

УДК 519.237.8+550.34

М. Г. Сидорова, А. А. Колдунова

Дніпропетровський національний університет імені Олеся Гончара

ЗАСТОСУВАННЯ МЕТОДІВ КЛАСТЕРНОГО АНАЛІЗУ ПРИ ДОСЛІДЖЕННІ КОМПОНЕНТ ВЕКТОРА ІНДУКЦІЇ ЗА ДАНИМИ ЯПОНСЬКИХ ГЕОМАГНІТНИХ ОБСЕРВАТОРІЙ

Розроблено інформаційну технологію кластерного аналізу значень векторів індукції на 4 обсерваторіях поблизу від катастрофічного землетрусу Тохоку 11.03.2011.

Ключові слова: *інформаційна технологія, кластерний аналіз, вектори індукції (Візе-Паркінсона).*

Разработана информационная технология кластерного анализа значений векторов индукции на 4 обсерваториях близости от катастрофического землетрясения Тохоку 11.03.2011.

Ключевые слова: *информационная технология, кластерный анализ, векторы индукции (Визе-Паркинсона).*

It has been developed the information technology of cluster analysis of the values of the components of the induction vectors for 4 observatories near to catastrophic Tohoku earthquake of 11.03.2011.

Key words: *information technology, cluster analysis, (Wieze-Parkinson) induction vectors.*

Вступ. Кластерний аналіз є одним з важливих завдань інтелектуального аналізу даних (Data Mining), дозволяє зрозуміти структуру багатовимірних даних, спростити подальшу обробку, скоротити вихідну вибірку, виявити нетипові об'єкти, сформулювати або перевірити гіпотези на підставі отриманих результатів. Не вимагаючи апріорних знань про сутність досліджуваних явищ, процесів або ситуацій та використовуючи мінімальну про них інформацію, методи кластерного аналізу дозволяють знаходити практично цінні знання про природу об'єкта дослідження (кластери об'єктів, ієрархічні структури в даних, еталонні об'єкти і «викиди») та є корисними в різних прикладних областях: економіці, екології, соціології, маркетингу, біології, медицині, інформатиці та ін.

Постановка проблеми. Матриця Візе-Паркінсона, що показує відношення вертикальної і горизонтальної компонент магнітного поля (складається з комплексних компонент W_{zx} і W_{zy}), впроваджена до геоелектрики Г. Візе, В. Паркінсоном, У. Шмукером, має високу чутливість до геоелектричних неоднорідностей далеко від точки спостереження; обробка і інтерпретація її компонент є досить складною

задачею, оскільки доводиться зустрічатися з високим рівнем промислових перешкод і складною геоелектричною будовою середовища.

Проведений геофізиками кропіткий аналіз матеріалів японських геомагнітних обсерваторій показав наявність посиленних шумів на періодах, менших за 500 с. [5], а застосування стійких до шумів алгоритмів лише частково послабило їх вплив. У ході обробки було знайдено різні похибки, природа яких не завжди зрозуміла. Моменти аномалій амплітуди і фази змінюються для різних періодів варіацій, різних компонент функцій відгуку, різних станцій. Така мінливість може бути зумовлена впливом геомагнітних явищ та внеском перешкод, а також відображати міграцію літосферної емісії, зміни електропровідності у літосфері Землі (виникає припущення про наявність впливу провісників і косейсмічних явищ). Очевидна перспективність використання як провісника землетрусів геомагнітних функцій відгуку (вектора індукції та горизонтального тензора), що реагують на літосферну емісію і на зміни електропровідності. Актуальність дослідження визначається необхідністю застосування нових сучасних підходів для визначення глибинної будови розломних зон і земної кори в цілому, і зокрема. Дослідження глибинних розломних зон дає ключ до розуміння характеру тектонічних рухів, сейсмічної активності, флюїдного режиму, локалізації родовищ корисних копалин.

Математична інтерпретація цих аномалій за допомогою використання методів статистичного аналізу (пошук кореляцій, регресій та, за наявності, їх вилучення, встановлення залежностей, кластеризація тощо) може дати цінний матеріал для тлумачення природи підготовки землетрусів і виникнення провісників.

Аналіз останніх досліджень і публікацій. Існує безліч методів аналізу МТ даних, що дозволяють локалізувати аномалії, оцінити розмірність середовища, визначити напрямки простягання структур, а також наявні різні підходи, що реалізують програми вирішення 1D, 2D і 3D обернених задач. Для інтерпретації (інверсії) магнітотелуричних даних використовується ряд програмних пакетів (WinGlink, GeoTools MT, ZondMT2D та багато інших).

Щодо прогресивних методик на сьогодні активно застосовують методики послідовних часткових інверсій, що надають можливість виявлення глибинної структури земної кори, а в поєднанні з іншими методами (GPS, структурним аналізом) дозволяють контролювати прояви сейсмічності внутрішньоконтинентальних орогенічних зон [6].

Аналіз матриці імпедансу, фазового тензору, матриці Візе-Паркінсона дозволяє навіть за окремими профільними спостереженнями визначити простягання виявлених аномалій утворюючих об'єктів. Безпосередньо до даних матриці Візе-Паркінсона був засто-

сований новий метод візуалізації у вигляді променів, реалізований у програмі Іванова MT_Array [8].

Застосування методів кластерного аналізу із залученням нових програмних засобів та високочотних даних за фазами імпедансу і тіперу розширює можливість магнітоваріаційного зондування і підвищує достовірність інтерпретації результатів.

Постановка задачі. Метою роботи є розробка інформаційної технології обробки компонент вектора індукції на основі методів кластерного аналізу. Дані представлені у вигляді матриці дійсних чисел $X = \{x_{ij}; i = 1, n, j = 1, p\}$, де n – кількість спостережень, p – кількість досліджуваних ознак, x_{ij} – значення j -ї ознаки, у i -й момент спостереження. Необхідно виявити кластерну структуру досліджуваних даних, її структурні зміни у діапазоні спостережень, враховуючи різні періоди варіацій геомагнітного поля, дослідити динамічні зміни значень векторів індукції на кожній з досліджуваних магнітоваріаційних станцій.

Основний матеріал. Для вирішення поставленої задачі пропонується інформаційна технологія, що складається з наступних етапів:

1. Визначення способу обчислення близькості досліджуваних об'єктів (евклідова метрика, манхеттенська, Чебишева) та налаштування параметрів методів.

2. Вибір методу кластерного аналізу. Для отримання найбільш повної інформації про угруповання, притаманні досліджуванім даним, пропонується застосовувати різноманітні алгоритми кластеризації, а саме: алгоритми швидкої ієрархічної агломеративної кластеризації (одиночного зв'язку (ближнього сусіда), повного зв'язку (дальнього сусіда), середнього зв'язку, центрального та Уорда), графовий метод найкоротшого незамкненого шляху, генетичний алгоритм, методи розділової кластеризації К-середніх у варіантах Болла-Холла та Мак-Кіна, а також метод Forel [1, 2].

3. Визначення наявності кластерної структури на основі критерію Дуда – Харта:

$$F_{DH} = w_2 / w_1,$$

де w_2 – сума квадратів внутрішньокластерних відстаней у випадку, коли дані розподілені на 2 кластери, w_1 – сума квадратів внутрішньокластерних відстаней у випадку одного кластеру.

Гіпотеза існування єдиного кластеру однорідних даних відхиляється, якщо значення критерію менше критичного значення, обчисленого за формулою

$$F_{kp} = 1 - \frac{2}{\pi p} - u_{(1-\alpha)} \sqrt{2(1 - 8 / (\pi^2 p)) / (np)},$$

де $u_{(1-\alpha)}$ – квантиль стандартного нормального розподілу рівня $(1-\alpha)$.

4. У випадку наявності кластерної структури проведення кластерного аналізу різними методами. Визначення оптимальної кількості кластерів та найкращого розбиття на основі багатокритеріальної оцінки якості [3], що полягає у порівнянні результатів кластеризації за критеріями якості (сума внутрішньокластерних дисперсій за всіма ознаками, сума квадратів відстаней до центрів класів, сума внутрішньокластерних відстаней, відношення середньої внутрішньокластерної і середньої міжкластерної відстаней, індекси Калінського-Гарабача, Данна, Беджека-Данна), застосування до отриманих оцінок колективних методів прийняття рішень (Борда, Коупленда, плюралітарної процедури, множинного аналізу) та агрегування найбільш якісних результатів на основі ансамблю алгоритмів [4].

5. Візуалізація та аналіз отриманих кластерів.

Розглянемо результати проведених досліджень. Аналізувалися значення векторів індукції Візе-Паркінсона, отримані шляхом перетворення геомагнітного поля магнітоваріаційних станцій на території Японії: Esashi, Kakioka, Kanozan окремо при різних значеннях періодів варіацій за спостереженнями 2011 та 2012 років. Результати показали, що досліджувані дані не є однорідними. При аналізі даних, що відповідають періоду 8 варіацій геомагнітного поля як у 2011 (рис. 1), так і у 2012 (рис. 3а) роках, отримали три відокремлених кластери, що відповідають спостереженням за геомагнітним полем магнітоваріаційних станцій Esashi, Kakioka, Kanozan відповідно. Тобто можна зробити висновок, що геомагнітне поле за значеннями векторів індукції Візе-Паркінсона значно відрізняється на кожній з магнітоваріаційних станцій. Тому подальші дослідження з метою виявлення передвісників землетрусів необхідно проводити окремо на кожній зі станцій.

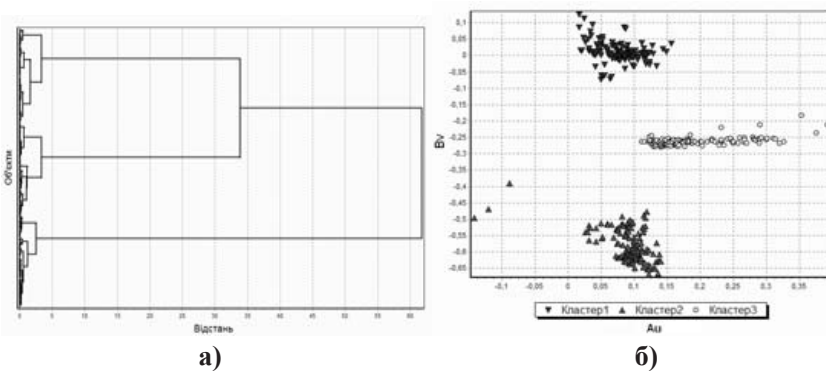


Рис. 1. Результати кластерного аналізу значень векторів індукції Візе-Паркінсона за 2011 рік при періоді 8 варіацій геомагнітного поля:
а) дендрограма; б) діаграма розсіювання

Для отриманих кластерів було обчислено середні значення досліджуваних показників. Результати за 2011 р. наведено у табл. 1.

Таблиця 1

Середні значення показників у кожному кластері

	Au	Bu	Av	Bv
Кластер 1	0,0773	-0,1304	-0,0008	0,0099
Кластер 2	0,0901	-0,4795	-0,1106	-0,5772
Кластер 3	0,1926	-0,3935	0,0037	-0,2630

Аналізуючи результати досліджень, можна зробити висновок, що значення періоду варіацій геомагнітного поля значним чином впливає на кластерну структуру. Наприклад, при періоді 8000 для даних спостережень 2011 року кластерна структура являє собою 2 групи (рис. 2). До першої групи увійшли значення векторів індукції Візе-Паркінсона, отриманих на магнітоваріаційній станції Esashi, а спостереження двох інших станцій склали другий кластер. А у 2012 році (рис. 3б) при тому самому значенні періоду було виявлено змішану кластерну структуру.

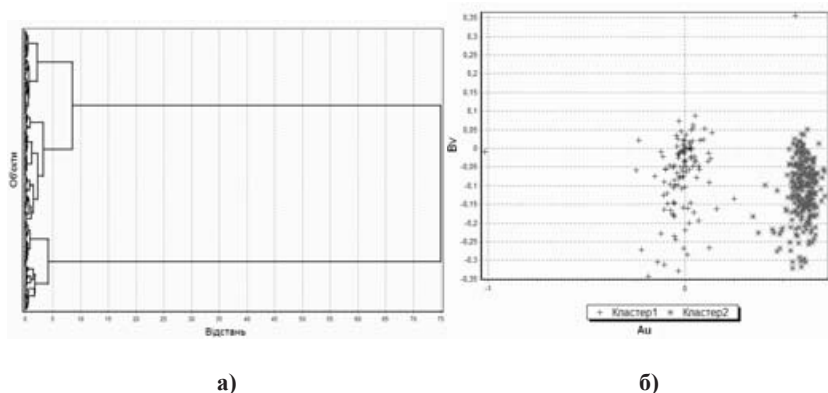


Рис. 2. Результати кластерного аналізу значень векторів індукції Візе-Паркінсона за 2011 рік при періоді 8000 варіацій геомагнітного поля: а) дендрограма; б) діаграма розсіювання

Також досліджувалися динамічні зміни значень векторів індукції Візе-Паркінсона на кожній з магнітоваріаційних станцій окремо у 2009, 2011 та 2012 роках. Було зроблено висновок, що даним притаманна кластерна структура, яка різниться в залежності від обраного значення періоду варіацій геомагнітного поля.

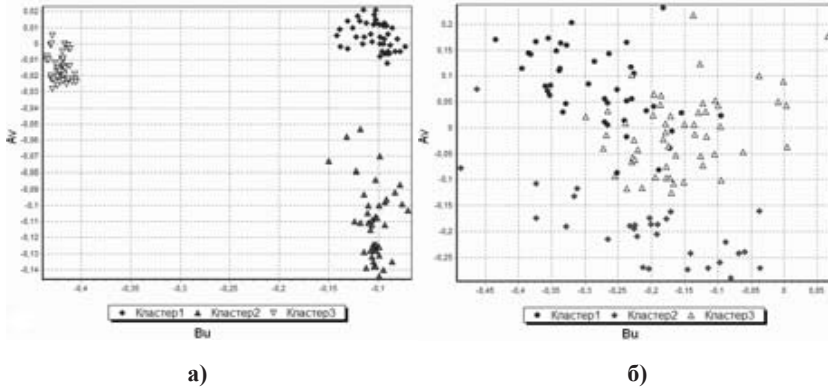


Рис. 3. Результати кластерного аналізу значень векторів індукції Візе-Паркінсона за 2012 рік при періоді варіацій геомагнітного поля: а) рівного 8; б) рівного 8000

Висновки. У роботі запропоновано інформаційну технологію кластерного аналізу значень векторів індукції Візе-Паркінсона, отриманих шляхом перетворення геомагнітного поля магнітоваріаційних станцій на території Японії: Esashi, Kakioka, Kanozan при різних значеннях періодів варіацій. Досліджено кластерну структуру даних, її структурні зміни у діапазоні спостережень, з урахуванням різних періодів варіацій геомагнітного поля, а також динамічних змін значень векторів індукції на кожному з досліджуваних об'єктів. Виявлено, що геомагнітне поле за значеннями векторів індукції Візе-Паркінсона значно відрізняється на кожній з магнітоваріаційних станцій. Тому подальші дослідження з метою виявлення передвісників землетрусів необхідно проводити окремо на кожній зі станцій.

Бібліографічні посилання

1. **Мандель И. Д.** Кластерный анализ / И. Д. Мандель. – М., 1988. – 176 с.
2. **Айвазян С. А.** Классификация многомерных наблюдений / С. А. Айвазян, З. И. Бежаева, О. В. Староверов. – М., 1974. – 240 с.
3. **Байбуз О. Г.** Інформаційна технологія кластеризації даних у часовому періоді спостережень / О. Г. Байбуз, М. Г. Сидорова // Системні дослідження та інформаційні технології. – 2013. – № 4. – С. 59–66.
4. **Сидорова М. Г.** Застосування ансамблів алгоритмів для підвищення стійкості результатів кластеризації / М. Г. Сидорова // Актуальні проблеми автоматизації та інформаційних технологій : зб. наук. праць. – Дніпропетровськ, 2013. – Т. 17. – С. 22–29.
5. Вариации компонент вектора индукции и горизонтального тензора перед землетрясением Тохоку 11 марта 2011 г. по данным японских геомагнитных об-

серваторий / [И. И. Рокитянский, В. И. Трегубенко, В. И. Бабак, А. В. Терешин] // Геофизический журнал. – Киев, 2013. – Т. 35. – № 3. – С. 115–130.

6. **Баталёва Е. А.** Глубинная структура крупнейших разломных зон западной части Киргизского Тянь-Шаня и современная геодинамика по данным магнитотеллурических зондирований: дис. ... кандидата геолого-минералогических наук : 25.00.03, 25.00.10. – Бишкек, 2005. – 200 с.

7. Магнитовариационное зондирование: новые возможности / [М. Н. Бердичевский, В. И. Дмитриев, Н. С. Голубцова и др.] // Физика Земли. – 2003. – № 9. – С. 3–30.

8. **Иванов П. В.** Программное обеспечение МТ_Аггау для анализа магнитотеллурических данных / П. В. Иванов // Материалы докл. XVII Междунар. науч. конф. студентов, аспирантов и молодых учёных «Ломоносов». – М., 2010.

9. **Куликов В. А.** Возможности магнитотеллурических методов при проведении работ на высокоомных разрезах / В. А. Куликов, А. Г. Яковлев // Разведка и охрана недр. – 2008. – № 12. – С. 3–6.

Надійшла до редколегії 19.10.14