

УДК 311.17 : 621.317.38

Л. Л. Полтавець,

кандидат економічних наук,
головний спеціаліст-економіст,
управління статистики у м. Кременчуці
Головного управління статистики у Полтавській області,
poltavetsll@mail.ru

Л. Л. Полтавец,

кандидат экономических наук,
главный специалист-экономист,
управление статистики в г. Кременчуге
Главного управления статистики в Полтавской области

L. L. Poltavets,

PhD (Economics),
Chief Specialist-Economist,
Department of Statistics in the Kremenchuh city,
Chief Department of Statistics in Poltava Region

Використання методів багатомірного групування при статистичному аналізі енергоспоживання на регіональному рівні

У статті обґрунтовано використання методів кластерного аналізу при дослідженні енергоспоживання на регіональному рівні. З використанням агломеративної та дивізімної процедури кластеризації виділено три групи регіонів України за рівнем споживання енергетичних ресурсів. Визначено основні характеристики отриманих кластерів.

Ключові слова: *споживання енергетичних ресурсів, кластерний аналіз, агломеративна та дивізімна кластеризація, ієрархічна дендрограма, регіональний рівень.*

Сьогодні зростає роль статистичного аналізу соціально-економічного розвитку регіонів. Серед основних показників, які аналізуються, важливе місце належить показникам ефективного використання енергетичних ресурсів. Аналіз цих показників повинен стати важливим інформаційним інструментом у вирішенні задач енергоефективності регіональної економіки завдяки наданню інформації, потрібної для вирішення завдань, пов'язаних із більш економічним споживанням наявних енергоресурсів та визначенням основних напрямів підвищення загальної ефективності використання палива та енергії.

Виходячи з вищевикладеного, метою дослідження є аналіз споживання основних видів енергетичних ресурсів (вугілля кам'яного, газу природного та електроенергії) регіонами України з використанням методів кластерного аналізу.

Для досягнення мети дослідження вирішуються такі завдання:

- 1) обґрунтування використання методів багатомірного групування при аналізі енергоспоживання на мезорівні;
- 2) проведення кластеризації обсягів енергоспоживання регіонами України.

Аналізуючи кінцеве споживання енергетичних ресурсів на регіональному рівні, обов'язково слід дотримуватися принципу однорідності: будь-яке оброблення емпіричних даних (усеред-

нення, вимірювання щільності зв'язку, прогнозування тощо) можна здійснювати лише у близьких за атрибутивними і кількісними ознаками групах спостережень.

Як відомо, оцінювання подібності здійснюється на основі однієї чи кількох ознак, які, на думку автора [1], формують "образ класу". У традиційній схемі класифікації ці ознаки ієрархічно впорядковуються за їх вагомістю. На кожному кроці поділу сукупності до уваги береться лише одна ознака, тобто відбувається послідовне формування, покрокове уточнення, більш детальний опис класів [1].

Створення однорідних сукупностей об'єктів базується на використанні методів групувань. Проблема групувань об'єктів за декількома кількісними показниками успішно вирішується за допомогою багатомірних методів аналізу, зокрема кластерного аналізу. Кластер – це група, клас однорідних одиниць сукупності. Основне завдання кластерного аналізу – формування таких груп у багатовимірному просторі [1].

На нашу думку, об'єкти досліджуваної сукупності (24 області України, АР Крим, м. Київ і м. Севастополь) з погляду енергоспоживання є більш-менш однорідними, хоч і відрізняються один від одного за площею, рівнем соціально-економічного розвитку, кількістю населення. Тому для нашого дослідження доцільно використовувати абсолютні показники, що характеризують обсяги споживан-

ня газу, вугілля та електроенергії на регіональному рівні. З метою отримання адекватних результатів дослідження варто використовувати показники з чітким економічним наповненням [1; 2]. Виділені автором статті показники дозволяють вимірювати рівень використання енергетичних ресурсів на регіональному рівні.

Вважаємо, що в контексті цього дослідження доцільно вказати на так званий “порог складності” [3], відповідно до якого адекватний аналіз певної проблемної ситуації можливий у результаті цільового спостереження та одночасного порівняльного оцінювання відносно невеликої кількості параметрів (не більше 7–10). Саме тому необхідність та доцільність застосування кластерного аналізу полягає в його здатності оперувати великою кількістю ознак для порівняння об’єктів [3].

Поставимо перед собою завдання виконати класифікацію регіонів України за обсягами споживання вугілля кам’яного, газу природного та електроенергії. Сенс аналізу полягає у перевірці наступної гіпо-

тези: внаслідок географічних і економічних відмінностей між регіонами обсяги споживання названих ресурсів розподіляються між ними (між регіонами) нерівномірно, внаслідок чого регіони мають створювати однорідні/гомогенні групи за показниками споживання цих ресурсів. Виявлення і аналіз таких груп дозволить детальніше дослідити характер споживання вищезгаданих ресурсів у регіонах і створити більш адекватні моделі для прогнозування їх споживання на короткотермінову перспективу

Як об’єкт аналізу було обрано 27 регіонів України. Розглядаємо три кількісні змінні: обсяг споживання газу природного (млн т у п.), обсяг споживання вугілля кам’яного (тис. т у п.), обсяг споживання електроенергії (млн т у п.) регіонами України (табл. 1).

Оброблення даних і моделювання проведемо в середовищі універсального статистичного пакета Statgraphics Centurion XV.

Виконавши візуалізацію початкових даних у просторі трьох змінних і дослідивши результати

Таблиця 1

Обсяги загального використання основних видів енергетичних ресурсів регіонами України у 2013 році

Регіон	X1, газ природний, тис. т у п.	X2, вугілля кам’яне, тис. т у п.	X3, електроенергія, млн т у п.
АР Крим	1734,40	144,60	631,40
Вінницька	1281,50	2260,40	443,10
Волинська	699,70	40,10	199,30
Дніпропетровська	5718,30	8409,30	7551,90
Донецька	6387,10	23033,60	5459,70
Житомирська	997,90	17,20	414,30
Закарпатська	732,20	14,80	148,80
Запорізька	1767,00	3940,20	2135,70
Івано-Франківська	1814,70	3809,20	367,70
Київська	2875,80	1678,20	839,10
Кіровоградська	728,80	120,20	827,20
Луганська	3400,50	5987,80	2795,30
Львівська	2751,60	933,70	724,60
Миколаївська	1424,50	128,20	539,90
Одеська	2991,30	43,40	790,00
Полтавська	3387,30	21,30	1295,10
Рівненська	1229,30	146,10	898,50
Сумська	1390,50	74,70	424,10
Тернопільська	890,90	12,20	134,50
Харківська	3698,00	2990,10	1314,50
Херсонська	613,90	36,30	318,70
Хмельницька	1084,10	225,10	342,40
Черкаська	3002,80	354,70	650,10
Чернівецька	547,80	19,10	110,10
Чернігівська	1053,30	398,00	268,70
м. Київ	4683,20	359,60	1413,30
м. Севастополь	238,00	4,90	99,70

Джерело: сформовано автором за даними [4–6]

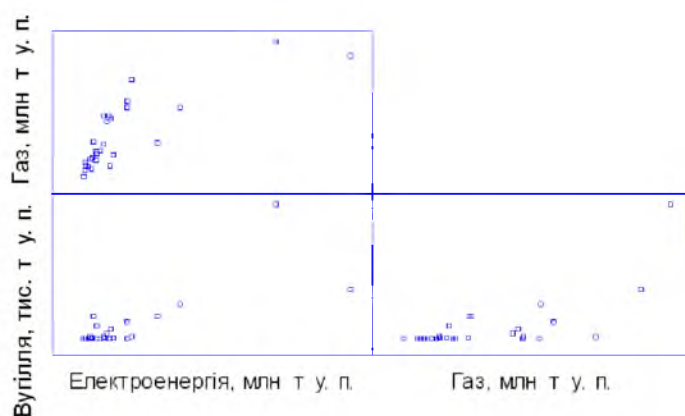


Рис. 1. Двовимірні проєкції простору трьох змінних

на предмет існування природної сегментації на початковому етапі, ми виявили, що, по-перше, спостерігаються два суттєвих викиди щодо великих обсягів споживання для всіх трьох видів ресурсів, які відповідають Донецькій та Дніпропетровській областям; по-друге, має місце наявність як мінімум двох природних сегментів, спричинених явною дискримінацією за двома видами ресурсів: природним газом та електроенергією (рис. 1). Виходячи з цього можна припустити існування двох-трьох сегментів (кластерів, гомогенних груп) у сукупності регіонів за трьома досліджуваними показниками.

Вважається, що у разі наявності природної сегментації (кластеризації) різні методи мають її виявляти [2]. Таким чином, для “чистоти експерименту” автором було обрано два різних за природою, але добре відомих алгоритми: на основі агломеративної ієрархічної кластеризації за методом “найбільш віддаленого сусіда” та на основі дивізивної сегментації за методом *k*-середніх. Метод “найбільш віддаленого сусіда” було обрано тому, що він дає гарні результати саме у випадках, коли

кластери достатньо віддалені один від одного, тобто має місце природна сегментація. Обидва алгоритми показали майже однакові результати.

Зупинимось детальніше на процедурі агломеративної кластеризації.

На першому етапі було виконано стандартизацію початкових даних для приведення їх до одного порядку (матриця *Z*), але рамки цієї статті не дають можливості показати значення цієї матриці. На основі матриці *Z* сформовано матрицю відстаней *D* між об’єктами за манхеттенською метрикою. На основі матриці *D* виконано процедуру ієрархічної кластеризації, результати якої наведено нижче.

На рис. 2 наведено дерево кластеризації (дендрограму) регіонів України за рівнем споживання енергоресурсів.

Як видно з рисунка, припущення щодо наявності трьох основних кластерів виявилось правильним. На це вказують чітко виражені стрибки мінімальної міжкластерної відстані на 22-му (з 4217,289 до 16459,06), на 24-му (з 2,98898 на 4,76806) і на 26-му (з 4,8258 на 11,8818) кроках. Візуальну зміну мінімальної величини міжклас-

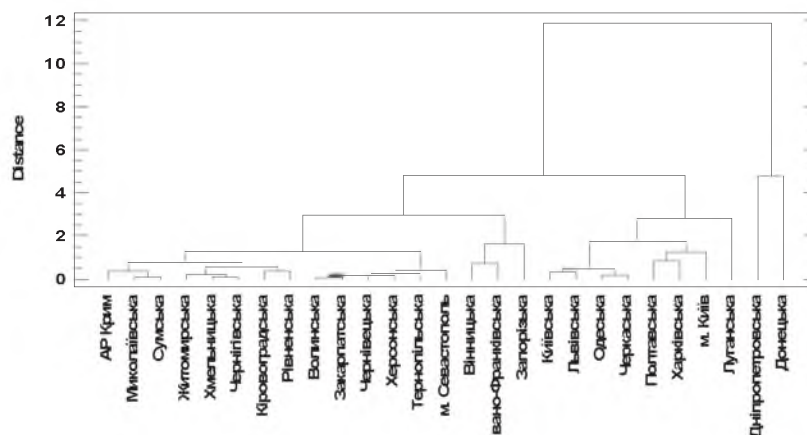


Рис. 2. Ієрархічна дендрограма процесу кластеризації регіонів України за показниками використання енергетичних ресурсів

Джерело: побудовано автором за результатами розрахунків

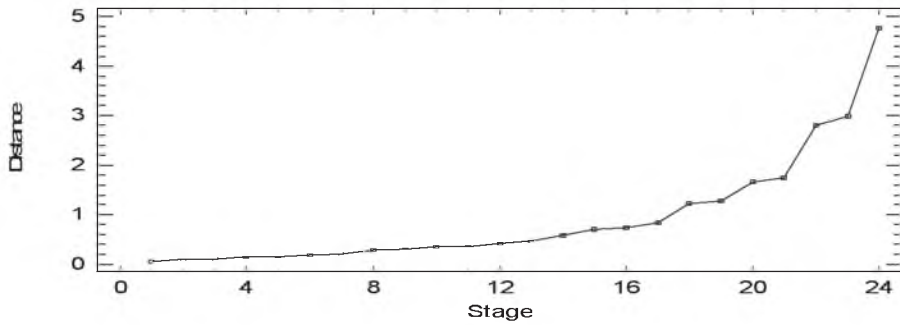


Рис. 3. Графік послідовності агломерації регіонів України за 2013 рік

Джерело: побудовано автором на основі розрахунків

терної відстані можна простежити на графіку послідовності агломерації регіонів (рис. 3).

У результаті проведеного аналізу нами отримано три кластери (табл. 2).

На рис. 4–6 показано двовимірні проєкції кластерів, які дозволяють визначити дискримінантні змінні й охарактеризувати кластери. Виявлено, що найпотужнішими дискримінантними змінними, які суттєво розрізняють регіони і, власне, ділять їх на три сегменти (кластери), є змінні X_1 і X_3 – споживання електроенергії та природного газу.

Таким чином, регіони України у 2013 році, які складають явно виражений кластер № 1, можна характеризувати як “регіони з низьким споживанням енергоресурсів”, кластер № 2 – як “регіони з високим споживанням енергетичних ресурсів”, кластер № 3 – як “регіони із середнім споживанням енергетичних ресурсів” (табл. 3).

На останньому етапі було реалізовано алгоритм k -середніх. На відміну від агломеративних алгоритмів кластерного аналізу, принципова особливість цього методу полягає в тому, що кількість

Таблиця 2

Розподіл регіонів України у 2013 році за кластерами, отриманий агломеративною процедурою

№ кластера	Частка у загальній кількості об'єктів, %	Кількість та перелік об'єктів у кластерах
1	63,0	17 (АР Крим, Вінницька, Волинська, Житомирська, Закарпатська, Запорізька, Івано-Франківська, Кіровоградська, Миколаївська, Рівненська, Сумська, Тернопільська, Херсонська, Хмельницька, Чернівецька, Чернігівська області та м. Севастополь)
2	7,4	2 (Дніпропетровська та Донецька області)
3	29,6	8 (Київська, Луганська, Львівська, Одеська, Полтавська, Харківська, Черкаська області та м. Київ)

Джерело: побудовано автором на основі розрахунків

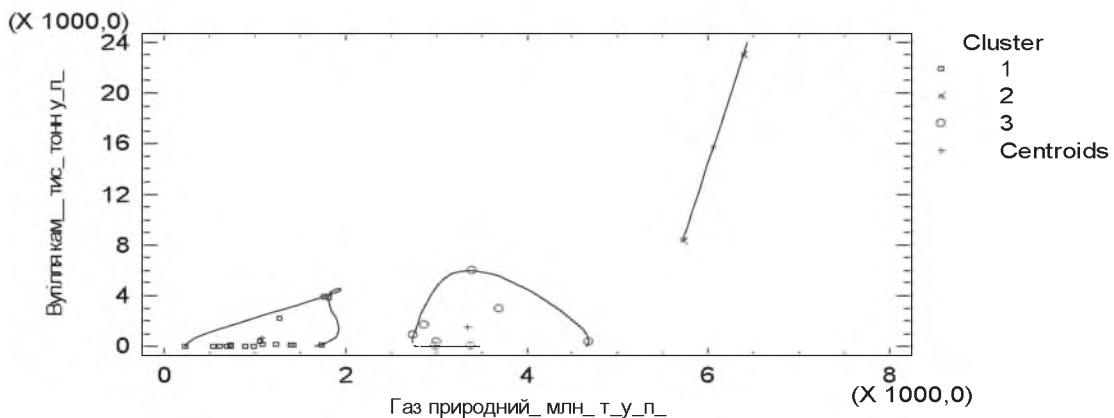


Рис. 4. Двовимірна проєкція кластерів на осі X_2 – X_1

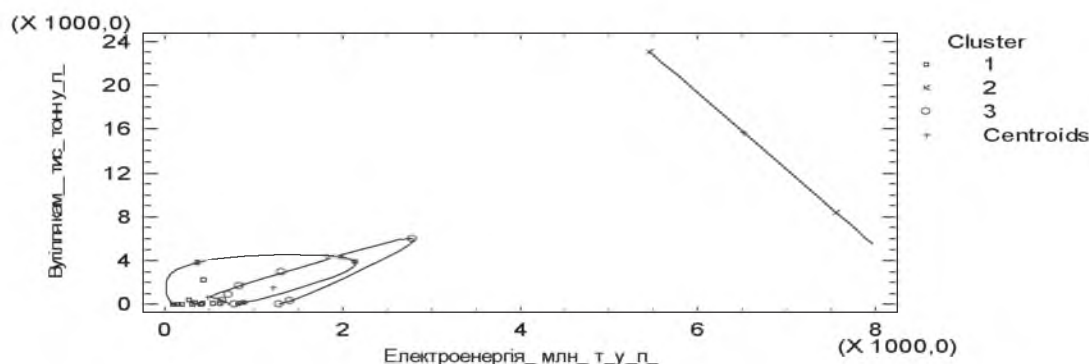


Рис. 5. Двовимірна проекція кластерів $X_2 - X_3$

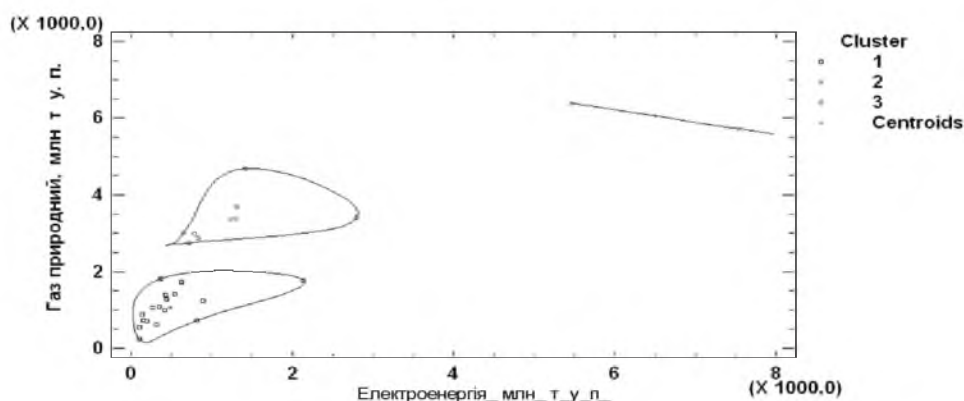


Рис. 6. Двовимірна проекція кластерів $X_1 - X_3$

кластерів має бути задана на початку [2]. Тому кількість груп було прийнято рівною трьом, як у попередньому випадку, з метою порівняння результатів кластеризації різними алгоритмами.

Сутність роботи алгоритму в цьому випадку є такою. Спочатку із 27 точок досліджуваної сукупності автором було задано 3 точки (об'єкти). Ці точки вважаються еталонами – центрами тяжіння майбутніх кластерів на першому кроці процедури. Кожному еталону присвоєно номер L ($L = 1, 2, \dots, k$), який одночасно є і номером кластера. На першому кроці із двох об'єктів, які залишилися ($n - k$), вибирається точка z_i і перевіряється, до якого з еталонів (центрів тяжіння) вона належить.

Досліджуваний об'єкт приєднується до того центру тяжіння, якому відповідає мінімальна евклідова відстань. Еталон замінюється на новий, пере-

рахований з урахуванням приєднаної точки, і його статистична вага (кількість об'єктів, що входить у цей кластер), збільшується на одиницю і т. д.

На рис. 7–9 представлено результати сегментації (у цьому випадку коректним є саме такий термін, оскільки дивізімні алгоритми не будують дерева). Внаслідок такої сегментації було отримано майже ідентичні результати.

Слід звернути увагу на те, що автоматична процедура Statgraphics змінила нумерацію сегментів.

У цьому випадку об'єкт 8 (Запорізька область) потрапив у кластер “середнього споживання ресурсів”. У результаті аналізу значень показників для цього регіону отриману класифікацію слід вважати більш вдалою. Характеристики сегментів в основному схожі з попередніми результатами, наведеними у табл. 3, та представлені у табл. 4.

Таблиця 3

Характеристики кластерів

№ кластера	Кількість регіонів	Частка у загальній кількості об'єктів, %	Координати центру тяжіння кластерів		
			Вугілля кам'яне, тис. т у. п.	Газ природний, тис. т у. п.	Електроенергія, млн т у. п.
1	17	62,96	670,076	1072,26	488,476
2	2	7,41	15721,4	6052,7	6505,8
3	8	29,63	1546,1	3348,81	1227,75

Джерело: побудовано автором на основі розрахунків

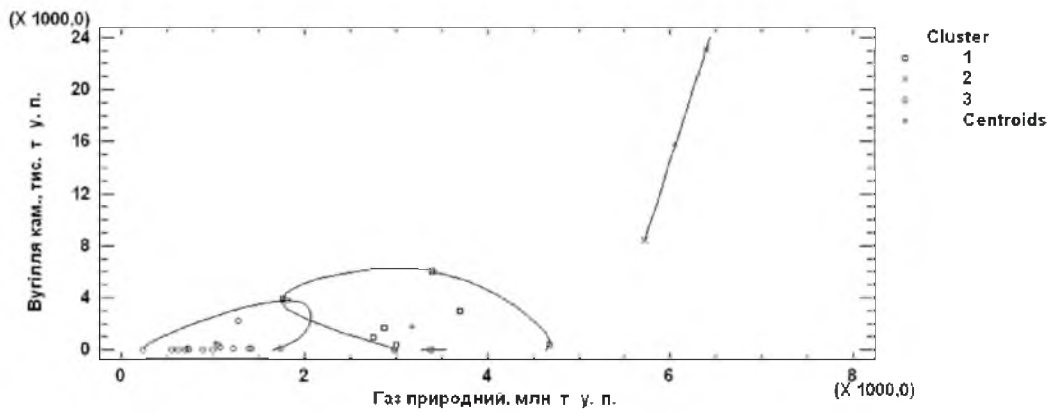


Рис. 7. Двовимірна проекція кластерів $X_2 - X_3$

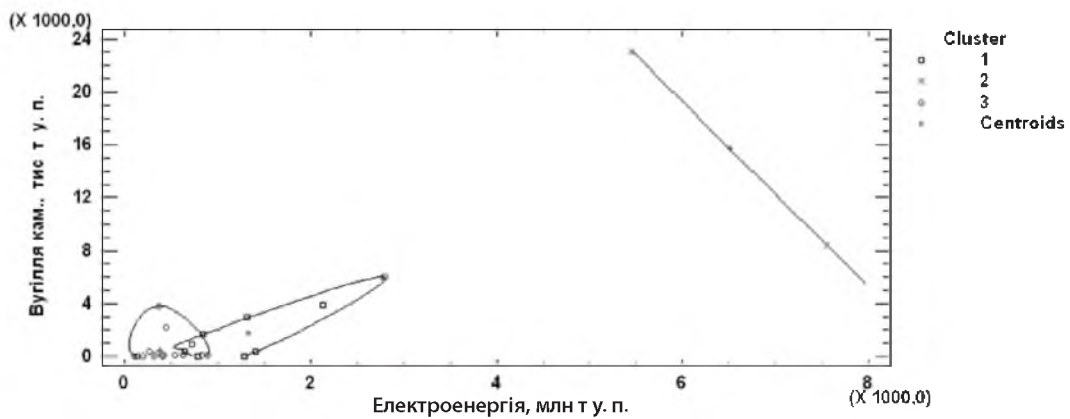


Рис. 8. Двовимірна проекція кластерів $X_2 - X_3$

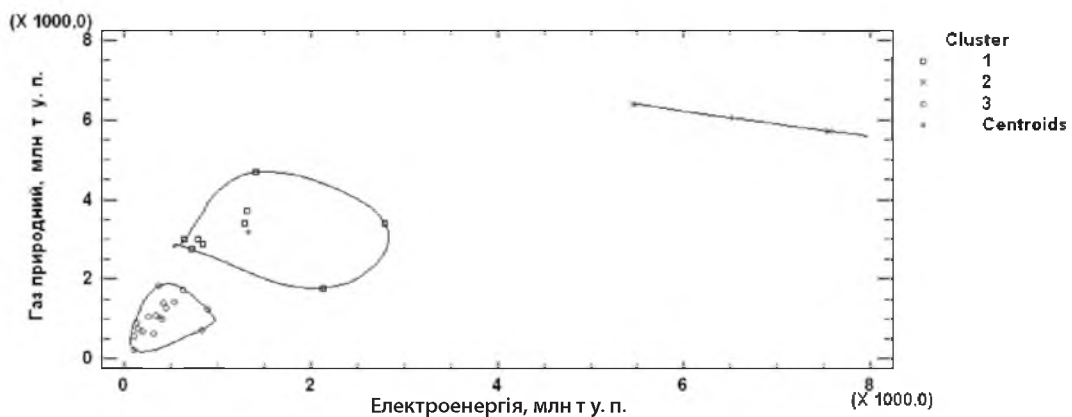


Рис. 9. Двовимірна проекція кластерів $X_1 - X_2$

На основі проведених досліджень можна зробити такі висновки. Регіони України прогнозовано не є однорідними з погляду енергоспоживання. З позиції енерговитрат за трьома видами досліджуваних енергоресурсів можна виділити три основні природні сегменти (кластери), характеристики яких наведено у табл. 5.

Вирішуючи завдання побудови, наприклад, моделей прогнозування і планування енерговитрат за регіонами, необхідно враховувати отриманий результат кластеризації і будувати моделі

для кожного сегменту окремо з метою підвищення точності та достовірності прогнозу. При цьому слід зазначити, що основними дискримінантними факторами для регіонів є споживання електроенергії і газу.

Здійснене багатомірне групування регіонів України свідчить про доволі високу ефективність його застосування при визначенні показників споживання енергетичних ресурсів на регіональному рівні за умов збереження попередніх тенденцій розвитку.

Характеристики сегментів

№ клас-теру	Кількість та перелік об'єктів	Частка у загальній кількості об'єктів, %	Координати центру тяжіння кластерів		
			Вугілля кам'яне, тис. т у. п.	Газ природний, тис. т у. п.	Електроенергія, млн т у. п.
1	9 (Київська, Запорізька, Луганська, Львівська, Одеська, Полтавська, Харківська, Черкаська області та м. Київ)	33,3	1812,1	3173,1	1328,6
2	2 (Дніпропетровська та Донецька області)	7,4	15721,4	6052,7	6505,8
3	16 (АР Крим, Вінницька, Волинська, Житомирська, Закарпатська, Івано-Франківська, Кіровоградська, Миколаївська, Рівненська, Сумська, Тернопільська, Херсонська, Хмельницька, Чернівецька, Чернігівська області та м. Севастополь)	59,3	465,7	1028,8	385,5

Джерело: побудовано автором на основі розрахунків

Розподіл регіонів України у 2013 році за кластерами, отриманий дивізійною процедурою

Кластер	Назва регіону
1. Регіони з середнім споживанням природного газу і електроенергії: • газу природного – приблизно 2000–5000 млн т у. п.; • електроенергії – приблизно 1500–2100 млн т у. п.	Запорізька
	Київська
	Луганська
	Львівська
	Одеська
	Полтавська
	Харківська
	Черкаська м. Київ
2. Регіони з високим споживанням енергетичних ресурсів: • газу природного – приблизно 6000 млн т у. п.; • електроенергії – приблизно 5000–8000 млн т у. п.; • кам'яного вугілля – приблизно 8000–24000 тис. т у. п.	Дніпропетровська
	Донецька
3. Регіони з низьким споживанням енергоресурсів: • газу природного – менше 2000 млн т у. п.; • електроенергії – менше 1000 млн т у. п.; • кам'яного вугілля – менше 4000 тис. т у. п.	АР Крим
	Вінницька
	Волинська
	Житомирська
	Закарпатська
	Івано-Франківська
	Кіровоградська
	Миколаївська
	Рівненська
	Сумська
	Тернопільська
	Херсонська
	Хмельницька
	Чернівецька
Чернігівська м. Севастополь	

Джерело: побудовано автором на основі розрахунків

Слід зазначити, що за теперішніх складних соціально-економічних умов, що склалися в Україні, результати кластеризації регіонів України за ознакою енергоспоживання можуть дещо відрізнятися від запропонованих автором статті. Причиною для цього є кардинальні зміни у територіальному устрої України, які відбулися за останні два роки:

- АР Крим є тимчасово окупованою територією;
- в окремих районах Донецької та Луганської областей проводиться антитерористична операція, що негативно позначається на соціально-економічному розвитку як усього Донбасу, так і України загалом.

Крім того, свої корективи в отриманні результату може внести і адміністративно-територіальна реформа, яка має на меті поступове укрупнення об'єктів територіального устрою.

З урахуванням вищезазначених факторів результати проведеного аналізу споживання енергетичних ресурсів (за відповідного коригування) можуть бути використані в подальшому для впровадження передових програм з енергозбереження у регіонах України з урахуванням специфіки та рівня їх соціально-економічного розвитку

Список використаних джерел

1. Єріна А. М. Статистичне моделювання та прогнозування : [навч. посіб.] / А. М. Єріна. – К. : КНЕУ, 2001. – 170 с.
2. Кендалл М. Многомерный статистический анализ и временные ряды / М. Кендалл, А. Стьюарт ; [пер. с англ. Э. Л. Пресмана, В. И. Ротаря] ; под ред. А. Н. Колмогорова, Ю. В. Прохорова. – М. : Наука, Главная редакция физико-математической литературы, 1976. – 736 с.
3. Шматковська Т. О. Обґрунтування доцільності застосування кластерного аналізу з метою кластеризації регіонів в аспекті соціальних умов функціонування особистих господарств [Електронний ресурс] / Т. О. Шматковська. – Режим доступу : <http://nauka.kushnir.mk.ua/?p=71181>
4. Виробництво електроенергії та окремі техніко-економічні показники роботи електростанцій в Україні за 2013 рік. : [стат. бюлетень] [Електронний ресурс]. – Режим доступу : <http://www.ukrstat.gov.ua> – Назва з титул. екрана.
5. Залишки і використання енергетичних матеріалів та продуктів перероблення нафти за 2011 (2012, 2013) рік : [стат. бюлетень] / відп. за вип. А. О. Фризоренко ; Державна служба статистики України. – К. : ДП “Інформ.-аналіт. агентство”, 2012 (2013, 2014). – 62 с.
6. Паливно-енергетичні ресурси України : [стат. зб.] / відп. за вип. А. О. Фризоренко ; Державна служба статистики України. – К. : ДП “Інформ.-аналіт. агентство”, 2013. – 334 с.