УДК 556.166

М.Г. Сербов, *к.г.н*
Одеський державний екологічний університет
С.Г. Кіріяк
Херсонський гідрометеорологічний технікум ОДЕКУ

ЗАСТОСУВАННЯ МЕТОДІВ БАГАТОВИМІРНОГО СТАТИСТИЧНОГО АНАЛІЗУ ДО ГІДРОЛОГІЧНОГО РАЙОНУВАННЯ (НА ПРИКЛАДІ РІЧОК РІВНИННОЇ ТЕРИТОРІЇ УКРАЇНИ)

Застосовані факторний та кластерний аналіз для проведення гідрологічного районування. Використані матеріали спостережень по 190 гідрологічним постам, розташованим на рівнинній території України. Приведені кластерна схема об'єднання гідрологічних об'єктів в однорідні групи та карта-схема гідрологічного районування рівнинної території України.
Ключові слова: *факторний, кластерний багатовимірний аналіз, гідрологічне районування, максимальний стік весняного водопілля*

Для аналізу структури кореляційних матриць в останні роки інтенсивно застосовуються методи факторного аналізу. Факторний аналіз передбачає існування таких зовнішніх причин (факторів), які одночасно впливають на декілька змінних величин і цим обумовлюють наявність видимості кореляційного зв'язку між ними. Задача факторного аналізу полягає у представленні данні спостережень у вигляді лінійних комбінацій факторів і додаткових залишків [2].

В процесі районування досліджуваних територій факторний аналіз може застосовуватись в двох варіантах [3, 4]: *Q*- модифікація, яка дозволяє визначити ступінь взаємної близькості *n* об'єктів шляхом дослідження кореляції між *m* ознаками і *R*- модифікація, яка встановлює зв'язок між *m* ознаками і *p* головними компонентами.

R і *Q* –модифікація аналогічні, але *Q* –модифікація дозволяє виділити групи близьких об'єктів, а *R* – модифікація виділяє групи близьких ознак.

Визначення районів за подібними умовами формування весняного водопілля виконувалось на основі R-модифікації статистичної моделі факторного аналізу [3], яка дозволяє виявити більш значимі фактори з сукупності ознак і використовувати факторні навантаження для подальшого застосуванням дистанційного аналізу [1].

У загальному вигляді факторна модель може бути представлена у вигляді [1]

$$Z_{ij} = \sum_{l=1}^{p} A_{jl} F_{il} + v_j \,, \qquad (1)$$

де Z_{ij} – нормована величина j – ої змінної i –го об'єкту; A_{jl} – факторне навантаження j – го загального фактору на l – ої змінної; F_{il} – значення l – го загального фактору у i – го об'єкта; v_j – залишок змінної, який пов'язаний з похибками вимірювань та зі специфікою тієї чи іншої змінної; p - кількість факторів.

В матричній формі вираз (1) записується як

$$Z=AF+V, \qquad (2)$$

де *Z,A,F,V*- матриця стандартизованих змінних загальних факторів, значень всіх факторів і незалежних залишків.

Рівняння (1) та (2) вказують на основне припущення факторного аналізу, що безліч корельованих рядів спостережень Z_{ij} можна описати меншою кількістю гіпотетичних змінних (факторів) F_{il} та багатьма незалежними залишками v_j.

Вихідною інформацією для визначення невідомих p, A_{jl}, F_{il}, v_i є значення вихідних ознак Z_{ij} (з матриці спостережень x_{ij})

Змінні x_{ij} повинні бути нормовані для приведення їх до однієї розмірності. Відповідно рекомендаціям [1], нормування виконується за формулою

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j}, \tag{3}$$

де $\bar{x}_j$ та σ_j - середнє значення і стандарт *j-oї* змінної.

Вирішення факторної задачі здійснюється на основі матриці кореляційних ознак

$$R = \begin{vmatrix} 1....r12...r13.....r1m \\ r21....1...r23.....r2m \\ r31....r32....1.....r3m \\ ..... \\ rm1....rm2....rm3.....1 \end{vmatrix}, \tag{4}$$

де m-кількість рядів спостережень.

З системи (4) видно, що елементи кореляційної матриці R, факторні навантаження α_{jl}^2 і дисперсії незалежних залишків v_j, пов'язані співвідношенням [3]:

$$r_{jj} = \sum_{l=1}^{p} \alpha_{jl}^2 + v_j \,, \tag{5}$$

$$r_{ij} = \sum_{l=1}^{p} \alpha_{il}^2 + \alpha_{jl} \,, \qquad (j \neq i). \tag{6}$$

Таким чином задача факторного аналізу полягає у лінійному перетворенні кореляційної матриці R в матрицю факторних навантажень та вирішення рівняння узагальнених факторів [3]. Між навантаженнями на фактори, частинний числами і відповідними їм значеннями власних векторів матриці кореляцій розміром m x m існує зв'язок [3]

$$\alpha_{jl} = \alpha_{jl}\sqrt{\lambda_l} \,/ \sqrt{\sum_{l=1}^{k} \alpha_{1l}^2} + \alpha_{2l}^2 + \ldots\ldots \alpha_{ml}^2 \,, \tag{7}$$

де α_{jl} - значення вектора, що відповідає власному числу λ_l, яке показує внесок l-ої компоненти в *j-y* змінну.

Визначення факторних навантажень може бути зведене до вирішення задачі власних векторів і власних значень матриці кореляцій R.

Ефективне число перших факторів p визначається з умови

$$\frac{\sum_{i=1}^{p}\lambda_i}{m}100\rangle 70-80\%, \qquad (8)$$

де λ_i - перші власні числа кореляційної матриці R; m - число змінних .

Більш розповсюдженим в наш час є метод обертання "Варімакс", який було запропоновано Кайзером [3]. Відповідно до цього методу, простота фактора визначається дисперсією квадратів його навантажень. Критерій "Варімакс" Кайзера можна записати у вигляді

$$m\sum_{i=1}^{r}\sum_{j=1}^{m}\left(\frac{b_{jl}}{h_j}\right)^4-\sum_{l=1}^{r}\left[\sum_{j=1}^{m}\frac{b_{jl}}{h_j^2}\right]^2=\max, \qquad (9)$$

де $B=(b_{jl})$ – фінальна факторна матриця, отримана після ортогонального обертання

$$B=AT, \qquad (10)$$

$A=(a_{jl})$ – вихідна ортогональна факторна матриця; T- перетворена ортогональна матриця.

Якщо $B=AT$, то значення узагальненості будь-якої змінної (h_j^2) залишається постійним [3], отже

$$\sum_{l=1}^{r}b_{jl}^2=\sum_{l=1}^{r}a_{jl}^2=h_j^2. \qquad (11)$$

При максимізації функції (11) знайдене положення системи координат задовольняє вимогам ортогональної простої структури.

Розрахунок значень узагальнених факторів виконується за виразом

$$F_{l,i}=\frac{1}{\lambda_l}\sum_{j=1}^{m}a_{jl}\cdot z_{ij}. \qquad (12)$$

Останнім етапом під час проведення факторного аналізу є оцінка значень отриманих факторів. Це досягається за допомогою застосування багаточисельного регресійного аналізу [5]. Таким чином, використання факторного аналізу дозволяє скоротити вихідну інформацію і використовувати факторні навантаження для подальшого групування об’єктів. З цією метою використовується метод кластерного аналізу (дистанційного), який дозволяє об’єднувати в одну групу декілька об’єктів, "близьких" один до одного за комплексними характеристиками [1].

Показником міри схожості служить так званий коефіцієнт близькості (d_{ij}), який представляє відстань між двома точками в m-мірному просторі

$$d_{ij}=\sqrt{\sum_{q=1}^{m}(P_{iq}-P_{jq})^2} \qquad ij=1,2...,n; \quad q=1,2...,m, \qquad (13)$$

де P_{ijq} –матриця ознак об’єкта; ij- номера об’єктів (водозборів); q- номера ознак .

Таким чином, задача факторного аналізу полягає у лінійному перетворенні кореляційної матриці в матрицю факторних навантажень та вирішення рівняння узагальнених факторів [5].

Особливе місце в теорії факторного аналізу займає питання про отримання більш простої структури системи факторних навантажень, що дозволяє об'єктивно інтерпретувати отримані узагальнені фактори.

Відповідно [3], простота факторів визначається дисперсією квадратів їх навантажень. Якщо дисперсія максимальна, то окремі їх навантаження близькі або до нуля, або до одиниці.

Аналіз факторного рішення пов'язаний, як правило, з графічними побудовами. Застосування діаграм річок в площині узагальнених факторів (декартовій системі координат) вже у випадку трьох ефективних факторів викликає суттєві складнощі, так як стає необхідним одночасний аналіз трьох графіків *(f_1,f_2), (f_2,f_3), (f_1,f_3)*, оскільки кожний з графіків описує різний відсоток дисперсії.

Трактування результатів можна суттєво спростити, якщо представити навантаження не в декартовій, а в полярній системі координат, яка дозволяє одночасно аналізувати навантаження всіх факторів.

Перехід від декартової системи координат до полярної виконується за формулами [7]:

$$d_i = \sqrt{a_{1i}^2 + a_{2i}^2} + a_{3i}^2, \tag{14}$$

$$\varphi = \arcsin \frac{a_{3i}}{d_i}, \tag{15}$$

$$\lambda = \arcsin \frac{a_{2i}}{\sqrt{a_{1i}^2 + a_{2i}^2}}, \tag{16}$$

де d_i - одиничний вектор (0≤d≤0), який відображає повноту описання дисперсії *i*-го ряду спостережень в факторах $f_к$; a_i- значення факторних навантажень *к*-го фактору для *i*- го ряду спостережень.

В основу районування рівнинної території України за умовами формування весняного водопілля покладено *R*- модифікацію факторної моделі [3]. Модель будується з припущення про те, що процеси і явища характеризуються великою кількістю визначаючих їх змінних (факторів), які мають різну фізичну природу і просторово-часовий розподіл, з яких повинні бути одібрані основні або значимі фактори.

Для характеристики процесів формування максимального весняного стоку прийняті дані, які визначають місце знаходження та фізико-географічні особливості річкових басейнів (широта φ° центрів тяжіння водозборів), їх морфометричні характеристики (середня висота водозборів *H*), ландшафтні характеристики (заболоченість $f_б$, лісистість $f_л$, розораність $f_р$), а також показники „коефіцієнта дружності весняного водопілля" K_0.

В цілому під час проведення розрахунків використані дані по 190 гідрологічних постах мережі Державної гідрометеорологічної служби України з періодом спостережень за максимальним стоком весняного водопілля не менше 20 років.

Відбір головних, за їх внеском, факторів на процес формування весняної повені здійснювався на основі аналізу матриці парних коефіцієнтів кореляції вищезазначених ознак (табл.1).

Таблиця 1 - Кореляційна матриця ландшафтно-гідрологічних характеристик рівнинної території України

Ознака	Середнє значення	K_0	*H*	*fл*	*fp*	*φ*
K_0	35.2	1,00	0,41	0,42	-0,30	0,38
H	224		1,00	0,54	-0,23	-0,06
fл	10.3			1,00	-0,64	0,16
fp	59.3				1,00	-0,35
φ	49.5					1,00

Апарат факторного аналізу дозволяє стиснути вхідну інформацію до двох узагальнених факторів (табл. 2), які описують 71% сумарної дисперсії кореляційної матриці.

Аналіз факторних навантажень на кожний з предикторів показує, що перший фактор містить у собі, головним чином, інформацію про басейнові чинники стоку: висоту водозборів та лісистість. Внесок першого фактора становить 41% від загальної дисперсії вихідних даних. Другий фактор містить у собі інформацію про зональність (a_φ*=0,922)* і описує 30% вихідної інформації (табл. 2).

Таблиця 2 - Факторні навантаження та внесок сумарної дисперсії

Ознака	1	2
K_0	- 0,550	- 0,474
H	- 0,866	0,191
fл	- 0,842	- 0,252
fp	0,532	0,565
φ	0,043	- 0,922
Внесок сумарної дисперсії, %	41	30

Відповідно [6], практично не існує жодного з методів районування території, який би давав можливість достатньо точно проводити межі відокремлених районів. В нашому досліджені межі районів проводились по вододілах річок, які були віднесені за результатами кластерного аналізу до однієї групи (рис. 1).

За результатами районування рівнинної території України за ландшафтно-гідрологічними характеристиками з урахуванням сумісного застосування факторного та кластерного аналізу виділено 8 гідрологічних районів, які наведені на рис. 2.

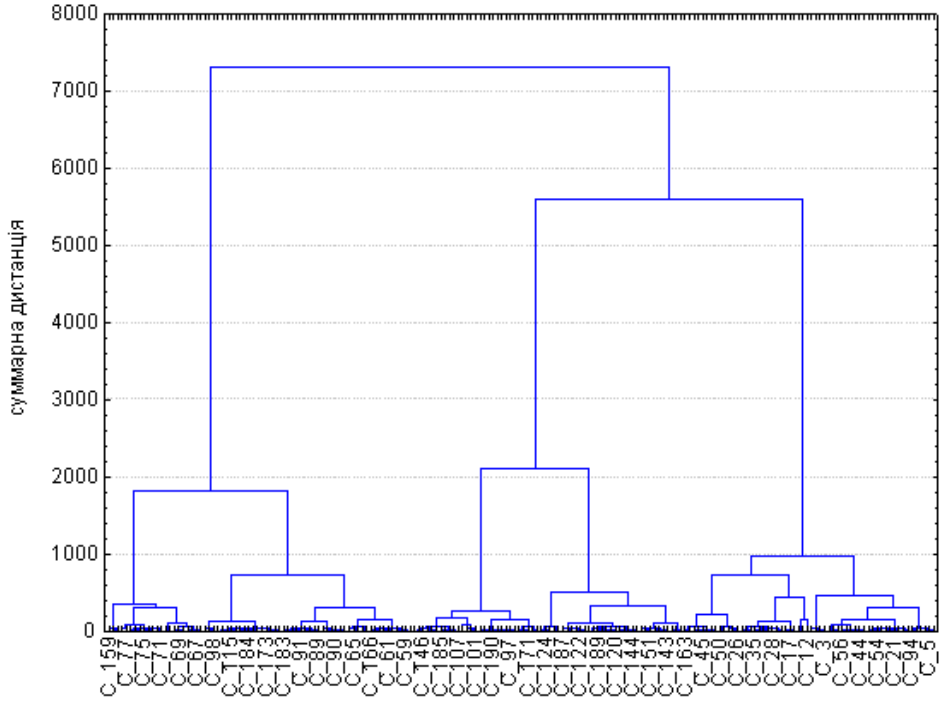

Рис. 1 - Об’єднання гідрологічних об’єктів в однорідні групи.

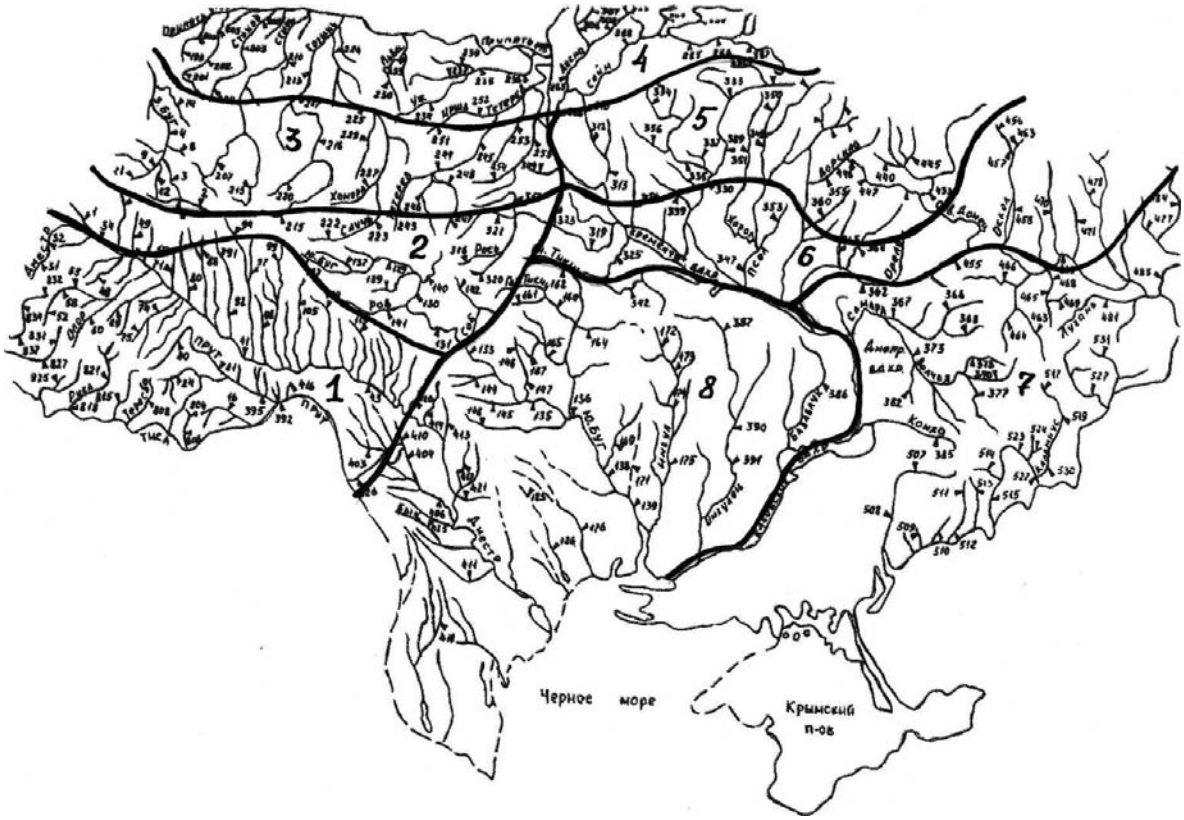

Рис. 2 - Карта-схема районування рівнинної території України.

Список літератури

1. *Архипов Ю.Р., Блажко Н.И.* Математические методы в географии// Из-во Казанского ун-та.-1976.-325с.
2. *Болгов М.В., Лобода Н.С., Николаевич Н.Н.* Пространственное обобщение коэффициентов автокорреляции годового стока Украины//Труды Укр НИГМИ.-1993.-Вып.245.-С.22-29.
3. *Иберла К.* Факторный анализ: Пер.с англ..-М.: Статистика.-1980.-397с.
4. *Комаров В.Д.* Весенний сток равнинных рек Европейской части СССР, условия его формирования и методы прогнозов. -М.: Гидрометеоиздат.-1959.-295с.
5. *Крюков В.Ф.* Исследование территориальной неоднородности максимального дождевого стока на примере Среднерусской возвышенности// Метеорология и гидрология.- 1967.-№5.- С.69-74.
6. *Христофоров А.В.* Надёжность расчетов речного стока. - М.: Изд-во МГУ.-1993.-168с.
7. *Школьный Е.П., Волконская Н.К.* Обобщение факторов температурно-влажностного режима //Межвед.научн.сб. Украины.-Метеорология, климатология и гидрология.-Одесса.-1982.- Вып.18.-С.3-16.

Применение методов многомерного статистического анализа к гидрологическому районированию (на примере рек равнинной территории Украины). Сербов Н.Г., Кирияк С.Г.

Применены факторный и кластерный анализ для проведения гидрологического районирования. Использованы материалы наблюдений по 190 гидрологическим постам, размещенным на равнинной территории Украины. Приведены кластерная схема объединения гидрологических объектов в однородные группы и карта-схема гидрологического районирования равнинной территории Украины.

***Ключевые слова**: факторный, кластерный многомерный анализ, гидрологическое районирование, максимальный сток весеннего половодья*

Use of multidimensional statistic analysis methods at hydrological zoning (by the example of the flat territory in the Ukraine). M.G. Serbov, S.G. Kiriyak

Factor and cluster analyses at conducting hydrological zoning are applied. Monitoring data for 190 gaging stations throughout the flat territory of the Ukraine are used. A cluster chart of unification of hydrological objects into homogeneous groups and a map-chart of hydrological zoning of the flat territory of Ukraine are resulted.

***Keywords:** factor, multidimensional cluster analysis, hydrological zoning, maximum flow of spring tide.*