

ДОСЛІДЖЕННЯ МЕТОДІВ КЛАСТЕРІЗАЦІЇ ДІЛЯНОК РІЗНОЇ ПОТУЖНОСТІ ВУГІЛЬНОГО ПЛАСТА З МЕТОЮ СТВОРЕННЯ ЇХ ПРИРОДНОЇ ТИПІЗАЦІЇ ЗА ВМІСТОМ ГЕРМАНІЮ (НА ПРИКЛАДІ ПЛАСТА C₈^H ШАХТИ «ДНІПРОВСЬКА», УКРАЇНА)

Ішков Валерій Валерійович,

кандидат геолого-мінералогічних наук, доцент
Національний ТУ «Дніпровська політехніка», Україна
старший науковий співробітник
інституту геотехнічної механіки ім. М.С. Полякова НАН України, Україна

Чернобук Олександр Іванович

заступник директора, департамент стратегічного планування виробництва,
Грузинський марганець, Грузія

Козар Микола Антонович

кандидат геологічних наук, старший науковий співробітник
інституту геохімії, мінералогії і рудоутворення ім. М.П. Семененка НАН
України, Україна

Березняк Олена

Дьячков Павло Анатолійович

старший викладач, Національний ТУ «Дніпровська політехніка», Україна

Актуальність дослідження вмісту германію у вугільних пластах обумовлена можливістю його промислового вилучення та використання в якості цінного попутного компонента. Зараз вугілля є основним оціненим джерелом Ge в Україні, Китаї, Узбекистані, а також в Росії. Ge-вугільні родовища розробляються в Англії, Канаді, США, Україні, Росії та ін.

Особливу актуальність виконаним дослідженням надає рішення Ради національної безпеки та оборони України від 16 липня 2021 року «Про стимулювання пошуку, видобутку та збагачення корисних копалин, які мають стратегічне значення для сталого розвитку та обороноздатності держави» та Указ Президента України №306/2021, який вводить в дію це рішення. В цих документах руди Ge включені до переліку, що мають стратегічне значення для сталого розвитку та обороноздатності держави.

Для об'єктивної геолого-економічної оцінки можливості попутного вилучення германію з вугілля, відходів і продуктів його переробки та планування найбільш ефективних організаційно-технічних заходів з цього приводу, перш за все необхідно мати відомості про характер розподілу і рівень концентрації цього елемента у вугіллі. З метою одержання такої інформації авторами були виконані

детальні дослідження розподілу германія по площі і в розрізі вугільного пласта c_8^H поля шахти «Дніпровська».

Останні досягнення. Раніше були досліджені особливості розподілу «малих елементів», які відносяться до групи «токсичних та потенційно токсичних елементів» у вугільних пластах деяких шахт та геолого-промислових районів Донбасу[1-21]. Були обґрунтовані методи природної типізації вугільних родовищ за вмістом супутніх елементів та родовищ нафти Дніпровсько-Донецької западини за вмістом металів. У той же час, аналіз методів об'єктивної (природної) типізації ділянок вугільного пласта різної потужності за концентраціями Ge у вугільному пласті раніше не виконувався.

Мета досліджень – встановити та обґрунтувати найефективніший метод створення об'єктивної (природної) типізації ділянок вугільного пласта c_8^H шахти «Дніпровська» різної потужності за концентраціями Ge.

Методи дослідження. Фактологічною основою роботи були результати 370 аналізів Ge, виконаних після 1981р. в центральних сертифікованих лабораторіях виробничих геологорозвідувальних організацій України з матеріалу пластових проб отриманих виробничими і науково-дослідницькими підприємствами і організаціями та вимірювань потужності пласту. У ряді випадків вони доповнювались аналізами пластових проб відібраних борозновим методом із дублікатів керна і гірничих виробках за участю авторів та співробітників геологічної служби вугледобувного підприємства і виробничих геологорозвідувальних організацій в період з 1981 по 2013 рік.

Вміст Ge визначався кількісним емісійним спектральним аналізом. На внутрішній лабораторний контроль направлено 7% дублікатів проб. Зовнішньому лабораторному контролю піддано 10% дублікатів проб. Якість результатів аналізів (правильність і відтворюваність) оцінювалася як значимість середньої систематичної похибки, яка перевіряється за допомогою критерію Стюдента і значимість середньої випадкової похибки, яка перевіряється за допомогою критерію Фішера. Оскільки вказані вище похибки при рівні значимості 0,95 є не значимими, якість аналізів визнано задовільною.

За допомогою програм Excel 2016 і Statistica 13.3 на початковому етапі обробки первинної геохімічної інформації розраховувалися значення основних описових статистичних показників, виконувалась побудова частотних гістограм вмісту і встановлення закону розподілу германію. В ході побудови графіків всі значення концентрацій Ge та потужності пласта нормувались для приведення вибірки до одного масштабу незалежно від одиниць виміру та розмаху вибірок.

Як відомо, процедура типізації – це систематизація об'єктів по апріорно заданим ознакам. Зазвичай для цього використовуються кластерний аналіз, таксономія, розпізнавання образів, факторний аналіз.

Важливо що на відміну від інших методів, які використовують при вирішенні задач типізації, кластерний аналіз не вимагає апріорних припущень про набір даних, що не накладає обмеження на подання досліджуваних об'єктів, дозволяє аналізувати природні показники різних типів даних (інтервальних даних, частот, бінарних даних, тощо). Використання кластерного аналізу з метою типізації має ряд переваг, оскільки дозволяє виконати розбиття безлічі досліджуваних об'єктів

і ознак на однорідні у відповідному розумінні групи або кластери, а також виявити внутрішню структуру (на різних ієрархічних рівнях) вибіркової сукупності. У той же час, як і будь-який інший метод, кластерний аналіз має певні недоліки. Зокрема, склад і кількість кластерів залежить від обраних критеріїв угруповання («стратегії класифікації»), а застосування різних методів, що відповідають різним концептуальним підходам до виділення таксонів, до одних і тих же вибірок, може призвести до суттєво відмінних результатів. Таким чином, характерною особливістю кластерного аналізу, на відміну від інших методів багатомірної статистики, є сильна залежність одержуваних результатів від апріорних установок дослідника на змістовному рівні. У нашому випадку, до апріорних установок відносяться: відсутність гіпотез щодо числа кластерів, їх структури на форми; досягнення максимальної візуалізації розбиття ділянок вугільного пласта різної потужності по класам на різних масштабних рівнях; встановлення методу (алгоритму) кластеризації для найбільш стійкого поділу всієї сукупності проб, що розглядаються.

У кластерному аналізі вважається, що: а) вибрані характеристики допускають в принципі бажане розбиття на кластери; б) одиниці виміру (масштаб) обрані правильно. Таким чином, вибір масштабу у процедурах класифікації відіграє значну роль. Для приведення до одного масштабу вихідні дані нормують, що ми й зробили.

Для досягнення поставленої в роботі мети у процесі досліджень було здійснено кластеризацію різними методами, які реалізовані у найпопулярніших професійних статистичних програмних платформах «STATISTICA» та «SPSS»; виконано їх аналіз та обґрунтовано вибір найбільш оптимального з них.

У роботі використовувалися версії програм Excel 2016, STATISTICA 13.3 та IBM SPSS Statistics 22.

Результати дослідження та їх обговорення. Для розробки методики вибору найефективнішого методу створення об'єктивної (природної) типізації ділянок вугільного пласта різної потужності за концентраціями Ge загальна вибірка із врахуванням кількості проб була поділена на 10 приватних вибірок, основні характеристики яких наведені у таблиці 1.

Для виконання кластерного аналізу в професійних програмних платформах «STATISTICA» та «SPSS» пропонується сімейство ієрархічних агломеративних методів, двовходового об'єднання та ітеративний дивізімний метод середніх.

Таблиця 1
Основні характеристики приватних вибірок вмістів Ge

Приватні вибірки	1	2	3	4	5	6	7	8	9	10
Інтервал потужності, м	$\leq 0,45$	0,45 – 0,50	0,50 – 0,55	0,55 – 0,60	0,60 – 0,65	0,65 – 0,70	0,70 – 0,75	0,75 – 0,80	0,80 – 0,85	$\geq 0,85$
Кількість проб	35	40	35	19	26	44	58	47	27	22
Нормовані медіанні значення вмісту Ge	0,73	0,60	0,48	0,44	0,34	0,28	0,22	0,18	0,14	0,12

Метод двохходового об'єднання відносно рідко використовується для одночасної кластеризації як спостережень і змінних. У цьому випадку очікується, що і спостереження і змінні одночасно роблять внесок у виявлення кластерів, які далі інтерпретуються в геологічних поняттях. Головним недоліком методу є проблеми з понятійною інтерпретацією результатів, які є наслідком того, що відстань між різними кластерами може визначатися відмінностями в змінних. Проблематичність понятійної інтерпретації результатів аналізу не дає можливості для його використання як оптимального методу вирішення поставленої задачі типізації.

Використання ітеративного дивізимного методу K-середніх для оптимальної типізації ділянок за вмістом Ge теж має суттєві недоліки. Для нього характерна властива всім ітеративним дивізимним методам проблема субоптимальних рішень, які полягають у іноді невдалих вихідних розбиттях вибірових сукупностей. Його використання має на увазі існування апріорних гіпотез щодо числа кластерів, а результат кластеризації, представлений у вигляді системи таблиць, не дозволяє наочно і однозначно виявити і візуалізувати структуру типізації. Вказані недоліки методу не дозволяють розглядати його як оптимальний для типізації ділянок пласту за вмістом Ge.

Сімейство ієрархічних об'єднуючих методів реалізованих у використаних програмах відноситься до методів кластерного аналізу, що найчастіше використовуються. Всі вони полягають у послідовному об'єднанні найбільш схожих об'єктів, яке можна візуалізувати у вигляді деревоподібної діаграми – дендрограми (яка графічно відображає ієрархічну структуру матриці подібності об'єктів). Така наочність результатів кластеризації є суттєвою перевагою цих методів. Як правило, в дендрограмі по горизонталі вказуються об'єкти, що кластеризуються, а по вертикалі - значення міжкласових відстаней, при яких відбувається їх об'єднання (коефіцієнт злиття або дистанція об'єднання). При цьому в результаті аналізу формуються групи кластерів, що не перекриваються, причому кожен кластер є елементом ширшого кластера на більш високому рівні подібності. За способом угруповання всі ієрархічні агломеративні методи у

використаних програмах поділяються на наступні: метод одиночного зв'язку («найближчого сусіда»); метод повного зв'язку («найвіддаленішого сусіда»); різновиди методу «середнього зв'язку» – незважений метод «середнього зв'язку» («незважене попарне середнє») та зважений метод «середнього зв'язку»; зважений центроїдний метод і метод Уорда. Крім того, у всіх перерахованих методах можуть бути використані міжкласові відстані – евклідова відстань (або її квадрат), манхеттенська відстань («відстань міських кварталів»), метрики Чебишева і Мінковського, лінійний коефіцієнт кореляції. Стосовно особливостей вирішуваної задачі угруповання найбільш оптимальним є використання як міри подібності евклідової відстані.

Розглянемо можливості застосування ієрархічних агломеративних методів кластеризації до побудови типізації ділянок пласта за вмістом Ge більш докладно.

Метод одиночного зв'язку формує кластери із принципу наявності хоча б одного зв'язку між об'єктами. В результаті роботи цього методу кластери видаються довгими «ланцюжками» «зчепленими разом» лише окремими елементами, які випадково виявилися ближчими від інших один до одного. Незважаючи на те, що його результати інваріантні до монотонних перетворень матриці подібності і використання методу не обмежує присутність «збігу» в даних, практичне його застосування з метою типізації викликає певні труднощі.

На прикладі результатів кластеризації родовищ за вмістом Ge (рис. 1) видно, що уся сукупність родовищ по мірі наближення до завершення процесу кластеризації формує кластери 1 та 2, які у свою чергу складаються з відповідно вкладених підкластерів 1.1, 1.2, 1.1.1, 1.1.2, тощо. Аналіз рис. 1 не дозволяє визначити кількість та структуру кластерів, що містяться у вихідних даних, а також суттєво ускладнюється встановлення кількості та структури кластерів на інших масштабних рівнях.

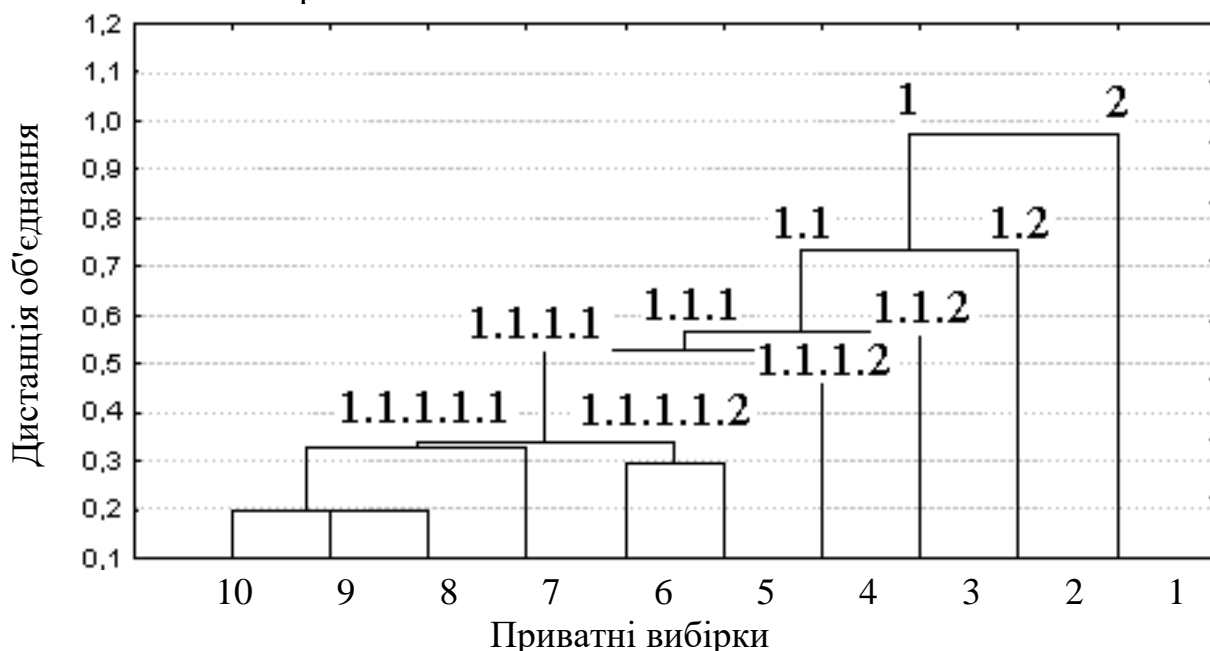


Рис. 1. Дендрограма результатів кластеризації методом одиночного зв'язку ділянок по вмісту Ge. Умовні позначення: 1, 2, 1.1, 1.2, 1.1.1, 1.1.2, 1.1.1.1, 1.1.1.2, 1.1.1.1.1, 1.1.1.1.2 – кластери

Метод повного зв'язку (рис. 2), на відміну від розглянутого вище методу накладає більш жорсткі вимоги до об'єднання об'єктів в один кластер. У цьому випадку з'являється тенденція до виявлення відносно компактних гіперсферичних (у багатовимірному просторі) кластерів, що поєднують схожі об'єкти. Тут відстані між кластерами визначаються найбільшою відстанню між будь-якими двома об'єктами в різних кластерах (тобто найбільш віддаленими сусідами).

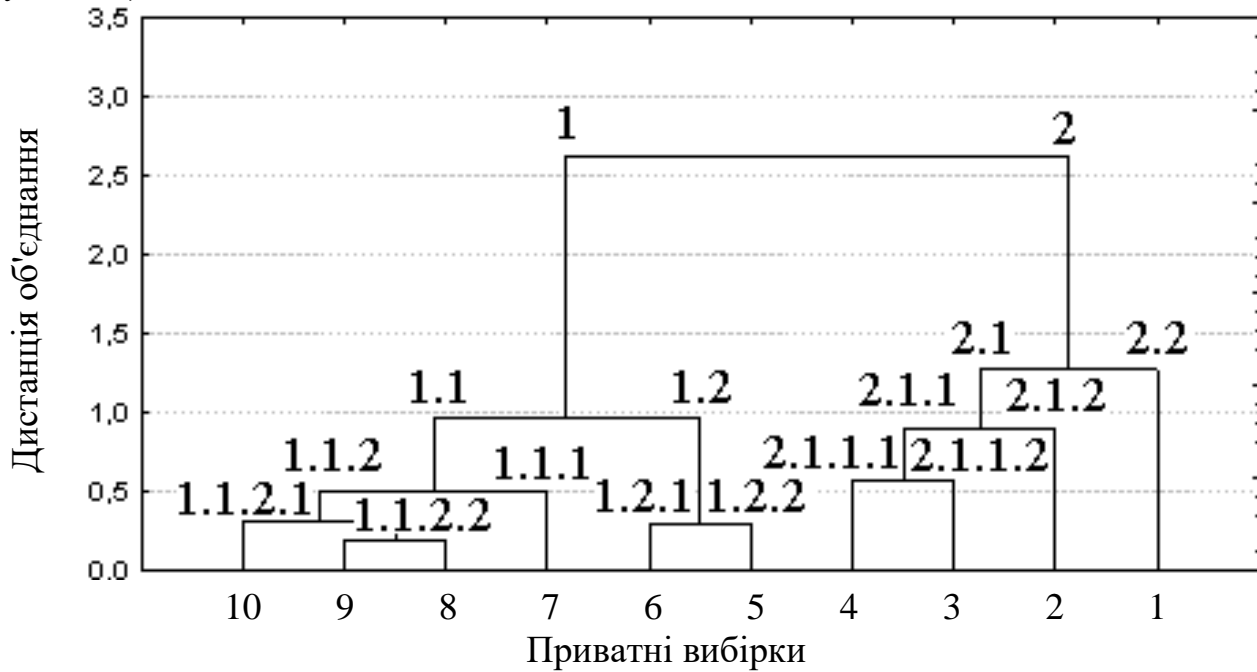


Рис. 2. Дендрограма результатів кластеризації методом повного зв'язку ділянок по вмісту Ge. Умовні позначення: 1, 2, 1.1, 1.2, 1.1.1, 1.1.2, 1.1.2.1, 1.1.2.2, 1.2.1., 1.2.2, 2.1, 2.2, 2.1.1, 2.1.2, 2.1.1.2, 2.1.1.1 – кластери

Зіставлення рис. 1 з рис. 2 дозволяє виявити ряд переваг кластеризації з використанням методу повних зв'язків. У той же час, якщо при кластеризації методом одиночного зв'язку отримана дендрограма досить переконливо вказує на кількість та структуру кластерів (рис. 1), то при кластеризації методом повних зв'язків спостерігається деяка зміна ділянок між кластерами і загалом, при порівнянні результатів обох методів виявлення остаточної структури та кількості родовищ у кластерах (рис. 1, рис. 2) не так очевидно. Крім того, в обох випадках використання тільки дендрограми без залучення первинних даних ускладнює віднесення окремих об'єктів до тих чи інших кластерів, а сама структура кластерів виявлена досить слабо.

Метод «середнього зв'язку» розроблений Сокелом та Мінченером у 1958 р. як компроміс між методами одиночного та повного зв'язку. У версії програм STATISTICA 13.3 та IBM SPSS Statistics 22 реалізовано два різновиди методу: незважений метод «середнього зв'язку» («незважене попарне середнє») та зважений метод «середнього зв'язку». У першому різновиді методу відстань між двома кластерами обчислюється як середня відстань між усіма парами об'єктів у них, а у другій - крім того, розмір кластерів (тобто кількість об'єктів, що містяться в них) використовується як ваговий коефіцієнт. Використання кількості об'єктів, що містяться в кластері, як ваговий коефіцієнт передбачає

«хорошу якість» аналізу за наявності у вибірці кластерів нерівного розміру. Результати кластеризації ділянок пласта за вмістом Ge обома методами у вигляді дендрограм наведені відповідно на рис. 3 та рис. 4. Аналіз наведених дендрограм дозволяють прийти до висновку, що у випадку з поділом ділянок пласта за вмістом Ge результати кластерного аналізу відповідно методами незваженого середнього зв'язку і зваженого середнього зв'язку у нашому випадку повністю співпадають.

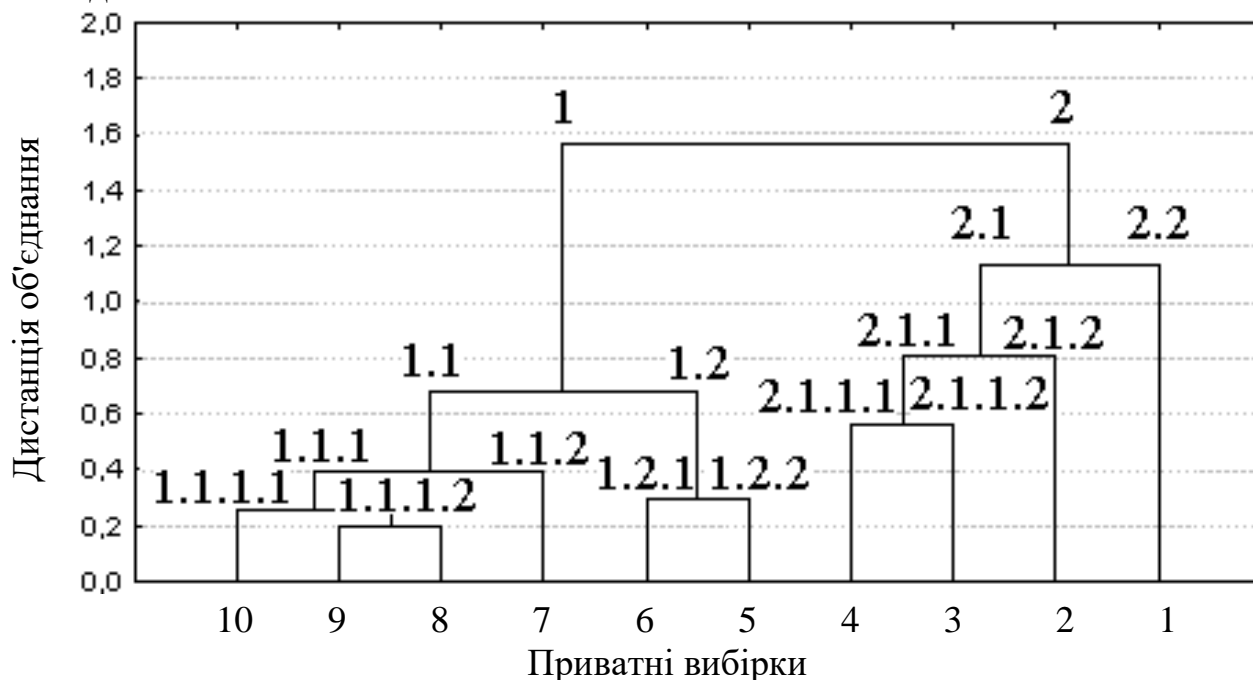


Рис. 3. Дендрограма результатів кластеризації методом незваженого середнього зв'язку ділянок по вмісту Ge. Умовні позначення: 1, 2, 1.1, 1.2, 1.1.1, 1.1.2, 1.1.1.1, 1.1.1.2, 1.2.1., 1.2.2, 2.1, 2.2, 2.1.1, 2.1.2, 2.1.1.2, 2.1.1.1 – кластери

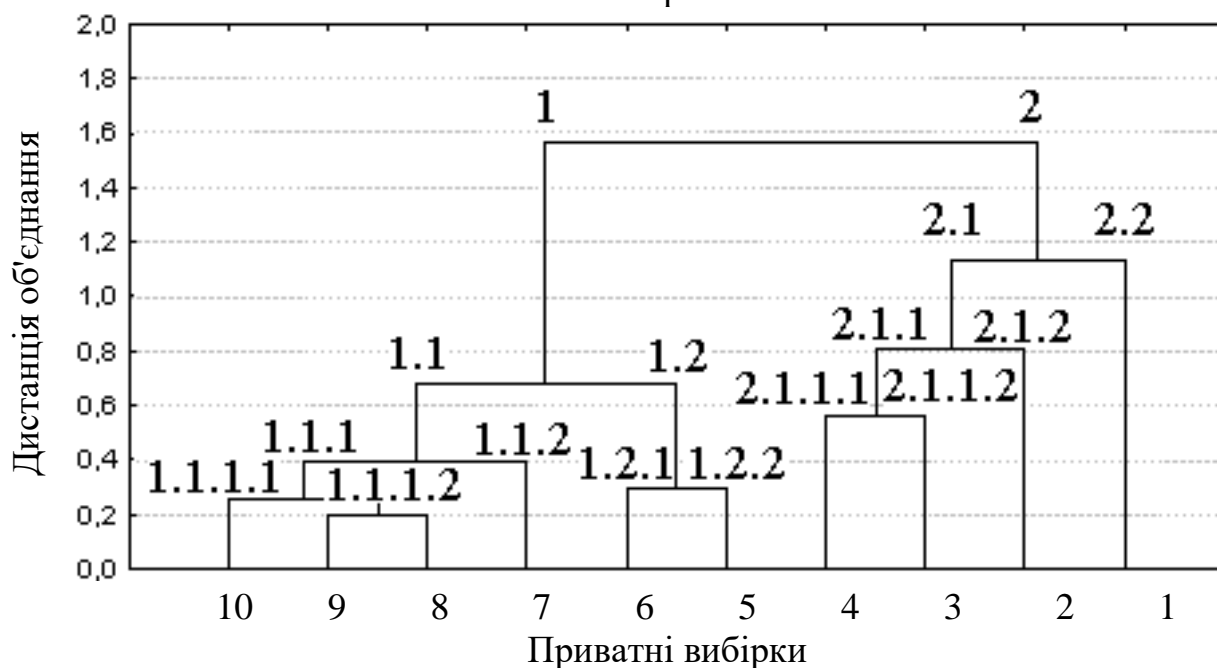


Рис. 4. Дендрограма результатів кластеризації методом зваженого середнього зв'язку ділянок по вмісту Ge. Умовні позначення: 1, 2, 1.1, 1.2, 1.1.1, 1.1.2, 1.1.1.1, 1.1.1.2, 1.2.1, 1.2.2, 2.1, 2.2, 2.1.1, 2.1.2, 2.1.1.2, 2.1.1.1 – кластери

Зважений центроїдний метод використовує як відстань між кластерами (об'єктами) відстань між їх центрами тяжіння. На рис. 5 наведено дендрограму результатів кластеризації зваженим центроїдним медіанним методом ділянок пласта за вмістом Ge. Аналіз цієї дендрограми на відміну від наведених раніше (рис. 1, рис. 2, рис. 3 та рис. 4) дозволяє не тільки досягти найбільш стійкого поділу всієї сукупності ділянок, що розглядаються, але й максимальної візуалізації розбиття об'єктів по класам на різних масштабних рівнях при апріорній відсутності гіпотез щодо числа кластерів та їх форми. При цьому впевнено виділяється наявність та структура кластерів незалежно від масштабного рівня їх формування, чітко простежується послідовність об'єднання окремих родовищ та їх груп у результуючий кластер. Наявність цих переваг у даному випадку дозволяє максимально використовувати вже існуючу інформацію для розробки природних типізацій ділянок вугільних пластів за вмістом Ge та інтерпретувати отримані результати у геологічних поняттях.

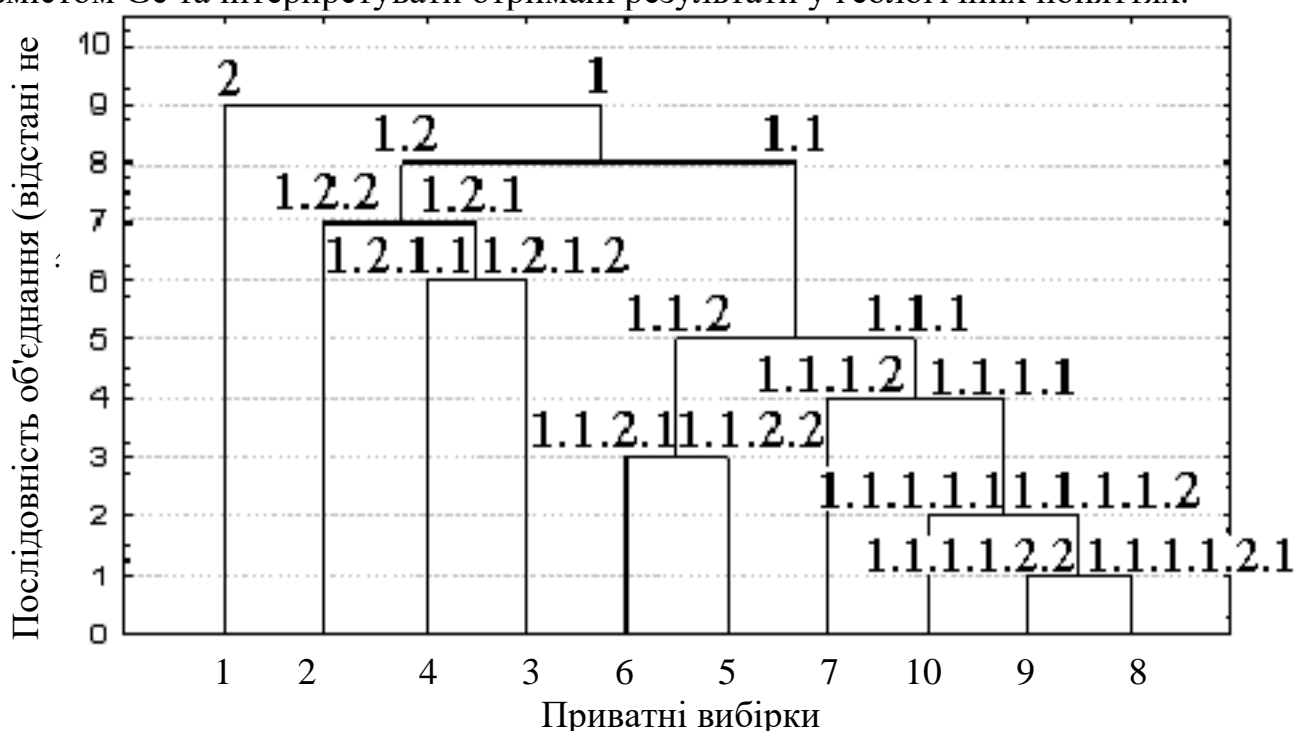


Рис. 5. Дендрограма результатів кластеризації зваженим центроїдним медіанним методом ділянок по вмісту Ge. Умовні позначення: 1, 2, 1.1, 1.2, 1.1.1, 1.1.2, 1.1.1.1, 1.1.1.2, 1.2.1, 1.2.2, 1.2.1.1, 1.1.2.1, 1.1.2.2, 1.1.1.1.1, 1.1.1.1.2, 1.1.1.1.2.1, 1.1.1.1.2.2 – кластери

Висновок. В результаті виконаних досліджень встановлено що використання зваженого центроїдного медіанного методу кластерного аналізу, який запроваджено у професійних статистичних програмних платформах «STATISTICA» та «SPSS» є найбільш оптимальним для вільного від суб'єктивізму дослідника поділу ділянок вугільного пласта за вмістом Ge на таксони.

Побудовані дендрограми кластеризації ділянок вугільного пласта за вмістом Ge можуть бути використані як основа для розробки природної типізації вугільних пластів для геолого-економічної оцінки. Це дозволить максимально використовувати вже наявну інформацію та інтерпретувати отримані результати

у геолого-генетичних поняттях, що надає можливість її використання для комплексного використання мінеральної сировини, вирішення стратегічних питань сталого розвитку України.

Список літератури

1. Козій Є.С., Ішков В.В. (2017). Класифікація вугілля основних робочих пластів Павлоград-Петропавлівського геолого-промислового району за вмістом токсичних та потенційно токсичних елементів. Збірник наукових праць «Геотехнічна механіка». № 136, С. 74 - 86
2. Ишков В.В., Сердюк Е.А., Слипенький Е.В. Особенности применения методов кластерного анализа для классификации угольных пластов по содержанию токсичных и потенциально токсичных элементов (на примере Красноармейского геолого-промышленного района) // Сб. науч. тр. НГУ. -2003. - № 19, Т.1. - С. 5 - 16.
3. Ishkov V.V., Koziy E.S., Lozovoy A.L. Definite peculiarities of toxic and potentially toxic elements distribution in coal seams of Pavlograd-Petropavlovka region / Збірник наукових праць НГУ . 2013. № 42. С.18 - 23.
4. Ишков В.В., Козий Е.С. О распределении токсичных и потенциально-токсичных элементов в угле пласта с₆^н шахты «Терновская» Павлоград - Петропавловского геолого-промышленного района / Матеріали міжнародної конференції «Форум гірників». 2013. С. 49 – 55.
5. Ишков В.В., Козий Е.С. Новые данные о распределении токсичных и потенциально токсичных элементов в угле пласта с₆^н шахты «Терновская» Павлоград-Петропавловского геолого-промышленного района / Збірник наукових праць НГУ. 2013. № 41. С. 201 – 208.
6. Ишков В.В., Козий Е.С. О распределении золы, серы, марганца в угле пласта с₄ шахты «Самарская» Павлоград - Петропавловского геолого-промышленного района / Збірник наукових праць НГУ. 2014. № 44. С. 178 - 186.
7. Ишков В.В., Козий Е.С. О классификации угольных пластов по содержанию токсичных элементов с помощью кластерного анализа / Збірник наукових праць НГУ. 2014. № 45. С. 209 – 221.
8. Kozar, M.A., Ishkov, V.V., Kozii, Ye.S., Pashchenko P.S. (2020). New data about the distribution of nickel, lead and chromium in the coal seams of the Donetsk-Makiivka geological and industrial district of the Donbas. Journ. Geol. Geograph. Geocology. №. 29(4), pp. 722-730. <http://doi: 10.15421/112065>
9. Ишков В.В. Кобальт. Ванадий в угле основных рабочих пластов Алмазно-Марьевского геолого-промышленного района Донбасса. Науковий вісник НГУ. 2009. № 10. С. 48-53.
10. Ишков В.В., Козій Є.С., Труфанова М.О. Особливості онтогенезу уролітів жителів Дніпропетровської області. Мінерал. журн. 2020. 42, № 4. С. 50 - 59. <https://doi.org/10.15407/mineraljournal.42.04.050>
11. Ишков В.В., Нагорный В.Н. О закономерностях накопления ртути в угольных пластах Красноармейского геолого-промышленного района // Науковий вісник Національної гірничої академії України. № 2. – Днепропетровск, 2005. – С. 84-88.

12. Ишков В.В. Мышьяк и фтор в угольных пластах Лисичанского геолого-промышленного района // Збірник наукових праць Національного гірничого університету № 33, т. 1. - Днепропетровск, 2009. – С. 5 - 16.
13. Ишков В.В., Козій Є.С. Розподіл ртуті у вугільному пласті с₇^н поля шахти «Павлоградська» / Наукові праці Донецького національного технічного університету, Серія: «Гірничо-геологічна». 2020. №1 (23) - 2(24). С. 26 - 33.
14. Козар М.А., Ишков В.В., Козій Є.С., Стрельник Ю.В. Токсичні елементи мінеральної та органічної складової вугілля нижнього карбону Західного Донбасу / Геологічна наука в незалежній Україні: Збірник тез наукової конференції Ін-ту геохімії, мінералогії та рудоутворення ім. М.П. Семененка НАН України. 2021. С.55 - 58.
15. Ишков В.В., Козій Є.С. Накопление Со и Мп на примере пласта С5 Западного Донбасса как результат их миграции из кор выветривания Украинского кристаллического щита / Материалы XVI Международного совещания по геологии россыпей и месторождений кор выветривания «Россыпи и месторождения кор выветривания XXI века: задачи, проблемы, решения»ю 2021. С. 160 - 162.
16. Ишков В.В., Козій Є.С., Стрельник Ю.В. Результати досліджень розподілу кобальту у вугільному пласті k5 поля ВП «шахта «Капітальна» / Збірник праць Всеукраїнської конференції «Від мінералогії і геогнозії до геохімії, петрології, геології та геофізики: фундаментальні і прикладні тренди XXI століття» (MinGeoIntegration XXI). 2021. С. 178 - 181.
17. Ишков В.В., Козій Є.С. Аналіз поширення хрому і ртуті в основних вугільних пластах Красноармійського геолого-промислового району / Вид-во ІГН НАН України. Серія тектоніка і стратиграфія. 2019. № 46. С. 96 - 104.
18. Ишков В.В., Козій Є.С. Деякі особливості розподілу берилію у вугільному пласті k5 шахти «Капітальна» Красноармійського геолого-промислового району Донбасу / Вісник ОНУ. Сер.: Географічні та геологічні науки. 2020. Т. 25, вип. 1(36). С. 214-227.
19. New data about the distribution of nickel, lead and chromium in the coal seams of the Donetsk - Makiiivka geological and industrial district of the Donbas / Kozar M.A., Ishkov V.V., Kozii Ye.S., Pashchenko P.S. / Journ. Geol. Geograph. Geoecology. 2020. № 29(4). pp. 722 - 730. <http://doi: 10.15421/112065>
20. Ишков В.В., Козій Є.С. Особливості розподілу свинцю у вугільних пластах Донецько-Макіївського геолого-промислового району Донбасу / Вид-во ІГН НАН України, Серія тектоніка і стратиграфія. 2020. № 47. С. 77 - 90.
21. Ишков, В.В., Козій, Є.С. Розподіл арсену та ртуті у вугільному пласті k5 шахти "Капітальна", Донбас / Мінерал. журн. 2021. Вип.43, № 4. С. 73 - 86. <https://doi.org/10.15407/mineraljournal.43.04.073>
22. Ишков В.В. Проблеми геохімії «малих» і токсичних елементів у вугіллі України // Наук. вісник НГА України. -№1. –Дніпропетровськ, НГАУ, 1999. – с. 128 – 132.