

УДК 628.17:658.26

Н.В. Давиденко

Національний університет водного господарства та природокористування, Україна

ФОРМУВАННЯ ГРУП ОДНОТИПНИХ ДОБОВИХ ГРАФІКІВ ВИТРАТИ ВОДИ З МЕРЕЖІ ВОДОПОСТАЧАННЯ МЕТОДОМ К-СЕРЕДНІХ

Обговорюються питання аналізу фактичного режиму водоспоживання в системах комунального водопостачання як початкового етапу планування енергоефективного режиму водоподачі. Метою дослідження є формування принципів урахування впливу сезонних і соціальних чинників. За допомогою кластерного аналізу виконано класифікацію добових графіків витрати води як характеристики водоспоживання та перевірку правильності кластеризації та надійності виділених кластерів. Побудовані кластери подібних графіків відображають вплив сезонів року і зміни ритму життя населення в робочий та вихідний дні.

Ключові слова: система комунального водопостачання, графік витрати води, моніторинг, кластерний аналіз.

Постановка проблеми

Важливим елементом управління в системі комунального водопостачання (СКВ) є встановлення оптимальних експлуатаційних режимів по кожному її об'єкту. Цільовою функцією оптимізації технологічних режимів насосних станції (НС) є мінімізація енерговитрат за умови безперебійної подачі води споживачу відповідно до режиму водоспоживання та дотримання заданого напору в контрольних точках водопровідної мережі [1]. Основним принципом організації ефективного режиму водопостачання має бути такий: водоподача повинна максимально відповідати водоспоживанню. Недотримання обмежень зумовить погіршення надійності та якості режиму водопостачання, а їх перевищення – додаткові витрати електроенергії: прямі, зумовлені перекачуванням надлишкових об'ємів води, та непрямі, зумовлені зростанням витоків води за рахунок надлишкового тиску в мережі, витрат води на власні потреби. Тобто, забезпечення ефективного електроспоживання вимагає оптимізації режиму електроспоживання об'єктів водопостачання та технологічного процесу водоподачі. Разом з тим, водоспоживання залежить від багатьох факторів: поверховості забудови, ступеня благоустрою будинків, стану встановленого водорозбірного обладнання, числа споживачів, кліматичних умов, пори року. Режими водоспоживання, тобто зміни величин витрати води з мережі водопостачання впродовж доби, місяця, року нерівномірні [2]. На характер водоспоживання впливають: сезон, температура повітря, державні та релігійні свята тощо [3]. Нерівномірність споживання води протягом року зумовлена зміною укладу життя населення; відключенням опалення, гарячого водопостачання тощо. Тобто, на водоспоживання вплива-

ють кліматичні (відображають залежність від погодних явищ, що виражаються в зміні температури повітря, виникненні атмосферних явищ і опадів) та соціальні чинники (відображають залежність від показників, зумовлених соціальним укладом життєдіяльності: типу дня - робочі, вихідні; час доби). Зовнішнє середовище представляє собою сукупність факторів прямої та непрямої дії, які потребують урахування під час планування добової водоподачі, визначають організацію технологічного процесу водопостачання, та впливають на ефективність режиму електроспоживання. Не врахування впливу чинників зовнішнього середовища призводить до зниження ефективності організації режиму водоподачі, а значить споживання електроенергії.

Забезпечення ефективного електроспоживання вимагає як оптимізації самого режиму електроспоживання об'єктів водопостачання, так і оптимізації технологічного процесу водопостачання. При цьому, основним принципом організації ефективного режиму водопостачання має бути такий: водоподача повинна максимально відповідати водоспоживанню. Але водоспоживання – випадковий процес. На нього впливає низка факторів - час доби, зміна укладу життя населення протягом року, у вихідні та святкові дні, сезонні відключення опалення та гарячого водопостачання, зміна температури повітря [3].

Отже, соціальні та кліматичні чинники належать до збурюючих зовнішніх чинників, які визначають режим водоподачі, а отже й ефективність режиму електроспоживання НС. Урахування впливу соціальних та сезонних чинників зовнішнього середовища на характер добового водоспоживання є необхідною умовою реалізації процедур-алгоритмів планування ефективних режимів водоподачі та електроспоживання для кожного з характерних днів.

Аналіз останніх досліджень і публікацій

Основним режимним показником процесу водопостачання є добовий графік витрати води (ГВВ) з мережі водопостачання [4]. Системи автоматизованого управління та системи моніторингу дозволяють накопичити достатній об'єм адекватного статистичного матеріалу та створити базу даних добової витрати води, що забезпечує можливість здійснення аналізу добових ГВВ для виявлення впливу зовнішніх чинників на характер водоспоживання.

Задача пошуку прихованих закономірностей на основі дослідження параметрів добового графіка витрати води може бути розв'язана шляхом формування класів подібних ГВВ з використанням методів класифікації [5]. Зважаючи на відсутність інформації щодо можливих класів доцільним є використання процедур автоматичної класифікації «без учителя», до якої належить кластерний аналіз (КА). Доцільність використання КА для аналізу ГВВ пояснюється тим, що він дозволяє розглядати великий об'єм даних, будувати обґрунтовані класифікації, досліджуючи внутрішні зв'язки між одиницями спостереженої сукупності, та виявляти раніше невідомі закономірності. КА не потребує апріорних припущень про набір даних. Класифікація будується у ознаковому просторі [6] шляхом об'єднання об'єктів в кластери з урахуванням принципу найбільшої подібності в групах і різниці між групами, а її результатом є розбиття об'єктів на групи, що задовольняють деякий критерій [6]. У більшості задач КА немає початкової інформації щодо кількості кластерів. Для ієрархічних методів кількість кластерів не є вхідним параметром та може бути визначена шляхом аналізу дендрограми. Використання ієрархічного КА дозволяє виявити приховані закономірності у ГВВ, зумовлені сезонними та соціальними чинниками [5]. Об'єктами для проведення кластерного аналізу є добові ГВВ. Ознаками, якими описано об'єкти класифікації, є параметри добового ГВВ [4]. Кластером подібних ГВВ є група ГВВ, що мають подібні риси. Утворені кластери є диференційованим відповідно до сезонів року і відображають специфіку водоспоживання в робочий та вихідний дні [5]. Оскільки методи ієрархічного КА не передбачають визначення оптимальної кількості отриманих кластерів, а лише послідовно групують кластери, то алгоритм ієрархічного КА слід розглядати як попередній для визначення числа класів, а одним з актуальних питань постає оцінювання результатів та пошук розбиття, що найкраще відповідає структурі даних, та перевірка обґрунтованості кластерного рішення.

Метою статті є розробка принципів обґрунтованого формування сукупностей однотипних ГВВ, які відображають вплив сезонів року і зміни ритму життя населення в робочий та вихідний дні.

Виклад основного матеріалу

Для визначення належності об'єктів до кластерів використано метод k -середніх, який належить до ітеративних методів еталонного типу, заснований на принципі мінімізації внутрішньокластерної дисперсії, прийнятний для використання у випадку великої кількості спостережень [6] та має найпоширеніше використання серед неієрархічних методів. Класичний алгоритм k -середніх широко використовується для інтелектуального аналізу даних для вирішення задач класифікації даних або кластеризації [7, 8]. Основна ідея алгоритму - розбиття n спостережень на k кластерів. При цьому, необхідно мати гіпотезу про ймовірну кількість кластерів. Тоді, алгоритм кластеризації, заснований на мінімізації внутрішньокластерної дисперсії, дозволить знайти кластери так, щоб вони максимально різнилися один від одного, тобто, розташовувались на максимально можливо великих відстанях один від одного.

Кластеризація методом k -середніх розподіляє вхідний набір векторів, які містять їх ключові характеристики, за k кластерами з кожним з яких пов'язано центроїд \bar{X} . Процес класифікації починається з задання початкових умов: кількості кластерів і їх початкових центрів - центроїдів.

Нехай, $S = \{X\}$, $|S| = n$ - множина вхідних векторів; $D = (X, \bar{X})$ - відстань між вектором X і центроїдом \bar{X} . Для кожного кластеру задають початкове розташування центроїдів \bar{X} - еталонів. Від початкового вибору центроїдів залежить похибка кластеризації та кількість ітерацій [6]. Кожне багатовимірне спостереження сукупності відноситься до того кластеру, центр якого найближче до цього спостереження. Еталон замінюється новим, перерахованим з урахуванням приєднаної точки, вага його (кількість об'єктів, що входять в даний кластер) збільшується на одиницю. Потім виконується перевірка на стійкість класифікації. Нове розбиття порівнюється з попереднім. Якщо вони збігаються - класифікація стійка, процес зупиняється. Інакше відбувається чергова процедура розбиття об'єктів по кластерам. На кожній ітерації об'єкти переміщуються в різні кластери. Алгоритм збігається за скінченну кількість ітерацій. Остаточне розбиття має центри ваги $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_k$, які не збігаються з еталонами. При цьому, кожна точка X_i буде належати до такого кластеру l , для якого $\min_l d(X_i, \bar{X}_l)$.

Для опису добового ГВВ використано параметри, які відображають абсолютні характеристики режиму водоподачі, та класичні показники нерівномірності добових графіків: середнє, максимальне (мінімальне) значення; дисперсію, середньоквадратичне відхилення, коефіцієнти форми, максимуму, за-

повнення та нерівномірності [4]. Для більш детальної оцінки нерівномірності ГВВ застосовано морфометричний підхід [9]. Добовий ГВВ представлено у вигляді діаграми радарного типу (ДРТ) та використано базові (периметр; площа; центр ваги; округлість; компактність; видовження; випуклість) та похідні (радіус вписаного та описаного кіл; довжина, ширина; головна та додаткова вісь видовження; периметр випуклості; площа випуклості; кут осей видовження) морфометричні параметри.

Оскільки одночасне врахування багатьох ознак, що описують ГВВ, ускладнює інтерпретацію отриманих результатів, то групування добових ГВВ виконувалось у два етапи: 1 етап - за впливом сезонності; 2 етап – за впливом соціальних чинників. На першому етапі використано характеристику ГВВ за об'ємами водоподачі та класичні показники нерівномірності. Оскільки, добові ГВВ відмінністю для різних днів тижня, то на другому етапі для врахування форми ГВВ використано морфометричні показники [4]. Відбір інформативних класифікаційних ознак виконано шляхом виключення взаємозалежних ознак на основі кореляційного аналізу. У результаті сформовано набори за впливом: 1) сезонності: x_1 - об'єм добової витрати води з мережі; x_2 – максимальне значення добової витрати води; x_3 – мінімальне значення добової витрати води; x_5 – мінімальне значення витрати води протягом денних годин доби; x_9 – дисперсія за день; 2) соціальних чинників: x_2 - площа ДРТ; x_5 - округлість; x_6 - компактність; x_7 - видовження; x_{10} - кут видовження; x_{11} - випуклість; x_{15} - зміщення координат центра ваги.

Із застосуванням модулю «Cluster Analysis» пакету ПП Statistica 6.0. (StatSoft, Inc., 2001) проведено процедуру класифікації k-means. Кількість кластерів прийнято: для першого етапу 4, для другого - 3 [5]. Формування початкових центрів кластерів виконано таким чином: спочатку упорядковано відстані між усіма об'єктами, а потім як початкові кластерні центри вибрано спостереження на постійних інтервалах [10]. Для формування міри відстані між векторами використано незважену евклідову відстань між вектором та центроїдом. Результатом процедури є визначення належності ГВВ до певного кластеру (рис. 1). Аналіз відстаней об'єктів, що утворюють кластер, до центру кластеру дозволяє визначити, які об'єкти розташовані найближче до центру, тобто є найбільш типовими для свого кластера.

Слід зазначити, що за результатами дисперсійного аналізу змінна x_3 , визначена як інформативна класифікаційна змінна для першого етапу, а також змінні x_6 та x_{15} , визначені як інформативні класифікаційні змінні для другого етапу, мають рівень значущості $p > 0,05$ (їх вплив на класифікацію є незначимим) і виключені з розгляду.

а)

	витрата води рік				КЛАСТЕР	РАССТ.
	x_1	x_2	x_3	x_5		
02.01.пн	0.019135	-0.60229	0.275996	-0.19296	2	0.74
03.01.вт	0.107608	0.180319	0.383961	-0.67673	2	0.40
04.01.ср	0.279476	-0.88581	0.499124	-0.69536	3	1.00
05.01.чт	0.062169	-1.05119	0.873403	-1.66081	2	1.03
06.01.пт	0.920705	1.219865	0.981368	0.550138	4	0.73
07.01.сб	-1.79897	-0.50484	-1.49103	0.485873	4	0.63
08.01.нд	-0.95754	-0.72042	0.344374	-1.0457	2	0.32
09.01.пн	-1.02383	-1.74816	0.146438	-1.3903	2	0.49
10.01.вт	-0.37726	-0.62592	0.686264	-0.99629	2	0.65
11.01.ср	0.113221	0.020843	1.020956	-0.77084	2	0.67
12.01.чт	0.36688	-0.40738	0.898595	-0.95868	2	0.48
13.01.пт	0.182984	-0.22723	-1.68897	1.072322	2	0.98
14.01.сб	-1.85323	-1.92831	-0.29262	-1.52677	2	0.82
15.01.нд	-1.38467	0.026749	0.286793	-0.61527	2	0.98
16.01.пн	-3.64167	-0.8356	-3.42361	2.815434	4	1.16
17.01.вт	-5.26546	-2.52486	-5.6045	1.182994	2	2.31
18.01.ср	-4.95514	-3.45809	-3.70072	-0.42618	2	2.17
19.01.чт	-3.33188	-2.95899	-5.19424	0.639444	3	1.79
20.01.пт	-0.53549	0.339794	-0.86484	1.08988	2	0.70
21.01.сб	-0.74906	-0.49302	-1.08436	-0.135789	2	0.87
22.01.нд	-0.47642	0.10354	-0.52654	0.575226	4	0.50
23.01.пн	-1.00432	-0.20656	0.236499	-0.32332	2	0.64
24.01.вт	-0.66139	0.003123	0.218415	-0.35763	2	0.52
25.01.ср	-1.09172	-1.19885	-0.58413	-0.48175	2	0.59
26.01.чт	-0.76269	-0.76177	0.106851	-0.52052	2	0.24
27.01.пт	-0.52453	0.552429	0.045671	0.238787	2	0.51
28.01.сб	-0.82577	0.310262	-1.9121	0.803182	2	0.74
29.01.нд	-0.98882	-0.85037	0.150037	-0.90253	2	0.21
30.01.пн	-1.04922	0.564242	-0.962	0.88037	2	0.85
31.01.вт	-0.54191	0.162599	0.258002	0.186548	2	0.36

б)

	1	2	3	4	5	КЛАСТЕР	РАССТ.
	x_2	x_5	x_7	x_{10}	x_{11}		
02.01.пн	0.8959	1.0621	0.33506	1.42107	-0.4646	1	0.73
03.01.вт	0.0537	0.2736	-0.2948	0.19766	0.41749	1	0.22
04.01.ср	1.6394	1.3775	-0.4748	0.04936	1.29957	1	0.74
05.01.чт	1.6961	1.0621	0.1551	0.19766	0.41749	1	0.59
06.01.пт	-0.382	1.3775	0.96492	1.01327	0.12346	3	0.76
07.01.сб	-0.864	1.0621	1.41482	-1.0258	1.00554	2	0.85
08.01.нд	0.0843	1.0621	-0.3848	-0.4326	0.71151	2	0.47
09.01.пн	1.8837	0.2736	-0.3848	-0.5067	1.00554	1	5.48
10.01.вт	0.4636	0.589	-0.1148	0.12351	0.71151	1	0.15
11.01.ср	0.4213	-0.042	-0.4748	-0.3584	0.12346	1	0.40
12.01.чт	1.0338	0.7467	-0.4748	-0.0248	0.41749	1	0.37
13.01.пт	0.7247	3.4277	0.69498	0.53131	2.47567	1	1.52
14.01.сб	1.405	-0.673	-0.3848	0.16058	0.12346	2	0.61
15.01.нд	-1.028	-0.988	-0.8347	0.16058	-0.1706	2	0.26
16.01.пн	-2.191	-0.83	-1.8245	-0.0989	-0.1706	3	0.77
17.01.вт	-2.303	-0.357	-1.2846	-0.3955	-1.6407	1	0.83
18.01.ср	-0.496	-0.042	-0.2948	0.71688	0.71151	1	0.49
19.01.чт	1.5095	0.7467	0.51502	0.34595	0.41749	3	0.51
20.01.пт	-0.763	-0.042	-0.5647	-0.4326	0.12346	2	0.55
21.01.сб	0.1421	-0.042	0.87494	1.05034	0.71151	2	0.54
22.01.нд	-0.44	-1.619	-1.2846	0.08644	0.71151	3	0.73
23.01.пн	-0.5	-1.461	-1.0146	0.08644	-0.4646	1	0.34
24.01.вт	-0.585	-0.83	-0.5647	1.19863	0.12346	1	0.52
25.01.ср	0.8024	-1.461	-0.3848	0.53131	-0.1706	1	0.76
26.01.чт	0.4352	0.1159	1.32484	-1.137	-0.4646	1	0.37
27.01.пт	-0.865	-0.515	-0.6547	0.19766	0.12346	1	0.37
28.01.сб	-0.96	0.7467	-0.0249	-0.655	0.71151	3	0.72
29.01.нд	0.364	-0.2	-0.5647	0.49424	0.12346	2	0.38
30.01.пн	-1.465	-1.934	-0.8347	0.56839	-1.0526	1	0.57
31.01.вт	-0.559	-1.146	-0.8347	0.04936	0.12346	1	0.40

Рис. 1 – Визначення належності спостережень до кластерів (фрагмент) для виявлення впливу: а) сезонності; б) соціальних чинників (сезон «зима»)

Після отримання результатів КА слід оцінити, наскільки кластери відрізняються один від одного. Найбільш важливими властивостями кластерів є:

1) щільність розподілу точок всередині кластеру - дає можливість визначити наскільки даний кластер є компактним, або ж розрідженим. Показником, що характеризує щільність «упаковки» багатомірних спостережень в кластері, є дисперсія відстані від центру кластера до окремих точок кластера. Чим менша дисперсія цієї відстані, тим ближче до центру кластера знаходяться спостереження та тим більша щільність кластера; чим більша дисперсія відстані, тим більше розріджений даний кластер;

2) локальність кластерів - характеризує ступінь перекриття і віддаленості кластерів один від одного в багатомірному просторі.

Характеристиками кластера є внутрішня однорідність та зовнішня ізоляваність. Отже, оптималь-

ність розбиття може бути визначена як вимога максимізації щільності кластерів (міри концентрації) або мінімізації середньоквадратичної відстані між центром кластера і всіма його об'єктами (міри внутрішньокласового розсіювання).

Основними характеристиками кластерів є [6]:

– середнє значення кожного фактору:

$$\bar{X}_{kj} = \frac{1}{n_k} \sum_{i=1}^{n_k} X_{kji} \quad (1)$$

– середнє квадратичне відхилення (стандарт) для кожного фактору:

$$\sigma_{kj} = \frac{1}{n_k - 1} \sum_{i=1}^{n_k} X_{kji}^2 - \bar{X}_{kj}^2 \quad (2)$$

де n_k – кількість об'єктів, які увійшли у k -й кластер, j – номер фактору, для якого знаходиться ця характеристика ($1 \leq j \leq m$).

Сукупність середніх дає координати центра кластера у гіперпросторі, а сукупність стандартів – міру розсіювання. Важливо також перевірити статистичну достовірність відмінності середніх для кожного фактору для різних кластерів.

Для перевірки правильності кластеризації методом k -середніх розраховують середні значення ознак для кожного кластера (табл. 1), щоб оцінити, наскільки кластери відрізняються один від одного. В ідеалі отримані середні повинні для всіх вимірювань або хоча б більшій їх частини істотно відрізнятися.

Таблиця 1

Середні значення ознак кластерів				
Виявлення впливу сезонності				
	Кластер 1	Кластер 2	Кластер 3	Кластер 4
x_1	0,408	-0,837	0,691	-1,838
x_2	0,964	-0,969	0,221	-0,543
x_5	-0,411	-0,107	0,761	-2,532
x_9	1,089	-0,678	-0,378	1,2
Виявлення впливу соціальних чинників (сезон «Зима»)				
	Кластер 1	Кластер 2	Кластер 3	
x_2	0,465	-0,797	0,021	
x_5	0,352	-0,801	0,339	
x_7	-0,016	-0,734	1,782	
x_{10}	0,263	0,084	-1,754	
x_{11}	0,528	-0,713	-0,525	

Для більшості ознак середні значення параметрів в кластерах істотно відрізняються. Це свідчить про задовільну якість формування кластерів.

Кількісна однорідність класифікаційних ознак об'єктів, що об'єднані в один кластер, підтверджується результатами аналізу внутрішньогрупової варіації для кожної із ознак (табл. 2).

Результати розрахунку евклідової відстані між кластерами (табл. 3), також підтверджує правомірність проведеної процедури побудови кластерів.

Таблиця 2

Описові статистики та варіація ознак кластерів

Номер кластера	Фактор	Середнє \bar{x}	Стандарт σ	Дисперсія σ^2
Виявлення впливу сезонності				
1	x_1	0,408	0,665	0,442
	x_2	0,964	0,715	0,512
	x_5	-0,41	0,781	0,611
	x_9	1,088	0,725	0,526
2	x_1	-0,836	0,498	0,248
	x_2	-0,969	0,561	0,314
	x_5	-0,107	0,535	0,286
	x_9	-0,678	0,594	0,353
3	x_1	0,691	0,471	0,221
	x_2	0,221	0,515	0,265
	x_5	0,761	0,529	0,281
	x_9	-0,378	0,568	0,323
4	x_1	-1,838	1,558	2,429
	x_2	-0,543	1,291	1,667
	x_5	-2,532	1,239	1,535
	x_9	1,202	0,884	0,782
Виявлення впливу соціальних чинників (Зима)				
1	x_2	0,465	0,702	0,493
	x_5	0,352	0,765	0,585
	x_7	-0,015	0,583	0,341
	x_{10}	0,263	0,611	0,374
	x_{11}	0,528	0,676	0,457
2	x_2	-0,797	0,889	0,791
	x_5	-0,801	0,692	0,479
	x_7	-0,734	0,537	0,288
	x_{10}	0,0843	0,596	0,356
	x_{11}	-0,713	0,829	0,688
3	x_2	0,021	1,043	1,088
	x_5	0,339	0,764	0,583
	x_7	1,782	0,894	0,799
	x_{10}	-1,754	1,274	1,625
	x_{11}	-0,525	0,862	0,744

Таблиця 3

Евклідова відстань між кластерами

Виявлення впливу сезонності				
	Кластер 1	Кластер 2	Кластер 3	Кластер 4
Кластер 1	0,000	1,458	1,019	2,960
Кластер 2	2,126	0,000	1,148	2,649
Кластер 3	1,038	1,071	0,000	5,079
Кластер 4	1,721	1,627	2,253	0,000
Виявлення впливу соціальних чинників («Зима»)				
	Кластер 1	Кластер 2	Кластер 3	
Кластер 1	0,000	0,997	1,722	
Кластер 2	0,998	0,000	2,499	
Кластер 3	1,312	1,581	0,000	

Перевагою методу k-середніх є можливість перевірки статистичної значимості відмінностей між виділеними кластерами, яка полягає у визначенні статистичної достовірності належності дисперсії кожного фактора для різних кластерів до однієї генеральної сукупності. Це досягається за допомогою F-тесту за критерієм Фішера і здійснюється на основі припущення про характер розподілу ознак у генеральній сукупності [10]. Спостереження розглядаються як вибіркові незалежно від того, як вони були отримані. Тоді статистична значимість (істотність) відмінності між виділеними групами тестується в такий спосіб: нульова гіпотеза базується на припущенні, що в генеральній сукупності групові центри ваги рівні між собою, тобто об'єкти груп отримано з однієї генеральної сукупності. Альтернативною гіпотезою виступає протилежне припущення: об'єкти груп отримано з різних генеральних сукупностей.

Дисперсійний аналіз дає змогу визначити наявність значущих відмінностей кількісних характеристик між класами. При цьому, перевіряється гіпотеза про нерівність дисперсій між кластерами і всередині них. Якщо гіпотеза відкидається (приймається, що дисперсії рівні), то отримана класифікація не має сенсу, так як дані є статистично однорідними і не можуть бути розділені на різні групи. Ефективність класифікації на задане число кластерів оцінюється за допомогою дисперсійної таблиці (табл. 4).

Таблиця 4

Дисперсійна таблиця

Виявлення впливу сезонності				
	Міжгрупова дисперсія	Внутрішньогруповая дисперсія	F-критерій Фішера	p-рівень значущості
x_1	221,36	142,09	186,94	0,0000
x_2	213,75	149,29	171,81	0,0000
x_5	208,60	155,39	161,08	0,0000
x_9	215,80	147,52	175,54	0,0000
Виявлення впливу соціальних чинників («Зима»)				
	Міжгрупова дисперсія	Внутрішньогруповая дисперсія	F-критерій Фішера	p-рівень значущості
x_2	118,80	58,76	44,02	0,0000
x_5	98,02	52,67	47,82	0,0000
x_7	108,02	68,14	141,07	0,0000
x_{10}	94,31	85,46	80,64	0,0000
x_{11}	101,37	64,01	56,20	0,0000

Для обґрунтування прийнятності класифікації виконано порівняння мінливості всередині кластерів з мінливістю поза ними. Для перевірки гіпотези про нерівність дисперсій між кластерами та всередині них використано F-критерій Фішера [6, 10]. Чим менша внутрішньогруповая дисперсія та більша міжгруповая, тим краще ознака характеризує належність об'єктів до кластеру і тим більш обґрунтованою є кластеризація. Параметри **F** (значення закону розпо-

ділу Фішера) та **p** (**p-level** - рівень значущості) також характеризують вклад ознаки в розподіл об'єктів на класи. **P**-рівень - ймовірність похибки, якщо стверджувати, що дана ознака впливає на кластеризацію. Кращій кластеризації відповідає більші значення першого та менші значення другого параметру.

Слід зазначити, що дані таблиці 1 щодо середніх по кожній змінній узгоджуються із значеннями F-показників критерію Фішера (табл. 4). Найбільш високими значеннями F-показників критерію Фішера характеризується x_1 – подача (для етапу 1) та x_7 – видовження (для етапу 2), що свідчить про наявність великих відстаней по цих змінних між кластерами. Значення рівня значущості для всіх класифікаційних ознак значно менше ніж 0.05 (навіть для ознак, середні значення яких виявились достатньо близькими) свідчить про значущість впливу цих ознак, оскільки їх середні статистично є відмінними, а також дозволяє прийняти гіпотезу про нерівність дисперсій, тобто, що дисперсії не належать до однієї генеральної сукупності. Таким чином, аналіз наведених в таблиці даних вказує на задовільну якість кластеризації, тобто, розбиття сукупності спостережень на попередньо задану кількість кластерів цілком обґрунтоване та проведено правильно.

Висновки

Використання кластерного аналізу дозволило отримати важливу інформацію щодо особливостей водоспоживання залежно від сезону року і зміни ритму життя населення в робочий та вихідний дні. За допомогою методу k-середніх підтверджено припущення щодо відмінностей у ГВВ за сезоном і типом дня та сформовано групи однотипних ГВВ, які відображають вплив сезонних та соціальних чинників зовнішнього середовища на специфіку витрати води з мережі. Отримані результати вказують на обґрунтованість результатів класифікації, а також однорідність класифікаційних ознак об'єктів, що об'єднані в один кластер, та значимість їх внеску в розподіл об'єктів на групи. Організація системи моніторингу на постійній основі сприятиме накопиченню інформації про витрату води протягом доби, що дозволить виконати уточнення показників добових ГВВ та їх класифікацію.

Література

1. Романчук, С.М. Алгоритмы управления технологическими режимами водоснабжения городов [Текст] / С.М. Романчук // Вісник Донецького національного ун-ту. Серія А: Природничі науки. - 2014. - № 1. - С. 103-110.
2. Євтушенко, А.О. Визначення оптимального складу насосної станції системи комунального водопостачання [Текст] / А.О. Євтушенко, В.Г. Неня, М.І. Сотник, С.О. Хованський // Вісник КДПУ ім. М. Остроградського. – 2008. - Вип. 4/2008 (51), Ч. 1. - С. 158-162.

3. Шушкевич, Е.В. Эффективное управление системой подачи и распределения воды Московского мегаполиса [Текст] / Е.В. Шушкевич // Водоснабжение и санитарная техника. - 2011. - № 1. - С. 24-30.
4. Розен, В.П. Формування множини характеристик фактичного режиму водоспоживання в системах комунального водопостачання [Текст] / В.П. Розен, Н.В. Давиденко // Енергетика: економіка, технології, екологія. - 2015. - № 3 (41). - С. 85-92.
5. Розен, В.П. Процедура побудови базового рівня електроспоживання об'єктів водопостачання з урахуванням впливу зовнішніх чинників [Текст] / В.П. Розен, Л.В. Давиденко, Н.В. Давиденко // Енергетика: економіка, технології, екологія. - 2017. - №3 (49). - С. 31-37.
6. Пістунів, І. М. Кластерний аналіз в економіці [Текст] / І. М. Пістунів, О.П. Антонюк. – Дніпропетровськ: Національний гірничий університет, 2008.– 84 с.
7. Aggarwal, N., Aggarwal, K. (2012) An Improved K-means Clustering Algorithm for Data Mining. *Lambert Academic Publishing*, 72.
8. Hastie, T., Tibshirani, R., Friedman, J. (2003) The Elements of Statistical Learning: Data Mining, Inference, and Prediction. *Springer*, 764.
9. Pavlinov, I. Ya. (2001) Geometric morphometrics, a new analytical approach to comparison of digitized images. *Information technologies in biodiversity research. St. Petersburg*, 41-90.
10. Янковой, А.Г. Многомерный анализ в системе STATISTICA [Текст] / А.Г. Янковой. - Одесса: Оптимум, 2001. Вып.1. – 216 с.
3. Shushkevich, E.V. (2011) Effective management of the water supply and distribution system of the Moscow metropolis. *Water supply and sanitary engineering*, 1, 24-30.
4. Rosen, V.P., Davydenko, N.V. Formation of the set of characteristics of the actual regime of water consumption in municipal water supply systems. *Energy: Economics, Technology, Ecology*, 3 (41), 85-92.
5. Rosen, V.P., Davydenko, L.V., Davydenko, N.V. (2017) Procedure for building the basic level of electricity consumption of water supply facilities taking into account the influence of external factors. *Energy: Economics, Technology, Ecology*, 3 (49), 31-37.
6. Pistunov, I.M., Antonyuk, O.P. (2008) Cluster analysis in economics. *Dnipropetrovsk: National Mining University*, 84.
7. Aggarwal, N., Aggarwal, K. (2012) An Improved K-means Clustering Algorithm for Data Mining. *Lambert Academic Publishing*, 72.
8. Hastie, T., Tibshirani, R., Friedman, J. (2003) The Elements of Statistical Learning: Data Mining, Inference, and Prediction. *Springer*, 764.
9. Pavlinov, I. Ya. (2001) Geometric morphometrics, a new analytical approach to comparison of digitized images. *Information technologies in biodiversity research. St. Petersburg*, 41-90.
10. Yankovoy, A.G. (2011) Multidimensional analysis in the system STATISTICA. *Odessa: Optimum*, 1, 216.

References

1. Romanchuk, S.M. (2014) Algorithms for managing technological regimes of water supply of cities. *Bulletin of the Donetsk National University. Series A: Natural Sciences*, 1, 103-110.
2. Evtushenko, A.O., Nenyа, V.G., Sotnik, M.I., Khovansky, S.O. (2008) Determination of the optimal composition of the pumping station of the municipal water supply system. *Bulletin of the KDPU them. M. Ostrogradsky*, 4/2008 (51), 1, 158-162.

Рецензент: д-р техн. наук, проф. В.П. Розен, Національний технічний університет України «Київський політехнічний інститут імені І. Сікорського», Україна.

Автор: ДАВИДЕНКО Ніна Володимирівна
асистент

Національний університет водного господарства та природокористування, Україна.

E-mail – ninadavydenko1992@gmail.com

ID ORCID: <http://orcid.org/0000-0002-9722-745X>

FORMATION OF THE GROUPS OF TYPICAL DAILY WATER CONSUMPTION SCHEDULES FROM THE WATER SUPPLY NETWORK BY K-MEDIUM METHOD

N.V. Davydenko

National University of Water and Environmental Engineering, Ukraine

The analysis issues of the actual regime of water consumption in public water supply systems as the initial stage of planning the water supply regime have been discussed in the article. The purpose of the research is forming principles of taking into consideration the influence of seasonal and social factors. A daily water consumption graph from the water supply network is used as a characteristic of water consumption. Searches for similar daily water consumption graphs have been performed to detect regularities in water consumption. The absolute characteristics of the water supply regime and the classic indicators of daily graphs unevenness have been used to describe the graph. These signs have been used to take into consideration the influence of seasonal factors. Morphometric indicators have been used to describe the form of the graph. These features have been used to take into consideration the influence of social factors. With the help of cluster analysis, the classification of daily water consumption graphs has been accomplished and groups of similar schedules have been formed. The correctness of the clustering and the reliability of the selected clusters have been checked. To do this, the dispersion analysis procedure have been used. Its results point out to the validity of the results of classification, the homogeneity of the classification features of one cluster objects, the significance of their contribution to the distribution of objects into groups. It has been established that the most significant variable for the season influence is the volume of daily water consumption, and for the influence of social factors is the lengthening. The constructed clusters reflect the influence of the seasons of the year and changes in the life rhythm of the population during the working day and the weekends. The obtained results are the basis for planning the energy efficient mode of water supply depending on the season of the year and the day of the week.

Keywords: public water supply system, water consumption schedule, monitoring, cluster analysis.