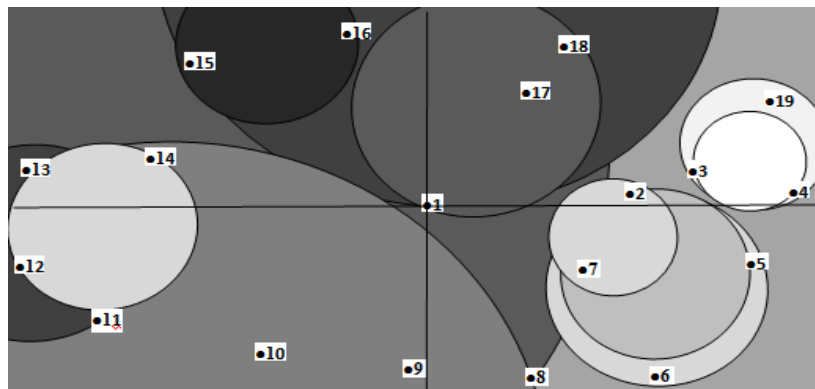


Рис. 7. Карта радіаційного фону станом на 31.05.17, побудована за методом Варда. Умовні позначення див. рис. 5

Порівнюючи між собою рис. 5–7, можна бачити, що передбачення усіх трьох методів кластеризації щодо побудови поля радіаційного фону на відносно невеликій території якісно і кількісно практично збігаються (особливо за методом дальнього сусіда і методом Варда). Це підтверджує достовірність застосування кластерного аналізу для завдань радіаційного моніторингу місцевості й побудови карт радіаційного забруднення.

Висновки. Таким чином, у даній роботі доведена можливість застосування кластерного аналізу для побудови поля потужності дози випромінювання, тобто поля радіаційного забруднення території. Уперше сформульовано поняття "радіаційного кластера". За своїм змістом послідовне проектування радіаційних кластерів на площину контрольованої території в певному сенсі аналогічно до процедури операторного відновлення зображення (формула 2). Надалі становить певний інтерес дослідження питань щодо інтеграції кластерного аналізу з геоінформаційними системами.

Список використаних джерел

- Василенко, Г., Тараторин, А. (1986). Восстановление изображений. М.: Радио и связь.
- Мандель, И. (1988). Кластерный анализ. М.: Финансы и статистика.
- Некос, В.Ю., Юшманова, І.П., Пеліхатий, М.М. (2007). Розробка систем радіаційного моніторингу довкілля. *Журнал наукових праць: Людина і довкілля. Проблеми неоекології*, 9, 52–62.
- Некос, В., Пеліхатий, М., Юшманова, І. (2008). Методи і алгоритми визначення радіаційного стану довкілля. *Журнал наукових праць: Людина і довкілля. Проблеми неоекології*, 1–2 (11–12), 90–98.
- Полищук, Ю.М. и др. (2005). Геоинформационная методология анализа радиационного загрязнения речных систем. *Проблемы анализа риска*, т. 2, 3, 208–220.
- Atsushi, I., Shintaro, S., Atshushi, M. et al. (2014). A cluster analysis of constant ambient air monitoring data from the Kato region of Japan. *Int. J. Res. Public Health*, 11 (7), 6844–6855.
- Duran, B., Odell, P. (1974). Cluster analysis. Springer-Verlag, Berlin.
- Pinero-Garcia F., Ferro-García, M., Chham, E. (2015). A cluster analysis of back trajectories to study the behavior of radioactive aerosols in the south-east of Spain. *J. Environ Radioact.*, 142–152.
- Saksena S. Joshi, V., Patil, R. (2003). Cluster analysis of Delhi's ambient air quality data. *J. Environ. Monit.*, 5, 491–499.

References

- Atsushi, I., Shintaro, S., Atshushi, M. et al. (2014). A cluster analysis of constant ambient air monitoring data from the Kato region of Japan. *Int. J. Res. Public Health*, 11 (7), 6844–6855.
- Duran B., Odell, P. (1974). Cluster analysis. Berlin: Springer-Verlag.
- Mandel, I. (1988). Klasternyi analiz. Moscow: Finansy i statistika. [In Russian]
- Nekos V., Pelikhatyi, M., Yushmanova, I. (2008). Metody i alhorytmy vyznachennia radiatsionnoho stanu dovkillia. *Zhurnal naukovykh prats: Liudyna i dovkillia. Problemy neoeekologii*, 1-2 (11-12), 90 – 98. [In Ukrainian]
- Nekos, V., Yushmanova, I., Pelikhatyi, M. (2007). Rozrobka system radiatsionnoho monitorynhu dovkillia. *Zhurnal naukovykh prats: Liudyna i dovkillia. Problemy neoeekologii*, 9, 52 – 62. [In Ukrainian]
- Pinero-Garcia, F., Ferro-Garcia, M., Chham, E. (2015). A cluster analysis of back trajectories to study the behavior of radioactive aerosols in the south-east of Spain. *J. Environ Radioact.*, 142–152.
- Polyschuk, Yu., Peremytna, T., Kocherhyn, H. (2014). Heoynformatsyonnaia metodolohiia analiza radyatsyonnoho zahriaznennia rechnykh system. *Problemy analiza ryska*, V. 2, 3, 208–220. [In Russian]
- Saksena, S., Joshi, V., Patil, R. (2003). Cluster analysis of Delhi's ambient air quality data. *J. Environ. Monit.*, 5, 491–499.
- Vasilenko G., Taratorin, A. (1986). Vosstanovleniye izobrazheny. Moscow: Radio i svayz. [In Russian]

Надійшла до редколегії 17.02.19

O. Getmanets, Cand. Sci. (Phys.-Math.), Assoc. Prof.
E-mail: getmanets54@gmail.com;
A. Nekos, Dr. Sci. (Geogr.), Prof.
E-mail: nekos@karazin.ua;
M. Pelikhatyi, Dr. Sci. (Phys.-Math.), Prof.
E-mail: nikolay42@gmail.com
V. Karazin Kharkiv National University

CLUSTER ANALYSIS AND RADIATION MONITORING OF ENVIRONMENT

Building a background radiation field on the ground on the basis of measurement data taken at a finite number of points is one of the most important tasks of radiation monitoring. The aim of the work: to study the possibility of applying cluster analysis for the tasks of radiation monitoring of the environment. Cluster analysis is a multidimensional statistical analysis. Its main purpose is to split the set of objects under study (observation points) into homogeneous groups or clusters, that is, the task of classifying data and identifying the corresponding structure in them is solved. Methods of research: the measurements of the power of the ambient dose of continuous X-ray and gamma radiation on the terrain by using the MKS-05 dosimeter "TERRA-0"; processing of the obtained data by cluster analysis methods using the computer program "Statistics-10", wherein each cluster point is characterized by three coordinates: two coordinates on the ground and the power of the ambient dose of radiation at a given point; Euclidean distance was chosen as the distance between two points. Results: after processing data using various clustering methods: the method of Complete Linkage, the method of Weighted pair-group average and the Ward's method, it was found that the results of the analysis practically coincide with each other, that proves the reliability of the application of cluster analysis for the tasks of radiation monitoring of the environment and mapping of radiation pollution. Conclusions: the concept of a "radiation cluster" was first formulated in this work, combining coordinates on a plane with an ambient dose rate; the possibility of using cluster analysis to construct a map of radiation pollution of the environment has been proved by sequential projection from more connected to less connected radiation clusters onto the plane of the controlled zone. In this sense, cluster analysis is similar to the operator approach to the construction of the radiation field. For further research, it is of some interest to study the issues of integration of cluster analysis with geographic information systems.

Keywords: radiation monitoring, X-ray and gamma radiation, cluster analysis.

O. Гетманец, канд. физ.-мат. наук, доц.
E-mail: getmanets54@gmail.com;
A. Некос, д-р геогр. наук, проф.
E-mail: nekos@karazin.ua;
Н. Пелихатий, д-р физ.-мат. наук, проф.
E-mail: nikolay42@gmail.com;
Харьковский национальный университет им. В. Каразина,
Пл. Свободы, 6, г. Харьков, 61022, Украина

КЛАСТЕРНЫЙ АНАЛИЗ И РАДИАЦИОННЫЙ МОНИТОРИНГ ОКРУЖАЮЩЕЙ СРЕДЫ

Построение поля радиационного фона на местности на основании данных измерений, проведенных в конечном числе точек – одна из важнейших задач радиационного мониторинга. Цель работы: исследование возможности применения кластерного анализа для задач радиационного мониторинга окружающей среды. Кластерный анализ является многомерным статистическим анализом. Его главное назначение – разбиение множества исследуемых объектов (точек наблюдений) на однородные группы или кластеры, то есть решается задача классификации данных и выявления в них соответствующей структуры. Методы исследований: измерение мощности ambientной дозы непрерывного рентгеновского и гамма-излучения на местности с помощью дозиметра MKS-05 "TERRA-0", обработка полученных данных методами кластерного анализа с применением компьютерной программы "Statistics-10", при этом каждая точка кластера характеризовалась тремя координатами: двумя координатами на местности и мощностью ambientной дозы излучения в данной точке; в качестве расстояния между двумя точками было выбрано евклидово расстояние. Результаты: после обработки данных с применением различных способов кластеризации: метода ближнего соседа, метода взвешенного попарного среднего и метода Варда было установлено, что результаты анализа качественно практически совпадают между собой, что доказывает достоверность применения кластерного анализа для задач радиационного мониторинга местности и построения карт радиационного загрязнения. Выводы: в работе впервые сформулировано понятие "радиационного кластера", объединяющего координаты на плоскости с мощностью ambientной дозы; доказана возможность применения кластерного анализа для построения карты радиационных загрязнений окружающей среды путем последовательного проектирования от более связанных до менее связанных радиационных кластеров на плоскость контролируемой зоны. В этом смысле кластерный анализ аналогичен операторному подходу к построению поля радиации. Для дальнейших исследований представляет определенный интерес изучение вопросов интеграции кластерного анализа с геоинформационными системами.

Ключевые слова: радиационный мониторинг, рентгеновское и гамма-излучение, кластерный анализ.