

УДК 528.94

## ВИКОРИСТАННЯ КЛАСТЕРНОГО АНАЛІЗУ ДЛЯ ПРОСТОРОВОГО МОДЕЛЮВАННЯ В ГЕОІНФОРМАЦІЙНИХ СИСТЕМАХ

В. Путренко<sup>1</sup>, В. Тихоход<sup>2</sup>

<sup>1</sup>*Інститут географії НАН України,*

<sup>2</sup>*Національний технічний університет України*

*“Київський політехнічний інститут”*

*вул. Володимирська, 44, м. Київ, 01034, Україна*

Розглянуто питання математичного аналізу географічної інформації за допомогою процедури кластерного аналізу. Розроблено програмний модуль для проведення різних видів кластерного аналізу географічних об'єктів та автоматизованої побудови тематичних карт у програмному середовищі ArcGIS.

*Ключові слова:* математичні методи, кластерний аналіз, ГІС, поводження з відходами.

Сучасні підходи до управління територіальним розвитком ґрунтуються на складних інформаційних системах моделювання розвитку природних, техногенних, суспільних явищ та процесів, де визначну роль відіграють геоінформаційні системи як засіб накопичення, обробки та аналізу просторової інформації. Однією з інтегральних тем взаємодії людства і природи є прогнозування та запобігання надзвичайним ситуаціям, які загрожують збереженню природного середовища, біологічному різноманіттю, безпеці життя людей. В Україні існує декілька актуальних екологічних проблем, які потенційно загрожують виникненню нових надзвичайних ситуацій різного територіального рівня, однією з яких є проблема поводження з відходами. Загальний обсяг накопичення відходів в Україні оцінюють у 35 млрд т. Обсяг їхнього щорічного утворення сягає 780–800 млн т. Переважну їхню частину становлять промислові відходи (розкривні супутні породи, шлами) продукти збагачення корисних копалин, металургійні шлаки тощо. Значну небезпеку та проблему становлять місця накопичення відходів, які у більшості випадків не мають відповідних умов для їхнього збереження. З огляду на це завдання прийняття науково обґрунтованих рішень у процесі управління відходами на різних територіальних та функціональних рівнях є актуальним і сьогодні.

Кластеризацію територіальних одиниць за комплексними групами показників вже багато років використовують для типізації території та моделювання системутворюючих зв'язків в економіці, суспільній географії та інших науках. Кластерний аналіз регіонів України за особливостями утворення та

накопичення відходів дає змогу прогнозувати найбільші екологічні небезпеки, пов'язані з цим.

Моделі кластерного аналізу під час розв'язання географічних завдань у геоінформатиці досліджували В. С. Тикунов, О. А. Євтеєв, С. Н. Сербенюк, О. О. Ішук, О. О. Світличний та інші дослідники геоінформаційного напрямку [1, 3–8].

У статті розроблено автоматизовані алгоритми багатовимірної кластеризації векторних геоінформаційних моделей даних на прикладі діагностування та управління небезпеками пов'язаними з накопиченням відходів у регіонах України.

Завданнями дослідження є опрацювання методичних основ проведення багатовимірних класифікацій, вибір форми реалізації програмних алгоритмів у геоінформаційному середовищі, апробація кластерних моделей на прикладі даних про утворення відходів у регіонах України, виділення основних ознак досліджених кластерних груп та небезпек, пов'язаних з накопиченням відходів у регіонах.

Процедура кластерного аналізу належить до базових методів математичної обробки даних, коли кожен об'єкт потрапляє до певної групи на підставі аналізу його ознак у багатовимірному, евклідовому просторі. Кластерний аналіз належить до значної кількості статистичних та моделюючих програмних систем, таких як STATISTICA, MatLab тощо. У більшості сучасних геоінформаційних пакетів процедури певної кластеризації (класифікації) об'єктів також наявні. Проте кожен з програмних продуктів має свої обмеження.

Для геоінформаційних продуктів притаманний одно- або двовимірний аналіз векторних або растрових моделей даних, коли дані групують на підставі одного чи двох параметрів та відображають за допомогою картограми. З іншого боку, статистичні пакети мають досконалі модулі кластерного аналізу, але при цьому нема змоги автоматизованого картографування даних та використання географічної інформації для цілей поділу об'єктів на групи. Сьогодні цю проблему вирішують завдяки інтеграції статистичних пакетів та геоінформаційних продуктів через протоколи взаємодії даних (експорт / імпорт даних, підтримка форматів), але така процедура потребує наявності декількох програмних продуктів на робочому місці та досвідченого користувача. Тому існує потреба у створенні додаткових модулів кластерного аналізу, які працюють безпосередньо у геоінформаційному середовищі та використовують його переваги.

Такий програмний модуль було розроблено у вигляді інструмента кластеризації на підставі розширень геоінформаційної системи ArcGis 9.3.

Алгоритм використання програмного модуля зображено на рис. 1. Однією з головних його переваг є можливість проведення операцій кластеризації у багатовимірному просторі, залучення географічної інформації як додаткових координатних вимірів об'єкта, побудова автоматизованих картографічних моделей та легенд за результатами аналізу. Використання географічних координат дає змогу врахувати взаємне розташування об'єктів у географічному просторі,

топологію їхніх зв'язків та, відповідно, виконувати районування території на підставі математично обґрунтованих даних.

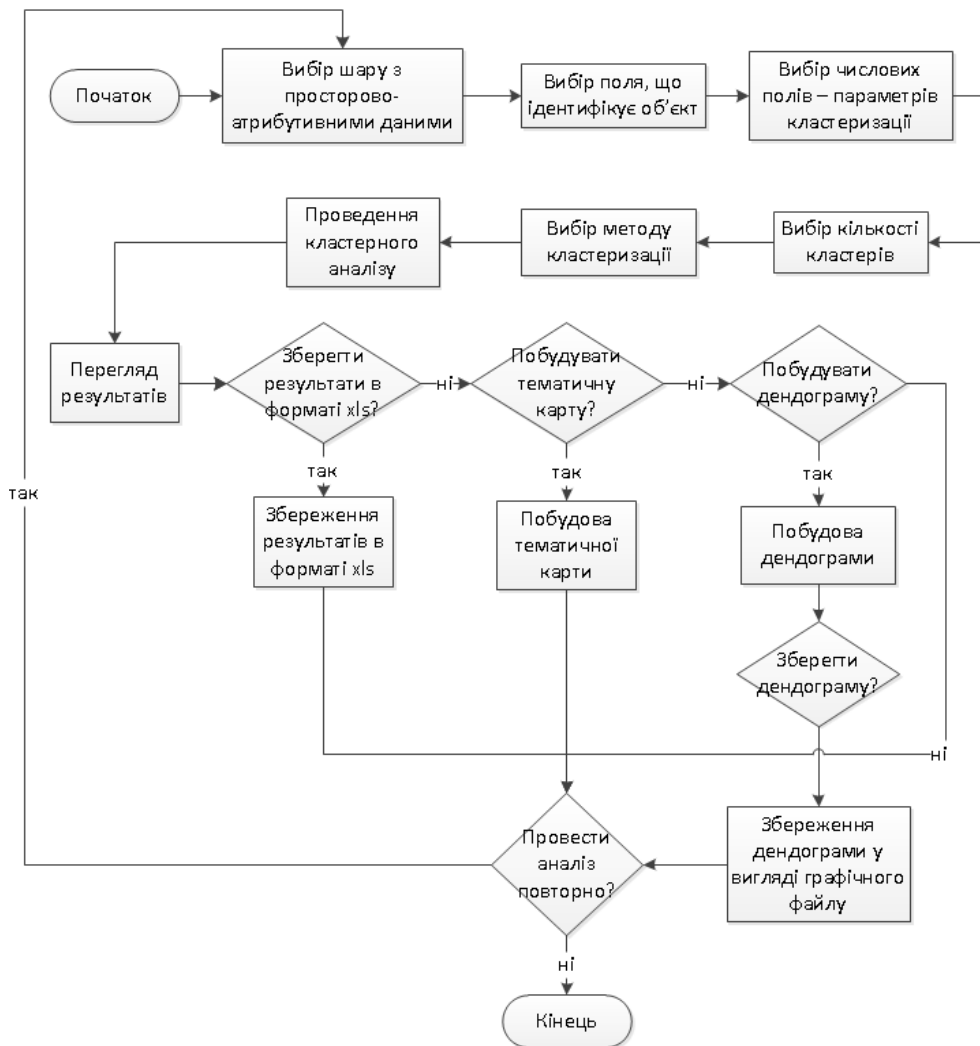


Рис. 1. Алгоритм використання модуля кластерного аналізу

Програмний модуль розроблено за допомогою мови C#, набору засобів розроблення ArcGis SDK для .NET архітектури. Модуль містить набір методів агрегації даних за методом ближнього сусіда, Уорда, середньої відстані, центральної відстані, дальнього сусіда, а також можливості задавати кількість вихідних кластерів. Результатом роботи модуля є тематична карта та легенда до неї з виділеними кластерами, таблиця з розбивкою за класами географічних об'єктів, дендрограма кластерного аналізу (рис. 2).

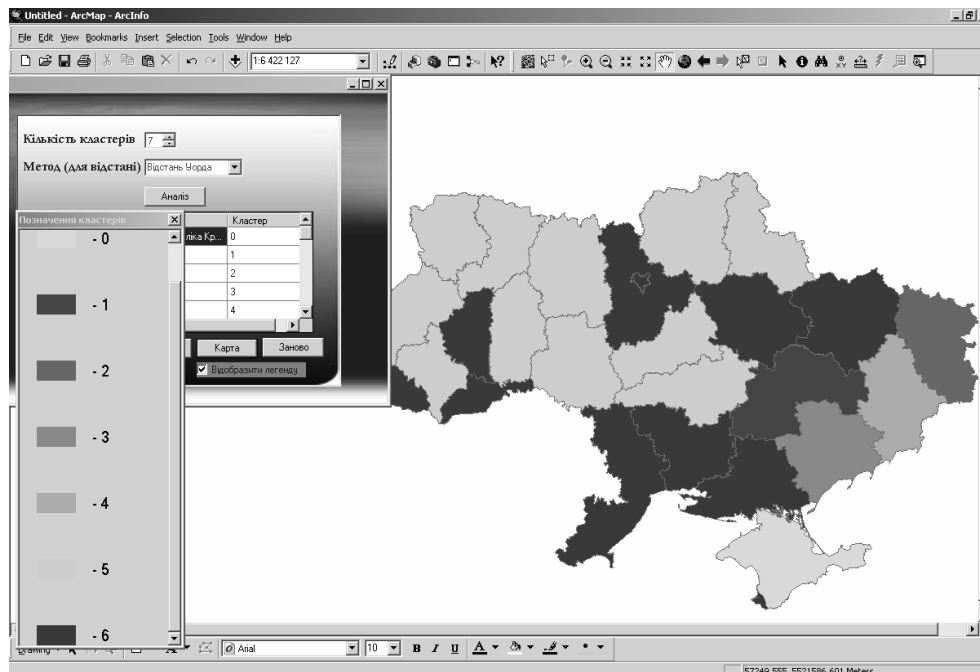


Рис. 2. Вікно ArcGis з результатами кластерного аналізу

Для проведення кластеризації обрано адміністративно-територіальний поділ України на 24 області, Автономну Республіку Крим і міста республіканського підпорядкування Київ та Севастополь. Як вхідні дані для кластеризації використано статистичні дані за 2010 р. про утворення та накопичення на території регіонів відходів 1–4 класів небезпеки, що дало змогу проведення аналізу у восьмивимірному просторі. Головним завданням кластеризації за цими показниками є виявлення основних типів регіонів за особливостями утворення відходів та небезпек, пов'язаних з ними.

Головними характерними особливостями розподілу відходів в Україні є їхнє накопичення в регіонах видобутку корисних копалин та промислового виробництва у східних, центральних та деяких західних областях України. Пов'язані з цим небезпеки залежать від обсягів утворення та накопичення відходів. Відходи першої групи небезпеки накопичені здебільшого в Донбасі, а також Кіровоградській, Сумській та Харківській областях, а за їхнім утворенням перші місця посідають Луганська, Харківська, Полтавська та Донецька області. Відходи другої групи накопичені у АР Крим, Донецькій та Луганській областях, а за їхнім утворенням очолюють список АР Крим, Дніпропетровська, Сумська та Донецька області. Найбільша кількість відходів третьої групи накопичена в Запорізькій та Донецькій області, а за їхнім щорічним утворенням перші місця посідає Полтавська, Миколаївська та Донецька області. Абсолютні обсяги накопичення відходів четвертого класу небезпеки

зафіксовані у Дніпропетровській, Донецькій та Луганській областях, які разом з Кіровоградщиною також мають найбільші обсяги щорічного утворення відходів цієї групи [2].

Після порівняльного аналізу методів кластеризації було вирішено обрати метод Уорда, який ґрунтується на принципі мінімізації дисперсії двох кластерів, що об'єднуються на кожному кроці [8]. У дослідженні проаналізовано випадок кластеризації у восьми- та десятивимірному просторі. У другому випадку використано як додаткові показники географічні координати центрів регіонів, які дають змогу врахувати близькість розміщення адміністративних одиниць та провести їхнє зонування.

У восьмивимірному просторі виділяють п'ять кластерів регіонів. Перший кластер – АР Крим, другий – Дніпропетровська та Луганська області, третій – Запорізька та Донецька області, четвертий – Сумська, Кіровоградська, Івано-Франківська, Чернігівська, Рівненська, Черкаська, Житомирська, Львівська, Хмельницька, Волинська, Вінницька області, п'ятий – Харківська, Херсонська, Київська, Одеська, Тернопільська, Закарпатська, Чернівецька, Полтавська, Миколаївська області, міста Київ і Севастополь.

У десятивимірному просторі виділяють сім кластерів, які демонструють географічне та тематичне поєднання (табл. 1). За цими даними розроблено карту території України, на якій виділено сім груп областей (рис. 3). Перший кластер охоплює Донецьку та Луганську області з високим рівнем накопичення відходів, зокрема, відходів першого та четвертого класів, що свідчить про високий ступінь небезпеки, пов'язаної з високотоксичними речовинами у місцях їхнього складування та небезпечними процесами, що пов'язані з відходами четвертого класу. До другого кластеру належить Дніпропетровська та Запорізька області з високим ступенем небезпеки хімічних відходів першого класу та відходів металургії. Третій кластер містить АР Крим, Львівську та Івано-Франківську області, які вирізняються наявністю відходів другого–третього класів небезпеки, що пов'язано з видобутком та промисловим виробництвом хімічних сполук. Четвертий кластер охоплює Київ, Київську, Черкаську та Кіровоградську області, де переважають токсичні відходи гірничого видобутку та відходи першого–другого класів небезпеки при хімічному виробництві. П'ятий кластер – Харківська, Полтавська, Миколаївська, Херсонська області та м. Севастополь, де переважають відходи третього–четвертого ступенів небезпеки. Шостий кластер охоплює Чернігівську, Тернопільську, Хмельницьку, Сумську, Одеську, Вінницьку області з порівняно низьким рівнем небезпеки відходів, які зазвичай належать до третьої та четвертої груп небезпеки. Останній, сьомий кластер, – це Волинська, Рівненська, Житомирська, Закарпатська, Чернівецька області з низькими обсягами утворення та складування відходів. Отримані кластери у своїй більшості мають компактну географічну структуру, яка дає змогу проводити не лише типізацію території, а й районування за певним набором ознак.

Застосування методів кластерного аналізу у поєднанні з аналітичними можливостями геоінформаційних систем значно розширює методичний інстру-

Таблиця 1

Кластерний аналіз (метод – відстань Уорда)

Кластер	1	2	3	4	5	6	7
Об'єкти	Волинська обл.	Київська обл.	Донецька обл.	Чернігівська обл.	Полтавська обл.	Дніпропетровська обл..	АР Крим
	Рівненська обл.	м. Київ	Луганська обл.	Тернопільська обл.	Харківська обл.	Запорізька обл.	Львівська обл.
	Житомирська обл.	Черкаська обл.		Хмельницька обл.	Миколаївська обл.		Івано-Франківська обл.
	Закарпатська обл.	Кіровоградська обл.		Сумська обл.	Херсонська обл.		
	Чернівецька обл.			Одеська обл.	м. Севастополь		
				Вінницька обл.			



Рис. 3. Результати кластеризації за просторово-атрибутивними ознаками поводження з відходами

ментарій сучасного географічного дослідження. Створені засоби проведення кластерного аналізу в середовищі ArcGis дають змогу класифікувати просторові об'єкти за просторово-атрибутивними параметрами. До основних переваг багатовимірного аналізу в середовищі ГІС належить автоматизоване картографування кластерів, використання додаткової географічної інформації, варіабельність у прийнятті рішень щодо аналізу. Унаслідок апробації програмного модуля кластеризації під час дослідження даних про поведження з відходами дійшли таких висновків:

порівняно з кластеризацією за атрибутивними параметрами застосування методу з просторово-атрибутивними параметрами дало змогу виділити більш рівномірні за кількістю адміністративні одиниці та територіально зв'язані кластери;

серед семи виділених кластерів перші чотири належать до регіонів з порівняно високим рівнем небезпеки виникнення НС, три інші кластери мають помірну небезпеку, пов'язану з поведженням з відходами;

Подальші дослідження пов'язані з формуванням баз даних потенційних регіональних небезпек та автоматизованих механізмів їхнього оброблення в геоінформаційному середовищі.

- 
1. Геоінформатика: учебник для студ. высш. учебн. заведений / Е. Г. Капралов, А. В. Кошкарев, В. С. Тикунов и др.; под ред. В. С. Тикунова. В 2 кн. Кн 1. – [2-е изд. перераб. и доп.]. – М.: Издательский центр Академия, 2008. – 384 с.
  2. Довкілля України–2010. – Державна служба статистики України, 2011. – 205 с.
  3. *Евтеев О. А.* Проектирование и составление социально-экономических карт: учебник / О. А. Евтеев. – М.: Изд-во Моск. ун-та, 1999. – 224 с.
  4. *Ищук О. О.* Просторовий аналіз і моделювання в ГІС: навч. посібн. / О. О. Іщук, М. М. Коржнев, О. Е. Кошляков; за ред. акад. Д. М. Гродзинського. – К.: Видавничо-поліграфічний центр “Київський університет”, 2003. – 200 с.
  5. *Світличний О. О.* Основи геоінформатики: навч. посібн. / О. О. Світличний, С. В. Злотницький; за заг. ред. О. О. Світличного. – Суми: ВТД Університетська книга, 2006. – 295 с.
  6. *Сербенюк С. Н.* Картография и геоинформатика – их взаимодействие / С. Н. Сербенюк. – М., 1990. – 159 с.
  7. *Тикунов В. С.* Классификации в географии: ренессанс или увядание? / В. С. Тикунов. – М.; Смоленск: Изд-во СГУ, 1997. – 367 с.
  8. *Murray Alan T.* Cluster analysis discovery techniques for exploratory spatial data analysis Alan T. Murray, Vladimir Estivill-Castro // International Journal of Geographical Information Science. – 1998. – Vol. 12, Iss. 5. – P. 431–443.

*Стаття: надійшла до редколегії 16.05.2013  
доопрацьована 12.07.2013  
прийнята до друку 25.09.2013*

## CLUSTER ANALYZES FOR SPATIAL MODELLING IN GEOGRAPHIC INFORMATION SYSTEMS

V. Putrenko<sup>1</sup>, V. Tyhohod<sup>2</sup>

<sup>1</sup>*Institute of Geography of the National Academy of Sciences of Ukraine,*

<sup>2</sup>*National Technical University of Ukraine "Kyiv Polytechnic Institute"*

*Str. Vladimirskaya, 44, Kyiv, 01034, Ukraine*

The problems of mathematical analysis of geographic information with using the procedure of cluster analysis were considered. A software module for various types of cluster analysis of geographical objects and the automated construction of thematic maps was programmed in ArcGIS. Based on the data about accumulation of waste in the regions of Ukraine method of cluster analysis using the geographical coordinates of the centroids of objects to take into account their position in content classification was tested. The results of cluster analysis identified groups of regions of Ukraine with similar indicators for waste management and the potential hazards to the environment.

*Key words:* mathematical methods, cluster analysis, GIS, waste management.

## ИСПОЛЬЗОВАНИЕ КЛАСТЕРНОГО АНАЛИЗА ДЛЯ ПРОСТРАНСТВЕННОГО МОДЕЛИРОВАНИЯ В ГЕОИНФОРМАЦИОННЫХ СИСТЕМАХ

В. Путренко<sup>1</sup>, В. Тихоход<sup>2</sup>

<sup>1</sup>*Институт географии НАН Украины,*

<sup>2</sup>*Национальный технический университет Украины*

*"Киевский политехнический институт"*

*ул. Владимирская, 44, г. Киев, 01034, Украина*

Рассмотрено вопросы математического анализа географической информации с помощью процедуры кластерного анализа. Разработан программный модуль для проведения кластерного анализа географических объектов и автоматизированного построения тематических карт в программной среде ArcGIS.

*Ключевые слова:* математические методы, кластерный анализ, ГИС, бытовые отходы.