

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ОДЕСЬКИЙ ДЕРЖАВНИЙ ЕКОЛОГІЧНИЙ УНІВЕРСИТЕТ**

Є.П.Шкільний, Е.М. Серга, Є.А. Галич

**БАГАТОВИМІРНИЙ СТАТИСТИЧНИЙ
АНАЛІЗ ГІДРОМЕТЕОРОЛОГІЧНОЇ
ІНФОРМАЦІЇ**

Навчальний посібник

Одеса – 2015

ББК 26.23
Ш 67
УДК 551.58

Рецензенти:

Лобода Н.С., доктор геогр.наук, професор кафедри гідроекології та водних досліджень, Одеський державний екологічний університет.

Матигін О.С., кандидат фіз.-мат. наук, начальник сектора методичних робіт Гідрометеорологічного центру Чорного та Азовського морів.

Багатовимірний статистичний аналіз гідрометеорологічної інформації. Навчальний посібник/ Шкільний Є.П., Серга Е.М., Галич Є.А. – Одеса , ТЕС, 2015. – 161 с.

Розглядаються методи багатовимірного статистичного аналізу гідрометеорологічних об'єктів. Для аспірантів, спеціалістів та наукових співробітників за напрямками «Науки про Землю» та «Екологію».

ББК 26.23
УДК 551.58
© Одеський державний
екологічний університет, 2015

ПЕРЕДМОВА

При проведенні досліджень у сферах природничих наук, особливо у гідрометеорології, також у практиці гідрометеорологічного обслуговування галузей економіки країни, постійно доводиться звертатися до всебічного вивчення процесів і об'єктів, які розвиваються в цих сферах. Їм притаманна багатовимірність як у просторі, так і в часі. Фізичні величини, що характеризують особливості зазначених багатовимірних процесів та об'єктів, є випадковими величинами. Тому їх властивості в повній мірі описуються багатовимірними законами розподілу ймовірностей. З іншого боку, багатовимірність потребує при побудові методів статистичного аналізу залучати такі поняття як простір у математичному розумінні, множина, поле, вектор тощо. Ці поняття широко використовуються в методах багатовимірного статистичного аналізу.

Інтенсивний розвиток методів цього дуже важливого розділу математики почався з другої половини минулого сторіччя після опублікування визначальних робіт Пірсона, Хотеллінга, Бартлетта, Спірмана, Андерсона. Суттєвий внесок у безпосереднє впровадження методів багатовимірного статистичного аналізу в гідрометеорологічну практику зробили дослідники шкіл М.І. Юдіна, Є.П. Борисєнкова, Н.А. Багрова. У цей час методи багатовимірного статистичного аналізу при дослідженнях статистичної структури випадкових полів, побудові статистичних моделей процесів і явищ, прогностичних моделей, тощо, широко використовується науковцями всіх розвинутих країн.

Однією з основних моделей, на якій засновуються методи багатовимірного статистичного аналізу, модель багатовимірного нормального розподілу або комбінації багатовимірних нормальних розподілів. Хоча існують багатовимірні статистичні проблеми, які не можуть бути описаними моделю нормального розподілу й існують статистичні методи, які застосовують для вибірок інших типів розподілів, наприклад поліноміального розподілу, у навчальному посібнику будуть розглядатися тільки методи, які ґрунтуються саме на моделі нормального закону розподілу.

Як відомо, будь-який розподіл ймовірностей визначається своїми параметрами. Параметрами багатовимірного нормального розподілу є вектор математичних сподівань і матриця коваріацій. Оскільки вектор математичних сподівань багатовимірний, тобто утримує відповідну множину координат, то він належить до багатовимірного простору, який визначається системою аксіом. Матриця коваріацій – другий параметр нормального розподілу – є результатом відповідних дій над багатовимірними випадковими об'єктами в цьому просторі. В якості цих

об'єктів можуть виступати випадкові поля, вертикальні профілі випадкових величин або множина векторів-предикторів. Матриці коваріацій, а також пов'язані з ними матриці кореляцій, несуть в собі великий об'єм інформації про статистичну структуру цих об'єктів. Тому вони є визначальними вхідними операторами всіх без винятку методів багатовимірної статистичного аналізу, які спираються на модель нормального розподілу. Це є підставою того, що саме алгоритмам розрахунків матриць коваріації, обговоренням кореляційних властивостей присвячується *перший розділ* навчального посібника.

У *другому розділі* розв'язується повна проблема власних значень матриць коваріації й послідовно крок за кроком розглядаються алгоритми методу компонентного аналізу, який часто в науковій літературі називають методом розкладу гідрометеорологічних полів по природних (або емпіричних) ортогональних функціях. Цей метод багатовимірної статистичного аналізу широко використовується в сучасних наукових працях, пов'язаних з дослідженням статистичної структури випадкових гідрометеорологічних об'єктів, розробкою статистичних методів прогнозів.

Цілком природно, що за розділом, присвяченим компонентному аналізу, представлений розділ, у якому розглядаються основи факторного аналізу. Це обумовлюється тим, що ці методи формально подібні, але розрізняються по цілях. У тих випадках, коли застосовується метод головних компонент, не потрібно висловлювати ніяких гіпотез відносно змінних. У протилежність методу головних компонент факторний аналіз заздалегідь пояснює матрицю коваріацій наявністю мінімальної кількості гіпотетичних змінних, тобто факторів, причому кожній вихідній змінній визначається її вага у факторі. Факторний аналіз у гідрометеорологічних дослідженнях набуває все більшу й більшу популярність.

При вивченні процесів в атмосфері чи океані часто виникає необхідність провести класифікацію множини багатовимірних векторів ситуацій (образів). Це означає, що за допомогою відповідного алгоритму, чи системи алгоритмів, розробляється вирішувальне правило або система правил, за допомогою яких із вихідної множини образів виділяються сукупності, які мають визначену міру подібності. Привабливість алгоритмів кластерного аналізу полягає у об'єктивності класифікації. Кластерному аналізу відводиться *четвертий розділ* навчального посібника.

При створенні статистичних моделей гідрометеорологічних процесів і, особливо, моделей прогнозу фізичних величин широко використовується багатовимірний лінійний регресійний аналіз. Основні його аспекти, такі як методи розрахунку коефіцієнтів регресії по вибірках предиктанта й предикторів, методи об'єктивного добору статистично значущих

предікторів із вихідної навчальної сукупності, оцінки статистичної значущості отриманих параметрів моделі та її адекватності процесу, для якого будується модель, розглядаються в *п'ятому розділі* посібника.

Регресійний аналіз дає можливість створити прогностичні моделі за допомогою яких із заданою завчасністю прогнозуються значення гідрометеорологічних величин. Але в атмосфері та гідросфері виникають умови, відносно яких необхідно приймати альтернативне рішення щодо можливості чи неможливості їх появи. Розв'язання таких задач здійснюється за допомогою методів розпізнавання образів. З нашої точки зору, найбільш обґрунтованим є метод, який називають дискримінантним аналізом. Саме цим методам присвячується *останній* розділ навчального посібника.

При підготовці цього навчального посібника автори намагалися зберегти необхідну строгість математичного опису методів багатовимірної статистичної аналізи, але зробити ці методи доступними для студентів магістратури й аспірантів і, нарешті, створити ланцюг логічного зв'язку розділів посібника, що сприятиме кращому засвоєнню матеріалів. З цією метою математичне обґрунтування кожного з розділів завершується розгляданням прикладів застосування відповідних методів й переліком запитань для самоперевірки рівня опанування розглянутих методів. У додатках для кожного розділу навчального посібника наводиться перелік задач, які можуть розв'язуватися студентами та аспірантами в терміни практичних занять й самостійної роботи, а також рекомендації щодо організації обчислювальних процедур.

Підготовку навчального посібника здійснював колектив викладачів кафедри фізики атмосфери та кліматології гідрометеорологічного інституту Одеського державного екологічного університету.

Розділ перший за винятком параграфу 1.3, розділ другий, п'ятий та шостий підготовлені професором Школьнім Є.П. Розділ третій за винятком параграфів 3.3 і 3.4 та розділ четвертий доцентом Сергою Е.М. Параграфи 1.3 першого розділу і 3.3.3.4 третього розділу, а також Додатки – доцентом Галич Є.А. Вона також взяла на себе важкий тягар підготовки рукопису до друку. Загальну наукову редакцію рукопису навчального посібника здійснив д.т.н., професор Школьний Є.П.

Автори висловлюють щире подяку професорам Лоевій І.Д. та Лободі Н.С. за надання інтересних матеріалів, які стали основою для типових задач четвертого й шостого розділів.

ЗМІСТ

Передмова	3
1 КОРЕЛЯЦІЙНИЙ АНАЛІЗ ГІДРОМЕТЕОРОЛОГІЧНИХ ПОЛІВ.....	8
1.1 Гідрометеорологічні поля як об'єкт багатовимірного статистичного аналізу.....	8
1.2 Коваріаційна структура випадкових гідрометеорологічних полів.....	12
1.2.1 Матриці коваріацій.....	12
1.2.2 Матриці кореляцій.....	15
1.3 Однорідність та ізотропність гідрометеорологічних полів.....	17
1.4 Приклад дослідження однорідності та ізотропності гідрометеорологічних полів.....	19
2 КОМПОНЕНТНИЙ АНАЛІЗ ГІДРОМЕТЕОРОЛОГІЧНИХ ПОЛІВ....	25
2.1 Компонентний аналіз як метод багатовимірного статистичного аналізу.....	25
2.2 Власні значення та власні вектори.....	27
2.3 Ортогональні компоненти метеорологічних полів.....	31
2.4 Головні компоненти метеорологічних полів. Задача стиску і фільтрації інформації.....	34
2.5 Приклад застосування компонентного аналізу.....	37
3 ФАКТОРНИЙ АНАЛІЗ.....	45
3.1 Основні ідеї факторного аналізу	45
3.2 Рівняння для оцінок факторів.....	47
3.3 Оцінки вагів узагальнених факторів і дисперсій залишків.....	49
3.4 Приклади застосування факторного аналізу гідрометеорологічної інформації	53
4 КЛАСТЕРНИЙ АНАЛІЗ.....	59
4.1 Загальні теоретичні положення кластерного аналізу.....	59
4.2 Універсальний ітераційний метод кластеризації даних (УІМКД).....	65
4.3 Приклад застосування кластерного аналізу.....	70

5	РЕГРЕСІЙНІ МОДЕЛІ МЕТЕОРОЛОГІЧНОГО ПРОГНОЗУ.....	75
5.1	Лінійні моделі множинної регресії.....	75
5.2	Перевірка гіпотези про статистичну значущість коефіцієнтів регресії.....	80
5.3	Множинний коефіцієнт кореляції.....	83
5.4	Методи відбору статистично значущих предикторів.....	87
5.4.1	Поняття про операцію «просіювання» предикторів.....	87
5.4.2	Просіювання предикторів за допомогою частинного коефіцієнта кореляції.....	88
5.4.2.1	Визначення частинного коефіцієнта кореляції....	88
5.4.2.2	«Просіювання» предикторів за методом «включення».....	91
5.4.2.3	«Просіювання» предикторів за методом покрокової регресії.....	94
5.4.3	Відбір статистично значущих предикторів на основі множинного коефіцієнта кореляції.....	96
5.5	Приклад застосування регресійної моделі прогнозу.....	98
6	МЕТОДИ ПРИЙНЯТТЯ АЛЬТЕРНАТИВНИХ РІШЕНЬ ВІДНОСНО СТАНУ ГІДРОМЕТЕОРОЛОГІЧНИХ ОБ'ЄКТІВ.....	102
6.1	Принципи побудови схем альтернативного прогнозу.....	102
6.2	Дискримінантна функція на основі багатовимірного нормального розподілу (дискримінантний аналіз).....	107
6.3	Приклад дискримінантного аналізу	111
	ДОДАТОК А. Завдання до виконання практичних робіт.....	114
	ДОДАТОК Б. Відомості про комп'ютерні програми.....	116
	ДОДАТОК В. Вихідні дані до виконання практичних робіт	121
	ДОДАТОК Г. Карта України.....	157
	ЛІТЕРАТУРА.....	158
	ПРЕДМЕТНИЙ ПОКАЖЧИК.....	159

1 КОРЕЛЯЦІЙНИЙ АНАЛІЗ ГІДРОМЕТЕОРОЛОГІЧНИХ ПОЛІВ

1.1 Гідрометеорологічні поля як об'єкт багатовимірного статистичного аналізу

У гідрометеорологічних дослідженнях, пов'язаних із задачами вивчення динаміки глобального чи регіонального кліматів, розробки статистичних моделей гідрометеорологічних прогнозів тощо, доводиться працювати не тільки з рядами даних метеорологічних чи гідрологічних величин, отриманих у результаті вимірювань або спостережень на мережі станцій та постів, а й з полями гідрометеорологічних величин. Ряди гідрометеорологічних величин є випадковими функціями одного аргументу – часу або відстані між точками вимірювання. Але досить часто ми маємо справу з розгляданням взаємозв'язків між ними в часі й просторі, тому що, як відомо, характеристики стану атмосфери чи гідросфери залежать від фізичних процесів, що розвиваються безперервно в цих сферах й визначаються значною мінливістю. Дуже велика складність цих процесів обумовлює те, що успішна спроба пізнання їх сутті потребує залучення складних математичних методів і, як наслідок, відповідної формалізації.

Як було зазначено, мінливість фізичних величин відбувається у просторі. Тому треба визначити що таке простір й у якому просторі будуть розглядатися гідрометеорологічні об'єкти. У математиці найбільш часто розглядаються такі багатовимірні простори: банахів, гільбертів, евклідів, риманів. Кожний з них визначається своєю аксіоматикою. Методи багатовимірного аналізу розроблено стосовно арифметичного простору.

Арифметичним багатовимірним простором називають множину всіляких n -вимірних систем упорядкованих дійсних чисел $a_1, a_2, \dots, a_n$. Саму систему чисел називають точкою n -вимірного арифметичного простору $A(a_1, a_2, \dots, a_n)$. Якщо визначити точку початку координат $O(0, 0, \dots, 0)$, то точка A може ототожнюватися з n -вимірним вектором $\vec{A}(a_1, a_2, \dots, a_n)$ з координатами – числами, що розташовуються в дужках.

Будемо у подальшому користуватися **метричним арифметичним простором**. **Метричним арифметичним простором** називають такий простір, у якому будь-яким двом точкам $A(a_1, a_2, \dots, a_n)$ та $B(b_1, b_2, \dots, b_n)$ ставиться у відповідність деяке **дійсне число** $\rho(A, B)$, що задовольняє таким вимогам:

- 1) $\rho(A, B) \geq 0$, при $A=B$ $\rho(A, B)=0$;
- 2) $\rho(A, B) = \rho(B, A)$;
- 3) $\rho(A, B) \leq \rho(A, C) + \rho(B, C)$.

(1.1)

Першу властивість називають властивістю невід'ємності, другу - властивістю симетричності, третю - властивістю трикутника. Число $\rho(A, B)$ називається **метрикою n -вимірного арифметичного простору**.

Серед метричних арифметичних просторів ми виберемо такий прості, метрикою якого є **евклідова відстань**

$$\rho(A, B) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + \dots + (b_n - a_n)^2} = \|\overrightarrow{AB}\|. \quad (1.2)$$

Цю **метрику** називають **нормою вектора $\overrightarrow{AB}$** і позначають $\|\overrightarrow{AB}\|$.

Перелічені вище властивості метрики $\rho(A, B)$ називають **аксіоматикою евклідового простору**. Для **евклідового простору** визначається **скалярний добуток** двох векторів:

$$\langle \overrightarrow{AB} \rangle = \overrightarrow{AB} = \sum_{i=1}^n a_i b_i. \quad (1.3)$$

Розглянемо точку в просторі з координатами (x, y, z) у часі t . Значення деякої метеорологічної величини p у цій точці визначається зазначеними координатами, тобто $p(x, y, z, t)$.

Упорядковану сукупність значень гідрометеорологічної чи гідрологічної величини на множині точок багатовимірного простору в деякий момент часу називають **полем цієї величини $p(x, y, z, t)$** . Таке поле є чотиривимірним. Якщо зафіксувати час $t = t_k$, то отримаємо тривимірне поле $p(x, y, z, t_k)$. Зміна гідрометеорологічних величин відбувається у горизонтальному та вертикальному напрямках по-різному. Наприклад, у середньому на рівні моря атмосферний тиск дорівнює 1013 гПа, а на висоті 5 км – 500 гПа у той час, коли на відстані 5 км у горизонтальному напрямку зміна тиску порівняно незначна і може становити лише декілька гПа. Це відноситься також і до інших фізичних величин. Тому, як правило, фіксується не тільки час, але й висота $z = z_l$. Тоді маємо двовимірне поле $p(x, y, z_l, t_k)$ на висоті z_l у час t_k . У подальшому й будуть розглядатися двовимірні поля гідрометеорологічних величин.

Гідрометеорологічні поля можуть бути скалярними й векторними. Наприклад, поле температури повітря, поле атмосферного тиску – поля скалярні, а поля швидкості океанічної течії – векторні.

Є всі підстави розглядати гідрометеорологічні величини як випадкові, незважаючи на те, що їх зміна обумовлюється певними фізичними причинами. Але зв'язки в складних гідрометеорологічних процесах виявляються залежними від багатьох факторів і, як правило, нема достатніх знань про те, який саме зв'язок або зв'язки обумовили зміну цієї величини в той чи інший час. У багатьох випадках не існує повних уявлень про причинно-наслідкові залежності в природних явищах, які спричиняють гідрометеорологічні процеси. Та чи інша міра невизначеності завжди має місце коли йдеться про майбутній стан атмосфери чи об'єктів гідросфери, тобто про гідрометеорологічне прогнозування. Все це дає підставу розглядати поля випадкових гідрометеорологічних величин як **випадкові поля**.

Поля гідрометеорологічних величин, як було зазначено, визначаються сукупністю їх значень на множині упорядкованих точок у багатовимірному просторі. Це означає, що кожна з точок поля має свій номер. Точками поля можуть бути, наприклад, метеорологічні станції чи вузли регулярної сітки точок.

Як відомо з лінійної алгебри, вектори можуть бути представлені упорядкованою системою координат (вектори будемо позначати великими літерами латиниці $A, B, \dots, X, Y, Z, \dots$, а значення координат вектора відповідними малими буквами), тобто **n -вимірний вектор-стовпець** має вигляд:

$$X_j = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \dots \\ x_{ij} \\ \dots \\ x_{nj} \end{pmatrix} \quad (i = \overline{1, n}). \quad (1.4)$$

Ми маємо всі підстави ототожнювати відповідну координату вектора (1.4) значенню метеорологічної величини в момент часу j у тій точці поля, яка має такий же номер, як і номер відповідної координати вектора X_j . Це означає, що двовимірне поле можна зіставити зі значеннями фізичних

величин x_{ij} ($i = \overline{1, n}$; $j = \overline{1, m}$) у j -тий момент часу вектором (1.4), і, оскільки термін спостереження пробігає значення від 1 до m , сукупність m полів означається як m векторів (1.4). Очевидно, розташувавши ці вектори у зазначеному порядку (у часовій послідовності), можна з сукупності m випадкових n -вимірних полів побудувати матрицю порядку $n \times m$ такого вигляду:

$$X = \begin{pmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1j} & \dots & x_{1m} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2j} & \dots & x_{2m} \\ x_{31} & x_{32} & x_{33} & \dots & x_{3j} & \dots & x_{3m} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ x_{i1} & x_{i2} & x_{i3} & \dots & x_{ij} & \dots & x_{im} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ x_{s1} & x_{s2} & x_{s3} & \dots & x_{sj} & \dots & x_{sm} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & x_{n3} & \dots & x_{nj} & \dots & x_{nm} \end{pmatrix}. \quad (1.5)$$

Матриця (1.5) містить великий об'єм вихідної інформації. Індeksi елементів x_{ij} ($i = \overline{1, n}$; $j = \overline{1, m}$) цієї матриці визначають, відповідно, номер метеорологічної станції (вузла сітки) та час, до якого відноситься j -те поле. Такий спосіб представлення сукупності полів є дуже зручним. По-перше, кожен вектор – стовпець матриці X – це *індивідуальне поле* метеорологічної величини, а кожний її рядок – *статистична сукупність* метеорологічної величини у відповідній точці поля (на i -тій станції). По-друге, у матриці X кожний її стовпець (1.4) можна розглядати як вектор-рядок, який є результатом *транспонування* вектора стовпця:

$$X'_j = (x_{j1}, x_{j2}, \dots, x_{ji}, \dots, x_{jn}) \quad i = \overline{1, n}, \quad (1.6)$$

де (') – позначення операції транспонування.

1.2 Коваріаційна структура випадкових гідрометеорологічних полів

1.2.1 Матриці коваріацій

Матриці коваріацій та кореляцій являють собою вихідні, найбільш важливі параметри, які визначають особливості статистичної структури гідрометеорологічних полів. Майже всі методи багатовимірної статистичного аналізу базуються на матрицях коваріацій та кореляцій. Будемо позначати матриці коваріацій K_x , а матриці кореляції R_x .

Основою для визначення матриць коваріацій є інформація про значення метеорологічних величин X у n точках поля, отриманих у результаті вимірювань протягом m інтервалів часу, які сконцентровані в матриці (1.5). Побудуємо матрицю (1.5) за допомогою оператора центрування. Для цього, на основі кожного рядка матриці (1.5) знайдемо середнє значення метеорологічної величини в кожній точці поля:

$$\bar{x}_i = \frac{1}{m} \sum_{j=1}^m x_{ij}, \quad (i = \overline{1, n}) \quad (1.7)$$

і побудуємо вектор середніх значень:

$$\bar{X} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_i \\ \vdots \\ \bar{x}_n \end{pmatrix}, \quad (i = \overline{1, n}). \quad (1.8)$$

Якщо кожне значення $\bar{x}_i$ нанести на карту у відповідний вузол регулярної сітки точок (чи метеорологічних станцій) і провести систему ізоліній, то отримаємо *середнє поле* цієї фізичної величини.

Операція *центрування* полягає у тому, що від кожного елемента рядка матриці (1.5) віднімається відповідне цьому рядку середнє значення елемента вектора (1.8). Отримаємо:

$$\Delta x_{ij} = x_{ij} - \bar{x}_i \quad (i = \overline{1, n}; j = \overline{1, m}) \quad (1.9)$$

і, відповідно, матрицю центрованих полів ΔX :

$$\Delta X = \begin{pmatrix} \Delta x_{11} & \Delta x_{12} & \Delta x_{13} & \dots & \Delta x_{1j} & \dots & \Delta x_{1m} \\ \Delta x_{21} & \Delta x_{22} & \Delta x_{23} & \dots & \Delta x_{2j} & \dots & \Delta x_{2m} \\ \Delta x_{31} & \Delta x_{32} & \Delta x_{33} & \dots & \Delta x_{3j} & \dots & \Delta x_{3m} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \Delta x_{i1} & \Delta x_{i2} & \Delta x_{i3} & \dots & \Delta x_{ij} & \dots & \Delta x_{im} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \Delta x_{s1} & \Delta x_{s2} & \Delta x_{s3} & \dots & \Delta x_{sj} & \dots & \Delta x_{sm} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \Delta x_{n1} & \Delta x_{n2} & \Delta x_{n3} & \dots & \Delta x_{nj} & \dots & \Delta x_{nm} \end{pmatrix}. \quad (1.10)$$

Матриця коваріації визначається за допомогою матричної операції:

$$K_x = \frac{1}{m} \Delta X \Delta X', \quad (1.11)$$

де (') – операція транспонування матриці.

Як відомо, добуток двох матриць можливий, коли кількість рядків першої матриці дорівнює кількості стовпців другої. У матриці ΔX є n рядків і m стовпців, а у матриці $\Delta X'$ - n стовпців і m рядків. Кожний з елементів матриці K_x є скалярним добутком вектора-рядка матриці ΔX на вектор – стовпець матриці $\Delta X'$.

Покажемо цю операцію на матриці ΔX меншого порядку $2 \times m$:

$$K_X = \frac{1}{m} \begin{pmatrix} \Delta x_{11} & \Delta x_{12} & \dots & \Delta x_{1j} & \dots & \Delta x_{1m} \\ \Delta x_{21} & \Delta x_{22} & \dots & \Delta x_{2j} & \dots & \Delta x_{2m} \end{pmatrix} \begin{pmatrix} \Delta x_{11} & \Delta x_{21} \\ \Delta x_{12} & \Delta x_{22} \\ \dots & \dots \\ \Delta x_{1j} & \Delta x_{2j} \\ \dots & \dots \\ \Delta x_{1m} & \Delta x_{2m} \end{pmatrix}. \quad (1.12)$$

Використовуючи для скалярного добутку рівняння (1.3) і рівняння (1.9), що визначає операцію центрування, маємо:

$$K_x = \begin{pmatrix} \frac{1}{m} \sum_{j=1}^m (x_{1j} - \bar{x}_1)^2 & \frac{1}{m} \sum_{j=1}^m (x_{1j} - \bar{x}_1)(x_{2j} - \bar{x}_2) \\ \frac{1}{m} \sum_{j=1}^m (x_{2j} - \bar{x}_2)(x_{1j} - \bar{x}_1) & \frac{1}{m} \sum_{j=1}^m (x_{2j} - \bar{x}_2)^2 \end{pmatrix}. \quad (1.13)$$

Визначимо сенс кожного з елементів матриці (1.13). Ясно, що елементи головної діагоналі є **дисперсії** випадкових величин x_1 і x_2 , а елементи побічної діагоналі – коваріації між цими випадковими величинами в точках поля x_1 і x_2 . Отже:

$$K_x = \begin{pmatrix} \sigma_{x_1}^2 & K_{x_1 x_2} \\ K_{x_2 x_1} & \sigma_{x_2}^2 \end{pmatrix}. \quad (1.14)$$

Аналогічним чином розраховуються елементи матриці розміром $n \times m$:

$$K_x = \begin{pmatrix} \sigma_1^2 & K_{12} & K_{13} & \dots & K_{1s} & \dots & K_{1n} \\ K_{21} & \sigma_2^2 & K_{23} & \dots & K_{2s} & \dots & K_{2n} \\ K_{31} & K_{32} & \sigma_3^2 & \dots & K_{3s} & \dots & K_{3n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ K_{i1} & K_{i2} & K_{i3} & \dots & K_{is} & \dots & K_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ K_{n1} & K_{n2} & K_{n3} & \dots & K_{ns} & \dots & \sigma_n^2 \end{pmatrix}. \quad (1.15)$$

З матриці (1.13) можна зробити висновок про властивості матриці коваріації K_x :

- вона симетрична $K_{x_i x_j} = K_{x_j x_i}$;
- на її головній діагоналі розташовуються дисперсії випадкової величини x ;
- її елементи є дійсними числами;

- матриця коваріацій є додатньо визначеною (детермінант матриці – додатне число $|K_x| > 0$).

Якщо елементи матриці (1.13), які розташовані на головній діагоналі, порівняти з рівнянням (1.2), можна зробити висновок, що дисперсії є зменшеними в m разів нормами векторів, які складаються з центрованих елементів. Це стосується і елементів матриці (1.15).

До речі, дуже важливим є те, що матриця коваріацій, поряд з вектором математичних сподівань, є параметром багатовимірного нормального розподілу ймовірностей:

$$p(X) = \frac{1}{(2\pi)^{\frac{n}{2}} |K_x|^{\frac{1}{2}}} \times \exp \left[-\frac{1}{2} (X - m_x)' K_x^{-1} (X - m_x) \right] \quad (1.16)$$

де індекс (-1) позначає операцію обернення матриці K_x , а $|K_x|$ - визначник матриці K_x . Статистичною оцінкою вектора *математичних сподівань* m_x є вектор середніх значень (1.8).

1.2.2 Матриці кореляцій

По елементах матриці коваріації (1.15), розташованих на головній діагоналі ($\sigma_{x_i} = \sqrt{\sigma_{x_i}^2}$), можна побудувати діагональну матрицю середніх квадратичних відхилів σ . Вона має такий вигляд:

$$\sigma = \begin{pmatrix} \sigma_{x_1} & 0 & \dots & 0 \\ 0 & \sigma_{x_2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_{x_n} \end{pmatrix}. \quad (1.17)$$

Кожний елемент матриці (1.17) має сенс *середнього квадратичного відхилення* гідрометеорологічної величини в точці поля з відповідним номером. Якщо нанести значення середніх квадратичних відхилів на карту й провести відповідні ізолінії, отримаємо поле характеристик мінливості цієї фізичної величини.

Обернена матриця середніх квадратичних відхилів σ^{-1} має вигляд:

$$\sigma^{-1} = \begin{pmatrix} \frac{1}{\sigma_{x_1}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_{x_2}} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{1}{\sigma_{x_n}} \end{pmatrix}. \quad (1.18)$$

Розглянемо тепер матричне рівняння, яке визначає матрицю кореляції:

$$R_x = \sigma^{-1} K_x \sigma^{-1}. \quad (1.19)$$

Структуру матриці R_x просто визначити, виконуючи операції множення на матрицях меншого порядку 2×2 :

$$\begin{aligned} R_x &= \begin{pmatrix} \frac{1}{\sigma_{x_1}} & 0 \\ 0 & \frac{1}{\sigma_{x_2}} \end{pmatrix} \begin{pmatrix} \sigma_{x_1}^2 & K_{x_1 x_2} \\ K_{x_2 x_1} & \sigma_{x_2}^2 \end{pmatrix} \begin{pmatrix} \frac{1}{\sigma_{x_1}} & 0 \\ 0 & \frac{1}{\sigma_{x_2}} \end{pmatrix} = \\ &= \begin{pmatrix} \frac{\sigma_{x_1}^2}{\sigma_{x_1}^2} & \frac{K_{x_1 x_2}}{\sigma_{x_1} \sigma_{x_2}} \\ \frac{K_{x_2 x_1}}{\sigma_{x_2} \sigma_{x_1}} & \frac{\sigma_{x_2}^2}{\sigma_{x_2}^2} \end{pmatrix} = \begin{pmatrix} 1 & r_{x_1 x_2} \\ r_{x_2 x_1} & 1 \end{pmatrix}, \end{aligned} \quad (1.20)$$

оскільки, як відомо

$$\frac{K_{x_i x_j}}{\sigma_{x_i} \sigma_{x_j}} = r_{x_i x_j}, \quad (1.21)$$

є коефіцієнт кореляції.

Отже, у результаті множення матриць в правій частині рівняння (1.19) ми отримаємо n -вимірну квадратну матрицю, яка називається **матрицею кореляції**:

$$R_x = \begin{pmatrix} 1 & r_{12} & r_{13} & \dots & r_{1s} & \dots & r_{1n} \\ r_{21} & 1 & r_{23} & \dots & r_{2s} & \dots & r_{2n} \\ r_{31} & r_{32} & 1 & \dots & r_{3s} & \dots & r_{3n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ r_{i1} & r_{i2} & r_{i3} & \dots & r_{is} & \dots & r_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ r_{n1} & r_{n2} & r_{n3} & \dots & r_{ns} & \dots & 1 \end{pmatrix}. \quad (1.22)$$

Як і матриця K_x , матриця кореляції є **дійсною, симетричною** $r_{x_i x_j} = r_{x_j x_i}$ ($i, j = \overline{1, n}$) і **додатньо визначеною**, а на **головній діагоналі розташовані одиниці**. Ці значення відповідають кореляційному зв'язку в означеній точці поля фізичної величини самої з собою, тому цей елемент головної діагоналі матриці кореляцій відноситься до точки поля, яку називають **полюсом** або **центром кореляції**. Інші елементи характеризують кореляційні зв'язки фізичної величини в різних точках поля. У зв'язку з симетричністю матриці кореляцій елементи її рядка і стовпця, що мають однаковий номер, дорівнюють одне одному. Отже, у кореляційній матриці порядку $n \times n$ утримується інформація про n полів кореляцій.

1.3 Однорідність та ізотропність гідрометеорологічних полів

При вивченні випадкових процесів дуже важливими є властивості стаціонарності. Стаціонарність значно спрощує опис випадкового процесу. Для просторових полів подібними є умови їх однорідності та ізотропності.

Випадкове поле називається однорідним, якщо всі n -вимірні закони розподілення не зміняться при зміні точок $\rho_1, \rho_2, \dots, \rho_n$ на той же самий вектор:

$$\begin{aligned} f_1(x; \rho_1) &= f_1(x; \rho_1 + \rho_0), \\ f_2(x_1; x_2; \rho_1; \rho_2) &= f_2(x_1; x_2; \rho_1 + \rho_0; \rho_2 + \rho_0). \end{aligned} \quad (1.23)$$

Якщо прийняти $\rho_0 = -\rho_1$, то будемо мати

$$\begin{aligned} f_1(x_1; \rho_1) &= f_1(x; 0) = f_1(x_1), \\ f_2(x_1; x_2; \rho_1; \rho_2) &= f_2(x_1; x_2; \rho_2 - \rho_0). \end{aligned} \quad (1.24)$$

Але, як відомо,

$$m_x(\rho) = \int_{-\infty}^{\infty} x f_1(x_1, \rho) dx = \int_{-\infty}^{\infty} x f_1(x) dx = m_x. \quad (1.25)$$

$$\begin{aligned} R_x(\rho_1, \rho_2) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} [x_1 - m_x(\rho_1)][x_2 - m_x(\rho_2)] f(x_1; x_2; \rho_1; \rho_2) dx_1 dx_2 = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (U_1 - m_x)(U_2 - m_x) f_2(x_1; x_2; \rho_2 - \rho_1) dx_1 dx_2 = \\ &= R_x(\rho_2 - \rho_1) = R_x(l), \end{aligned} \quad (1.26)$$

$$l = \rho_2 - \rho_1.$$

Це означає, що для однорідного випадкового поля математичне сподівання не залежить від координат точок поля, тобто є сталою величиною, а кореляційна функція $R_x(\rho_2 - \rho_1) = R_x(l)$ залежить тільки від різниці векторів $l = \rho_2 - \rho_1$.

Таким чином, **поле називають однорідним**, для якого математичне сподівання є сталим, а кореляційна функція залежить тільки від різниці векторів: відстаней точок поля від полюса кореляції. **Однорідні поля називають ізотропними**, якщо просторово кореляційна функція $K(l)$ не залежить від напрямків перетинів полів кореляції.

Гідрометеорологічні поля, як правило, не мають властивостей однорідності та ізотропності але часто цими властивостями володіють поля, якщо провести операції центрування й нормування випадкових величин, тобто розглядати поля величин:

$$\varphi_{ij} = \frac{x_{ij} - \bar{x}_i}{\sigma_{x_i}}. \quad (1.27)$$

Треба відмітити, що $m_\varphi = 0$, тобто є нульовим вектором, а матриця коваріацій $K_\varphi(\rho_1, \rho_2) = R_x(\rho_1, \rho_2)$ дорівнює матриці кореляцій вихідної змінної. Звідси висновок, що всі елементи діагональної матриці σ_φ дорівнюють одиниці.

Тепер, щоб зробити висновок про однорідність та ізотропність полів φ досить впевнитися у тому, що поля кореляцій, які містяться в матриці (1.22) для всіх полюсів кореляції мають подібний вигляд і визначаються системою майже концентричних ізокорелят. Також враховують ще й довірчі інтервали для просторових кореляційних функцій.

Якщо зробити перетини полів кореляцій у різних напрямках, це приведе до отримання просторових кореляційних функцій. І якщо всі кореляційні функції розташуються у достатньо вузькій смузі, то це є підставою вважати, що структура полів кореляцій не залежить від напрямку. Статистична структура метеорологічних полів буде характеризуватися просторовою кореляційною функцією, яка знаходиться шляхом осереднення кореляційних функцій для кількох перетинів. Для неї необхідно побудувати довірчі інтервали, щоб удосконалитися у вірогідності отриманих результатів.

1.4 Приклад дослідження однорідності та ізотропності гідрометеорологічних полів

Властивості однорідності та ізотропності випадкових полів є дуже важливими. Якщо ці властивості притаманні метеорологічним полям, то дуже спрощуються процедури їх статистичного аналізу.

Дослідження властивостей однорідності та ізотропності метеорологічних полів проводилось на основі часових рядів місячних сум опадів у грудні, отриманих на мережі метеорологічних станцій України (Додаток Г).

У результаті реалізації комп'ютерної програми "KOR.EXE (Додаток Б) було отримано:

- матрицю кореляцій;
- вектор середніх значень;
- вектор середніх квадратичних відхилень місячних сум опадів.

На рис.1.1, 1.2, наводяться поля кореляції місячних кількостей опадів на території України в грудні для полюсів кореляції у Дніпропетровську та Чернігові. Видно, що поля кореляції в обох випадках мають подібний вигляд. Отже, оскільки для центрованих і нормованих на середній квадратичний відхил місячних кількостей опадів, які визначаються рівністю (1.27), середні значення дорівнюють нулю, а дисперсії – одиниці для всіх точок поля, а поля кореляції не змінюються при змінюванні полюса кореляцій, то поля цих метеорологічних величин можна вважати квазіоднорідними.

Якщо зробити перетини полів кореляцій в різних напрямках (рис. 1.1 і 1.2), це приведе до отримання просторових кореляційних функцій. Кореляційні функції для зазначених напрямків перетинів містяться в табл.1.1.

Якщо всі кореляційні функції розташуються в достатньо вузькій смужі (рис. 1.3), то це є підставою вважати, що структура полів кореляцій не залежить від напрямку перетинів. Статистична структура полів опадів буде характеризуватися просторовою кореляційною функцією, яка знаходиться шляхом осереднення кореляційних функцій для кількох перетинів. Із рис.1.3 випливає, що з достатньою точністю осереднену кореляційну функцію можна апроксимувати співвідношенням:

$$r(l) = 1 - \frac{l}{l_0}, \quad (1.28)$$

де $l_0 = 13$ одиниць масштабу, а l - поточне значення одиниць масштабу.

Для неї можна побудувати довірчі інтервали, щоб удосконалитися у вірогідності отриманих результатів.

Таким чином, нормовані й центровані поля місячних сум опадів у грудні є не тільки квазіоднорідними, але й квазіізотропними. Їх статистична структура буде характеризуватися просторовою кореляційною функцією, яка знаходиться шляхом осереднення кореляційних функцій для декількох перетинів.

Таблиця 1.1 – Просторові кореляційні функції місячних кількостей опадів у грудні.

Номери одиниць масштабу (<i>l</i>)	Полюси						
	Чернігів			Дніпропетровськ			
	Номери напрямків перетинів полів						$\bar{r}(l)$
	1	2	3	1	2	3	
1	0,95	0,93	0,94	0,89	0,87	0,93	0,92
2	0,92	0,87	0,89	0,78	0,78	0,89	0,86
3	0,82	0,84	0,81	0,69	0,68	0,87	0,79
4	0,68	0,82	0,77	0,64	0,60	0,85	0,73
5	0,65	0,79	0,73	0,58	0,48	0,83	0,68
6	0,62	0,74	0,65	0,55	0,44	0,81	0,64
7	0,59	0,65	0,64	0,52	0,40	0,75	0,59
8	0,54	0,55	0,51	0,43	0,38	0,69	0,52
9	0,48	0,48	0,43	0,32	0,36	0,61	0,45
10	0,44	0,40	0,35	0,23	0,34	0,5	0,38
11	0,40	0,36	0,27		0,33		0,34
12		0,33	0,20		0,31		0,28
13		0,28			0,27		0,27

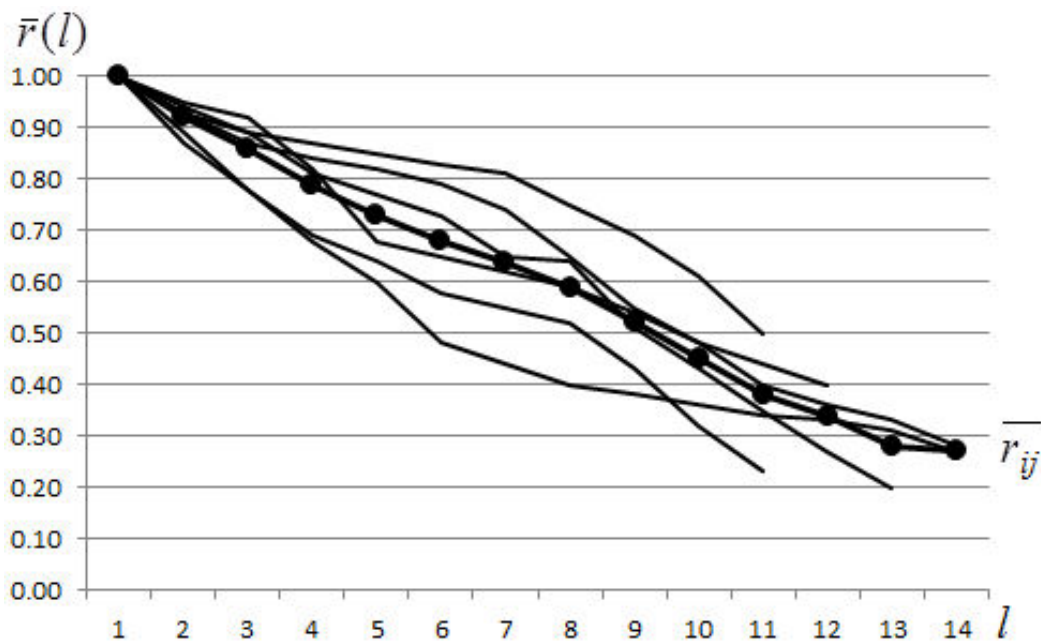


Рис. 1.3 – Просторово кореляційні функції місячних кількостей опадів у грудні з полюсами на метеорологічних станціях Дніпропетровськ та Чернігів

Висновок: на основі отриманих результатів дослідження метеорологічних полів місячних кількостей опадів у грудні на території України, можна зробити висновок, що вони володіють властивістю квазіоднорідності (поля кореляцій мають подібний вигляд при змінюванні полюса кореляцій і визначаються системою майже концентричних ізокорелят), а також квазіізотропності (всі просторово кореляційні функції незалежно від напрямку перетину полів кореляцій розташовуються у вузькій смузі).

Запитання для самоперевірки

1. Дати визначення багатовимірному «арифметичного», «метричного арифметичного» та «евклідового» просторів.
2. Дати визначення «метеорологічного поля». Перелічити види представлення метеорологічних полів при дослідженнях атмосферних процесів. Вміти представляти їх у матричному вигляді.
3. Як матрично представити сукупність метеорологічних полів? Чим є кожний рядок і стовпець такої вихідної матриці?
4. Як отримати матрицю центрованих значень ΔX , якщо є вихідна матриця X ? Чому дорівнює кожний елемент центрованої матриці?
5. В яких випадках застосовують операцію транспонування матриць та в чому її сенс?
6. Як отримати середнє метеорологічне поле та в чому його сенс?
7. Чому дорівнює скалярний добуток двох векторів евклідового простору?
8. Вміти записати норму вектора в евклідовому просторі. Чому дорівнює квадрат норми вектора, зменшений в m разів?
9. Як розрахувати середній квадратичний відхил випадкової величини на метеорологічній станції через відхилення від середнього арифметичного та через норму вектора?
10. Як розрахувати коваріаційний та кореляційний моменти між метеорологічною величиною в різних вузлах метеорологічного поля?
11. Яка інформація міститься в матриці коваріацій (кореляцій)?
12. Як отримати матрицю коваріацій (кореляцій)? Як представити матрицю коваріацій (кореляцій) у координатній формі? Якими властивостями володіють дані матриці?

2 КОМПОНЕНТНИЙ АНАЛІЗ ГІДРОМЕТЕОРОЛОГІЧНИХ ПОЛІВ

2.1 Компонентний аналіз як метод багатовимірної статистичної аналізу

Застосування статистичних методів аналізу й прогнозу метеорологічної ситуації й інших чисельних задач, пов'язаних з необхідністю врахування стану атмосфери чи об'єктів гідросфери спряжено із рядом труднощів. Це непередбачуваність в багатьох випадках метеорологічних величин нормальному закону розподілу ймовірностей, неоднорідність та неізотропність метеорологічних полів, нестационарність випадкових процесів, що характеризують зміну метеорологічних величин за часом і в просторі та інш. Існує ще одне суттєве обмеження, яке пов'язане з розробкою статистичних моделей різного призначення. Йдеться про дуже велику розсередженість вихідної інформації, яка необхідна для прийняття висновку про майбутній стан атмосферних процесів. Здається, що ці труднощі можливо обійти шляхом збільшення кількості впливаючих факторів, які називають *предикторами*, й розширення об'ємів їх статистичних сукупностей. Але чим більше предикторів, тим більшою стає розмірність матриці коваріації й тим гіршою буде її обумовленість. Це пояснюється тим, що збільшення кількості предикторів в статистичній моделі приводить до зростання тісноти кореляційних зв'язків між ними. У алгоритмах обчислення коефіцієнтів рівняння моделі, якою є, наприклад, рівняння регресії або параметрів дискримінантної функції, коли модель ґрунтується на теорії розпізнавання образів, обов'язково є процедура обернення коваріаційної матриці. Якщо вона погано обумовлена (це трапляється коли визначник є близьким до нуля), обернення матриці приводить до дуже ненадійних результатів. Як наслідок, параметри статистичної моделі будуть містити в собі великі похибки, і модель не буде в необхідній мірі адекватною процесу, для якого розробляється модель. Тому склад предикторів повинен бути в певній мірі оптимальним. Це досягається за допомогою методів відбору статистично значущих предикторів (методів «просіювання» предикторів).

Продуктивним може бути й інший шлях. Він полягає у тому, що спочатку проводиться параметризація впливаючих факторів, тобто в моделі замість вихідних факторів використовуються нові змінні, які є їх лінійними комбінаціями. Нові змінні повинні бути, по-перше, взаємно *некорельованими (ортогональними)*. По-друге, щоб при можливо меншій їх кількості враховувалася суттєва частина мінливості вихідних факторів.

Переліченим вимогам відповідає компонентний аналіз, який часто в метеорологічній літературі називають методом "емпіричних

ортогональних функцій (е. о. ф.)” або ”природних ортогональних функцій (п. о. ф.)”.

Існують й інші ортогональні функції, які можуть використовуватися для апроксимації двовимірних метеорологічних полів. Це, наприклад, поліноми Лежандра, Лаггера, Ерміта, Чебишева I і II роду, система експоненціальних поліномів й інші. Однак, розклади метеорологічних полів по цих поліномах є у великій мірі формальними у тому сенсі, що коефіцієнти розкладів не несуть смислової інформації. У назві ”природні ортогональні функції” підкреслюється той факт, що отримані на основі вихідних метеорологічних полів (або інших метеорологічних об’єктів, наприклад, вертикальних профілів метеорологічних величин) функції відбивають основні особливості їх статистичної структури.

Завдяки некорельованості нових змінних спрощується фізичний аналіз статистичних моделей, не виникає значних обчислювальних труднощів при розв’язаннях систем нормальних рівнянь, необхідних для визначення параметрів моделей. Не виникає особливих проблем при необхідності повернутися до вихідних змінних.

Окрім перелічених переваг, нові змінні, які називають головними компонентами, мають і самостійне значення. Вони, як правило, відбивають особливості тих чи інших фізичних процесів, які обумовлюють структуру метеорологічних об’єктів, тобто несуть у собі смислове навантаження.

Компонентний аналіз з успіхом використовується й при розв’язанні інших метеорологічних задач. Однією з них є задача **стиску метеорологічної інформації**. *Стиском інформації* називають суттєве скорочення кількості інформації при збереженні основного її змісту. Це є надзвичайно важливою обставиною при використанні її у практичних цілях.

Другою важливою задачею, яку дає можливість реалізувати компонентний аналіз, є **задача фільтрації метеорологічної інформації**. Суть її полягає у наступному. Як відомо, поля метеорологічних величин формуються під впливом атмосферних процесів різних масштабів: процесів макромасштабу, синоптичного масштабу, мезомасштабу та мікромасштабу. Метеорологічна інформація утримує й шумову компоненту, яка обумовлена дрібномасштабними флуктуаціями, похибками вимірювань та первинною обробкою результатів спостережень. Досить часто виникає необхідність, в залежності від характеру задачі при вивченні сутності атмосферних процесів, звернути увагу на процеси більших масштабів і не враховувати (відфільтрувати) складові, що обумовлені впливами процесів дрібних масштабів. Цю задачу також розв’язує компонентний аналіз.

2.2 Власні значення та власні вектори

Розглянемо поле центрованих значень метеорологічних величин ΔX_j , яке визначене у n його точках у j -тий момент часу:

$$\Delta X_j = \begin{pmatrix} \Delta x_{1j} \\ \Delta x_{2j} \\ \Delta x_{3j} \\ \dots \\ \Delta x_{nj} \end{pmatrix}. \quad (2.1)$$

Поставимо задачу здійснити **параметризацію** цього поля, *тобто виразити поле за допомогою декількох некорельованих параметрів, які лінійно зв'язані з компонентами випадкового вектора (поля) (2.1), які містять у собі основну інформацію про властивості поля.* В основі розв'язання цієї задачі лежить лінійне ортогональне перетворення вихідного поля (2.1) в базису власних векторів матриці коваріацій (або кореляцій) полів цієї метеорологічної величини. Отже, першим етапом у розв'язанні зазначеної задачі є визначення базису власних векторів. З цією метою розглянемо матричне **рівняння повної проблеми власних значень**:

$$K_x U_i = \lambda_i U_i, \quad (2.2)$$

де

K_x – n -вимірна матриця коваріацій;

U_i – i -тий власний вектор;

λ_i – відповідне власне значення матриці коваріації.

Рівняння (2.2) означає, що власний вектор U_i так перетворюється матрицею коваріацій, як він трансформується множенням його на власне число λ_i . Як відомо, таке перетворення вектора є **перетворенням розтягання**, коли $\lambda_i > 1$, і **перетворенням стиску**, якщо $\lambda_i < 1$.

Перетворення матричного рівняння (2.2) приводить його до вигляду:

$$(K_x - E\lambda_i)u_i = 0. \quad (2.3)$$

де

Якщо записати всі складові рівняння в координатній формі й виконати операцію множення, то матричне рівняння (2.3) перетвориться у систему однорідних лінійних алгебраїчних рівнянь:

[illegible]

$$u_i = \begin{pmatrix} u_{1i} \\ u_{2i} \\ \dots \\ u_{ki} \\ \dots \\ u_{ni} \end{pmatrix} \quad (i = \overline{1, n}). \quad (2.5)$$

28

$$\begin{vmatrix} \sigma_1^2 - \lambda & K_{12} & \dots & K_{1n} \\ K_{21} & \sigma_2^2 - \lambda & \dots & K_{2n} \\ \dots & \dots & \dots & \dots \\ K_{n1} & K_{n2} & \dots & \sigma_n^2 - \lambda \end{vmatrix} = 0. \quad (2.6)$$

Не важко показати, що детермінант (2.6) тотожний алгебраїчному рівнянню n -го степеня:

$$\lambda^n + A_1 \lambda^{n-1} + \dots + A_{n-1} \lambda + A_n = 0. \quad (2.7)$$

Коефіцієнти $A_1, A_2, \dots, A_n$ є відповідними комбінаціями елементів матриці коваріацій. Розв'язання рівняння (2.7) втілюється за допомогою відомих чисельних методів. Після визначення власних значень λ_i ($i = \overline{1, n}$), всі коефіцієнти системи (2.4) стають відомими, що дає змогу для кожного власного значення отримати відповідний власний вектор. Але алгебраїчна проблема власних значень розв'язується за допомогою відповідних методів. Найбільш популярним є **ітераційний метод обертань Якобі**, який характеризується точністю, надійністю й простотою розрахунків, порівняно з такими методами, як метод узагальнених обертань, ортогональний степеневий метод й інші ітераційні процедури.

Існує така теорема: *власні значення дійсної, додатно означеної, симетричної матриці є дійсними, додатними й простими числами.* Оскільки матриця коваріацій задовольняє умовам цієї теореми, то власні значення мають зазначені в теоремі властивості.

Власні значення розташовують у порядку їх зменшення:

$$\lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_n. \quad (2.8)$$

Означені властивості власних значень мають глибокий фізичний сенс, який буде розкритий пізніше.

Із системи рівнянь (2.4) випливає, що кожному власному значенню λ_i відповідає відповідний власний вектор (2.5).

Якщо матриця коваріацій (кореляцій) добре обумовлена, то перший власний вектор складається тільки з додатних компонент, у другого – спостерігається одна зміна знака компонент, у третього – дві зміни і т.д.

Власні вектори мають важливу властивість, яка визначається співвідношенням:

$$\langle U_i U_j \rangle = U_i' U_j = \delta_{ij} \|U_i\| \|U_j\|, \quad (2.9)$$

де

δ_{ij} – символ Кронекера.

Згадаємо, що дужки в лівій частині (2.9) означають скалярний добуток векторів, а подвійні дужки у правій частині – норми цих векторів.

Рівняння (2.9) свідчить про те, що **власні вектори ортогональні**.

Як правило, власні вектори нормуються: замість власних векторів $U_i (i = \overline{1, n})$ використовуються **ортонормовані власні вектори**:

$$W_i = \frac{U_i}{\|U_i\|} \quad (i = \overline{1, n}). \quad (2.10)$$

Скалярний добуток цих векторів, очевидно, дорівнює:

$$\langle W_i W_j \rangle = \frac{\langle U_i U_j \rangle}{\|U_i\| \|U_j\|} = \delta_{ij} \frac{\|U_i\| \|U_j\|}{\|U_i\| \|U_j\|} = \delta_{ij}. \quad (2.11)$$

Якщо власні вектори W_i розташувати у порядку нумерації, то ми отримаємо квадратну n – вимірну матрицю:

$$W = \{W_{ij}\} (i, j = \overline{1, n}), \quad (2.12)$$

яка є ортонормованою:

$$WW' = E. \quad (2.13)$$

як і раніше, індекс (') – позначає операцію транспонування матриці.

З лінійної алгебри відомо, що будь-яка система лінійно незалежних векторів багатовимірного простору являє собою базис цього простору. Оскільки власні вектори ортогональні, а значить лінійно незалежні, то вони складають базис n – вимірного евклідового простору.

Є підстави втілювати розкладання векторів – полів метеорологічних величин саме в цьому базису. Координати кожного власного вектора упорядковані. Це значить, що p -та координата k -того вектора W_{pk} належить p -тій точці вихідного поля, тобто є функцією координат цієї точки. Звідси виходить, що власний вектор можна зобразити як поле. Саме тому, як вже зазначалося, **власні вектори називають емпіричними**

ортogonalними функціями або природними ортogonalними функціями.

2.3 Ортogonalні компоненти метеорологічних полів

Отриманий в результаті розв'язання рівняння повної проблеми власних значень базис ортонормованих векторів W n -вимірного евклідового простору R^n може використовуватися для розкладання метеорологічних полів – векторів, що належать до простору R^n .

Будемо вважати, що вихідні поля метеорологічних величин ΔX_j є центрованими. Операція розкладання вектора ΔX_j у базису $W = \{W_{ij}\}$ ($i, j = \overline{1, n}$) визначається таким чином:

$$W' \Delta X_j = Z_j, \quad (2.14)$$

де

$$Z_j = \begin{pmatrix} z_{1j} \\ z_{2j} \\ \dots \\ z_{ij} \\ \dots \\ z_{nj} \end{pmatrix} \quad (i = \overline{1, n}), \quad (2.15)$$

вектор координат розкладу вихідного поля в базису власних векторів.

Оскільки базис власних векторів складається з ортogonalних векторів, то проєкції поля ΔX_j , які є елементами вектора (2.15) – **ортogonalні (незалежні, або у статистичному сенсі – некорельовані)**. Це означає, що

$$M[z_i' z_j] = \begin{cases} \sigma_{z_i}^2 & \text{çà òí î âè } i = j \\ 0 & \text{çà òí î âè } i \neq j \end{cases}, \quad (2.16)$$

де

M – операція математичного сподівання скалярного добутку зазначених векторів;

$\sigma_{z_i}^2$ – дисперсії проекцій z_{ij} метеорологічних полів.

Основою для розв’язання задачі повної проблеми власних значень є матриці коваріацій метеорологічних полів. Як було показано вище, вона розраховується на основі множини m полів, які складають матрицю:

$$X = \{\Delta X_{ij}\}, (i = \overline{1, n}; j = \overline{1, m}). \quad (2.17)$$

Якщо залучити операцію (2.14) до матриці (2.17):

$$W' \Delta X = Z, \quad (2.18)$$

то отримаємо матрицю:

$$Z = \{z_{ij}\}, (i = \overline{1, n}; j = \overline{1, m}) \quad (2.19)$$

ортогональних компонент всіх m вихідних метеорологічних полів у базису власних векторів.

Матричне рівняння (2.14) представимо у координатній формі:

$$\begin{pmatrix} W_{11} & W_{21} & \dots & W_{j1} & \dots & W_{n1} \\ W_{12} & W_{22} & \dots & W_{j2} & \dots & W_{n2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ W_{1i} & W_{2i} & \dots & W_{ji} & \dots & W_{ni} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ W_{1n} & W_{2n} & \dots & W_{jn} & \dots & W_{nn} \end{pmatrix} \begin{pmatrix} \Delta x_{1j} \\ \Delta x_{2j} \\ \dots \\ \Delta x_{ij} \\ \dots \\ \Delta x_{nj} \end{pmatrix} = \begin{pmatrix} z_{1j} \\ z_{2j} \\ \dots \\ z_{ij} \\ \dots \\ z_{nj} \end{pmatrix}. \quad (2.20)$$

Кожний компонент вектора, який розташовується у правій частині рівняння (2.20), дорівнює скалярному добутку відповідного вектора рядка матриці W на вектор-стовпець ΔX_j :

$$z_{ij} = \sum_{s=1}^n W_{si} \Delta x_{sj} \quad (i = \overline{1, n}; j = \overline{1, m}). \quad (2.21)$$

Із рівняння (2.22) видно, що оскільки середнє значення $\overline{\Delta x_{sj}} = 0$ ($\forall s = \overline{1, n}; j = \overline{1, m}$), то завдяки лінійності оператора осереднення $\overline{z_{ij}} = 0$ отримаємо дисперсії ортогональних компонент метеорологічних полів:

$$\sigma_{z_i}^2 = \frac{1}{m} \sum_{j=1}^m z_{ij}^2. \quad (2.22)$$

Підставивши під знак суми рівняння (2.21) отримаємо:

$$\sigma_{z_i}^2 = \sum_{s=1}^n \sum_{p=1}^n W_{si} W_{pi} K_{sp}, \quad (2.23)$$

де

K_{sp} — елемент матриці коваріацій K_x , який розташовується на перетині її s -того стовпця і p -того рядка. Можна показати, що

$$\sum_{s=1}^n \sum_{p=1}^n W_{si} W_{pi} K_{sp} = W_i' K_x W_i. \quad (2.24)$$

Очевидно, якщо помножити обидві частини основного рівняння повної проблеми власних значень:

$$K_x W_i = \lambda_i W_i \quad (2.25)$$

на вектор-рядок W_i' то отримаємо:

$$W_i' K_x W_i = \lambda_i W_i' W_i = \lambda_i. \quad (2.26)$$

Розглядаючи послідовно рівняння (2.23), (2.24) і (2.26), приходимо до висновку, що $\sigma_{z_i}^2 = \lambda_i$, тобто **власні значення є дисперсіями відповідних ортогональних координат розкладу метеорологічних полів у базису власних векторів.**

Отже, виходячи з нерівності (2.8) маємо:

$$\sigma_{z_1}^2 > \sigma_{z_2}^2 > \dots > \sigma_{z_n}^2. \quad (2.27)$$

Як відомо, значення дисперсії, а саме порядок, залежить від масштабів процесів. Наприклад, якщо метеорологічна величина – швидкість вітру, то у переважаючій планетарній західно-східній течії дисперсії мають порядок $10^2 - 10^3 \text{ м}^2/\text{с}^2$, у течіях синоптичного масштабу $10^2 - 10^1 \text{ м}^2/\text{с}^2$, а в турбулентних рухах $10^0 - 10^{-2} \text{ м}^2/\text{с}^2$. Отже, чим більший масштаб атмосферного процесу, під дією якого формуються поля метеорологічних величин, тим більша дисперсія їм притаманна. Тому $\sigma_{z_1}^2$ відбиває властивості полів найбільшого масштабу, $\sigma_{z_2}^2$ – полів меншого масштабу і інш. Ця властивість власних значень дає можливість отримати важливу інформацію про структуру метеорологічних об’єктів, що розглядаються.

Якщо виконати зазначену параметризацію на основі множини метеорологічних полів за рівнянням (2.18), то в матриці Z кожний її рядок має сенс статистичного ряду параметру z_{ij} .

Методи теорії випадкових процесів, що застосовані до часових рядів ортогональних компонент метеорологічних полів можуть дати важливу інформацію щодо їх властивостей і, як наслідок, властивостей метеорологічних процесів, що їх обумовлюють.

2.4 Головні компоненти метеорологічних полів. Задача стиску і фільтрації інформації

Дисперсії ортогональних компонент метеорологічних полів володіють властивістю (2.27), тобто зменшуються зі збільшенням номера ортогональної компоненти. Як було показано, кожна з дисперсій дорівнює відповідному власному значенню матриці коваріацій, але власні значення мають ще одну важливу властивість, окрім тих, що були сформульовані вище. Вона полягає у тому, що сума всіх власних значень дорівнює сумі дисперсій вихідних метеорологічних полів, тобто:

$$\sum_{i=1}^n \lambda_i = t_r K_x = \sum_{i=1}^n \sigma_{x_i}^2, \quad (2.28)$$

де

$t_r K_x$ – це сума всіх елементів головної діагоналі матриці K_x , тобто сума дисперсії метеорологічної величини на всіх точках поля (ця величина називається сумарною дисперсією поля).

Оскільки власні значення $\lambda_i (i = \overline{1, n})$ дорівнюють дисперсіям ортогональних компонент метеорологічних полів, то рівняння (2.28) можна переписати так:

$$\sum_{i=1}^n \sigma_{z_i}^2 = \sum_{i=1}^n \sigma_{x_i}^2. \quad (2.29)$$

Але дисперсії метеорологічних величин у точках поля не упорядковані, тобто можуть мати різні значення, у той час коли дисперсії ортогональних компонент розташовані строго в порядку зменшення, як це зазначається нерівністю (2.27). Причому вони володіють властивістю швидкої збіжності. Це дає можливість отримати інформацію про те як швидко процес збіжності відбувається. Очевидно, ця інформація полягає у тому, щоб визначити яку суму перших дисперсій ортогональних компонент, тобто проєкцій метеорологічних полів у базису власних векторів треба взяти, щоб вичерпати задану частку від сумарної дисперсії метеорологічних полів. Ця частка визначається рівнянням:

$$\eta_k = \frac{\sum_{i=1}^k \sigma_{z_i}^2}{\sum_{i=1}^n \sigma_{x_i}^2} = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^n \lambda_i} = \frac{\sum_{i=1}^k \lambda_i}{t_r K_x}, \quad (2.30)$$

де

k — кількість дисперсій ортогональних компонент власних значень матриці коваріацій, сума яких задовольняє заданій меті.

Ортогональні компоненти, сума дисперсій яких вичерпує більшу частку сумарної дисперсії вихідних метеорологічних полів, називаються головними компонентами.

Значення параметра η_k залежить від характеру задачі, яку ставить перед собою дослідник. Визначивши число головних компонент k гідрометеорологічних характеристик можна вирішити задачу **стиску інформації**: значне зменшення її об'єму при умові збереження основного фізичного сенсу.

Очевидно, ми досягнемо сформульованої мети, якщо сформуємо вектор головних компонент для j -того поля, дисперсії яких вичерпують задану частку від сумарної дисперсії вихідних полів за рівністю (2.30), а інші ортогональні компоненти обнулюємо. Такий вектор має вигляд:

$$\tilde{Z}_j = \begin{pmatrix} z_{1j} \\ z_{2j} \\ \dots \\ z_{kj} \\ 0 \\ \dots \\ 0 \end{pmatrix}. \quad (2.31)$$

Проведемо операцію обертання, оскільки

$$\tilde{Z}_j = W' \Delta \tilde{X}_j, \quad (2.32)$$

де

$\Delta \tilde{X}_j$ – n -вимірний вектор, компоненти якого утримують стислу інформацію.

Щоб його отримати, помножимо рівняння (2.32) на матрицю власних векторів. Будемо мати:

$$W \tilde{Z}_j = W W' \Delta \tilde{X}_j. \quad (2.33)$$

Оскільки $W W' = E$, маємо

$$\Delta \tilde{X}_j = W \tilde{Z}_j. \quad (2.34)$$

Дисперсії головних компонент розташовуються, як вже зазначалось, у порядку зменшення, тобто дисперсія першої головної компоненти $\sigma_{z_1}^2 = \lambda_1$ – є найбільшою, відбиває властивості гідрометеорологічних полів найбільшого масштабу, дисперсія другої головної компоненти $\sigma_{z_2}^2 = \lambda_2$ – значно менша і т.д. Отже, можна прийняти до висновку, що ортогональні компоненти, які у векторі Z_j розташовуються після головних компонент, відбивають властивості все меншого й меншого масштабу, а останні – похибки первинної обробки і вимірювань. Таким чином рівняння (2.31) – (2.34) дають можливість вирішити **задачу фільтрації інформації від дрібномасштабних флуктуацій та похибок**.

2.5 Приклад застосування компонентного аналізу метеорологічних об'єктів

Місячна кількість опадів q є однією із важливих кліматичних параметрів. Вона характеризує умови зволоження того чи іншого району, а її поля – умови зволоження територій, для яких ці поля побудовані. На рис.2.1 зображається поле місячних кількостей опадів у лютому 1980 року. Воно визначено значеннями цієї метеорологічної величини на мережі 21 метеорологічних станцій України. На карті методом лінійної екстраполяції проведені ізогієти – лінії однакових кількостей опадів. Таке поле, може бути зображене матрицею Q розмірності $21 \times n$:

$$Q = \begin{pmatrix} q_{11} & q_{12} & \dots & q_{1n} \\ q_{21} & q_{22} & \dots & q_{2n} \\ \dots & \dots & \dots & \dots \\ q_{21,1} & q_{21,2} & \dots & q_{21,n} \end{pmatrix} \quad (2.35)$$

В матриці (2.35) кожний стовпець визначає окреме поле місячних кількостей опадів, а кожний рядок - вибірку цієї метеорологічної величини на метеорологічній станції, умовний номер якої співпадає з номером рядка. В прикладі, що наводиться, здійснювався добір місячних кількостей опадів за період з 1951 до 1981 ($n = 31$). На основі матриці (2.35) була розрахована матриця коваріацій K_Q порядку 21×21 і відповідна матриця кореляцій R_Q , для якої розв'язувалася повна проблема власних значень. В табл.2.1 наводяться перші власні значення, які вичерпують основну долю сумарної дисперсії змінних, і відповідні їм власні вектори.

Із табл. 2.1 випливає, що перші два власні значення вичерпують біля 97% сумарної дисперсії центрованих і нормованих на середній квадратичний відхил місячних кількостей опадів (слід матриці кореляцій). Тому найбільш великомасштабні властивості статистичної структури полів опадів утримуються в перших двох власних векторах. Перший власний вектор має однорідну структуру, що обумовлюється впливом на поля опадів найбільш крупномасштабного опадоутворюючого процесу – переміщення через Україну циклонів і атмосферних фронтів в загальному напрямку із заходу на схід. Другий власний вектор має області максимуму в південно-східній частині і мінімум в північно-західній частині України. Зазначений максимум обумовлюється впливом циклонів, що

спостерігалися у цю пору і проходили з Чорного моря в північно-східному напрямку.

Таблиця 2.1 – Власні значення λ_i і відповідні їм власні вектори місячних кількостей опадів у лютому.

Номер компо- нент	Номери власних векторів						
	i	1	2	3	4	5	6
	λ_i	19,25	1,27	0,27	0,17	0,02	0,01
1		0,22	-0,28	0,07	-0,20	-0,25	0,06
2		0,21	-0,29	0,17	-0,44	0,04	-0,43
3		0,22	-0,27	0,18	-0,06	0,19	0,19
4		0,21	-0,29	0,20	-0,31	0,17	-0,19
5		0,22	-0,12	-0,27	0,13	-0,40	-0,01
6		0,22	-0,05	-0,36	0,11	-0,27	-0,24
7		0,22	-0,22	-0,02	0,06	-0,20	0,29
8		0,22	-0,14	-0,15	0,14	-0,20	0,14
9		0,22	-0,22	0,07	0,10	0,04	0,32
10		0,23	-0,03	-0,05	0,20	0,21	-0,03
11		0,22	-0,12	0,08	0,26	0,24	0,23
12		0,22	0,03	0,29	0,43	0,28	-0,10
13		0,22	0,17	0,11	0,15	0,13	-0,10
14		0,22	0,266	0,26	0,09	-0,12	-0,13
15		0,23	0,01	-0,27	0,10	0,05	-0,15
16		0,22	0,12	-0,28	-0,02	0,18	-0,18
17		0,22	0,19	-0,36	-0,18	0,23	-0,18
18		0,20	0,35	-0,07	-0,46	0,06	0,54
19		0,22	0,22	-0,12	-0,09	0,18	-0,01
20		0,21	0,32	0,17	-0,14	-0,14	-0,11
21		0,21	0,32	0,40	0,03	-0,45	-0,18

Відповідно до зазначеної структури власних значень і власних векторів, головна інформація про властивості множини полів опадів утримується в перших двох головних компонентах, рядки яких утворюються ортогональними перетвореннями матриці (2.35) першими двома власними векторами

$$\begin{pmatrix} z_{11} & z_{12} & \dots & z_{1n} \\ z_{21} & z_{22} & \dots & z_{2n} \end{pmatrix} = \begin{pmatrix} W_{11} & W_{21} & \dots & W_{21,1} \\ W_{12} & W_{22} & \dots & W_{21,2} \end{pmatrix} \begin{pmatrix} q_{11} & q_{12} & \dots & q_{1n} \\ q_{21} & q_{22} & \dots & q_{2n} \\ \dots & \dots & \dots & \dots \\ q_{21,1} & q_{21,2} & \dots & q_{21,n} \end{pmatrix}.$$

Таким чином, здійснюється операція стиску інформації про n двадцятиодновимірних полів місячних кількостей опадів.

Розглянемо тепер задачу, фільтрації поля опадів. Для цього необхідно втілити операцію, яка відповідає рівності (2.34). Приведемо її для поля місячних кількостей опадів, що міститься на рис.2.1. Відповідне фільтроване поле зображається на рис.2.2. Для освітлення результату фільтрації порівняємо вихідне та фільтроване поля. На рис.2.1 виразно виділяються два максимуми опадів. Перший з них (вторинний) розташовується в районі Карпат, другий в Придніпров'ї. Їх розділяє область мінімуму, вісь якої приходить від Ізмаїлу на Кам'янець-Подільський і далі у північному напрямку. Структура поля опадів яскраво відбиває вплив Карпат на опадоутворюючі процеси, а саме при перевалюванні циклонів і атмосферних фронтів, що надходять із південно-західного і західного напрямків, через Карпати, відбувається їх ослаблення, що обумовлює зону мінімуму опадів. При подальшому їх переміщенні на схід над Придніпров'єм вони знову набувають активність, і кількості опадів зростає. На величину і конфігурацію максимуму над східною Україною чинять вплив і циклони, траєкторії яких приходили у цьому місяці з Чорного моря на північний схід через цю частину України. На відмінну від цього, фільтроване поле має лише один максимум місячних кількостей опадів, що розташовується над південно-східною Україною. Причина появи його вже обговорювалась. Необхідно підкреслити, що в результаті операції фільтрації в полі опадів зникає вторинний процес – коливання кількості опадів, яке проявляється шляхом утворення мінімуму над зоною, що межується з східними схилами Карпат.

Другий приклад використання компонентного аналізу відноситься до досліджень вертикальної структури полів швидкості вітру в тропосфері і стратосфері західної півкулі.

У якості вихідних даних виступають результати радіо – та ракетного зондування атмосфери над районом, який розташовується в помірній (острови Воллоп: 37° 50' півн.ш., 75° 29' зах.д.). Північної Америки. Ряди зональної та меридіональної компонент швидкості вітру з тижневою дискретністю були сформовані для рівнів 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 і 55 км. Шар атмосфери 5-55 км за відомими властивостями поділявся на два шари: 5-25 км і 25-55 км. Значення компонент швидкості вітру на

зазначених рівнях, що відносяться до першого чи другого шару, розглядалися як компоненти 5-вимірною у першому шарі і 7-вимірною у другому шарі вектора. Послідовності векторів складають матриці вихідних даних типу (2.35). Для зазначеного пункту зондування атмосфери матриці зональної і меридіональної складових вітру мають порядок 5×528 - для першого, і порядок 7×528 – для другого шару атмосфери.

На основі цих матриць були розраховані матриці коваріацій, для яких розв'язувалася повна проблема власних значень. Розглянемо результати лише для зональної складової швидкості вітру.

В табл. 2.2 містяться значення перших двох власних значень матриці коваріацій, що вичерпують біля 90 сумарної дисперсії, і відповідні їм власні вектори для шару 5 - 25 км.

Таблиця 2.2 – Власні значення і власні вектори матриці коваріацій зональної складової швидкості вітру. Шар 5 - 25 км.

Висота, км	$\lambda_1 = 581.0$ $\eta = 0,76$	$\lambda_1 = 104.6$ ($\eta = 0,89$)
	W_1	W_2
5	0,3845	- 0,2241
10	0,6646	- 0,4549
15	0,4577	0,0339
20	0,2994	0,3892
25	0,3338	0,7683

В табл. 2.3 утримуються перші три власні значення матриці коваріацій для шару 25-55 км і відповідні їм власні вектори. Як видно, вони складають 85 сумарної дисперсії зональної складової швидкості вітру в цьому шарі атмосфери.

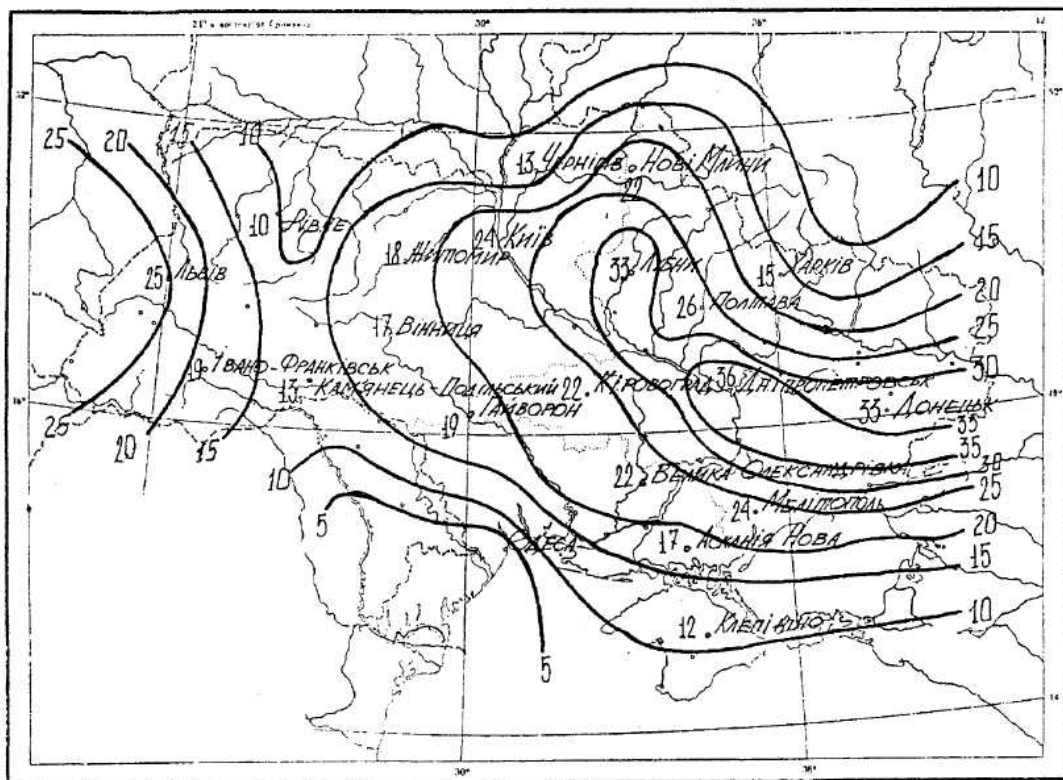


Рис. 2.1 – Фактичне поле місячних кількостей опадів на території України в лютому 1980 р.

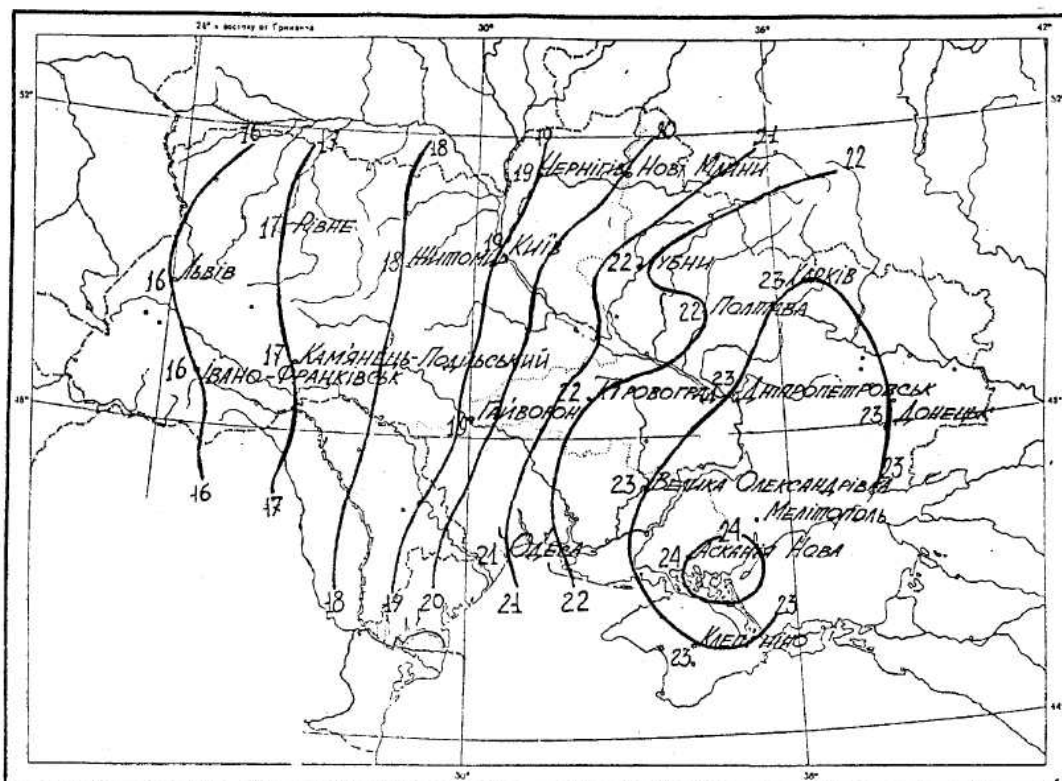


Рис. 2.2 – Відфільтроване поле місячних кількостей опадів на території України в лютому 1980 р. (відповідає полю опадів рис. 2.1)

Таблиця 2.3 – Власні значення і власні вектори матриці коваріацій зональної складової швидкості вітру. Шар 25-55 км.

Висота км	$\lambda_1 = 419.6$ ($\eta = 0.60$)	$\lambda_2 = 121.8$ ($\eta = 0.77$)	$\lambda_3 = 59.4$ ($\eta = 0.85$)
	W_1	W_2	W_3
25	0,0208	- 0,0763	0,0085
30	0,0922	- 0,1735	- 0,0511
35	0,1716	- 0,2637	- 0,0826
40	0,4162	- 0,7887	0,2881
45	0,4893	0,0391	- 0,5666
50	0,5263	0,2597	- 0,3437
55	0,5216	0,4511	0,6843

Табл. 2.2 і 2.3 свідчать про те, що основні риси вертикальної структури зональної складової швидкості вітру утримуються в перших двох головних компонентах в шарі 5 - 25 км і в перших трьох головних компонентах в шарі 25 - 55 км. Вони були отримані шляхом ортогонального перетворення матриць вихідних даних за допомогою приведених в табл.2.2 і 2.3 власних векторів. Часові послідовності головних компонент підлягали операції фільтрації за допомогою інтегрального перетворення Фур'є, в результаті чого виявлені періодичності, що сховані в часових рядах головних компонент.

Із таблиці 2.4 випливає, що зональній складовій швидкості вітру в шарі 5 - 25 км (тропосфера і нижня стратосфера) притаманні 3 періодичності: річна – в першій, річна і піврічна – в другій головній компоненті. Амплітуда річної періодичності у першій головній компоненті є досить великою (12,8 м/с). Річна періодичність, що проявляється в другій головній компоненті має у 4 рази меншу амплітуду ніж у першій компоненті. Третя ж головна компонента включає широкий спектр періодичних коливань.

В середній і верхній стратосфері (шар 25-55 км) перша головна компонента утримує річну і піврічну періодичності з дуже великими амплітудами. Друга і третя головні компоненти характеризуються наявністю широких спектрів періодичних коливань. У другій головній компоненті крім річної проявляються восьмимісячна, чотирьохмісячна періодичності, а також 2-3 місячні і хвилі з періодом 2-3 неділі (хвилі Кельвіна). Останні можуть бути віднесеними до хвиль Кельвіна. Цікавим є той факт, що у третій головній компоненті крім періодичностей, що утримуються в другій головній компоненті зональної складової швидкості вітру, проявляється квазідвохрічна періодичність. Це свідчить про те, що

квазідвохрічна періодичність швидкості вітру є притаманною стратосфері не тільки приєкваторіальної зони, а і стратосфері середніх широт.

Таблиця 2.4 – Періодичності, що утримуються в часових рядах головних компонент зональної складової швидкості вітру (острови Воллоп).

Номер головної компонен- ти	Частота (ω , 1/тижд.)	Період (T , тижд.)	Амплітуда, (A , $\text{м} \cdot \text{с}^{-1}$)	Фазовий зсув (ϕ , рад)
шар 5 - 25 км				
1	0,132	47,7	12,8	1,09
2	0,132	47,7	3,0	1,53
	0,251	25,0	1,8	0,67
3	0,284	22,1	1,2	0,11
	0,540	11,6	1,1	-1,07
	1,231	5,1	1,1	-1,22
	2,093	3,0	1,2	0,04
	2,166	2,9	1,1	0,91
шар 25 - 55 км				
1	0,132	47,7	62,0	1,42
	0,251	25,0	12,8	0,41
2	0,132	47,7	4,7	-0,35
	0,183	34,3	2,9	-1,11
	0,388	16,2	3,4	-0,88
	0,523	12,0	2,8	-1,50
	0,785	8,0	2,6	-1,29
	2,326	2,7	2,3	0,77
3	0,063	99,3	3,3	-0,67
	0,217	28,9	2,7	-0,76
	0,338	18,6	3,3	0,02
	0,388	16,2	2,9	-0,01
	0,491	12,8	2,9	0,23
	0,675	9,3	2,3	-1,29
	1,282	4,9	2,4	-0,55

Наведений приклад показує, що компонентний аналіз дає можливість провести ретельне дослідження властивостей великомасштабних

складових загальної циркуляції атмосфери, виявити шляхом статистичного аналізу головних компонент швидкості вітру періодичності, які притаманні швидкості вітру в тропосфері і стратосфері.

Контрольні запитання

1. У чому полягає сенс компонентного аналізу? Перелічить напрямки його використання?
2. Записати основні рівняння проблеми власних значень та прокоментувати їх.
3. У чому полягає основний смисл рівняння повної проблеми власних значень?
4. Що лежить в основі розв'язання задачі повної проблеми власних значень?
5. Як ви розумієте задачу параметризації метеорологічних об'єктів?
6. У чому полягає сенс задачі «стиску» метеорологічної інформації?
7. У чому полягає сенс задачі «фільтрації» метеорологічної інформації?
8. Властивості власних векторів матриці кореляцій (коваріацій)? Які вектори називаються ортонормованими?
9. Властивості власних значень матриці кореляцій (коваріацій)?
10. Сформулюйте теорему, на основі якої знаходять власні значення матриці коваріацій (кореляцій).
11. Який сенс мають власні значення?
12. Як отримати вектор ортогональних компонент? Які компоненти називають головними?
13. Як розв'язати задачу стиску інформації за допомогою компонентного аналізу?
14. Як можна розв'язати задачу фільтрації інформації за допомогою компонентного аналізу?

3 ФАКТОРНИЙ АНАЛІЗ

3.1 Основні ідеї факторного аналізу

Факторний аналіз є тією гілкою багатовимірного статистичного аналізу, яка розкриває внутрішню структуру матриць коваріацій або кореляцій. При аналізі структури цих матриць найбільш часто зустрічаються два напрями. Один з них – це розглянутий у попередньому розділі метод головних компонент, інший – факторний аналіз. Необхідно розрізняти ці два підходи, хоча вони, на перший погляд, мають деяку подібність.

Метод компонент співпадає з методом розчеплення коваріаційної або кореляційної матриці на сукупність ортогональних векторів (або напрямів) за кількістю змінних, по яких побудована матриця коваріації. Ці вектори, як було показано, відповідають власним значенням і власним векторам коваріаційної (кореляційної) матриці. За цим методом власні значення розташовуються у порядку зменшення, що є суттєвим, якщо для опису даних необхідно використовувати лише незначну кількість компонент. Ці декілька головних компонент можуть вміщувати більшу частину від сумарної дисперсії змінних, але для точного відтворення кореляцій між змінними необхідні всі компоненти. У випадках, коли застосовують метод головних компонент, немає необхідності формулювати будь-яких гіпотез про змінні.

У протилежність методу головних компонент, факторний аналіз заздалегідь пояснює матрицю коваріацій наявністю максимальної, або принаймні, невеликої кількості гіпотетичних змінних або факторів. Це означає, що дослідник, перш за все ставить питання, чи існує випадкова величина f_1 така, що попарні кореляції між змінними дорівнюють нулю, якщо вплив f_1 вже враховано. Якщо після цього кореляційна матриця ще не пояснена, дослідник розглядає вже дві випадкові величини f_1 і f_2 , такі, що їх вплив приводить до того, що парні кореляції між змінними дорівнюють нулю, і так далі. Отже, у той час, як *метод головних компонент орієнтується на дисперсії*, **факторний аналіз орієнтований на коваріаційний (або кореляційний) зв'язок**.

Сенс цих двох методів можна охарактеризувати іншим чином. Нехай $x_1, x_2, \dots, x_i, \dots, x_n$ – n -вимірна сукупність вихідних змінних. У компонентному аналізі шляхом ортогонального перетворення отримуються інші ортогональні змінні $z_1, z_2, \dots, z_i, \dots, z_n$. Вони вибираються так, щоб у

z_1 була максимальна дисперсія, у z_2 - максимальна дисперсія за умовою, що z_2 не корельовано з z_1 і т.п. До того ж відносно змінних x_i ($i = \overline{1, n}$) не потрібно формулювати будь-яких припущень.

У факторному аналізі основною передумовою є рівняння

$$x_i = \sum_{j=1}^k P_{ij} f_j + v_i, \quad (i = \overline{1, n}), \quad (3.1)$$

де

f_j – *узагальнений фактор*, $k \ll n$ - задане число;

v_i – залишки, що являють собою відхилення, які відносяться тільки до x_i .

Ці n залишків вважаються незалежними як між собою, так і з факторами f_j . Крім того, фактори f_j – теж незалежні величини, хоча в теорії факторного аналізу розглядаються випадки й з корельованими факторами.

P_{ij} – деякі коефіцієнти, які потрібно визначити. Їх називають *вагою j -того фактора* у i -тій змінній, або вагою i -тої змінної на j -тий фактор.

Вважається, що змінні x_i центровані. Тоді всі середні значення дорівнюють нулю, у тому числі й невідомі фактори f_j . Незважаючи на сформульовані гіпотези, рівняння (3.1) утримують у собі гіпотезу відносно коваріацій та дисперсій вихідних змінних x_i , яку можна перевірити.

Ще одна важлива відмінність факторного аналізу від компонентного полягає у тому, що другий із цих методів є лінійним й адаптованим, тоді як у факторному аналізі лінійність розглядається як гіпотеза. Але вона може бути відкинutoю, як лінійна модель (3.1), якщо фактичні дані не будуть задовольняти цій моделі.

$$P = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1k} \\ p_{21} & p_{22} & \dots & p_{2k} \\ \dots & \dots & \dots & \dots \\ p_{n1} & p_{n2} & \dots & p_{nk} \end{pmatrix} . \quad (3.6)$$

Тоді систему рівнянь (3.2) можна записати в матричній формі:

$$X = PF + V . \quad (3.7)$$

Як зазначалось у попередньому розділі, відносно факторів f_j й залишків v_i сформульовані такі гіпотези

$$M[F \cdot F'] = E , \quad (3.8)$$

$$M[V \cdot V'] = D , \quad (3.9)$$

де

E – одинична матриця;

D – діагональна матриця, що складається з дисперсій d_i залишків v_i :

$$D = \begin{pmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & d_n \end{pmatrix} \quad (3.10)$$

Це означає, що узагальнені фактори некорельовані й мають одиничну дисперсію. Крім того, залишки v_i є некорельовані з факторами. За таких умов знайдемо матрицю коваріацій вихідних змінних, пам'ятаючи, що вони центровані

$$\begin{aligned} K_x &= M[X \cdot X'] = M\{(PF + V)(F'P' + V')\} = \\ &= PM[F \cdot F']P' + M[V \cdot F'] \cdot P' + PM[F \cdot V'] + M[V \cdot V'] . \end{aligned} \quad (3.11)$$

Враховуючи перелічені властивості відносно змінних, що входять у рівняння (3.7), маємо

$$K_x = PP' + D. \quad (3.12)$$

Отже, коваріаційну матрицю випадкової величини X можна виразити через матрицю вагів факторів і діагональну матрицю дисперсій залишків.

3.3 Оцінки вагів узагальнених факторів і дисперсій залишків

Будемо вважати, що змінні $x_1, x_2, \dots, x_n$ підпорядковуються нормальному багатовимірному закону розподілу з n -вимірною матрицею коваріацій K_x . Це традиційна умова для багатовимірного статистичного аналізу.

З урахуванням рівнянь (3.1) і (3.7) можна записати, що діагональні елементи матриці коваріацій (k_{ii}) – дисперсії вихідної змінної, які дорівнюють

$$k_{ii} = \sum_{j=1}^k P_{ij}^2 + d_i, \quad (3.13)$$

а коваріації

$$k_{ij} = \sum_{j=1}^k P_{ij} P_{qj}, \quad (3.14)$$

Отже, основна модель полягає у виборі гіпотези H_0 відносно коваріаційної матриці K_x , а саме, що вона може бути представлена як сума діагональної матриці (3.13) з додатними елементами й матриці рангу k (3.14) з додатними власними значеннями.

Якщо дисперсії d_i будуть відомі, то треба щоб $k < n$. У загальному випадку, коли інформація про значення дисперсій залишків відсутня, ця умова приймає вигляд $(n + k) < (n - k)^2$.

Задача полягає у тому, щоб на основі вибіркової матриці коваріацій $\hat{K}$ знайти незсунені, ефективні та умотивовані оцінки факторних вагів p_{ij} і дисперсій залишків d_i ($i = \overline{1, n}; j = \overline{1, k}$).

Найбільш сприятливими для розв'язання цієї задачі є два методи: центроїдний метод і метод максимальної правдоподібності. Розглянемо останній. Він полягає у тому, що за певним правилом формується функція правдоподібності, на основі якої знаходять такі шукані параметри, котрі

надають максимум функції правдоподібності. Функція правдоподібності має такий вигляд:

$$L = -\frac{1}{2}n \ln |K| - \frac{n}{2} \text{tr} \left(K^{-1} \cdot \widehat{K} \right), \quad (3.15)$$

де

tr – слід матриці;

$|K|$ – **детермінант коваріаційної** матриці генеральної сукупності.

Функція правдоподібності, як свідчать рівняння (3.13 - 3.14), є функцією двох шуканих змінних p_{ij} і d_i . Максимум функції (3.15) визначається за допомогою рівнянь

$$\frac{\partial L}{\partial p_{ij}} = 0 ; \quad \frac{\partial L}{\partial d_i} = 0. \quad (3.16)$$

Диференціювання функції L приводить до матричних рівнянь

$$P' K' - P' K^{-1} \widehat{K} K^{-1} = 0, \quad (3.17)$$

$$\text{diag} \left(K^{-1} - K^{-1} \widehat{K} K^{-1} \right) = 0, \quad (3.18)$$

де

$\text{diag } B$ (B - матриця в дужках) позначає матрицю, яка має лише діагональні елементи.

Після матричних перетворень будемо мати

$$P' = N^{-1} P \ddot{A}^{-1} (\widehat{K} - \ddot{A}), \quad (3.19)$$

$$\text{diag} \left(K - \widehat{K} \right) = 0, \quad (3.20)$$

Рівняння (3.20) означає, що $i = j$ $K_{ii} - \widehat{K}_{ii}$, тобто діагональні елементи генеральної й вибіркової матриць коваріацій дорівнюють одне одному. Враховуючи це, за допомогою рівняння (3.19) отримаємо

$$d_i = \widehat{K}_{ii} - \sum_{j=1}^k P_{ij}^2. \quad (3.21)$$

Отже, оскільки вибіркова коваріаційна матриця відома, рівняння (3.21) дає можливість знайти дисперсії залишків після визначення вагів факторів. Для їх розрахунків помножимо рівняння (3.19) ліворуч на матрицю

$$N = P'D^{-1}P, \quad (3.22)$$

а праворуч на матрицю $D^{-1}P$. Отримаємо матрицю

$$N^2 = P' \ddot{A}^{-1} (\hat{K} - \ddot{A}) \ddot{A}^{-1} P, \quad (3.23)$$

яка є діагональною. Її елементи являють собою перші k власних значень матриці $\ddot{A}^{-1} (\hat{K} - \ddot{A})$, а рядки матриці P – її відповідні власні вектори.

Оскільки вибіркова коваріаційна матриця $\hat{K}$ може бути розрахованою, зазначені особливості дають можливість побудувати алгоритм ітераційного процесу розв'язання рівнянь (3.19) і (3.21). Ітераційний процес описується такою системою рівнянь:

$$\left\{ \begin{array}{l} h_{\mu(v)} = P'_{\mu(v)} \ddot{A}^{-1}_{(v)}; \\ \alpha_{\mu j(v)} = h'_{\mu(v)} P_{j(v)}; \\ g_{\mu(v)} = h_{\mu(v)} \hat{K} - P'_{\mu(v)} - \sum_{j=1}^{\mu-1} \alpha_{\mu j(v)} P'_{j(v+1)}; \\ \beta_{\mu(v)} = g'_{\mu(v)} h_{\mu(v)}; \\ P_{\mu(v+1)} = \frac{1}{\sqrt{\beta_{\mu(v)}}} g'_{\mu(v)}; \end{array} \right. \quad (3.24)$$

де

μ – номер узагальненого фактора;

$v = \overline{0, r}$ – номер ітерації;

P'_{μ} – μ -тий рядок матриці P' (де $'$, як і раніше, операція транспонування).

Для організації ітераційної процедури необхідно визначити кількість k факторів, які разом з залишком v_i описують множину вихідних змінних x_i . Ця задача вирішується шляхом перевірки гіпотези H_0 про наявність k факторів. Таким чином, основна модель полягає у виборі гіпотези H_0 про коваріаційну матрицю K_x . Як зазначалося вище, вона повинна бути такою, щоб виконувались умови (3.13) і (3.14), тобто вона повинна бути сумою діагональної матриці з додатніми елементами (3.13) матриці рангу k з додатніми власними значеннями. Якби v_i були відомі, тоді необхідно щоб $k < n$. Але коли v_i невідомі, умова приймає вигляд $(n+k) < (n-k)^2$, додатність числа

$$\frac{1}{2}[(n-k)^2 - (n+k)] \geq 0 \quad (3.25)$$

забезпечує не тривіальність гіпотези H_0 . Звідси виходить, що для коваріаційної матриці третього порядку $n=3$, **може існувати лише один фактор** $k=1$.

У загальному випадку гіпотеза H_0 перевіряється за допомогою критерію χ^2 , який розраховується за формулою:

$$\chi^2 = n \left(\frac{1}{2} \sum_{i=1}^n y_{ii}^2 + \sum_{i,j=1}^n y_{ij} y_{ji} \right), \quad (3.26)$$

де

y_{ij} ($i, j = \overline{1, n}$) – елементи матриці:

$$Y = \ddot{A}^{-1} (\hat{K}_0 - PM'), \quad (3.27)$$

а матриця M' визначається рівнянням:

$$M' = N^{-1} P' \ddot{A}^{-1} \hat{K}_0. \quad (3.28)$$

Матриця $\hat{K}_0$ формується шляхом заміни елементів головної діагоналі матриці коваріацій $\hat{K}$ найбільшими елементами її відповідних стовпців (при визначенні вагів першого фактора) або матриці залишків (останні утворюються в результаті виключення факторів, для яких оцінки факторів вже проведені).

Критерій χ^2 порівнюється з табличним для числа степенів вільності:

$$v = \frac{1}{2}[(n-k)^2 + (n+k)] \quad (3.29)$$

та вибраного рівня значущості α .

Число ітерацій визначається співвідношенням:

$$|P_{\mu j(m+1)} - P_{\mu j(m)}| < \varepsilon, \quad j = \overline{1, n}, \quad (3.30)$$

де

ε – задана точність розрахунків.

Після отримання матриць P і D , можна розрахувати вектор F значень узагальнених факторів для вектора X за формулою:

$$F = (E + N)^{-1} P' D^{-1} X. \quad (3.31)$$

Це дає можливість визначити сукупність n вихідних характеристик через невелику кількість k узагальнених факторів.

3.4 Приклади застосування факторного аналізу до гідрометеорологічної інформації

Факторний аналіз був започаткований відомим психологом Спірманом і спочатку використовувався при розв'язанні задач з психології. Але роботи відомих математиків Пірсона, Хотеллінга й інших, присвячені факторному аналізу, удосконалили цей метод й починаючи з першої половини минулого століття, використовують його при розв'язанні проблем статистичного аналізу і в інших науках, у тому числі і науках метеорологічного профілю. Наведемо два приклади застосування факторного аналізу.

Перший з них відноситься до проведення досліджень кліматичних особливостей температурно-вологісного режиму території України. Він може бути охарактеризований трьома параметрами: середньою за місяць відносною вологою x_1 , середньомісячною температурою повітря x_2 й місячною кількістю опадів x_3 . У якості вихідних даних використовувались перелічені характеристики за термін березень - липень за період 1900-1974 рр. для 14 міст України, рівномірно розподілених на території країни.

На основі вихідних даних були розраховані матриці коваріацій, які мають третій порядок. Означені кліматичні характеристики визначаються

одним фактором. На основі приведенного алгоритму за допомогою комп'ютерної програми «FACTOR» були розраховані для кожного пункту матриці коваріацій й вектори вагів фактора відносної вологості p_1 , середньомісячної температури повітря p_2 й місячної кількості опадів p_3 . Вони представлені в таблиці 3.1

Таблиця 3.1 - Ваги фактору температурно-вологісного режиму і дисперсії залишків.

Пункт	березень				липень			
	Ваги			d_i^2	Ваги			d_i^2
	p_1	p_2	p_3		p_1	p_2	p_3	
Львів	0,49	-0,31	0,71	0,45	0,47	-0,62	0,50	0,48
Чернівці	0,94	-0,28	0,38	0,39	0,80	-0,65	0,61	0,33
Житомир	0,79	-0,26	0,39	0,41	0,96	-0,57	0,53	0,22
Київ	0,70	-0,43	0,67	0,38	0,94	-0,69	0,70	0,16
Умань	0,58	-0,42	0,66	0,45	0,78	-0,56	0,64	0,26
Кіровоград	0,98	-0,40	0,47	0,28	0,98	-0,62	0,58	0,07
Миколаїв	0,79	-0,37	0,50	0,39	0,91	-0,67	0,54	0,12
Херсон	0,01	-0,73	0,06	0,49	0,84	-0,64	0,69	0,08
Одеса	0,91	-0,79	0,45	0,34	0,94	-0,54	0,64	0,18
Симферопіль	0,96	-0,46	0,38	0,33	0,87	-0,55	0,48	0,30
Полтава	-0,15	-0,93	-0,09	0,20	0,98	-0,66	0,64	0,05
Харків	-0,92	-0,10	-0,30	0,24	0,99	-0,66	0,60	0,05
Дніпропетровськ	-0,96	-0,38	-0,25	0,14	0,91	-0,65	0,66	0,11
Луганськ	-0,19	-0,96	-0,22	0,14	0,99	-0,63	0,62	0,05

За даними табл. 3.1 видно, що майже усі ваги фактора середньомісячної температури повітря мають протилежні знаки відносно вагів середніх за місяць відносних вологостей й місячних сум опадів. Це свідчить про те, що на території України у зазначений період більшість випадків при одному й тому ж значенні фактора від'ємні аномалії середньомісячної температури повітря супроводжуються додатними аномаліями середньомісячної відносної вологості повітря й місячної суми опадів, і навпаки. Значення вагів фактора визначають, що в Україні температурно-вологісний режим має свої особливості. Вони полягають, по-перше, у тому, що у березні ваги фактора відносної вологості й середньомісячної температури повітря над Лівобережною Україною мають знаки, протилежні знакам відповідних вагів для Правобережної України. Це значить, що у цьому місяці при одному й тому ж факторі від'ємним аномаліям температури й додатним аномаліям сум опадів в західній частині

країни у більшості випадків відповідають додатні аномалії температури й від'ємні аномалії відносної вологості у східній її частині. Відносно опадів, подібну закономірність можна побачити лише у Дніпропетровську й Хракові.

По-друге, східна частина України з травня по серпень виділяється однорідністю знаків вагів фактора, хоча значення вагів, особливо у сумах опадів, значно розрізняються як по регіонах країни, так і по місяцях.

По-третє, починаючи з червня, однорідністю значень вагів характеризуються Кіровоград, Херсон, Миколаїв, які відносяться до степової південної частини України. У серпні до них приєднуються Умань і Київ.

Отже, аналіз вагів узагальнених факторів дає можливість зробити висновок про необхідність розділення території України за однорідністю температурно-вологісного режиму у березні, квітні й травні на дві частини – Лівобережну й Правобережну, а у червні, липні й серпні – на три частини: східну (Лівобережну), північно-західну й західну степову. Таке виділення східного режиму зволоження пояснюється особливостями переважаючих над країною циркуляційних атмосферних процесів. В цей час східна частина України більш часто знаходиться під дією антициклонічного баричного поля, що приводить до більш низького температурного фону і меншої кількості опадів, ніж в західній частині країни. В липні циркуляційний режим території України однорідний. Наслідком чого і є однорідний температурно-вологісний режим території, що відбивається на однорідному розподілі вагів узагального фактору температури, вологості й опадів. Звертає на себе увагу той факт, що окрім західних областей (Львів, Чернівці) дисперсії залишків в липні суттєво менші, ніж в березні, що свідчить про значно меншу мінливість метеорологічних величин влітку, ніж в березні на більшості території України.

Після отриманих вагів факторів у змінних, що розглядаються, за допомогою рівняння (3.31) треба отримати значення факторів. Статистичний аналіз часових рядів факторів дає можливість отримати цікаві характеристики цих випадкових процесів.

Узагальнені фактори дають можливість виконати аналіз предикторів, які включаються у статистичну модель метеорологічного прогнозу. Добре відомо, що погоду Європи обумовлюють процеси, які розвиваються над Північною Атлантикою. Їх інтенсивність у великій мірі залежить від взаємодії атмосфери з поверхнею океану. Особливості цієї взаємодії характеризують наступні фізичні величини: температура повітря над поверхнею океану (T), температура океанічної поверхні (T_e), різниця цих температур (ΔT) і швидкість вітру (U). Всі означені параметри вимірювались на мережі «кораблів погоди», яка була за міжнародною

угодою створена в Атлантичному океані у 1953 р. і функціонувала до 1980р. Мережа в Атлантичному океані складалася з 9 кораблів, які постійно розташовувались в точках океану, показаних на рис. 3.1.

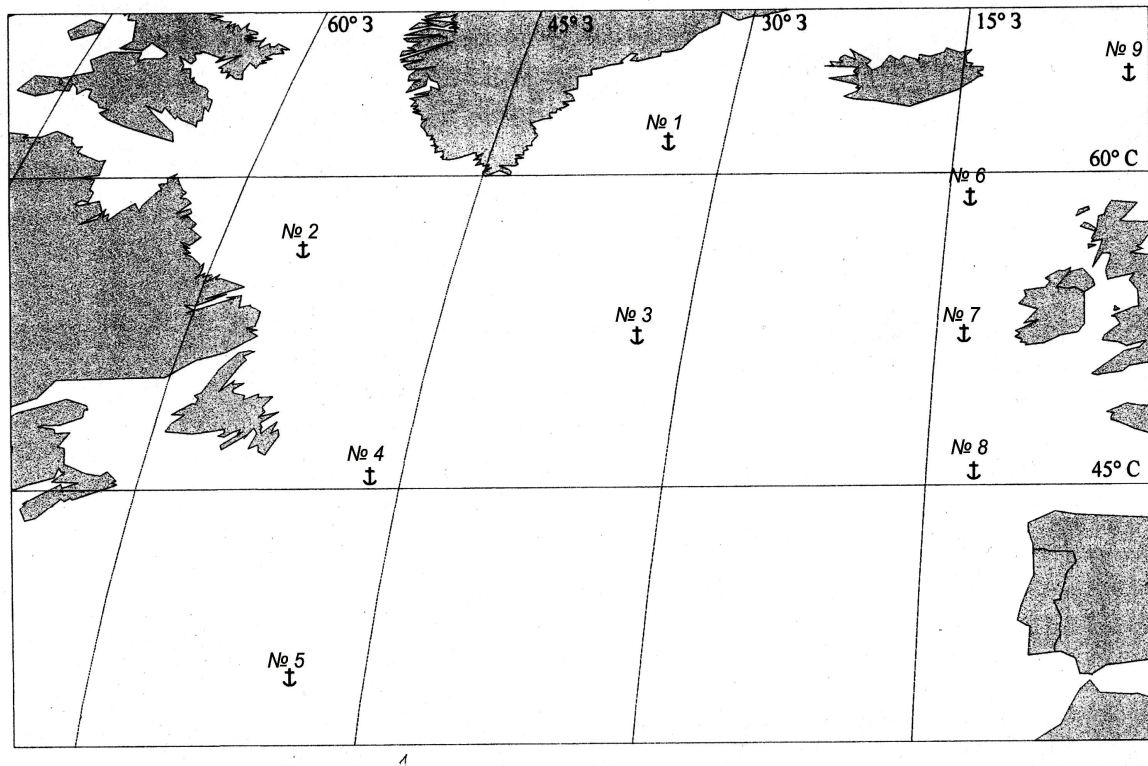


Рис. 3.1 Схема розташування кораблів погоди в північній Атлантиці

Положення «кораблів погоди» було вибрано таким чином, щоб вимірювання зазначених вище характеристик атмосфери і океану приходилися на акваторії енергетично активних зон океану. Результати вимірювань дали можливість створити для кожного пункту вимірювань статистичні ряди середньомісячних значень T , T_e , ΔT і U . Ряди були розширені даними із масиву ERA-40 до 1992 р. Ці параметри є безпосередніми характеристиками тепло і волого обміну між атмосферою і океаном. Тому перед включенням їх у статистичну модель у якості предикторів за допомогою факторного аналізу був проведений їх статистичний аналіз з метою стиску інформації.

Оскільки матриця коваріацій, що розрахована на основі означених вище фізичних параметрів, має четвертий порядок, була за допомогою χ^2 критерію перевірена гіпотеза про кількість факторів. Всі розрахунки проводилися для грудня. Перевірка гіпотези показала, що з ймовірністю 90% дисперсії температур повітря й поверхневих вод океану, різниці температур й швидкості вітру у районах розташування кораблів №№1-3, 5-9 обумовлюють два фактори F_1 і F_2 , а в районі корабля №4 тільки один

фактор F_1 . Ваги узагальнених факторів у змінних представлені у таблиці 3.2.

Таблиця 3.2 - Ваги загальних факторів й фізичних змінних. Грудень

Фізична змінна	Фактор	Номер корабля погоди								
		1	2	3	4	5	6	7	8	9
ΔT	F_1	0,89	0,95	0,94	0,80	0,16	0,8	0,99	-0,07	0,95
	F_2	-0,33	-0,02	-0,24	-	-0,87	-0,45	-0,04	-0,97	-0,11
T	F_1	0,92	0,86	0,82	0,97	0,91	0,96	0,81	0,57	0,99
	F_2	0,37	0,45	0,50	-	-0,40	0,03	0,56	-0,81	0,1
T_e	F_1	0,24	0,06	0,07	0,65	0,94	0,02	0,06	0,88	0,52
	F_2	0,88	0,99	0,99	-	0,23	0,93	0,91	-0,18	0,62
u	F_1	0,18	-0,78	-0,69	-0,77	0,07	0,21	-0,12	-0,74	-0,16
	F_2	-0,69	-0,02	-0,25	-	0,7	-0,50	-0,72	-0,00	0,88

Із табл. 3.2 випливає, що проявляються певні закономірності у розподіленні вагів факторів. Великі додатні ваги у різниці температур ΔT і у температурі повітря T , перший фактор проявляється у районах Північної Атлантики, які відносяться до середніх й високих широт (кораблі № 1, 2, 3, 6, 7) тобто там, де спостерігаються найбільш значні контрасти між температурами поверхні океану й повітря і, як наслідок, де процеси обміну теплом й вологою між океаном і атмосферою найбільш інтенсивні. Температура води T_e характеризується головним чином другим фактором F_2 . У центральній частині Північної Атлантики, де циркуляційні процеси в атмосфері й в океані обумовлені великими антициклонічними кругообертанням (кораблі №5 і №8), найбільшу вагу з від'ємним знаком в різниці температур вносить другий фактор F_2 , а у температуру води T_e – перший фактор F_1 , але вже з додатнім знаком. У швидкості вітру із значною вагою входить другий фактор F_2 (кораблі №1, 5, 7, 9).

Наведені приклади показують, що застосування факторного аналізу до різних метеорологічних задач дає дослідникам дуже важливу інформацію, яка може використовуватися в задачах кліматичного районування, побудова статистичних моделей метеорологічних прогнозів і т.п.

Запитання для самоперевірки

1. На що орієнтований факторний аналіз?
2. У чому полягає відмінність факторного аналізу від компонентного?
3. Що називають вагою фактора?
4. Що являють собою залишки u_i ?
5. Який фізичний сенс діагональних елементів матриці коваріацій?
6. Які два методи використовують для визначення факторних вагів p_{ij} і дисперсій залишків d_i ?
7. У чому полягає метод максимальної правдоподібності?
8. Функцією яких параметрів є функція правдоподібності?
9. Скільки факторів може існувати для коваріаційної матриці третього порядку?
10. За допомогою якого критерію перевіряється гіпотеза H_0 ?

4 КЛАСТЕРНИЙ АНАЛІЗ

4.1 Загальні теоретичні положення кластерного аналізу

Перше застосування кластерний аналіз знайшов у соціології. Назва кластерний аналіз походить від англійського слова cluster - гроно, скупчення. Вперше в 1939 був визначений предмет кластерного аналізу і зроблено його опис дослідником Тріон. Головне **призначення кластерного аналізу** - розбивка множини досліджуваних об'єктів і ознак на однорідні, у відповідному розумінні, групи або кластери. Це означає, що вирішується завдання класифікації даних і виявлення відповідної структури в ній. Методи кластерного аналізу можна застосовувати в самих різних випадках, навіть у тих випадках, коли мова йде про просте згрупування, в якому все зводиться до утворення груп за кількісною подібністю. Велика заслуга кластерного аналізу в тому, що він дозволяє проводити розбиття об'єктів не за одним параметром, а за цілим набором ознак. Крім того, кластерний аналіз на відміну від більшості математико-статистичних методів не накладає ніяких обмежень на вид розглянутих об'єктів і дозволяє розглядати безліч вихідних даних.

Кластерний аналіз дозволяє розглядати досить великий об'єм інформації і різко скорочувати, стискати великі масиви різних даних, робити їх компактними і наочними. Важливе значення кластерний аналіз має стосовно до сукупностей часових рядів, що характеризують фізичні процеси. Тут можна виділяти періоди, коли значення відповідних показників були досить близькими, а також визначати групи часових рядів, динаміка яких найбільш схожа.

Кластерний аналіз можна використовувати циклічно. У цьому випадку дослідження проводиться до тих пір, поки не будуть досягнуті необхідні результати. При цьому кожен цикл тут може давати інформацію, яка здатна сильно змінити спрямованість і підходи подальшого застосування кластерного аналізу. Цей процес можна представити системою оберненого зв'язку. В задачах прогнозування досить перспективно поєднання кластерного аналізу з іншими кількісними методами (наприклад, з регресійним аналізом).

Як і будь-який інший метод, кластерний аналіз має певні недоліки і обмеження: зокрема, склад і кількість кластерів залежить від обраних критеріїв розбиття. При зведенні вихідного масиву даних до більш компактного вигляду можуть виникати певні спотворення, а також можуть втрачатися індивідуальні риси окремих об'єктів за рахунок заміни їх характеристиками узагальнених значень параметрів кластера. При

проведенні класифікації об'єктів ігнорується дуже часто можливість відсутності в розглянутій сукупності будь-яких значень кластерів.

Існує безліч літератури, де докладно описані різні методи проведення кластерного аналізу. Однак ідеального алгоритму не існує і потенційно не може існувати. Можна створити досить якісний алгоритм, який буде давати добре обґрунтовані результати, наприклад, в медицині, але при застосуванні в інших галузях науки, отримані результати можуть бути сумнівними.

Для більш глибокого розуміння суті кластерного аналізу розглянемо класифікацію на основі мінімуму функції відстані.

Розглянемо вектори, компоненти яких характеризують ті чи інші параметри фізичного стану атмосфери, або ту чи іншу ситуацію, що склалася в атмосфері. Будемо називати ці вектори – **векторами ситуації** або образами. Припустимо також, що по деяких властивостях вектори ситуацій можна об'єднати у класи. Класи векторів, що мають деяку схожість називають **кластерами**. Серед векторів ситуацій кластера є такий образ, котрий є репрезентативним для усього кластера, тобто має деякі риси, притаманні всім образам цього кластера. Цей вектор (образ) називають **еталоном або центром кластера**. В деяких випадках кластер може мати декілька еталонів. Задача класифікації складається, по-суті, з двох задач: а) побудови правила, за яким можна визначити еталон і, таким чином, сформувати кластер; б) визначити міру близькості некластеризованого вектора до того чи іншого кластера.

Природною мірою близькості є відстань між некластеризованим вектором і кластером. За цією мірою, очевидно, вектор X_1 (рис.4.1) треба віднести до кластера V_i ($X_1 \in V_i$). Видимих підстав для віднесення вектора X_2 до кластерів V_i чи V_j нема, але це можна зробити, якщо розглянути відстані між вектором X_2 і центрами кластерів, які позначаються жирною крапкою на рис. 4.1.

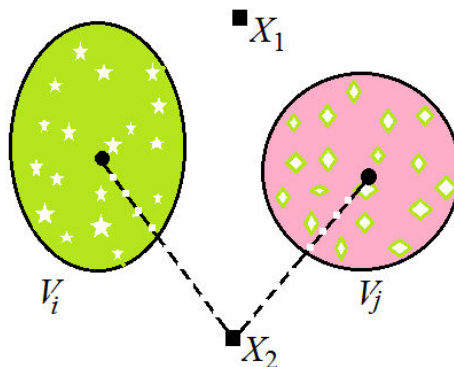


Рис. 4.1 – До поняття кластеру

Будемо відносити некластеризований вектор до того кластера, деяка функція відстані від котрого до центра цього кластера є найменшою порівняно з аналогічною функцією відстані до центрів інших кластерів. Це означає, що кластеризація проводиться на основі мінімуму функції відстані. Розглянемо два випадки:

а) Випадок єдиного еталону в кластері.

Будемо вважати, що в кластері тільки один еталон, або центр кластера. Знайдемо відстань між вектором X і центром Z_i i -того кластера U_i . У якості відстані будемо розглядати евклідову відстань D_i

$$D_i = \|X - Z_i\| = \sqrt{(X - Z_i)'(X - Z_i)}, \quad (4.1)$$

де вертикальними дужками позначається норма вектора.

Очевидно, коли

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_m \end{pmatrix} \quad (4.2)$$

і

$$Z_i = \begin{pmatrix} z_{1i} \\ z_{2i} \\ \dots \\ z_{mi} \end{pmatrix}, \quad (4.3)$$

тоді

$$D_i = \sqrt{(x_1 - z_{1i})^2 + (x_2 - z_{2i})^2 + \dots + (x_m - z_{mi})^2}. \quad (4.4)$$

Припустимо, що ми маємо M кластерів U_i ($i = \overline{1, M}$), кожний з яких визначається єдиним еталоном $Z_1, Z_2, \dots, Z_i, \dots, Z_j, \dots, Z_M$. Тоді, очевидно, $X \in U_i$, якщо $D_i < D_j \quad \forall j = \overline{1, M}, j \neq i$.

Можна розглянути й іншу функцію відстані, а саме:

$$D_i^2 = |X - Z_i|^2 = (X - Z_i)'(X - Z_i). \quad (4.5)$$

В матричному співвідношенні (4.5) виконаємо операцію множення:

$$D_i^2 = X'X - Z_i'X - X'Z_i + Z_i'Z_i. \quad (4.6)$$

Оскільки $Z_i'X = X'Z_i$, маємо:

$$D_i^2 = X'X + Z_i'Z_i - 2X'Z_i \quad (4.7)$$

або

$$D_i^2 = X'X - 2(X'Z_i - \frac{1}{2}Z_i'Z_i). \quad (4.8)$$

Впровадимо позначення:

$$d_i(X) = X'Z_i - \frac{1}{2}Z_i'Z_i. \quad (4.9)$$

Із формул (4.8) і (4.9) видно, що функція D_i^2 буде мінімальною, якщо функція $d_i(X)$ є максимальною. Такий же висновок відноситься і до евклідової відстані D_i . Отже, розв'язувальне правило має такий вигляд:

$$X \in U_i, \text{ якщо } d_i(X) > d_j(X), \forall j = \overline{1, M}; i \neq j. \quad (4.10)$$

Функція відстані $d_i(X)$ є лінійною формою відносно координат вектора X з коефіцієнтами - координатами відомого вектора Z_i тобто:

$$d_i(X) = x_1z_{1i} + x_2z_{2i} + \dots + x_mz_{mi} - \frac{1}{2} \sum_{s=1}^m z_{si}^2. \quad (4.11)$$

б) Випадок декількох еталонів у кластері.

Нехай маємо M кластерів, але кожний з них характеризується числом еталонів, тобто для i - того кластера U_i маємо еталони $z_i^{(1)}, z_i^{(2)}, \dots, z_i^{(p)}$. Тоді знайдемо евклідову відстань між вектором X і кожним з центрів i - того кластера. Виберемо з них ту, яка є найменшою, нехай це буде відстань до l - того центру $z_i^{(l)}$, тобто:

$$D_i = \min_l \left| X - z_i^{(K)} \right| \quad (K = 1, 2, \dots, l, \dots, p). \quad (4.12)$$

Теж саме зробимо для останніх кластерів. Тоді розв'язувальним правилом є

$$X \in U_i, \text{ якщо } D_i < D_j \quad \forall j = \overline{1, M}; (j \neq i). \quad (4.13)$$

Це розв'язувальне правило можна записати й для функції відстані

$$d_i(X) = \max_l \left\{ X' z_i^{(K)} - \frac{1}{2} \left(z_i^{(K)} \right)' z_i^{(K)} \right\}, \quad (K = 1, 2, \dots, l, \dots, p). \quad (4.14)$$

Воно має вид:

$$X \in U_i, \text{ якщо } d_i(X) > d_j(X) \quad \forall j = \overline{1, M}; (j \neq i). \quad (4.15)$$

Критерій мінімуму евклідової відстані (максимуму функції відстані) знаходить найбільш часте застосування в алгоритмах класифікації. Але треба мати на увазі, що встановлення критерія близькості, за допомогою якого ми можемо віднести некластеризований вектор X до одного з уже відомих кластерів, ще не вирішує задачу кластеризації. Необхідно, крім того, встановити правило або систему правил, які в системі даних дозволять ідентифікувати кожний з центрів кластерів і, таким чином, сформулювати кластери. Такі правила складають *міри схожості* образів і критерії кластеризації. Розрізняють метричні і неметричні міри схожості. Однією з метричних мір схожості може розглядатися вже знайома нам метрика - евклідова відстань $D = |X - Z|$. Іншою мірою схожості є число *Махаланобіса*:

$$\Delta^2 = (X - \mu)' K^{-1} (X - \mu), \quad (4.16)$$

де

μ і K - відповідно вектор математичних сподівань і матриця коваріацій.

Звичайно, останню міру схожості можна використовувати тільки тоді, коли нам відомі ймовірнісні характеристики образів (вектори математичних сподівань і матриця коваріацій) для кожного з кластерів. Вона є дуже зручною для вирішення задачі віднесення некластеризованого вектора до того чи іншого кластера, але не може використовуватися для

формування кластерів, якщо апіорі невідомі зазначені ймовірнісні характеристики.

Окрім метричних використовуються й неметричні міри схожості. Однією з них може бути міра, що визначається таким співвідношенням:

$$S(X, Z) = \frac{X'Z}{|X| \cdot |Z|}. \quad (4.17)$$

У складній формі рівність (5.17) має вигляд:

$$S(X, Z) = \frac{x_1 z_1 + x_2 z_2 + \dots + x_m z_m}{\sqrt{x_1^2 + x_2^2 + \dots + x_m^2} \sqrt{z_1^2 + z_2^2 + \dots + z_m^2}}. \quad (4.18)$$

Вираз (4.18), очевидно, має сенс *косінуса кута між векторами* X і Z , або *коефіцієнта кореляції* між випадковими величинами X і Z , що визначені відповідними векторами:

$$S(X, Z) = \cos \left(\hat{X, Z} \right) = r_{xz}. \quad (4.19)$$

Звичайно, ця міра ніякого відношення до метричних критеріїв не має. Існують й інші непараметричні міри схожості.

Вибір міри схожості ще не вирішує проблему кластеризації, оскільки треба ще визначити критерій, за яким відбувається розбиття простору даних на кластери, тобто необхідно ще визначити критерії кластеризації. Тут можливо використовувати два підходи. Перший з них є *евристичним*. В його основу вкладається досвід та інтуїція. Інтуїтивно зрозуміло, що критерієм близькості образів у кластері може бути евклідова відстань. Але цей критерій є відносним. Його треба доповнювати **порогом** - деякою величиною евклідової відстані, починаючи з якої треба образ відносити до того чи іншого кластера. Іншим підходом є підхід, що має в основі деякий критерій якості, який в залежності від характеру задачі підлягає мінімізації або максимізації. Прикладом такого критерія є сума квадратів відхилів векторів від деякого середнього вектора кластера

$$\Delta = \sum_{j=1}^{N_c} \sum_{X \in S_j} |X - m_j|^2, \quad (4.20)$$

де

$m_i = \frac{1}{N_i}$ – середній вектор i - того кластера;

N_c – кількість кластерів;

S_i – простір векторів, що відносяться до i - того кластера;

N_i – кількість векторів i - того кластера.

У 2003 році на кафедрі фізики атмосфери та кліматології був розроблений алгоритм УАІМКА (Універсальний **адаптивний** ітераційний метод кластерного аналізу), який добре показав себе при кластеризації територій, порівнянних з територією України, але при застосуванні його для більш великих просторів виникли деякі труднощі, пов'язані з виявленням невеликої кількості великих кластерів з потенційно неоднорідними районами. Тому початкові кластери доводилося «дробити», за допомогою цього ж методу, після чого можна було якісно обґрунтувати, з фізичної точки зору, отриману кластеризацію. Крім усього, іноді виявляються значущі коефіцієнти кореляції, які використовувалися в якості критерію в УАІМКА, могли б викликати неоднозначне тлумачення. Крім того, використання внутрикластерних і міжкластерних дисперсій для визначення відмінностей або ж критеріїв при кластеризації гідрометеорологічних даних не завжди доречно. Відповідно і критерій Фішера (відношення сумарної дисперсії до зменшеної за ранжированим рядом або ж відношення дисперсій двох випадкових величин) в якості єдиного, що визначає ті ж самі відмінності. Оскільки залежно від розглянутої характеристики, що знаходиться під впливом одного глобального процесу, можна отримати, наприклад, один кластер, там де їх декілька, що може викликати помилкові судження про фізику процесу.

4.2 Універсальний ітераційний метод кластеризації даних (УІМКД)

В універсальному ітераційному методі кластеризації даних застосований дещо інший підхід до критеріїв кластеризації. Вони акцентують увагу на виявленні неоднорідності кластерів за допомогою відомих параметричних і непараметричних критеріїв Фішера, Крамера-Уелча і Лемана-Розеблатта, застосування яких в математичній статистиці до незалежних безперервних випадкових величин з невідомими законами розподілу є найбільш аргументованим.

Відомо, що найвища ступінь однорідності досягається, якщо обидві вибірки взяті з однієї генеральної сукупності, тобто справедлива нульова гіпотеза $H_0: F(x) = G(x)$ при всіх x . Відсутність однорідності означає, що приймається альтернативна гіпотеза: $H_1: F(x) \neq G(x)$, хоча б при одному

значенні аргументу x . Якщо гіпотеза H_0 прийнята, то вибірки можна об'єднати в одну, якщо ні, то неможна.

Перш ніж приступити більш докладно до розгляду згаданих критеріїв, необхідно сказати, що гідрометеорологічні величини в більшості випадків не підкоряються нормальному закону розподілу і є незалежними безперервними випадковими величинами.

Наведемо кінцеві формули критеріїв Фішера і Крамера-Уелча, що їх визначають:

а) критерій Фішера: $F = \frac{S_x^2}{S_y^2};$

б) критерій (статистика) Крамера-Уелча: $T = \frac{\sqrt{mn}(\bar{x} - \bar{y})}{\sqrt{nS_x^2 + mS_y^2}},$

де

S_x^2 – не зсунена – оцінка дисперсії випадкової величини x ;

S_y^2 – не зсунена оцінка дисперсії випадкової величини y ;

$\bar{x}$ і $\bar{y}$ – середні значення випадкових величин x , y ;

m і n – об'єми вибірок випадкових величин x і y .

Необхідно зазначити, що застосування критерію Крамера-Уелча не менш обґрунтоване, ніж застосування критерію Стьюдента. Додаткова перевага – не потребує рівності дисперсій S_x^2 и S_y^2 . Тому, для перевірки однорідності математичних сподівань (гіпотеза H_0) доцільно застосовувати критерій Крамера-Уелча.

Критерії перевірки гіпотез про дисперсії на відміну від гіпотез про середні вельми чутливі до будь-яких відхилень від припущень, в умовах яких вони були отримані. І також відсутня або суперечлива інформація щодо потужності відповідних критеріїв.

Невідхилення перевіряємих гіпотез про рівність середніх і (або) рівність дисперсій ще не говорить про приналежність вибірок до однієї і тієї ж генеральної сукупності. Це свідчить лише про можливу рівність числових характеристик, але не законів розподілу. Вибір же критеріїв перевірки гіпотез щодо законів розподілу, що відповідають двом вибірками, більш скромний. Як правило, на практиці використовується або критерій Смірнова, або критерій Лемана-Розенблатта. Для критерію типу омега-квадрат (ω^2) немає вираженого ефекту відмінності між номінальними і реальними рівнями значущості. Тому рекомендовано для перевірки однорідності функцій розподілу (гіпотеза H_0) застосовувати

статистику типу A омега-квадрат, а за відсутності методичного, табличного або програмного забезпечення для статистики Лемана-Розенблатта, рекомендовано використовувати критерій Смірнова.

Статистика критерію типу омега-квадрат для перевірки однорідності двох незалежних вибірок має вигляд:

$$A = \frac{mn}{m+n} \int_{-\infty}^{\infty} (F_m(x) - G_m(x))^2 dH_{m+n}(x),$$

де

$H_{m+n}(x)$ – емпірична функція розподілу, побудована для об'єднаної вибірки:

$$H_{m+n}(x) = \frac{m}{m+n} F_m(x) + \frac{n}{m+n} G_n(x).$$

Значення статистики залежить лише від рангів елементів вибірки.

Статистика A типу омега-квадрат (ω^2) була запропонована Е. Леманом в 1951 р., вивчена М. Розенблаттом в 1952 р., а потім і іншими дослідниками. Вона залежить лише від рангів елементів двох вибірок в об'єднаній вибірці. Нехай $x_1, x_2, x_3, \dots, x_m$ – перша вибірка, $x'_1 < x'_2 < x'_3 < \dots < x'_m$ – відповідний варіаційний ряд, $y_1, y_2, y_3, \dots, y_n$ – друга вибірка, $y'_1 < y'_2 < y'_3 < \dots < y'_n$ – варіаційний ряд, що відповідає другій вибірці. Оскільки функції розподілу незалежних вибірок неперервні, то з ймовірністю $p=1$ всі вибіркові значення різні, збіги відсутні. Статистика A представляється у вигляді:

$$A = \omega^2 = \frac{1}{mn(m+n)} \left[m \sum_{i=1}^m (r_i - i)^2 + n \sum_{j=1}^n (s_j - j)^2 \right] - \frac{4mn-1}{6(m+n)},$$

де

r_i – ранг x'_i и s_j – ранг y'_j в загальному варіаційному ряді, побудованому для об'єднаної вибірки.

Правила прийняття рішень при перевірці однорідності двох вибірок на основі статистики типу омега-квадрат (ω^2), так само як і статистики Смірнова, на основі критичних значень залежно від рівнів значущості та обсягів сукупностей випадкових величин наведені в таблицях Смірнова.

У методі (УІМКД), в якості вихідної інформації виступає матриця $X = (x_{ij})_{n \times m}$, що містить n векторів-рядків розмірності m , що

характеризує статистичні ряди об'ємом m у n пунктах, які і повинні бути кластеризовані. В якості апріорної інформації, на відміну від інших методів, так само як і в УАІМКА, задається тільки мінімальна кількість векторів τ , які можуть скласти кластер.

Ітераційний процес в алгоритмі УІМКД складається з ряду кроків:

1-й крок: Розраховується квадратна матриця порядку n евклідових відстаней між усіма векторами матриці X :

$$D = (D_{ij})_{n \times n}, \quad D_{ij} = \sqrt{\sum_{s=1}^n (x_{js} - x_{is})^2}. \quad (4.21)$$

Матриця D є симетричною. На головній діагоналі цієї матриці розташовуються нулі.

2-й крок: У кожному рядку матриці D здійснюється ранжування її елементів, тобто її елементи розташовуються в порядку зростання. В результаті отримаємо матрицю D^1 вигляду:

$$D^1 = \begin{pmatrix} 0 & d_{12}^{(p)} & d_{13}^{(p)} & d_{14}^{(p)} & \dots & d_{1n}^{(p)} \\ 0 & d_{22}^{(p)} & d_{23}^{(p)} & d_{24}^{(p)} & \dots & d_{2n}^{(p)} \\ 0 & d_{32}^{(p)} & d_{33}^{(p)} & d_{34}^{(p)} & \dots & d_{3n}^{(p)} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & d_{n2}^{(p)} & d_{n3}^{(p)} & d_{n4}^{(p)} & \dots & d_{nn}^{(p)} \end{pmatrix}, \quad (4.22)$$

в якій проводиться перенумерація елементів кожного рядка (p - номер стовпчика, в якому розташовується елемент кожного рядка матриці (4.22) D ($p = \overline{1, n}$)).

Припустимо, що мінімальне число векторів, які можуть скласти кластер дорівнює τ . Тоді аналізу піддається блок матриці (4.22), що складається з перших $\tau = 3$ стовбців (на першому етапі, наприклад $3 = 3$).

3-й крок: Проводиться порівняння евклідових відстаней D_{lj} ($l, j = \overline{1, n}$) з евклідовими відстанями $d_{l3}^{(p)}$. Якщо, $D_{lj} \leq d_{l3}^{(p)}$ ($l = \overline{1, n}$), тоді j -й вектор може розглядатися як потенційний центр кластера, до якого входить l -й вектор (з урахуванням значення індексу p).

4-й крок: Для кожного такого j -го вектора визначається кількість l -х векторів (число входжень S_j), для яких він може бути центром кластеру.

5-й крок: Із загального числа j -х векторів виділяються ті, для яких $S_j \geq \tau$ (у нашому прикладі $S_j \geq 3$). Інші потенційні центри кластерів, для яких ця умова не виконується, ліквідуються.

6-й крок: Визначається число залишених j -х векторів як центрів кластерів. Нехай їх число дорівнює r ($j = r$).

7-й крок: Із загального числа m -х векторів встановлюються ті, які потрапили в s -й (V_s) і q -й (V_q) кластери одночасно ($s, q = \overline{1, r}; l \neq s, q$) і проводиться їх розведення за кластерами і відповідно з вирішальним правилом: $X_l \in V_s$, якщо $D_{ls} < D_{lq}$, при цьому $S_q = S_q - 1$.

8-й крок: Знаходяться евклідові відстані $\tilde{D}_{sq}$ між центрами s -го (V_s) і q -го (V_q) кластерів.

9-й крок: Знаходиться максимальне з відстаней між векторами, що входять в s -й кластер D_{ts} та q -й кластер D_{fq} ($t = \overline{1, s_s}; f = \overline{1, s_q}$). Припустимо це буде D_{fq} .

10-й крок: Евклідова відстань D_{fq} порівнюється з відстанями між центрами кластерів $\tilde{D}_{sq}$. Якщо $\tilde{D}_{sq} < D_{fq}$, тоді за умови $S_s \geq S_q$ ліквідується q -й кластер. Якщо $S_q > S_s$, то ліквідується s -й кластер.

11-й крок: Виробляється формування низки попередніх центрів кластерів z_j .

12-й крок: Проводиться розподіл за кластерами векторів вихідної вибірки відповідно з вирішальним правилом $X \in V_j$ якщо $D_{xz_j} < D_{xz_k}$.

13-й крок: визначається кількість векторів S_j , що входить у кожний j -й кластер V_j .

14-й крок: Розраховуються середні вектори для кожного j -го кластера R_j .

15-й крок: Для кожної пари s -го і q -го кластерів на базі R_j визначаються значення критерію Фішера F_j і критерію Стюдента t_j (статистика Крамера-Уелча).

16-й крок: Отримані значення критерію Фішера F_j і критерію Стюдента t_j порівнюються з критичними значеннями на рівні значущості $\alpha = 0,05$ $F_{кр}$ і $t_{кр}$.

17-й крок: Якщо всі j -і кластери на даному кроці виявляються неоднорідними по відношенню один до одного, проводиться додаткова

перевірка на однорідність за допомогою розрахунку критерію типу омега-квадрат (Лемана-Роземблатта) ω^2 і порівняння його з критичним значенням $\omega_{кр}^2$ на рівні значущості $\alpha = 0,05$.

18-й крок: За умови $\omega^2 > \omega_{кр}^2$ повертаємося до кроку 15 і ітераційна процедура продовжується знову, тільки вже для початкових центрів отриманих кластерів z_j аж до 17-го кроку включно.

19-й крок: Повертаємось до шагу 2 за умови $\vartheta = \vartheta + 1$, і ітераційна процедура триває знову аж до 18-го кроку включно.

20-й крок: У разі виявлення повної неоднорідності середніх векторів R_j і початкових центрів z_j кластерів, отриманих на етапі ϑ і етапі $\vartheta - g$, ($g \geq 1$) (згідно вищезазначеним критеріям), проводиться порівняння числа кластерів $N_{\vartheta}, N_{\vartheta-g}$ на відповідних етапах, за умови $N_{\vartheta-g} \geq 2$ и $N_{\vartheta} \geq 2$. Якщо $N_{\vartheta} \geq N_{\vartheta-g}$, тоді ϑ етап вважається основним, і кількість неоднорідних кластерів, отриманих на наступних етапах, порівнюється з N_{ϑ} . Якщо $N_{\vartheta-g} > N_{\vartheta}$, тоді $\vartheta - g$ вважається основним етапом.

21-й крок: Процедура повторюється, починаючи з кроку 2 по 20 крок до тих пір, доки $g \leq m$.

4.3 Приклад застосування кластерного аналізу гідрометеорологічної інформації

Як приклад застосування алгоритму УІМКД можна навести кластеризацію полів середньомісячної температури підстильної поверхні. Для вихідної вибірки були відібрані дані ре-аналізу ERA-40, задані у вузлах регулярної сітки точок $2,5^\circ \times 2,5^\circ$, за період з 1958 по 2002 роки, в районі північної частини Тихого океану (в секторі, обмеженому за широтою від 20° до 70° північної широти і меридіанально від 160° східної довготи до 120° західної довготи), для грудня.

Таким чином, в кожному вузлі сітки був сформований 45-ти мірний вектор середньомісячних значень зазначеної характеристики для грудня. Множину цих векторів і було представлено алгоритмом УІМКД для розбиття полів температури повітря на однорідні кластери.

Метою даного прикладу було отримання максимально наближене до об'єктивного районування полів температур, що повинно, проявиться у відповідності виділених відносно незалежних температурних зон побудованих середньомісячних карт основних характеристик, що

відображають властивості підстильної поверхні і умови взаємодії в нижньому шарі атмосфери розглянутих територій.

Результати проведеної кластеризації представлені на рис. 4.2, де літерними позначеннями латиниці виділені виявлені алгоритмом кластери (використовувалася програма "Surfer").

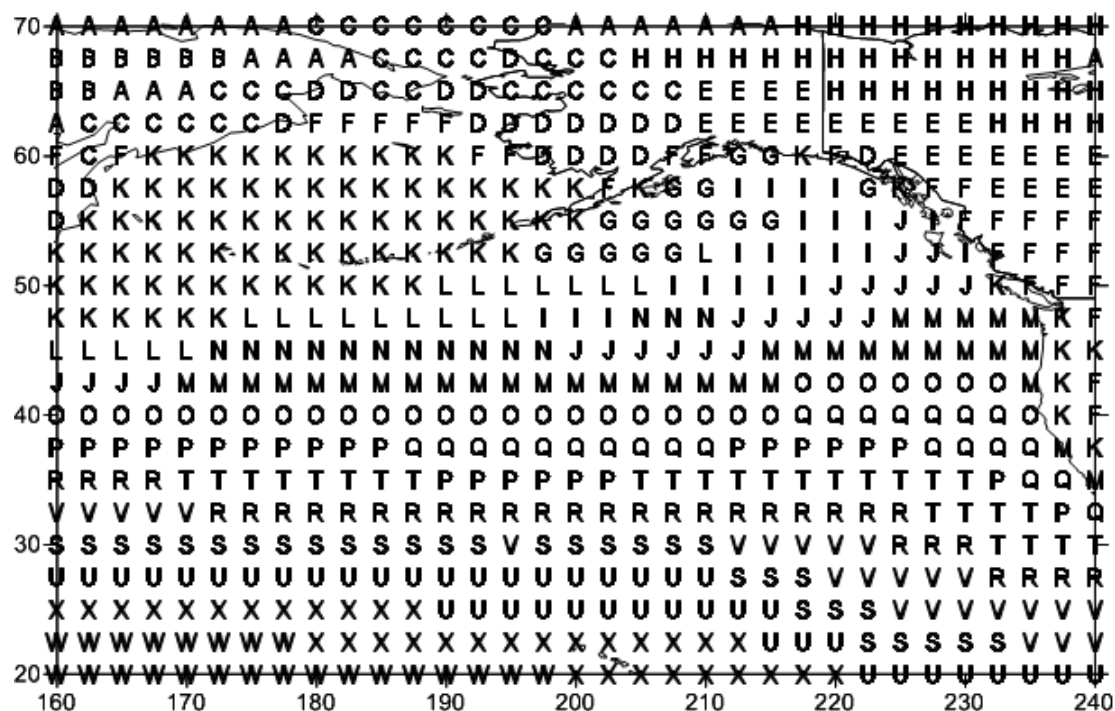


Рис. 4.2 - Карта розподілу кластерів середньомісячної температури підстильної поверхні в північній частині Тихого океану (грудень)

Відомо, що найбільш важливою закономірністю визначається не тільки термічний режим поверхневих вод, а й надається дуже великий вплив на зміну температури по вертикалі, є саме характер зміни температури води на поверхні океану в залежності від широти, який для північної частини Тихого океану формується під дією процесів, що визначають розподіл поверхневої температури води з широтою (різну кількість сонячної радіації, що поглинається діяльним шаром океану і процесів теплообміну, що відбуваються в ньому, різні умови обміну з атмосферою теплом і кількістю руху, танення та утворення льодів в районі Берингова моря) і має загальну зональну спрямованість, що порушується, в основному, тільки в районі розмежувальної лінії вода-суходіл.

Доцільно вказати, що в грудні добре проявляються в полі температури кластери циклонічної і антициклонної циркуляції вод океану. На користь об'єктивності проведеної кластеризації говорить і досить чіткий поділ у розглянутий період сукупності кластерів розподілу температури поверхонь суші (наявність льоду) і океану (наприклад, для суші кластери А, В, С, D, Е, Н, F (рис. 4.2)).

На схематичній карті розглянутого зимового місяця проявляється найбільший кластер (К). Він досить чітко визначає розташування Берингова моря та прилеглої до нього території. Основний кліматичний фон тут формується переважно циркуляційними чинниками, тобто морським циклонічним кругообігом і перенесенням повітряних мас, а не радіаційними складовими, що і призводить до утворення однорідної області.

Найбільшу акваторію північній частині Тихого океану (від екватора до району 40° півн.ш.) охоплює океанський північний субтропічний антициклонічний кругообіг, який, в нашому випадку, можна простежити за допомогою структурного розташування кластерів О, Р, Q, R, S, T, U, V, W, X. Зазначені кластери добре пов'язані з проходженням таких течій як: Північно-Пасатне (U, W, X), Тайванське (W), Курошіо (V, R, P), Північно-Тихоокеанське (О, Q) і Каліфорнійське (Q, T). З південного боку кругообіг обмежений північним тропічним океанічним фронтом, а з півночі - північним полярним. Конструктивне розташування кластерів в зоні зазначеного кругообігу відповідає цим фронтам і підкреслює зональні і меридіональні особливості циркуляційних процесів описаної вище області Тихого океану.

Структура кластерів на північ від субтропічного кругообігу (приблизна область від 40° півн.ш. до 60° півн.ш.) досить добре відбиває розташування субполярного циклонічного кругообігу. Цей кругообіг охоплює цілий ряд течій, таких як: Алеутських (на схематичній карті відповідає кластерам I, J, M, N), представляє собою північну гілку Північно-Тихоокеанської течії, Аляскінське (G), що огинає Аляскінську затоку проти годинникової стрілки, Камчатське (F) і Курильські холодні течії.

Із заходу на схід в районі 40° півн.ш. простягається лінія конвергенції між двома різними океанічними циркуляційними процесами, яка поширюючись на південний схід, перетворюється в зону позитивної горизонтальної дивергенції. Цю особливість розподілу температури поверхні води можна добре простежити на прикладі груп кластерів на представленій карті: M, O, Q, R, T (рис. 4.2), де чітко простежується розбіжність потоків на північ і південь. Цікаво, що межі між основними кластерними утвореннями чітко відповідають фізично обґрунтованим фактам.

Порівняння отриманого схематизованого уявлення розподілу середньомісячної температури поверхні води північної частини Тихого океану з існуючими картами метеорологічних і кліматичних характеристик, таких як температури повітря; температури води на поверхні; витрати тепла на випаровування; радіаційний баланс; кількість опадів; турбулентний теплообмін між океаном і атмосферою і т.д.,

представлених у вигляді розподілу ізоліній, показало хорошу узгодженість між отриманими кластерами і однорідними зонами полів розглянутих характеристик.

Статистичний аналіз отриманої кластеризації представлений у вигляді таблиць 4.1, 4.2. Згідно з отриманими результатами внутрікластерні дисперсії (табл. 4.1) для досліджуваної території у всі місяці мають чіткий поділ на дві частини. Великі дисперсії - характеризують материкову поверхню і поверхню моря, покриту льодом, пов'язані, з особливостями характеру цих поверхонь та взаємодії з нижнім шаром атмосфери. Незначні дисперсії - характеризують океанічну частина території, яка визначається великою інерційністю океану. Середні значення температури поверхні мають в основному меридіональну спрямованість, зменшуючись з півдня на північ, і тільки на материковій частині спостерігаються відмінності між береговими частинами і більш видаленими від води. Причому мінімальні середні значення спостерігаються в кластері територіально розташованому в глибині півострова Чукотка, а максимальні - у крайній південно-західній частині, знаходиться під впливом теплого Північно-Пасатної течії.

Таблиця 4.1 – Внутрікластерні дисперсії і середні значення середньомісячної температури підстильної поверхні кластерів в північній частині Тихого океану (грудень)

Утворення кластерів	Внутрікластерні дисперсії	Середні значення	Позначення кластерів	Внутрікластерні дисперсії	Середні значення
A	9.563	247.254	M	0.121	283.523
B	17.211	241.647	N	0.389	280.792
C	13.579	252.505	O	0.2	285.916
D	11.522	259.963	P	0.155	289.186
E	18.026	256.947	Q	0.105	287.799
F	11.846	249.198	R	0.083	292.155
G	2.328	267.685	S	0.083	294.653
H	0.213	278.493	T	0.069	290.583
I	0.227	279.805	U	0.068	296.312
J	0.165	281.775	V	0.09	293.725
K	0.13	276.416	W	0.145	299.364
L	0.212	279.095	X	0.09	297.817

Екстремальні значення евклідових відстаней (як одна з характеристик віддаленості) між кластерами, представлені в табл. 4.2 повністю відповідають кластерам з максимальними і мінімальними середніми значеннями (табл. 4.1).

Таблиця 4.2 - Екстремальні значення евклідових відстаней між кластерами середньомісячної температури підстильної поверхні в північній частині Тихого океану (грудень)

max	min
388.157	5.897
B-W	H-L
(2-23)	(12-8)

Запитання для самоперевірки

1. Яке призначення кластерного аналізу?
2. Що є вектори-ситуацій (образи)?
3. Що таке кластер?
4. Дати поняття репрезентативного образу кластера.
5. Що таке центр кластера або еталон кластера?
6. У чому полягає задача класифікації метеорологічної інформації?
7. Як визначається евклідова відстань між некластеризованим вектором і центром кластера?
8. Що таке функція відстані?
9. Яким є критерій кластеризації при використанні евклідової відстані?
10. Яким є критерій кластеризації при використанні функції відстані?
11. Що таке міра схожості та які міри схожості вам відомі?

5 РЕГРЕСІЙНІ МОДЕЛІ МЕТЕОРОЛОГІЧНОГО ПРОГНОЗУ

5.1 Лінійні моделі множинної регресії

При розв'язанні задач дослідження та прогнозування метеорологічних величин і їх полів необхідно розглядати не тільки особливості їх статистичної структури, але створювати на цій основі відповідні статистичні моделі. Найбільш часто використовуються лінійні моделі. До таких моделей пред'являється ряд вимог. По-перше, модель повинна бути достатньо простою, по-друге, вона повинна у достатній мірі бути адекватною процесу, що прогнозується. Однією з таких моделей є модель авторегресії. У цій моделі поточне значення випадкового процесу виражається як скінчена лінійна комбінація попередніх його значень й випадкового імпульсу. Нехай $x(t-1), x(t-2), \dots, x(t-p)$ є значення випадкового процесу $x(t)$ зі значеннями з дискретністю $t-1, t-2, \dots, t-p$, а $Z(t)$ - білий шум. Якщо виконати відому операцію центрування $\circ$ $\bar{x}(t) = x(t) - m_x$, тоді лінійна форма:

$$\bar{x}(t) = \alpha_1 \bar{x}(t-1) + \alpha_2 \bar{x}(t-2) + \dots + \alpha_p \bar{x}(t-p) + Z(t) \quad (5.1)$$

називається процесом авторегресії порядку p . Модель має $p+2$ параметрів: $m_x, \alpha_1, \alpha_2, \dots, \alpha_p, \sigma_z^2$, де σ_z^2 - дисперсія білого шуму $Z(t)$.

Наступна лінійна модель випадкового процесу є модель ковзного середнього. Вона полягає у тому, що значення процесу $x(t)$ виражаються через попередні значення $z(t-1), z(t-2), \dots, z(t-q)$ білого шуму $Z(t)$, а саме:

$$\bar{x}(t) = Z(t) + \beta_1 z(t-1) + \beta_2 z(t-2) + \dots + \beta_q z(t-q). \quad (5.2)$$

Процес, який відображається рівнянням (4.2), називається процесом ковзного середнього. Модель (4.2) утримує $q+2$ параметра: $m_x, \beta_1, \beta_2, \dots, \beta_q, \sigma_z^2$. У ряді випадків використовуються змішані моделі: авторегресії-ковзного середнього. Вона має такий вигляд:

$$\begin{aligned} \bar{x}(t) = & \alpha_1 \bar{x}(t-1) + \alpha_2 \bar{x}(t-2) + \dots + \alpha_p \bar{x}(t-p) + Z(t) + \\ & + \beta_1 z(t-1) + \beta_2 z(t-2) + \dots + \beta_q z(t-q) \end{aligned} \quad (5.3)$$

Ця модель визначається $p + q + 2$ параметрами: $m_x, \alpha_1, \alpha_2, \dots, \alpha_p, \beta_1, \beta_2, \dots, \beta_q, \sigma_z^2$.

Перевага цих методів полягає у їх економічності. Розглянуті стохастичні моделі мають й суттєві недоліки. По-перше, вони не можуть використовуватись при наявності прогалин у часових рядах. По-друге, модель дає можливість здійснити прогноз тільки на один крок вперед, тобто завчасність прогнозу дорівнює інтервалу дискретності ряду, на якому він базується. По-третє, прогнозуєме значення дуже чутливе до передісторії процесу.

Методи прогнозу метеорологічних величин повинні враховувати не тільки їх передісторію, але й у значній мірі вплив інших факторів, які обумовлюються процесами, що розвиваються в атмосфері (у подальшому, *прогнозні характеристики* будемо називати **предіктантами**, а фактори, від яких вони залежать – **предікторами**). Зазначеним вимогам задовольняють прогностичні моделі, побудовані на основі лінійних множинних рівняннях регресії.

Будемо вважати, що предіктант y й предіктори x_i центровані величини. Тоді рівняння лінійної регресії має вигляд:

$$y = b_1 x_1 + b_2 x_2 + \dots + b_i x_i + \dots + b_n x_n. \quad (5.4)$$

де $b_1, b_2, \dots, b_i, \dots, b_n$ – коефіцієнти (параметри) регресії.

Коефіцієнти регресії повинні бути визначеними на основі статистичних сукупностей предіктанта і предікторів x_i ($i = \overline{1, n}$) – факторів, які впливають на формування величини, що прогнозується.

Розглянемо як розраховуються коефіцієнти регресії на спрощеному рівнянні:

$$y = b_1 x_1 + b_2 x_2. \quad (5.5)$$

Якщо маємо статистичні сукупності предіктанта y і предікторів x_1, x_2 з об'ємом m , тоді для визначення коефіцієнтів b_1, b_2 застосуємо метод найменших квадратів. Його метрика має вигляд:

$$\Delta^2 = \sum_{j=1}^m \left[y_j - (b_1 x_{1j} + b_2 x_{2j}) \right]^2. \quad (5.6)$$

Потрібно отримати такі значення b_1, b_2 , котрі доставляли б мінімуму метрики Δ^2 . Це можливо, коли виконуються умови:

$$\frac{\partial \Delta^2}{\partial b_1} = 0, \quad \frac{\partial \Delta^2}{\partial b_2} = 0 \quad (5.7)$$

Застосування цієї умови приводить до системи неоднорідних лінійних рівнянь:

$$\begin{cases} b_1 \sum_{j=1}^m x_{1j}^2 + b_2 \sum_{j=1}^m x_{1j}x_{2j} = \sum_{j=1}^m y_j x_{1j} \\ b_1 \sum_{j=1}^m x_{1j}x_{2j} + b_2 \sum_{j=1}^m x_{2j}^2 = \sum_{j=1}^m y_j x_{2j} \end{cases} \quad (5.8)$$

Коефіцієнти при шуканих величина b_1, b_2 , які мають форму сум, визначені. Тому система (5.8) може бути розв'язаною. Зазначені суми пов'язані з коваріаціями випадкових величин y_i, x_{1i} й x_{2i} . Це дає можливість записати систему рівнянь (5.8) у більш простій формі:

$$\begin{cases} b_1 \sigma_{x_1} + b_2 \sigma_{x_2} r_{x_1 x_2} = \sigma_y r_{yx_1} \\ b_1 \sigma_{x_1} r_{x_1 x_2} + b_2 \sigma_{x_2} = \sigma_y r_{yx_2} \end{cases}, \quad (5.9)$$

де

$\sigma_{x_1}, \sigma_y, \sigma_{x_1}$ – середні квадратичні відхили;

$r_{x_1 x_2}, r_{yx_1}, r_{yx_2}$ – коефіцієнти кореляції відповідних випадкових величин.

Розв'язок системи (5.9) має вигляд:

$$b_1 = \frac{r_{yx_1} - r_{yx_2} r_{x_1 x_2}}{1 - r_{x_1 x_2}^2} \frac{\sigma_y}{\sigma_{x_1}}, \quad (5.10)$$

$$b_2 = \frac{r_{yx_2} - r_{yx_1} r_{x_1 x_2}}{1 - r_{x_1 x_2}^2} \frac{\sigma_y}{\sigma_{x_2}} \quad (5.11)$$

Просто показати, що систему рівнянь (5.9) можна записати й у матричній формі:

$$R_x \sigma B = \sigma_y R_{yx}, \quad (5.12)$$

де

R_x – кореляційна матриця величин x_1, x_2 ;

R_{yx} – вектор кореляцій між предиктантом y і предикторами x_1, x_2 ;

σ – діагональна матриця середніх квадратичних відхилів x_1, x_2 ;

B – вектор шуканих коефіцієнтів регресії (5.5).

Такий же вигляд буде мати й система рівнянь відносно шуканих коефіцієнтів рівняння (5.4):

$$\begin{cases} b_1 \sigma_{x_1} + b_2 \sigma_{x_2} r_{x_1 x_2} + \dots + b_n \sigma_{x_n} r_{x_1 x_n} = \sigma_y r_{yx_1} \\ b_1 \sigma_{x_1} r_{x_2 x_1} + b_2 \sigma_{x_2} + \dots + b_n \sigma_{x_n} r_{x_2 x_n} = \sigma_y r_{yx_2} \\ \dots \\ b_1 \sigma_{x_1} r_{x_n x_1} + b_2 \sigma_{x_2} r_{x_n x_2} + \dots + b_n \sigma_{x_n} = \sigma_y r_{yx_n} \end{cases}, \quad (5.13)$$

але матриці R_x , σ й вектори R_{yx} і B будуть мати вже n -ий порядок.

Розв'язання систем (5.9) і (5.13) у матричній формі мають однаковий вигляд:

$$B = \sigma_y \sigma^{-1} R_x^{-1} R_{yx}, \quad (5.14)$$

де індекс (-1) позначає операцію обертання матриці.

Отже маємо n -вимірний вектор:

$$B' = (b_1, b_2, \dots, b_n), \quad (5.15)$$

коефіцієнтів рівняння (5.4).

Якщо на початку операцію центрування предиктанта й предикторів не проводити, то отримаємо модель вигляду:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_n X_n, \quad (5.16)$$

де

b_0 – вільний член, який дорівнює:

$$b_0 = \bar{Y} - (b_1 \bar{X}_1 + b_2 \bar{X}_2 + \dots + b_n \bar{X}_n). \quad (5.17)$$

В деяких випадках з'являється можливість будувати регресійну модель на статистично незалежних (ортогональних) предикторах, коли $r_{X_i X_j} = 0, \forall i, j = \overline{1, n}$. Тоді у системі (5.13) відрізняються від нуля лише діагональні елементи і тоді формули, що визначають коефіцієнти регресії, значно спрощуються і мають вигляд:

$$\begin{cases} b_1 = r_{yx_1} \frac{\sigma_y}{\sigma_{x_1}}, \\ b_2 = r_{yx_2} \frac{\sigma_y}{\sigma_{x_2}}, \\ \dots\dots\dots \\ b_n = r_{yx_n} \frac{\sigma_y}{\sigma_{x_n}}. \end{cases} \quad (5.18)$$

Прикладом такого випадку є побудова регресії вигляду:

$$y = a_1 z_1 + a_2 z_2 + \dots + a_k z_k, \quad (5.19)$$

між деякою метеорологічною величиною y і головними компонентами z_i векторів предикторів, дисперсіями яких, як зазначалося вище, є власні значення матриці коваріації λ_i . Тоді коефіцієнти a_i визначаються співвідношеннями:

$$\begin{cases} a_1 = r_{yz_1} \frac{\sigma_y}{\sqrt{\lambda_1}} \\ a_2 = r_{yz_2} \frac{\sigma_y}{\sqrt{\lambda_2}}, \\ \dots\dots\dots \\ a_k = r_{yz_k} \frac{\sigma_y}{\sqrt{\lambda_k}} \end{cases} \quad (5.20)$$

де

$r_{yz_i} (i = \overline{1, k})$ – коефіцієнти кореляції між предиктором y й відповідними головними компонентами z_i .

Система нормальних рівнянь (5.13) не обов'язково розв'язується шляхом застосування матричного рівняння (5.14). У більшості випадків для цього використовується чисельний метод, наприклад метод Гаусса.

5.2 Перевірка гіпотези про статистичну значущість коефіцієнтів регресії

Як зазначалося вище, шляхом застосування метода найменших квадратів отримується стовпець параметрів множинного коефіцієнта регресії (5.4). Матрична формула розв'язання системи нормальних рівнянь (5.13) має вигляд (5.14). Але можна показати, що ця задача розв'язується й за допомогою інформаційної матриці Фішера F , тобто:

$$B = F^{-1}XY. \quad (5.21)$$

Матриця Фішера є добутком матриць:

$$F = XY. \quad (5.22)$$

де

X – матриця порядку $n \times t$ центрованих значень предикторів;

Y – вектор-стовпець центрованих значень предиктанта.

З іншого боку, як можна показати:

$$F^{-1} = \frac{1}{\sigma_y^2} \begin{pmatrix} \sigma^2(b_1) & k_{b_1b_2} & \dots & k_{b_1b_i} & \dots & k_{b_1b_n} \\ k_{b_2b_1} & \sigma^2(b_2) & \dots & k_{b_2b_i} & \dots & k_{b_2b_n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ k_{b_ib_1} & k_{b_ib_2} & \dots & \sigma^2(b_i) & \dots & k_{b_ib_n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ k_{b_nb_1} & k_{b_nb_2} & \dots & k_{b_nb_i} & \dots & \sigma^2(b_n) \end{pmatrix}, \quad (5.23)$$

де

$\sigma^2(b_i)$ – дисперсія i -того коефіцієнта регресії;

$k_{b_i b_n}$ – коваріація між коефіцієнтами b_i й b_n .

Позначимо для простоти подальших розрахунків обернену інформаційну матрицю таким чином:

$$C = \begin{pmatrix} C_{11} & C_{12} & \dots & C_{1n} \\ C_{21} & C_{22} & \dots & C_{2n} \\ \dots & \dots & \dots & \dots \\ C_{i1} & C_{i2} & \dots & C_{in} \\ \dots & \dots & \dots & \dots \\ C_{n1} & C_{n2} & \dots & C_{nn} \end{pmatrix}. \quad (5.24)$$

Діагональні елементи матриці C дорівнюють:

$$C_{ii} = \frac{\sigma^2(b_i)}{\sigma_y^2}, \quad (i = \overline{1, n}). \quad (5.25)$$

Звідси,

$$\sigma(b_i) = \sqrt{C_{ii}} \cdot \sigma_y, \quad (5.26)$$

можна показати, що добуток

$$XY = \begin{pmatrix} \sum_{j=1}^m x_{1j} y_j \\ \sum_{j=1}^m x_{2j} y_j \\ \vdots \\ \sum_{j=1}^m x_{nj} y_j \end{pmatrix} = m \sigma_y \sigma R_{yx}. \quad (5.27)$$

де

σ – діагональна матриця середніх квадратичних відхилів;

R_{yx} – вектор кореляцій між предиктантом й предикторами.

Компануючи рівняння (5.21), (5.24) й (5.27) отримаємо:

$$B = m \sigma_y C \sigma R_{yx}. \quad (5.28)$$

Розглянувши рівняння (5.4) й порівнюючи ліву його частину з лівою частиною рівняння (5.27), отримаємо:

$$m\sigma_y C \sigma R_{yx} = \sigma_y \sigma^{-1} R_x^{-1} R_{yx}, \quad (5.29)$$

тоді

$$C = \frac{1}{m} \sigma^{-1} R_x^{-1} \sigma^{-1}, \quad (5.30)$$

Розраховуючи обернену матрицю R_x^{-1} , можна знайти матрицю C , а значить і її діагональні елементи C_{ii} ($i = \overline{1, n}$). Тоді за формулою (5.25) отримаємо:

$$\sigma(b_i) = \sqrt{C_{ii}} S_y \quad (5.31)$$

де

S_y – середній квадратичний відхил відтвореності процесу. **Дисперсія відтвореності** – це є **дисперсія нев'язки** рівняння:

$$\hat{\sigma}_y^2 = \frac{1}{k} \sum_{j=1}^k [y_j - (b_1 x_{1j} + b_2 x_{2j} + \dots + b_n x_{nj})]^2. \quad (5.32)$$

Її можна розрахувати шляхом проведення k експериментів з побудованою моделлю.

Зазначені розрахунки дають можливість побудувати довірчий інтервал для коефіцієнтів регресії

$$\Delta b_i = \sigma(b_i) t_{кр} \left(\frac{\alpha}{2}, \nu \right), \quad (5.34)$$

де

$t_{кр}$ – критичне значення критерію Стъдента.

Для рівня значущості $\frac{\alpha}{2}$ і числа степенів вільності $\nu = m - n$, де n - кількість коефіцієнтів регресії.

Гіпотеза H_0 про статистичну значущість коефіцієнта регресії не відхиляється, якщо

$$|b_i| > \Delta b_i \quad (5.35)$$

При цьому треба переконатися, що коефіцієнти кореляції:

$$r_{b_i b_j} = \frac{K_{b_i b_j}}{\sigma(b_i)\sigma(b_j)}, \quad (5.36)$$

які основані на елементах матриці C , близькі до нуля, тобто параметри рівняння регресії є некорельованими.

Оцінку статистичної значущості коефіцієнтів регресії можна проводити й за допомогою критерію Стюдента:

$$t = \frac{|b_i|}{\sigma(b_i)}. \quad (5.37)$$

Гіпотеза про статистичну значущість коефіцієнтів регресії не відхиляється, коли $t > t_{кр}(\alpha, \nu)$.

Треба мати на увазі, що вилучати коефіцієнти регресії, для яких гіпотеза про статистичну значущість відхиляється треба дуже обережно. Може статися, що їх вилучення приведе до зниження інформаційності й адекватності моделі. Щоб уникнути цієї ситуації, треба включати в модель й ті коефіцієнти, для яких критерій Стюдента не дуже відрізняється від критичного значення.

5.3 Множинний коефіцієнт кореляції

У регресійному аналізі велике значення має множинний коефіцієнт кореляції. Він характеризує тісноту лінійного кореляційного зв'язку не з одним, а з цілою системою предикторів $x_1, x_2, \dots, x_n$, тобто він може розглядатися як міра адекватності регресійної прогностичної моделі. Обґрунтування цієї властивості проведемо на основі елементарного рівняння множинної регресії вигляду:

$$y = b_1 x_1 + b_2 x_2. \quad (5.38)$$

Як і раніше, будемо вважати, що всі змінні *центровані*.

Розглянемо метрику найменших квадратів з позначенням її лівої частини у вигляді суми квадратів:

$$\sum_{j=1}^m S_{y \bullet x_1 x_2}^2 = \sum_{j=1}^m [y_j - (b_1 x_{1j} + b_2 x_{2j})]^2. \quad (5.39)$$

Проведені математичні операції в правій частині рівняння (5.39) показали, що два останні члени правої частини дорівнюють нулю. Отже маємо

$$\sum_{j=1}^m S_{y \bullet x_1 x_2}^2 = \sum_{j=1}^m y_j^2 - b_1 \sum_{j=1}^m y_j x_{1j} - b_2 \sum_{j=1}^m y_j x_{2j}, \quad (5.40)$$

Оскільки всі змінні правої частини цього рівняння є центрованими величинами, тоді, очевидно

$$\sum_{j=1}^m S_{y \bullet x_1 x_2}^2 = m\sigma_y^2 - b_1 m\sigma_{x_1} \sigma_y r_{yx_1} - b_2 m\sigma_{x_2} \sigma_y r_{yx_2}. \quad (5.41)$$

Якщо рівняння (5.41) поділити на об'єм статистичної сукупності m і σ_y , а замість коефіцієнтів b_1, b_2 підставимо їх значення з рівнянь (5.10) і (5.11), тоді будемо мати:

$$\tilde{S}_{y \bullet x_1 x_2}^2 = 1 - \frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1} r_{yx_2} r_{x_1 x_2}}{1 - r_{x_1 x_2}^2}. \quad (5.42)$$

Ліва частина рівняння (5.42):

$$\tilde{S}_{y \bullet x_1 x_2}^2 = \frac{1}{\sigma_y^2} \frac{\sum_{i=1}^n S_{y \bullet x_1 x_2}^2}{m}, \quad (5.43)$$

є відносною дисперсією нев'язки моделі.

Другий член правої частини рівняння (5.42):

$$R_{y \bullet x_1 x_2}^2 = \frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1} r_{yx_2} r_{x_1 x_2}}{1 - r_{x_1 x_2}^2} \quad (5.44)$$

називають **коефіцієнтом множинної детермінації**. **Коефіцієнт множинної кореляції**, який характеризує тісноту лінійного кореляційного зв'язку між предиктантом y і предикторами x_1 і x_2 , є коренем квадратним з коефіцієнта множинної детермінації:

$$R_{y \bullet x_1 x_2} = \sqrt{\frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1} r_{yx_2} r_{x_1 x_2}}{1 - r_{x_1 x_2}^2}} \quad (5.45)$$

Зв'язок між відносною дисперсією нев'язки моделі й коефіцієнтом множинної кореляції визначається так:

$$R_{y \bullet x_1 x_2} = \sqrt{1 - \tilde{S}_{y \bullet x_1 x_2}^2} \quad (5.46)$$

Для того, щоб перейти від двовимірного рівняння (4.5) до n -вимірного (5.4) рівняння лінійної регресії, створимо розширену матрицю вигляду:

$$\tilde{R} = \begin{pmatrix} 1 & r_{yx_1} & r_{yx_2} \\ r_{yx_1} & 1 & r_{x_1 x_2} \\ r_{yx_2} & r_{x_1 x_2} & 1 \end{pmatrix} \quad (5.47)$$

й розрахуємо його визначник, поділивши обидві його частини на визначник двовимірної матриці кореляцій $|R_x| = 1 - r_{x_1 x_2}^2$, отримаємо:

$$\frac{|\tilde{R}|}{|R_x|} = 1 - \frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1} r_{yx_2} r_{x_1 x_2}}{1 - r_{x_1 x_2}^2} \quad (5.48)$$

Звідси виходить, що

$$R_{y \bullet x_1 x_2} = \sqrt{1 - \frac{|\tilde{R}|}{|R_x|}} \quad (5.49)$$

Поширюючи отриманий результат на n предикторів, запишемо формулу для визначення відповідного множинного коефіцієнта кореляції. Вона має вигляд:

$$R_{y \bullet x_1 x_2 \dots x_n} = \sqrt{1 - \frac{|\tilde{R}|}{|R_x|}}. \quad (5.50)$$

Але тепер матриця R_x є n -вимірною, а матриця $\tilde{R}$ $-(n+1)$ -вимірною, тобто

$$\tilde{R} = \begin{pmatrix} 1 & r_{yx_1} & r_{yx_1} & \dots & r_{yx_n} \\ r_{yx_1} & 1 & r_{x_1 x_2} & \dots & r_{x_1 x_n} \\ r_{yx_2} & r_{x_1 x_2} & 1 & \dots & r_{x_2 x_n} \\ \dots & \dots & \dots & \dots & \dots \\ r_{yx_n} & r_{x_n x_1} & r_{x_n x_2} & \dots & 1 \end{pmatrix}. \quad (5.51)$$

Множний коефіцієнт кореляції може приймати значення в межах:

$$0 \leq R_{y \bullet x_1 x_2 \dots x_n} \leq 1. \quad (5.52)$$

Оскільки він характеризує тісноту кореляційного зв'язку між предиктантом і всіма предикторами одночасно $x_1, x_2, \dots, x_n$, тоді можна розглядати його як міру адекватності моделі метеорологічного прогнозу $y = f(x_1, x_2, \dots, x_n)$. Чим його значення є більшим, тим тісніший зв'язок між предиктантом і системою виділених предикторів й тим більш адекватною є прогностична модель.

Оцінити міру адекватності моделі можна шляхом перевірки статистичної гіпотези H_0 про те, що залишкова дисперсія $\sigma_{y \bullet x_1 x_2 \dots x_n}^2$ не значуще відрізняється від дисперсії предиктанта. Якщо гіпотеза H_0 відхиляється, то це значить, що прогноз по моделі не відрізняється від випадкового. Перевірка гіпотези H_0 проводиться за критерієм Фішера:

$$F = \frac{\sum_{j=1}^m (y_i - \bar{y})^2 / m - 1}{\sum_{j=1}^{m-n} (y_j - y_{p_i})^2 / m - n - 1}, \quad (5.53)$$

де

m – об'єм вибірки;

n – кількість предикторів, що включені в модель;

y і y_p – фактичні й розрахункові значення предиктанта.

Із рівності (5.53) видно, що залишкова дисперсія розташовується в знаменнику. Гіпотеза H_0 не відхиляється, якщо $F < F_{кр}(\alpha, v_1, v_2)$, де $v_1 = m - 1$; $v_2 = m - n - 1$.

Вірогідність моделі прогнозу можна оцінити й на основі коефіцієнта кореляції між фактичними та розрахунковими значеннями предиктанта. Коефіцієнти кореляції розраховуються як по навчаючій, так по перевіряючій (незалежній) сукупності предикторів. Модель тим краще відповідає величині, що прогнозується, чим більшим є коефіцієнт кореляції.

5.4 Методи відбору статистично значущих предикторів

5.4.1 Поняття про операцію «просіювання» предикторів

Побудова лінійного рівняння множинної регресії в якості моделі для гідрометеорологічного прогнозування є суто математичною задачею за умови, що встановлений склад предикторів $x_1, x_2, \dots, x_n$. Цей склад предикторів називають *навчальною сукупністю* або *сукупністю потенційних предикторів*. Складання навчальної сукупності – це розв'язання вже фізичної задачі, тобто предиктори відбираються на основі знань дослідника про фізичні особливості атмосферних чи гідрологічних процесів, які, як вважає дослідник, із заданою завчасністю будуть обумовлювати значення предиктанта.

Може скластися враження, що чим більше предикторів вибрано для побудови прогностичної моделі, тим більш обумовленою вона буде й тим кращими будуть результати прогнозування. Але таке враження є помилковим. Включення до моделі великої кількості предикторів збільшує,

по-перше, порядок їх матриць кореляцій й, по-друге, значно збільшується в цій матриці кількість великих значень кореляційних зв'язків між предикторами, що зменшує обумовленість матриць кореляцій, і звичайно, матриць коваріацій. Оскільки елементи цих матриць є статистичними оцінками, які, як відомо, мають похибки, то перелічені обставини можуть суттєво впливати на точність параметрів моделі і, як наслідок, на вірогідність прогнозу. Щоб усунути перелічені проблеми, із множини потенційних предикторів добирають ті, які виявляються статистично значущими. Таку операцію *добору статистично значущих (оптимальних) предикторів називають операцією "просіювання"*. Розглянемо деякі алгоритми "просіювання" предикторів.

5.4.2 Просіювання предикторів за допомогою частинного коефіцієнта кореляції

5.4.2.1 Визначення частинного коефіцієнта кореляції

У попередніх розділах розглядалися парні r_{xy} і множинні коефіцієнти кореляції $r_{y \bullet x_1 x_2 \dots x_n}$. Парний коефіцієнт кореляції визначає тісноту лінійної кореляційної залежності між двома випадковими величинами x й y . При цьому ми нехтуємо впливами на випадкову величину y інших випадкових величин. В дійсності предиктори, як правило, статистично зв'язані між собою й при побудові статистичних моделей ці зв'язки треба враховувати. Це можна здійснити за допомогою **частинних коефіцієнтів кореляції**. Приведемо визначення частинного коефіцієнта кореляції.

Припустимо, що на випадкову величину y чинять вплив дві випадкові величини x_1 і x_2 . **Частинним коефіцієнтом кореляції** між випадковими величинами y і x_1 називають коефіцієнт кореляції $r_{yx_1 \bullet x_2}$ між ними такий, що вплив другої випадкової величини x_2 на величину y вже враховано. Таким же чином формулюється й визначення частинного коефіцієнта кореляції $r_{yx_2 \bullet x_1}$. Розглянемо алгоритм розрахунку частинного коефіцієнта кореляції.

Будемо вважати, що відома матриця кореляції:

$$R_x = \begin{pmatrix} 1 & r_{x_1 x_2} \\ r_{x_2 x_1} & 1 \end{pmatrix} \quad (5.54)$$

і вектор парних кореляцій:

$$R_{yx} = \begin{pmatrix} r_{yx_1} \\ r_{yx_2} \end{pmatrix}. \quad (5.55)$$

На їх основі сформуємо розширену матрицю кореляцій:

$$\tilde{R} = \begin{pmatrix} 1 & r_{yx_1} & r_{yx_2} \\ r_{yx_1} & 1 & r_{x_1x_2} \\ r_{yx_2} & r_{x_1x_2} & 1 \end{pmatrix}. \quad (5.56)$$

Очевидно, ця матриця створюється шляхом додавання до матриці (5.54) рядка та стовпця з вектора (5.55). На основі матриці $\tilde{R}$ розрахуємо мінори $\ddot{A}_{yx_i}, \ddot{A}_{yx_i}^-$ ($i=1,2$). Мінори $\ddot{A}_{yx_i}$ створюють таким чином: стовпець, на першому місці якого розташовується парна кореляція r_{yx_i} , переставляється на перше місце, а на його місце – перший стовпець і, після цього, викреслюються перший рядок і стовпець. Очевидно, що

$$\ddot{A}_{yx_1} = r_{yx_1} - r_{yx_2} r_{x_1x_2}, \quad (5.57)$$

$$\ddot{A}_{yx_2} = r_{yx_2} - r_{yx_1} r_{x_1x_2}. \quad (5.58)$$

Мінори $\ddot{A}_{yx_i}^-$ ($i=1,2$) – це мінори визначника $|\tilde{R}|$, які не мають парної кореляції r_{yx_i} . Ми їх отримаємо, якщо викреслемо із мінора $|\tilde{R}|$ рядок і стовпець, що утримують цю парну кореляцію. Отже

$$\ddot{A}_{yx_1}^- = \begin{vmatrix} 1 & r_{yx_2} \\ r_{yx_2} & 1 \end{vmatrix} = 1 - r_{yx_2}^2, \quad (5.59)$$

$$\ddot{A}_{yx_2}^- = \begin{vmatrix} 1 & r_{yx_1} \\ r_{yx_1} & 1 \end{vmatrix} = 1 - r_{yx_1}^2. \quad (5.60)$$

Міnor матриці кореляцій, очевидно, дорівнює:

$$|R_x| = \begin{vmatrix} 1 & r_{x_1x_2} \\ r_{x_1x_2} & 1 \end{vmatrix} = 1 - r_{x_1x_2}^2, \quad (5.61)$$

Частинні коефіцієнти кореляції визначаються таким чином:

$$r_{yx_1 \bullet x_2} = \frac{D_{yx_1}}{\sqrt{|R_x| D_{yx_1}^-}} = \frac{r_{yx_1} - r_{yx_2} r_{x_1x_2}}{\sqrt{(1 - r_{x_1x_2}^2)(1 - r_{yx_2}^2)}}, \quad (5.62)$$

$$r_{yx_2 \bullet x_1} = \frac{D_{yx_2}}{\sqrt{|R_x| D_{yx_2}^-}} = \frac{r_{yx_2} - r_{yx_1} r_{x_1x_2}}{\sqrt{(1 - r_{x_1x_2}^2)(1 - r_{yx_1}^2)}}. \quad (5.63)$$

Розв'яжемо елементарну задачу, яка пояснить суть частинних коефіцієнтів кореляції. Нехай парні коефіцієнти кореляції між випадковими величинами мають значення: $r_{yx_1} = 0.70$; $r_{yx_2} = 0.90$; $r_{x_1x_2} = 0.80$. Це значить, що тісний кореляційний зв'язок спостерігається не тільки між предіктантом y і предікторами x_1 , x_2 , але і між предікторами. Для того, щоб визначити, як кожний предіктор x_1 , x_2 впливає на предіктант y , треба розрахувати частинні коефіцієнти кореляції за допомогою формул (5.62) і (5.63):

$$r_{yx_1 \bullet x_2} = -0.08, \quad r_{yx_2 \bullet x_1} = 0.79.$$

Таким чином ясно, що в дійсності на випадкову величину y чинить вплив випадкова величина x_2 , у той час, коли зв'язок y з величиною x_1 , при врахуванні її зв'язку з x_2 , є не тільки незначним, але навіть має протилежний знак.

Для поширення розглянутого алгоритму розрахунку частинних коефіцієнтів кореляції для n -ої кількості випадкових величин, створимо розширену матрицю:

$$\tilde{R} = \begin{pmatrix} 1 & r_{yx_1} & r_{yx_2} & r_{yx_3} & \dots & r_{yx_k} & \dots & r_{yx_n} \\ r_{yx_1} & 1 & r_{x_1x_2} & r_{x_1x_3} & \dots & r_{x_1x_k} & \dots & r_{x_1x_n} \\ r_{yx_2} & r_{x_2x_1} & 1 & r_{x_2x_3} & \dots & r_{x_2x_k} & \dots & r_{x_2x_n} \\ r_{yx_3} & r_{x_3x_1} & r_{x_3x_2} & 1 & \dots & r_{x_3x_k} & \dots & r_{x_3x_n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ r_{yx_k} & r_{x_kx_1} & r_{x_kx_2} & r_{x_kx_3} & \dots & 1 & \dots & r_{x_kx_n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ r_{yx_n} & r_{x_nx_1} & r_{x_nx_2} & r_{x_nx_3} & \dots & r_{x_nx_k} & \dots & 1 \end{pmatrix} \quad (5.64)$$

і на її основі визначити мінори $|R_x|, D_{yx_k}, D_{yx_k}^-$ ($k = \overline{1, n}$) за тими ж правилами, як це було зроблено в попередньому випадку для матриці третього порядку:

$$r_{yx_k \bullet x_1, x_2, \dots, x_{k-1}, x_{k+1}, \dots, x_n} = \frac{D_{yx_k}}{\sqrt{|R_x| D_{yx_k}^-}} \quad (5.65)$$

Мінори $|R_x|, D_{yx_k}, D_{yx_k}^-$ мають порядок n , матриця $\tilde{R}$ - порядок $n+1$.

Частинні коефіцієнти кореляції використовуються у ряду методів об'єктивного добору оптимальних предикторів. Розглянемо два з них, які є найбільш популярними.

5.4.2.2 «Просіювання» предикторів за методом «включення»

Метод включення базується на проведенні ряду послідовних матричних операцій, які полягають у розрахунках частинних коефіцієнтів кореляцій, що враховують вплив нового предиктора на кожному кроці. Для спрощення, запровадимо позначення $r_{yx_j} = r_{0j}$, $r_{x_i x_j} = r_{ij}$ і розширену матрицю кореляцій побудуємо таким чином:

$$\tilde{R} = \begin{pmatrix} 1 & r_{12} & r_{13} & \dots & r_{1n} & r_{01} \\ r_{21} & 1 & r_{23} & \dots & r_{2n} & r_{02} \\ r_{31} & r_{32} & 1 & \dots & r_{3n} & r_{03} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r_{n1} & r_{n2} & r_{n3} & \dots & 1 & r_{0n} \\ r_{01} & r_{02} & r_{03} & \dots & r_{0n} & 1 \end{pmatrix}. \quad (5.66)$$

На відмінну від матриці (5.64), розширення матриці кореляцій між потенційними предикторами $R_x = (r_{ij})_{n \times n}$ відбулось шляхом додавання рядка та стовпця кореляцій між предиктантом і предикторами не на перших, як у матриці (5.64), а на останніх місцях.

У якості першого предиктора, приймається той k -тий предиктор із множини всіх потенційних предикторів x_j ($j = \overline{1, n}$), якому відповідає найбільше за модулем значення парного коефіцієнта кореляції між предиктантом і предиктором: $|r_{0k}| = \max_j |r_{0j}|$. Йому приписується *перший номер*. Після цього на основі матриці (5.66) розраховується матриця частинних коефіцієнтів кореляції між предиктантом і предикторами за умови, що вплив k -того предиктора, вже враховано. Ця матриця має вигляд:

$$R^{(1)} = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & r_{23 \bullet 1} & r_{24 \bullet 1} & \dots & r_{2n \bullet 1} & r_{02 \bullet 1} \\ 0 & r_{32 \bullet 1} & 1 & r_{34 \bullet 1} & \dots & r_{3n \bullet 1} & r_{03 \bullet 1} \\ 0 & r_{42 \bullet 1} & r_{43 \bullet 1} & 1 & \dots & r_{4n \bullet 1} & r_{04 \bullet 1} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & r_{n2 \bullet 1} & r_{n3 \bullet 1} & r_{n4 \bullet 1} & \dots & 1 & r_{0n \bullet 1} \\ 0 & r_{02 \bullet 1} & r_{03 \bullet 1} & r_{04 \bullet 1} & \dots & r_{0n \bullet 1} & 1 \end{pmatrix}. \quad (5.67)$$

Не трудно зрозуміти, чому всі елементи першого рядка і першого стовпця матриці $R^{(1)}$, окрім першого елемента, дорівнюють нулю. Дійсно, наприклад, другий елемент першого рядка, є частинним коефіцієнтом кореляції $r_{12 \bullet 1}$. Якщо його розрахувати за формулою (5.63), в якій $y = x_1$, то отримаємо:

$$r_{12\bullet 1} = \frac{r_{12} - r_{11}r_{12}}{\sqrt{(1-r_{12}^2)(1-r_{11}^2)}}. \quad (5.68)$$

Оскільки $r_{11}=1$, тоді приходимо до невизначеності вигляду $0/0$. Якщо здійснити граничний перехід в рівнянні (5.68) й застосувати *правило Лопіталя*, тоді будемо мати:

$$\begin{aligned} \lim_{r_{11} \rightarrow 1} r_{12\bullet 1} &= \lim_{r_{11} \rightarrow 1} \frac{\frac{\partial}{\partial r_{11}}(r_{12} - r_{11}r_{12})}{\frac{\partial}{\partial r_{11}} \left[\sqrt{(1-r_{12}^2)(1-r_{11}^2)} \right]} = \\ &= \lim_{r_{11} \rightarrow 1} \frac{-r_{12} \sqrt{(1-r_{12}^2)(1-r_{11}^2)}}{-2r_{11}(1-r_{12}^2)} = 0. \end{aligned}$$

Таким же чином пояснюється рівність нулю й інших членів першого рядка і стовпця матриці $R^{(1)}$.

Після цього розглядаються частинні коефіцієнти кореляції між предиктантом і предикторами, що містяться в останньому стовпці (або рядку) матриці $R^{(1)}$ й вибирається з них той, який є найбільшим. Відповідаючий йому предиктор включається до складу статистично значущих предикторів. Йому приписується *номер другий*. Після цього розраховується матриця $R^{(2)}$, елементами якої є частинні коефіцієнти кореляції визначені за умови, що вплив першого і другого (тобто відібраних) предикторів вже врахований. Якщо відкинути перший рядок і стовпець матриці (5.67), тоді матриця $R^{(2)}$ приймає вигляд:

$$R^{(2)} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & r_{34\bullet 12} & \dots & r_{3n\bullet 12} & r_{03\bullet 12} \\ 0 & r_{43\bullet 12} & 1 & \dots & r_{4n\bullet 12} & r_{04\bullet 12} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & r_{n3\bullet 12} & r_{n4\bullet 12} & \dots & 1 & r_{0n\bullet 12} \\ 0 & r_{03\bullet 12} & r_{04\bullet 12} & \dots & r_{0n\bullet 12} & 1 \end{pmatrix} \quad (5.69)$$

Третій предиктор визначається шляхом порівняння частинних коефіцієнтів кореляції між предиктантом і предикторами, які містяться в останньому стовпці (рядку) матриці (5.69) за зазначеним вище правилом. Процедура продовжується до тих пір, доки на деякому $s+1$ етапі всі частинні коефіцієнти кореляції, що залишилися, втрачають статистичну значущість. Гіпотеза H_0 про статистичну незначущість цих частинних коефіцієнтів кореляції на рівні значущості α перевіряється за допомогою критерію Стюдента:

$$t = \frac{|r_{0(s+1) \bullet 12 \dots s}|}{\sigma_r}. \quad (5.70)$$

Вона не відхиляється, якщо $t < t(\alpha, \nu)$, де $\nu = m - s$. Отже, ці s предикторів, що вийшли до складу статистично значущих, й являють собою основу, на якій будуються статистичні моделі гідрометеорологічних прогнозів.

5.4.2.3 «Просіювання» предикторів за методом покрокової регресії

Метод покрокової регресії отримав таку назву тому, що сукупність статистично значущих предикторів відбирається із сукупності потенційних предикторів за допомогою деякої кількості послідовних кроків.

Спочатку, на основі об'єма m статистичних сукупностей всіх n потенційних предикторів, які визначені при розв'язанні фізичної задачі, а також сукупності такого ж об'єму предиктанта, розраховується вектор парних коефіцієнтів кореляції між предиктантом і предикторами:

$$R_{yx} = \begin{pmatrix} r_{yx_1} \\ r_{yx_2} \\ \dots \\ r_{yx_k} \\ \dots \\ r_{yx_n} \end{pmatrix}. \quad (5.71)$$

Аналіз компонент цього вектора дає можливість визначити найбільший з них за модулем. Предиктор x_k , для якого відповідає така

умова $|r_{yk}| = \max$, вважається статистично значущим, і йому приписують номер перший ($x_k = x_1$). Інші предиктори перенумеровують й переходять до виконання подальшої покрокової процедури.

1. За допомогою методу найменших квадратів (МНК) знаходять оцінку рівняння регресії предиктанта y на перший предиктор x_1 (щойно визначений), тобто:

$$y^{(1)} = a_1^{(1)}x_1 \quad (5.72)$$

(число в дужках визначає номер кроку).

Після цього, знаходять **нев'язки предиктанта** з його моделлю:

$$\varepsilon_{1i} = y_i - y_i^{(1)}, \quad \forall i = \overline{1, m} \quad (5.73)$$

і далі розраховують коефіцієнти кореляції $r_{\varepsilon_1 x_j} (j = \overline{2, n})$. Як свідчить рівняння (5.73) коефіцієнти $r_{\varepsilon_1 x_j}$ визначає кореляційний зв'язок між предиктантом і предикторами за умови, що вплив першого предиктора вилучається. Отже, коефіцієнти кореляції $r_{\varepsilon_i x_j} (j = \overline{2, n})$ мають сенс частинних коефіцієнтів кореляції $r_{yx_j \bullet x_1}$, оскільки величина ε_{1i} характеризує значення предиктанта, з якого вилучається вплив першого предиктора.

Далі відбирається найбільший за модулем частинний коефіцієнт кореляції другого кроку, що дає підстави вважати відповідний йому предиктор статистично значущим і віднести його до сукупності оптимальних предикторів. Йому присвоюється номер другий ($x_s = x_2$). Інші предиктори знову перенумеровують.

2. Будується рівняння регресії вже з двома відібраними предикторами:

$$y^{(2)} = a_1^{(2)}x_1 + a_2^{(2)}x_2, \quad (5.74)$$

коефіцієнти якого отримують за допомогою МНК.

Визначають другі невязки:

$$\varepsilon_{2i} = y_i - y_i^{(2)}, \quad (5.75)$$

тобто вилучаються із значень предиктанта впливи предикторів x_1 і x_2 . Далі розраховуються коефіцієнти кореляції $r_{\varepsilon_2 x_j} (j = \overline{3, n})$. Вони також є аналогами частинних коефіцієнтів кореляцій $r_{yx_j \bullet x_1 x_2} (j = \overline{3, n})$, аналіз яких дає підставу для визначення третього предиктора x_3 і т.д. Інші предиктори знову перенумеровують. Процедура триває до тих пір, доки отримані на деякому l -тому кроці всі частинні коефіцієнти кореляції $r_{\varepsilon_l x_j} (j = \overline{l+1, n})$ втрачають статистичну значущість. Гіпотеза H_0 про статистичну незначущість цих частинних коефіцієнтів кореляції перевіряється за допомогою відомого критерію Стюдента, який визначається таким же чином, як і у попередньому розділі.

Треба мати на увазі, що регресійна модель створюється за умови, що об'єми статистичних сукупностей є досить великими. Це дає підставу вважати, що розподіл коефіцієнтів кореляції близький до нормального, що дуже важливо при здійсненні процедури оцінки статистичної значущості предикторів рівняння регресії, яке будується на основі «просіяних» предикторів:

$$y = a_1 x_1 + a_2 x_2 + \dots + a_l x_l. \quad (5.76)$$

Отже, при виконанні покрокової процедури на кожному кроці, як допоміжний контролюючий параметр, розраховується відповідний множинний коефіцієнт кореляції.

Два розглянутих метода відбору предикторів із множини потенційних мало розрізняються. Відмінність між ними полягає лише в формі математичної постановки задачі.

До речі, множинний коефіцієнт кореляції може використовуватися й для відбору статистично значущих предикторів. Розглянемо цей метод у подальшому.

5.4.3 Відбір статистично значущих предикторів на основі множинного коефіцієнта кореляції

Коефіцієнти множинної кореляції можуть використовуватись не тільки для оцінки адекватності моделі але й для операції «просіювання» предикторів.

Ця процедура починається із аналізу вектора парних коефіцієнтів кореляції (5.71) між предиктантами і предикторами. Аналіз координат цього

вектора дає можливість визначити той предіктор, парний коефіцієнт кореляції якого з предіктантом є найбільшим за абсолютною величиною. Він зараховується до складу оптимальних предікторів і йому задають номер *перший*. Інші предіктори перенумеровуються. Після цього розраховуються множинні коефіцієнти кореляції $R_{y \bullet x_1 x_j} (j = \overline{2, n})$.

Очевидно, що у якості другого оптимального предіктора повинен виступати той s -тий предіктор, для якого $R_{y \bullet x_1 x_j} = \max_j R_{y \bullet x_1 x_j} (j = \overline{2, n})$ є найбільшим. Йому приписується номер *другий*, а інші предіктори знову перенумеровуються. Далі проводять розрахунки множинних коефіцієнтів кореляції $R_{y \bullet x_1 x_2 x_j} (j = \overline{3, n})$ і визначається той l -тий предіктор x_l , для якого $R_{y \bullet x_1 x_2 x_l} = \max_j R_{y \bullet x_1 x_2 x_j} (j = \overline{3, n})$. Він включається до складу оптимальних предікторів за номером *третім*, а інші предіктори знову перенумеровуються. Виникає питання, до яких пір треба продовжувати цю процедуру? Справа у тому, що з включенням нового предіктора до складу статистично значущих предікторів значення відповідного множинного коефіцієнта кореляції збільшується, і при включенні деякого k -того предіктора він переходить до стану насичення.

Стан насичення виникає тоді, коли гіпотеза H_0 про значущість розбіжностей між $R_{y \bullet x_1 x_2 \dots x_k}$ і $R_{y \bullet x_1 x_2 \dots x_k x_{k+1}}$ відхиляється. Перевірка цієї статистичної гіпотези проводиться за допомогою критерію Фішера, який визначається за формулою:

$$F = \frac{R_{y \bullet x_1 x_2 \dots x_k x_{k+1}}^2 - R_{y \bullet x_1 x_2 \dots x_k}^2}{1 - R_{y \bullet x_1 x_2 \dots x_k}^2} \frac{m - k + 1}{k - 1}. \quad (5.77)$$

Зазначена гіпотеза відхиляється, коли виконується умова $F < F_{кр}(\alpha, v_1, v_2)$, де $v_1 = k - 1$, а $v_2 = m - k + 1$, α - рівень значущості. Гіпотеза H_0 перевіряється після кожного кроку розрахунків множинних коефіцієнтів кореляції. Незначущість розбіжностей між $R_{y \bullet x_1 x_2 \dots x_k}$ і $R_{y \bullet x_1 x_2 \dots x_k x_{k+1}}$ значить що включення до складу факторів прогностичної моделі предіктора x_{k+1} не приводить до суттєвого зменшення остаточної дисперсії порівняно з випадком, коли модель базується на k попередньо відібраних предікторах.

5.5 Приклад застосування регресійного аналізу

Розглянемо приклад, що ілюструє процес побудови лінійного рівняння регресії як моделі прогнозу концентрації SO_2 для одного з районів Одеси. У якості предикторів були розглянуті 30 характеристик стану граничного шару атмосфери, які впливають на формування полів концентрації інгредієнта. Методом "включення" було визначено дев'ять таких найбільш значущих предикторів: середня концентрація SO_2 за попередню добу (предиктор №1), температура повітря (предиктор №2), напрямок вітру (предиктор №3), швидкість вітру (предиктор №4), відносна вологість повітря (предиктор №5), верхня границя приземної інверсії (предиктор №6), нижня границя припіднятої інверсії (предиктор №7), верхня її границя (предиктор №8), вертикальний градієнт температури повітря в пограничному шарі (предиктор №9). Перелік цих метеорологічних величин характеризує перенос і турбулентну дифузію домішок в приземному шарі атмосфери.

Оскільки термін прогнозу дорівнював 12 годин, то всі предиктори, окрім першого, добиралися за попередній дванадцяти годинний термін. Вектор кореляцій між предиктантом y і переліченими предикторами X має такий вигляд:

$$R_{xy} = \begin{pmatrix} 0.837 \\ -0.214 \\ 0.310 \\ 0.196 \\ 0.507 \\ 0.210 \\ 0.414 \\ 0.374 \\ -0.196 \end{pmatrix}.$$

Лінійне рівняння регресії будувалося за допомогою метода покрокової регресії. Нижче наводяться результати відповідного ітераційного процесу:

1) Крок 1-й

Вільний член $b_0 = 5,527 * 10^{-12}$

Коефіцієнти регресії

$b_1 = 0,837$, коефіцієнт Стьюдента $t = 8,79$. Предиктор 1

Множинний коефіцієнт кореляції $R = 0,837$.

2) Крок 2-й

Вільний член $b_0 = 5,380 * 10^{-12}$

Коефіцієнт регресії

$b_1 = 0,847$, коефіцієнт Стюдента $t = 9,19$ Предіктор 1

$b_2 = 0,148$, коефіцієнт Стюдента $t = 1,50$ Предіктор 8

Множинний коефіцієнт кореляції $R = 0,848$.

3) Крок 3

Вільний член $b_0 = 5,32 * 10^{-12}$

Коефіцієнти регресії

$b_1 = 0,872$, коефіцієнт Стюдента $t = 9,81$ Предіктор 1

$b_2 = 0,173$, коефіцієнт Стюдента $t = 1,80$ Предіктор 8

$b_3 = 0,145$, коефіцієнт Стюдента $t = 1,64$ Предіктор 2

Множинний коефіцієнт кореляції $R = 0,860$.

4) Крок 4

Вільний член $b_0 = 5,33 * 10^{-12}$

Коефіцієнт регресії

$b_1 = 0,867$, коефіцієнт Стюдента $t = 9,92$ Предіктор 1

$b_2 = 0,199$, коефіцієнт Стюдента $t = 2,11$ Предіктор 8

$b_3 = 0,116$, коефіцієнт Стюдента $t = 1,33$ Предіктор 2

$b_4 = -0,099$, коефіцієнт Стюдента $t = 1,29$ Предіктор 9

Множинний коефіцієнт кореляції $R = 0,865$.

5) Крок 5

Вільний член $b_0 = 5,74 * 10^{-12}$

Коефіцієнт регресії

$b_1 = 0,826$, коефіцієнт Стюдента $t = 9,53$ Предіктор 1

$b_2 = 0,197$, коефіцієнт Стюдента $t = 2,10$ Предіктор 8

$b_3 = 0,103$, коефіцієнт Стюдента $t = 1,19$ Предіктор 2

$b_4 = -0,086$, коефіцієнт Стюдента $t = 0,99$ Предіктор 9

$b_5 = 0,079$, коефіцієнт Стюдента $t = 0,91$ Предіктор 5

Множинний коефіцієнт кореляції $R = 0,867$.

6) Крок 6

Вільний член $b_0 = -6,075 * 10^{-12}$

Коефіцієнти регресії

$b_1 = 0,826$, коефіцієнт Стюдента $t = 9,49$. Предіктор 1

$b_2 = 0,215$, коефіцієнт Стюдента $t = 2,28$. Предіктор 8

$b_3 = 0,095$, коефіцієнт Стюдента $t = 1,09$. Предіктор 2

$b_4 = 0,094$, коефіцієнт Стюдента $t = 1,08$. Предіктор 9

$b_5 = 0,078$, коефіцієнт Стюдента $t = 0,90$. Предіктор 5

$b_6 = 0,078$, коефіцієнт Стюдента $t = 0,55$. Предіктор 3

Множинний коефіцієнт кореляції $R = 0,866$.

7) Крок 7

Вільний член $b_0 = -7,673 * 10^{-12}$

Коефіцієнти регресії

$b_1 = 0,815$, коефіцієнт Стюдента $t = 9,45$. Предіктор 1
 $b_2 = 0,263$, коефіцієнт Стюдента $t = 3,89$. Предіктор 8
 $b_3 = 0,074$, коефіцієнт Стюдента $t = 0,86$. Предіктор 2
 $b_4 = -0,087$, коефіцієнт Стюдента $t = 1,01$. Предіктор 9
 $b_5 = 0,086$, коефіцієнт Стюдента $t = 1,00$. Предіктор 5
 $b_6 = 0,098$, коефіцієнт Стюдента $t = 0,68$. Предіктор 3
 $b_7 = 0,176$, коефіцієнт Стюдента $t = 1,89$. Предіктор 7
Множинний коефіцієнт кореляції $R = 0,869$.

8) Крок 8

Вільний член $b_0 = -9,180 * 10^{-12}$

Коефіцієнт регресії

$b_1 = 0,814$, коефіцієнт Стюдента $t = 9,41$. Предіктор 1
 $b_2 = 0,363$, коефіцієнт Стюдента $t = 3,87$. Предіктор 8
 $b_3 = 0,071$, коефіцієнт Стюдента $t = 0,82$. Предіктор 2
 $b_4 = -0,98$, коефіцієнт Стюдента $t = 1,13$. Предіктор 9
 $b_5 = 0,080$, коефіцієнт Стюдента $t = 0,93$. Предіктор 5
 $b_6 = 0,118$, коефіцієнт Стюдента $t = 0,83$. Предіктор 3
 $b_7 = -0,179$, коефіцієнт Стюдента $t = 1,91$. Предіктор 7
 $b_8 = 0,033$, коефіцієнт Стюдента $t = 0,38$. Предіктор 4
Множинний коефіцієнт кореляції $R = 0,868$

9) Крок 9

Вільний член $b_0 = -9,40 * 10^{-12}$

Коефіцієнт регресії

$b_1 = 0,815$, коефіцієнт Стюдента $t = 9,41$. Предіктор 1
 $b_2 = 0,362$, коефіцієнт Стюдента $t = 3,86$. Предіктор 8
 $b_3 = 0,073$, коефіцієнт Стюдента $t = 0,84$. Предіктор 2
 $b_4 = -0,097$, коефіцієнт Стюдента $t = 1,13$. Предіктор 9
 $b_5 = 0,080$, коефіцієнт Стюдента $t = 0,92$. Предіктор 5
 $b_6 = 0,120$, коефіцієнт Стюдента $t = 0,85$. Предіктор 3
 $b_7 = -0,177$, коефіцієнт Стюдента $t = 1,89$. Предіктор 7
 $b_8 = 0,035$, коефіцієнт Стюдента $t = 0,40$. Предіктор 4
 $b_9 = 0,057$, коефіцієнт Стюдента $t = 0,16$. Предіктор 6.
Множинний коефіцієнт кореляції $R = 0,868$.

Порівнюючи 8-й і 9-й кроки, приходимо до висновку, що ітераційний процес усталюється. Дійсно, невеликі різниці між коефіцієнтами регресії восьмого і дев'ятого кроків виявляються тільки у третьому знаку, коефіцієнти Стюдента теж змінюються незначно, а множинний коефіцієнт регресії зовсім не змінюється.

У попередньому розділі зазначалось, що формальне відкидання коефіцієнтів регресії у лінійній моделі по значеннях коефіцієнта Стюдента може привести до втрати адекватності й інформативності

моделі. Це особливо відноситься до випадку, коли вибірки предиктанта і предикторів відносно невеликі, внаслідок чого елементи матриці кореляцій предикторів можуть утримувати значні похибки. Тому остаточний висновок відносно виду моделі у такому випадку треба робити на основі чисельних експериментів з моделлю. У прикладі, що розглядається, чисельні експерименти показали, що задовільну адекватність показує прогностична модель для прогнозу концентрації $SO_2 - \hat{q}$ з предикторами, для яких коефіцієнт Стюдента $t \geq 1$. Як видно це такі предиктори: середня концентрація SO_2 за попередню добу (x_1), верхня границя припіднятої інверсії (x_8), нижня її границя (x_7) і вертикальний градієнт температури повітря у пограничному шарі атмосфери (x_9).

Отже прогностична модель має вигляд:

$$\hat{q} = 0.8x_1 - 0.177x_7 + 0.362x_8 - 0.097x_9.$$

В рівнянні регресії предиктори і предиктанти є центрованими і нормованими на середній квадратичний відхил. Значення множинного коефіцієнта кореляції свідчить про те, що адекватність моделі є задовільною. Легко можна побачити, що дисперсія похибки прогнозу складає біля 24% від дисперсії інгредієнта.

Запитання для самоперевірки

1. Як називається регресійна модель метеорологічного прогнозу, якщо вона будується на множині предикторів?
2. Який метод використовується при визначенні коефіцієнтів лінійного множинного рівняння регресії?
3. Як побудувати регресійну модель метеорологічного прогнозу на статистично незалежних предикторах?
4. Який сенс має частинний коефіцієнт кореляції і яку суттєву інформацію він утримує?
5. Дати визначення множинного коефіцієнта кореляції і на яких етапах побудови регресійної моделі він використовується?
6. Які предиктори називаються потенційними, а які – оптимальними?
7. Які методи використовуються в основі добору оптимального складу предикторів при побудові регресійної моделі метеорологічного прогнозу?
8. У чому полягає процедура “просіювання” предикторів?
9. Який сенс методу покрокової регресії?
10. Як можна оцінити міру адекватності побудованої прогностичної регресійної моделі метеорологічного прогнозу?

6 МЕТОДИ ПРИЙНЯТТЯ АЛЬТЕРНАТИВНИХ РІШЕНЬ ВІДНОСНО СТАНУ ГІДРОМЕТЕОРОЛОГІЧНИХ ОБ'ЄКТІВ

6.1 Принципи побудови схем альтернативного прогнозу

Розглянуті вище алгоритми регресійного аналізу дають можливість побудувати прогностичну модель, на основі якої отримують прогностичне значення метеорологічної величини, яка виражається числом. Але в практиці прогнозування досить часто виникає необхідність передбачення гідрометеорологічного явища. Наприклад, гроза може виникати чи не виникати, весняний повінь може утворитися або ні і т.п. Відповідно до цих обставин необхідно розробляти прогноз здійснення чи не здійснення того чи іншого гідрометеорологічного явища. Такий прогноз називається **альтернативним**.

Як і в інших статистичних моделях прогнозу, модель альтернативного прогнозу ґрунтується на множині навчальних сукупностей предікторів. Позначимо її вектором:

$$X_j = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \dots \\ x_{ij} \\ \dots \\ x_{nj} \end{pmatrix}, \quad (j = \overline{1, m}). \quad (6.1)$$

Вектор предікторів називають **образом** або **зображенням**. Отже, сформульована задача по суті являє собою *задачу розпізнавання образів*.

Нехай ми маємо множину V векторів-предікторів (6.1), що складають простір зображень R . Припустимо, що цей простір розділяється на два підпростори R_{V_1} і R_{V_2} . У першому з них розташовується множина V_1 образів X , а у другому множина V_2 образів X . При множині образів V_1 явище відбувається, а при множині V_2 – не відбувається. Очевидно

$$V = V_1 \cup V_2; \quad V_1 \cap V_2 = \emptyset. \quad (6.2)$$

Поверхня, що розділяє простір R на два зазначених підпростори називається **розділяючою гіперповерхнею**. Її рівняння має вигляд:

$$F(x_1, x_2, \dots, x_n) = 0. \quad (6.3)$$

Отримання рівняння розділяючої гіперповерхні – це *перший етап задачі розпізнавання*. Його називають **етапом навчання** системи розпізнавання. Другою задачею є отримання правила, за допомогою якого є підстави віднести вектор X , що підлягає розпізнаванню, до простору R_{V_1} ($X \in R_{V_1}$) чи до простору R_{V_2} ($X \in R_{V_2}$). Цю задачу називають **етапом розпізнавання**. Основні ідеї теорії розпізнавання образів полягають у наступному.

Позначимо *гіпотезу* H_1 про те, що образ $X \in R_{V_1}$. *Альтернативною* є *гіпотеза* H_2 , що образ $X \in R_{V_2}$. Задача полягає у тому, що треба знайти правило, за яким можна обґрунтовано прийняти гіпотезу H_1 чи H_2 . Процедура перевірки будь-якої гіпотези передбачає, що приймаючи те чи інше рішення, можливо зробити помилку *першого* чи *другого роду*. **Помилка I роду** буде тоді, коли відхиляється правильна гіпотеза. Коли приймається не вірна гіпотеза, то це **помилка II роду**.

Кожний підпростір R_{V_1} і R_{V_2} утримує відповідну множину V_1 й V_2 образів.

Будемо вважати, що відомими є умовні ймовірності класів V_1 і V_2 :

$$p(x_1, x_2, \dots, x_n / V_1), \quad (6.4)$$

$$p(x_1, x_2, \dots, x_n / V_2), \quad (6.5)$$

Позначимо ймовірність помилки *I роду* через p_a , а помилку *II роду* через p_b . Знаючи ймовірність класів V_1 і V_2 , які позначимо $p(V_1)$ і $p(V_2)$ (їх оцінки легко здобути на основі навчаючої сукупності V), можна отримати ймовірності помилок I і II роду:

$$p_a = p(V_1) \int \dots \int_{R_{V_2}} p(x_1, x_2, \dots, x_n / V_1) dx_1 dx_2 \dots dx_n \quad (6.6)$$

$$p_b = p(V_2) \int_{R_{V_1}} \dots \int p(x_1, x_2, \dots, x_n / V_2) dx_1 dx_2 \dots dx_n \quad (6.7)$$

Із теорії статистичних рішень відомо, що класифікуюче рішення повинно бути таким, щоб середній ризик прийняття рішення W був мінімальний. Він визначається формулою:

$$W = \delta_a p_a + \delta_b p_b, \quad (6.8)$$

де

δ_a і δ_b - ціни помилок I і II роду.

Підставивши до рівняння (6.8) відповідні ймовірності (6.6) і (6.7), з урахуванням того, що

$$\begin{aligned} & \int_{R_{V_1}} \dots \int p(x_1, x_2, \dots, x_n / V_1) dx_1 dx_2 \dots dx_n + \\ & + \int_{R_{V_2}} \dots \int p(x_1, x_2, \dots, x_n / V_1) dx_1 dx_2 \dots dx_n = 1, \end{aligned} \quad (6.9)$$

маємо

$$\begin{aligned} W = & \int_{R_{V_1}} \dots \int [\delta_b p(V_2) p(x_1, x_2, \dots, x_n / V_2) dx_1 dx_2 \dots dx_n - \\ & - \delta_a p(V_1) p(x_1, x_2, \dots, x_n / V_1) dx_1 dx_2 \dots dx_n] + \delta_a p(V_1) \quad . \end{aligned} \quad (6.10)$$

Із рівняння (6.10) видно, що W може досягати мінімуму лише тоді, коли підінтегральна функція від'ємна, тобто якщо

$$\delta_a p(V_1) p(x_1, x_2, \dots, x_n / V_1) > \delta_b p(V_2) p(x_1, x_2, \dots, x_n / V_2). \quad (6.11)$$

Звідси виходить, що

$$\frac{p(x_1, x_2, \dots, x_n / V_1)}{p(x_1, x_2, \dots, x_n / V_2)} > \frac{\delta_b p(V_2)}{\delta_a p(V_1)}. \quad (6.12)$$

Ліва частина рівняння (6.12) є функцією вектора-предиктора:

$$\lambda(x_1, x_2, \dots, x_n) = \frac{p(x_1, x_2, \dots, x_n / V_1)}{p(x_1, x_2, \dots, x_n / V_2)}. \quad (6.13)$$

Її називають **функцією подібності**.

Величина

$$\Theta = \frac{\delta_b p(V_2)}{\delta_a p(V_1)}, \quad (6.14)$$

називається **порогом**. Отже, ми прийшли до такого розв'язувального правила:

$$X \in V_1, \text{ якщо } \lambda(x_1, x_2, \dots, x_n) > \Theta, \quad (6.15)$$

$$X \in V_2, \text{ якщо } \lambda(x_1, x_2, \dots, x_n) < \Theta, \quad (6.16)$$

Безумовні ймовірності можуть бути однаковими $p(V_1) = p(V_2)$. Це трапиться тоді, коли множини векторів-предикторів класів V_1 і V_2 будуть мати однакову їх кількість. Якщо, крім того, можна вважати, що ціни помилок теж однакові $\delta_a = \delta_b$, то розв'язувальне правило спрощується і має вигляд:

$$X \in V_1, \text{ якщо } \lambda(x_1, x_2, \dots, x_n) > 1, \quad (6.17)$$

$$X \in V_2, \text{ якщо } \lambda(x_1, x_2, \dots, x_n) < 1, \quad (6.18)$$

Очевидно, рівняння:

$$\lambda(x_1, x_2, \dots, x_n) = \Theta, \quad (6.19)$$

є рівнянням *розділяючої гіперповерхні*.

Очевидним висновком умови (6.12) є співвідношення:

$$\ln \frac{p(x_1, x_2, \dots, x_n / V_1)}{p(x_1, x_2, \dots, x_n / V_2)} > \ln \frac{\delta_b p(V_2)}{\delta_a p(V_1)}. \quad (6.20)$$

Після логарифмування отримаємо замість функції подібності функцію

$$F(x_1, x_2, \dots, x_n) = \ln p(x_1, x_2, \dots, x_n / V_1) - \ln p(x_1, x_2, \dots, x_n / V_2) + \ln \frac{\delta_a p(V_1)}{\delta_b p(V_2)}, \quad (6.21)$$

яку називають **дискримінантною функцією**. Якщо використовується дискримінантна функція, то вирішувальне правило має вигляд:

$$X \in V_1, \text{ якщо } F(x_1, x_2, \dots, x_n) > 0, \quad (6.22)$$

$$X \in V_2, \text{ якщо } F(x_1, x_2, \dots, x_n) < 0, \quad (6.23)$$

Умова

$$F(x_1, x_2, \dots, x_n) = 0, \quad (6.24)$$

є рівнянням розділяючої гіперповерхні.

Застосування для розв'язання задачі розпізнавання образів функції подібності (6.13) або дискримінантної функції (6.21) має суттєве обмеження. Воно полягає у тому, що, як правило, багатовимірні умовні розподіли векторів-предикторів класів V_1 і V_2 невідомі. Тому вважають, що вид законів розподілу умовних ймовірностей є відомим. Тоді задача зводиться до необхідності, на основі вибірок векторів-предикторів, знайти оцінки параметрів цих законів. Ця процедура називається **відновленням** законів розподілу. При практичних реалізаціях зазначених вище вирішувальних правил для розпізнавання образів найбільш часто вважають, що класи векторів-предикторів підпорядковуються багатовимірним умовним нормальним розподілам. У дійсності, ці припущення строго не виконуються. Але, як показує досвід, це не вносить суттєвих похибок, якщо більшість предикторів має одномодальний розподіл. Ця умова, як правило, виконується.

6.2 Дискримінантна функція на основі багатовимірного нормального розподілу (дискримінантний аналіз)

Будемо вважати, що вектори-предіктори:

$$X_j = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \dots \\ x_{ij} \\ \dots \\ x_{nj} \end{pmatrix}, \quad (j = \overline{1, m}). \quad (6.25)$$

мають багатовимірний нормальний розподіл. Параметрами його є вектори математичних сподівань:

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \dots \\ \mu_i \\ \dots \\ \mu_n \end{pmatrix}, \quad (i = \overline{1, n}). \quad (6.26)$$

і матриця коваріацій $K = \{K_{ij}\}_{n \times m}$. Тому щільності умовних розподілів класів V_1 і V_2 визначаються рівняннями:

$$\begin{aligned} p(x_1, x_2, \dots, x_n / V_1) &= \\ &= \frac{1}{(2\pi)^{n/2} |K_1|^{1/2}} \exp \left[-\frac{1}{2} (X - \mu_1)' K_1^{-1} (X - \mu_1) \right], \end{aligned} \quad (6.27)$$

$$\begin{aligned} p(x_1, x_2, \dots, x_n / V_2) &= \\ &= \frac{1}{(2\pi)^{n/2} |K_2|^{1/2}} \exp \left[-\frac{1}{2} (X - \mu_2)' K_2^{-1} (X - \mu_2) \right], \end{aligned} \quad (6.28)$$

де

μ_1, μ_2, K_1, K_2 – вектори математичних сподівань і матриці коваріацій першого та другого класів.

Підставимо щільність ймовірностей (6.27) і (6.28) до дискримінантної функції (6.21). Після математичних перетворень будемо мати:

$$F(x) = \frac{1}{2}[(X - \mu_2)'K_2^{-1}(X - \mu_2) - (X - \mu_1)'K_1^{-1}(X - \mu_1) + \ln \frac{|K_2|}{|K_1|}] + \ln \frac{\delta_a p(V_1)}{\delta_b p(V_2)}. \quad (6.29)$$

Дискримінантна функція (6.29) називається квадратичною дискримінантною функцією, тому що члени у квадратних дужках являють собою квадратичну форму. Щоб використовувати дискримінантну функцію (6.29) при складанні альтернативного прогнозу гідрометеорологічного явища, необхідно здійснити етап «навчання». Він полягає в наступному. Нехай ми відібрали m векторів-предикторів $X \in V_1$, тобто коли явище, що прогнозується, здійснювалося, а також k векторів-предикторів $X \in V_2$, коли воно не спостерігалось. Загальна множина векторів X , яка дорівнює $N = m + k$, являє собою навчальну сукупність. На її основі визначаються оцінки векторів математичних сподівань μ_1 і μ_2 , якими є вектори середніх значень $\bar{X}_1$ і $\bar{X}_2$ предикторів класів V_1 і V_2 , матриць коваріацій цих класів $\hat{K}_1$ і $\hat{K}_2$, а також оцінки апіорних ймовірностей класів $\hat{p}(V_1) = \frac{m}{N}$ і $\hat{p}(V_2) = \frac{k}{N}$. Зазначені характеристики підставляють до рівняння (6.29), що дає можливість отримати коефіцієнти у квадратичних формах. Проведення цих обчислювань й завершує етап «навчання». Після цього можна переходити до етапу розпізнавання, тобто прогнозу. Він полягає в розрахунках значення дискримінантної функції (6.29) й застосуванні вирішувальних правил (6.22) і (6.23). Перше з них дає підстави прогнозувати явище, а друге – не прогнозувати.

Дискримінантна функція спрощується, якщо підготувати навчальну сукупність таким чином, щоб класи V_1 і V_2 мали однакову кількість векторів-предикторів. Тоді $\hat{p}(V_1) = \hat{p}(V_2)$. Якщо ще є підстави вважати, що оцінки помилок першого й другого роду однакові ($\delta_a = \delta_b$), тоді останній член правої частини функції (6.29) дорівнює нулю.

Дискримінантна функція (6.29) утримує операцію обернення коваріаційних матриць $\hat{K}_1$ і $\hat{K}_2$. Ця операція може привести до негативних наслідків, коли матриці коваріацій погано обумовлені. У такому випадку похибки, що містяться у коваріаціях, можуть привести до значних похибок коефіцієнтів квадратичних форм рівняння (6.29) і, як наслідок, до помилок на етапі розпізнавання. У ряді випадків зазначені похибки можуть бути більшими ніж похибки, які ми зробимо, вважаючи, що

$$K = \frac{K_1 + K_2}{2}, \quad (6.30)$$

і в рівнянні (6.29) ввести припущення, що $K_1 = K_2 = K$.

Тоді, при зазначених вище умовах ($\hat{p}(V_1) = \hat{p}(V_2), \delta_a = \delta_b$), дискримінантна функція приймає вигляд:

$$F(x) = (\mu_1 - \mu_2)K^{-1}X + \frac{1}{2}(\mu_2'K^{-1}\mu_2 - \mu_1'K^{-1}\mu_1). \quad (6.31)$$

тобто стає лінійною.

Якщо у коваріаційній матриці значення коваріацій значно менші її діагональних елементів, тобто дисперсій, настільки, що ними можна знехтувати, тоді матриця коваріацій (6.30) стає діагональною:

$$K = \begin{pmatrix} \sigma_1^2 & 0 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & \sigma_n^2 \end{pmatrix}, \quad (6.32)$$

а її обернена матриця, як відомо, дорівнює

$$K^{-1} = \begin{pmatrix} \frac{1}{\sigma_1^2} & 0 & 0 \dots 0 \\ 0 & \frac{1}{\sigma_2^2} & 0 \dots 0 \\ \dots & \dots & \dots \\ 0 & 0 & 0 \dots \frac{1}{\sigma_n^2} \end{pmatrix}, \quad (6.33)$$

Якщо, крім того, виконується умова (6.30), значно спрощується вид лінійної дискримінантної функції (6.31). У такому випадку вона дорівнює:

$$F(x) = \sum_{i=1}^n \frac{\mu_{1i} - \mu_{2i}}{\sigma_i^2} x_i + \frac{1}{2} \sum_{i=1}^n \frac{\mu_{2i} - \mu_{1i}}{\sigma_i^2}. \quad (6.34)$$

Як було зазначено вище, при практичному використанні розглянутих дискримінантних функцій замість елементів векторів математичних сподівань і дисперсій предикторів використовують їх оцінки. Це відноситься також і до функції (6.34). Дискримінантну функцію (6.34) можна впроваджувати в прогностичних цілях тільки тоді, коли виконуються всі умови, на підставі яких відбувалося спрощення функції (6.31).

Важливим є етап перевірки вірогідності прогнозу. Він вирішується шляхом проведення чисельних експериментів з побудованими моделями на вибіркових предикторах, за допомогою яких отримується кількість вірних і невірних прогнозів у класах V_1 і V_2 . Це дає змогу знайти оцінку ймовірності правильного розпізнавання події ($X \in V_1$ і $X \in V_2$), тобто ймовірність справджування прогнозу явища. Якщо дозволяє об'єм вихідних даних, то окрім навчальної сукупності створюють контрольну сукупність векторів-предикторів, з якими проводять зазначені вище чисельні експерименти (перевірка вірогідності за незалежними даними). Отримані ймовірності порівнюють з ймовірностями правильного розпізнавання, отриманих за навчальною сукупністю.

Ймовірність правильного розпізнавання образів при застосуванні лінійної дискримінантної функції можна отримати за допомогою критерію Махаланобіса:

$$\Delta = (\mu_1 - \mu_2)' K^{-1} (\mu_1 - \mu_2). \quad (6.35)$$

Ймовірність правильного розпізнавання визначається за формулою:

$$p = \frac{1}{2} \int_0^{\sqrt{\Delta}/2} e^{-t^2/2} dt = \hat{O}\left(\frac{\sqrt{\Delta}}{2}\right), \quad (6.36)$$

де

$$\hat{O}\left(\frac{\sqrt{\Delta}}{2}\right) - \text{відповідне значення інтеграла ймовірностей.}$$

6.3 Приклад застосування дискримінантного аналізу

Більшість характеристик річок Поділля, які впадають у р.Дністер ліворуч, значно відрізняються від параметрів річного стоку інших лівобережних приток. Така різниця може бути пояснена умовами формування стоку, а саме особливостями підстильної поверхні, такими як середня висота водозборів $\bar{H}(x_1)$, уклін $I(x_2)$, заболоченість $f_{\bar{o}}(x_3)$, залісненість $f_l(x_4)$. Ставиться задача виконати ідентифікацію окремих водозборів лівобережної частини басейну р.Дністер, характеристики яких відрізняються від інших за властивостями водозборів.

1) Етап «навчання». За зазначеними вище особливостями водозборів річок, які розглядаються, сформовані сукупності чотирьохвимірних векторів-предикторів двох класів V_1 і V_2 . До першого класу увійшли 17, а до другого – 18 векторів-предикторів.

На їх основі отримані середні вектори класів $\hat{\mu}_1$ і $\hat{\mu}_2$:

$$\hat{\mu}_1 = \begin{pmatrix} \bar{x}_{11} \\ \bar{x}_{21} \\ \bar{x}_{31} \\ \bar{x}_{41} \end{pmatrix} = \begin{pmatrix} 324,0 \\ 67,7 \\ 5,1 \\ 23,4 \end{pmatrix}, \quad (6.37)$$

$$\hat{\mu}_2 = \begin{pmatrix} \bar{x}_{12} \\ \bar{x}_{22} \\ \bar{x}_{32} \\ \bar{x}_{42} \end{pmatrix} = \begin{pmatrix} 283,0 \\ 66,2 \\ 1,4 \\ 10,2 \end{pmatrix}, \quad (6.38)$$

а також матриці коваріацій $\hat{K}_1$, $\hat{K}_2$ класів V_1 і V_2 :

$$\hat{K}_1 = \begin{pmatrix} 482 & 401 & -19,0 & 135 \\ 401 & 913 & -35,1 & 131 \\ -19,0 & -35,1 & 4,8 & -2,4 \\ 135 & 131 & -2,4 & 49,6 \end{pmatrix}, \quad (6.39)$$

$$\hat{K}_2 = \begin{pmatrix} 505 & 777 & -65,2 & 196 \\ 777 & 315 & -11,6 & 458 \\ -65,2 & -11,6 & 3,1 & -0,4 \\ 196 & 458 & -0,4 & 29,0 \end{pmatrix}. \quad (6.40)$$

Коваріаційні моменти, що складають матриці $\hat{K}_1$ і $\hat{K}_2$ розрізняються, але характеристики їх структури подібні. Маючи на увазі невеликі об'єми вихідної інформації доцільно провести осереднення матриць $\hat{K}_1$ і $\hat{K}_2$, що приведе дискримінантну функцію до лінійної. Осереднена матриця коваріацій має вигляд:

$$\hat{K} = \begin{pmatrix} 493,5 & 589 & -42,1 & 165,5 \\ 589 & 614 & -23,4 & 294,5 \\ -42,1 & -23,4 & 4,0 & -1,4 \\ 165,5 & 294,5 & -1,4 & 39,3 \end{pmatrix}, \quad (6.41)$$

Підставивши вектори (6.37) і (6.38), а також матрицю (6.41), після операції її обернення, до рівняння (6.31), отримаємо лінійну дискримінантну функцію:

$$F(X) = -9,32 - 0,001x_1 + 0,01x_2 + 1,14x_3 + 0,37x_4. \quad (6.42)$$

Вирішальне правило, як було показано вище, має вигляд:

$X \in V_1$, якщо $F(X) > 0$; $X \in V_2$, якщо $F(X) < 0$.

Вирогідність правильного розпізнавання визначається відстанню Махаланобіса Δ , яка характеризується рівнянням (6.35). Рівняння (6.35) при приведених значеннях $\hat{\mu}_1$, $\hat{\mu}_2$ і $\hat{K}$ дає значення відстані Махаланобіса $\Delta = 11,1$. Отже, ймовірність правильного розпізнавання

$$\Delta = \Phi\left(\frac{\sqrt{\Delta}}{2}\right) = \Phi(1,665) = 0,90.$$

Це дає підстави використовувати отриману дискримінантну функцію для розпізнавання типів водозборів.

2) Етап «розпізнавання». У якості прикладу розглядається водозбір річки Стржвяк – с. Луки, який не ввійшов до навчальної сукупності образів й для якого необхідно прийняти рішення про його належність до класу V_1 чи V_2 .

Характеристики цього водозбору такі:

висота $\bar{H} = 500$ і (x_1), уклін $I = 156$ (x_2), заболоченість $f_a = 0,5\%$ (x_3), залісненість $f_{\bar{e}} = 44\%$ (x_4). Дискримінантна функція (6.42) при перелічених характеристиках водозбору має значення $F(X) = 8,39$, тобто вона більше нуля. Таким чином, за вирішувальним правилом $F(X) > 0$ річка Стржвяк – с.Луки належить до класу V_1 .

Запитання для самоперевірки

1. Який прогноз називають альтернативним?
2. Який вектор називають образом або зображенням?
3. Що називають розділяючою гіперповерхнею?
4. Яка суть помилки першого роду?
5. Що називається порогом?
6. У чому сенс процедури відновлення закону розподілу?

ДОДАТОК А

Завдання до виконання практичних робіт

Завдання до виконання практичної роботи №1

Провести дослідження однорідності та ізотропності вихідних рядів місячних сум опадів (Додаток В) для мережі із 33 метеорологічних станцій України (Додаток Г). Для цього:

1. Отримати у викладача варіант задачі для виконання практичної роботи та комп'ютерну програму "KOR.EXE" (призначення програми та особливості користування дивись нижче);
2. Виконати на ПК необхідні розрахунки та отримати матрицю кореляцій місячних сум опадів;
3. Побудувати поле кореляцій для двох метеорологічних станцій, які визначаються викладачем;
4. Зробити в різних напрямках перетини отриманих полів кореляції та побудувати просторові кореляційні функції;
5. Для осередненої кореляційної функції побудувати довірчі інтервали. Рівень значущості α прийняти 0,05;
6. Провести аналіз результатів дослідження статистичної структури полів опадів та зробити висновок по роботі у цілому.

Завдання до виконання практичної роботи №2

1. Отримати у викладача завдання (Додаток В) для виконання практичної роботи та комп'ютерну програму "Е.О.Ф" (призначення програми та особливості користування дивись у Додатку Б).
2. Реалізувати на ПК алгоритм розв'язання задачі повної проблеми власних значень і отримати власні значення та власні вектори матриці кореляцій.
3. Розв'язати задачу стиску інформації про поле опадів для території України за умови $\eta \geq 95\%$:
 - визначити скільки відсоток сумарної дисперсії поля вичерпує перше власне значення;
 - визначити кількість власних значень та власних векторів ($\eta \geq 95\%$);
 - побудувати та проаналізувати поля перших власних векторів;
 - визначити вектор ортогональних компонент та отримати головні компоненти.
4. Розв'язати задачу фільтрації інформації про поле місячної кількості

- опадів для території України:
- побудувати та проаналізувати вихідне поле опадів для 33 станцій України;
 - отримати відфільтроване поле опадів за визначений термін та побудувати його на карті.
5. Провести аналіз результатів компонентного аналізу метеорологічного поля місячної кількості опадів для 33 станцій України та зробити висновок по роботі у цілому.

Завдання до виконання практичної роботи №3

1. Отримати у викладача варіант задачі (Додаток В) для виконання практичної роботи та комп'ютерну програму "СТЕР" (призначення програми та особливості користування дивись у Додатку Б).
2. Реалізувати на ПК алгоритм розв'язання задачі покрокової регресії та отримати:
 - середні значення та середні квадратичні відхили предикторів та предиктанта;
 - матрицю центрованих та нормованих значень предикторів та предиктанта;
 - матрицю кореляцій парних коефіцієнтів кореляцій між предиктантом і кожним з потенційних предикторів (r_{yx_i}) та між предикторами ($r_{x_i x_s}$);
 - фактичні значення критерію Стюдента для перевірки гіпотези про статистичну значущість отриманих парних коефіцієнтів кореляцій;
 - результати ітераційної покрокової процедури;
 - множинний коефіцієнт кореляції на кожному кроці.
3. Провести аналіз отриманих результатів та записати рівняння множинної лінійної регресії.
4. На рівні значущості $\alpha = 0,05$ оцінити міру адекватності побудованої прогностичної моделі метеорологічного прогнозу одним із відомих методів.
5. Оформити виконану роботу згідно з викладеним порядком та захистити у викладача.

ДОДАТОК Б

Відомості про комп'ютерну програму "KOR.EXE"

Програма "**KOR.EXE**" дозволяє отримати матрицю кореляції n -го порядку, середнє арифметичне значення та середній квадратичний відхил випадкової величини на i -тій метеорологічній станції.

Файл даних до комп'ютерної програми формується таким чином щоб кожний стовпець вихідної матриці X був часовим рядом метеорологічної величини на i -тій станції (i -тій точці поля).

Приклад формування файлу вихідних даних місячної кількості опадів для 5-ти станцій України ($n = 5$) за період 7 років ($m = 7$) наводиться нижче:

24	79	66	70	24
36	24	16	25	26
49	37	16	23	21
48	62	27	72	30
66	42	66	52	40
45	91	75	44	62
49	45	36	21	29

У результаті розрахування отримаємо матрицю кореляцій місячної кількості опадів 5-го порядку.

Відомості про комп'ютерну програму „eof”

Програма „eof” дозволяє реалізувати алгоритм повної проблеми власних значень і отримати власні значення та відповідні їм власні вектори матриці кореляцій. Також розраховуються: середній квадратичний відхил випадкової величини на *i*-тій метеорологічній станції (*i*-тому вузлі сітки точок), матриця нормованих і центрованих значень випадкової величини та матриця кореляцій *n*-го порядку.

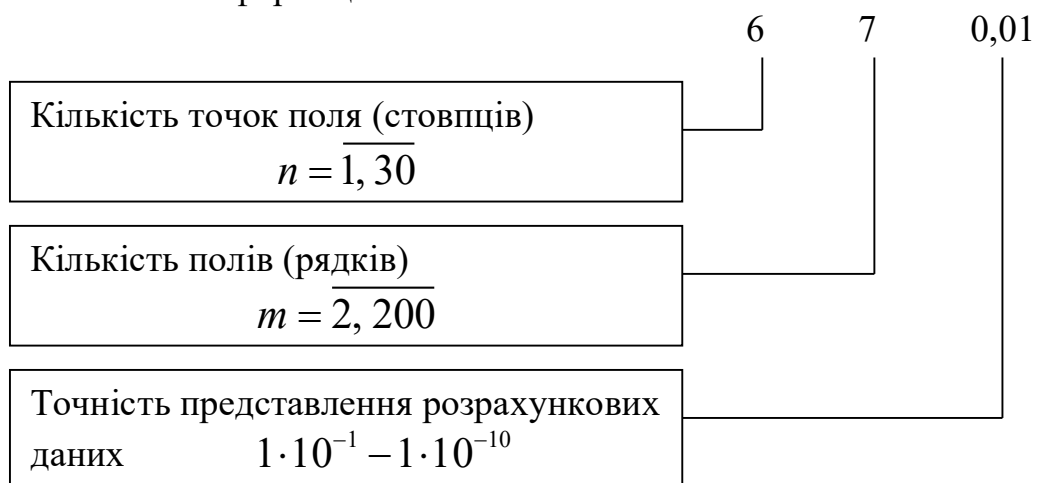
Програма дає можливість параметризувати вихідні поля ортогональними компонентами та розв'язувати задачі стиску й фільтрації інформації.

Файл даних до комп'ютерної програми формується таким чином, щоб кожний стовпець вихідної матриці *X* був часовим рядом метеорологічної величини на *i*-тій станції (*i*-тій точці поля).

Приклад формування файлу вихідних даних місячної кількості опадів для 6-ти станцій України (*n* = 6) за період 7 років (*m* = 7) наводиться нижче:

6	7	0,01				
4	39	23	56	47	89	
93	64	62	23	58	96	
18	62	31	12	78	58	
29	36	20	63	12	47	
12	56	74	28	36	45	
5	48	13	2	56	96	
45	3	78	96	85	12	

Обов'язкова інформація:



Відомості про комп'ютерну програму "FACTOR"

Комп'ютерна програма "FACTOR" дозволяє розрахувати середні значення вихідних векторів, стандартні відхилення, коефіцієнти кореляції між векторами, власні значення і власні вектори, а також ваги (навантаження) факторів у кожній змінній (у файлі результатів "**Обернена матриця факторів**"), сумарні вклади і частку сумарної дисперсії.

1. Порядок формування вихідної вибірки.

Програма передбачає формування вихідної вибірки в двох варіантах:

а) значення вихідних векторів знаходяться в строках, наприклад,

$$\begin{array}{ccccccccccc} x_{11}, & x_{12}, & x_{13}, & x_{14}, & x_{15}, & \dots, & x_{1k}, & \dots, & x_{1n}; \\ x_{21}, & x_{22}, & x_{23}, & x_{24}, & x_{25}, & \dots, & x_{2k}, & \dots, & x_{2n}; \\ x_{31}, & x_{32}, & x_{33}, & x_{34}, & x_{35}, & \dots, & x_{3k}, & \dots, & x_{3n}; \\ x_{41}, & x_{42}, & x_{43}, & x_{44}, & x_{45}, & \dots, & x_{4k}, & \dots, & x_{4n}. \end{array}$$

б) значення вихідних векторів знаходяться в стовпцях.

2. Послідовність відповідей на запити програми-фактор:

а) "**Назва масиву**": як правило, вводиться назва вихідного файлу, що містить вибірку з вихідними векторами.

б) "**Довжина ряду**": кількість значень містяться в кожному вихідному векторі (статистична сукупність).

в) "**Число змінних**": кількість вихідних векторів.

г) "**Якщо дані набиті по рядках введіть 1, ні -1**":

наприклад, якщо вихідна вибірка для 4 векторів має вигляд:

$$\begin{array}{ll} X_1: & x_{11}, \quad x_{12}, \quad x_{13}, \quad x_{14}, \quad x_{15}, \quad \dots, \quad x_{1k}, \quad \dots, \quad x_{1n}; \\ X_2: & x_{21}, \quad x_{22}, \quad x_{23}, \quad x_{24}, \quad x_{25}, \quad \dots, \quad x_{2k}, \quad \dots, \quad x_{2n}; \\ X_3: & x_{31}, \quad x_{32}, \quad x_{33}, \quad x_{34}, \quad x_{35}, \quad \dots, \quad x_{3k}, \quad \dots, \quad x_{3n}; \\ X_4: & x_{41}, \quad x_{42}, \quad x_{43}, \quad x_{44}, \quad x_{45}, \quad \dots, \quad x_{4k}, \quad \dots, \quad x_{4n}, \end{array}$$

тобто дані в векторах розташовані в строках – ставимо 1, в протилежному випадку (дані в векторах розташовані в стовпцях) -1.

д) "**Введіть файл даних**": вводиться назва вихідного файлу даних.

е) "**Введіть файл результатів**": вводиться назва кінцевого файлу для виводу результатів.

ж) Процедура повторюється, якщо є необхідність розрахунку для введення наступного файлу вихідних даних, в протилежному випадку натискається комбінація "**Ctrl^z**", а потім "**Enter**".

Відомості про комп'ютерну програму "Clast_uimkd"

Комп'ютерна програма "Clast_uimkd" дозволяє здійснити розбиття вихідної вибірки, що складається з певної кількості фізичних векторів на групи однорідних векторів. В результаті виконання програми можна визначити номери фізичних центрів і значення середніх векторів кластерів (еталони), номери векторів вихідної вибірки входять в певні кластери, середні значення і дисперсії (внутрікластерні) середніх векторів (еталонів), а також евклідові відстані між еталонами.

Програма передбачає формування вихідної вибірки тільки в одному варіанті - значення вихідних векторів знаходяться в рядках.

Вигляд файлу вихідної вибірки:

n	m								
x_{11} ,	x_{12} ,	x_{13} ,	x_{14} ,	x_{15} ,	...	x_{1k} ,	...	x_{1n} ;	
x_{21} ,	x_{22} ,	x_{23} ,	x_{24} ,	x_{25} ,	...	x_{2k} ,	...	x_{2n} ;	
...	...	...	...	...	...	...	...	...	...
x_{m1} ,	x_{m2} ,	x_{m3} ,	x_{m4} ,	x_{m5} ,	...	x_{mk} ,	...	x_{mn} ,	

де

n – кількість вихідних векторів;

m – кількість значень, що міститься в кожному вихідному векторі (статистична сукупність);

x_{ij} – значення вихідних векторів ($i = \overline{1, n}$; $j = \overline{1, m}$).

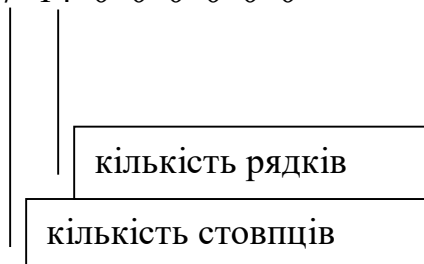
Відомості про комп'ютерну програму “STEP”:

Зразок файлу вихідних даних до програми “STEP”:

7 14 0 0 0 0 0 0

1017.7	997.5	1009.2	1021	1024	1008.0	3.4
1006.2	994.0	1011.0	1024.0	1026.0	1019.0	-2.6
1003.0	1008.7	1001.5	1023.0	1022.0	1025.0	-0.5
1004.5	1007.8	1009.3	1014.0	1020.0	1010.0	-3.5
1005.5	1001.5	998.8	1019.0	1028.0	1006.0	0.2
1003.5	1011.8	999.8	1021.0	1021.0	1015.8	-0.5
1008.2	997.9	995.0	1027.1	1023.1	1020.8	-2.8
1000.2	987.6	996.7	1023.7	1022.1	1021.7	1.0
1010.7	1004.9	977.7	1019.8	1024.6	1021.0	0.8
999.5	992.5	988.9	1021.3	1025.3	1023.4	-0.4
1000.5	1005.8	999.1	1023.2	1021.2	1024.8	-5.5
1000.0	1012.1	1010.6	1023.0	1016.9	1018.6	3.5
1002.6	999.6	996.9	1017.3	1014.7	1028.4	-0.6
992.3	1005.9	1012.0	1023.9	1011.7	1014.0	-1.0

Обов'язкова інформація: 7 14 0 0 0 0 0 0



Приклад для 6 предікторів та одного предіктанта з об'ємами вибірок $m = 14$. Останній стовпчик (7-й) – предіктант.

ДОДАТОК В

Вихідні дані до виконання практичних робіт

Таблиця В.1 - Місячна кількість опадів (січень)

Станція		Рік									
		1976	1977	1978	1979	1980	1981	1982	1983	1984	1985
1	Луцьк	80	22	32	76	35	42	18	38	28	20
2	Рівне	64	24	19	69	23	36	16	26	26	20
3	Львів	153	20	43	83	37	45	25	63	30	24
4	Тернопіль	100	13	16	78	52	35	18	28	25	18
5	Ужгород	91	95	43	133	35	55	69	71	52	61
6	Івано-Франківськ	66	14	21	55	46	40	35	28	34	24
7	Житомир	59	18	25	76	37	38	9	33	17	40
8	Хмельницький	120	22	21	99	52	40	18	36	34	36
9	Чернівці	68	18	9	58	54	43	24	27	29	24
10	Вінниця	65	27	34	76	44	44	14	34	34	43
11	Київ	52	37	48	71	64	38	16	39	42	74
12	Чернігів	38	42	11	52	61	26	19	47	41	66
13	Семенівка	37	46	16	53	49	34	33	51	45	63
14	Суми	52	55	10	62	39	56	40	42	45	59
15	Гадяч	49	54	13	78	54	60	27	34	48	70
16	Полтава	48	41	12	75	73	73	39	37	52	50
17	Харків	62	32	10	74	59	56	49	40	61	65
18	Черкаси	30	44	16	38	44	34	17	19	35	43
19	Ізюм	100	44	19	87	59	53	58	35	71	53
20	Луганськ	41	12	43	38	30	31	53	25	31	45
21	Дніпропетровськ	91	26	22	69	52	68	62	28	54	47
22	Запоріжжя	85	37	17	84	54	70	70	26	56	65
23	Донецьк	60	15	46	44	34	53	39	41	51	41
24	Бердянськ	73	27	34	57	52	46	33	20	16	22
25	Мелітополь	67	29	29	68	66	68	48	15	19	43
26	В.Олександрівка	56	31	10	53	43	59	30	14	46	33
27	Кіровоград	32	41	10	64	34	44	21	12	21	35
28	Вознесенськ	46	36	24	86	43	45	18	17	30	51
29	Любашівка	45	26	33	97	44	51	19	19	44	59
30	Херсон	36	35	14	78	44	50	17	12	36	29
31	Миколаїв	42	27	16	85	32	46	20	12	39	40
32	Роздільна	45	25	13	62	36	45	13	8	46	48
33	Одеса	33	24	18	66	48	55	16	11	42	49

Продовження таблиці В.1

Станція		Рік									
		1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
1	Луцьк	20	39	26	9	11	13	13	29	20	9
2	Рівне	25	41	26	16	12	15	18	29	20	14
3	Львів	35	47	54	16	16	23	32	33	39	29
4	Тернопіль	25	46	47	10	7	16	19	13	27	30
5	Ужгород	99	75	77	32	35	32	23	29	63	64
6	Івано-Франківськ	19	35	49	7	7	17	19	19	16	26
7	Житомир	56	43	20	10	14	17	17	23	37	39
8	Хмельницький	29	43	48	10	9	8	22	25	32	39
9	Чернівці	17	43	66	4	1	23	20	20	13	24
10	Вінниця	58	49	22	10	12	15	25	16	37	39
11	Київ	58	66	8	13	25	18	20	36	40	35
12	Чернігів	0	71	15	21	46	18	18	40	39	28
13	Семенівка	50	45	20	30	58	26	25	44	42	40
14	Суми	50	42	30	21	29	22	29	29	30	51
15	Гадяч	60	85	25	27	29	21	35	33	41	58
16	Полтава	69	43	27	16	31	26	21	27	44	72
17	Харків	47	39	27	21	30	27	17	20	27	61
18	Черкаси	44	49	18	10	18	22	22	23	44	36
19	Ізюм	99	66	34	34	49	40	30	34	32	80
20	Луганськ	95	63	12	24	46	40	22	20	34	47
21	Дніпропетровськ	85	67	17	24	31	28	27	21	16	48
22	Запоріжжя	90	75	17	18	27	30	31	16	16	57
23	Донецьк	96	59	15	21	34	19	26	20	36	66
24	Бердянськ	94	98	12	17	22	26	24	15	35	62
25	Мелітополь	95	60	17	17	15	16	19	15	25	44
26	В.Олександрівка	67	54	9	7	18	19	14	5	12	62
27	Кіровоград	36	44	13	10	9	15	28	18	24	34
28	Вознесенськ	37	49	11	5	9	15	9	8	15	38
29	Любашівка	36	42	20	7	4	10	10	11	27	39
30	Херсон	67	62	13	6	20	18	13	4	12	62
31	Миколаїв	61	59	17	3	15	12	11	5	10	44
32	Роздільна	55	46	34	8	13	10	14	9	21	39
33	Одеса	68	61	69	4	9	7	8	5	10	51

Продовження таблиці В.1

Станція		Рік									
		1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
1	Луцьк	21	3	19	16	39	46	18	20	39	32
2	Рівне	29	15	22	17	31	55	26	28	40	27
3	Львів	27	13	41	27	61	86	35	47	42	71
4	Тернопіль	28	9	24	20	42	31	17	24	30	35
5	Ужгород	35	23	36	31	59	62	37	61	74	57
6	Івано-Франківськ	38	17	22	26	25	30	17	31	35	40
7	Житомир	42	7	14	34	36	35	20	35	56	61
8	Хмельницький	48	6	26	25	35	42	25	36	44	40
9	Чернівці	41	10	26	27	18	28	14	25	24	48
10	Вінниця	32	7	22	27	24	35	11	33	43	39
11	Київ	37	9	36	48	31	33	22	37	55	50
12	Чернігів	16	10	24	45	31	33	23	22	74	51
13	Семенівка	24	9	32	24	28	51	38	22	84	61
14	Суми	22	13	25	49	19	40	17	47	54	48
15	Гадяч	31	14	39	53	28	38	24	63	69	44
16	Полтава	42	14	34	63	37	29	14	58	72	33
17	Харків	21	15	22	33	30	33	12	40	56	47
18	Черкаси	16	8	22	41	35	17	9	65	80	33
19	Ізюм	27	29	27	22	64	20	15	82	85	54
20	Луганськ	12	30	16	27	39	15	13	43	80	72
21	Дніпропетровськ	48	29	28	48	71	20	15	63	106	57
22	Запоріжжя	36	28	24	44	53	16	10	68	110	49
23	Донецьк	9	23	15	23	30	16	12	42	114	56
24	Бердянськ	15	28	38	20	40	31	23	53	169	62
25	Мелітополь	21	31	29	21	47	24	13	51	106	51
26	В.Олександрівка	30	16	32	32	38	21	9	59	92	29
27	Кіровоград	28	12	17	44	41	16	7	77	101	21
28	Вознесенськ	31	14	19	31	43	9	8	70	101	18
29	Любашівка	34	16	24	32	41	12	13	72	68	35
30	Херсон	24	12	33	43	28	17	10	48	60	25
31	Миколаїв	27	7	21	38	33	9	7	41	74	21
32	Роздільна	41	3	13	23	45	16	11	58	109	17
33	Одеса	27	15	22	28	61	12	9	74	113	25

Таблиця В.2 - Місячна кількість опадів (лютень)

Станція		Рік									
		1976	1977	1978	1979	1980	1981	1982	1983	1984	1985
1	Луцьк	1	59	42	25	16	25	21	33	19	22
2	Рівне	4	66	26	24	10	31	12	23	16	18
3	Львів	3	94	48	30	25	35	41	31	33	65
4	Тернопіль	3	50	50	10	21	46	31	14	39	54
5	Ужгород	3	100	37	40	38	32	13	48	44	56
6	Івано-Франківськ	2	37	52	17	19	47	36	13	82	58
7	Житомир	10	62	41	12	18	38	28	24	41	25
8	Хмельницький	7	49	54	9	17	42	30	22	63	51
9	Чернівці	3	28	67	12	14	26	29	8	77	36
10	Вінниця	8	69	68	10	18	32	22	14	62	39
11	Київ	6	104	64	17	24	47	31	41	56	35
12	Чернігів	2	66	29	42	13	49	26	58	38	27
13	Семенівка	7	53	35	37	26	51	22	53	10	29
14	Суми	4	36	48	33	34	54	30	66	17	42
15	Гадяч	3	81	73	31	26	76	33	62	18	49
16	Полтава	2	84	65	26	26	64	26	50	16	29
17	Харків	4	73	31	25	15	54	20	44	17	34
18	Черкаси	1	84	50	10	22	31	12	40	27	31
19	Ізюм	5	108	52	38	53	62	23	43	53	59
20	Луганськ	5	55	16	74	9	70	23	27	12	90
21	Дніпропетровськ	3	50	57	35	39	51	18	50	36	79
22	Запоріжжя	3	42	48	59	33	57	11	38	31	75
23	Донецьк	4	62	43	75	33	53	17	36	12	51
24	Бердянськ	3	45	41	67	42	42	8	42	13	72
25	Мелітополь	1	45	48	95	24	59	8	35	20	85
26	В.Олександрівка	3	25	64	32	22	51	18	51	39	72
27	Кіровоград	2	56	55	24	22	36	15	34	34	30
28	Вознесенськ	3	43	70	30	9	41	23	25	47	57
29	Любашівка	6	63	54	19	13	36	19	19	62	38
30	Херсон	2	40	57	41	10	34	15	24	41	69
31	Миколаїв	3	42	71	34	11	43	25	42	42	70
32	Роздільна	8	50	54	32	6	28	38	22	73	44
33	Одеса	11	52	59	23	4	38	33	26	69	66

Продовження таблиці В.2

Станція		Рік									
		1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
1	Луцьк	19	6	13	17	25	25	22	19	20	34
2	Рівне	21	14	12	23	25	25	23	25	18	38
3	Львів	31	17	34	14	42	36	39	39	34	65
4	Тернопіль	27	11	23	18	26	21	28	26	26	36
5	Ужгород	32	20	81	50	35	72	32	40	45	80
6	Івано-Франківськ	21	8	20	5	19	15	16	28	33	33
7	Житомир	22	5	11	21	21	36	24	43	13	36
8	Хмельницький	17	16	14	19	27	28	34	40	25	38
9	Чернівці	37	21	15	3	29	23	17	31	24	21
10	Вінниця	20	16	23	18	27	27	18	45	17	21
11	Київ	33	11	17	57	54	52	29	47	13	45
12	Чернігів	25	20	29	38	39	28	29	20	15	60
13	Семенівка	20	28	42	29	42	39	34	21	18	52
14	Суми	34	16	29	17	49	30	18	17	18	42
15	Гадяч	30	22	43	26	80	43	32	16	20	52
16	Полтава	46	13	40	22	53	25	14	27	16	41
17	Харків	27	14	30	20	34	34	17	26	7	43
18	Черкаси	33	8	34	16	52	28	19	28	15	31
19	Ізюм	38	24	35	34	30	29	37	29	29	62
20	Луганськ	17	36	37	15	7	33	58	27	16	41
21	Дніпропетровськ	48	14	24	14	39	30	44	36	23	54
22	Запоріжжя	51	18	21	19	47	28	46	39	14	43
23	Донецьк	20	25	28	20	28	19	41	7	13	44
24	Бердянськ	37	51	18	21	31	21	54	17	95	31
25	Мелітополь	43	31	14	35	25	17	38	34	4	26
26	В.Олександрівка	48	11	15	13	32	24	29	34	7	36
27	Кіровоград	42	5	40	24	35	19	9	25	12	22
28	Вознесенськ	51	7	17	20	14	24	14	28	4	24
29	Любашівка	33	11	12	22	26	30	14	40	3	20
30	Херсон	37	13	16	16	36	27	29	24	4	26
31	Миколаїв	49	10	28	19	25	17	13	29	3	21
32	Роздільна	52	6	23	13	18	21	23	29	2	19
33	Одеса	55	13	15	10	20	24	23	33	2	20

Продовження таблиці В.2

Станція		Рік									
		1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
1	Луцьк	17	8	19	44	23	27	24	38	43	60
2	Рівне	26	16	24	56	21	29	35	28	43	61
3	Львів	69	35	31	100	39	51	27	40	83	40
4	Тернопіль	23	20	18	60	25	28	30	26	48	56
5	Ужгород	44	45	19	120	65	36	44	29	78	62
6	Івано-Франківськ	28	19	30	79	19	30	33	34	40	41
7	Житомир	37	19	17	48	33	38	34	27	56	58
8	Хмельницький	22	15	16	20	24	34	49	29	50	59
9	Чернівці	34	7	15	77	20	24	16	29	32	63
10	Вінниця	36	16	13	36	30	21	30	21	39	40
11	Київ	43	24	14	41	37	54	38	18	45	62
12	Чернігів	29	27	17	56	38	50	41	18	57	54
13	Семенівка	27	32	23	50	56	48	50	12	69	50
14	Суми	27	26	19	39	31	37	42	46	55	40
15	Гадяч	39	30	18	52	36	49	40	19	51	64
16	Полтава	46	37	21	38	31	34	29	29	37	61
17	Харків	54	49	20	49	24	45	28	11	50	42
18	Черкаси	39	18	7	27	25	35	17	13	55	55
19	Ізюм	70	41	27	68	34	54	30	24	109	43
20	Луганськ	58	20	24	47	19	32	37	49	84	36
21	Дніпропетровськ	65	54	26	49	49	67	21	45	87	81
22	Запоріжжя	57	50	17	41	40	70	15	47	57	87
23	Донецьк	33	23	28	48	19	29	23	27	99	37
24	Бердянськ	72	19	42	64	24	58	14	58	128	42
25	Мелітополь	73	24	23	34	28	55	11	53	55	52
26	В.Олександрівка	50	24	13	27	21	33	26	51	40	87
27	Кіровоград	48	24	8	26	26	41	12	22	30	68
28	Вознесенськ	53	18	15	33	35	27	5	45	40	79
29	Любашівка	46	11	11	37	23	19	2	30	41	64
30	Херсон	42	24	19	29	25	38	22	49	33	79
31	Миколаїв	37	19	11	35	16	25	17	33	36	71
32	Роздільна	61	17	12	51	27	17	5	51	38	84
33	Одеса	74	18	17	47	40	36	32	65	30	95

Таблиця В.3 - Місячна кількість опадів (березень)

Станція		Рік									
		1976	1977	1978	1979	1980	1981	1982	1983	1984	1985
1	Луцьк	54	44	31	33	32	23	10	73	24	16
2	Рівне	50	22	34	42	24	23	8	60	22	23
3	Львів	84	79	30	62	58	53	10	65	40	31
4	Тернопіль	56	35	14	58	57	65	8	53	30	10
5	Ужгород	54	71	18	78	11	57	24	70	36	61
6	Івано-Франківськ	33	36	23	48	64	48	19	41	46	15
7	Житомир	44	26	51	63	67	40	10	44	23	13
8	Хмельницький	55	36	37	51	67	56	8	35	32	5
9	Чернівці	22	24	26	26	66	66	18	31	37	7
10	Вінниця	44	22	33	55	82	50	6	32	33	9
11	Київ	55	22	33	70	66	42	8	39	49	12
12	Чернігів	47	25	22	67	33	31	10	47	27	8
13	Семенівка	71	23	54	48	26	34	18	53	30	8
14	Суми	49	18	77	48	57	40	12	56	22	8
15	Гадяч	67	17	99	64	68	52	6	52	24	7
16	Полтава	51	12	90	64	42	34	5	38	24	8
17	Харків	37	9	51	33	15	37	6	32	18	7
18	Черкаси	36	20	44	26	61	40	2	32	32	5
19	Ізюм	48	16	44	27	29	48	17	37	50	19
20	Луганськ	17	34	12	19	23	31	44	24	19	15
21	Дніпропетровськ	36	16	74	41	30	35	16	27	35	9
22	Запоріжжя	38	20	77	43	35	55	24	15	40	8
23	Донецьк	27	22	14	33	22	52	27	24	17	11
24	Бердянськ	22	23	18	28	37	50	40	15	22	11
25	Мелітополь	16	29	37	41	32	29	34	12	23	7
26	В.Олександрівка	32	13	56	35	43	37	12	12	44	16
27	Кіровоград	30	7	63	27	65	36	3	14	27	5
28	Вознесенськ	21	11	56	60	107	43	12	15	43	6
29	Любашівка	35	11	43	56	88	43	9	22	51	10
30	Херсон	26	15	54	46	88	23	21	11	45	16
31	Миколаїв	30	8	57	42	89	40	14	15	63	8
32	Роздільна	21	6	50	29	64	41	15	13	47	3
33	Одеса	32	8	56	31	85	38	20	7	48	14

Продовження таблиці В.3

Станція		Рік									
		1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
1	Луцьк	9	34	29	8	20	1	17	14	38	36
2	Рівне	7	30	38	14	20	2	14	19	48	38
3	Львів	12	50	71	13	42	5	25	61	49	27
4	Тернопіль	11	43	39	6	13	11	10	17	37	45
5	Ужгород	27	84	75	31	16	24	27	43	35	42
6	Івано-Франківськ	9	43	57	14	27	11	22	46	24	41
7	Житомир	7	28	55	35	30	4	32	30	37	32
8	Хмельницький	13	26	33	12	12	8	16	24	32	34
9	Чернівці	12	27	42	17	14	14	23	76	27	44
10	Вінниця	10	28	46	43	10	8	56	43	20	23
11	Київ	11	36	55	35	7	7	41	51	72	33
12	Чернігів	12	41	48	42	27	6	41	33	57	37
13	Семенівка	13	21	63	50	27	9	42	30	53	26
14	Суми	20	20	72	33	19	6	23	35	53	33
15	Гадяч	21	42	98	31	23	2	29	37	80	42
16	Полтава	8	25	79	31	22	8	29	47	66	57
17	Харків	2	42	49	20	17	9	20	39	36	57
18	Черкаси	16	31	53	23	8	4	42	59	33	50
19	Ізюм	3	63	44	25	34	10	24	47	80	50
20	Луганськ	2	41	21	19	29	14	28	23	50	19
21	Дніпропетровськ	4	51	87	29	15	23	31	38	84	61
22	Запоріжжя	4	52	82	25	15	25	25	29	60	66
23	Донецьк	4	52	29	21	30	13	25	19	48	13
24	Бердянськ	3	67	67	14	11	31	30	27	48	61
25	Мелітополь	4	60	84	17	11	19	34	27	38	45
26	В.Олександрівка	3	23	53	13	2	7	19	39	34	42
27	Кіровоград	6	44	46	23	6	5	23	55	24	48
28	Вознесенськ	2	25	68	11	11	2	23	61	13	37
29	Любашівка	11	15	71	19	10	33	31	45	12	25
30	Херсон	2	22	55	15	3	20	23	38	22	49
31	Миколаїв	4	21	64	13	5	2	17	37	17	30
32	Роздільна	3	19	81	19	2	1	39	53	8	20
33	Одеса	3	15	91	12	1	1	27	64	10	63

Продовження таблиці В.3

Станція		Рік									
		1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
1	Луцьк	13	17	22	18	56	48	26	14	46	39
2	Рівне	30	18	32	21	59	59	30	23	38	33
3	Львів	35	26	51	12	77	108	40	20	49	54
4	Тернопіль	17	15	45	13	65	88	26	26	25	42
5	Ужгород	20	12	31	23	91	119	29	20	25	22
6	Івано-Франківськ	21	13	32	10	62	97	60	23	22	17
7	Житомир	24	23	49	50	48	67	17	29	28	89
8	Хмельницький	25	14	39	28	65	56	17	34	23	50
9	Чернівці	25	17	47	8	39	35	54	24	14	24
10	Вінниця	19	20	42	35	50	56	16	29	12	31
11	Київ	45	30	49	43	34	89	17	26	21	52
12	Чернігів	37	31	42	34	47	66	12	24	33	35
13	Семенівка	33	31	64	27	48	44	10	29	42	27
14	Суми	19	22	88	35	54	84	21	24	53	19
15	Гадяч	30	28	76	40	49	99	26	36	43	26
16	Полтава	20	24	69	53	47	74	28	29	36	22
17	Харків	21	24	46	25	65	63	26	28	43	25
18	Черкаси	44	21	91	52	42	74	26	21	16	15
19	Ізюм	27	65	88	49	90	65	41	27	30	41
20	Луганськ	24	44	46	28	52	36	73	22	26	32
21	Дніпропетровськ	28	64	90	48	69	64	51	40	40	35
22	Запоріжжя	20	57	77	33	46	50	53	41	31	24
23	Донецьк	25	46	71	32	77	35	43	14	24	33
24	Бердянськ	39	58	107	52	89	45	39	46	19	30
25	Мелітополь	23	66	88	38	42	34	54	40	13	10
26	В.Олександрівка	26	36	75	42	29	60	64	27	23	18
27	Кіровоград	36	13	78	62	40	55	17	21	17	16
28	Вознесенськ	28	24	68	32	32	56	37	26	21	13
29	Любашівка	27	14	57	30	28	52	49	17	8	24
30	Херсон	23	40	46	31	28	66	49	45	14	15
31	Миколаїв	28	38	40	19	8	57	39	30	19	12
32	Роздільна	18	32	63	20	11	55	65	24	18	30
33	Одеса	26	50	40	19	12	61	70	39	19	19

Таблиця В.4 - Місячна кількість опадів (квітень)

Станція		Рік									
		1976	1977	1978	1979	1980	1981	1982	1983	1984	1985
1	Луцьк	39	47	40	31	61	7	36	36	6	63
2	Рівне	84	70	65	59	34	7	52	31	9	52
3	Львів	43	48	88	63	100	29	60	58	11	65
4	Тернопіль	93	60	72	62	81	19	41	59	11	50
5	Ужгород	38	69	61	68	51	13	32	66	16	73
6	Івано-Франківськ	114	72	69	66	92	28	81	71	12	71
7	Житомир	58	84	39	73	47	21	88	77	9	45
8	Хмельницький	69	111	63	70	74	14	71	37	15	55
9	Чернівці	58	104	71	96	63	26	78	38	27	39
10	Вінниця	48	120	73	71	35	24	111	51	22	71
11	Київ	65	128	13	96	62	33	100	54	14	35
12	Чернігів	52	47	13	54	67	32	74	49	14	37
13	Семенівка	36	40	22	49	59	38	80	52	8	40
14	Суми	29	36	18	59	48	58	78	29	5	19
15	Гадяч	54	23	26	67	45	62	89	32	17	23
16	Полтава	78	27	31	90	51	65	80	22	37	15
17	Харків	34	40	29	57	42	30	39	37	10	17
18	Черкаси	36	50	34	82	50	34	44	27	30	25
19	Ізюм	28	40	50	87	76	65	34	26	14	8
20	Луганськ	73	67	46	41	121	60	45	17	12	17
21	Дніпропетровськ	82	47	33	75	79	57	45	30	31	28
22	Запоріжжя	62	44	35	44	70	52	27	11	54	20
23	Донецьк	32	51	50	47	66	62	41	26	39	31
24	Бердянськ	31	48	30	31	64	54	67	24	69	28
25	Мелітополь	12	57	50	26	38	38	34	8	67	35
26	В.Олександрівка	40	121	44	37	57	70	33	23	45	33
27	Кіровоград	53	69	44	67	49	47	68	13	48	18
28	Вознесенськ	50	108	42	37	42	37	45	28	57	38
29	Любашівка	56	110	83	64	51	38	64	23	68	54
30	Херсон	17	93	51	34	73	52	36	22	63	41
31	Миколаїв	24	104	41	45	54	52	36	16	59	25
32	Роздільна	22	84	47	50	76	20	48	28	120	32
33	Одеса	19	138	44	44	75	27	38	20	63	27

Продовження таблиці В.4

Станція		Рік									
		1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
1	Луцьк	33	39	22	66	34	28	19	17	19	51
2	Рівне	32	34	17	49	35	27	15	27	20	50
3	Львів	38	15	48	50	87	38	67	28	25	70
4	Тернопіль	55	14	29	61	42	35	55	25	13	44
5	Ужгород	54	38	48	73	52	55	46	24	92	69
6	Івано-Франківськ	61	12	55	72	44	58	53	42	21	40
7	Житомир	39	26	30	54	37	51	56	53	65	39
8	Хмельницький	46	18	43	59	34	47	59	60	40	69
9	Чернівці	36	36	72	78	51	50	50	77	15	42
10	Вінниця	39	16	41	37	66	63	58	70	22	38
11	Київ	31	72	36	28	92	88	84	36	29	69
12	Чернігів	35	38	22	19	96	26	48	45	29	67
13	Семенівка	44	20	11	28	69	16	44	36	23	72
14	Суми	67	49	25	38	102	33	44	45	42	90
15	Гадяч	50	43	44	48	77	29	26	37	45	95
16	Полтава	52	37	32	56	44	17	24	41	71	100
17	Харків	72	34	20	59	68	3	21	18	39	63
18	Черкаси	21	39	31	31	72	45	27	39	28	68
19	Ізюм	55	41	22	64	41	17	45	29	33	74
20	Луганськ	15	44	31	43	12	19	22	43	19	35
21	Дніпропетровськ	43	58	38	32	43	11	44	52	80	77
22	Запоріжжя	42	65	39	17	37	16	40	70	95	84
23	Донецьк	46	49	38	41	19	14	23	68	36	62
24	Бердянськ	20	35	28	9	21	55	15	46	52	73
25	Мелітополь	67	45	35	12	22	27	16	86	63	69
26	В.Олександрівка	48	31	16	23	52	17	11	33	85	68
27	Кіровоград	13	29	21	26	80	34	12	49	45	72
28	Вознесеньск	14	32	25	50	75	24	10	44	22	56
29	Любашівка	19	19	45	25	40	45	20	43	27	36
30	Херсон	37	33	44	18	34	18	6	50	69	55
31	Миколаїв	17	43	23	18	33	19	5	43	26	64
32	Роздільна	8	27	52	35	35	50	10	18	11	43
33	Одеса	15	23	40	14	28	31	14	37	8	51

Продовження таблиці В.4

Станція		Рік									
		1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
1	Луцьк	13	41	65	67	50	25	40	45	24	52
2	Рівне	28	46	81	63	71	22	35	27	21	47
3	Львів	23	71	105	92	63	73	41	31	31	80
4	Тернопіль	27	36	70	35	44	49	30	16	18	63
5	Ужгород	30	44	114	45	65	39	22	30	36	71
6	Івано-Франківськ	66	56	44	89	41	46	35	12	16	56
7	Житомир	30	20	83	42	48	52	41	23	20	53
8	Хмельницький	71	53	49	35	48	58	39	12	24	43
9	Чернівці	63	56	45	97	28	65	37	11	13	63
10	Вінниця	84	36	32	30	43	42	39	12	22	62
11	Київ	42	41	99	20	29	65	40	24	21	68
12	Чернігів	28	64	110	23	41	69	29	27	28	27
13	Семенівка	8	66	106	23	30	32	40	33	36	39
14	Суми	29	72	47	14	57	51	33	38	41	23
15	Гадяч	56	56	54	29	50	48	40	57	33	14
16	Полтава	68	81	45	17	24	74	28	24	19	14
17	Харків	38	99	27	18	21	48	12	27	28	11
18	Черкаси	48	87	41	29	51	43	21	26	6	20
19	Ізюм	61	58	37	45	43	54	21	23	11	20
20	Луганськ	48	48	24	21	15	64	26	23	17	24
21	Дніпропетровськ	58	91	47	43	20	49	27	31	14	20
22	Запоріжжя	54	72	43	38	17	58	24	38	3	28
23	Донецьк	42	56	27	25	36	60	31	25	17	47
24	Бердянськ	27	68	31	21	24	65	26	28	19	21
25	Мелітополь	32	70	23	29	15	62	36	30	15	39
26	В.Олександрівка	30	61	34	44	30	38	21	22	6	21
27	Кіровоград	33	69	57	26	42	37	29	26	17	8
28	Вознесенськ	28	63	29	70	28	44	11	22	17	9
29	Любашівка	46	77	23	76	31	42	8	30	9	33
30	Херсон	24	51	26	26	31	39	12	14	13	16
31	Миколаїв	27	42	38	28	46	40	12	40	9	9
32	Роздільна	33	56	10	35	26	31	8	40	22	19
33	Одеса	23	53	14	35	21	26	11	39	29	3

Таблиця В.5 - Місячна кількість опадів (травень)

Станція		Рік									
		1976	1977	1978	1979	1980	1981	1982	1983	1984	1985
1	Луцьк	40	56	85	43	69	43	52	101	136	49
2	Рівне	24	50	91	29	74	36	48	98	87	29
3	Львів	68	23	172	44	54	91	38	40	100	123
4	Тернопіль	42	106	97	25	54	91	29	78	126	39
5	Ужгород	139	21	142	34	51	58	31	99	118	103
6	Івано-Франківськ	102	115	178	50	73	88	30	70	137	92
7	Житомир	24	43	40	60	43	70	91	109	64	43
8	Хмельницький	43	35	55	54	83	70	15	236	77	21
9	Чернівці	77	48	177	22	93	108	49	84	166	59
10	Вінниця	39	79	55	47	55	40	14	133	65	53
11	Київ	60	14	46	28	43	26	25	108	63	87
12	Чернігів	51	18	92	6	32	26	11	43	84	34
13	Семенівка	46	71	63	12	68	19	23	55	79	39
14	Суми	31	73	86	6	66	34	21	107	36	60
15	Гадяч	47	68	91	11	92	13	21	37	43	56
16	Полтава	39	55	81	22	58	34	36	51	46	108
17	Харків	26	92	103	1	46	39	20	78	19	64
18	Черкаси	36	57	26	35	70	31	12	29	31	58
19	Ізюм	32	113	138	6	70	58	50	40	14	26
20	Луганськ	13	37	43	3	55	14	62	54	46	37
21	Дніпропетровськ	26	25	138	8	50	15	26	33	33	32
22	Запоріжжя	13	50	80	11	45	25	19	36	29	41
23	Донецьк	31	54	116	8	66	25	11	70	31	104
24	Бердянськ	11	58	48	37	30	79	7	18	7	74
25	Мелітополь	39	38	49	38	20	61	21	34	9	89
26	В.Олександрівка	39	40	69	31	45	74	22	55	2	88
27	Кіровоград	35	100	31	16	58	43	41	19	37	49
28	Вознесенськ	34	61	59	28	68	62	17	47	22	20
29	Любашівка	55	37	84	56	55	63	30	97	39	45
30	Херсон	13	47	48	104	52	78	17	33	7	42
31	Миколаїв	20	26	68	65	32	68	24	59	4	55
32	Роздільна	26	18	79	59	67	30	20	51	24	54
33	Одеса	51	19	62	46	40	46	9	24	5	34

Продовження таблиці В.5

Станція		Рік									
		1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
1	Луцьк	58	56	30	63	46	65	82	18	71	45
2	Рівне	23	57	40	74	36	69	71	52	71	57
3	Львів	52	81	57	149	89	169	94	44	57	79
4	Тернопіль	25	75	100	72	42	80	72	61	79	69
5	Ужгород	50	87	69	143	48	51	41	76	81	39
6	Івано-Франківськ	26	112	91	131	51	115	46	41	68	38
7	Житомир	6	54	47	51	59	95	116	25	65	63
8	Хмельницький	27	50	60	55	30	115	72	37	48	70
9	Чернівці	24	110	76	99	25	198	58	61	99	76
10	Вінниця	13	65	67	59	44	78	35	44	54	50
11	Київ	17	91	75	29	46	117	73	25	61	106
12	Чернігів	1	89	42	30	75	109	42	53	70	109
13	Семенівка	20	65	53	53	60	110	25	40	72	102
14	Суми	32	30	73	29	51	113	53	60	48	50
15	Гадяч	32	21	51	80	61	94	36	24	57	19
16	Полтава	21	36	43	23	71	107	77	30	77	33
17	Харків	55	35	82	32	78	104	85	27	103	20
18	Черкаси	1	23	51	33	33	95	53	130	80	89
19	Ізюм	31	21	51	52	34	95	130	13	47	36
20	Луганськ	49	39	29	55	31	67	119	52	25	25
21	Дніпропетровськ	37	49	34	37	47	74	52	63	51	32
22	Запоріжжя	12	25	54	33	33	57	56	37	32	40
23	Донецьк	36	31	9	32	33	63	132	71	55	29
24	Бердянськ	14	83	23	23	20	71	63	42	47	53
25	Мелітополь	19	43	39	29	16	60	79	43	40	44
26	В.Олександрівка	8	36	56	21	34	40	76	62	22	48
27	Кіровоград	2	46	49	38	41	74	67	39	45	42
28	Вознесенськ	11	30	67	30	30	103	56	115	27	51
29	Любашівка	10	31	30	34	50	106	78	82	55	31
30	Херсон	36	24	45	38	41	59	35	34	28	59
31	Миколаїв	9	31	75	10	24	90	37	69	16	40
32	Роздільна	2	28	32	33	40	106	37	80	19	25
33	Одеса	5	32	46	13	48	80	34	35	18	29

Продовження таблиці В.5

Станція		Рік									
		1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
1	Луцьк	46	54	49	30	27	31	36	69	96	52
2	Рівне	51	67	35	25	23	67	118	52	73	79
3	Львів	92	132	65	27	66	55	91	104	85	89
4	Тернопіль	59	55	31	17	61	46	26	52	72	53
5	Ужгород	96	139	88	28	39	36	45	32	83	121
6	Івано-Франківськ	56	60	88	45	61	62	97	35	61	50
7	Житомир	63	69	52	36	44	43	28	12	57	101
8	Хмельницький	44	70	46	30	40	33	62	17	36	42
9	Чернівці	28	67	110	23	26	41	48	17	61	68
10	Вінниця	39	37	43	37	47	44	74	30	31	72
11	Київ	28	48	31	48	73	33	65	49	53	60
12	Чернігів	56	43	39	56	76	43	61	62	58	50
13	Семенівка	45	42	54	100	71	56	67	62	43	57
14	Суми	90	71	24	35	63	54	124	12	80	19
15	Гадяч	54	67	18	35	39	64	114	12	76	19
16	Полтава	50	65	26	55	50	48	110	16	89	26
17	Харків	39	75	41	43	46	28	48	12	1022	26
18	Черкаси	25	76	31	31	76	72	52	12	82	38
19	Ізюм	29	48	39	81	49	39	25	9	132	45
20	Луганськ	60	33	49	79	39	57	19	2	37	34
21	Дніпропетровськ	24	38	34	47	21	52	26	5	145	15
22	Запоріжжя	21	35	35	58	8	64	39	1	108	11
23	Донецьк	59	32	58	38	30	75	25	2	72	5
24	Бердянськ	36	61	97	67	9	57	38	0	28	11
25	Мелітополь	44	11	85	68	23	31	8	0	73	89
26	В.Олександрівка	35	23	53	63	17	66	7	32	210	26
27	Кіровоград	17	80	44	25	41	98	41	14	58	18
28	Вознесенськ	24	76	102	46	55	88	41	22	88	30
29	Любашівка	30	27	46	31	34	85	26	12	77	38
30	Херсон	27	27	143	48	21	38	7	54	97	17
31	Миколаїв	23	59	47	36	5	41	20	48	156	24
32	Роздільна	33	21	52	31	14	53	17	43	104	44
33	Одеса	23	31	34	37	28	56	2	4	104	39

Таблиця В.6 - Місячна кількість опадів (червень)

Станція		Рік									
		1976	1977	1978	1979	1980	1981	1982	1983	1984	1985
1	Луцьк	58	31	65	34	101	45	75	49	53	86
2	Рівне	54	40	77	40	118	100	66	28	68	148
3	Львів	66	55	97	87	133	85	135	64	98	161
4	Тернопіль	49	113	64	92	99	79	114	85	70	128
5	Ужгород	32	80	86	48	220	57	97	82	118	85
6	Івано-Франківськ	81	102	113	69	205	71	68	107	64	103
7	Житомир	41	83	102	19	134	122	124	60	102	118
8	Хмельницький	43	124	84	104	112	74	119	117	64	167
9	Чернівці	82	104	85	146	113	58	65	92	49	166
10	Вінниця	37	67	115	14	193	75	103	82	76	175
11	Київ	40	92	98	10	131	49	75	62	140	84
12	Чернігів	45	120	55	39	130	100	56	41	153	42
13	Семенівка	47	113	57	20	118	83	67	47	122	76
14	Суми	68	102	19	46	85	81	109	76	134	93
15	Гадяч	45	127	44	42	80	128	94	52	87	107
16	Полтава	44	95	85	6	109	33	71	73	84	82
17	Харків	61	88	67	7	88	15	51	39	25	119
18	Черкаси	62	80	56	22	110	62	53	51	131	99
19	Ізюм	49	82	76	17	72	20	56	51	23	84
20	Луганськ	65	98	54	5	40	13	54	50	24	85
21	Дніпропетровськ	87	153	93	7	96	5	36	34	30	87
22	Запоріжжя	68	135	92	21	55	26	38	92	9	120
23	Донецьк	112	68	65	8	38	41	33	44	15	52
24	Бердянськ	30	117	53	12	30	27	22	42	20	89
25	Мелітополь	35	84	47	2	27	6	71	90	15	77
26	В.Олександрівка	56	141	64	13	168	10	24	32	56	56
27	Кіровоград	46	83	34	8	216	76	81	34	113	134
28	Вознесенськ	81	110	33	41	96	68	99	33	99	81
29	Любашівка	28	26	47	92	87	49	78	27	145	124
30	Херсон	55	60	50	4	61	5	65	55	27	96
31	Миколаїв	91	82	44	10	74	11	44	29	65	121
32	Роздільна	61	108	22	89	112	41	74	70	117	162
33	Одеса	45	38	46	37	78	11	54	37	128	63

Продовження таблиці В.6

Станція		Рік									
		1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
1	Луцьк	86	71	150	131	51	75	28	146	26	51
2	Рівне	76	56	211	157	62	62	90	90	52	54
3	Львів	94	58	96	157	44	65	80	83	48	57
4	Тернопіль	64	66	124	123	80	57	68	95	46	48
5	Ужгород	50	81	68	143	71	71	29	63	27	75
6	Івано-Франківськ	100	53	136	159	36	107	120	58	67	61
7	Житомир	63	51	137	159	121	75	77	96	78	62
8	Хмельницький	105	63	167	154	138	68	-	85	72	56
9	Чернівці	76	55	137	128	63	59	123	70	132	57
10	Вінниця	64	94	162	119	120	109	45	83	66	129
11	Київ	97	67	104	162	77	97	29	87	88	55
12	Чернігів	95	94	136	59	145	65	50	54	59	66
13	Семенівка	107	133	88	68	84	141	100	86	63	95
14	Суми	13	66	107	74	81	69	29	94	38	76
15	Гадяч	104	73	129	26	61	55	44	100	41	51
16	Полтава	36	117	108	105	90	73	43	48	15	94
17	Харків	65	75	110	111	80	43	46	82	38	78
18	Черкаси	59	31	101	43	104	98	47	77	63	80
19	Ізюм	38	74	114	70	78	81	44	78	28	96
20	Луганськ	77	101	107	122	36	45	65	66	42	115
21	Дніпропетровськ	23	90	61	99	96	47	89	129	52	54
22	Запоріжжя	10	117	117	120	66	107	104	80	37	43
23	Донецьк	57	145	239	47	74	16	49	31	47	140
24	Бердянськ	67	103	71	50	48	13	97	54	63	21
25	Мелітополь	41	158	43	69	59	84	118	34	47	6
26	В.Олександрівка	38	114	53	80	58	85	95	44	40	141
27	Кіровоград	35	49	125	64	153	85	65	83	64	125
28	Вознесенськ	59	40	41	71	134	21	78	55	22	154
29	Любашівка	73	40	65	45	58	103	55	110	55	92
30	Херсон	47	34	71	30	24	81	61	32	45	31
31	Миколаїв	64	27	35	66	50	48	53	61	36	39
32	Роздільна	44	46	53	77	53	138	47	55	63	87
33	Одеса	60	16	32	65	53	61	58	54	65	42

Продовження таблиці В.6

Станція		Рік									
		1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
1	Луцьк	37	57	125	147	46	92	93	22	30	76
2	Рівне	72	63	94	59	54	122	77	59	31	95
3	Львів	79	55	168	71	71	157	114	54	33	84
4	Тернопіль	58	53	131	43	41	151	96	20	37	65
5	Ужгород	24	75	136	74	23	108	63	32	43	56
6	Івано-Франківськ	39	60	196	75	63	232	62	40	46	54
7	Житомир	61	136	125	54	18	165	121	38	33	55
8	Хмельницький	56	67	159	71	47	146	115	40	43	80
9	Чернівці	117	42	131	48	33	145	104	18	42	74
10	Вінниця	70	134	41	16	92	155	144	28	28	58
11	Київ	73	83	63	47	65	152	190	26	7	111
12	Чернігів	61	94	45	22	62	78	29	31	5	124
13	Семенівка	47	78	41	9	56	95	26	25	31	190
14	Суми	75	92	10	44	34	162	87	20	13	82
15	Гадяч	90	134	15	53	38	220	83	29	45	113
16	Полтава	56	115	9	40	41	204	53	24	45	101
17	Харків	74	95	13	25	13	116	50	89	29	104
18	Черкаси	65	95	50	64	57	136	86	46	64	102
19	Ізюм	63	86	34	19	52	123	70	104	99	55
20	Луганськ	70	85	27	12	49	87	48	83	126	83
21	Дніпропетровськ	50	90	29	45	100	113	33	52	106	104
22	Запоріжжя	21	177	22	18	75	114	34	72	68	100
23	Донецьк	29	97	22	41	71	190	24	109	120	56
24	Бердянськ	8	69	35	9	83	115	48	24	119	45
25	Мелітополь	35	97	33	27	63	90	35	41	58	72
26	В.Олександрівка	95	74	76	26	45	97	50	38	53	87
27	Кіровоград	19	140	30	46	32	65	59	65	35	136
28	Вознесенськ	16	97	53	24	29	112	103	52	49	87
29	Любашівка	35	79	56	61	75	129	64	49	11	87
30	Херсон	17	101	75	53	67	58	63	39	54	80
31	Миколаїв	4	89	37	57	26	55	52	31	52	75
32	Роздільна	75	77	62	60	29	67	28	39	64	49
33	Одеса	15	99	38	57	58	56	83	44	37	51

Таблиця В.7 - Місячна кількість опадів (липень)

Станція		Рік									
		1976	1977	1978	1979	1980	1981	1982	1983	1984	1985
1	Луцьк	36	102	52	50	107	70	49	66	75	47
2	Рівне	19	79	94	46	108	93	90	50	89	54
3	Львів	78	86	116	63	143	139	48	100	180	76
4	Тернопіль	74	71	68	86	130	114	90	143	73	83
5	Ужгород	98	123	103	72	157	137	88	69	95	43
6	Івано-Франківськ	64	118	145	71	157	141	87	56	93	29
7	Житомир	64	142	128	110	161	80	208	69	55	121
8	Хмельницький	30	85	86	76	146	176	139	71	48	70
9	Чернівці	106	91	94	80	58	156	51	62	101	32
10	Вінниця	32	83	131	15	171	62	170	61	89	68
11	Київ	91	55	190	61	97	105	113	32	74	101
12	Чернігів	65	76	128	85	156	62	77	43	125	122
13	Семенівка	66	71	169	175	84	43	167	118	159	114
14	Суми	76	100	135	167	39	24	84	26	58	125
15	Гадяч	95	115	77	122	67	70	105	38	88	121
16	Полтава	103	65	107	91	56	92	72	78	99	99
17	Харків	77	70	77	120	55	65	102	47	59	66
18	Черкаси	79	67	87	66	96	88	75	53	61	136
19	Ізюм	130	101	78	68	64	45	125	73	76	65
20	Луганськ	65	110	28	41	89	16	61	134	50	43
21	Дніпропетровськ	129	98	90	78	35	91	62	29	63	116
22	Запоріжжя	100	25	88	34	12	117	95	36	54	72
23	Донецьк	76	87	31	73	59	50	72	127	17	60
24	Бердянськ	36	51	27	6	74	8	45	141	15	55
25	Мелітополь	100	39	67	60	2	71	94	95	18	122
26	В.Олександрівка	107	109	43	49	23	65	78	102	30	85
27	Кіровоград	52	68	94	40	58	56	135	66	82	152
28	Вознесенськ	50	144	89	47	43	24	65	54	43	149
29	Любашівка	87	164	244	47	60	49	71	100	60	89
30	Херсон	51	110	76	27	4	31	125	65	25	118
31	Миколаїв	42	136	49	52	59	76	121	84	63	87
32	Роздільна	46	98	101	57	67	68	79	122	23	76
33	Одеса	74	78	46	35	49	30	39	74	7	61

Продовження таблиці В.7

Станція		Рік									
		1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
1	Луцьк	63	68	105	22	108	66	64	148	11	34
2	Рівне	110	32	80	34	114	151	61	190	25	27
3	Львів	51	45	86	31	104	82	74	180	22	10
4	Тернопіль	88	119	68	28	58	57	33	211	11	43
5	Ужгород	84	39	58	51	86	51	56	74	16	18
6	Івано-Франківськ	120	70	84	46	51	143	52	136	7	17
7	Житомир	102	53	130	48	74	78	26	168	17	33
8	Хмельницький	102	100	105	79	51	128	85	147	117	38
9	Чернівці	101	29	134	24	48	321	31	138	39	36
10	Вінниця	60	78	96	58	47	179	42	73	31	12
11	Київ	93	65	56	65	50	63	62	108	66	63
12	Чернігів	62	16	106	44	46	49	24	91	34	18
13	Семенівка	128	41	78	68	85	62	36	138	37	15
14	Суми	161	38	113	69	44	82	74	130	22	32
15	Гадяч	107	15	100	46	20	62	55	87	6	17
16	Полтава	48	41	151	38	32	14	27	28	24	31
17	Харків	35	42	114	80	15	79	77	54	5	34
18	Черкаси	94	23	121	20	67	53	9	53	5	31
19	Ізюм	25	44	144	53	12	27	88	78	5	26
20	Луганськ	25	43	94	93	5	35	234	105	1	59
21	Дніпропетровськ	85	36	87	13	27	11	52	21	6	1
22	Запоріжжя	57	25	79	13	19	17	45	31	13	6
23	Донецьк	45	53	46	47	15	18	113	83	2	23
24	Бердянськ	79	18	35	25	10	36	76	18	3	24
25	Мелітополь	86	18	16	26	44	35	30	27	1	26
26	В.Олександрівка	79	8	57	27	102	62	13	23	11	33
27	Кіровоград	87	57	90	7	33	95	13	49	14	10
28	Вознесенськ	52	37	110	31	43	85	34	77	20	12
29	Любашівка	58	34	122	46	41	142	31	68	24	73
30	Херсон	34	19	87	1	97	38	34	22	2	2
31	Миколаїв	54	26	49	14	41	25	5	40	3	36
32	Роздільна	70	93	89	48	35	83	15	36	31	7
33	Одеса	66	33	142	10	44	16	54	42	19	4

Продовження таблиці В.7

Станція		Рік									
		1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
1	Луцьк	53	146	156	67	148	45	80	142	50	78
2	Рівне	67	178	190	97	190	77	71	161	54	68
3	Львів	73	145	207	52	131	137	46	118	159	87
4	Тернопіль	39	164	149	25	168	129	62	165	144	55
5	Ужгород	47	138	161	97	119	111	75	80	73	80
6	Івано-Франківськ	49	125	110	73	101	135	75	85	131	79
7	Житомир	85	154	164	46	91	56	40	102	88	15
8	Хмельницький	60	236	151	31	229	74	52	154	106	103
9	Чернівці	42	212	93	78	181	101	111	209	110	41
10	Вінниця	68	205	129	50	134	42	76	164	90	35
11	Київ	69	92	88	81	82	6	14	61	113	29
12	Чернігів	98	146	98	29	131	42	26	32	80	28
13	Семенівка	54	44	108	115	197	90	20	96	71	37
14	Суми	51	80	116	64	80	39	30	85	92	68
15	Гадяч	59	59	140	63	77	90	8	91	104	85
16	Полтава	27	95	120	32	99	69	41	222	128	69
17	Харків	23	87	80	31	59	4	91	158	75	91
18	Черкаси	46	75	105	46	84	113	40	109	70	27
19	Ізюм	6	83	40	47	41	39	50	122	70	62
20	Луганськ	7	120	33	31	79	83	64	116	84	95
21	Дніпропетровськ	18	118	66	85	64	22	69	101	76	55
22	Запоріжжя	13	156	77	119	43	10	99	122	75	36
23	Донецьк	5	196	36	31	24	24	81	70	27	109
24	Бердянськ	1	160	12	7	11	0	26	70	33	66
25	Мелітополь	2	91	56	40	15	12	60	129	39	48
26	В.Олександрівка	20	87	46	24	92	57	25	124	128	59
27	Кіровоград	13	71	77	38	85	107	89	76	128	73
28	Вознесенськ	35	150	76	50	64	14	32	122	126	42
29	Любашівка	17	114	102	31	83	36	81	62	167	17
30	Херсон	12	124	20	9	93	19	39	68	102	35
31	Миколаїв	8	83	47	49	55	27	32	58	67	23
32	Роздільна	8	115	92	28	97	114	59	123	91	56
33	Одеса	6	124	88	5	39	2	19	44	70	36

Таблиця В.8 - Місячна кількість опадів (серпень)

Станція		Рік									
		1976	1977	1978	1979	1980	1981	1982	1983	1984	1985
1	Луцьк	52	121	80	86	71	58	30	29	11	38
2	Рівне	36	96	54	45	54	61	46	25	28	36
3	Львів	38	131	73	137	41	62	52	65	31	46
4	Тернопіль	71	60	47	35	92	52	35	80	60	65
5	Ужгород	26	197	102	119	158	84	46	35	50	123
6	Івано-Франківськ	62	89	76	83	77	27	75	119	21	85
7	Житомир	17	106	26	72	57	90	36	46	18	28
8	Хмельницький	68	126	40	64	52	45	57	115	32	32
9	Чернівці	54	37	20	103	110	53	37	115	18	61
10	Вінниця	110	94	24	82	40	16	42	54	32	28
11	Київ	63	119	39	77	32	49	63	44	8	31
12	Чернігів	93	173	61	81	78	64	75	23	19	57
13	Семенівка	85	118	53	81	90	76	36	22	33	118
14	Суми	106	80	35	66	138	84	106	35	42	71
15	Гадяч	52	129	55	55	136	134	87	37	31	86
16	Полтава	64	84	35	65	40	94	51	45	40	56
17	Харків	38	112	31	72	49	65	21	26	46	34
18	Черкаси	168	93	20	52	57	70	52	97	17	17
19	Ізюм	113	113	22	58	76	36	50	36	64	33
20	Луганськ	47	94	26	28	60	60	83	25	60	34
21	Дніпропетровськ	71	98	13	52	20	50	14	54	64	25
22	Запоріжжя	81	126	34	34	43	76	30	30	87	35
23	Донецьк	122	63	10	43	93	32	38	56	54	21
24	Бердянськ	127	109	15	38	27	47	40	40	25	1
25	Мелітополь	81	69	26	50	44	35	25	165	12	61
26	В.Олександрівка	138	53	12	48	8	63	56	92	42	16
27	Кіровоград	90	91	25	48	41	39	93	98	42	22
28	Вознесенськ	123	90	19	24	29	69	41	75	100	6
29	Любашівка	137	110	50	91	44	35	52	63	36	68
30	Херсон	105	96	16	31	28	34	27	92	74	10
31	Миколаїв	179	82	30	17	17	26	27	35	26	6
32	Роздільна	163	59	53	12	24	38	64	25	29	20
33	Одеса	90	51	25	31	18	37	25	16	29	7

Продовження таблиці В.8

Станція		Рік									
		1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
1	Луцьк	87	65	40	87	27	63	45	43	70	41
2	Рівне	88	61	54	100	7	58	23	31	67	41
3	Львів	81	65	90	109	64	104	45	69	94	40
4	Тернопіль	138	82	73	82	37	91	35	65	73	70
5	Ужгород	50	64	62	137	30	72	23	31	76	54
6	Івано-Франківськ	87	91	69	163	52	103	57	47	76	64
7	Житомир	167	89	73	169	42	98	37	34	66	49
8	Хмельницький	58	97	46	91	26	148	70	30	64	87
9	Чернівці	65	61	53	129	39	90	31	34	107	47
10	Вінниця	113	112	69	181	41	124	12	48	89	70
11	Київ	113	86	77	55	65	55	36	54	76	75
12	Чернігів	104	96	114	56	48	40	34	34	52	109
13	Семенівка	161	117	107	30	52	54	16	35	78	93
14	Суми	83	75	76	41	64	78	4	56	39	48
15	Гадяч	34	87	48	17	43	34	19	72	27	52
16	Полтава	8	34	72	5	51	70	10	16	56	40
17	Харків	22	37	47	24	46	97	13	131	67	21
18	Черкаси	45	108	40	17	28	85	4	42	44	53
19	Ізюм	3	49	49	86	46	80	19	32	25	33
20	Луганськ	2	36	20	1	25	45	36	62	30	48
21	Дніпропетровськ	2	13	67	13	24	103	12	21	32	75
22	Запоріжжя	4	8	31	25	11	49	18	39	16	21
23	Донецьк	6	45	10	2	55	45	21	73	17	75
24	Бердянськ	1	56	8	20	2	39	1	20	13	15
25	Мелітополь	0	32	17	25	12	24	4	30	19	9
26	В.Олександрівка	4	28	15	44	18	37	61	41	60	51
27	Кіровоград	9	47	39	19	41	132	9	18	60	76
28	Вознесенськ	29	151	24	19	32	83	1	36	94	52
29	Любашівка	26	104	54	35	44	63	15	68	111	110
30	Херсон	3	33	4	25	15	24	0	10	80	56
31	Миколаїв	13	65	28	11	40	45	18	30	62	25
32	Роздільна	26	67	33	30	37	71	44	26	127	38
33	Одеса	48	47	14	4	26	51	20	4	113	25

Продовження таблиці В.8

Станція		Рік									
		1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
1	Луцьк	71	55	40	62	25	34	14	25	144	72
2	Рівне	83	70	43	91	9	53	21	36	118	93
3	Львів	63	79	65	82	21	79	26	26	170	80
4	Тернопіль	56	65	52	187	11	60	32	43	73	91
5	Ужгород	82	57	84	130	5	84	75	27	70	153
6	Івано-Франківськ	57	47	46	97	23	56	53	24	138	141
7	Житомир	89	70	77	135	17	33	89	36	125	90
8	Хмельницький	25	731	56	67	24	40	58	52	93	219
9	Чернівці	95	64	46	77	52	60	95	13	126	307
10	Вінниця	23	53	56	56	16	21	117	53	244	76
11	Київ	46	35	58	132	20	16	101	87	130	86
12	Чернігів	60	70	85	84	12	53	52	96	49	55
13	Семенівка	34	26	72	91	18	36	59	70	21	133
14	Суми	53	19	90	47	53	38	22	107	21	11
15	Гадяч	101	16	85	20	14	26	82	60	48	9
16	Полтава	74	64	56	27	12	10	53	43	150	24
17	Харків	43	3	33	45	2	31	20	138	68	42
18	Черкаси	143	72	72	34	129	13	55	55	64	66
19	Ізюм	30	39	35	26	31	21	34	64	64	33
20	Луганськ	122	70	54	15	42	33	52	21	42	22
21	Дніпропетровськ	70	111	12	93	4	14	71	32	167	32
22	Запоріжжя	30	92	28	28	14	19	93	31	95	27
23	Донецьк	52	68	34	29	21	22	31	53	110	16
24	Бердянськ	88	166	1	77	129	16	25	6	202	22
25	Мелітополь	54	132	6	82	58	28	23	40	149	14
26	В.Олександрівка	118	98	26	40	9	10	95	9	228	29
27	Кіровоград	85	90	79	32	31	11	35	20	94	54
28	Вознесенськ	67	72	63	68	35	15	37	11	64	79
29	Любашівка	37	59	62	38	52	35	18	30	98	58
30	Херсон	25	116	14	13	39	16	18	47	120	59
31	Миколаїв	18	141	31	6	29	14	34	2	49	134
32	Роздільна	43	145	54	48	27	5	87	12	53	74
33	Одеса	20	111	38	23	41	13	151	14	36	88

Таблиця В.9 - Місячна кількість опадів (вересень)

Станція		Рік									
		1976	1977	1978	1979	1980	1981	1982	1983	1984	1985
1	Луцьк	114	42	155	11	67	28	18	39	93	55
2	Рівне	150	48	151	8	32	27	10	36	42	37
3	Львів	98	70	105	36	36	46	18	59	67	46
4	Тернопіль	167	71	180	9	58	37	14	54	61	41
5	Ужгород	142	54	41	11	44	45	109	61	102	37
6	Івано-Франківськ	167	74	146	18	47	51	19	28	72	38
7	Житомир	109	29	76	12	49	27	46	49	110	79
8	Хмельницький	105	54	174	15	59	33	30	32	44	45
9	Чернівці	172	71	148	7	37	48	6	15	36	24
10	Вінниця	106	46	82	1	33	28	7	53	85	62
11	Київ	98	40	68	26	59	19	34	23	41	91
12	Чернігів	41	13	79	18	28	34	21	40	43	57
13	Семенівка	36	38	123	63	44	27	20	9	79	141
14	Суми	49	34	88	51	48	24	17	19	55	91
15	Гадяч	115	30	103	41	19	29	16	9	77	66
16	Полтава	84	29	123	21	36	14	9	5	28	70
17	Харків	49	35	88	56	40	10	8	8	18	48
18	Черкаси	69	17	32	8	36	63	9	31	73	64
19	Ізюм	115	32	83	38	74	15	17	19	9	38
20	Луганськ	59	46	41	11	56	8	15	2	4	53
21	Дніпропетровськ	75	34	75	11	84	29	6	3	19	36
22	Запоріжжя	50	22	44	15	65	23	11	2	5	45
23	Донецьк	81	23	34	7	46	27	4	23	3	42
24	Бердянськ	85	29	37	34	34	27	12	3	1	50
25	Мелітополь	66	16	19	63	34	19	15	10	19	33
26	В.Олександрівка	48	17	38	14	65	20	19	1	19	52
27	Кіровоград	20	12	74	5	36	21	21	1	44	54
28	Вознесенськ	39	23	41	14	45	25	8	0	11	52
29	Любашівка	74	16	27	13	36	70	22	2	88	39
30	Херсон	95	13	38	42	58	22	17	2	10	54
31	Миколаїв	21	10	56	9	56	21	12	6	29	50
32	Роздільна	75	10	27	31	46	49	26	3	41	46
33	Одеса	64	13	27	12	34	36	1	1	11	26

Продовження таблиці В.9

Станція		Рік									
		1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
1	Луцьк	28	40	45	57	67	48	86	41	57	43
2	Рівне	55	45	50	55	68	34	105	47	62	62
3	Львів	11	100	114	38	101	61	147	58	57	87
4	Тернопіль	13	61	74	20	45	36	122	59	56	77
5	Ужгород	3	49	94	56	97	30	82	60	120	86
6	Івано-Франківськ	5	51	64	49	48	81	80	132	76	92
7	Житомир	21	72	25	56	73	18	33	84	21	93
8	Хмельницький	14	45	54	35	35	15	149	85	48	165
9	Чернівці	6	28	42	86	36	36	60	79	22	87
10	Вінниця	12	34	10	153	46	17	49	108	29	137
11	Київ	22	66	57	100	54	7	60	129	18	101
12	Чернігів	64	74	50	35	97	16	43	119	26	120
13	Семенівка	45	87	110	30	109	26	49	95	22	141
14	Суми	30	94	49	67	81	15	86	108	5	61
15	Гадяч	34	70	51	42	57	11	82	92	3	84
16	Полтава	21	40	167	67	38	10	45	76	5	73
17	Харків	41	64	82	69	26	6	24	106	10	101
18	Черкаси	20	15	11	113	33	10	40	66	14	129
19	Ізюм	46	48	75	65	48	47	77	64	5	84
20	Луганськ	11	53	51	74	39	30	122	49	14	58
21	Дніпропетровськ	16	26	68	117	61	29	17	66	2	24
22	Запоріжжя	11	18	48	107	43	11	36	26	3	37
23	Донецьк	135	51	77	49	41	45	34	63	9	50
24	Бердянськ	22	20	45	227	58	52	22	14	15	14
25	Мелітополь	12	20	78	121	51	29	8	36	5	10
26	В.Олександрівка	11	25	118	85	61	26	14	42	5	46
27	Кіровоград	3	25	63	103	33	3	13	86	1	85
28	Вознесенськ	5	52	99	95	56	9	11	70	0	80
29	Любашівка	4	32	64	156	49	13	38	60	4	85
30	Херсон	7	58	118	85	61	10	32	35	5	72
31	Миколаїв	11	26	126	141	52	10	16	47	2	53
32	Роздільна	1	12	75	121	64	38	27	78	0	22
33	Одеса	2	5	56	140	65	18	11	30	0	80

Продовження таблиці В.9

Станція		Рік									
		1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
1	Луцьк	58	40	102	72	42	150	49	36	26	13
2	Рівне	70	48	79	59	48	172	51	34	41	17
3	Львів	147	47	56	47	42	113	84	39	42	30
4	Тернопіль	88	38	86	60	28	142	60	39	33	11
5	Ужгород	129	34	91	45	89	86	77	96	93	88
6	Івано-Франківськ	105	59	31	52	43	97	125	55	53	18
7	Житомир	88	51	53	56	106	136	75	42	62	11
8	Хмельницький	132	54	62	56	61	72	99	36	42	9
9	Чернівці	146	34	84	47	32	130	59	41	51	11
10	Вінниця	181	52	45	40	128	77	101	34	60	2
11	Київ	128	71	26	8	141	53	76	52	79	7
12	Чернігів	108	31	42	47	112	54	116	58	37	3
13	Семенівка	122	72	67	62	104	50	132	65	50	19
14	Суми	76	81	7	20	95	64	98	17	65	7
15	Гадяч	81	87	15	8	127	64	91	21	113	10
16	Полтава	118	69	2	3	71	108	73	31	58	1
17	Харків	122	55	3	12	66	44	104	20	47	2
18	Черкаси	74	58	40	12	118	53	122	26	63	5
19	Ізюм	108	69	6	21	69	31	75	3	90	11
20	Луганськ	127	65	1	21	67	142	63	6	48	9
21	Дніпропетровськ	100	32	8	4	112	21	137	7	37	1
22	Запоріжжя	172	25	16	18	118	34	74	11	22	5
23	Донецьк	126	35	5	3	94	41	70	2	28	4
24	Бердянськ	169	38	4	29	36	21	87	3	29	7
25	Мелітополь	112	24	6	8	31	28	80	5	60	6
26	В.Олександрівка	67	17	16	9	129	80	109	11	33	4
27	Кіровоград	76	32	21	35	103	46	70	10	84	1
28	Вознесенськ	85	18	27	40	150	81	67	49	68	4
29	Любашівка	131	38	34	36	139	81	73	8	43	0
30	Херсон	10	8	22	68	120	75	39	5	21	10
31	Миколаїв	106	10	25	14	107	61	38	24	14	8
32	Роздільна	187	14	38	23	148	85	67	62	38	1
33	Одеса	108	2	31	10	108	119	42	57	22	1

Таблиця В.10 - Місячна кількість опадів (жовтень)

Станція		Рік									
		1976	1977	1978	1979	1980	1981	1982	1983	1984	1985
1	Луцьк	11	5	38	13	80	76	20	34	44	21
2	Рівне	9	2	31	13	82	88	21	35	40	34
3	Львів	23	11	51	60	78	97	30	47	43	32
4	Тернопіль	24	5	15	45	71	76	22	44	26	18
5	Ужгород	42	3	28	19	99	97	77	38	45	21
6	Івано-Франківськ	56	5	10	60	56	88	15	34	22	18
7	Житомир	12	4	23	30	59	68	9	48	45	20
8	Хмельницький	12	9	15	23	65	82	7	25	24	20
9	Чернівці	61	9	9	40	69	51	5	9	18	8
10	Вінниця	15	5	10	36	62	62	22	25	30	6
11	Київ	11	5	32	50	36	101	16	42	22	11
12	Чернігів	20	11	32	14	62	94	14	42	15	14
13	Семенівка	36	24	37	23	48	122	34	39	38	20
14	Суми	12	26	53	73	51	82	36	44	33	20
15	Гадяч	14	18	50	68	27	76	38	55	41	17
16	Полтава	13	30	50	72	13	87	69	21	13	7
17	Харків	8	32	33	46	27	83	96	21	27	16
18	Черкаси	6	10	30	83	19	89	11	22	18	4
19	Ізюм	8	24	27	43	20	68	40	22	55	13
20	Луганськ	4	17	12	22	14	41	19	16	63	4
21	Дніпропетровськ	8	16	37	39	21	44	101	22	27	24
22	Запоріжжя	12	44	22	38	20	66	79	25	40	31
23	Донецьк	11	13	11	43	7	69	24	22	66	4
24	Бердянськ	7	8	11	52	27	47	8	22	30	15
25	Мелітополь	28	1	6	47	31	55	24	21	39	64
26	В.Олександрівка	1	13	20	36	53	24	44	21	23	42
27	Кіровоград	16	18	24	48	11	109	7	16	22	6
28	Вознесенськ	3	22	29	33	20	72	31	7	26	24
29	Любашівка	6	13	14	44	51	93	32	6	21	44
30	Херсон	5	15	27	50	52	32	45	22	35	32
31	Миколаїв	4	12	23	42	30	35	28	9	29	22
32	Роздільна	23	11	35	24	51	77	29	7	12	47
33	Одеса	14	4	29	38	42	54	70	3	20	41

Продовження таблиці В.10

Станція		Рік									
		1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
1	Луцьк	13	28	14	54	26	34	75	26	42	4
2	Рівне	9	31	14	51	26	48	64	27	45	9
3	Львів	17	41	22	35	46	78	94	36	111	22
4	Тернопіль	7	40	22	31	17	65	54	36	31	14
5	Ужгород	26	24	21	33	72	101	136	44	84	14
6	Івано-Франківськ	25	33	12	52	13	107	60	24	42	10
7	Житомир	11	16	14	49	22	50	64	10	12	9
8	Хмельницький	9	41	21	30	21	64	53	13	26	15
9	Чернівці	24	58	15	45	13	53	30	15	33	8
10	Вінниця	19	33	10	37	13	37	46	6	18	6
11	Київ	24	15	15	57	25	61	60	19	29	11
12	Чернігів	26	4	18	61	38	63	70	26	35	14
13	Семенівка	24	0	20	68	43	83	71	37	50	10
14	Суми	58	14	30	52	40	78	126	15	108	4
15	Гадяч	61	0	42	54	33	75	127	22	73	4
16	Полтава	110	0	56	72	61	38	94	13	70	7
17	Харків	74	0	11	60	56	54	53	10	29	27
18	Черкаси	29	0	33	75	33	73	97	64	49	5
19	Ізюм	60	5	12	30	61	44	72	37	67	71
20	Луганськ	10	2	12	23	25	31	27	18	21	46
21	Дніпропетровськ	74	2	7	16	58	71	36	7	33	18
22	Запоріжжя	45	7	15	26	47	66	41	4	37	17
23	Донецьк	8	2	24	25	39	48	53	12	33	52
24	Бердянськ	11	9	18	7	21	23	46	49	46	37
25	Мелітополь	21	8	17	5	13	40	66	7	24	35
26	В.Олександрівка	85	7	18	17	34	58	28	5	13	7
27	Кіровоград	55	0	29	23	40	38	56	5	60	6
28	Вознесенськ	44	6	36	18	39	23	47	2	36	9
29	Любашівка	10	26	33	47	26	65	43	2	48	11
30	Херсон	66	8	57	21	25	31	55	10	10	17
31	Миколаїв	58	9	44	28	31	36	45	6	40	19
32	Роздільна	42	42	31	51	24	24	56	3	23	6
33	Одеса	60	29	64	27	32	27	44	3	47	4

Продовження таблиці В.10

Станція		Рік									
		1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
1	Луцьк	18	50	59	41	3	16	76	80	30	29
2	Рівне	22	50	68	30	3	15	64	89	36	57
3	Львів	46	65	84	38	6	26	79	88	39	26
4	Тернопіль	28	40	55	22	2	18	34	84	32	30
5	Ужгород	76	52	94	28	0	32	103	113	56	11
6	Івано-Франківськ	27	45	46	38	3	26	59	86	33	40
7	Житомир	31	37	74	64	1	22	56	106	38	37
8	Хмельницький	30	48	74	33	1	18	64	64	25	40
9	Чернівці	57	40	77	49	2	20	49	50	16	68
10	Вінниця	32	31	100	50	3	21	37	82	16	41
11	Київ	42	56	76	52	0	23	74	110	31	78
12	Чернігів	37	49	56	64	5	34	100	138	22	57
13	Семенівка	39	85	71	65	6	57	91	96	25	58
14	Суми	52	86	89	26	9	33	59	85	24	58
15	Гадяч	55	88	101	18	17	28	68	91	25	100
16	Полтава	62	76	90	19	7	27	63	132	27	66
17	Харків	44	60	70	36	9	39	55	81	27	51
18	Черкаси	32	80	89	30	0	8	62	104	23	55
19	Ізюм	35	67	57	44	14	27	52	82	32	47
20	Луганськ	53	90	33	71	24	22	60	37	65	58
21	Дніпропетровськ	34	66	45	27	7	18	78	48	30	38
22	Запоріжжя	31	51	53	27	1	12	47	40	41	43
23	Донецьк	38	102	30	43	26	19	37	40	46	27
24	Бердянськ	18	106	16	42	9	19	80	10	68	30
25	Мелітополь	31	68	39	30	5	20	54	13	78	36
26	В.Олександрівка	10	44	31	26	1	18	53	52	57	52
27	Кіровоград	18	58	57	24	0	16	66	79	21	40
28	Вознесенськ	15	39	52	52	1	22	76	91	40	31
29	Любашівка	33	66	67	26	1	17	79	100	26	28
30	Херсон	8	45	25	13	7	35	73	4	43	11
31	Миколаїв	2	46	61	23	3	29	63	81	19	10
32	Роздільна	30	35	76	44	10	31	87	67	25	18
33	Одеса	7	48	64	15	5	19	26	65	53	7

Таблиця В.11 - Місячна кількість опадів (листопад)

Станція		Рік									
		1976	1977	1978	1979	1980	1981	1982	1983	1984	1985
1	Луцьк	31	44	18	13	41	73	14	15	15	41
2	Рівне	27	48	12	32	43	59	21	18	18	48
3	Львів	22	77	22	35	77	67	10	26	24	69
4	Тернопіль	20	51	23	28	61	65	14	13	15	57
5	Ужгород	62	83	23	96	120	31	38	58	56	97
6	Івано-Франківськ	29	47	41	20	74	44	7	13	25	52
7	Житомир	42	86	13	52	49	44	32	30	21	88
8	Хмельницький	49	81	12	45	56	77	25	17	17	60
9	Чернівці	23	61	40	23	69	55	15	15	38	50
10	Вінниця	36	65	5	44	60	64	24	26	17	70
11	Київ	24	76	12	60	66	59	33	34	30	86
12	Чернігів	17	99	13	57	53	62	4	44	30	76
13	Семенівка	22	90	16	68	47	80	53	55	34	54
14	Суми	14	83	12	51	140	81	20	29	36	67
15	Гадяч	14	83	13	59	138	71	18	31	29	71
16	Полтава	12	66	9	42	121	78	11	24	37	83
17	Харків	15	93	7	33	121	72	13	18	20	49
18	Черкаси	19	54	8	71	87	79	13	25	36	63
19	Ізюм	44	61	7	38	81	82	14	14	34	85
20	Луганськ	31	7	5	26	64	90	17	24	23	50
21	Дніпропетровськ	22	96	5	28	90	101	11	18	52	56
22	Запоріжжя	27	98	5	25	85	93	9	15	76	62
23	Донецьк	21	89	8	32	63	76	13	18	30	43
24	Бердянськ	22	63	4	37	34	95	0	28	69	24
25	Мелітополь	30	77	4	24	42	74	6	16	98	71
26	В.Олександрівка	17	73	10	37	72	84	9	12	76	43
27	Кіровоград	21	43	6	56	114	77	5	11	42	41
28	Вознесенськ	24	36	16	63	103	109	6	4	36	49
29	Любашівка	44	26	29	83	91	91	12	20	32	40
30	Херсон	21	55	13	44	61	96	8	15	64	47
31	Миколаїв	14	30	14	55	70	110	8	6	73	53
32	Роздільна	55	22	19	77	80	136	7	6	24	46
33	Одеса	24	31	12	111	80	115	6	7	50	46

Продовження таблиці В.11

Станція		Рік									
		1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
1	Луцьк	12	34	48	58	42	26	65	23	29	35
2	Рівне	22	31	36	49	35	22	65	20	37	33
3	Львів	20	43	43	44	42	33	112	23	47	65
4	Тернопіль	20	39	38	21	27	21	60	27	25	26
5	Ужгород	18	51	32	43	64	64	90	45	40	51
6	Івано-Франківськ	22	45	31	43	56	22	47	20	26	48
7	Житомир	25	48	39	44	52	21	88	35	40	34
8	Хмельницький	26	48	29	23	46	24	67	49	22	42
9	Чернівці	11	46	32	16	27	13	33	43	26	64
10	Вінниця	18	62	27	24	39	9	73	37	19	23
11	Київ	31	56	37	40	47	17	98	22	43	39
12	Чернігів	30	55	46	36	45	25	87	20	34	41
13	Семенівка	25	54	44	35	80	42	85	14	47	39
14	Суми	18	43	28	41	97	21	55	10	26	87
15	Гадяч	12	46	40	50	75	18	73	14	38	95
16	Полтава	17	43	37	73	58	17	63	11	43	79
17	Харків	14	42	29	60	76	12	60	6	18	136
18	Черкаси	19	58	31	37	30	20	61	19	22	51
19	Ізюм	10	41	61	86	51	11	70	8	28	125
20	Луганськ	23	29	71	41	30	30	47	3	32	55
21	Дніпропетровськ	23	37	48	50	35	18	76	15	20	128
22	Запоріжжя	7	28	63	27	45	18	86	15	21	128
23	Донецьк	14	34	62	45	27	22	72	4	24	66
24	Бердянськ	10	50	52	26	31	25	69	4	45	88
25	Мелітополь	4	36	68	33	7	25	59	10	28	90
26	В.Олександрівка	2	29	30	11	33	23	60	19	43	90
27	Кіровоград	15	33	23	33	24	25	34	19	11	42
28	Вознесенськ	10	47	17	16	9	12	44	27	17	47
29	Любашівка	9	64	15	21	23	14	39	41	13	40
30	Херсон	3	42	47	14	15	8	54	11	33	76
31	Миколаїв	4	43	22	6	13	4	37	30	26	44
32	Роздільна	10	46	29	9	4	8	44	33	36	33
33	Одеса	8	58	29	7	6	2	44	27	28	56

Продовження таблиці В.11

Станція		Рік									
		1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
1	Луцьк	23	36	50	38	32	38	22	22	66	25
2	Рівне	29	54	48	42	32	41	25	25	63	25
3	Львів	39	65	60	75	17	70	24	37	96	27
4	Тернопіль	27	30	54	45	22	37	24	24	65	27
5	Ужгород	19	94	89	85	39	81	48	31	65	29
6	Івано-Франківськ	24	45	66	20	3	65	26	19	41	22
7	Житомир	80	34	43	89	51	50	32	30	44	32
8	Хмельницький	43	48	61	58	39	41	23	28	55	33
9	Чернівці	28	23	57	44	16	73	40	14	48	33
10	Вінниця	76	34	24	70	27	55	30	35	35	31
11	Київ	84	56	73	47	44	70	50	30	44	40
12	Чернігів	73	43	52	34	40	80	57	22	51	47
13	Семенівка	50	37	53	56	90	59	54	18	38	64
14	Суми	40	23	62	52	23	48	31	33	30	78
15	Гадяч	58	36	59	72	27	53	43	33	35	81
16	Полтава	34	48	60	68	7	74	34	26	59	67
17	Харків	28	38	43	33	4	58	36	42	67	87
18	Черкаси	33	56	59	56	36	66	35	26	66	31
19	Ізюм	9	64	44	55	2	80	43	34	37	85
20	Луганськ	8	46	55	56	17	62	38	46	44	59
21	Дніпропетровськ	22	57	42	96	4	102	34	31	74	60
22	Запоріжжя	19	53	37	74	1	74	50	45	58	66
23	Донецьк	7	30	37	37	12	78	37	44	43	87
24	Бердянськ	11	52	43	80	3	58	55	50	38	92
25	Мелітополь	18	55	45	62	10	62	56	48	46	60
26	В.Олександрівка	37	43	55	81	16	36	59	21	54	56
27	Кіровоград	49	50	53	70	36	61	31	18	91	23
28	Вознесенськ	43	48	81	53	50	90	45	11	65	29
29	Любашівка	51	61	61	77	113	93	57	18	58	26
30	Херсон	30	50	62	65	13	48	41	24	48	40
31	Миколаїв	33	40	56	42	18	67	48	8	40	42
32	Роздільна	39	60	72	34	61	79	62	12	36	37
33	Одеса	36	60	83	53	61	73	59	25	50	43

Таблиця В.12 - Місячна кількість опадів (грудень)

Станція		Рік									
		1976	1977	1978	1979	1980	1981	1982	1983	1984	1985
1	Луцьк	42	18	52	48	43	77	34	39	22	36
2	Рівне	42	17	48	36	42	57	30	29	28	60
3	Львів	59	49	73	51	70	60	59	22	28	76
4	Тернопіль	30	17	61	20	62	57	35	11	35	51
5	Ужгород	108	42	110	73	50	135	88	18	46	84
6	Івано-Франківськ	29	26	88	23	50	39	28	7	40	56
7	Житомир	39	16	40	30	62	77	17	23	31	42
8	Хмельницький	40	16	51	27	75	76	23	17	36	39
9	Чернівці	37	18	63	17	54	37	23	12	33	37
10	Вінниця	39	13	48	43	65	58	14	44	42	37
11	Київ	26	27	44	45	69	110	28	21	47	40
12	Чернігів	41	23	46	40	46	93	39	17	38	35
13	Семенівка	41	26	35	55	54	88	57	24	38	45
14	Суми	71	39	102	55	83	123	52	17	23	27
15	Гадяч	74	38	86	50	92	118	44	27	44	38
16	Полтава	61	18	120	58	66	127	31	23	38	33
17	Харків	76	16	96	32	60	100	32	15	18	27
18	Черкаси	29	24	64	38	76	90	31	31	30	24
19	Ізюм	92	29	128	38	49	68	35	30	25	28
20	Луганськ	79	33	66	23	63	51	22	32	26	34
21	Дніпропетровськ	61	17	120	31	31	120	25	29	19	35
22	Запоріжжя	66	14	91	33	36	110	29	27	17	41
23	Донецьк	113	30	90	22	66	55	21	35	19	39
24	Бердянськ	89	57	77	20	66	74	16	33	29	34
25	Мелітополь	66	43	59	14	43	90	16	48	24	29
26	В.Олександрівка	51	28	49	22	20	100	14	41	21	21
27	Кіровоград	46	30	73	31	57	65	32	17	12	28
28	Вознесенськ	61	35	53	21	66	81	20	14	18	20
29	Любашівка	51	32	46	35	76	79	46	7	32	18
30	Херсон	43	48	36	22	25	71	18	18	13	18
31	Миколаїв	39	77	31	17	42	73	15	29	11	19
32	Роздільна	70	40	36	28	61	79	22	19	20	17
33	Одеса	61	72	49	13	63	72	19	20	15	15

Продовження таблиці В.12

Станція		Рік									
		1986	1987	1988	1989	1990	1991	1992	1993	1994	1995
1	Луцьк	18	43	30	23	35	26	12	29	61	17
2	Рівне	17	42	33	22	37	27	17	32	59	23
3	Львів	51	77	65	49	44	43	28	40	64	47
4	Тернопіль	32	55	39	20	40	25	15	31	53	34
5	Ужгород	79	83	92	23	79	58	33	103	61	108
6	Івано-Франківськ	28	77	44	23	62	22	14	34	45	69
7	Житомир	29	27	40	25	51	20	15	53	60	26
8	Хмельницький	21	49	35	25	68	29	25	50	45	32
9	Чернівці	9	58	55	17	50	23	25	29	37	62
10	Вінниця	27	28	47	23	66	14	20	47	44	37
11	Київ	70	39	56	32	40	20	12	74	44	31
12	Чернігів	56	35	70	39	38	19	17	66	40	18
13	Семенівка	40	53	66	57	29	30	10	73	45	19
14	Суми	22	28	52	44	16	37	11	58	31	25
15	Гадяч	49	45	78	41	16	34	10	70	48	38
16	Полтава	44	31	66	38	24	22	12	66	39	43
17	Харків	29	27	50	26	19	18	16	61	21	38
18	Черкаси	38	28	53	31	49	26	12	48	35	34
19	Ізюм	60	27	63	44	41	36	25	90	30	68
20	Луганськ	52	18	48	38	42	28	19	76	19	63
21	Дніпропетровськ	39	32	60	32	61	32	22	81	38	51
22	Запоріжжя	68	26	50	23	33	51	19	72	40	59
23	Донецьк	40	21	52	43	41	27	25	105	22	58
24	Бердянськ	76	40	67	18	68	30	20	66	34	59
25	Мелітополь	66	33	49	18	32	45	19	72	27	43
26	В.Олександрівка	52	26	37	9	40	39	10	36	24	45
27	Кіровоград	33	27	49	30	73	18	8	21	32	36
28	Вознесенськ	33	30	35	2	72	19	11	50	31	53
29	Любашівка	33	25	50	5	79	17	15	43	42	17
30	Херсон	43	34	49	14	59	36	13	38	17	37
31	Миколаїв	44	29	46	5	85	23	8	35	20	38
32	Роздільна	29	22	50	4	73	35	29	20	31	45
33	Одеса	46	18	65	4	73	26	13	23	28	35

Продовження таблиці В.12

Станція		Рік									
		1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
1	Луцьк	38	71	24	44	41	18	19	20	15	53
2	Рівне	32	65	26	38	44	25	22	18	18	61
3	Львів	49	66	39	61	39	42	34	27	9	55
4	Тернопіль	45	65	16	51	35	21	20	20	20	57
5	Ужгород	67	73	68	98	76	33	56	48	39	110
6	Івано-Франківськ	71	52	17	46	29	25	29	24	20	25
7	Житомир	34	75	30	54	30	29	19	29	21	71
8	Хмельницький	37	76	18	57	30	28	17	25	33	59
9	Чернівці	71	50	19	53	23	29	11	27	10	29
10	Вінниця	31	53	14	54	29	23	11	25	22	50
11	Київ	28	84	33	58	24	39	11	32	16	78
12	Чернігів	34	82	66	50	44	45	16	25	35	74
13	Семенівка	29	42	20	49	57	42	25	28	42	56
14	Суми	23	55	19	36	33	25	13	39	23	60
15	Гадяч	30	65	28	48	31	40	9	34	19	64
16	Полтава	23	63	31	70	27	33	8	39	26	70
17	Харків	23	52	37	45	45	30	8	39	36	61
18	Черкаси	11	66	10	55	12	18	5	23	16	75
19	Ізюм	27	71	43	51	44	47	20	40	40	95
20	Луганськ	33	57	48	58	22	51	22	11	49	49
21	Дніпропетровськ	33	68	23	55	33	42	11	36	33	74
22	Запоріжжя	42	78	30	62	27	35	8	39	49	80
23	Донецьк	36	49	41	67	23	37	21	24	57	80
24	Бердянськ	67	63	38	76	19	48	16	28	43	75
25	Мелітополь	77	67	54	64	17	38	9	24	49	69
26	В.Олександрівка	49	86	12	70	22	19	5	23	42	89
27	Кіровоград	12	68	10	70	22	15	8	19	12	77
28	Вознесенськ	39	76	15	79	14	22	2	26	19	52
29	Любашівка	40	74	11	74	17	33	4	15	17	48
30	Херсон	52	82	13	69	16	26	10	20	53	80
31	Миколаїв	41	62	15	59	12	34	4	19	28	65
32	Роздільна	40	71	10	81	10	22	1	27	16	37
33	Одеса	47	63	18	67	6	22	3	27	33	52

ДОДАТОК Г



Рис. Г1 – Карта розташування метеорологічних станцій на території України.

ЛІТЕРАТУРА

1. Андерсон Т.В. Введение в многомерный статистический анализ.: М.Физматгиз, 1963. – 500 с.
2. Багров Н.В. Аналитическое представление последовательности метеорологических полей посредством естественных ортогональных функций // Труды ЦИП, 1959. Вып. 74. – С.133-138.
3. Багров Н.В. Преобразование и отбор предсказателей в корреляционном анализе. – Л.: Труды ГМЦ, 1970, вып. 64. – С. 3-23.
4. Борисенков Е.П. Физико-статистические методы анализа и предвычислений метеорологических полей.- Л.: Морской транспорт, 1963.- С.320.
5. Вапник В.Н., Червоненкис В.Я. Теория распознавания образов. – М.: Наука, 1974. – 252 с.
6. Гирко В.Л. Многомерный статистический анализ.:К.- Высшая школа, 1988.
7. Горелик В.Л., Скрипник В.В. Распознавания.-М.: Высшая школа, 1984.
8. Ивахненко В.Г., Долгосрочное прогнозирование и управление сложными системами.-К.: Техніка, 1975..
9. Кендалл М.Дж., Стюарт В. Многомерный статистический анализ и временные ряды.-М.: Наука, 1976.
10. Лоева И.Д. Применение кластерного анализа при решении задачи оценки достоверности измеренных концентраций двуокиси серы в атмосфере // Метеорология, климатология и гидрология.-К.:Высшая школа, 1985.- Вып. 21.- С43-49.
11. Лоули Д., Максвелл В. Факторный анализ как статистический метод.- М.: Мир, 1967.
12. Мещерская В.В., Репинская Р.П., Юдин М.И. Естественное ортогональное составляющее метеорологических полей. – Л.: Гидрометеиздат, 1970. – 205 с.
13. Многомерные статистические модели климатов. – Л.: Гидрометеиздат, 1989. – 183 с.
14. Райзин Дж.Вэн. Класификация и кластер. – М.: Мир, 1980. – 389 с.
15. Себестиан Г.С. Процессы принятия решений при распознавании образов. – К.: Техника, 1965. – 310 с.
16. Себер Дж. Линейный регрессионный анализ . – М.: Мир, 1980. – 456 с.
17. Школьный Є.П., Лоева І.Д., Гончарова Л.Д. Обробка та аналіз гідрометеорологічної інформації: підручник. – К.: Міносвіти України, 1999. – 600 с.

ПРЕДМЕТНИЙ ПОКАЖЧИК

А	К
альтернативний прогноз – 102	кластер – 60
арифметичним простором – 8	коефіцієнт - кореляції – 16, 64 - множинної детермінації – 85 - множинної кореляції – 85 - регресії – 76 - частинний – 88
В	Критерій - Лемана-Розенблатта – 66 - Фішера – 66, 86, 97 - Крамера-Уелча – 66 - омега-квадрат – 66 - Стюдента – 66, 69, 82, 83 - Смирнова – 67
вага фактора – 46	
вектори ситуації – 60	
випадкове поле – 10	
власні - значення – 27, 29, 33 - вектори – 27, 29	
Г	
головні компоненти – 34, 35	
Д	М
детермінант коваріації – 50	матриця - коваріації – 12, 13 - кореляції – 15, 16 - Фішера – 81
дисперсія – 14 - відтвореності – 82 - залишків – 49, 86 - нев'язки – 82	математичне сподівання – 15
дискримінантна функція – 106	метрика – 9
Е	метричний арифметичний простір – 8
евклідова відстань – 9, 61	мінори – 89
евристичний – 64	міра схожості – 63
емпіричні ортогональні ф-ції – 31	Н
еталон кластера – 61	навчальна сукупність – 87
З	нев'язка предиктанта – 95
задача фільтрації – 26, 36	норма вектора – 9
І	О
ізотропність – 18	однорідне поле – 18
індивідуальне поле – 11	ортогональні вектори – 25, 30
ітераційний процес – 51, 68	ортонормовані вектори – 30
ітераційний метод Якобі – 29	

П

параметризація поля – 27
перетворення
- розтягання – 27
- стиску – 27
поле – 9,
полнос кореляції – 17
помилка першого роду – 103
помилка другого роду – 103
предіктант – 76
предіктор – 27, 76

Р

розпізнавання образів – 102
розділяюча гіперповерхня – 103, 105

С

середнє поле – 12
середньо квадратичне відхилення – 15
скалярний добуток – 9
статистична сукупність – 11
стиском інформації – 26, 35

Т

транспонування – 11

У

узагальнений фактор – 46, 47

Ф

Функція

- відстані – 61
- дискримінантна – 36
- правдоподібності – 50
- подібності – 105

Ц

центрування – 12
центр кореляції – 17
центр кластера – 60

Ч

частинний коефіцієнт – 88
число Махалобіса – 63

Навчальне видання

БАГАТОВИМІРНИЙ СТАТИСТИЧНИЙ АНАЛІЗ ГІДРОМЕТЕОРОЛОГІЧНОЇ ІНФОРМАЦІЇ

Навчальний посібник

**ШКОЛЬНИЙ Євген Павлович,
СЕРГА Едуард Миколайович,
ГАЛИЧ Єлизавета Анатоліївна**

Підп. до друку _____ Формат 60×84/16
Умовн. друк. арк. 31,04. Наклад _____ прим. Зам № _____
Видавництво та друкарня "ТЕС"
(Свідоцтво ДК № 771) Одеса, Канатна 81/2
Надруковано з готового оригінал-макета

Одеський державний екологічний університет
65016, Одеса, вул. Львівська, 15
