

Л.Д. ГОНЧАРОВА

**МЕТОДИ БАГАТОВИМІРНОГО
СТАТИСТИЧНОГО АНАЛІЗУ
МЕТЕОРОЛОГІЧНИХ ПОЛІВ ТА
АТМОСФЕРНИХ ПРОЦЕСІВ**

Навчальний посібник



Гончарова Л.Д.

***МЕТОДИ БАГАТОВИМІРНОГО
СТАТИСТИЧНОГО АНАЛІЗУ
МЕТЕОРОЛОГІЧНИХ ПОЛІВ ТА
АТМОСФЕРНИХ ПРОЦЕСІВ***

Навчальний посібник

ОДЕСА – 2016

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ОДЕСЬКИЙ ДЕРЖАВНИЙ ЕКОЛОГІЧНИЙ УНІВЕРСИТЕТ

Гончарова Л.Д.

***МЕТОДИ БАГАТОВИМІРНОГО
СТАТИСТИЧНОГО АНАЛІЗУ
МЕТЕОРОЛОГІЧНИХ ПОЛІВ ТА
АТМОСФЕРНИХ ПРОЦЕСІВ***

*Затверджено Вченою радою
Одеського державного екологічного університету
Міністерства освіти і науки України
як навчальний посібник
для здобувачів вищої освіти
за спеціальністю «Науки про Землю»
(протокол № 8 від 24 вересня 2015 р.).*

ББК 26.23
Г65
УДК. 551.58

Гончарова Л.Д.

Методи багатовимірного статистичного аналізу метеорологічних полів та атмосферних процесів: Навчальний посібник. – Одеса: _____, 2016. – 195 с.

Навчальний посібник містить основні методи багатовимірного статистичного аналізу з обробки гідрометеорологічної інформації. Велика увага приділяється базовим теоретичним положенням та прикладним аспектам кореляційного, компонентного, факторного, кластерного та регресійного аналізу. В ньому наводяться алгоритми розв'язання великої кількості спеціальних задач з фізичним аналізом одержаних результатів, для розв'язання яких залучається сучасний емпіричний матеріал. Структура навчального посібника наближена, з необхідною деталізацією та методологією, до досліджень, якими займається сучасна гідрометеорологічна наука.

Навчальний посібник розрахований на здобувачів вищої освіти за спеціальністю «Науки про Землю».

Рецензенти:

Світличний О.О. – професор кафедри фізичної географії і природокористування Одеського національного університету ім. І.І. Мечникова, доктор географічних наук, професор.

Іванов С.В. – провідний науковий співробітник Одеського державного екологічного університету, доктор географічних наук.

© Л.Д. Гончарова, 2016
© Одеський державний
екологічний університет, 2016

ПЕРЕДМОВА

Проблема клімату нашої планети та його мінливості є однією з центральних у сучасній гідрометеорологічній науці. Інтерес до неї зростає з кожним роком. Це пов'язано, перш за все, з тим, що ефект впливу на клімат діяльності людини як цілеспрямований, так і ненавмисний, особливо з другої половини ХХ-го століття, різко зріс та в ряді випадків став співвимірним з впливом основного кліматоутворювального фактора – притоку сонячної радіації. З іншого боку, в умовах швидкого науково-технічного прогресу залежність людства від клімату не зменшилась, а, навпаки, суттєво зросла, завдяки чому дослідження мінливості клімату набули чітко визначеної практичної значущості.

На сьогоднішній день досягнуті достатньо великі успіхи в зборі, систематизації та узагальненні емпіричної інформації, яка характеризує клімат різного просторового та часового масштабу. Але розосередженість вихідної інформації ускладнює вивчення структури та мінливості сучасного клімату і потребує статистичного підходу, за допомогою якого можна використати для дослідження багаторічний архів даних, а це є серйозним резервом покращення якості довгострокових прогнозів погоди (ДПП).

Атмосфера (гідросфера), як відомо, є турбулентним середовищем, і поля метеорологічних (гідрологічних) величин дуже мінливі як у просторі, так і в часі. Тому для метеорологічних (гідрологічних) полів плідним є використання статистичного підходу, що опирається на методи теорії ймовірності та математичної статистики. В різних випадках величина може набувати різних значень. Сукупність усіх значень для об'єму простору, що розглядається в даному інтервалі часу, представляє собою реалізацію випадкового поля. Відбираючи всі об'єми та інтервали часу, які характеризуються в деякому сенсі однаковими зовнішніми умовами, можна отримати нескінченний набір реалізацій такого поля. Статистичний підхід потребує відмови від спроби роздільного розгляду індивідуальних властивостей випадкових полів. Розглядаються лише їх статистичні характеристики, які дозволяють встановити загальні особливості, характерні для всього набору реалізацій. Ці загальні особливості прийнято називати статистичною структурою випадкового поля.

Розвиток макропроцесів у просторі та параметрів гідрометеорологічних величин представляє собою багатовимірну область, в якій кожній точці стан макропроцесу визначено їх координатами. Для дослідження таких n -вимірних геофізичних систем необхідно залучати методи багатовимірного статистичного аналізу (БСА). Тому в навчальні плани Одеського державного екологічного університету (ОДЕКУ), в якому здійснюється підготовка магістрів-кліматологів, включена дисципліна «Багатовимірний аналіз метеорологічних полів та атмосферних процесів».

Метою дисципліни є пізнання студентами методів багатовимірної статистичного аналізу, які використовуються для дослідження статистичної структури гідрометеорологічних полів та побудови статистичних моделей гідрометеорологічних прогнозів.

Практична значущість дисципліни полягає у тому, щоб сформувати у магістрів систему знань, умінь та навичок, які дозволять фахівцю-кліматологу виконувати на сучасному рівні дослідження статистичної структури випадкових полів.

Після вивчення дисципліни «Багатовимірний аналіз метеорологічних полів та атмосферних процесів» магістр повинен вміти досліджувати метеорологічні поля на однорідність та ізотропність, використовуючи кореляційний аналіз, розв'язувати задачі стиску та фільтрації гідрометеорологічної інформації за допомогою компонентного аналізу; класифікувати вихідну інформацію, використовуючи кластерний аналіз та вміти будувати статистичні моделі у вигляді лінійного множинного рівняння регресії.

Один з останніх підручників «Обробка та аналіз гідрометеорологічної інформації» (автори Школьний Є.П., Лоева І.Д., Гончарова Л.Д.), опублікований ще у 1999 році, розкриває, в основному, теоретичні основи методів багатовимірної статистичного аналізу, але в ньому мало приділяється уваги шляхам реалізації цих методів при вирішенні практичних задач. До того ж, необхідність створення цього навчального посібника зумовлена стрімким розвитком наук про Землю, спричиненим суттєвими глобальними кліматичними змінами.

При написанні цього навчального посібника автор намагалася передати багаторічний особистий досвід наукових досліджень з застосуванням методів багатовимірної статистичного аналізу, а також наукових робіт, що виконувались та виконуються в Одеському державному екологічному університеті.

Структура посібника така, що кожен розділ починається з викладення базових теоретичних положень щодо теми, яка підлягає вивченню, і наведення прикладу розв'язання типової задачі з фізичним аналізом отриманих результатів. Наводяться запитання для самоперевірки по кожній темі та даються рекомендації щодо виконання практичних вправ. Тому він є необхідним доповненням до підручника «Обробка та аналіз гідрометеорологічної інформації».

У посібнику наводяться великі масиви різних гідрометеорологічних та кліматичних даних по 33 станціях України (перелік станцій та карта України наводяться у Додатку А). На жаль, не використані станції Автономної Республіки Крим, оскільки порушена безперервність даних. В Додатках наводяться вихідні дані (Додаток Б), допоміжні таблиці та довідковий матеріал (Додатки В, Г, Д, Е, Ж). Це дає можливість використовувати цей навчальний посібник при розв'язанні спеціальних

задач безпосередньо без будь-яких додаткових матеріалів, що дуже зручно у навчальному процесі.

Важливим є ще й те, що розв'язання типових задач передбачає не тільки виконання необхідних розрахунків із застосуванням того чи іншого методу БСА, а й фізичного аналізу одержаних результатів. На нашу думку, такий задачник повинен у значній мірі активізувати самостійну та індивідуальну роботу студентів.

Автор сподівається, що книга допоможе широкому колу спеціалістів-гідрометеорологів у засвоєнні основних ідей та методів дослідження випадкових просторових полів. Слід підкреслити, що у навчальному посібнику розглядаються особливості метеорологічних полів за умови просторових масштабів порядку сотен та більше кілометрів, що описуються статистичною макроструктурою.

Представлений у навчальному посібнику матеріал не претендує на викладення всіх методів, що використовуються в гідрометеорології для дослідження метеорологічних полів та атмосферних процесів. Він написаний на основі курсу лекцій з дисципліни «Багатовимірний аналіз метеорологічних полів та атмосферних процесів», які читаються магістрам на кафедрі «Метеорології та кліматології» Одеського державного екологічного університету. Крім того, він може використовуватися здобувачами вищої освіти за спеціальністю «Науки про Землю», яким необхідні знання особливостей макропроцесів у різних регіонах планети та на регіональному рівні. Викладені в навчальному посібнику базові теоретичні положення та приклади застосування методів БСА в реалізації різних практичних задач, на нашу думку, є підґрунтям для проведення наукових досліджень не тільки в ОДЕКУ, а й в інших вищих навчальних закладах України.

Зміст навчального посібника відповідає робочій програмі з дисципліни «Багатовимірний аналіз метеорологічних полів та атмосферних процесів».

Автор виражає щире подяку професору, доктору географічних наук О.О. Світличному та доктору географічних наук С.В. Іванову, які зробили ряд корисних зауважень щодо змісту, структури книги та висловили ряд побажань, що були враховані автором, а також висловлює подяку начальнику редакційного відділу ОДЕКУ Соколенко О.Д. за редагування тексту рукопису при підготовці цього навчального посібника до видання. Автор вдячна к.геогр.н., доценту Галич Є.А. за велику роботу по підготовці ілюстрацій, що наведені в посібнику, та аспіранту кафедри метеорології та кліматології ОДЕКУ Косолаповій Н.І. за технічну обробку тексту цієї книги.

1 ОСНОВНІ ЗАСАДИ БАГАТОВИМІРНОГО СТАТИСТИЧНОГО АНАЛІЗУ

1.1 Базові теоретичні положення

1.1.1 Елементи лінійної алгебри

Багатовимірний статистичний аналіз є важливим математичним методом дослідження просторово-часових змін складних і мінливих процесів, що відбуваються в кліматичній системі Землі.

Розв'язання чисельних задач сучасної гідрометеорології потребує вивчення та удосконалення статистичних методів при складанні довгострокових прогнозів погоди, при вивченні динаміки глобального чи регіонального кліматів, при розробці статистичних моделей метеорологічних чи гідрологічних прогнозів тощо.

Якщо йдеться про статистичні моделі, то вихідною інформацією є статистичні сукупності випадкових величин, які відіграють роль факторів, що впливають на фізичні процеси. Оцінки параметрів статистичної моделі метеорологічного прогнозу визначаються за допомогою тих чи інших методів багатовимірного статистичного аналізу на основі цієї інформації.

Методи багатовимірного статистичного аналізу базуються на визначених аксіомах і теоремах лінійної алгебри та теорії ймовірності.

Всі математичні об'єкти будемо розглядати і визначати у багатовимірному просторі [1, 5, 17, 18, 24, 25, 27].

Арифметичним багатовимірним простором називають множину будь-яких n -вимірних систем упорядкованих дійсних чисел a_1, a_2, \dots, a_n . Саму систему чисел називають точкою n -вимірного арифметичного простору. Якщо визначити початок координат у $m.O (0, 0, \dots, 0)$, то $m.A (a_1, a_2, \dots, a_n)$ може ототожнюватися з n -вимірним вектором $\vec{A}(a_1, a_2, \dots, a_n)$. Числа, що розташовуються в дужках, у цьому випадку мають сенс координат вектора \vec{A} .

Велике практичне застосування має **метричний арифметичний простір**. Це такий простір, де будь-яким двом точкам цього простору ($m.A$ і $m.B$) ставиться у відповідність **дійсне число** $\rho(A, B)$, яке **називають метрикою** й вона задовольняє такі вимоги:

- 1) $\rho(A, B) \geq 0$ (властивість невід'ємності); за умови $A=B$, $\rho(A, B)=0$.

- 2) $\rho(A, B) = \rho(B, A)$ (властивість симетричності).
- 3) Властивість трикутника: для будь-яких трьох точок цього простору (т.т. A, B, C)

$$\rho(A, B) \leq \rho(A, C) + \rho(B, C).$$

Метричний арифметичний n -вимірний простір називають евклідовим (R^n) , якщо для точок $A(a_1, a_2, \dots, a_n)$ та $\hat{A}(b_1, b_2, \dots, b_n)$ метрикою $\rho(A, B)$ є евклідова відстань, що визначається формулою:

$$\rho(A, B) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + \dots + (b_n - a_n)^2} = \|\overrightarrow{AB}\|. \quad (1.1)$$

Цю метрику називають **нормою вектора \overrightarrow{AB}** . Властивості метрики $\rho(A, B)$, що перелічені вище, носять назву **аксіоматики метричного простору**. Для евклідового простору **скалярний добуток** двох векторів є скалярною величиною й визначається рівнянням:

$$\langle \overrightarrow{AB} \rangle = \overrightarrow{AB} = \sum_{i=1}^n a_i b_i, \quad (1.2)$$

тобто скалярний добуток двох векторів, що складаються з n компонент (координат), визначається як сума добутку координат відповідних векторів.

1.1.2 Метеорологічні поля як об'єкти багатовимірною статистичного аналізу

Розвиток теорії фомування та змін глобального клімату і довгострокового прогнозування погоди потребує кількісного врахування взаємодій процесів в атмосфері, океані, суходолі та в кріосфері Земної кулі, в зв'язку з чим стало необхідним вивчення складної кліматичної системи за визначений термін із залученням вірогідних статистичних характеристик і відповідних методів дослідження. Наприклад, кліматичне поле температури повітря є характеристикою стану однієї з ланок кліматичної системи. Їх структура залежить від особливостей розподілу сонячної радіації на поверхні землі, а також від процесів взаємодії між ланками кліматичної системи (розподіл суші та океану, полів морського льоду, рельєфу поверхні та т.п.). Всі ці фактори впливають на значні

коливання полів температури у просторі та часі. Тому формування будь-яких метеорологічних полів відбувається під дією комплексу кліматоутворювальних процесів та географічних факторів клімату, певним чином взаємодіючих між собою [9, 11, 17].

Як відомо, характеристики стану атмосфери чи гідросфери (гідрометеорологічні величини) залежать від фізичних процесів, що безперервно розвиваються в найбільш мінливих оболонках Землі.

Для гідрометеорологічних досліджень, а також безпосереднього використання метеорологічної інформації в практичних цілях формується вихідна база випадкових величин зі значними просторовими та часовими масштабами.

Статистичні сукупності, як відомо, складаються не тільки з величин безпосередньо вимірних. Це можуть бути і величини, що отримані в результаті узагальнення первинних вимірювань чи спостережень.

Метеорологічні та гідрологічні вимірювання виконуються у чітко встановлені моменти часу на мережі станцій чи постів. Таким чином маємо значення гідрометеорологічної величини на множині точок деякої території. Положення цих точок визначено географічними координатами і їм можуть надаватись конкретні номери 1, 2, 3, ..., n . Мережа точок може бути *нерегулярною*, якщо відстані між ними різні, та – *регулярною*, коли ці відстані в деякій системі координат однакові. Регулярна мережа точок може будуватись шляхом інтерполяції значень гідрометеорологічної величини в її вузлах. Прикладом нерегулярної мережі точок є мережа метеорологічних станцій та постів. Але в обох випадках *система точок є упорядкованою* в тому сенсі, що вони розташовані відповідно до географічних координат станцій у нерегулярній мережі точок або прямокутними умовними координатами при регулярній мережі.

*Упорядковану сукупність значень метеорологічної чи гідрологічної величини на множині точок багатовимірного простору в фіксований (певний) момент часу називають **полем цієї величини**.*

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

Відомо, що метеорологічні поля мають різні форми представлення. Інколи їх зображають графічно у вигляді відповідних карт, на яких спочатку наносять значення тієї чи іншої метеорологічної величини, а потім проводять ізолінії (ізотерми, ізобари, ізогіети, ізогіпси, ізалобари, ізотахи і т.п.). Такий вигляд представлення метеорологічних полів неможливий, коли в процесі дослідження треба використовувати статистичні методи.

Крім того, поля гідрометеорологічних величин, як було зазначено, визначені як сукупність їх значень на множині упорядкованих точок

багатовимірному простору у фіксований час. А це дає підстави представляти гідрометеорологічні поля у векторній формі при застосуванні методів багатовимірної статистичної аналізу [6, 27].

Як відомо з лінійної алгебри [24], вектори зображають упорядкованою системою координат (вектори будемо позначати великими літерами латиниці $A, B, \dots, X, Y, Z, \dots$, а значення координат вектора – відповідними малими літерами $a, b, \dots, x, y, z, \dots$).

Розглянемо n -вимірний вектор-стовпець X_j з координатами

$$X_j = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \dots \\ x_{ij} \\ \dots \\ x_{nj} \end{pmatrix}, \quad (i = \overline{1, n}). \quad (1.3)$$

У статистичних дослідженнях є всі підстави розглядати вектор X_j як поле метеорологічної величини [27]. Дійсно, оскільки це поле визначається сукупністю її значень на мережі метеорологічних станцій (на множині вузлів сітки точок), а положення кожної станції (вузла) добре відомо, то розташувавши значення метеорологічної величини відповідно до визначеного порядку нумерації станцій (вузлів), отримаємо упорядковану систему дійсних чисел. Вони й складають вектор-стовпець n -вимірному евклідовому простору вигляду (1.3) або, якщо провести **операцію транспонування**, то отримаємо вектор-рядок:

$$X'_j = (x_{j1}, x_{j2}, \dots, x_{ji}, \dots, x_{jn}), \quad (i = \overline{1, n}), \quad (1.4)$$

де (') – позначення **операції транспонування**.

Індекс j біля координат цих векторів позначає час, до якого відноситься дане поле, або номер поля. Якщо у процесі дослідження необхідно розглянути m полів, то $j = \overline{1, m}$. Таким чином, ми використовуємо m векторів вигляду (1.3) або (1.4).

Як відомо, значення метеорологічної величини має випадковий характер. Тому поля метеорологічних величин будемо у подальшому розглядати як

випадкові вектори n -вимірною евклідового простору [6, 14, 27]. Випадковий вектор може також відбивати не тільки метеорологічне поле, а й вертикальний профіль метеорологічної (гідрологічної) величини, що характеризує стан того чи іншого гідрометеорологічного процесу.

Інформацію про множину полів (вертикальних профілів чи векторів-предикторів) можна розглядати як відповідну множину з m випадкових n -вимірних векторів або, якщо розташувати цю часову послідовність векторів (1.3) один за одним, то як 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. Отже, стовпці матриці X – це індивідуальні поля гідрометеорологічної величини, а рядки – їх статистичні сукупності на i -й станції.

Така форма представлення інформації про гідрометеорологічні поля дуже зручна, тому що дає можливість побудувати прості алгоритми дослідження їх статистичної структури.

Важливим видом представлення метеорологічних об'єктів є вектор середніх значень гідрометеорологічної величини на i -й станції [6, 27]. Він має сенс середнього поля чи середнього вертикального профілю випадкової величини

$$\bar{X} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \vdots \\ \bar{x}_i \\ \vdots \\ \bar{x}_n \end{pmatrix}, \quad (i = \overline{1, n}). \quad (1.6)$$

Кожен елемент вектора (1.6) розраховується за формулою (1.7):

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

де

m – об'єм статистичних сукупностей.

1.1.3 Матричне представлення метеорологічних полів

Матриці коваріацій та кореляцій представляють собою математичні об'єкти, які містять найбільш важливу інформацію про статистичну структуру випадкових метеорологічних полів. Вона є основою для втілення багатьох методів багатовимірного статистичного аналізу в гідрометеорології [1, 6, 27].

1.1.3.1 Матриця коваріацій та алгоритм її розрахування

Матриця коваріацій визначається таким матричним рівнянням [27]:

$$K_x = \frac{1}{m} \Delta X \Delta X', \quad (1.8)$$

де

ΔX – матриця центрованих елементів;

$\Delta X'$ – транспонована матриця від ΔX ;

m – об'єм статистичних сукупностей.

Якщо розглядається матриця X вигляду (1.5), тоді матриця ΔX в координатній формі має такий вигляд:

$$\Delta X = \begin{pmatrix} \Delta x_{11} & \Delta x_{12} & \Delta x_{13} & \dots & \Delta x_{1j} & \dots & \Delta x_{1m} \\ \Delta x_{21} & \Delta x_{22} & \Delta x_{23} & \dots & \Delta x_{2j} & \dots & \Delta x_{2m} \\ \Delta x_{31} & \Delta x_{32} & \Delta x_{33} & \dots & \Delta x_{3j} & \dots & \Delta x_{3m} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \Delta x_{i1} & \Delta x_{i2} & \Delta x_{i3} & \dots & \Delta x_{ij} & \dots & \Delta x_{im} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \Delta x_{s1} & \Delta x_{s2} & \Delta x_{s3} & \dots & \Delta x_{sj} & \dots & \Delta x_{sm} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \Delta x_{n1} & \Delta x_{n2} & \Delta x_{n3} & \dots & \Delta x_{nj} & \dots & \Delta x_{nm} \end{pmatrix}, \quad (1.9)$$

де кожен елемент Δx_{ij} визначається як

$$\Delta x_{ij} = x_{ij} - \bar{x}_i, \quad (i = \overline{1, n}; \quad j = \overline{1, m}). \quad (1.10)$$

Операцію, проведену над матрицею (1.5), **називають операцією центрування**.

Розглянемо вектор-рядок ΔX_i матриці (1.9):

$$\Delta X_i = (\Delta x_{i1} \quad \Delta x_{i2} \quad \dots \quad \Delta x_{ij} \quad \dots \quad \Delta x_{im}). \quad (1.11)$$

Знайдемо норму вектора (1.11):

$$\|\Delta X_i\| = \sqrt{\Delta x_1^2 + \Delta x_2^2 + \dots + \Delta x_m^2} = \sqrt{\sum_{j=1}^m \Delta x_{ij}^2}. \quad (1.12)$$

Очевидно квадрат норми вектора дорівнює:

$$\|\Delta X_i\|^2 = \sum_{j=1}^m \Delta x_{ij}^2 = \sum_{j=1}^m (x_{ij} - \bar{x}_i)^2. \quad (1.13)$$

Якщо поділити квадрат норми вектора ΔX_i на m (кількість координат, об'єм статистичної сукупності), то отримаємо дисперсію метеорологічної величини на i -й станції [13]:

$$\frac{\|\Delta X_i\|^2}{m} = \frac{\sum_{j=1}^m (x_{ij} - \bar{x}_i)^2}{m} = \sigma_{x_i}^2. \quad (1.14)$$

Таким чином, *дисперсію метеорологічної величини в i -й точці поля можна трактувати як зменшений у m разів квадрат норми вектора (1.11). Тоді середній квадратичний відхил метеорологічної величини в i -й точці* (на i -й метеорологічній станції) є

$$\sigma_{x_i} = \frac{\|\Delta X_i\|}{\sqrt{m}}. \quad (1.15)$$

Розглянемо ще один вектор-рядок матриці (1.9), наприклад, ΔX_s :

$$\Delta X_s = (\Delta x_{s1} \quad \Delta x_{s2} \quad \dots \quad \Delta x_{sj} \quad \dots \quad \Delta x_{sm}). \quad (1.16)$$

Знайдемо скалярний добуток векторів ΔX_i та ΔX_s [24], враховуючи визначення скалярного добутку двох векторів евклідового простору (1.2). Будемо мати:

$$\begin{aligned} \langle \Delta X_i \cdot \Delta X_s \rangle &= \Delta X_i \cdot \Delta X'_s = \begin{pmatrix} \Delta x_{i1} & \Delta x_{i2} & \Delta x_{ij} & \dots & \Delta x_{im} \end{pmatrix} \begin{pmatrix} \Delta x_{1s} \\ \Delta x_{2s} \\ \vdots \\ \Delta x_{js} \\ \vdots \\ \Delta x_{ms} \end{pmatrix} = \\ &= \sum_{j=1}^m \Delta x_{ij} \Delta x_{sj} = \sum_{j=1}^m (x_{ij} - \bar{x}_i)(x_{sj} - \bar{x}_s). \end{aligned} \quad (1.17)$$

Вираз

$$\frac{1}{m} \sum_{j=1}^m (x_{ij} - \bar{x}_i)(x_{sj} - \bar{x}_s) = \frac{\langle \Delta X_i \cdot \Delta X_s \rangle}{m} = k_{is} \quad (1.18)$$

є *коваріацією* між метеорологічною величиною в точках i та s метеорологічного поля.

Щоб показати структуру матриці коваріації, реалізуємо алгоритм (1.8) за умови, що матриця ΔX має порядок $2 \times m$:

$$K_x = \frac{1}{m} \begin{pmatrix} \Delta x_{11} & \Delta x_{12} & \dots & \Delta x_{1j} & \dots & \Delta x_{1m} \\ \Delta x_{21} & \Delta x_{22} & \dots & \Delta x_{2j} & \dots & \Delta x_{2m} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \Delta x_{j1} & \Delta x_{j2} & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \Delta x_{m1} & \Delta x_{m2} & \dots & \dots & \dots & \dots \end{pmatrix}. \quad (1.19)$$

Використовуючи для скалярного добутку рівняння (1.2), отримаємо матрицю коваріацій порядку 2×2 :

$$K_x = \begin{pmatrix} \frac{1}{m} \sum_{j=1}^m (x_{1j} - \bar{x}_1)^2 & \frac{1}{m} \sum_{j=1}^m (x_{1j} - \bar{x}_1)(x_{2j} - \bar{x}_2) \\ \frac{1}{m} \sum_{j=1}^m (x_{2j} - \bar{x}_2)(x_{1j} - \bar{x}_1) & \frac{1}{m} \sum_{j=1}^m (x_{2j} - \bar{x}_2)^2 \end{pmatrix}. \quad (1.20)$$

Визначимо сенс кожного з елементів матриці (1.20). З урахуванням рівняння (1.14), елементи головної діагоналі є дисперсіями випадкової величини σ_1^2 , σ_2^2 [13], а відповідно до рівняння (1.18), елементи побічної діагоналі – коваріаційні моменти у зазначених точках поля [13].

Отже отримана матриця коваріації має вигляд:

$$K_x = \begin{pmatrix} \sigma_1^2 & k_{12} \\ k_{21} & \sigma_2^2 \end{pmatrix}. \quad (1.21)$$

Аналогічним чином розраховуються елементи матриці коваріації n -го порядку і в координатній формі вона має такий вигляд:

$$K_x = \begin{pmatrix} \sigma_1^2 & k_{12} & \dots & k_{1j} & \dots & k_{1s} & \dots & k_{1n} \\ k_{21} & \sigma_2^2 & \dots & k_{2j} & \dots & k_{2s} & \dots & k_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ k_{i1} & k_{i2} & \dots & k_{ij} & \dots & k_{is} & \dots & k_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ k_{n1} & k_{n2} & \dots & k_{nj} & \dots & k_{ns} & \dots & \sigma_n^2 \end{pmatrix}. \quad (1.22)$$

Порядковий номер дисперсії на головній діагоналі матриці (1.22) відповідає номеру метеорологічної станції. Інші елементи k_{ij} матриці (1.22) – відповідні коваріації на i -й та j -й метеорологічних станціях.

Матриця коваріацій K_x випадкової величини X має такі властивості: вона є **симетричною** ($k_{ij} = k_{ji}$); на головній діагоналі розташовуються дисперсії σ_i^2 ($i = \overline{1, n}$); **її елементи – дійсні числа**; вона є **додатно визначеною** (детермінант матриці – додатне число ($|K_x| > 0$)).

Матриця коваріацій, поряд з вектором математичних сподівань m_x відіграє роль параметра щільності ймовірності багатовимірного нормального розподілу [6, 14, 27]

$$f(X) = \frac{1}{(2\pi)^{\frac{n}{2}} |K_x|^{\frac{1}{2}}} \times \exp \left[-\frac{1}{2} (X - m_x) K_x^{-1} (X - m_x) \right], \quad (1.23)$$

де

індекс (-1) позначає **операцію обернення матриці** K_x ;

$|K_x|$ – визначник матриці K_x .

1.1.3.2 Матриця кореляцій та алгоритм її розрахування

Матриця кореляцій визначається таким матричним рівнянням [27]:

$$R_x = \sigma^{-1} K_x \sigma^{-1}, \quad (1.24)$$

де

K_x – матриця коваріацій;

σ^{-1} – *обернена матриця* від діагональної матриці (σ) середніх квадратичних відхилів.

Діагональна матриця [24] *середніх квадратичних відхилів* має вигляд:

$$\sigma = \begin{pmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_n \end{pmatrix}. \quad (1.25)$$

Тоді *обернена матриця* σ^{-1} є:

$$\sigma^{-1} = \begin{pmatrix} \frac{1}{\sigma_1} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{1}{\sigma_n} \end{pmatrix}. \quad (1.26)$$

Отримаємо спочатку матрицю кореляцій другого порядку в координатній формі. Для цього реалізуємо матричне рівняння (1.24):

$$\begin{aligned} \sigma^{-1} K_x \sigma^{-1} &= \begin{pmatrix} \frac{1}{\sigma_1} & 0 \\ 0 & \frac{1}{\sigma_2} \end{pmatrix} \begin{pmatrix} \sigma_1^2 & k_{12} \\ k_{21} & \sigma_2^2 \end{pmatrix} \begin{pmatrix} \frac{1}{\sigma_1} & 0 \\ 0 & \frac{1}{\sigma_2} \end{pmatrix} = \\ &= \begin{pmatrix} \frac{\sigma_1^2}{\sigma_1^2} & \frac{k_{12}}{\sigma_1 \sigma_2} \\ \frac{k_{21}}{\sigma_2 \sigma_1} & \frac{\sigma_2^2}{\sigma_2^2} \end{pmatrix} = \begin{pmatrix} 1 & r_{12} \\ r_{21} & 1 \end{pmatrix} = R_x. \end{aligned} \quad (1.27)$$

Оскільки [13]

$$\frac{k_{is}}{\sigma_{x_i} \sigma_{x_s}} = \frac{k_{is} m}{\sqrt{m} \sigma_{x_i} \sqrt{m} \sigma_{x_s}} = \frac{\langle \Delta X_i \Delta X_s \rangle}{\|\Delta X_i\| \|\Delta X_s\|} = r_{is}. \quad (1.28)$$

Вираз (1.28) має сенс **кореляції**, яка визначає міру лінійного кореляційного зв'язку між метеорологічною величиною в точках i та s поля, i є нормованим скалярним добутком векторів ΔX_i та ΔX_s . Тому **кореляцію** й називають **нормованим коваріаційним моментом**.

З іншого боку, за визначенням скалярного добутку двох векторів евклідового простору [24] маємо:

$$\langle \Delta X_i \Delta X_s \rangle = \|\Delta X_i\| \|\Delta X_s\| \cos(\Delta X_i, \Delta X_s). \quad (1.29)$$

Відповідно

$$\frac{\langle \Delta X_i \Delta X_s \rangle}{\|\Delta X_i\| \|\Delta X_s\|} = \cos(\Delta X_i, \Delta X_s) = r_{is}. \quad (1.30)$$

Таким чином, **кореляцію можна трактувати як косинус кута між векторами** ΔX_i та ΔX_s .

За умови некорельованості випадкових величин X_i та X_s $r_{is} = 0$. Вектори ортогональні, якщо скалярний добуток двох векторів дорівнює нулю $\langle \Delta X_i \Delta X_s \rangle = 0$, $i \neq s$.

Таким чином, **поняття ортогональності та некорельованості є однозначними**, коли йдеться про випадкові величини [27].

Система координат у n -вимірному просторі, що представлена сукупністю статистично незалежних векторів, є ортогональною системою координат та носить назву **базису евклідового простору** [24].

Якщо розглянути нерівність Коші-Буняковського (1.31), яка в лінійній алгебрі чітко обґрунтовується [24, 27]

$$\| \langle \vec{A} \vec{B} \rangle \| \leq \| \vec{A} \| \| \vec{B} \|, \quad (1.31)$$

для векторів ΔX_i та ΔX_s отримаємо:

$$\| \langle \Delta X_i \Delta X_s \rangle \| \leq \| \Delta X_i \| \| \Delta X_s \|. \quad (1.32)$$

Врахування формул (1.15) і (1.18) дає можливість отримати нерівність (1.33):

$$|k_{ij}| \leq \sigma_{x_i} \sigma_{x_s}, \quad (1.33)$$

а розділивши обидві її частини на $\sigma_{x_i} \sigma_{x_s}$ маємо:

$$|r_{is}| \leq 1. \quad (1.34)$$

Узагальнюючи результат (1.27) на матрицю кореляцій n -го порядку, отримаємо:

$$R_x = \begin{pmatrix} 1 & r_{12} & \dots & r_{1j} & \dots & r_{1s} & \dots & r_{1n} \\ r_{21} & 1 & \dots & r_{2j} & \dots & r_{2s} & \dots & r_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ r_{i1} & r_{i2} & \dots & r_{ij} & \dots & r_{is} & \dots & r_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ r_{n1} & r_{n2} & \dots & r_{nj} & \dots & r_{ns} & \dots & 1 \end{pmatrix}. \quad (1.35)$$

Матриця кореляцій R_x випадкової величини X має властивості, аналогічні властивостям матриці коваріацій, тобто **вона є симетричною** ($r_{is} = r_{si}$), **її елементи – дійсні числа; вона є додатно визначеною** ($|K_x| > 0$), але **на її головній діагоналі розташовуються одиниці**.

У матриці кореляцій (1.35) міститься інформація про структуру n полів кореляцій. Дійсно, елемент r_{is} цієї матриці характеризує лінійний кореляційний зв'язок між метеорологічною величиною на i -й та s -й станціях, тобто рядок чи відповідний стовпець матриці R_x складає **поле кореляцій**. **Полюсом** цього **поля (центром кореляцій)** буде i -та метеорологічна станція, елемент кореляційної матриці для якої розташовується на перетину i -го рядка та i -го стовпця матриці R_x .

Як бачимо, матриці коваріацій та кореляцій містять важливу інформацію про особливості статистичної структури випадкових об'єктів. Якщо, наприклад, йдеться про гідрометеорологічні поля, то поряд із полем середніх значень, яке характеризує вектор (1.6), можна побудувати за допомогою матриці коваріацій поле дисперсій метеорологічної величини або поле середніх квадратичних відхилів.

1.2 Розв'язання типової задачі. Структура матриць коваріацій та кореляцій

На основі вихідної матриці $X = \begin{pmatrix} 12 & 14 & 10 & 5 \\ 25 & 30 & 27 & 15 \end{pmatrix}$, в якій кожен рядок є статистичною сукупністю ($i = \overline{1,2}; j = \overline{1,4}$), розрахувати:

- ❖ дисперсії випадкової величини X в 1-й та 2-й точках поля ($\sigma_{x_1}^2, \sigma_{x_2}^2$);
- ❖ середні квадратичні відхили випадкової величини X в 1-й та 2-й точках поля ($\sigma_{x_1}, \sigma_{x_2}$);
- ❖ коваріаційний момент між випадковою величиною X в 1-й та 2-й точках поля ($k_{x_1x_2}$);
- ❖ кореляційний момент між випадковою величиною X в 1-й та 2-й точках поля ($r_{x_1x_2}$);
- ❖ матрицю коваріацій K_x , використовуючи матричне рівняння

$$K_x = \frac{1}{m} \Delta X \Delta X';$$

- ❖ матрицю кореляцій R_x , використовуючи матричне рівняння

$$R_x = \sigma^{-1} K_x \sigma^{-1}.$$

1) За визначенням, **дисперсію випадкової величини X в i -й точці поля** можна трактувати як зменшений у m разів квадрат норми вектора ΔX_i

$$\sigma_{x_i}^2 = \frac{\|\Delta X_i\|^2}{m}, \quad (i = \overline{1,2}).$$

Для знаходження норми вектора ΔX_i треба визначити вектор-рядок центрованих значень

$$\Delta X_i = (\Delta x_{i1} \Delta x_{i2} \dots \Delta x_{ij} \dots \Delta x_{im}),$$

де $\Delta x_{ij} = (x_{ij} - \bar{x}_i)$ – операція центрування;

$\bar{x}_i = \frac{1}{m} \sum_{j=1}^m x_{ij}$ – середнє значення випадкової величини X ($i = \overline{1,2}; j = \overline{1,4}$).

Розрахувавши середні значення в 1-й та 2-й точках поля, сформуємо вектор середніх значень:

$$\bar{X} = \begin{pmatrix} 10,3 \\ 24,3 \end{pmatrix}.$$

Далі отримаємо вектори-рядки центрованих значень випадкової величини X для відповідних точок поля:

$$\Delta X_1 = (1,7 \ 3,7 \ -0,3 \ -5,3),$$

$$\Delta X_2 = (0,7 \ 5,7 \ 2,7 \ -9,3).$$

Норми векторів $\|\Delta X_1\|$ та $\|\Delta X_2\|$ визначимо за допомогою рівняння

$$\|\Delta X_i\| = \sqrt{\Delta x_{i1}^2 + \Delta x_{i2}^2 + \dots + \Delta x_{im}^2} = \sqrt{\sum_{j=1}^m \Delta x_{ij}^2}.$$

Отримаємо:

$$\|\Delta X_1\| = \sqrt{1,7^2 + 3,7^2 + (-0,3)^2 + (-5,3)^2} = \sqrt{44,76} = 6,69;$$

$$\|\Delta X_2\| = \sqrt{0,7^2 + 5,7^2 + 2,7^2 + (-9,3)^2} = \sqrt{126,76} = 11,26.$$

Дисперсії випадкової величини X для 1-ї та 2-ї точок поля дорівнюють відповідно:

$$\sigma_{x_1}^2 = \frac{\|\Delta X_1\|^2}{4} = \frac{(6,69)^2}{4} = 11,19;$$

$$\sigma_{x_2}^2 = \frac{\|\Delta X_2\|^2}{4} = \frac{(11,26)^2}{4} = 31,70.$$

2) Використовуючи рівняння (1.15), **середні квадратичні відхили випадкової величини X** у зазначених точках поля є:

$$\sigma_{x_1} = \frac{\|\Delta X_1\|}{\sqrt{m}} = \frac{6,69}{\sqrt{4}} = 3,35;$$

$$\sigma_{x_2} = \frac{\|\Delta X_2\|}{\sqrt{m}} = \frac{11,26}{\sqrt{4}} = 5,63.$$

3) Знайдемо коефіцієнт коваріації $k_{x_1x_2}$ випадкової величини X , використовуючи рівняння:

$$k_{x_1x_2} = \frac{\langle \Delta X_1 \Delta X_2 \rangle}{m}.$$

Скалярний добуток двох векторів евклідового простору визначається як:

$$\langle \Delta X_1 \Delta X_2 \rangle = \Delta X_1 \Delta X_2' = (\Delta x_{11} \Delta x_{12} \dots \Delta x_{1j} \dots \Delta x_{1m}) \begin{pmatrix} \Delta x_{12} \\ \Delta x_{22} \\ \vdots \\ \Delta x_{j2} \\ \vdots \\ \Delta x_{m2} \end{pmatrix}.$$

Для вказаних векторів скалярний добуток дорівнює:

$$\Delta X_1 \Delta X_2' = (1,7 \quad 3,7 \quad -0,3 \quad -5,3) \begin{pmatrix} 0,7 \\ 5,7 \\ 2,7 \\ -9,3 \end{pmatrix} =$$

$$= 1,19 + 21,09 + (-0,81) + 49,29 = 70,76.$$

Визначимо коефіцієнт коваріації між випадковою величиною X в 1-й та 2-й точках поля

$$k_{x_1x_2} = \frac{\Delta X_1 \Delta X_2'}{m} = \frac{70,76}{4} = 17,69.$$

4) Коефіцієнт кореляції $r_{x_1x_2}$ між випадковою величиною X в 1-й та 2-й точках поля визначається як:

$$r_{x_1x_2} = \frac{\Delta X_1 \Delta X_2'}{\|\Delta X_1\| \|\Delta X_2\|} = \frac{70,76}{6,69 \cdot 11,26} = 0,94.$$

5) Розрахуємо **матрицю коваріацій**, використовуючи матричне рівняння:

$$K_x = \frac{1}{m} \Delta X \Delta X',$$

де

$$m = 4;$$

$$\Delta X = \begin{pmatrix} 1,7 & 3,7 & -0,3 & -5,3 \\ 0,7 & 5,7 & 2,7 & -9,3 \end{pmatrix}; \quad \Delta X' = \begin{pmatrix} 1,7 & 0,7 \\ 3,7 & 5,7 \\ -0,3 & 2,7 \\ -5,3 & -9,3 \end{pmatrix}.$$

Розрахунок K_x :

$$K_x = \frac{1}{4} \begin{pmatrix} 1,7 & 3,7 & -0,3 & -5,3 \\ 0,7 & 5,7 & 2,7 & -9,3 \end{pmatrix} \begin{pmatrix} 1,7 & 0,7 \\ 3,7 & 5,7 \\ -0,3 & 2,7 \\ -5,3 & -9,3 \end{pmatrix} = \begin{pmatrix} 11,19 & 17,69 \\ 17,69 & 31,70 \end{pmatrix}.$$

6) Щоб отримати **матрицю кореляцій**, застосуємо матричне рівняння:

$$R_x = \sigma^{-1} K_x \sigma^{-1},$$

де

$$\sigma = \begin{pmatrix} 3,35 & 0 \\ 0 & 5,63 \end{pmatrix}, \quad \sigma^{-1} = \begin{pmatrix} \frac{1}{3,35} & 0 \\ 0 & \frac{1}{5,63} \end{pmatrix}, \quad K_x = \begin{pmatrix} 11,19 & 17,69 \\ 17,69 & 31,70 \end{pmatrix}.$$

Тоді

$$R_x = \begin{pmatrix} \frac{1}{3,35} & 0 \\ 0 & \frac{1}{5,63} \end{pmatrix} \begin{pmatrix} 11,19 & 17,69 \\ 17,69 & 31,70 \end{pmatrix} \begin{pmatrix} \frac{1}{3,35} & 0 \\ 0 & \frac{1}{5,63} \end{pmatrix} = \begin{pmatrix} \frac{11,19}{3,35 \cdot 3,35} & \frac{17,69}{3,35 \cdot 5,63} \\ \frac{17,69}{3,35 \cdot 5,63} & \frac{31,70}{5,63 \cdot 5,63} \end{pmatrix} = \begin{pmatrix} 1 & 0,94 \\ 0,94 & 1 \end{pmatrix}.$$

Висновок: Матриця коваріації K_x є симетричною ($k_{x_1x_2} = k_{x_2x_1} = 17,69$); на головній діагоналі розташовуються дисперсії випадкової величини X у визначених точках поля; її елементи є дійсними числами і вона є додатно визначеною.

Матриця кореляції R_x є симетричною ($r_{x_1x_2} = r_{x_2x_1} = 0,94$); на головній діагоналі розташовуються одиниці; її елементи є дійсними числами і вона є додатно визначеною.

1.3 Запитання для самоперевірки

1. Дати визначення багатовимірних «арифметичного», «метричного арифметичного» та «евклідового» просторів.
2. Дати визначення «метеорологічного поля». Перелічити форми представлення метеорологічних полів при дослідженнях атмосферних процесів. Вміти представляти поля у матричному вигляді.
3. Як матрично представити сукупність метеорологічних полів? Чим є кожен рядок і стовпець такої вихідної матриці?
4. Як отримати матрицю центрованих значень ΔX , якщо є вихідна матриця X ? Чому дорівнює кожен 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 у m разів (m – об'єм статистичної сукупності)?
9. Як розрахувати середній квадратичний відхил випадкової величини на метеорологічній станції 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?
12. Яка інформація міститься в матриці кореляцій?
13. Як отримати матрицю коваріацій (кореляцій)? Як представити матрицю коваріацій (кореляцій) у координатній формі? Які властивості мають ці матриці?

1.4 Завдання для закріплення теми

На основі матриці X , в якій кожен рядок є статистичною сукупністю (табл.1.1), знайти: дисперсію випадкової величини X в двох точках поля $(\sigma_{x_1}^2, \sigma_{x_2}^2)$; середні квадратичні відхили $(\sigma_{x_1}, \sigma_{x_2})$; коефіцієнт коваріації $(k_{x_1x_2})$; коефіцієнт кореляції $(r_{x_1x_2})$; отримати матриці коваріації (K_x) та кореляції (R_x) другого порядку.

Таблиця 1.1 – Вихідна інформація до розв’язання задач

Варіант	Матриця	Варіант	Матриця
1	$X = \begin{pmatrix} 2 & 4 & 6 & 1 & 8 \\ 45 & 43 & 40 & 38 & 35 \end{pmatrix}$	10	$X = \begin{pmatrix} 7 & 15 & 18 & 35 & 12 \\ 21 & 42 & 34 & 5 & 63 \end{pmatrix}$
2	$X = \begin{pmatrix} 12 & 14 & 10 & 5 & 2 \\ 50 & 55 & 49 & 52 & 51 \end{pmatrix}$	11	$X = \begin{pmatrix} 72 & 60 & 50 & 33 & 18 \\ 16 & 21 & 7 & 82 & 11 \end{pmatrix}$
3	$X = \begin{pmatrix} 4 & 7 & 10 & 8 & 11 \\ 30 & 25 & 41 & 20 & 3 \end{pmatrix}$	12	$X = \begin{pmatrix} 42 & 47 & 31 & 29 & 15 \\ 61 & 79 & 17 & 19 & 110 \end{pmatrix}$
4	$X = \begin{pmatrix} 33 & 42 & 51 & 15 & 3 \\ 12 & 18 & 4 & 21 & 37 \end{pmatrix}$	13	$X = \begin{pmatrix} 33 & 14 & 41 & 27 & 51 \\ 11 & 18 & 25 & 38 & 49 \end{pmatrix}$
5	$X = \begin{pmatrix} 12 & 15 & 25 & 38 & 10 \\ 34 & 14 & 19 & 13 & 47 \end{pmatrix}$	14	$X = \begin{pmatrix} 27 & 19 & 31 & 52 & 73 \\ 4 & 23 & 75 & 94 & 82 \end{pmatrix}$
6	$X = \begin{pmatrix} 20 & 38 & 40 & 12 & 22 \\ 1 & 34 & 10 & 23 & 51 \end{pmatrix}$	15	$X = \begin{pmatrix} 10 & 11 & 30 & 16 & 43 \\ 13 & 34 & 29 & 5 & 11 \end{pmatrix}$
7	$X = \begin{pmatrix} 7 & 25 & 12 & 13 & 45 \\ 26 & 22 & 40 & 1 & 68 \end{pmatrix}$	16	$X = \begin{pmatrix} 39 & 12 & 47 & 48 & 42 \\ 14 & 20 & 29 & 8 & 22 \end{pmatrix}$
8	$X = \begin{pmatrix} 27 & 34 & 11 & 14 & 18 \\ 25 & 77 & 40 & 84 & 1 \end{pmatrix}$	17	$X = \begin{pmatrix} 12 & 23 & 8 & 63 & 3 \\ 60 & 6 & 21 & 24 & 9 \end{pmatrix}$
9	$X = \begin{pmatrix} 9 & 11 & 20 & 36 & 29 \\ 15 & 33 & 23 & 38 & 24 \end{pmatrix}$	18	$X = \begin{pmatrix} 13 & 54 & 8 & 12 & 93 \\ 33 & 7 & 42 & 19 & 22 \end{pmatrix}$

2 КОРЕЛЯЦІЙНИЙ АНАЛІЗ МЕТЕОРОЛОГІЧНИХ ПОЛІВ

2.1 Базові теоретичні положення

2.1.1 Основні ідеї кореляційного аналізу

Кореляційний аналіз є одним з методів багатовимірного статистичного аналізу, за допомогою якого вирішується багато прикладних задач при дослідженні випадкових метеорологічних полів.

Для довгострокових прогнозів погоди необхідно глибоке вивчення реально існуючих просторово-часових зв'язків між гідрометеорологічними величинами та ролі фізико-географічних факторів клімату. Для дослідження кліматичного режиму та статистичної структури основних метеорологічних полів залучається кореляційний аналіз.

У сучасну епоху в науках про Землю широко використовується математичний апарат теорії випадкових функцій та випадкових полів [3, 6, 13, 14]. В основі цього є розгляд фіксованих значень гідрометеорологічних процесів та просторових полів як окремих реалізацій деякого поля. Такий підхід дозволяє відмовитися від розгляду особливостей окремих миттєвих значень гідрометеорологічних полів, залежність яких від просторових координат, а також їх часовий хід носять доволі складний характер, та перейти до розгляду деяких осереднених властивостей статистичної сукупності їх реалізацією, що відповідає деякій сукупності фіксованих зовнішніх умов. При цьому мається на увазі, що статистичне осереднення проводиться по всьому набору можливих реалізацій поля.

Повний статистичний опис поведінки випадкового поля в різних точках потребує використання багатовимірних функцій розподілу і є дуже складним. У багатьох випадках можна обмежитися більш простими характеристиками статистичної структури випадкового поля. Найчастіше використовуються кореляційні, структурні та спектральні функції для опису статистичних зв'язків [6] в основних оболонках нашої планети.

Найбільш простою характеристикою статистичного зв'язку є *коваріаційна функція* величини X , яка для будь-якої пари точок s_i та s_k визначається за формулою:

$$K_x(s_i, s_k) = \overline{[x(s_i) - \bar{x}(s_i)][x(s_k) - \bar{x}(s_k)]}. \quad (2.1)$$

Кореляційна функція

$$R_x(s_i, s_k) = \frac{K_x(s_i, s_k)}{\sigma_x(s_i)\sigma_x(s_k)}, \quad (2.2)$$

яка для кожної пари точок s_i та s_k дає коефіцієнт кореляції значень величини X .

Якщо розглядається зв'язок між значеннями однієї і тієї ж величини, замість термінів «коваріаційна та кореляційна функції», інколи використовуються терміни «автоковаріаційна та автокореляційна функції» [14].

З визначення функцій $K_x(s_i, s_k)$ та $R_x(s_i, s_k)$ випливають їх властивості:

1) Властивість симетричності:

$$K_x(s_i, s_k) = K_x(s_k, s_i); \quad (2.3)$$

$$R_x(s_i, s_k) = R_x(s_k, s_i). \quad (2.4)$$

2) Властивість обмеженості:

$$K_x(s_i, s_k) \leq \sigma_x(s_i)\sigma_x(s_k); \quad (2.5)$$

$$K_{xg}(s_i, s_k) \leq \sigma_x(s_i)\sigma_g(s_k); \quad (2.6)$$

$$|R_x(s_i, s_k)| \leq 1; \quad (2.7)$$

$$|R_{xg}(s_i, s_k)| \leq 1. \quad (2.8)$$

В граничному випадку, коли аргументи функцій співпадають, мають місце рівності:

$$K_x(s_i, s_i) \leq \sigma_x^2(s_i); \quad (2.9)$$

$$R_x(s_i, s_i) = 1. \quad (2.10)$$

Для опису статистичної структури випадкового поля можна використовувати **структурну функцію**, що представляє собою для двох заданих точок середній квадрат різниці значень x :

$$B_x(s_i, s_k) = \overline{[x(s_i) - x(s_k)]^2}. \quad (2.11)$$

Структурна функція зв'язана з коваріаційною та кореляційною функціями такою формулою:

$$\begin{aligned} B_x(s_i, s_k) &= \sigma_x^2(s_i) + \sigma_x^2(s_k) - 2K_x(s_i, s_k) + [\overline{x(s_i)} - \overline{x(s_k)}]^2 = \\ &= 2\sigma_x(s_i)\sigma_x(s_k)[1 - R_x(s_i, s_k)] + [\sigma_x(s_i) - \sigma_x(s_k)]^2 + [\overline{x(s_i)} - \overline{x(s_k)}]^2. \end{aligned} \quad (2.12)$$

Із визначення структурної функції виходить її невід'ємність.

Крім того:

$$B_x(s_i, s_k) = B_x(s_k, s_i); \quad (2.13)$$

$$R_x(s_i, s_k) = 0; \quad (2.14)$$

$$B_x(s_i, s_k) \leq [\sigma_x(s_i) + \sigma_x(s_k)]^2 + [\overline{x(s_i)} - \overline{x(s_k)}]^2. \quad (2.15)$$

Розглянуті вище функції будуть використані у подальшому при розв'язанні спеціальних задач.

2.1.2 Напрями використання кореляційного аналізу в гідрометеорології

У сучасних умовах практичною основою вивчення клімату того чи іншого регіону та його динаміки є база емпіричних даних, яка включає не тільки ряди спостережень, але й вибір статистик та методів, за допомогою яких визначаються просторово-часові масштаби метеорологічних полів, їх структура та мінливість. Статистики, що отримані в результаті застосування методів багатовимірного аналізу, поряд з традиційною кліматичною інформацією набувають в теперішній час широкого застосування при плануванні та оптимізації мережі станцій, на яких виконуються спостереження, в прогностичній діяльності, кліматичному моделюванні, розв'язанні задач моніторингу клімату і т.п.

У природі досить часто виникають ситуації, коли між деякими випадковими процесами проявляються статистичні зв'язки.

Дослідник має всі підстави розглядати гідрометеорологічні величини як випадкові величини [13], незважаючи на те, що їх зміни зумовлені певними фізичними закономірностями. Така їх властивість пов'язана з тим, що зв'язки між випадковими величинами спричиняються взаємодією багатьох факторів, і у нас немає, як правило, достатніх знань про те, який саме зв'язок або зв'язки, зумовили змінювання тієї чи іншої величини у термін її вимірювання. Більше того, у деяких випадках ще є відсутніми

повні уявлення про причинно-наслідкові залежності у природних явищах, що спричиняють гідрометеорологічні процеси. Дослідити такі складні статистичні зв'язки допомагає кореляційний аналіз.

Просторові кореляційні функції, які представлені у вигляді ізокорелят, або осереднені за різними напрямками чи районами, дозволяють більш коректно оцінити локальні особливості мінливості клімату, а також є необхідною основою для оптимального осереднення метеорологічних полів. Як відомо, найбільш загальне представлення про мінливість клімату того чи іншого регіону дають карти дисперсій або середніх квадратичних відхилів основних метеорологічних полів. Крім того, перехід кореляційних функцій через нуль характеризує масштаб гідрометеорологічних полів, а їх орієнтування за різними румбами дозволяє визначати сезонні зміни в цих полях [6].

Перед сучасною кліматологією стоїть цілий ряд актуальних проблем, однією з яких є наддовгостроковий («кліматичний») прогноз. Необхідність в ньому зростає у зв'язку з зростанням впливу людства на навколишнє середовище. Важливим етапом в розв'язанні цієї проблеми є вивчення причин, які зумовлюють коливання клімату, а в цьому допомагають знання та правильне застосування методів багатовимірної статистичної аналізу.

В останні роки значна увага приділяється перспективам використання статистично значущих кореляційних зв'язків між випадковими процесами районів, значно віддалених один від одного (так званих «далеких зв'язків») стосовно задач, перш за все, довгострокового та наддовгострокового прогнозів по осередках додатних і від'ємних значень кореляції.

Само поняття «наддовгостроковий прогноз» означає вихід за межі внутрішньорічної завчасності і використання з цією метою в якості предикторів показників, які мають більш тривалу, ніж річну, часову мінливість. У зв'язку з цим необхідною умовою при розробці наддовгострокового прогнозу є вивчення часової структури коливань прогнозованих характеристик і факторів, що формують їх багаторічну мінливість, а це можливо дослідити тільки за допомогою методів багатовимірної статистичної аналізу і кореляційного аналізу в тому числі [17].

Детермінований підхід до повного розв'язання задачі наддовгострокового прогнозу геофізичних процесів в даний час ще неможливий із-за надзвичайної складності взаємозв'язку природних процесів та його недостатньої вивченості. Тому при розв'язанні саме таких задач широко застосовується фізико-статистичний, а точніше, статистично-фізичний підхід. Сенс його полягає у наступному: реалізуються кореляційний та поряд з ним інші багатовимірні аналізи просторово-часових рядів геофізичних характеристик, по можливості з'ясовується фізичний сенс виявлених зв'язків та циклів, які потім і використовуються для цілей прогнозу [6].

Наведемо приклад використання кореляційного аналізу для дослідження статистичної структури вертикальних профілів температури повітря у вільній атмосфері [11, 12].

Як відомо, своєрідність атмосферних процесів, що розвиваються над різними районами Земної кулі, впровадження в оперативну практику регіональних та глобальних чисельних моделей обумовлюють необхідність мати детальні відомості про просторову та часову структуру полів основних метеорологічних величин за всіма широтними зонами. Розвиток ракетних та супутникових методів виміру параметрів атмосфери сприяють постійному уточненню за емпіричними даними будови та циркуляції повітряної оболонки Землі на висотах більше 25-30 км.

Відомо, що особливості атмосферних процесів проявляються в статистичних характеристиках, зокрема в значеннях та знаках коефіцієнтів кореляції. Наприклад, розігрів стратосфери в період зимових потеплінь супроводжується ослабленням тропосферно-стратосферних зв'язків. При східному типі циркуляції мають місце великі від'ємні зв'язки між зміненнями температури на рівнях середньої стратосфери та більш високих рівнях, тобто розігріву більш низьких шарів стратосфери відповідає охолодження більш високих шарів. При західному типі циркуляції кореляційні зв'язки між зміненнями температури слабнуть зі збільшенням відстані між поверхнями, що порівнюються.

В якості статистичних характеристик, що описують вертикальну структуру полів температури в густих шарах атмосфери, розраховувалися середні значення, середні квадратичні відхили та автокореляційні функції по часових рядах, що були сформовані за холодний період на стандартних рівнях 5, 10, 15, 20, 25, 30, 35, 40, 45 та 50 км. Така дискретність з висотою пов'язана з необхідністю збереження міжрівневних коефіцієнтів кореляції статистично значущими [12].

Нижче наводяться результати кореляційного аналізу, застосованого до різних районів Західної півкулі, що включають приекваторіальні, субтропічні, помірні та високі широти, відповідно пункти 1, 2, 3, 4 (табл.2.1).

Таблиця 2.1 – Перелік пунктів ракетного і радіозондування атмосфери, дані яких використовувались у дослідженні

Пункт	Координати		Об'єм вибірки
	широта північна	довгота західна	
1	09°20'	80°00'	117
2	28°27'	80°32'	183
3	37°50'	75°29'	127
4	58°44'	93°49'	166

Як випливає з рис.2.1, кореляційні функції температури повітря над районом дослідження мають ряд характерних рис.

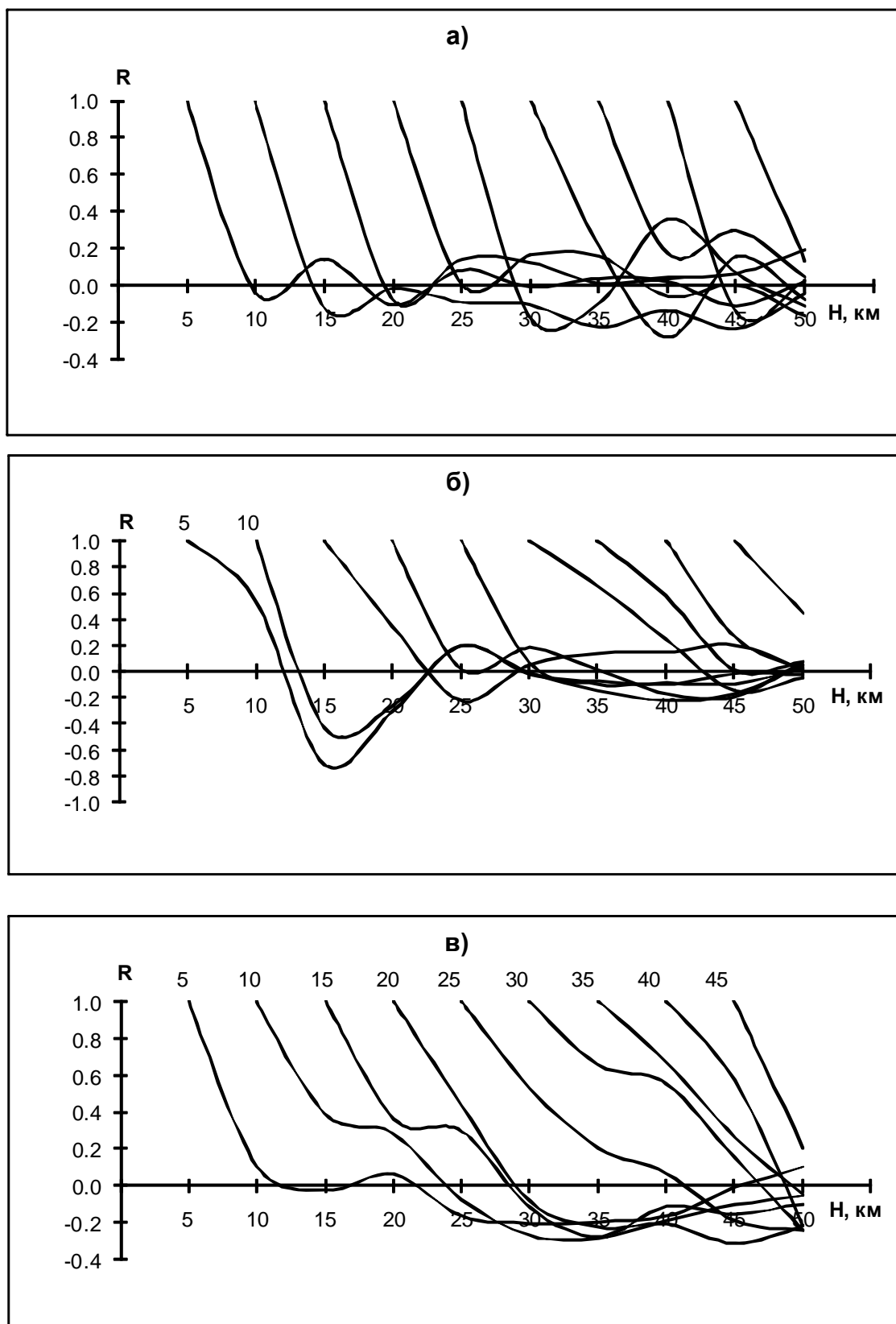


Рис.2.1 – Автокореляційні функції температури повітря в приекваторіальних (а), помірних (б) та високих широтах (в)

У приєкваторіальній зоні (рис. 2.1а) кореляційний зв'язок швидко падає з висотою. Кореляційні функції температури повітря середніх широт (рис. 2.1б) характеризуються тим, що з висотою їх значення також швидко зменшуються, досягаючи нуля на висотах 10-12 км. Вище цього рівня зв'язок стає зворотним і максимум його спостерігається на висоті 15 км. Пояснюється це тим, що зростання температури повітря на висоті 15 км найчастіше супроводжується її падінням на нижчих рівнях. У холодний період верхні стратосферні функції та функції з вихідними рівнями кореляції 15-20 км свідчать про достатньо швидке зменшення кореляційного зв'язку з висотою. У високих широтах (рис. 2.1в) кореляційні функції температури, що відповідають вихідним рівням 10-15 км також швидко зменшуються з висотою, проте зворотний кореляційний зв'язок, максимум якого в середніх широтах припадає на 15 км, є слабким.

Аналіз індивідуальних профілів температури повітря дає підстави зробити висновок про те, що теплій верхній стратосфері (приблизно 45 км) у ряді випадків відповідає більш холодна нижня стратосфера. Цей факт добре підтверджується коефіцієнтами вертикальної автокореляції температури в інтервалі висот 45-50 км. Від'ємний кореляційний зв'язок означає, що процесам, які вихолоджують один шар атмосфери, відповідають процеси, що нагрівають інший шар, і навпаки.

Наведемо ще один приклад використання кореляційного аналізу для дослідження «далеких зв'язків» в Атлантико-Європейському секторі [11].

Той факт, що Північна Атлантика в холодну пору року є головним джерелом тепла, під дією якого знаходяться процеси над великими районами Європи та Азії, був зрозумілим ще в середині XIX століття, але детальне вивчення процесів в Атлантико-Європейському секторі розпочалося тільки у XX сторіччі.

В комплексній схемі прогнозу для Європи необхідно враховувати інформацію про характер макропроцесів у районі Північної Атлантики (регіональні предиктори) та про атмосферну циркуляцію інших районів Північної півкулі. Для розв'язання такої задачі можна використати кореляційний аналіз для побудови карт-схем «далеких зв'язків» («teleconnection») в полях основних кліматичних характеристик.

Гончарова Л.Д. дослідила «далекі зв'язки» (схеми телеконекцій) в полях температури повітря, геопотенціалу та складових швидкості вітру в вільній атмосфері на важливих для ДПП ізобаричних поверхнях [11]. Враховуючи процеси в Атлантико-Європейському секторі, було виділено регіон, розташований між меридіанами 60° зах.д. та 60° сх.д. позатропічних широт Північної півкулі (30°-70°).

Як відомо, великомасштабні зміни кліматичного режиму над великими територіями мають квазіперіодичний характер і зараз висловлюються різні думки щодо причин таких змін. Однією з гіпотез є

існування автоколивальної системи безпосередньо в земних умовах.

Для перевірки ряду гіпотез, що були висунуті Гончаровою Л.Д. [11] для раціонального пояснення та врахування в схемах ДПП, був застосований кореляційний аналіз, який дозволив отримати надійне якісне представлення напрямків взаємодій у зимовій атмосфері. Для опису цих напрямків використовувалися парні коефіцієнти кореляції між однойменними параметрами атмосфери та між усіма можливими парами вузлів регулярної сітки точок ($2^\circ \times 2^\circ$) на ізобаричних поверхнях 700 та 500 гПа.

На основі вихідної інформації, а саме середньої місячної температури повітря, геопотенціалу, зональної та меридіональної складових швидкості вітру отримані матриці кореляції 65-го порядку (за кількістю вузлів регулярної сітки). Було проаналізовано більше 2000 коефіцієнтів кореляції на кожному стандартному рівні тропосфери для окремого параметра атмосферної циркуляції. Цей підхід дав можливість отримати статистичні схеми «далеких зв'язків», які більш детально описані у монографії [11].

Відомо, що від'ємні кореляції вказують на характерні телеконекції в визначених параметрах.

Перевірка статистичних гіпотез про значущість отриманих парних коефіцієнтів кореляції проводилася за відомою методикою [13] на рівні значущості $\alpha=0,05$.

Додатні коефіцієнти кореляції аналізувалися для ситуацій, які потребують опису районів з типовими (однорідними) погодними умовами.

Результати кореляційного аналізу дали змогу визначити райони, в яких розташовані центри диполів на атлантико-європейській території, що вказує на наявність статистично обґрунтованих автоколивальних систем у параметрах загальної циркуляції атмосфери у нижніх шарах атмосфери.

Отримані схеми телеконекцій, які побудовані в полях складових швидкості вітру, геопотенціалу та температури повітря на основних поверхнях тропосфери, дозволять врахувати напрямки переносу основних субстанцій, а це в свою чергу допоможе краще зрозуміти вклад різних районів Північної півкулі в енергетичний баланс системи океан-атмосфера.

2.1.3 Однорідність та ізотропність випадкових полів

Крім наведених вище прикладів, за допомогою кореляційного аналізу є можливість провести дослідження властивостей однорідності та ізотропності гідрометеорологічних полів, використовуючи матриці коваріацій та кореляцій.

Властивості однорідності та ізотропності випадкових процесів значно спрощують аналіз просторових полів та дозволяють без значних зусиль аналізувати їх структуру [6].

Поле називають однорідним, для якого математичне сподівання є сталим, а коваріаційна функція залежить тільки від різниці векторів: відстаней точок поля від полюса кореляції. Однорідні поля **називають ізотропними**, якщо просторова кореляційна функція $K(l)$ не залежить від напрямків перетинів полів кореляції [6].

Для однорідних полів $\bar{x}_i = \text{const}$, $\sigma_i^2 = \text{const}$, $\forall_i = \overline{1, n}$, а k_{is} залежить тільки від відстані між i -ю та s -ю точками поля.

Однорідне випадкове поле X_j , що визначене на n -вимірному евклідовому просторі R^n , називається **ізотропним**, якщо його коваріаційна матрична функція $K(l)$ залежить тільки від норми вектора $|l|$ в R^n (n - позначає кількість точок, у яких визначається поле). Властивість ізотропності потребує ще й незалежності просторової кореляційної функції від напрямків перетинів полів кореляцій [27].

Метеорологічні поля, як правило, не мають властивостей однорідності та ізотропності. Завдяки тому, що їх структура залежить від циркуляційних факторів і місцевих умов, які суттєво розрізняються в різних районах території, яку охоплює метеорологічне поле, а середні значення й дисперсії метеорологічних величин у різних точках поля мають різні значення. Але досить часто можна використати просте лінійне перетворення вихідних випадкових величин x_{ij} і перейти до інших випадкових величин φ_{ij} , для яких вимоги однорідності та ізотропності виконуються, або майже виконуються [27]. Поля таких випадкових величин називають **квазіоднорідними та квазіізотропними**.

Зазначене **лінійне перетворення** має такий вигляд:

$$x_{ij} = \bar{x}_i + \sigma_i \varphi_{ij}, \quad (2.16)$$

з якого визначимо

$$\varphi_{ij} = \frac{x_{ij} - \bar{x}_i}{\sigma_i}. \quad (2.17)$$

Величини φ_{ij} є центровані та нормовані.

Розглянемо властивості таких випадкових величин. Для цього знайдемо середнє значення $\overline{\varphi_{ij}}$ та дисперсію σ_{φ}^2 .

Очевидно, що

$$\overline{\varphi_{ij}} = \frac{1}{\sigma_i} \overline{\Delta x_{ij}} = \frac{1}{\sigma_i} (\overline{x_{ij} - \bar{x}_i}) = \frac{1}{\sigma_i} (\bar{x}_i - \bar{x}_i) = 0; \quad (2.18)$$

$$\sigma_\varphi^2 = \frac{\sum_{j=1}^m \varphi_{ij}^2}{m} = \frac{\sum_{j=1}^m \frac{\Delta x_{ij}^2}{\sigma_i^2}}{m} = \frac{1}{\sigma_i^2} \frac{\sum_{j=1}^m (x_{ij} - \bar{x}_i)^2}{m} = \frac{\sigma_i^2}{\sigma_i^2} = 1 = const. \quad (2.19)$$

Рівняння (2.18) та (2.19) свідчать про те, що середнє значення та дисперсія випадкової величини φ у всіх точках поля однакові.

Матриця $\varphi = \{\varphi_{ij}\}$ перетворюється з матриці ΔX за допомогою матричної рівності

$$\varphi = \sigma^{-1} \Delta X. \quad (2.20)$$

Знайдемо матрицю коваріацій випадкової величини φ за допомогою матричної рівності (1.8):

$$K_\varphi = \frac{1}{m} \varphi \varphi' \quad (2.21)$$

та враховуючи формулу (2.20), будемо мати:

$$\begin{aligned} K_\varphi &= \frac{1}{m} \sigma^{-1} \Delta X (\sigma^{-1} \Delta X)' = \frac{1}{m} \sigma^{-1} \Delta X \Delta X' \sigma^{-1} = \\ &= \sigma^{-1} \left(\frac{1}{m} \Delta X \Delta X' \right) \sigma^{-1} = \sigma^{-1} K_x \sigma^{-1} = R_x \\ K_\varphi &= R_x. \end{aligned} \quad (2.22)$$

Таким чином, *відповідь на запитання, чи виконуються **властивості однорідності та ізотропності** для випадкових полів φ , залежить від структури матриці кореляцій R_x* . Структура цієї матриці така, що кожен її рядок, який має номер i (чи стовпець із тим же номером),

представляє собою поле кореляцій, причому номер елемента цього рядка (цього стовпця), розташованого на головній діагоналі матриці R_x , визначає номер вузла сітки точок (номер метеорологічної станції), котрий є *центром кореляції*. Очевидно, що кореляційний момент у цьому вузлі дорівнює одиниці. Таким чином, кореляційна матриця порядку n вміщує у собі інформацію про n полів кореляцій.

Щоб зробити висновок відносно однорідності та ізотропності полів φ , досить упевнитися у тому, що поля кореляцій для всіх полюсів кореляцій мають подібний вигляд і визначаються системою майже концентричних ізокорелят. Властивість ізотропності визначається за виглядом просторових кореляційних функцій та їх розташування відносно довірчого коридору, який будується на заданому рівні значущості [13].

2.2 Розв’язання типової задачі. Дослідження однорідності та ізотропності метеорологічних полів

Дослідження властивостей однорідності та ізотропності випадкових полів проводилось на основі часових рядів місячної кількості опадів у лютому за період з 1975 по 2005 рр., отриманих на мережі 33 метеорологічних станцій України (Додаток Б).

В результаті реалізації алгоритму кореляційного аналізу за допомогою комп’ютерної програми «KOR.EXE» (Додаток В) було отримано:

- 33-вимірний вектор середніх значень місячної кількості опадів;
- 33-вимірний вектор середніх квадратичних відхилів;
- матрицю кореляцій 33-го порядку.

Для двох полюсів кореляцій (ст. Одеса та ст. Київ) на основі отриманої матриці кореляції були побудовані поля ізокорелят, які представлені на рис. 2.2 та рис. 2.3.

Як видно з цих рисунків, поля кореляцій для вказаних полюсів мають подібний вигляд, тобто зміна центра кореляції не призводить до зміни конфігурації полів і вони визначаються системою майже концентричних ізокорелят.

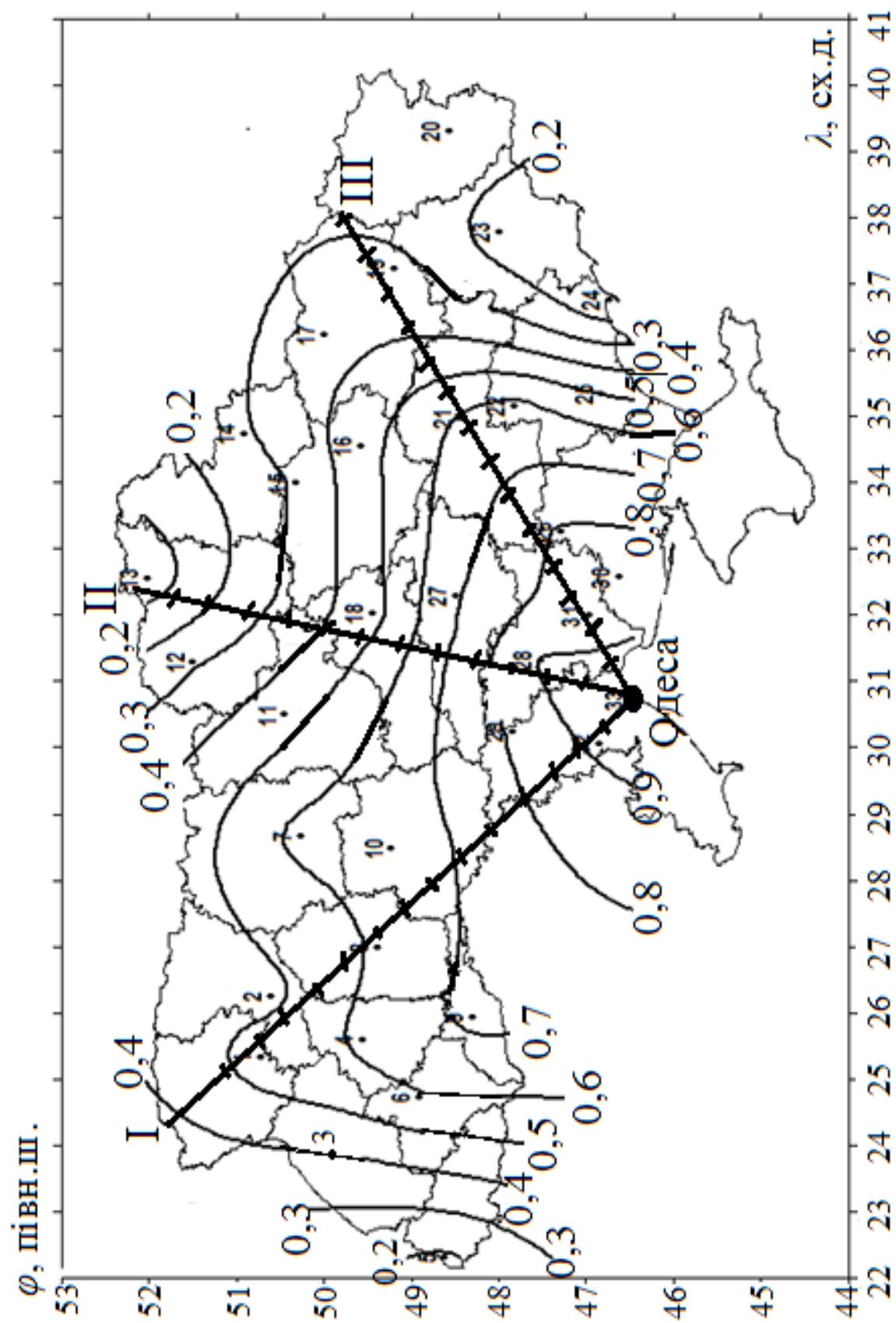


Рис. 2.2 – Поле кореляцій місячної кількості опадів у лютому (з полюсом кореляції на ст. Одеса)

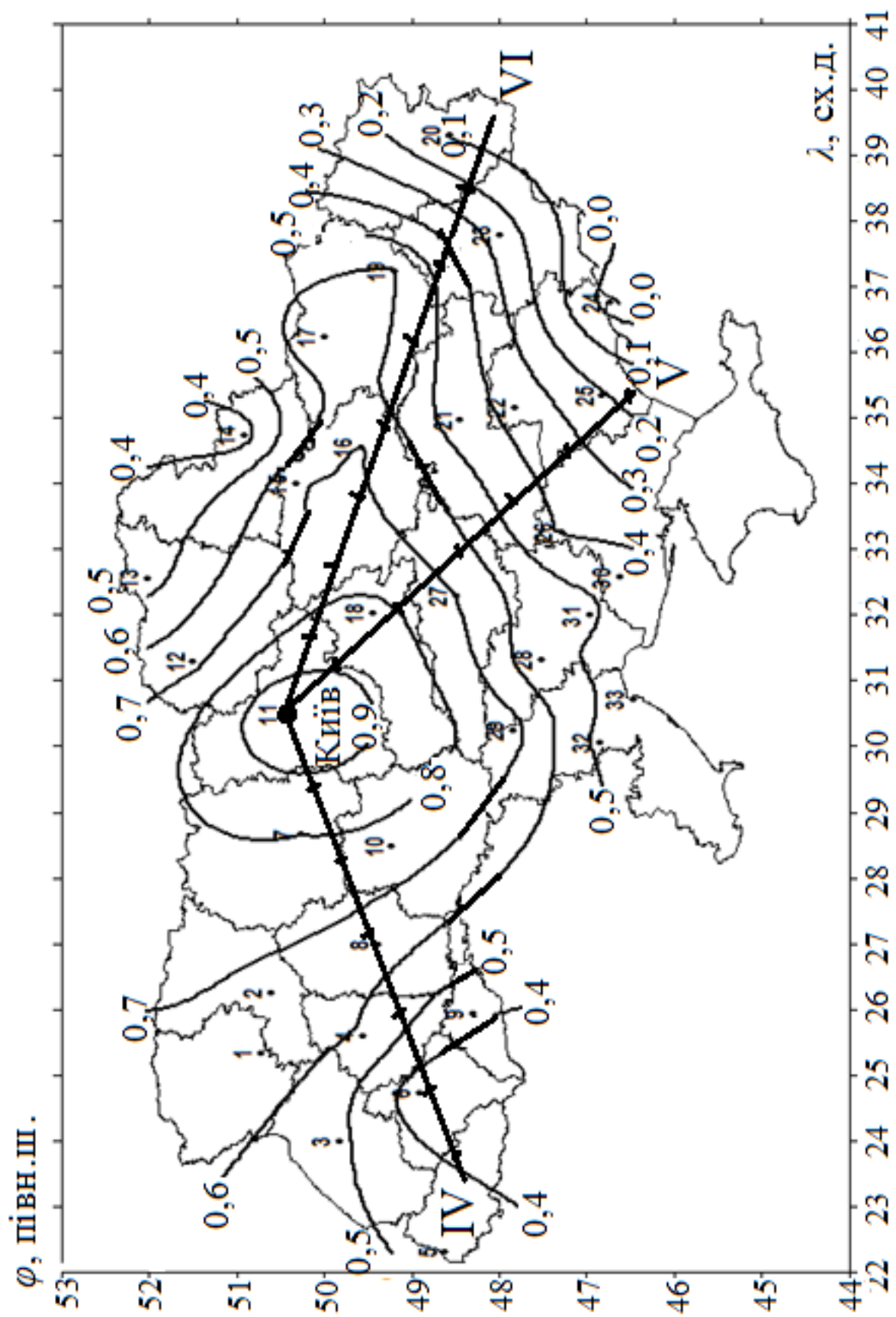


Рис. 2.3— Поле кореляцій місячної кількості опадів у лютому (з полюсом кореляції на ст. Київ)

Це дає підстави стверджувати, що поле опадів у лютому на території України є квазіоднорідним, оскільки для центрованих і нормованих на середній квадратичний відхил значень місячної кількості опадів, які визначаються рівністю (2.17), середні значення дорівнюють нулю, а дисперсії – одиниці, для всіх точок поля.

Для подальшого дослідження квазіоднорідного поля на ізотропність були зроблені перетини полів кореляцій у трьох довільних напрямках (рис. 2.2 та рис. 2.3). З кожного із шести напрямків перетину полів кореляції через 1 см від полюса кореляції були зняті значення коефіцієнтів кореляцій. Результати цієї процедури представлені в табл. 2.2. На їх основі були побудовані сім просторових кореляційних функцій місячної кількості опадів, які представлені на рис. 2.4.

Таблиця 2.2 – Значення коефіцієнтів кореляції місячної кількості опадів для різних напрямків перетину (відповідають рис 2.2 та рис. 2.3)

№ кроку	Напрямок перетину						$\overline{r_{is}}$
	I	II	III	IV	V	VI	
1	0,90	0,91	0,88	0,88	0,89	0,84	0,89
2	0,85	0,84	0,84	0,82	0,85	0,80	0,83
3	0,80	0,78	0,81	0,77	0,80	0,77	0,79
4	0,75	0,68	0,76	0,71	0,72	0,74	0,73
5	0,71	0,58	0,72	0,66	0,64	0,71	0,67
6	0,67	0,49	0,67	0,59	0,55	0,66	0,60
7	0,63	0,41	0,62	0,53	0,46	0,61	0,54
8	0,60	0,34	0,54	0,45	0,39	0,57	0,48
9	0,57	0,28	0,46	0,37	0,32	0,54	0,43
10	0,53	0,20	0,41	0,37	0,25	0,49	0,37
11	0,49	0,13	0,36	0,38	0,18	0,45	0,38

Статистична структура полів опадів буде характеризуватися просторовою кореляційною функцією $\overline{r_{is}}$, яка визначається шляхом осереднення кореляційних функцій шести перетинів.

Аналіз рис. 2.4 вказує на те, що всі кореляційні функції розташуються у достатньо вузький смугі, а це є підставою вважати, що структура полів кореляцій не залежить від напрямку їх перетину.

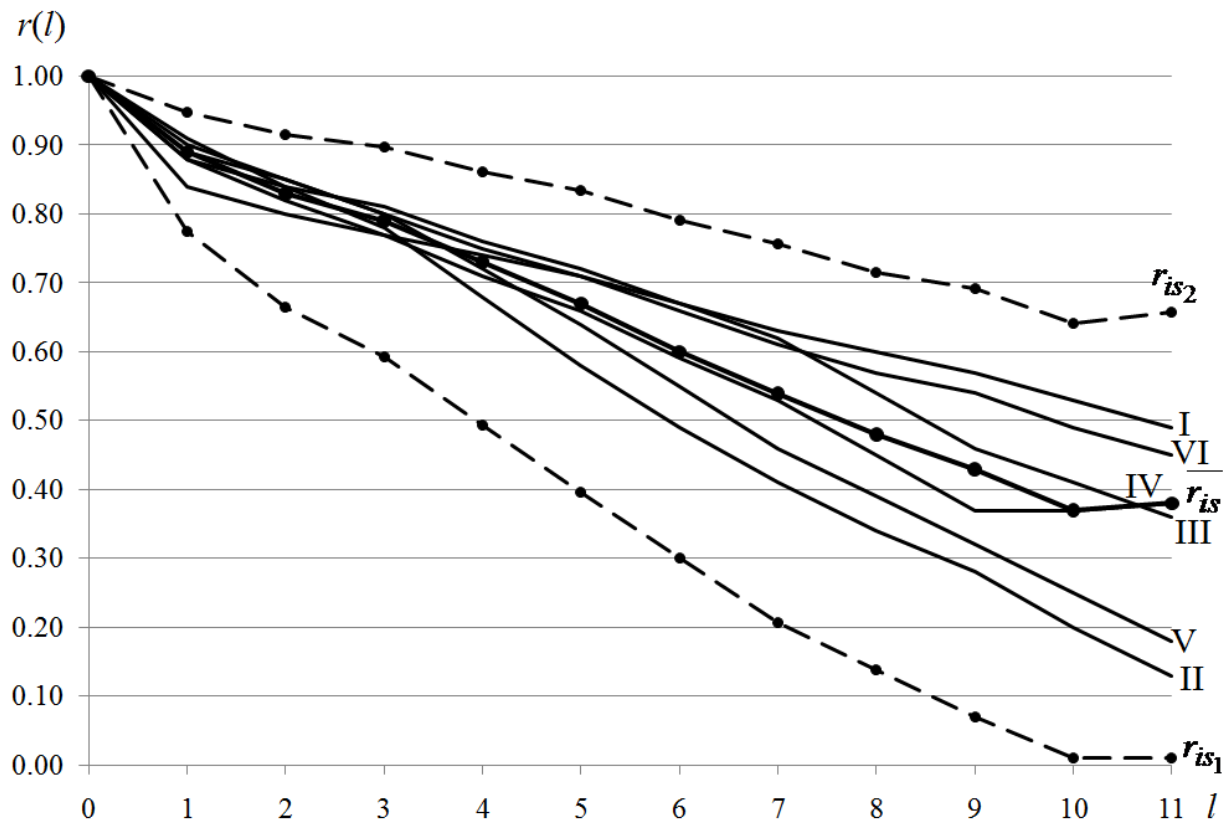


Рис. 2.4 – Просторові кореляційні функції місячної кількості опадів у лютому з полюсами кореляцій на метеорологічних станціях Одеса та Київ

Щоб переконатися у вірогідності отриманих результатів та висновку щодо ізотропності поля, треба на заданому рівні значущості ($\alpha = 0,05$) побудувати довірчі інтервали для коефіцієнтів кореляції.

Оскільки об'єми вибірок невеликі ($m < 50$), при побудові довірчих інтервалів для коефіцієнтів кореляцій r_{is} окремого кроку було застосовано Z -перетворення Фішера [13].

Нелінійне логарифмічне Z -перетворення Фішера дорівнює:

$$\hat{z} = \frac{1}{2} \ln \frac{1 + r_{is}}{1 - r_{is}}.$$

Спочатку будуються довірчі інтервали для Z -перетворення:

$$\hat{z} - t_{kp}(\alpha, \nu) \sigma_z < z < \hat{z} + t_{kp}(\alpha, \nu) \sigma_z, \quad \text{тобто} \quad z_1 < z < z_2,$$

де

$$z_1 = \hat{z} - t_{kp}(\alpha, \nu) \sigma_z;$$

$$z_2 = \hat{z} + t_{kp}(\alpha, \nu) \sigma_z;$$

$$\sigma_z = \frac{1}{\sqrt{n-3}} = \frac{1}{\sqrt{30-3}} = 0,19.$$

Критичне значення критерію Стюдента за умов $\alpha = 0,05$ та $\nu = m = 30$ знаходимо з Додатку Г:

$$t_{kp}(\alpha, \nu) = t_{kp}(0,05; 30) = 2,04.$$

Довірчий інтервал для коефіцієнта кореляції:

$$r_{is1} < \rho_{is} < r_{is2},$$

де

$$r_{is1} = \frac{e^{2z_1} - 1}{e^{2z_1} + 1}, \quad r_{is2} = \frac{e^{2z_2} - 1}{e^{2z_2} + 1}.$$

Результати всіх розрахунків для побудови довірчих інтервалів $[z]$ і $[\rho_{is}]$ представлені в табл. 2.3.

Таблиця 2.3 – Довірчі інтервали для Z-перетворення Фішера та коефіцієнтів кореляції

№ кроку	$\overline{r_{is}}$	\hat{z}	z_1	z_2	r_{is1}	r_{is2}
1	0,89	1,422	1,03	1,81	0,774	0,948
2	0,83	1,188	0,80	1,56	0,664	0,915
3	0,79	1,071	0,68	1,46	0,592	0,898
4	0,73	0,929	0,54	1,30	0,493	0,862
5	0,67	0,811	0,42	1,20	0,397	0,834
6	0,60	0,693	0,31	1,07	0,300	0,790
7	0,54	0,604	0,21	0,99	0,207	0,757
8	0,48	0,523	0,14	0,90	0,139	0,716
9	0,43	0,460	0,07	0,85	0,070	0,691
10	0,37	0,388	0,01	0,76	0,010	0,641
11	0,38	0,40	0,01	0,79	0,010	0,658

Аналіз рис. 2.4 вказує на те, що всі кореляційні функції (незалежно від напрямку перетину полів кореляцій) розташовуються у достатньо вузькій смузі та вкладаються в отриманий довірчий інтервал, а це є підставою вважати, що структура полів не залежить від напрямку перетину і поля місячної кількості опадів володіють не тільки властивістю квазіоднорідності, а й властивістю квазіізотропності.

Висновок: поле місячної кількості опадів у лютому на території України з ймовірністю 95% є квазіоднорідним та квазіізотропним.

2.3 Запитання для самоперевірки

1. Які задачі можливо вирішити в гідрометеорології, використовуючи кореляційний аналіз при дослідженні випадкових полів?
2. Які зв'язки в геофізичних процесах називають «статистично значущими»?
3. Яка інформація міститься в середньому n -вимірному полі?
4. Яка інформація міститься на картах дисперсій (середніх квадратичних відхилів) n -вимірних метеорологічних полів?
5. Яка інформація міститься в матриці коваріації?
6. Яка інформація міститься в матриці кореляції?
7. Як отримати матрицю коваріації? Як представити матрицю коваріації у координатній формі? Які властивості має ця матриця?
8. Як отримати матрицю кореляції? Як представити матрицю кореляції у координатній формі? Які властивості має ця матриця?
9. Яке випадкове поле називають однорідним?
10. Яке випадкове поле називають ізотропним?
11. Яке лінійне перетворення можна використати, щоб отримати випадкові величини, для яких вимоги однорідності та ізотропності майже виконуються?
12. За яких умов випадкове поле буде однорідним та ізотропним? Які при цьому повинні бути поля ізокорелят та просторові кореляційні функції?
13. Як за допомогою кореляційного аналізу виконують дослідження метеорологічних полів на однорідність та ізотропність?

2.4 Завдання для закріплення теми

- Провести дослідження однорідності та ізотропності випадкового поля на основі вихідних рядів місячної кількості опадів (Додаток Б) для мережі з 33 метеорологічних станцій України (Додаток А). Для цього:
 - отримати у викладача варіант задачі для виконання практичної роботи та комп'ютерну програму «KOR.EXE» (призначення програми та особливості користування дивись у Додатку В);
 - виконати на ПК необхідні розрахунки та отримати середні значення, середні квадратичні відхили та матрицю кореляцій місячної кількості опадів;
 - побудувати на карті n -вимірне середнє поле опадів;
 - побудувати на карті поле середніх квадратичних відхилів випадкової величини, що досліджується;
 - побудувати поля кореляцій для двох центральних станцій, які визначаються викладачем;
 - зробити у різних напрямках перетини отриманих полів кореляцій та побудувати просторові кореляційні функції;
 - для осередненої кореляційної функції побудувати довірчі інтервали. Рівень значущості α взяти 0,05;
 - провести аналіз результатів дослідження статистичної структури полів опадів та зробити висновок щодо роботи у цілому.

3 КОМПОНЕНТНИЙ АНАЛІЗ МЕТЕОРОЛОГІЧНИХ ПОЛІВ

3.1 Базові теоретичні положення

3.1.1 *Напрями використання компонентного аналізу в гідрометеорології*

Застосування статистичних методів аналізу й прогнозу метеорологічної ситуації та інших численних задач, що мають на меті необхідність врахування стану атмосфери чи об'єктів гідросфери, пов'язано з рядом труднощів. Це не підпорядкування у багатьох випадках випадкових величин нормальному закону розподілу, неоднорідність та неізотропність метеорологічних полів, нестаціонарність процесів, що характеризують зміну метеорологічних величин за часом і у просторі та т.п. Існує ще одне суттєве обмеження, яке пов'язане з розробкою статистичних моделей різного призначення. Йдеться про велику розосередженість вихідної інформації, яка необхідна для прийняття висновку про майбутній стан атмосферних процесів. Здається, що ці труднощі можливо обійти шляхом збільшення кількості предикторів (факторів, що впливають на предиктант) й розширення об'ємів їх статистичних сукупностей. Але чим більше предикторів у моделі, тим більшою стає розмірність матриці коваріації (кореляції) і тим гіршою буде її обумовленість. Це пояснюється тим, що збільшення в статистичній моделі кількості предикторів приводить до зростання тісноти кореляційних зв'язків між ними, що ускладнює побудову прогностичних моделей [27].

В алгоритмах обчислення коефіцієнтів рівняння моделі, якою є, наприклад, рівняння регресії або параметрів дискримінантної функції, коли модель ґрунтується на теорії розпізнавання образів [20, 23], обов'язковою є процедура «*обернення*» коваріаційної матриці. Якщо вона погано обумовлена (це трапляється за умови наближення визначника матриці до нуля), обернення матриці приводить до 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 в тому, що спочатку проводиться *параметризація* предикторів, 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 та II роду або система експоненціальних поліномів. Однак, розклади метеорологічних полів за цими поліномами є у великій мірі формальними в тому сенсі, що коефіцієнти розкладів не несуть смислової інформації. У назві «*природні ортогональні функції*» підкреслюється той факт, що вони отримані на основі вихідних метеорологічних полів та відбивають основні властивості їх статистичної структури [27].

Завдяки некорельованості нових змінних спрощується фізичний аналіз статистичних моделей, не виникає обчислювальних труднощів при розв'язаннях систем нормальних рівнянь, необхідних для визначення параметрів моделей, а за необхідності можна повернутися до вихідних змінних. Окрім перелічених переваг, нові параметри, 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 [27].

Компонентний аналіз з успіхом використовується й при розв'язанні інших гідрометеорологічних задач. Однією з них є задача ***стиску вихідної інформації***. ***Стиском інформації*** називають суттєве скорочення кількості вихідної інформації при збереженні основного її змісту. Це є надзвичайно важливою обставиною в задачі збереження інформації та її безпосереднього використання у практичних цілях [2, 7-9].

Іншою важливою задачею, котру дає можливість реалізувати компонентний аналіз, є ***задача фільтрації вихідної інформації***. ***Сенс її полягає у наступному***. Як відомо, поля метеорологічних величин, а також інші метеорологічні об'єкти, формуються під впливом атмосферних процесів різних масштабів: процесів макромасштабу, синоптичного масштабу, мезомасштабу та мікромасштабу. Метеорологічна інформація утримує і шумову компоненту, яка обумовлена дрібномасштабними флуктуаціями, похибками вимірювань та первинної обробки результатів спостережень.

Досить часто виникає необхідність (в залежності від характеру задачі) при дослідженні атмосферних процесів звернути увагу саме на великомасштабні і не враховувати ті, що обумовлені впливами процесів мезомасштабу та мікромасштабу. Тому, щоб вилучити з вихідної інформації дрібномасштабні збурення та похибки вимірювань, вирішують задачу фільтрації гідрометеорологічної інформації, яку можна також реалізувати за допомогою компонентного аналізу [27].

3.1.2 Власні значення та власні вектори матриці коваріації (кореляції)

Розглянемо n -вимірне поле центрованих значень метеорологічної величини X у j -й момент часу:

$$\Delta X_j = \begin{pmatrix} \Delta x_{1j} \\ \Delta x_{2j} \\ \Delta x_{3j} \\ \dots \\ \Delta x_{ij} \\ \dots \\ \Delta x_{nj} \end{pmatrix}, \quad (i = \overline{1, n}). \quad (3.1)$$

Треба здійснити *параметризацію* об'єкта, тобто представити поле, яке визначене значеннями метеорологічної величини на множині точок простору, за допомогою декількох некорельованих параметрів, які лінійно зв'язані з компонентами випадкового вектора (3.1), і які утримують основну інформацію про властивості цього поля. В основі розв'язання цієї задачі лежить лінійне ортогональне перетворення поля (3.1) в базисі власних векторів матриці коваріації (або кореляції). Нагадаємо, що система координат у n -вимірному просторі, представлена сукупністю статистично незалежних векторів, є ортогональною системою координат та носить назву *базису евклідового простору* [24].

Отже, першим етапом параметризації випадкового поля є визначення базису власних векторів, який можна отримати, розв'язавши рівняння *повної проблеми власних значень* вигляду (3.2) або (3.3):

$$R_x U_i = \lambda_i U_i, \quad (3.2)$$

де

R_x – n -вимірна матриця кореляцій;

U_i – i -й власний вектор;

Якщо використовується матриця коваріацій, то матричне рівняння цієї проблеми власних значень має вигляд

Рівняння (3.3) означає, що власний вектор U_i так перетворюється матрицею коваріацій, як він трансформується шляхом множення його на власне число λ_i . Як відомо, таке перетворення вектора є **перетворенням розтягу** за умови $\lambda_i > 1$, і – **перетворенням стиску**, якщо $\lambda_i < 1$ [27].

де

0 – нуль-вектор, тобто n -вимірний вектор, координати якого дорівнюють нулю.

Якщо записати всі складові матричного рівняння (3.4) у координатній формі та використати операцію множення, то отримаємо систему однорідних лінійних алгебраїчних рівнянь відносно невідомих координат u_{si} власного вектора U_i :

46

В рівняннях системи (3.5) $\sigma_i^2 (i = \overline{1, n})$ – дисперсії метеорологічної величини у n точках поля, $k_{si} (i, s = \overline{1, n})$ – коваріації, а $u_{si} (s = \overline{1, n})$ – s -та координата i -го власного вектора

$$U_i = \begin{pmatrix} u_{1i} \\ u_{2i} \\ \dots \\ u_{si} \\ \dots \\ u_{ni} \end{pmatrix}. \quad (3.6)$$

Тепер всі елементи матриці коваріації $K_x = \{k_{is}\} \quad (i, s = \overline{1, n})$ визначені, але система рівнянь (3.5) не може бути розв’язаною, оскільки коефіцієнти цієї системи містять власні значення λ_i матриці коваріації, які поки що є невідомими. Розрахувати невідомі власні значення $\lambda_i \quad (i = \overline{1, n})$ допомагає теорема про розв’язок системи лінійних однорідних алгебраїчних рівнянь [27]: *система однорідних лінійних алгебраїчних рівнянь має нетривіальний (не нульовий) розв’язок, якщо визначник цієї системи дорівнює нулю*. Оскільки нас цікавлять саме такі розв’язки, прирівнюємо до нуля визначник системи (3.5):

$$\begin{vmatrix} \sigma_1^2 - \lambda & k_{12} & \dots & k_{1n} \\ k_{21} & \sigma_2^2 - \lambda & \dots & k_{2n} \\ \dots & \dots & \dots & \dots \\ k_{s1} & k_{s2} & \dots & k_{sn} \\ \dots & \dots & \dots & \dots \\ k_{n1} & k_{n2} & \dots & \sigma_n^2 - \lambda \end{vmatrix} = 0. \quad (3.7)$$

Детермінант (3.7) тотожний алгебраїчному рівнянню n -го степеня відносно невідомого λ

$$\lambda^n + A_1 \lambda^{n-1} + \dots + A_{n-1} \lambda + A_n = 0. \quad (3.8)$$

Коефіцієнти A_1, A_2, \dots, A_n є відповідними комбінаціями елементів матриці коваріацій. Розв'язання рівняння (3.8) втілюється за допомогою відомих чисельних методів (наприклад, метод хорд або метод ітерацій) [27].

Після визначення власних значень λ_i ($i = \overline{1, n}$), всі коефіцієнти системи (3.5) стають відомими, що дає змогу для кожного власного значення отримати відповідний власний вектор.

В прикладній математиці розроблені чисельні методи, які значно спрощують розв'язання повної проблеми власних значень. Найбільш популярним є ітераційний метод обертань Якобі [27].

Існує теорема, яка визначає **характер власних значень**: *власні значення додатно визначеної, симетричної та дійсної матриці є дійсними, додатними й простими числами*. Оскільки матриця коваріацій (кореляцій) задовольняє умовам цієї теореми, то власні значення мають зазначені в теоремі властивості [27].

Власні значення розташовуються в порядку їх зменшення:

$$\lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_n. \quad (3.9)$$

Із системи рівнянь (3.5) випливає, що кожному власному значенню λ_i відповідає власний вектор (3.6). Тобто першому власному значенню λ_1 відповідає перший власний вектор

$$U_1 = \begin{pmatrix} u_{11} \\ u_{21} \\ u_{31} \\ \dots \\ u_{i1} \\ \dots \\ u_{n1} \end{pmatrix}, \quad (i = \overline{1, n}); \quad (3.10)$$

другому власному значенню λ_2 відповідає другий власний вектор

$$U_2 = \begin{pmatrix} u_{12} \\ u_{22} \\ u_{32} \\ \dots \\ u_{i2} \\ \dots \\ u_{n2} \end{pmatrix}, \quad (i = \overline{1, n}) \quad (3.11)$$

і т.п.

Якщо матриця коваріацій (кореляцій) добре зумовлена, то перший власний вектор складається тільки з додатних компонент, у другого – спостерігається одна зміна знака компонент, у третього – дві зміни і т.п. [27].

Власні вектори мають важливу властивість, яка визначається співвідношенням:

$$\langle U_i U_j \rangle = U_i' U_j = \delta_{ij} \|U_i\| \|U_j\|, \quad (3.12)$$

де

δ_{ij} – символ Кронекера;

$$\delta_{ij} = \begin{cases} 1, & \text{за умови } i = j; \\ 0, & \text{за умови } i \neq j. \end{cases} \quad (3.13)$$

Рівняння (3.12) означає, що скалярний добуток різних власних векторів дорівнює нулю, тобто **власні вектори ортогональні** [27].

Ця **властивість дає змогу розглядати сукупність власних векторів як базис n -вимірного евклідового простору** та шукати розкладання i -го випадкового вектора чи поля в базисі власних векторів. Матрицю U в координатній формі, що відповідає цьому базису, отримаємо, якщо розташуємо власні вектори один за одним:

$$U = \begin{pmatrix} u_{11} & u_{12} & \dots & u_{1j} & \dots & u_{1n} \\ u_{21} & u_{22} & \dots & u_{2j} & \dots & u_{2n} \\ u_{31} & u_{32} & \dots & u_{3j} & \dots & u_{3n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ u_{i1} & u_{i2} & \dots & u_{ij} & \dots & u_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ u_{n1} & u_{n2} & \dots & u_{nj} & \dots & u_{nn} \end{pmatrix}, (i, j = \overline{1, n}). \quad (3.14)$$

Очевидно

$$U'U = UU' = E. \quad (3.15)$$

Як правило, власні вектори нормуються [27]: замість векторів $U_i (i = \overline{1, n})$ використовуються **ортонормовані власні вектори**

$$W_i = \frac{U_i}{\|U_i\|}, \quad (i = \overline{1, n}). \quad (3.16)$$

Скалярний добуток ортонормованих власних векторів дорівнює

$$\langle W_i W_j \rangle = \frac{\langle U_i U_j \rangle}{\|U_i\| \|U_j\|} = \delta_{ij} \frac{\|U_i\| \|U_j\|}{\|U_i\| \|U_j\|} = \delta_{ij}. \quad (3.17)$$

Сукупність ортонормованих власних векторів складає ортогональну n -вимірну матрицю $W = \{w_{ij}\} (i, j = \overline{1, n})$, яка в координатній формі має вигляд:

$$W = \begin{pmatrix} w_{11} & w_{12} & \dots & w_{1j} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2j} & \dots & w_{2n} \\ w_{31} & w_{32} & \dots & w_{3j} & \dots & w_{3n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ w_{i1} & w_{i2} & \dots & w_{ij} & \dots & w_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ w_{n1} & w_{n2} & \dots & w_{nj} & \dots & w_{nn} \end{pmatrix}. \quad (3.18)$$

Матриця (3.18) є ортонормованою

$$W'W = WW' = E. \quad (3.19)$$

Якщо використовується ортонормований власний вектор W_i , то матричне рівняння повної проблеми власних значень може бути представлено як

$$R_x W_i = \lambda_i W_i \quad (3.20)$$

або

$$K_x W_i = \lambda_i W_i. \quad (3.21)$$

Розв'язки рівнянь (3.2) і (3.20) дають змогу отримати власні значення та власні вектори матриці кореляцій. Власні значення та власні вектори матриці коваріацій можна отримати за допомогою основних рівнянь повної проблеми власних значень вигляду (3.3) або (3.21), а далі сформувати ортогональну матрицю (3.14) чи (3.18) та розглядати їх як базис n -вимірного евклідового простору при параметризації метеорологічних об'єктів. Координати кожного власного вектора упорядковані, що дає можливість зобразити власний вектор як поле. Саме тому, як вже зазначалося, власні вектори називають **емпіричними ортогональними функціями** або **природними ортогональними функціями** [27].

3.1.3 Ортогональні компоненти випадкових метеорологічних об'єктів

Властивість ортогональності власних векторів дає можливість розглядати сукупність власних векторів як базис n -вимірного евклідового простору R^n та шукати розкладання вектора ΔX_j (чи X_j) у цьому базисі. Вектор ΔX_j являє собою той чи інший метеорологічний об'єкт (поле, вертикальний профіль, вектор предикторів і т.п.). Необхідно здійснити розклад вектора ΔX_j у деякому ортогональному базисі таким чином, щоб отримати вектор меншого розміру, в котрому би залишалися всі основні фізичні властивості вихідного вектора, тобто поля. Операція розкладання вектора ΔX_j у базисі $W = \{w_{ij}\}$ ($i, j = \overline{1, n}$) визначається таким чином [27]:

$$W' \Delta X_j = Z_j, \quad (3.22)$$

де

$$Z_j = \begin{pmatrix} z_{1j} \\ z_{2j} \\ \dots \\ z_{ij} \\ \dots \\ z_{nj} \end{pmatrix}, \quad (i = \overline{1, n}) \quad (3.23)$$

– вектор координат розкладання вихідного поля у базисі власних векторів.

Оскільки базис власних векторів є ортогональним, то компоненти z_{ij} вектора Z_j є лінійно незалежними. Вектор (3.23) називається **вектором ортогональних компонент**.

Як відомо, власні вектори ортогональні і відповідно головні компоненти поля є величинами статистично незалежними [27], тобто:

$$M[Z_i' Z_j] = \begin{cases} 0, & \text{за умови } i \neq j; \\ \sigma_{z_i}^2, & \text{за умови } i = j, \end{cases} \quad (3.24)$$

де

M – операція математичного сподівання скалярного добутку зазначених векторів;

$\sigma_{z_i}^2$ – дисперсії проекцій z_{ij} метеорологічних об'єктів.

Основою для розв'язання задачі повної проблеми власних значень (3.21) є матриця коваріацій метеорологічних полів. Як було показано в першому розділі, вона розраховується на основі множини m полів, які складають матрицю $\Delta X = \{\Delta x_{ij}\}, (i = \overline{1, n}; j = \overline{1, m})$. Залучивши операцію (3.22) до матриці ΔX отримаємо ортогональне перетворення цієї матриці [27]:

$$W' \Delta X = Z, \quad (3.25)$$

де

$Z = \{z_{ij}\}, (i = \overline{1, n}; j = \overline{1, m})$ – матриця ортогональних компонент.

Кожен рядок цієї матриці – це статистична сукупність i -ї ортогональної компоненти.

Запишемо рівняння (3.25) у координатній формі:

$$\begin{pmatrix} w_{11} & w_{21} & \dots & w_{j1} & \dots & w_{n1} \\ w_{12} & w_{22} & \dots & w_{j2} & \dots & w_{n2} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ w_{1i} & w_{2i} & \dots & w_{ji} & \dots & w_{ni} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ w_{1n} & w_{2n} & \dots & w_{jn} & \dots & w_{nn} \end{pmatrix} \begin{pmatrix} \Delta x_{1j} \\ \Delta x_{2j} \\ \dots \\ \Delta x_{ij} \\ \dots \\ \Delta x_{nj} \end{pmatrix} = \begin{pmatrix} z_{1j} \\ z_{2j} \\ \dots \\ z_{ij} \\ \dots \\ z_{nj} \end{pmatrix}. \quad (3.26)$$

Після визначення матричного добутку в лівій частині (3.26), отримаємо вектор Z , кожна компонента якого дорівнює скалярному добутку відповідного вектора-рядка матриці W на вектор-стовпець ΔX_j , тобто

$$z_{ij} = \sum_{i=1}^n w_{ji} \Delta x_{ij}, \quad (j = \overline{1, m}). \quad (3.27)$$

Із рівняння (3.27) випливає, що оскільки середнє значення $\overline{\Delta x_{ij}} = 0$ ($\forall i = \overline{1, n}; j = \overline{1, m}$), то завдяки лінійності оператора осереднення середнє значення i -ї ортогональної компоненти $\overline{z_i} = 0$, а дисперсія $\sigma_{z_i}^2$ i -ї ортогональної компоненти метеорологічного об'єкта визначається рівнянням [27]:

$$\sigma_{z_i}^2 = \frac{1}{m} \sum_{j=1}^m z_{ij}^2. \quad (3.28)$$

Врахувавши рівняння (3.27), отримаємо

$$\sigma_{z_i}^2 = \sum_{s=1}^n \sum_{p=1}^n w_{si} w_{pi} k_{sp}. \quad (3.29)$$

де

k_{sp} – елемент матриці коваріації K_x , який розташовується на перетині її s -го стовпця і p -го рядка.

Покажемо, що

$$\sum_{s=1}^n \sum_{p=1}^n w_{si} w_{pi} k_{sp} = W_i' K_x W_i. \quad (3.30)$$

Використовуючи рівняння повної проблеми власних значень вигляду (3.21), знайдемо праву частину рівності (3.30). Для цього помножимо обидві частини рівняння (3.21) ліворуч на вектор W_i' .

Будемо мати:

$$W_i' K_x W_i = \lambda_i W_i' W_i = \lambda_i. \quad (3.31)$$

Розглядаючи послідовно рівняння (3.29), (3.30) та (3.31), приходимо до висновку [2, 27], що

$$\sigma_{z_i}^2 = \lambda_i. \quad (3.32)$$

Тобто власні значення є дисперсіями відповідних ортогональних компонент розкладання метеорологічних полів у базисі власних векторів.

Згідно зі співвідношенням (3.9), маємо:

$$\sigma_{z_1}^2 > \sigma_{z_2}^2 > \dots > \sigma_{z_n}^2. \quad (3.33)$$

Крім того, власні значення мають ще одну важливу властивість [2, 27]: сума всіх власних значень матриці коваріації дорівнює сумі дисперсій вихідних метеорологічних полів

$$\sum_{i=1}^n \lambda_i = t_r K_x = \sum_{i=1}^n \sigma_{x_i}^2. \quad (3.34)$$

$t_r K_x$ – це сума всіх елементів головної діагоналі матриці K_x , тобто сума дисперсій метеорологічної величини в n точках поля. Ця величина називається **сумарною дисперсією поля**. А остання розподіляється таким чином, що найбільша її частина являє собою дисперсію першої ортогональної компоненти (перше власне значення). Зі збільшенням номеру ортогональної компоненти доля сумарної дисперсії, що їй належить, зменшується. Наприклад, власні значення матриці кореляції місячної кількості опадів у грудні (табл. 3.2) мають такі значення: $\lambda_1=18,39$; $\lambda_2=2,08$; $\lambda_3=0,37$; $\lambda_4=0,12$; $\lambda_5=0,03$, а решта ($i = \overline{6;21}$) наближаються до нуля.

Отже, чим більшим є масштаб атмосферного процесу, під дією якого формуються поля метеорологічних величин, тим більша дисперсія їм притаманна [2, 27]. Тому $\sigma_{z_1}^2$ відбиває властивості полів найбільшого масштабу, $\sigma_{z_2}^2$ – полів меншого масштабу і т.п. Ця властивість власних значень дає можливість отримати важливу інформацію про структуру досліджуваних метеорологічних об'єктів і, як наслідок, властивостей метеорологічних процесів, що їх зумовлюють.

3.1.4 Головні компоненти метеорологічних полів. Задачі стиску та фільтрації вихідної інформації

У попередньому підрозділі були представлені шляхи розв'язання рівняння повної проблеми власних значень, що дозволило отримати власні значення та власні вектори матриці коваріації (кореляції), які у подальшому були залучені до параметризації вихідних випадкових полів.

Оскільки власні значення λ_i ($i = \overline{1, n}$) дорівнюють дисперсіям ортогональних компонент метеорологічних полів, то рівняння (3.34) можна переписати так:

$$\sum_{i=1}^n \sigma_{z_i}^2 = \sum_{i=1}^n \sigma_{x_i}^2. \quad (3.35)$$

Але дисперсії метеорологічної величини у точках поля не упорядковані, тобто можуть мати різні значення, у той час коли дисперсії ортогональних компонент розташовані чітко у порядку зменшення, як це зазначається нерівністю (3.33). Причому вони *мають властивість швидкої збіжності*.

Необхідно визначити, яку суму дисперсій перших ортогональних компонент, тобто проєкцій метеорологічних полів у базисі власних векторів треба взяти, щоб вичерпати задану частку від сумарної дисперсії метеорологічного поля [2, 3, 27].

Як було показано в підрозділі 3.3, декілька перших власних значень матриці коваріації (чи кореляції) завжди складають більшу частину сумарної дисперсії поля. Це означає, що *основні властивості полів вичерпуються відповідною кількістю перших q ортогональних компонент*.

Ортогональні компоненти, сума дисперсій яких вичерпує більшу частку сумарної дисперсії вихідних метеорологічних полів, **називаються головними компонентами**. Саме вони містять найбільш суттєву інформацію про структуру метеорологічного об'єкта, що підлягає дослідженню.

Щоб визначити кількість головних компонент, треба, відповідно до сенсу поставленої задачі, визначити частку η_q сумарної дисперсії метеорологічного поля (або вектора предиктора). Ця частка визначається рівнянням

$$\eta_q = \frac{\sum_{i=1}^q \sigma_{z_i}^2}{\sum_{i=1}^n \sigma_{x_i}^2} = \frac{\sum_{i=1}^q \lambda_i}{\sum_{i=1}^n \lambda_i} = \frac{\sum_{i=1}^q \lambda_i}{t_r K_x} \cdot 100\% . \quad (3.36)$$

Число q , при якому $\eta_q \geq \eta$ (η задається дослідником) і визначає ту кількість головних компонент вектора (3.23), котрі містять найбільш важливу інформацію про статистичну структуру метеорологічних полів.

Отже, ми можемо тепер замість n -вимірного вихідного вектора використати q -вимірний вектор головних компонент, яких у цьому векторі лише декілька ($q \ll n$). Тоді останні $(n - q)$ ортогональних компонент відносяться до дрібномасштабних збурень та різних похибок, які містяться у вихідній інформації.

Визначивши кількість головних компонент гідрометеорологічних характеристик, можна вирішити задачу *стиску вихідної інформації* (значне зменшення її об'єму при умові збереження фізичного сенсу).

Очевидно, ми досягнемо поставленої мети, якщо сформуємо вектор головних компонент для j -го поля, дисперсії яких вичерпують задану частку від сумарної дисперсії вихідних полів за рівністю (3.36), а інші ортогональні компоненти обнулюємо. Такий модифікований вектор має вигляд

$$\tilde{Z}_j = \begin{pmatrix} z_{1j} \\ z_{2j} \\ \dots \\ z_{qj} \\ 0 \\ \dots \\ 0 \end{pmatrix}. \quad (3.37)$$

Таким чином, компонентний аналіз дає можливість перейти від сукупності, що складається з декількох десятків потенційних предикторів, до невеликої кількості головних компонент, які є незалежними. Це суттєво спрощує структуру прогностичної моделі й збільшує її вірогідність.

Задача фільтрації метеорологічних полів, у певному сенсі, є оберненою відносно задачі стиску інформації. Якщо у першому випадку треба було сконцентрувати найбільш суттєву інформацію у декількох перших ортогональних (головних) компонентах, то **в задачі фільтрації необхідно вихідну інформацію «звільнити» від інформації, що відноситься до дрібномасштабних збурень і похибок.**

Очевидно, справедливим є рівняння:

$$\frac{\sum_{i=1}^q \lambda_i}{t_r K_x} + \frac{\sum_{i=q+1}^n \lambda_i}{t_r K_x} = \eta + \delta = 100\%. \quad (3.38)$$

Оскільки η_q , як було визначено в першій частині підрозділу 3.4, характеризує ту найбільш суттєву інформацію, яку треба залишити, і яка міститься в перших q ортогональних (головних) компонентах метеорологічного поля, тоді

$$\delta = \frac{\sum_{i=q+1}^n \lambda_i}{t_r K_x} \cdot 100\% \quad (3.39)$$

відноситься до тієї інформації, від якої треба «звільнитися».

Ми досягнемо цієї мети, якщо будемо вважати, що всі ортогональні компоненти Z_j від $q+1$ до n -ї дорівнюють нулю (3.37).

Проведемо операцію обернення, оскільки

$$\tilde{Z}_j = W' \Delta \tilde{X}_j, \quad (3.40)$$

де

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

Для визначення вектора $\Delta \tilde{X}_j$ помножимо рівняння (3.40) на матрицю ортонормованих власних векторів.

Отримаємо

$$W \tilde{Z}_j = W W' \Delta \tilde{X}_j. \quad (3.41)$$

Оскільки $W W' = E$, маємо

$$\Delta \tilde{X}_j = W \tilde{Z}_j \quad (3.42)$$

або

$$\tilde{X}_j = W \tilde{Z}_j, \quad (3.43)$$

тобто від вектора ортогональних компонент поля перейшли до самого поля.

У координатній формі матричне рівняння (3.43) має вигляд:

$$\begin{pmatrix} w_{11} & w_{12} & \dots & w_{1p} & \dots & w_{1n} \\ w_{21} & w_{22} & \dots & w_{2p} & \dots & w_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ w_{p1} & w_{p2} & \dots & w_{pp} & \dots & w_{pn} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ w_{n1} & w_{n2} & \dots & w_{np} & \dots & w_{nn} \end{pmatrix} \begin{pmatrix} z_{1j} \\ z_{2j} \\ \dots \\ z_{qj} \\ 0 \\ \dots \\ 0 \end{pmatrix} = \begin{pmatrix} \sum_{q=1}^p w_{q1} z_{qj} \\ \sum_{q=1}^p w_{q2} z_{qj} \\ \dots \\ \sum_{q=1}^p w_{qp} z_{qj} \\ \dots \\ \sum_{q=1}^p w_{qn} z_{qj} \end{pmatrix} = \begin{pmatrix} \tilde{x}_{1j} \\ \tilde{x}_{2j} \\ \dots \\ \tilde{x}_{qj} \\ \dots \\ \tilde{x}_{nj} \end{pmatrix}. \quad (3.44)$$

$$\tilde{x}_{ij} = \sum_{q=1}^p w_{qi} z_{qj}, \quad (i = 1, \bar{n}). \quad (3.45)$$

Кожна компонента \tilde{x}_{ij} відфільтрованого поля, тобто відфільтроване значення метеорологічної величини на i -й метеорологічній станції в j -й момент часу визначається рівнянням (3.45).

Дисперсії головних компонент розташовуються, як вже зазначалось, у порядку зменшення, тобто дисперсія першої головної компоненти $\sigma_{z_1}^2 = \lambda_1$ – є найбільшою та відбиває властивості гідрометеорологічних полів найбільшого масштабу; дисперсія другої головної компоненти $\sigma_{z_2}^2 = \lambda_2$ – значно менша і т.д. Отже, можна прийти до висновку, що ортогональні компоненти, які у векторі Z_j розташовуються після головних компонент, відбивають властивості процесів все меншого й меншого масштабу, а останні – похибки первинної обробки і вимірювань.

Таким чином рівняння (3.40) – (3.45) дають можливість вирішити *задачу фільтрації вихідної інформації, тобто «звільнитися» від дрібномасштабних флуктуацій та похибок вимірювань.*

3.2 Розв’язання типової задачі. Стиск та фільтрація вихідної інформації за допомогою компонентного аналізу

Місячна кількість опадів є однією з важливих кліматичних параметрів. Вона характеризує умови зволоження того чи іншого району, а її поля – умови зволоження територій, для яких ці поля побудовані. На рис. 3.1 зображено поле місячної кількості опадів у грудні 1980 року. Воно визначене значеннями цієї метеорологічної величини на мережі 21 метеорологічної станції України, перелік яких наведено в табл. 3.1.

На рис. 3.1 методом лінійної інтерполяції проведені *ізогісти* – лінії однакових кількостей опадів. Таке поле може зображатись двадцятиодновимірним вектором, а сукупність полів за m років (з 1951 по 1980 рр.) – матрицею розмірністю 21×30 (табл. 3.1). В цій матриці кожен стовпець визначає окреме поле місячної кількості опадів, а кожен рядок – вибірку цієї величини на метеорологічній станції, умовний номер якої співпадає з номером рядка.

На основі вказаних даних була розрахована матриця кореляцій, за допомогою якої розв’язувалося рівняння повної проблеми власних значень вигляду (3.20). В табл. 3.2 наводяться власні значення та відповідні їм власні вектори матриці кореляції.

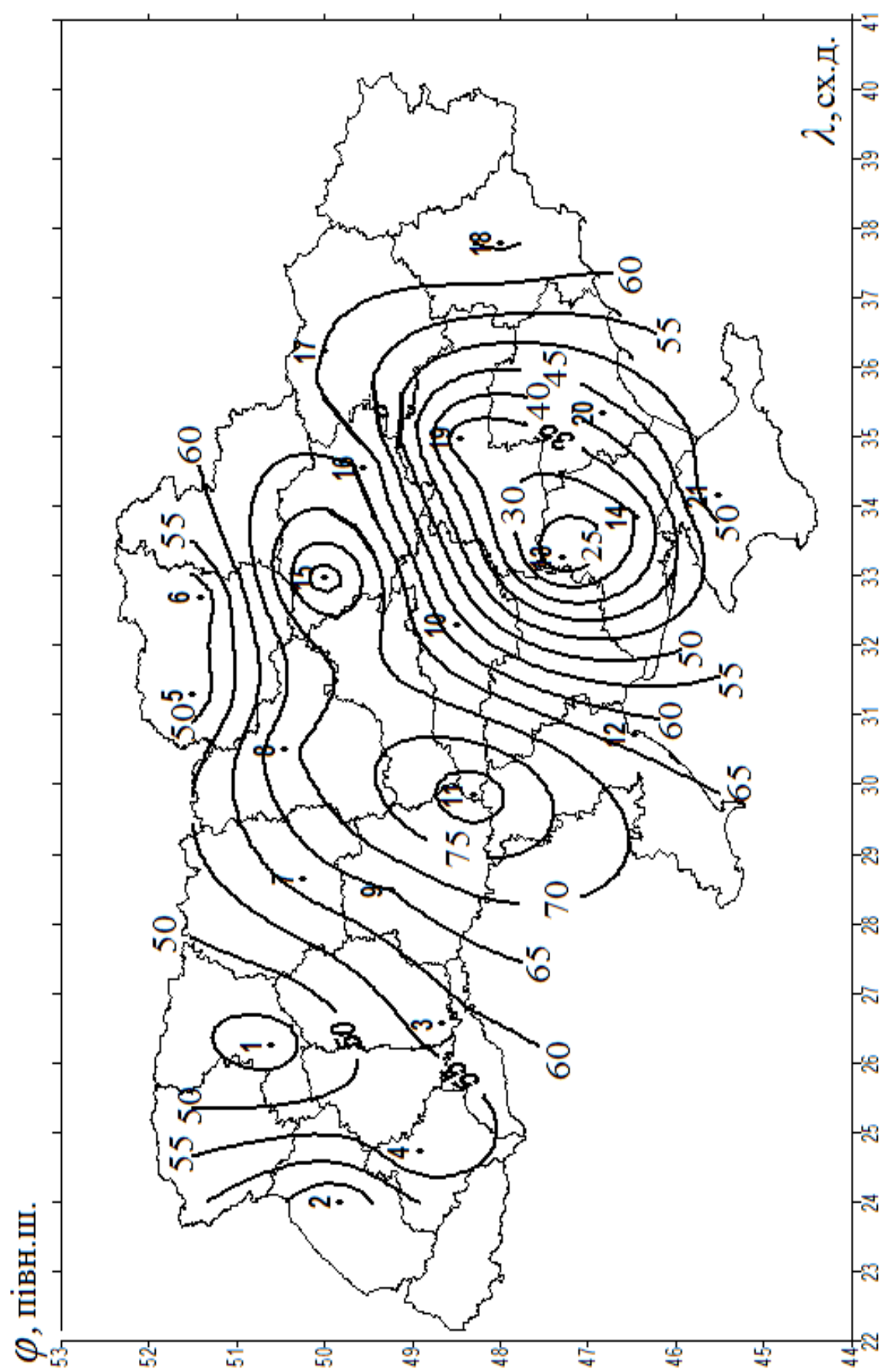


Рис. 3.1 – Фактичне поле місячної кількості опадів (мм) у грудні 1980 року на території України

Таблиця 3.1 – Місячна кількість опадів (мм) на станціях України у грудні

№ п/п	Назва станції	Рік														
		1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965
1	Рівне	21	28	14	58	52	39	31	39	61	39	29	25	15	80	33
2	Львів	27	51	18	65	80	91	33	62	96	52	48	50	30	34	61
3	Кам'янець-Подільський	18	52	15	37	30	68	18	27	62	43	23	40	53	36	30
4	Івано-Франківськ	18	38	34	42	21	54	21	25	54	46	28	40	36	46	66
5	Чернігів	22	24	6	45	42	46	30	38	42	58	24	12	26	62	54
6	Нові Млини	19	26	77	33	44	46	20	24	36	66	21	20	20	65	61
7	Житомир	20	42	16	43	50	42	22	36	46	62	25	17	25	67	47
8	Київ	20	30	14	31	49	52	20	49	45	79	23	15	25	48	58
9	Вінниця	21	32	11	35	47	57	20	51	48	71	22	20	46	64	37
10	Кіровоград	17	29	32	30	35	22	19	21	30	56	15	18	59	34	72
11	Гайворон	14	41	5	12	18	42	20	22	48	59	14	24	92	27	43
12	Одеса	4	28	4	32	29	37	34	11	39	70	63	53	98	19	57
13	Вел. Олександрівка	14	28	4	57	24	10	26	34	15	56	36	70	78	21	30
14	Асканія-Нова	10	31	5	57	36	5	34	41	13	62	28	96	56	21	13
15	Лубни	24	42	6	67	48	35	24	36	43	35	30	27	66	71	81
16	Полтава	14	35	5	56	50	31	18	33	36	65	20	34	58	42	64
17	Харків	14	44	10	68	53	22	13	39	47	48	25	42	52	45	62
18	Донецьк	16	72	9	55	76	29	80	92	33	65	49	12	99	15	53
19	Дніпропетровськ	6	30	6	52	46	17	24	34	20	65	33	110	70	49	36
20	Мелітополь	16	48	15	60	56	14	59	60	31	64	42	130	77	33	31
21	Клепініно	16	20	7	38	50	13	35	41	28	69	41	62	29	18	22

Продовження табл. 3.1

№ п/п	Назва станції	Рік														
		1966	1967	1968	1969	1970	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980
1	Рівне	36	52	41	33	68	34	6	20	43	2	42	17	46	36	42
2	Львів	51	84	52	52	63	51	8	28	66	10	59	49	73	51	70
3	Кам'янець-Подільський	32	40	36	68	70	68	6	14	47	2	41	14	46	26	58
4	Івано-Франківськ	42	41	41	55	63	50	4	16	25	3	29	26	88	23	50
5	Чернівці	106	56	61	75	54	53	6	57	48	4	41	23	46	40	46
6	Нові Млини	100	63	44	57	54	47	11	60	54	5	47	26	46	52	46
7	Житомир	39	50	42	44	50	49	4	27	43	3	39	16	39	30	62
8	Київ	83	85	54	100	74	66	5	105	42	5	26	27	44	45	69
9	Вінниця	42	56	41	72	73	79	7	60	41	3	39	13	48	43	65
10	Кіровоград	98	60	47	64	39	28	5	43	28	1	46	30	73	31	57
11	Гайворон	52	51	46	59	55	48	1	68	18	1	51	15	57	46	84
12	Одеса	111	72	60	96	36	46	3	36	29	1	61	72	49	13	63
13	Вел. Олександрівка	112	39	45	68	37	30	5	27	58	2	51	28	49	22	20
14	Асканія-Нова	75	56	33	69	30	35	1	29	64	3	58	33	57	27	30
15	Лубни	92	90	52	71	63	63	6	83	57	6	65	40	89	61	84
16	Полтава	105	68	36	66	43	30	6	58	47	4	61	18	120	58	65
17	Харків	132	72	37	72	23	26	7	54	46	3	76	16	96	32	60
18	Донецьк	103	96	27	37	40	23	3	42	91	3	113	30	90	20	66
19	Дніпропетровськ	109	60	37	98	49	31	5	41	41	3	61	17	120	31	31
20	Мелітополь	68	70	22	120	61	38	4	34	54	3	66	42	59	14	43
21	Клепініно	42	91	27	48	23	49	7	28	77	2	44	49	42	17	55

Таблиця 3.2 – Власні значення λ_i та відповідні їм власні вектори W_i матриці кореляції місячної кількості опадів (грудень)

Назва станції	λ_i i	Власні значення та власні вектори									
		1	2	3	4	5	6	7	8	9	10
		18.39	2.08	0.37	0.12	0.03	0.01	0.00	0.00	0.00	0.00
Рівне	1	0.20	-0.33	0.12	-0.10	0.31	0.14	-0.06	0.57	-0.10	-0.35
Львів	2	0.19	-0.37	0.36	-0.32	0.06	-0.35	-0.22	0.15	-0.06	0.41
Кам'янець-Подільський	3	0.21	-0.28	0.18	0.13	-0.26	0.19	0.20	-0.34	0.04	-0.34
Івано-Франківськ	4	0.20	-0.35	0.33	-0.14	-0.17	-0.17	0.10	-0.39	0.21	-0.05
Чернігів	5	0.22	-0.12	-0.35	0.03	0.34	-0.04	-0.19	-0.18	0.33	-0.08
Нові Млини	6	0.23	-0.05	-0.38	-0.07	0.20	-0.22	-0.18	-0.28	0.13	-0.13
Житомир	7	0.22	-0.22	-0.11	0.13	0.18	0.29	0.06	0.14	-0.08	-0.02
Київ	8	0.23	-0.12	-0.26	0.10	0.15	0.10	0.02	-0.03	-0.01	0.28
Вінниця	9	0.22	-0.19	-0.03	0.22	-0.10	0.31	0.15	-0.01	-0.19	0.05
Кіровоград	10	0.23	0.08	-0.09	0.11	-0.18	-0.12	0.17	0.20	0.07	0.29
Гайворон	11	0.23	-0.05	-0.05	0.32	-0.29	0.13	-0.01	0.05	-0.11	0.27
Одеса	12	0.22	0.14	0.13	0.45	-0.26	-0.20	-0.64	0.10	0.02	-0.18
Вел. Олександрівка	13	0.22	0.20	0.09	0.11	-0.04	-0.16	0.15	0.22	0.29	-0.09
Асканія-Нова	14	0.21	0.26	0.22	0.13	0.19	-0.12	0.10	0.02	0.10	-0.14
Лубни	15	0.23	0.03	-0.25	-0.07	-0.07	-0.18	0.12	-0.03	-0.14	0.25
Полтава	16	0.23	0.11	-0.18	-0.20	-0.17	-0.18	0.17	0.06	-0.18	-0.03
Харків	17	0.22	0.15	-0.17	-0.38	-0.24	-0.06	-0.08	-0.03	-0.46	-0.39
Донецьк	18	0.21	0.26	0.10	-0.42	-0.10	0.61	-0.40	-0.09	0.18	0.21
Дніпропетровськ	19	0.22	0.19	-0.01	-0.18	-0.12	-0.04	0.24	0.18	0.19	-0.07
Мелітополь	20	0.21	0.27	0.21	-0.07	0.18	0.11	0.23	0.01	0.27	0.01
Клепініно	21	0.20	0.30	0.33	0.15	0.46	-0.04	0.03	-0.32	-0.51	0.10

Продовження табл. 3.2

Назва станції	λ_i i	Власні значення та власні вектори																			
		11	12	13	14	15	16	17	18	19	20	21									
		0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00									
Рівне	1	0.21	0.35	0.01	-0.03	0.07	0.05	0.09	0.18	-0.06	0.08	0.14									
Львів	2	0.09	-0.37	0.04	0.04	0.00	0.07	-0.04	-0.07	0.07	-0.13	-0.19									
Кам'янець-Подільський	3	0.27	0.03	0.07	-0.06	0.20	0.20	-0.09	0.05	0.17	-0.37	-0.34									
Івано-Франківськ	4	-0.33	0.29	-0.10	-0.04	-0.07	-0.18	0.06	0.01	-0.12	0.27	0.33									
Чернігів	5	-0.11	-0.13	-0.03	0.04	0.43	0.27	0.25	-0.26	-0.24	0.14	-0.12									
Нові Млини	6	0.47	-0.07	-0.05	0.07	-0.32	-0.38	-0.15	0.14	0.08	-0.09	0.16									
Житомир	7	-0.37	-0.09	0.03	-0.09	-0.05	-0.35	-0.41	-0.49	0.18	-0.12	-0.00									
Київ	8	-0.42	0.03	-0.03	-0.08	-0.09	0.17	0.13	0.60	0.22	-0.30	0.06									
Вінниця	9	0.06	-0.26	-0.02	0.28	-0.50	0.04	0.29	-0.01	-0.31	0.28	-0.20									
Кіровоград	10	0.14	0.24	-0.30	-0.31	0.15	-0.44	0.32	-0.07	-0.13	-0.12	-0.31									
Гайворон	11	0.28	-0.23	-0.03	-0.03	0.38	0.05	-0.08	0.03	0.09	0.19	0.56									
Одеса	12	-0.17	0.11	0.29	0.01	-0.10	-0.05	0.03	0.00	-0.03	-0.06	-0.09									
Вел. Олександрівка	13	-0.06	-0.10	-0.46	0.08	-0.11	0.26	-0.49	0.10	-0.35	-0.12	0.02									
Асканія-Нова	14	0.02	-0.09	-0.28	-0.07	-0.17	0.16	0.24	-0.16	0.66	0.27	-0.02									
Лубни	15	0.07	0.41	0.27	-0.03	-0.04	0.26	-0.38	-0.01	0.11	0.44	-0.28									
Полтава	16	0.00	0.16	0.15	0.04	-0.22	0.31	0.25	-0.42	-0.06	-0.43	0.35									
Харків	17	-0.21	-0.31	-0.16	-0.27	0.14	-0.07	-0.01	0.16	-0.05	0.16	-0.08									
Донецьк	18	0.08	0.19	-0.16	0.05	-0.06	0.05	-0.04	-0.06	0.03	-0.03	-0.02									
Дніпропетровськ	19	-0.12	-0.07	0.26	0.67	0.26	-0.28	0.07	0.14	0.14	-0.00	-0.05									
Мелітополь	20	0.05	-0.24	0.55	-0.47	-0.09	-0.05	0.00	0.10	-0.20	0.04	0.07									
Клепініно	21	0.03	0.15	-0.06	0.18	0.19	-0.08	-0.03	0.02	-0.20	-0.11	0.01									

Для розв'язання задачі стиску вихідної інформації про поле опадів у грудні взяли $\eta \geq 95\%$.

З табл. 3.2 випливає, що перші два власні значення вичерпують близько 97% сумарної дисперсії місячної кількості опадів ($\eta \geq 95\%$):

$$\begin{aligned}\lambda_1 &= 18,39; & \eta_1 &= 87,6\%; \\ \lambda_2 &= 2,08; & \eta_2 &= 9,9\%; \\ t_r R_x &= 21.\end{aligned}$$

Тому найбільш великомасштабні властивості статистичної структури полів опадів містяться в перших двох власних векторах.

Перший власний вектор (рис. 3.2) має однорідну структуру, що зумовлюється впливом на поле опадів найбільш великомасштабного опадоутворювального процесу – переміщення через Україну циклонів та атмосферних фронтів у загальному напрямку – з заходу на схід. Тобто опадоутворювальні процеси, що мають найбільший масштаб, є однорідними для всієї території України.

Другий власний вектор (рис. 3.3) має області максимуму в південно-східній, а мінімуму – у північно-західній частинах України. Зазначений максимум зумовлюється впливом циклонів, що спостерігаються в цю пору року й проходять з Чорного моря в північно-східному напрямку. Тобто другий власний вектор відбиває вплив процесів синоптичного масштабу на утворення опадів. Це є деякий коливальний процес з протилежними тенденціями в західних та східних частинах України, тобто, процеси, які зумовлюють збільшення опадів над західною частиною, пов'язані з процесами, що не викликають опадів над східною її частиною.

Відповідно до зазначеної структури власних значень та власних векторів головна інформація про властивості множини полів опадів міститься в перших двох головних компонентах (z_{1j}, z_{2j}). Отримаємо їх, якщо до вихідного поля опадів X_j (наприклад за 1980 рік) застосуємо ортогональне перетворення першими двома ортонормованими векторами (W_1, W_2). Одна з координат першої головної компоненти – це результат скалярного добутку першого ортонормованого власного вектора на вихідне поле 1980 року:

$$z_{1j} = W_1' \cdot X_j = 253,56.$$

Відповідно одна з координат другої головної компоненти визначається як:

$$z_{2j} = W_2' \cdot X_j = -24,78.$$

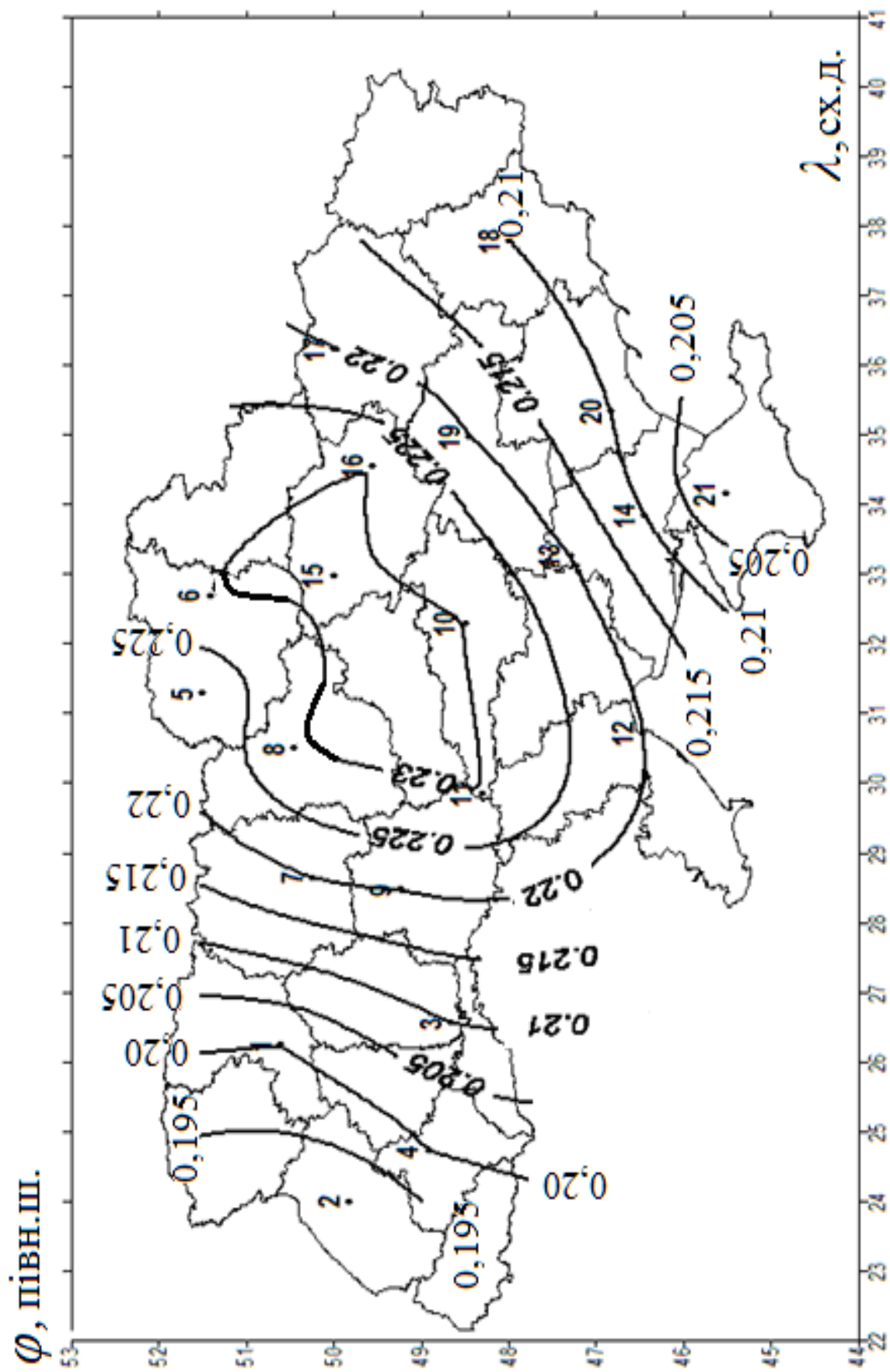


Рис. 3.2 – Поле першого власного вектора матриці кореляції місячної кількості опадів у грудні

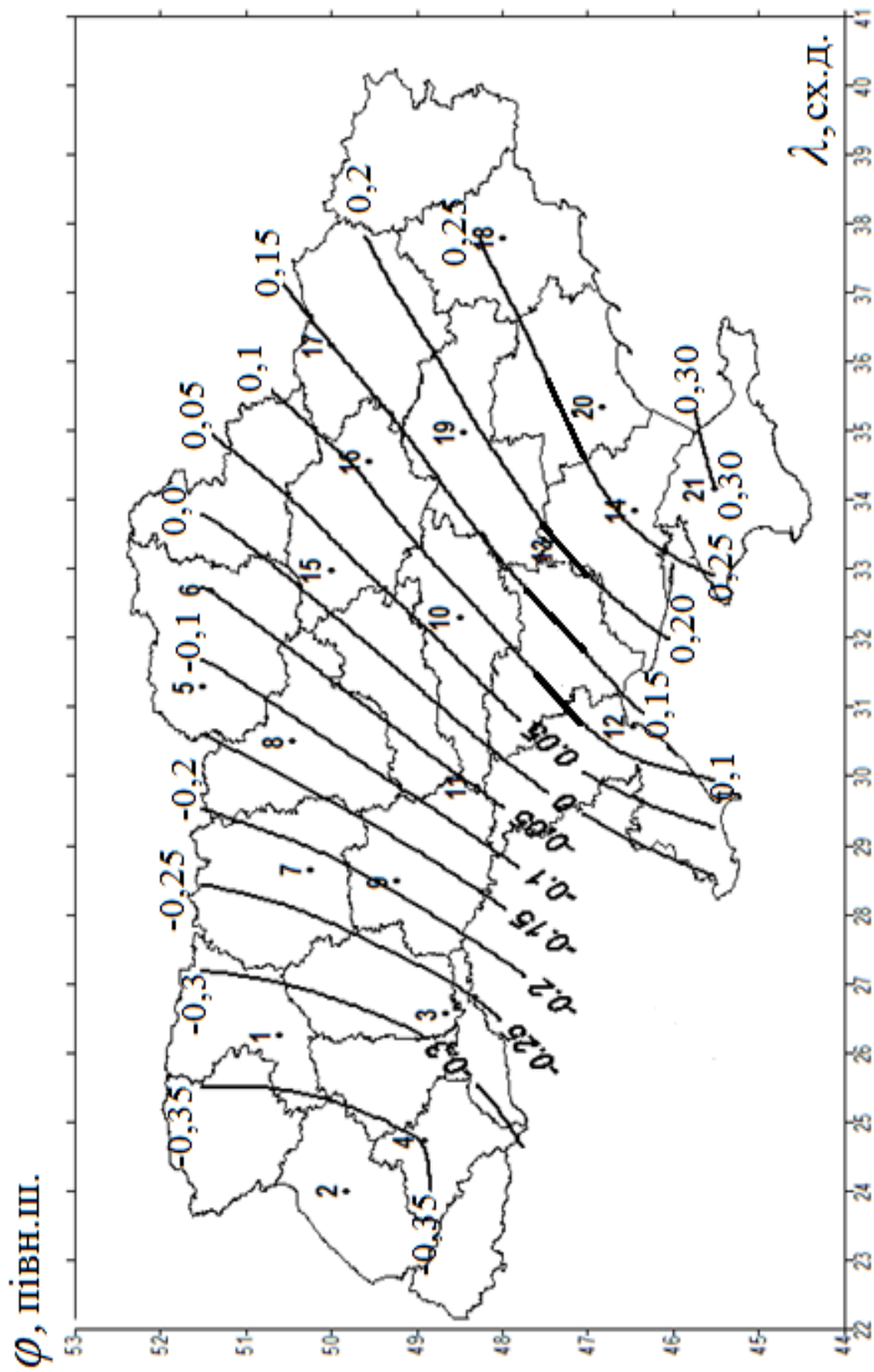


Рис. 3.3 – Поле другого власного вектора матриці кореляції місячної кількості опадів у грудні

Розрахувавши аналогічним чином для першої та другої головних компонент (в нашому прикладі) 30 координат, ми отримаємо часові ряди цих головних компонент з 21 ортогональної.

Таким чином ми здійснили операцію стиску вихідної інформації про поле місячної кількості опадів на території України, яке можна описати двома головними компонентами.

Отримані нові змінні, що називаються головними компонентами, мають самостійне значення, оскільки «несуть у собі» велике смислове навантаження. Їх можна включати до складу потенційних предикторів при побудові прогностичних моделей, що спрощує структуру та збільшує їх вірогідність.

Розв'яжемо тепер задачу фільтрації поля опадів за грудень 1980 р. Для цього необхідно втілити операцію, яка відповідає рівнянню (3.43). Відповідне відфільтроване поле зображене на рис. 3.4 та в табл. 3.3.

Таблиця 3.3 – Відфільтровані значення місячної кількості опадів у грудні 1980 р.

1980	Номер станції										
	1	2	3	4	5	6	7	8	9	10	11
\widetilde{X}_j	59	57	60	59	59	60	61	61	61	56	60

Продовження табл. 3.3

1980	Номер станції									
	12	13	14	15	16	17	18	19	20	21
\widetilde{X}_j	52	51	47	58	56	52	47	51	47	43

Як впливає з рис. 3.1, структура фактичного поля опадів за 1980 р. яскраво відбиває вплив процесів різних масштабів на його формування у грудні. Задача ж фільтрації дозволяє визначити вплив найбільш великомасштабних атмосферних процесів при утворенні опадів на Україні – глобального західного переносу помірних широт.

Як бачимо, у грудні (рис. 3.4) максимум опадів спостерігається у західній та центральній частинах України, який, як відомо, є результатом регенерації циклонів внаслідок перевалювання їх через Карпати, а вплив рельєфу вже нівелюється.

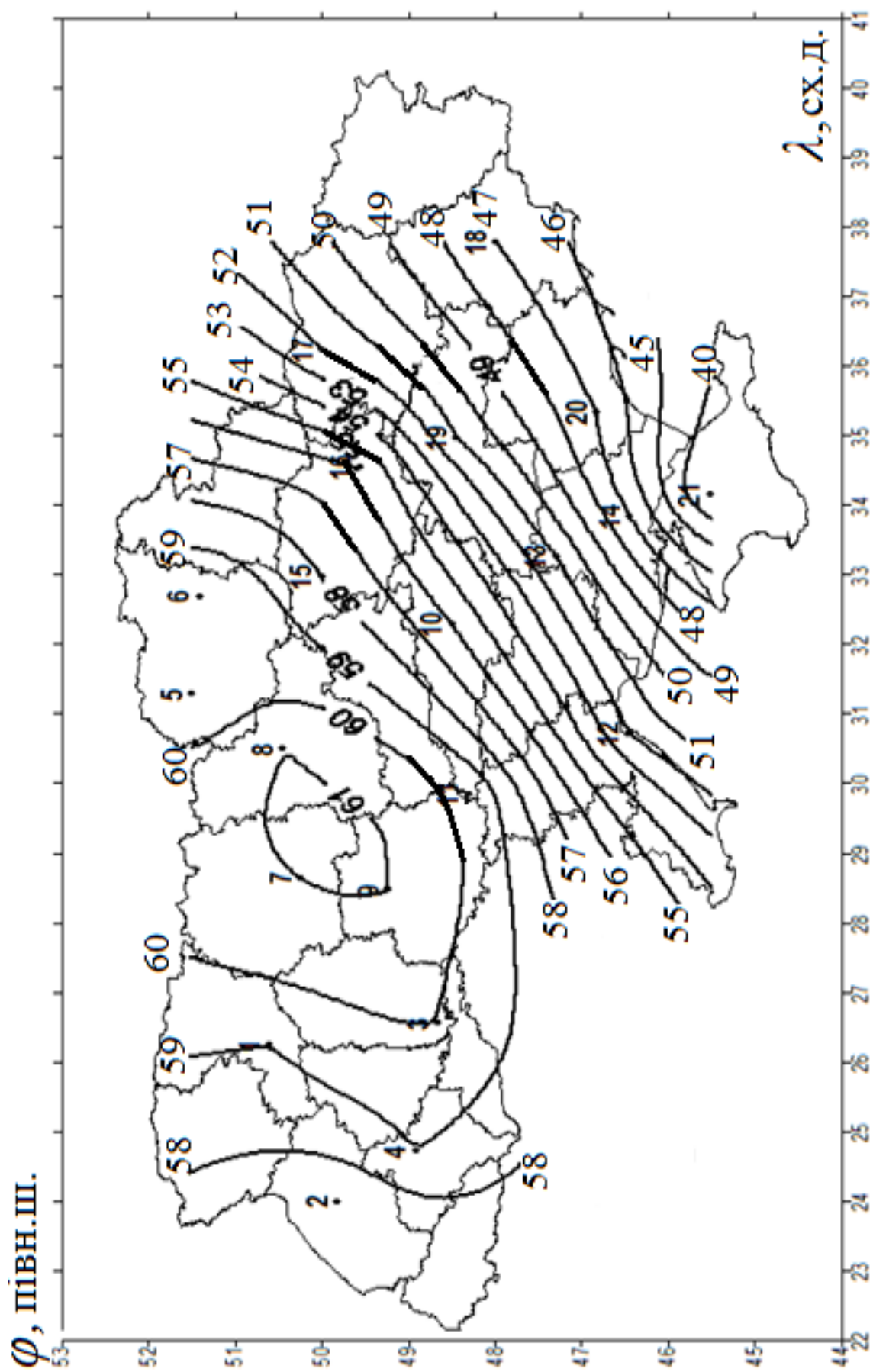


Рис. 3.4 – Відфільтроване поле місячної кількості опадів за 1980 р., мм (відповідає полю опадів рис. 3.1)

3.3 Запитання для самоперевірки

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

3.4 Завдання для закріплення теми

1. Отримати у викладача варіант задачі для виконання практичної роботи (Додаток Б) та комп'ютерну програму «EOF» (призначення програми та особливості користування дивись у Додатку В).
2. Реалізувати на ПК алгоритм розв'язання задачі повної проблеми власних значень і отримати власні значення та власні вектори матриці кореляції.
3. Розв'язати задачу стиску вихідної інформації про поле місячної кількості опадів для території України за умови $\eta \geq 70\%$:
 - визначити, який відсоток сумарної дисперсії заданого поля вичерпує перше власне значення;

- визначити кількість власних значень та власних векторів за умови $\eta \geq 70\%$;
 - побудувати та проаналізувати поля перших власних векторів матриці кореляції;
 - визначити вектор ортогональних компонент та отримати головні компоненти.
6. Розв'язати задачу фільтрації вихідної інформації про поле місячної кількості опадів для території України:
- побудувати та проаналізувати вихідне поле місячної кількості опадів для 33 станцій України (Додаток Б);
 - отримати відфільтроване поле місячної кількості опадів та побудувати його на карті.
7. Провести аналіз результатів застосування компонентного аналізу до поля місячної кількості опадів для 33 станцій України та зробити висновок щодо роботи у цілому.

4 ФАКТОРНИЙ АНАЛІЗ

4.1 Базові теоретичні положення

4.1.1 Основні засади факторного аналізу

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 та, починаючи з першої половини минулого століття, використовують його при розв'язанні проблем статистичного аналізу і в інших науках, у тому числі й науках метеорологічного профілю [15, 27].

При аналізі структури коваріаційних (або кореляційних) матриць найчастіше використовуються два підходи, формально деякою мірою схожих один до одного, однак різних за метою.

Одним з них є метод головних компонент за Пірсоном (Pearson, 1901) і Хотеллінгом (Hotelling, 1933), інший – факторний аналіз, що виник як метод в роботах Спірмана (Spearman, 1904, 1926) [15]. Необхідно розрізняти ці два підходи, хоча вони, на перший погляд, мають деяку подібність.

Метод емпіричних ортогональних компонент було розглянуто у попередньому розділі.

Коли застосовується метод головних компонент, значення компонент можна розрахувати; дійсно, між ними та значеннями змінних, що спостерігаються, діє взаємна однозначна відповідність. Навпаки, в факторному аналізі, де повна дисперсія змінних не вичерпується простими факторами f_j , проблема значно складніша, оскільки значення факторів неможливо оцінити в звичайному статистичному сенсі.

У факторному аналізі процес оцінювання можливо розглядати як двостадійний процес. Спочатку визначається факторна структура; під цим розуміється необхідна кількість факторів для пояснення кореляцій між змінними та навантаження факторів і цих змінних.

Після того, як це зроблено, залишається проблема оцінки значень індивідуальних членів вибірки для самих факторів.

Розглянемо систему метеорологічних величин x_i ($i = \overline{1, n}$), які є центрованими, тобто $M[x_i] = 0$. Задача полягає у тому, щоб виразити кожную з них за допомогою k деяких випадкових величин (*узагальнених факторів*) f_j ($j = \overline{1, k}$), які є однаковими для всіх метеорологічних величин, але входять у кожную з них із своєю вагою P_{ij} . Фактори f_j не повністю

Отже

При цьому $k \ll n$. Якщо в рівнянні (4.1) розгорнути суму і придати всі значення індексу i , то прийдемо до системи рівнянь:

Всі змінні x_i , як зазначалося вище, виражаються через однакові фактори f_j ($j = \overline{1, k}$). Проте вони входять в різні величини x_i з різними вагами p_{ij} ($i = \overline{1, n}; j = \overline{1, k}$).

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \cdots \\ x_i \\ \cdots \\ x_n \end{pmatrix}, \quad (4.3)$$

$$F = \begin{pmatrix} f_1 \\ f_2 \\ \dots \\ f_j \\ \dots \\ f_k \end{pmatrix}, \quad (4.4)$$

$$V = \begin{pmatrix} v_1 \\ v_2 \\ \dots \\ v_i \\ \dots \\ v_n \end{pmatrix} \quad (4.5)$$

і матрицю вагів

$$P = \begin{pmatrix} p_{11} & p_{12} & \dots & p_{1j} & \dots & p_{1k} \\ p_{21} & p_{22} & \dots & p_{2j} & \dots & p_{2k} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ p_{i1} & p_{i2} & \dots & p_{ij} & \dots & p_{ik} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ p_{n1} & p_{n2} & \dots & p_{nj} & \dots & p_{nk} \end{pmatrix}. \quad (4.6)$$

Тоді систему рівнянь (4.2) можна записати в матричній формі:

$$X = PF + V. \quad (4.7)$$

Фактори f_j будемо називати узагальненими факторами і вважатимемо їх такими, що :

$$M[F \cdot F'] = E, \quad (4.8)$$

де

E — одинична матриця.

Умова (4.8) означає, що узагальнені фактори некорельовані та мають одиничну дисперсію.

Крім того, будемо вважати, що залишки v_i є незалежними, тобто

$$M[V \cdot V'] = D, \quad (4.9)$$

де:

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

$$D = \begin{pmatrix} d_1 & 0 & \dots & 0 \\ 0 & d_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & d_n \end{pmatrix}. \quad (4.10)$$

Залишки U_i є некорельовані з факторами. За таких умов знайдемо матрицю коваріацій вихідних змінних, які є центрованими:

$$\begin{aligned} K_x &= M[X \cdot X'] = M\{(PF + V)(F'P' + V')\} = \\ &= PM[F \cdot F']P' + M[V \cdot F'] \cdot P' + PM[F \cdot V'] + M[V \cdot V']. \end{aligned} \quad (4.11)$$

Враховуючи перелічені властивості узагальнених факторів і залишків, отримаємо:

$$K_x = PP' + D. \quad (4.12)$$

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

4.1.2 Оцінки вагів узагальнених факторів та дисперсій залишків

Будемо вважати, що змінні x_1, x_2, \dots, x_n підлягають нормальному багатовимірному закону розподілу з n -вимірною матрицею коваріацій K_x . Це традиційна умова для багатовимірного статистичного аналізу [15, 27].

З урахуванням рівнянь (4.1) та (4.7) можна записати, що діагональні елементи (k_{ii}) матриці коваріацій – дисперсії вихідної змінної, які дорівнюють:

$$k_{ii} = \sum_{j=1}^k P_{ij}^2 + d_i, \quad (4.13)$$

а коваріації

$$k_{ij} = \sum_{j=1}^k P_{ij} P_{qj}. \quad (4.14)$$

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 (4.13) з додатними елементами й матриці рангу k (4.14) 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 набуває вигляду $(n+k) < (n-k)^2$.

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 невідомі параметри, котрі надають максимум функції правдоподібності.

Функція правдоподібності має такий вигляд:

$$L = -\frac{1}{2}n \ln |K| - \frac{n}{2} \text{tr} (K^{-1} \cdot \hat{K}), \quad (4.15)$$

де

tr – слід матриці;

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

Функція правдоподібності, як свідчать рівняння (4.13-4.14), є функцією двох невідомих змінних p_{ij} та d_i . Максимум функції (4.15) визначається за допомогою рівнянь:

$$\frac{\partial L}{\partial p_{ij}} = 0; \quad \frac{\partial L}{\partial d_i} = 0. \quad (4.16)$$

Результати диференціювання функції L приводять до матричних рівнянь:

$$P'K' - P'K^{-1}\widehat{K}K^{-1} = 0, \quad (4.17)$$

$$diag\left(K^{-1} - K^{-1}\widehat{K}K^{-1}\right) = 0, \quad (4.18)$$

де

$diag Q$ (Q – матриця в дужках) позначає матрицю, яка має лише діагональні елементи.

Введемо матрицю

$$N = P'D^{-1}P \quad (4.19)$$

і використаємо тотожність:

$$P'K^{-1} = (E + N)^{-1}P'D^{-1}. \quad (4.20)$$

Підставимо рівності (4.19) та (4.20) у (4.17) і після математичних перетворень будемо мати:

$$P' = N^{-1}P'D^{-1}(\widehat{K} - D). \quad (4.21)$$

Рівняння (4.21) помножимо ліворуч на матрицю $D = K - PP'$. Отримаємо

$$diag(K - \widehat{K}) = 0. \quad (4.22)$$

Рівняння (4.22) означає, що за умови $i = j$, $K_{ii} = \widehat{K}_{ii}$, тобто діагональні елементи генеральної та вибіркової матриць коваріацій дорівнюють одне одному.

Враховуючи це, за допомогою рівняння (4.21) отримаємо:

$$d_i = \widehat{K}_{ii} - \sum_{j=1}^k P_{ij}^2. \quad (4.23)$$

Отже, оскільки вибіркова коваріаційна матриця відома, рівняння (4.23) дає можливість знайти дисперсії залишків після визначення вагів узагальнених факторів.

Для визначення вагів узагальнених факторів помножимо рівняння (4.21) ліворуч на матрицю (4.19), а праворуч – на матрицю $D^{-1}P$.

Будемо мати:

$$N^2 = P'D^{-1}(\widehat{K} - D)D^{-1}P. \quad (4.24)$$

Матриця (4.24) є діагональною. Її елементи являють собою перші k власних значень матриці $\hat{A}^{-1}(\hat{K} - \hat{A})$, 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 (4.21) та (4.23) [27].

Ітераційний процес описується такою системою рівнянь:

$$\begin{cases} h_{\mu(v)} = P'_{\mu(v)} D_{(v)}^{-1}; \\ \alpha_{\mu j(v)} = h'_{\mu(v)} P_{j(v)}; \\ g_{\mu(v)} = h_{\mu(v)} \hat{K} - P'_{\mu(v)} - \sum_{j=1}^{\mu-1} \alpha_{\mu j(v)} P'_{j(v+1)}; \\ \beta_{\mu(v)} = g'_{\mu(v)} h_{\mu(v)}; \\ P_{\mu(v+1)} = \frac{1}{\sqrt{\beta_{\mu(v)}}} g'_{\mu(v)}, \end{cases} \quad (4.25)$$

де

μ – номер узагальненого фактора;

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

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

Для організації ітераційної процедури необхідно визначити кількість факторів k , які разом з залишком ν_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 (4.13) та (4.14), тобто вона повинна бути сумою діагональної матриці з додатними елементами (4.13) матриці рангу k з додатними власними значеннями.

Якби ν_i були відомі, тоді необхідно, щоб $k < n$. Але коли ν_i невідомі, умова набуває вигляду $(n + k) < (n - k)^2$, додатність числа

$$\frac{1}{2}[(n-k)^2 - (n+k)] \geq 0 \quad (4.26)$$

забезпечує нетривіальність гіпотези H_0 .

У загальному випадку гіпотеза H_0 перевіряється за допомогою критерію χ^2 [27], який розраховується за формулою:

$$\chi^2 = n \left(\frac{1}{2} \sum_{i=1}^n y_{ii}^2 + \sum_{i,j=1}^n y_{ij} y_{ji} \right), \quad (4.27)$$

де

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

$$Y = D^{-1}(\widehat{K}_0 - PM'), \quad (4.28)$$

а матриця M' визначається рівнянням:

$$M' = N^{-1}P'D^{-1}\widehat{K}_0. \quad (4.29)$$

Матриця \widehat{K}_0 формується шляхом заміни елементів головної діагоналі матриці коваріацій \widehat{K} найбільшими елементами її відповідних стовпців (при визначенні вагів першого узагальненого 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

$$\nu = \frac{1}{2}[(n-k)^2 + (n+k)] \quad (4.30)$$

та заданого рівня значущості α .

Число ітерацій визначається співвідношенням:

$$\left| P_{\mu j(m+1)} - P_{\mu j(m)} \right| < \varepsilon, \quad j = \overline{1, n}, \quad (4.31)$$

де

ε – задана точність розрахунків;

n – порядок матриці коваріації.

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

$$F = (E + N)^{-1} P' D^{-1} X. \quad (4.32)$$

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

4.2 Приклади застосування факторного аналізу до гідрометеорологічної інформації

Наведемо декілька прикладів застосування факторного аналізу.

Одним з прикладів застосування факторного аналізу є дослідження статистичної структури швидкості вітру у тропосфері та стратосфері. В ранніх роботах автора (Гончарова Л.Д., 1995) [7] шар атмосфери 5-55 км був розділений на два підшари з межею на висоті 25 км. Для цих підшарів за даними ракетного зондування атмосфери була вирішена повна проблема власних значень і проведено дослідження статистичної структури часових рядів головних компонент. Однак поділ тропо-стратосфери на вказані підшари було проведено з евристичних міркувань. Попередній аналіз вертикальних профілів амплітуд довгоперіодних коливань компонент швидкості вітру показує, що є необхідність ретельного дослідження вертикальної статистичної структури швидкості вітру, яке дозволило б на цій основі провести поділ тропо-стратосфери на шари, що характеризуються деякими загальними властивостями. Для вирішення цієї задачі було вибрано факторний аналіз [10].

Зазначена вище методика (підрозділ 4.1) була застосована для дослідження вертикальної статистичної структури компонент швидкості вітру на основі матриць коваріацій вертикальних профілів зональної та меридіональної складових швидкості вітру. Вертикальні профілі визначалися значеннями компонент швидкості вітру на кожному п'ятикілометровому рівні в шарі атмосфери 5-55 км. Використовувалися дані ракетного зондування у п'яти пунктах Західної півкулі (табл. 4.1 і табл. 4.2).

Вертикальні коваріаційні матриці 11-го порядку для складових швидкості вітру розраховувались за статистичними сукупностями вертикальних профілів з тижневим інтервалом. Загальні їх об'єми склали:

для пунктів Ф.Шерман, о.Антигуа і о.Вознесенія – 384, о-ва Воллоп – 528, Уайн Сендз – 480.

До матриць вертикальних коваріацій складових швидкості вітру був застосований факторний аналіз, результати якого містяться в табл. 4.1 і табл. 4.2.

Таблиця 4.1 – Факторні ваги зональної компоненти швидкості вітру

Висота, км	Ф.Шерман 9°20' півн.ш. і 80°00' зах.д			о.Антигуа 17°12' півн.ш. і 61°48' зах.д			о.Вознесенія 8°00' півд.ш. і 14°24' зах.д		
	1	2	3	1	2	3	1	2	3
5	0.209	-0.815	-0.002	0.256	-0.780	0.136	-0.087	0.707	0.813
10	-0.060	-0.829	-0.074	0.596	-0.960	0.174	0.847	-0.177	0.881
15	-0.063	-0.752	-0.060	0.083	-0.907	0.202	0.003	0.322	0.882
20	0.239	-0.352	-0.749	0.227	-0.617	0.506	0.078	-0.714	0.865
25	0.103	0.675	-0.340	0.214	-0.211	0.859	0.165	-0.597	-0.058
30	-0.108	0.569	-0.600	0.228	-0.204	0.806	-0.259	-0.708	0.089
35	0.271	0.680	-0.328	0.439	-0.297	0.633	0.312	-0.768	0.056
40	0.749	-0.230	-0.158	0.800	-0.165	0.282	0.722	0.367	0.028
45	0.783	-0.090	0.249	0.851	-0.219	0.256	0.636	0.159	0.078
50	0.883	-0.035	-0.006	0.884	0.072	0.174	0.872	0.093	0.061
55	0.324	0.197	-0.091	0.775	-0.157	0.109	0.770	0.014	-0.060

З табл. 4.1 випливає, що в усіх пунктах, що відносяться до тропічної зони, в зональній компоненті швидкості вітру (також як і в меридіональній) структуру матриці коваріації пояснюють три узагальнені фактори. Розподіл факторних вагів свідчить про те, що шар атмосфери 5-55 км за особливостями вертикальної статистичної структури слід розділити не на два, а на три підшари.

Перший з них обмежується висотами 5-20 км, тобто охоплює верхню тропосферу і нижню стратосферу. Цей підшар у тропічній зоні Північної півкулі (о.Антигуа, ф.Шерман) виділяється за характером вагів другого, а в тропічній зоні Південної півкулі (о.Вознесенія) – третього узагальненого фактора.

Другий підшар має товщину 20-35 км. Цей підшар виділяється однорідним характером розподілу вагів у пункті Ф.Шерман – другого і третього, у пункті о.Вознесенія – другого узагальненого фактора.

Нарешті останній підшар 35-55 км має однакові за структурою ваги першого узагальненого фактора (високі додатні значення).

У середніх широтах (Уайт Сендз і о-ва Воллоп) матриця коваріації зональної компоненти швидкості вітру пояснюється тільки двома узагальненими факторами (табл. 4.2). Аналіз вагів першого показує, що

вони розбиваються на дві групи: підшар 5-20 км (високі від'ємні ваги другого узагальненого фактора); підшар 20-55 км (високі додатні ваги першого узагальненого фактора).

Таблиця 4.2 – Факторні ваги компонент швидкості вітру

Висота, км	Уайт Сендз (32°23' півн.ш. і 106°29' зах.д.)					
	Компонента					
	зональна			меридіональна		
	1	2	3	1	2	3
5	0.214	-0.811	-	0.850	-0.116	0.016
10	0.260	-0.917	-	0.897	-0.198	0.105
15	0.393	-0.833	-	0.828	-0.328	-0.122
20	0.607	-0.531	-	0.649	-0.013	0.119
25	0.867	-0.279	-	-0.215	-0.421	-0.615
30	0.861	-0.212	-	-0.278	0.087	-0.617
35	0.910	-0.297	-	0.139	0.154	-0.828
40	0.919	-0.244	-	0.114	0.446	-0.765
45	0.898	-0.267	-	-0.209	0.636	-0.271
50	0.866	-0.316	-	0.187	0.731	-0.238
55	0.841	-0.358	-	0.182	0.781	0.092
Висота, км	о-ва Воллоп (37°50' півн.ш. і 75°29' зах.д.)					
	Компонента					
	зональна			меридіональна		
	1	2	3	1	2	3
5	0.250	-0.900	-	-0.159	-0.828	-0.260
10	0.196	-0.911	-	0.097	-0.955	-0.032
15	0.368	-0.852	-	0.147	-0.905	-0.065
20	0.615	-0.608	-	0.079	-0.459	-0.757
25	0.869	-0.308	-	0.413	-0.120	-0.578
30	0.819	-0.387	-	0.067	0.003	-0.904
35	0.877	-0.354	-	0.721	0.136	-0.364
40	0.917	-0.296	-	0.828	0.220	-0.228
45	0.934	-0.260	-	0.906	-0.088	0.125
50	0.925	-0.248	-	0.824	-0.262	0.024
55	0.877	-0.243	-	0.729	-0.130	-0.339

Таким чином, за особливостями вертикальної статистичної структури зонального вітру тропо-стратосферу середніх широт слід розділити на два підшари. Однак, якщо залучити до розгляду меридіональну компоненту, то від цього висновку доводиться відмовитися. Як видно з табл. 4.2, для

пояснення коваріаційної матриці меридіональної компоненти швидкості вітру необхідно розглядати не два, а три узагальнені фактори.

З табл. 4.2 випливає, що за особливостями вертикальної статистичної структури цієї компоненти тропо-стратосферу середніх широт, як і в тропічній атмосфері, слід розділити не на два, а на три підшари: 5-20 км, 20-35 км і 35-55 км. Перший з них виділяється за однорідною структурою вагів першого узагальненого фактора в пункті Уайт Сендз і другого – у пункті о-ва Воллоп. Підшар 20-35 км характеризується однорідною структурою вагів третього фактора в обох пунктах, а третій – однорідною структурою вагів другого узагальненого фактора в першому з розглянутих пунктів і першого узагальненого фактора – у другому.

Таким чином, факторний аналіз вертикальної статистичної структури компонент швидкості вітру дозволив виявити у шарі атмосфери 5-55 км три підшари, кожен з яких потребує самостійного розгляду для пояснення фізичних особливостей динамічних процесів, що в них розвиваються.

Іншим прикладом, який можна навести для ілюстрації можливостей факторного аналізу, є дослідження на його основі особливостей температурно-вологісного режиму території України [26]. Він може бути охарактеризований трьома параметрами: середньою за місяць відносною вологістю x_1 , середньомісячною температурою повітря x_2 та місячною кількістю опадів x_3 .

В якості вихідних даних використовувались перелічені характеристики за період 1900-1974 рр. для 14 довгорядних метеорологічних станцій України, рівномірно розміщених на території країни.

На основі вихідних даних були розраховані матриці коваріацій. Означені кліматичні характеристики визначаються одним фактором. В результаті реалізації наведеного алгоритму факторного аналізу (підр. 4.1) були розраховані матриці коваріацій та вектори вагів фактора відносної вологості p_1 , середньомісячної температури повітря p_2 і місячної кількості опадів p_3 , які представлені для центральних місяців весняного та літнього сезонів в табл. 4.3.

Аналіз даних табл. 4.3 вказує на те, що майже всі ваги фактора середньомісячної температури повітря мають протилежні знаки відносно вагів середньомісячних значень відносної вологості та місячних сум опадів. Це свідчить про те, що на території України у зазначений період у більшості випадків (за умови однакового значення фактора) від'ємні аномалії середньомісячної температури повітря супроводжуються додатними аномаліями середньомісячної відносної вологості повітря та місячної суми опадів, і навпаки.

Таблиця 4.3 – Ваги фактора температурно-вологісного режиму
і дисперсії залишків

Пункт	Квітень				Липень			
	Ваги			d_i^2	Ваги			d_i^2
	p_1	p_2	p_3		p_1	p_2	p_3	
Львів	0,46	-0,62	0,56	0,47	0,47	-0,62	0,50	0,48
Чернівці	0,69	-0,50	0,53	0,44	0,80	-0,65	0,61	0,33
Житомир	0,98	-0,47	0,64	0,38	0,96	-0,57	0,53	0,22
Київ	0,90	-0,52	0,78	0,22	0,94	-0,69	0,70	0,16
Умань	0,77	-0,47	0,75	0,33	0,78	-0,56	0,64	0,26
Кіровоград	0,98	-0,36	0,72	0,18	0,98	-0,62	0,58	0,07
Миколаїв	0,68	-0,45	0,97	0,20	0,91	-0,67	0,54	0,12
Херсон	0,51	-0,40	0,89	0,27	0,84	-0,64	0,69	0,08
Одеса	0,38	-0,46	0,97	0,21	0,94	-0,54	0,64	0,18
Симферопіль	0,70	-0,45	0,68	0,40	0,87	-0,55	0,48	0,30
Полтава	0,98	-0,38	0,64	0,08	0,98	-0,66	0,64	0,05
Харків	0,98	-0,46	0,71	0,07	0,99	-0,66	0,60	0,05
Дніпропетровськ	0,86	-0,38	0,69	0,09	0,91	-0,65	0,66	0,11
Луганськ	0,97	-0,34	0,56	0,10	0,99	-0,63	0,62	0,05

Значення вагів фактора свідчать про те, що над Україною температурно-вологісний режим має свої особливості і вони залежать від сезону року.

У липні циркуляційний режим території України більш однорідний. Наслідком чого і є однорідний температурно-вологісний режим території, що відбивається на однорідному розподілі вагів узагальненого фактора температури, вологості та опадів.

Крім того, як впливає з табл. 4.3, в структурі полів залишків проявляються визначені вище особливості температурно-вологісного режиму різних частин України. Дійсно, ряду пунктів, які відносяться до східної частини України, притаманні відносно невеликі значення метрики за вказані місяці. У липні проявляється друга область невеликих значень метрики, що виділяє райони, які відносяться до степової частини півдня України. Крім того, у липні дисперсії залишків суттєво менші, ніж у березні (винятком є ст. Львів), що свідчить про значно меншу мінливість метеорологічних величин влітку на більшості території України.

Отже, аналіз вагів узагальнених факторів дає можливість зробити висновок про необхідність поділу території України за однорідністю температурно-вологісного режиму у квітні на дві частини – західну та східну (лівобережну та правобережну), а у липні – на три частини: східну

(лівобережну), північно-західну та південно-західну степову. Такий характер режиму зволоження пояснюється особливостями переважних над країною циркуляційних атмосферних процесів.

Наведені результати свідчать про те, що узагальнений фактор є об'єктивною кількісною характеристикою температурно-вологісного режиму території.

Ще одним прикладом для з'ясування можливостей факторного аналізу в гідрометеорології є дослідження на основі цього апарату вертикальної статистичної структури поля температури у вільній атмосфері [11].

Для чотирьох станцій (табл. 4.4) на основі емпіричних даних були сформовані вертикальні профілі температури повітря на кожній п'ятикілометровій поверхні в шарі атмосфери 5-50 км і розраховані 10-вимірні матриці вертикальних коваріацій.

Таблиця 4.4 – Перелік пунктів ракетного та радіозондування атмосфери Західної півкулі, дані яких використовувались у дослідженні

Пункт	Координати		Об'єм вибірки
	широта північна	довгота західна	
1	09°20'	80°00'	117
2	28°27'	80°32'	183
3	37°50'	75°29'	127
4	58°44'	93°49'	166

На основі здобутих матриць розраховувалися ваги узагальнених факторів та дисперсії залишків. Ці результати представлені в табл. 4.5. та в табл. 4.6

Таблиця 4.5 – Частка сумарної дисперсії та сумарний внесок узагальнених факторів температури повітря

Пункт	Частка сумарної дисперсії. Фактори				Сумарний внесок. Фактори			
	1	2	3	4	1	2	3	4
1	16,48	15,87	13,80	12,05	1,65	1,59	1,38	1,21
2	24,31	17,25	12,90	14,82	2,43	1,73	1,29	1,48
3	24,45	19,98	15,01	12,84	2,45	2,00	1,50	1,28
4	24,69	21,07	17,94	—	2,47	2,11	1,79	—

Таблиця 4.6 – Ваги узагальнених факторів (P) температури повітря в тропосфері та стратосфері Західної півкулі

H, км	Пункт															
	1				2				3				4			
	P ₁	P ₂	P ₃	P ₄	P ₁	P ₂	P ₃	P ₄	P ₁	P ₂	P ₃	P ₄	P ₁	P ₂	P ₃	P ₄
5	0,11	0,05	-0,54	0,37	0,87	-0,02	0,08	0,07	0,87	-0,09	-0,04	0,07	0,10	-0,37	0,17	
10	-0,61	-0,21	0,01	-0,23	0,86	0,07	-0,16	-0,16	0,74	-0,02	0,02	-0,03	0,05	-0,22	0,79	
15	-0,58	0,68	-0,07	0,87	-0,80	-0,12	-0,14	0,11	-0,85	0,08	0,10	-0,00	0,10	0,16	0,78	
20	0,21	-0,08	0,64	-0,04	-0,40	0,51	-0,16	-0,42	-0,50	0,02	-0,07	0,72	0,32	0,40	0,58	
25	0,02	0,72	0,08	0,24	-0,15	0,32	0,13	-0,79	0,37	0,02	0,08	0,52	0,02	0,84	0,19	
30	0,35	-0,55	0,29	0,41	0,20	-0,28	0,03	-0,75	0,01	0,82	-0,14	0,31	-0,50	0,69	-0,23	
35	0,75	0,03	0,11	-0,03	-0,10	-0,76	-0,18	-0,09	-0,09	0,92	0,04	-0,06	-0,80	0,31	-0,27	
40	0,11	0,83	-0,03	-0,06	-0,13	-0,80	0,12	0,08	-0,13	0,67	0,18	-0,46	-0,85	0,23	-0,08	
45	0,70	-0,14	-0,16	-0,15	-0,08	-0,11	0,91	0,07	-0,13	-0,01	0,82	-0,33	-0,85	-0,31	-0,03	
50	0,18	-0,10	-0,74	-0,11	0,18	0,17	0,57	0,31	0,04	0,02	0,86	0,26	-0,01	-0,56	-0,16	

Як впливає з табл.4.5 та табл. 4.6, вертикальні профілі температури повітря визначаються 3 або 4 узагальненими факторами та їх кількість залежить від широти. Шар атмосфери 5-50 км за однорідними температурними умовами можна розділити на ряд шарів.

В приєкваторіальній зоні чітко виділяється шар 5-20 км. Що стосується більш високих рівнів, то спостерігається розкид вагів факторів по полю їх значень.

В субтропічних та високих широтах (пункти 2 і 4) врахування вагів перших двох факторів дають підстави розділити шар 5-50 км на три шари: 5-10 км; 15-25 км; 30-45 км.

В помірних широтах (пункт 3) врахування вагів першого та другого факторів чіткої картини розділення на шари не дала. Врахування вагів першого та третього узагальнених факторів дає підстави в помірних широтах виділити чотири шари: 5-10 км; 15-20 км; 25-40 км; 45-50 км. А це ще раз свідчить про те, що кожен з виділених шарів атмосфери потребує окремого подальшого вивчення для пояснення фізичних процесів, що в них розвиваються.

4.3 Запитання для самоперевірки

1. Які задачі можливо вирішувати за допомогою факторного аналізу?
2. У чому полягає відмінність факторного аналізу від компонентного?
3. Що називають «вагою» фактора?
4. Що являють собою залишки ν_i ?
5. Який фізичний сенс діагональних елементів матриці коваріації?
6. Які два методи використовують для визначення факторних вагів p_{ij} і дисперсій залишків d_i ?
7. Основні засади методу максимальної правдоподібності?
8. Функцією яких параметрів є функція правдоподібності?

5 КЛАСИФІКАЦІЯ І КЛАСТЕР

5.1 Базові теоретичні положення

5.1.1 Основи кластерного аналізу

У статистичних методах, здатних за допомогою швидкодіючих ПК обробляти багаторічний архів метеорологічних даних, міститься великий резерв поліпшення якості довгострокових прогнозів погоди (ДПП).

Гнучким інструментом виявлення прогностичних залежностей є такий прийом як кластерний аналіз. Головне **призначення кластерного аналізу** – кластеризація множини об'єктів і ознак, що досліджуються, на однорідні (у відповідному розумінні) групи або кластери. Це означає, що вирішується завдання класифікації даних і виявлення відповідної їх структури [19, 27].

Велика перевага кластерного аналізу при дослідженні випадкових величин в тому, що він дозволяє виконувати розбиття об'єктів (кластеризувати) не за одним параметром, а за деякою визначеною кількістю ознак.

Кластерний аналіз дозволяє розглядати досить великий об'єм вихідної інформації та різко скорочувати, стискати масиви даних, представляти їх компактно та наочно. Кластерний аналіз має важливе значення для 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 виконується до тих пір, поки не будуть досягнуті необхідні результати. При цьому кожен цикл тут може давати інформацію, яка може суттєво змінити спрямованість та підходи щодо подальшого застосування кластерного аналізу. Цей процес можна представити системою оберненого зв'язку.

Для поглибленого розуміння сенсу 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 повітряній оболонці Землі. Будемо називати ці вектори – **векторами ситуацій** 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 з двох задач:

- а) побудови правила, за яким можна визначити еталон і, таким чином, сформулювати 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.5.1) 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 точкою на рис. 5.1.

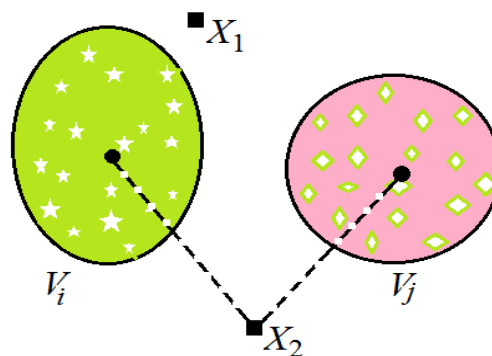


Рис. 5.1 – До поняття кластера

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.

1) **Випадок єдиного еталона в кластері** [27].

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

$$D_i = \|X - Z_i\| = \sqrt{(X - Z_i)'(X - Z_i)}, \quad (5.1)$$

де вертикальними дужками позначається норма вектора.

Очевидно, коли

$$X = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_j \\ \dots \\ x_m \end{pmatrix} \quad (5.2)$$

i

$$Z_i = \begin{pmatrix} z_{1i} \\ z_{2i} \\ \dots \\ z_{ji} \\ \dots \\ z_{mi} \end{pmatrix}, \quad (5.3)$$

Отримаємо:

$$D_i = \sqrt{(x_1 - z_{1i})^2 + (x_2 - z_{2i})^2 + \dots + (x_m - z_{mi})^2}. \quad (5.4)$$

Припустимо, що ми маємо M кластерів U_i ($i = \overline{1, m}$), кожен 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:

$$D_i^2 = |X - Z_i|^2 = (X - Z_i)'(X - Z_i). \quad (5.5)$$

У матричному співвідношенні (5.5) виконаємо операцію множення:

$$D_i^2 = X'X - Z_i'X - X'Z_i + Z_i'Z_i. \quad (5.6)$$

Оскільки $Z_i'X = X'Z_i$, маємо:

$$D_i^2 = X'X + Z_i'Z_i - 2X'Z_i \quad (5.7)$$

або

$$D_i^2 = X'X - 2(X'Z_i - \frac{1}{2}Z_i'Z_i). \quad (5.8)$$

Впровадимо позначення

$$d_i(X) = X'Z_i - \frac{1}{2}Z_i'Z_i. \quad (5.9)$$

Із формул (5.8) та (5.9) 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:

$$X \in U_i, \text{ якщо } d_i(X) > d_j(X), \forall j = \overline{1, m}; i \neq j. \quad (5.10)$$

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

$$d_i(X) = x_1 z_{1i} + x_2 z_{2i} + \dots + x_m z_{mi} - \frac{1}{2} \sum_{s=1}^m z_{si}^2. \quad (5.11)$$

2) **Випадок декількох еталонів у кластері** [27].

Нехай маємо M кластерів, але кожен з них характеризується деякою кількістю еталонів, тобто для i -го 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 -го кластера. Виберемо з них ту, яка є найменшою, нехай це буде відстань до l -го центра $z_i^{(l)}$, тобто

$$D_i = \min_l \left| X - z_i^{(K)} \right| \quad (K = 1, 2, \dots, l, \dots, p). \quad (5.12)$$

Теж саме зробимо для інших кластерів.

Тоді **розв'язувальним правилом** є:

$$X \in U_i, \text{ якщо } D_i < D_j \quad \forall j = \overline{1, m}; (j \neq i). \quad (5.13)$$

Це розв'язувальне правило можна записати й для функції відстані:

$$d_i(X) = \max_l \{X' z_i^{(K)} - \frac{1}{2} (z_i^{(K)})' z_i^{(K)}\}, \quad (K = 1, 2, \dots, l, \dots, p). \quad (5.14)$$

Воно має вигляд:

$$X \in U_i, \text{ якщо } d_i(X) > d_j(X) \quad \forall j = \overline{1, m}; \quad (j \neq i). \quad (5.15)$$

Найчастіше в алгоритмах класифікації застосовується **критерій мінімуму евклідової відстані (максимуму функції відстані)**. Але треба мати на увазі, що встановлення критерію 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 у масиві даних дозволять ідентифікувати кожен з центрів кластерів і, таким чином, сформувати кластери. Такі правила складають **міри схожості образів і критерії кластеризації**.

Розрізняють **метричні** та **неметричні** міри схожості [27]. Однією з метричних мір схожості може розглядатися вже знайома нам метрика – **евклідова відстань** $D = |X - Z|$. Іншою **мірою схожості є число Махаланобіса**

$$\Delta^2 = (X - \mu)' K^{-1} (X - \mu), \quad (5.16)$$

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 апріорі невідомі зазначені ймовірнісні характеристики.

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

$$S(X, Z) = \frac{X' Z}{|X| \cdot |Z|}. \quad (5.17)$$

У скалярній формі рівняння (5.17) має вигляд

$$S(X, Z) = \frac{x_1 z_1 + x_2 z_2 + \dots + x_m z_m}{\sqrt{x_1^2 + x_2^2 + \dots + x_m^2} \sqrt{z_1^2 + z_2^2 + \dots + z_m^2}}. \quad (5.18)$$

Вираз (5.18), очевидно, має сенс косинуса кута між векторами X і Z або коефіцієнта кореляції між випадковими величинами X і Z , що визначені відповідними векторами:

$$S(X, Z) = \cos \left(\hat{X, Z} \right) = r_{xz}. \quad (5.19)$$

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 *критерій кластеризації*.

Тут можливо використовувати два підходи [27]. 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 критерію є сума квадратів відхилів векторів від деякого середнього вектора кластера:

$$\Delta = \sum_{j=1}^{N_c} \sum_{X \in S_i} |X - m_i|^2, \quad (5.20)$$

де

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

N_c – кількість кластерів;

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

N_i – кількість векторів i -го кластера.

5.1.2 Універсальний адаптивний ітераційний метод кластерного аналізу («УАІМКА»)

У 2003 році на кафедрі фізики атмосфери та кліматології ОДЕКУ був розроблений алгоритм «УАІМКА» (Універсальний адаптивний ітераційний метод кластерного аналізу), який добре показав себе при кластеризації територій, порівнянних з територією України [22].

У методі «УАІМКА», в якості вихідної інформації виступає матриця $X = (x_{ij})_{n \times m}$, яка містить n векторів-рядків розмірності m , що характеризує статистичні ряди об'ємом m у n пунктах, які й повинні бути кластеризованими.

В якості апріорної інформації, на відміну від інших методів, задається тільки мінімальна кількість векторів τ , які можуть скласти кластер.

Ітераційний процес в алгоритмі «УАІМКА» складається з ряду кроків:

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

$$D = (D_{ij})_{n \times n}, \quad D_{ij} = \sqrt{\sum_{s=1}^n (x_{js} - x_{is})^2}. \quad (5.21)$$

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 напрямку збільшення. Отже, ми отримуємо матрицю D' вигляду (5.22), де здійснюється зміна нумерації елементів кожного рядка (p – номер стовпчика, в якому розташовується елемент кожного рядка матриці (5.22) в матриці $D(p = \overline{1, n})$).

Нехай, припустимо, що мінімальна кількість векторів, яка може скласти кластер, дорівнює τ .

Тоді аналізу піддається блок матриці (5.22), яка складається з перших $\tau = \theta$ стовпців (на першому етапі, наприклад, $\theta = 3$).

3-й крок. Далі проводиться порівняння евклідових відстаней $D_{ij}(i, j = \overline{1, n})$ з евклідовими відстанями $d_{i3}^{(p)}$. Якщо $D_{ij} \leq d_{i3}^{(p)} (i = \overline{1, n})$, то j -й вектор може розглядатися як потенційний центр кластера для i -го вектора (з урахуванням значення індексу p).

$$D' = \begin{pmatrix} 0 & d_{12}^{(p)} & d_{13}^{(p)} & d_{14}^{(p)} & \dots & d_{1n}^{(p)} \\ 0 & d_{22}^{(p)} & d_{23}^{(p)} & d_{24}^{(p)} & \dots & d_{2n}^{(p)} \\ 0 & d_{32}^{(p)} & d_{33}^{(p)} & d_{34}^{(p)} & \dots & d_{3n}^{(p)} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & d_{n2}^{(p)} & d_{n3}^{(p)} & d_{n4}^{(p)} & \dots & d_{nn}^{(p)} \end{pmatrix} \quad (5.22)$$

4-й крок. Для кожного j -го вектора визначається кількість i -х векторів (кількість входжень S_j), для котрих він може бути центром кластера.

5-й крок. Із загальної кількості j -х векторів виділяються ті, для яких виконується умова $S_j \geq \tau$ (у нашому випадку $S_j \geq 3$). Решта потенційних центрів кластерів, які не відповідають цій умові, ліквідується.

6-й крок. Визначається кількість j -х векторів, що залишилися, як центрів кластерів. Будемо вважати, що їх кількість дорівнює r ($j = r$).

7-й крок. Із загальної кількості m -х векторів устанавлюються ті, що потрапили в s -й (V_s) та g -й (V_g) кластери одночасно ($s, g = \overline{1, r}; i \neq s, g$). Після чого відбувається їхнє розведення за кластерами **за вирішальним правилом**: що $X_i \in V_s$, якщо $D_{is} < D_{ig}$, при цьому $S_g = S_g - 1$.

8-й крок. Знаходяться евклідові відстані \tilde{D}_{sg} між центрами s -го (V_s) та g -го (V_g) кластерів.

9-й крок. Визначаємо максимальну відстань між векторами, що потрапили в s -й кластер D_{ts} та g -й кластер D_{fg} ($t = \overline{1, S_s}; f = \overline{1, S_g}$). Будемо вважати це D_{fg} .

10-й крок. Евклідова відстань D_{fg} порівнюється з відстанями між центрами кластерів \tilde{D}_{sg} . Якщо $\tilde{D}_{sg} < D_{fg}$, то при $S_s \geq S_g$ ліквідується g -й кластер. Якщо $S_g > S_s$, то ліквідується s -й кластер.

11-й крок. Проводиться формування ряду попередніх центрів кластерів z_j ($j \leq \overline{1, m/3}$), де $m/3$ округляється до меншого цілого числа.

12-й крок. Відбувається розподіл по кластерах векторів вихідної вибірки відповідно до **вирішального правила**: $X \in V_j$, якщо $D_{xz_j} < D_{xz_k}$.

13-й крок. Визначається кількість векторів S_j , що увійшли в кожен j -й кластер V_j .

14-й крок. Розраховуються коефіцієнти кореляції $r_{z_j z_p}$ між центрами всіх кластерів.

15-й крок. Розраховується осереднене значення коефіцієнта кореляції \bar{r}_z по усіх виділених кластерах.

16-й крок. Здійснюється порівняння коефіцієнтів кореляції між центрами кластерів $r_{z_j z_p}$ з осередненим коефіцієнтом кореляції.

17-й крок. Якщо $r_{z_j z_p} > \bar{r}_z$, то центри z_j та z_p знаходяться у тісному кореляційному зв'язку і є підстава для об'єднання кластерів V_j та V_p . Після їх об'єднання ліквідується той центр кластера, якому належить менша кількість векторів (ліквідується центр z_j , якщо $S_j < S_p$), і підраховується кількість кластерів N_θ , що залишилися.

18-й крок. Повертаємося до 2 кроку за умови $\theta = \theta + 1$, і ітераційна процедура продовжується знову аж до 17 кроку включно.

19-й крок. Порівнюються кількість кластерів, які отримані на етапі θ та етапі $\theta - 1$. Якщо на етапі θ $N_{\theta-1} \geq 2, N_\theta < 2$, а також $N_{\theta-1} \geq 2, N \geq 2$ і $N_{\theta-1} > N_\theta$, то процедура завершується і здійснюється розподіл векторів вихідної вибірки за кластерами відповідно вирішального правила, зазначеного в кроці 12.

Якщо $N_{\theta-1} \geq 2, N_\theta \geq 2$ та $N_{\theta-1} \leq N_\theta$, то повертаємося до кроку 2 за умови $\theta = \theta + 1$.

20-й крок. Проводиться уточнення центрів кластерів. Як уточнений центр кластера розглядається осереднений вектор по усіх векторах, що входять у даний кластер.

5.2 Розв'язання типової задачі. Кластеризація гідрометеорологічної інформації

За інформацією 33 довгорядних станцій України про місячні кількості опадів за тридцятирічний період вимірювань (1966-1995 рр.) у лютому була сформована матриця вихідних даних порядку 33×30 .

Треба здійснити кластеризацію 33 векторів даної кліматичної характеристики та провести аналіз отриманих результатів.

Для реалізації кластерного аналізу до вказаних емпіричних даних був застосований «Універсальний адаптивний ітераційний метод кластерного аналізу» («УАІМКА»), алгоритм якого викладено в підрозділі 5.2.

Реалізація цього алгоритму виконувалась за допомогою комп'ютерної програми «Cla2003».

Призначення програми та особливості користування дивись у Додатку В.

Нижче наводяться результати реалізації алгоритму кластерного аналізу «УАІМКА», застосованого до кліматичних рядів місячної кількості опадів.

Це схематично представлено на рис 5.2.

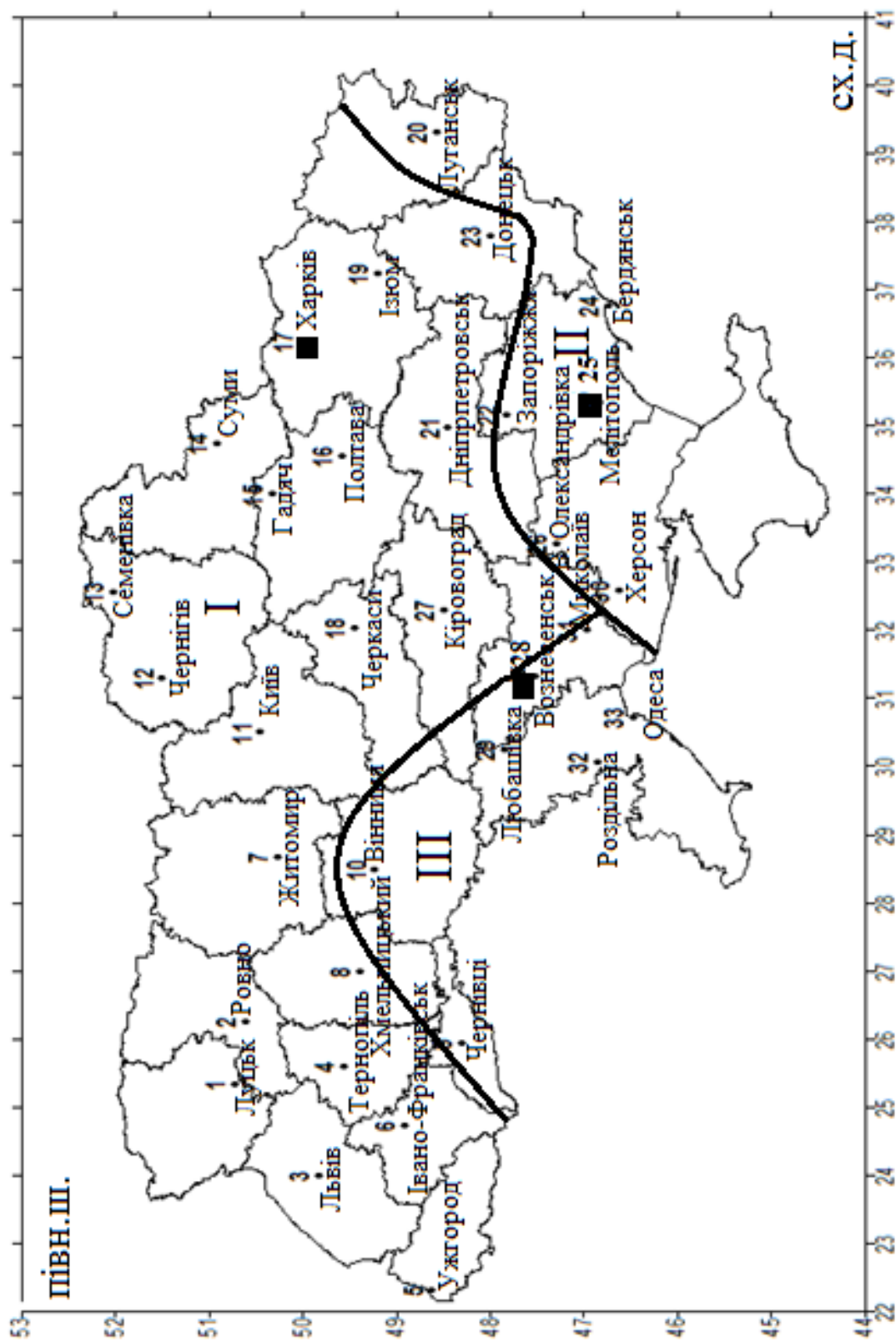


Рис.5.2 – Результати кластеризації місячної кількості опадів на території України у лютому

Розподіл векторів за кластерами

<p>центр 1-го кластера: 17-й вектор</p> <p>номери векторів, що входять в цей кластер:</p> <p>1 2 3 4 5 6 7 8 11 12 13 14 15 16 17 18 19 21 23 27</p>
<p>центр 2-го кластера: 25-й вектор</p> <p>номери векторів, що входять в цей кластер:</p> <p>20 22 24 25 26 30</p>
<p>центр 3-го кластера: 28-й вектор</p> <p>номери векторів, що входять в цей кластер:</p> <p>9 10 28 29 31 32 33</p>
Середній вектор 1-го кластера
<p>57,80 39,30 49,90 35,25 57,95 32,85 7,05 59,35 18,30 57,80</p> <p>3,80 68,70 47,60 28,00 24,85 46,25 24,25 39,25 33,55 42,25</p> <p>29,30 14,85 29,95 24,20 37,75 31,95 26,80 27,95 20,65 45,40</p>
Середній вектор 2-го кластера
<p>36,17 26,50 38,17 35,33 42,17 34,67 9,17 46,17 23,67 36,17</p> <p>2,83 42,00 45,67 61,33 23,33 52,17 13,83 36,17 26,00 77,17</p> <p>38,83 26,67 20,17 19,83 29,67 25,00 42,33 29,33 9,17 33,83</p>
Середній вектор 3-го кластера
<p>35,86 64,29 47,86 87,14 50,71 41,14 5,71 61,14 15,86 35,86</p> <p>6,00 49,57 63,29 22,86 10,71 34,86 27,00 22,29 61,71 50,00</p> <p>42,43 12,00 19,00 15,00 22,71 23,71 17,43 33,57 7,86 20,86</p>

В результаті реалізації алгоритму «УАІМКА» ми отримали розподіл 33 векторів за трьома кластерами.

Як видно з рис. 5.2, до першого кластера увійшли станції західної, північної, центральної та північно-східної України; до другого – східної та південно-східної; до третього – південно-західної України.

Такий розподіл, по-перше, відповідає загальному кліматичному розподілу опадів по території України у лютому і, по-друге, відділяє зони з різними характерними типами погоди при випадінні опадів по регіону, що досліджувався.

5.3 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?

5.4 Завдання для закріплення теми

- Розв'язати задачу кластеризації інформації про структуру полів опадів на території України у такій послідовності:
 - отримати у викладача вихідні дані для виконання практичної роботи (Додаток Б) та комп'ютерну програму «Cla2003.EXE» розроблену за алгоритмом «УАІМКА» (призначення програми та особливості користування дивись у Додатку В);
 - виконати на ПК необхідні розрахунки та отримати результати реалізації алгоритму «УАІМКА»;
 - за результатами розрахунків підготувати матеріали, необхідні для фізичного аналізу отриманих кластерів місячної кількості опадів для території України;
 - провести аналіз результатів кластеризації та зробити висновки щодо роботи у цілому.

6 РЕГРЕСІЙНІ МОДЕЛІ МЕТЕОРОЛОГІЧНОГО ПРОГНОЗУ

6.1 Базові теоретичні положення

Регресійний аналіз є одним з методів багатовимірної статистичного аналізу, за допомогою якого реалізується задача створення прогностичної моделі метеорологічного прогнозу у вигляді рівняння множинної лінійної регресії.[21, 27].

6.1.1 Прогностична модель лінійної множинної регресії

6.1.1.1 Постановка задачі

Раніше (в дисципліні «Методи обробки та аналізу гідрометеорологічної інформації») вже розглядалося рівняння лінійної регресії вигляду $\bar{y}(x) = ax + b$, яке відбиває кореляційний зв'язок між двома випадковими величинами X і Y . Але в метеорології дуже рідко реалізується зв'язок такої форми, оскільки стан того чи іншого процесу, а також тієї чи іншої метеорологічної величини, що його відбиває, зумовлюється великою кількістю **предикторів**. Наприклад, при побудові моделі для прогнозу місячної кількості опадів у деякому пункті на визначений термін треба враховувати, що опади зумовлюються, по-перше, полем атмосферного тиску біля земної поверхні над досить великим районом, по-друге, вони залежать від структури термобаричних полів на ряді рівнів вільної атмосфери. В число предикторів треба включити й деякі диференціальні характеристики метеорологічних полів, наприклад, адвекцію тепла, адвекцію вихору швидкості вітру та інші. Характер циклонічної діяльності, яка й зумовлює, головним чином, інтенсивність опадів, залежить великою мірою від тепловмісту поверхневих вод Північної Атлантики, особливо в їх частині, що називається енергоактивними зонами океану. Великий вплив здійснює також і положення меж морського льоду. Можна перелічувати й інші гідрометеорологічні процеси, які можуть відігравати роль предикторів.

Отже, якщо раніше йшлося про вплив однієї випадкової величини на іншу, то тепер необхідно будувати математичну модель, яка б відбивала впливи множини предикторів на прогнозовану величину (**предиктант**). В якості такої статистичної моделі може бути лінійна модель множинної регресії.

Побудова лінійного рівняння множинної регресії в якості моделі для

гідрометеорологічного прогнозування є суто математичною задачею за умови, що встановлено склад предикторів X_1, X_2, \dots, X_n . Цей склад предикторів називають *сукупністю потенційних предикторів*. Їх визначення – 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 зумовлювати значення предиктанта.

Може скластися враження, що чим більше предикторів вибрано для побудови прогностичної моделі, тим більш зумовленою 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 і, по-друге, значно збільшується у 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, котрі, 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»*. Деякі алгоритми «просіювання» предикторів будуть розглянуті у наступних підрозділах.

6.1.1.2 Структура моделі лінійної множинної регресії

При складанні прогнозів метеорологічних величин і явищ прогностичними центрами використовуються чисельні моделі двох типів: гідродинамічні та статистичні. Досить часто статистичні моделі метеорологічних прогнозів розробляються у вигляді лінійного множинного рівняння регресії.

Будемо вважати, що предиктанта Y й предиктори X_i ($i = \overline{1, n}$) центровані величини. Тоді рівняння лінійної регресії має вигляд:

$$Y = b_1 X_1 + b_2 X_2 + \dots + b_i X_i + \dots + b_n X_n, \quad (6.1)$$

де

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

Коефіцієнти регресії визначаються на основі статистичних сукупностей предиктанта і предикторів – факторів, які впливають на формування величини, що прогнозується.

Розглянемо методику отримання коефіцієнтів регресії спочатку для рівняння

$$Y = b_1 X_1 + b_2 X_2, \quad (6.2)$$

а потім отриманий результат розповсюдимо на рівняння (6.1).

Статистичні сукупності предиктанта Y і предикторів X_1, X_2 мають об'єми m . Для визначення коефіцієнтів b_1, b_2 застосуємо **метод найменших квадратів** [13]. Його метрика має вигляд:

$$\Delta^2 = \sum_{j=1}^m [y_j - (b_1 x_{1j} + b_2 x_{2j})]^2 = \min. \quad (6.3)$$

Необхідною і достатньою умовою мінімуму цієї метрики є

$$\frac{\partial \Delta^2}{\partial b_1} = 0 ; \quad \frac{\partial \Delta^2}{\partial b_2} = 0. \quad (6.4)$$

Застосування умови (6.4) приводить до системи неоднорідних лінійних рівнянь:

$$\begin{cases} b_1 \sum_{j=1}^m x_{1j}^2 + b_2 \sum_{j=1}^m x_{1j} x_{2j} = \sum_{j=1}^m y_j x_{1j} \\ b_1 \sum_{j=1}^m x_{1j} x_{2j} + b_2 \sum_{j=1}^m x_{2j}^2 = \sum_{j=1}^m y_j x_{2j} \end{cases}. \quad (6.5)$$

Для центрованих величин (предиктанта та предикторів) справедливими є співвідношення:

$$\begin{cases} \sum_{j=1}^m x_{1j}^2 = m\sigma_{x_1}^2; \\ \sum_{j=1}^m x_{2j}^2 = m\sigma_{x_2}^2; \\ \sum_{j=1}^m x_{1j}x_{2j} = m\sigma_{x_1}\sigma_{x_2}r_{x_1x_2}; \\ \sum_{j=1}^m y_jx_{1j} = m\sigma_y\sigma_{x_1}r_{yx_1}; \\ \sum_{j=1}^m y_jx_{2j} = m\sigma_y\sigma_{x_2}r_{yx_2}. \end{cases} \quad (6.6)$$

Підставивши рівняння (6.6) у систему (6.5), отримаємо:

$$\begin{cases} b_1\sigma_{x_1} + b_2\sigma_{x_2}r_{x_1x_2} = \sigma_y r_{yx_1} \\ b_1\sigma_{x_1}r_{x_1x_2} + b_2\sigma_{x_2} = \sigma_y r_{yx_2} \end{cases}, \quad (6.7)$$

де

σ_{x_1} , σ_y , σ_{x_2} – середні квадратичні відхили випадкових величин X_1 , Y та X_2 відповідно;

$r_{x_1x_2}$, r_{yx_1} , r_{yx_2} – парні коефіцієнти кореляції між вказаними випадковими величинами.

Розв'язок системи (6.7) має вигляд:

$$b_1 = \frac{r_{yx_1} - r_{yx_2}r_{x_1x_2}}{1 - r_{x_1x_2}^2} \frac{\sigma_y}{\sigma_{x_1}}; \quad (6.8)$$

$$b_2 = \frac{r_{yx_2} - r_{yx_1}r_{x_1x_2}}{1 - r_{x_1x_2}^2} \frac{\sigma_y}{\sigma_{x_2}}. \quad (6.9)$$

Систему рівнянь (6.7) можна записати у матричній формі:

$$R_x \sigma B = \sigma_y R_{yx}, \quad (6.10)$$

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 величин X_1, X_2 ;

B – вектор коефіцієнтів регресії, що підлягають визначенню.

Якщо матричне рівняння (6.10) є тотожнім системі (6.7), то воно є відображенням і системи нормальних рівнянь будь-якого порядку n [27]:

$$\begin{cases} b_1 \sigma_{x_1} + b_2 \sigma_{x_2} r_{x_1 x_2} + \dots + b_n \sigma_{x_n} r_{x_1 x_n} = \sigma_y r_{yx_1} \\ b_1 \sigma_{x_1} r_{x_2 x_1} + b_2 \sigma_{x_2} + \dots + b_n \sigma_{x_n} r_{x_2 x_n} = \sigma_y r_{yx_2} \\ \dots \\ b_1 \sigma_{x_1} r_{x_n x_1} + b_2 \sigma_{x_2} r_{x_n x_2} + \dots + b_n \sigma_{x_n} = \sigma_y r_{yx_n} \end{cases} \quad (6.11)$$

Розв'язки систем (6.7) та (6.11) у матричній формі мають однаковий вигляд:

$$B = \sigma_y \sigma^{-1} R_x^{-1} R_{yx}, \quad (6.12)$$

де

індекс (-1) позначає операцію обернення матриці;

$B = \{b_i\}_{n \times 1}$;

$R_x = \{r_{x_i x_j}\}_{n \times n}$;

$R_{yx} = \{r_{yx_i}\}_{n \times 1}$;

$\sigma = \{\sigma_{x_i}\}_{n \times n}$;

σ_y – скаляр, середнє квадратичне відхилення предиктанта.

Отже маємо n -вимірний вектор

$$B' = (b_1, b_2, \dots, b_n) \quad (6.13)$$

коефіцієнтів рівняння (6.1).

Якщо модель будується на вихідних (не центрованих) значеннях предиктанта й предикторів, то отримаємо модель вигляду:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_n X_n, \quad (6.14)$$

де

b_0 – вільний член, який дорівнює:

$$b_0 = \bar{y} - (b_1 \bar{x}_1 + b_2 \bar{x}_2 + \dots + b_n \bar{x}_n). \quad (6.15)$$

Вільний член b_0 характеризує ту частку середнього значення предиктанта, яка не вичерпується предикторами, тобто він характеризує визначену міру адекватності прогностичної моделі [27].

В деяких випадках з'являється можливість будувати регресійну модель на статистично незалежних (ортогональних) предикторах, за умови $r_{x_i x_j} = 0, \forall i, j = \overline{1, n}$.

Тоді в системі (6.11) залишаються лише діагональні елементи, а формули, що визначають коефіцієнти регресії, значно спрощуються [27]:

$$\begin{cases} b_1 = r_{yx_1} \frac{\sigma_y}{\sigma_{x_1}}; \\ b_2 = r_{yx_2} \frac{\sigma_y}{\sigma_{x_2}}; \\ \dots\dots\dots \\ b_n = r_{yx_n} \frac{\sigma_y}{\sigma_{x_n}}. \end{cases} \quad (6.16)$$

Прикладом такого випадку є побудова рівняння регресії вигляду:

$$Y = a_1 Z_1 + a_2 Z_2 + \dots + a_i Z_i + \dots + a_k Z_k \quad (6.17)$$

між деякою метеорологічною величиною Y і головними компонентами Z_i векторів предикторів, дисперсіями яких є власні значення λ_i матриці коваріації.

Тоді коефіцієнти a_i визначаються співвідношеннями:

$$\begin{cases} a_1 = r_{yz_1} \frac{\sigma_y}{\sqrt{\lambda_1}}; \\ a_2 = r_{yz_2} \frac{\sigma_y}{\sqrt{\lambda_2}}; \\ \dots\dots\dots \\ a_k = r_{yz_k} \frac{\sigma_y}{\sqrt{\lambda_k}}, \end{cases} \quad (6.18)$$

де

r_{yz_i} ($i = \overline{1, k}$) – парні коефіцієнти кореляції між предиктантом Y та відповідними головними компонентами Z_i .

Система нормальних рівнянь (6.11) не обов'язково розв'язується шляхом застосування матричного рівняння (6.12). У більшості випадків для цього використовується один з чисельних методів, наприклад, метод Гаусса.

6.1.1.3 Множинний коефіцієнт кореляції

На відміну від парного коефіцієнта кореляції, **множинний коефіцієнт** характеризує тісноту лінійного кореляційного зв'язку не з одним, а з цілою системою предикторів X_1, X_2, \dots, X_n , тобто є мірою адекватності регресійної прогностичної моделі [27]. Для того, щоб обґрунтувати це, а також отримати алгоритм розрахування множинного коефіцієнта кореляції [27], розглянемо рівняння регресії вигляду

$$Y = b_1 X_1 + b_2 X_2. \quad (6.19)$$

Запишемо метрику найменших квадратів

$$\Delta^2 = \sum_{j=1}^m S_{y \bullet x_1 x_2}^2 = \sum_{j=1}^m [y_j - (b_1 x_{1j} + b_2 x_{2j})]^2 \quad (6.20)$$

і розкриємо праву його частину

$$\sum_{j=1}^m S_{y \bullet x_1 x_2}^2 = \sum_{j=1}^m (y_j^2 - b_1 y_j x_{1j} - b_1 y_j x_{1j} - b_2 y_j x_{2j} - b_2 y_j x_{2j} + b_1^2 x_{1j}^2 + b_1 b_2 x_{1j} x_{2j} + b_1 b_2 x_{1j} x_{2j} + b_2^2 x_{2j}^2). \quad (6.21)$$

Якщо провести групування членів правої частини таким чином:

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

то прийдемо до формули

$$\sum_{j=1}^m S_{y \bullet x_1 x_2}^2 = \sum_{j=1}^m y_j^2 - b_1 \sum_{j=1}^m y_j x_{1j} - b_2 \sum_{j=1}^m y_j x_{2j}, \quad (6.22)$$

оскільки останні два члени дорівнюють нулю.

У рівнянні (6.22) будемо мати

$$\sum_{j=1}^m S_{y \bullet x_1 x_2}^2 = m \sigma_y^2 - b_1 m \sigma_{x_1} \sigma_y r_{yx_1} - b_2 m \sigma_{x_2} \sigma_y r_{yx_2}. \quad (6.23)$$

Розділимо обидві частини рівності (6.23) на загальний об'єм статистичної сукупності m і введемо позначення:

$$\sum_{j=1}^m S_{y \bullet x_1 x_2}^2 / m = \sigma_{y \bullet x_1 x_2}^2. \quad (6.24)$$

Очевидно *рівність (6.24) має сенс дисперсії нев'язки рівняння (6.19)*. Якщо її поділити на дисперсію предиктанта, то будемо мати *відносну дисперсію нев'язки прогностичного рівняння [27]*:

$$\tilde{\sigma}_{y \bullet x_1 x_2}^2 = \frac{\sigma_{y \bullet x_1 x_2}^2}{\sigma_y^2}. \quad (6.25)$$

Отже, з урахуванням формул (6.24) та (6.19), маємо:

$$\tilde{\sigma}_{y \bullet x_1 x_2}^2 = 1 - \frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1} r_{yx_2} r_{x_1 x_2}}{1 - r_{x_1 x_2}^2}. \quad (6.26)$$

Величина

$$R_{y \bullet x_1 x_2}^2 = \frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1} r_{yx_2} r_{x_1 x_2}}{1 - r_{x_1 x_2}^2} \quad (6.27)$$

і характеризує тісноту лінійного кореляційного зв'язку між предиктантом Y і предикторами X_1 і X_2 . Її називають *коефіцієнтом множинної детермінації*. Корінь квадратний з цієї величини

$$\sqrt{R_{y \bullet x_1 x_2}^2} = R_{y \bullet x_1 x_2} \quad (6.28)$$

є *коефіцієнтом множинної кореляції*.

З формул (6.26)-(6.28) випливає, що

$$R_{y \bullet x_1 x_2} = \sqrt{1 - \tilde{\sigma}_{y \bullet x_1 x_2}^2}. \quad (6.29)$$

Формула (6.29) є мірою адекватності прогностичної моделі.

Дійсно, $\tilde{\sigma}_{y \bullet x_1 x_2}^2$ характеризує ту частину предиктанта, котра не вичерпується моделлю. Чим вона менша, тим більш адекватною є модель. Якщо припустити, що $\tilde{\sigma}_{y \bullet x_1 x_2}^2 = 0$ (цього на практиці, звичайно, не спостерігається), то модель тотожно відбиває процес, що моделюється. У цьому випадку $R_{y \bullet x_1 x_2} = 1$. Навпаки, за умови $\tilde{\sigma}_{y \bullet x_1 x_2}^2 = \sigma_y^2$, $\tilde{\sigma}_{y \bullet x_1 x_2}^2 = 1$ (такий прогноз називають випадковим), а $R_{y \bullet x_1 x_2} = 0$ [27].

Отже множинний коефіцієнт кореляції змінюється в межах

$$0 < R_{y \bullet x_1 x_2} < 1. \quad (6.30)$$

Таким чином, чим більшим є значення множинного коефіцієнта кореляції, тим більш зумовленою є статистична модель.

Для того, щоб перейти від двовимірного рівняння (6.2) до n -вимірного рівняння лінійної регресії (6.1), створимо *розширену матрицю* вигляду (6.31), додаючи рядок і стовпець коефіцієнтів кореляції між предиктантом та предикторами до кореляційної матриці другого порядку:

$$\tilde{R} = \begin{pmatrix} 1 & r_{yx_1} & r_{yx_2} \\ r_{yx_1} & 1 & r_{x_1x_2} \\ r_{yx_2} & r_{x_2x_1} & 1 \end{pmatrix}. \quad (6.31)$$

Визначник такої матриці дорівнює:

$$|\tilde{R}| = (1 - r_{x_1x_2}^2) \left(1 - \frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1}r_{yx_2}r_{x_1x_2}}{1 - r_{x_1x_2}^2} \right). \quad (6.32)$$

Перший співмножник правої частини рівності (6.32) є визначником матриці кореляції $|R_x|$. Якщо на нього поділити обидві частини співвідношення (6.32), то будемо мати:

$$\frac{|\tilde{R}|}{|R_x|} = 1 - \frac{r_{yx_1}^2 + r_{yx_2}^2 - 2r_{yx_1}r_{yx_2}r_{x_1x_2}}{1 - r_{x_1x_2}^2}. \quad (6.33)$$

Другий член правої частини рівності (6.33) є *множинним коефіцієнтом детермінації* $R_{y \bullet x_1x_2}^2$. Тоді, як свідчить рівність (6.29), ліва частина (6.33) дорівнює *відносній залишковій дисперсії* $\tilde{\sigma}_{y \bullet x_1x_2}^2$. Отже з врахуванням цього маємо:

$$R_{y \bullet x_1x_2} = \sqrt{1 - \frac{|\tilde{R}|}{|R_x|}}. \quad (6.34)$$

Поширюючи отриманий результат на n предикторів, запишемо формулу для визначення *множинного коефіцієнта кореляції* [27]:

$$R_{y \bullet x_1x_2 \dots x_n} = \sqrt{1 - \frac{|\tilde{R}|}{|R_x|}}, \quad (6.35)$$

де

\tilde{R} – розширена $(n + 1)$ -вимірна матриця для кореляційної матриці R_x n -го порядку. Вона має вигляд:

$$\tilde{R} = \begin{pmatrix} 1 & r_{yx_1} & r_{yx_2} & \dots & r_{yx_n} \\ r_{yx_1} & 1 & r_{x_1x_2} & \dots & r_{x_1x_n} \\ r_{yx_2} & r_{x_2x_1} & 1 & \dots & r_{x_2x_n} \\ \dots & \dots & \dots & \dots & \dots \\ r_{yx_n} & r_{x_nx_1} & r_{x_nx_2} & \dots & 1 \end{pmatrix}. \quad (6.36)$$

У цьому випадку

$$\frac{|\tilde{R}|}{|R_x|} = \tilde{\sigma}_{y \bullet x_1 x_2 \dots x_n}^2 \quad (6.37)$$

і рівність (6.35) можна записати так:

$$R_{y \bullet x_1 x_2 \dots x_n} = \sqrt{1 - \tilde{\sigma}_{y \bullet x_1 x_2 \dots x_n}^2}. \quad (6.38)$$

Множинний коефіцієнт кореляції $R_{y \bullet x_1 x_2 \dots x_n}$ характеризує тісноту кореляційного зв'язку між предиктантом Y та системою предикторів X_1, X_2, \dots, X_n , тобто є мірою адекватності моделі метеорологічного прогнозу $Y = f(X_1, X_2, \dots, X_n)$, яка має вигляд лінійного рівняння множинної регресії [27]. Чим ближчим до одиниці є множинний коефіцієнт кореляції, тим більш вірогідною буде регресійна модель метеорологічного прогнозу.

6.1.1.4 Частинний коефіцієнт кореляції

В дисципліні «Методи обробки та аналізу гідрометеорологічної інформації» ми розглядали парні коефіцієнти кореляції, які, як відомо, визначають лінійний кореляційний зв'язок між двома випадковими величинами (наприклад X та Y). При цьому ми вважали, що на випадкову величину Y діють й інші випадкові величини. Але нас не цікавило питання,

саме які це величини. В дійсності, предиктори, як правило, статистично зв'язані між собою і при побудові статистичних моделей такі зв'язки треба враховувати. Це можна здійснити за допомогою частинних коефіцієнтів кореляції.

Дамо визначення частинного коефіцієнта кореляції й розглянемо алгоритм його оцінки [27].

Припустимо, що на випадкову величину Y діють дві випадкові величини X_1 та X_2 . **Частинним коефіцієнтом кореляції між випадковими величинами Y та X_1 ($r_{yx_1 \bullet x_2}$)** називають коефіцієнт кореляції між ними за умови, що вплив іншої випадкової величини X_2 на Y вже є врахованим. Таким же чином визначається частинний коефіцієнт кореляції $r_{yx_2 \bullet x_1}$.

Будемо вважати, що нам відома матриця кореляцій

$$R_x = \begin{pmatrix} 1 & r_{x_1 x_2} \\ r_{x_2 x_1} & 1 \end{pmatrix} \quad (6.39)$$

і вектор парних коефіцієнтів кореляцій між Y і X_1 та X_2

$$R_{yx} = \begin{pmatrix} r_{yx_1} \\ r_{yx_2} \end{pmatrix}. \quad (6.40)$$

На їх основі сформуємо **розширену матрицю кореляцій**

$$\tilde{R} = \begin{pmatrix} 1 & r_{yx_1} & r_{yx_2} \\ r_{yx_1} & 1 & r_{x_1 x_2} \\ r_{yx_2} & r_{x_2 x_1} & 1 \end{pmatrix}. \quad (6.41)$$

Вона утворюється з матриці R_x шляхом додавання до неї рядка та стовпця, що складаються з координат вектора R_{yx} .

На основі матриці (6.41) розрахуємо **мінори** $|R_x|$, \ddot{A}_{yx_i} , $\ddot{A}_{yx_i}^-$ ($i=1,2$). Мінори \ddot{D}_{yx_i} складаються таким чином: стовпець, на першому місці котрого розташовується парна кореляція r_{yx_i} , переставляється на перше місце, а на його місце ставиться перший стовпець і, після цього,

викреслюють перші рядок і стовпець.

Очевидно

$$D_{yx_1} = r_{yx_1} - r_{yx_2} r_{x_1 x_2}; \quad (6.42)$$

$$D_{yx_2} = r_{yx_2} - r_{yx_1} r_{x_1 x_2}. \quad (6.43)$$

Означення мінора $D_{yx_1}^-$ має такий сенс: це мінор визначника $|\tilde{R}|$, який не містить парної кореляції r_{yx_1} . Очевидно ми його будемо мати, якщо викреслимо з мінора $|\tilde{R}|$ рядок і стовпець, що містять цю парну кореляцію.

Отже

$$D_{yx_1}^- = \begin{vmatrix} 1 & r_{yx_2} \\ r_{yx_2} & 1 \end{vmatrix} = 1 - r_{yx_2}^2; \quad (6.44)$$

$$D_{yx_2}^- = \begin{vmatrix} 1 & r_{yx_1} \\ r_{yx_1} & 1 \end{vmatrix} = 1 - r_{yx_1}^2. \quad (6.45)$$

Частинні коефіцієнти кореляції визначаються таким чином:

$$r_{yx_1 \bullet x_2} = \frac{D_{yx_1}}{\sqrt{|R_x| D_{yx_1}^-}} = \frac{r_{yx_1} - r_{yx_2} r_{x_1 x_2}}{\sqrt{(1 - r_{x_1 x_2}^2)(1 - r_{yx_2}^2)}}; \quad (6.46)$$

$$r_{yx_2 \bullet x_1} = \frac{D_{yx_2}}{\sqrt{|R_x| D_{yx_2}^-}} = \frac{r_{yx_2} - r_{yx_1} r_{x_1 x_2}}{\sqrt{(1 - r_{x_1 x_2}^2)(1 - r_{yx_1}^2)}}. \quad (6.47)$$

Виникає питання, яка суттєва інформація міститься в частинних коефіцієнтах кореляції? Щоб відповісти на нього, розглянемо такий приклад. Нехай парні коефіцієнти кореляції між випадковими величинами мають такі значення: $r_{yx_1} = 0.70$; $r_{yx_2} = 0.90$; $r_{x_1 x_2} = 0.80$. Що можна сказати про ці випадкові величини? Звісно те, що випадкова величина Y характеризується дуже тісними кореляційними зв'язками і з величиною

X_1 , і з величиною X_2 . Але треба звернути увагу на те, що дві останні випадкові величини теж зв'язані дуже тісним кореляційним зв'язком між собою. Отже, щоб визначити, яка з величин X_i дійсно чинить вплив на величину Y , треба розрахувати частинні коефіцієнти кореляції за допомогою формул (6.46) і (6.47). Розрахунки дають такі значення: $r_{yx_1 \bullet x_2} = -0.08$; $r_{yx_2 \bullet x_1} = 0.79$. Таким чином, в дійсності на випадкову величину Y впливає випадкова величина X_2 . Кореляційний зв'язок Y з величиною X_1 , якщо врахувати її зв'язок з величиною X_2 , є не тільки незначним, але навіть має обернений характер.

Щоб розповсюдити отриманий алгоритм розрахунків частинних коефіцієнтів кореляції на n змінних, треба побудувати розширену матрицю

$$\tilde{R} = \begin{pmatrix} 1 & r_{yx_1} & r_{yx_2} & r_{yx_3} & \dots & r_{yx_k} & \dots & r_{yx_n} \\ r_{yx_1} & 1 & r_{x_1x_2} & r_{x_1x_3} & \dots & r_{x_1x_k} & \dots & r_{x_1x_n} \\ r_{yx_2} & r_{x_2x_1} & 1 & r_{x_2x_3} & \dots & r_{x_2x_k} & \dots & r_{x_2x_n} \\ r_{yx_3} & r_{x_3x_1} & r_{x_3x_2} & 1 & \dots & r_{x_3x_k} & \dots & r_{x_3x_n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ r_{yx_k} & r_{x_kx_1} & r_{x_kx_2} & r_{x_kx_3} & \dots & 1 & \dots & r_{x_kx_n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ r_{yx_n} & r_{x_nx_1} & r_{x_nx_2} & r_{x_nx_3} & \dots & r_{x_nx_k} & \dots & 1 \end{pmatrix} \quad (6.48)$$

і на її основі визначити мінори $|R_x|$, D_{yx_k} , $D_{yx_k}^-$ ($k=1,2,\dots,n$) за тими ж правилами, як це було зроблено у попередньому випадку для матриці третього порядку.

Отримаємо:

$$r_{yx_k \bullet x_1, x_2, \dots, x_{k-1}, x_{k+1}, \dots, x_n} = \frac{D_{yx_k}}{\sqrt{|R_x| D_{yx_k}^-}}. \quad (6.49)$$

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

Частинні коефіцієнти кореляції використовуються в ряді методів об'єктивного добору предикторів, які будуть розглянуті в пункті 6.1.3.

6.1.2 Перевірка гіпотези про статистичну значущість коефіцієнтів регресійної моделі

Як зазначалося у попередньому пункті, на основі метода найменших квадратів вектор оцінок коефіцієнтів лінійної множинної регресії визначається матричним рівнянням (6.12). Але цю задачу можна реалізувати і за допомогою інформаційної матриці Фішера F [21, 27], тобто

$$B = F^{-1}XY. \quad (6.50)$$

Матриця Фішера є добутком матриць X та Y :

$$F = XY, \quad (6.51)$$

де

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

Y – вектор-стовпець центрованих значень предиктанта.

З іншого боку, обернена інформаційна матриця Фішера визначається як:

$$F^{-1} = \frac{1}{\sigma_y^2} \begin{pmatrix} \sigma^2(b_1) & k_{b_1b_2} & \dots & k_{b_1b_i} & \dots & k_{b_1b_n} \\ k_{b_2b_1} & \sigma^2(b_2) & \dots & k_{b_2b_i} & \dots & k_{b_2b_n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ k_{b_ib_1} & k_{b_ib_2} & \dots & \sigma^2(b_i) & \dots & k_{b_ib_n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ k_{b_nb_1} & k_{b_nb_2} & \dots & k_{b_nb_i} & \dots & \sigma^2(b_n) \end{pmatrix}, \quad (6.52)$$

де

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

$k_{b_ib_n}$ – коваріація між коефіцієнтами b_i та b_n .

Для простоти подальших розрахунків позначимо обернену інформаційну матрицю таким чином:

$$F^{-1} = C = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1i} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2i} & \dots & c_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ c_{i1} & c_{i2} & \dots & c_{ii} & \dots & c_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ c_{n1} & c_{n2} & \dots & c_{ni} & \dots & c_{nn} \end{pmatrix}. \quad (6.53)$$

Діагональні елементи матриці C дорівнюють:

$$c_{ii} = \frac{\sigma^2(b_i)}{\sigma_y^2}, \quad (i = \overline{1, n}). \quad (6.54)$$

На основі рівняння (6.54)

$$\sigma(b_i) = \sqrt{c_{ii}} \cdot \sigma_y. \quad (6.55)$$

Визначимо добуток матриць X та Y :

$$XY = \begin{pmatrix} \sum_{j=1}^m x_{1j} y_j \\ \sum_{j=1}^m x_{2j} y_j \\ \vdots \\ \sum_{j=1}^m x_{nj} y_j \end{pmatrix} = m \sigma_y \sigma R_{yx}, \quad (6.56)$$

де

m – об'єм вибірок;

σ_y – середній квадратичний відхил предиктанта;

σ – діагональна матриця середніх квадратичних відхилів предикторів;

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

Компонуючи рівняння (6.50), (6.53) та (6.56), отримаємо:

$$B = m\sigma_y C \sigma R_{yx}. \quad (6.57)$$

Розглянувши рівняння (6.12) та (6.57), маємо:

$$m\sigma_y C \sigma R_{yx} = \sigma_y \sigma^{-1} R_x^{-1} R_{yx}. \quad (6.58)$$

Тоді

$$C = \frac{1}{m} \sigma^{-1} R_x^{-1} \sigma^{-1}. \quad (6.59)$$

Розраховуючи обернену матрицю R_x^{-1} , можна знайти матрицю C і відповідні їй діагональні елементи c_{ii} ($i = \overline{1, n}$).

Далі за формулою (6.55), замінюючи σ_y на S_y , отримаємо:

$$\sigma(b_i) = \sqrt{c_{ii}} S_y, \quad (6.60)$$

де

S_y – середній квадратичний відхил відтвореності процесу.

Дисперсія відтвореності – це є дисперсія нев'язки рівняння:

$$\hat{\sigma}_y^2 = \frac{1}{k} \sum_{j=1}^k [y_j - (b_1 x_{1j} + b_2 x_{2j} + \dots + b_n x_{nj})]^2. \quad (6.61)$$

Її можна розрахувати шляхом проведення k експериментів з побудованою моделлю.

Довірчий інтервал для коефіцієнтів регресії

$$\Delta b_i = \sigma(b_i) t_{kp} \left(\frac{\alpha}{2}, \nu \right), \quad (6.62)$$

де

t_{kp} – критичне значення критерію Стюдента для рівня значущості $\frac{\alpha}{2}$ та числа степенів вільності $\nu = m - n$ (m – об'єм вибірок; n – кількість коефіцієнтів регресії).

Гіпотеза H_1 про статистичну значущість коефіцієнта регресії не відхиляється, якщо

$$|b_i| > \Delta b_i. \quad (6.63)$$

При цьому треба переконатися, що коефіцієнти кореляції

$$r_{b_i b_j} = \frac{k_{b_i b_j}}{\sigma(b_i) \sigma(b_j)}, \quad (6.64)$$

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:

$$t = \frac{|b_i|}{\sigma(b_i)}. \quad (6.65)$$

Гіпотеза H_1 про статистичну значущість коефіцієнтів регресії не відхиляється, якщо $t > t_{кр}(\alpha, \nu)$.

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 є близьким до його критичного значення.

6.1.3 Методи відбору статистично значущих предикторів

Задача створення прогностичної моделі вигляду (6.1) чи (6.14) полягає в тому, щоб на основі статистичних сукупностей предиктанта Y і системи предикторів $X_1, X_2, \dots, X_i, \dots, X_n$ отримати значення коефіцієнтів регресії. В модель повинні бути включені тільки ті предиктори, які з заданою ймовірністю чинять найбільший вплив на формування предиктанта Y . Визначений таким чином склад предикторів називають **оптимальним**, а предиктори, що входять до цього складу – *статистично значущими*.

Процедури, які дають можливість визначити зі складу потенційних статистично значущі предиктори, називають *процедурами «просіювання»* предикторів. Існує декілька методів, що дозволяють розв'язати цю задачу (за допомогою множинного коефіцієнта кореляції, частинного коефіцієнта кореляції, методом «включення» або методом покрокової регресії) [21, 27].

6.1.3.1 *Відбір статистично значущих предикторів на основі множинного коефіцієнта кореляції*

Основою даного методу є вектор парних коефіцієнтів кореляцій між предиктантом і предикторами. Аналіз координат цього вектора дає можливість визначити той предиктор, парний коефіцієнт кореляції якого з предиктантом є найбільшим за абсолютною величиною. Він зараховується до складу оптимальних предикторів і йому задають перший номер. Інші предиктори перенумеровуються. Після цього розраховуються множинні коефіцієнти кореляції $R_{y \bullet x_1 x_j}$ ($j = \overline{2, n}$). Очевидно, що в якості другого оптимального предиктора повинен виступати той s -й предиктор, для якого $R_{y \bullet x_1 x_j} = \max_j R_{y \bullet x_1 x_j}$ ($j = \overline{2, n}$). Йому приписується другий номер, а інші предиктори 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 -й предиктор 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 предикторів за третім номером, а інші предиктори знову перенумеровуються. Виникає питання, до яких пір треба продовжувати цю процедуру? Справа у тому, що з включенням нового предиктора до складу статистично значущих, значення відповідного множинного коефіцієнта кореляції збільшується, і при включенні деякого k -го предиктора він переходить до стану насичення. Характер зміни множинного коефіцієнта кореляції на кожному кроці розрахунків схематично зображується на рис. 6.1.

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:

$$F = \frac{R_{y \bullet x_1 x_2 \dots x_k x_{k+1}}^2 - R_{y \bullet x_1 x_2 \dots x_k}^2}{1 - R_{y \bullet x_1 x_2 \dots x_k}^2} \times \frac{m - k + 1}{k - 1}. \quad (6.66)$$

Зазначена гіпотеза не відхиляється за умови $F < F_{кр}(\alpha, \nu_1, \nu_2)$, де $\nu_1 = k - 1$, а $\nu_2 = m - k + 1$, 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 $R_{y \bullet x_1 x_2 \dots x_k x_{k+1}}$ означає, що включення до складу факторів прогностичної моделі предиктора X_{k+1} не приводить до суттєвого зменшення остаточної дисперсії порівняно з випадком, коли модель базується на k попередньо вибраних предикторах.

6.1.3.2 «Просіювання» предикторів за методом «включення»

Метод «включення» [27] 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:

$$\tilde{R} = \begin{pmatrix} 1 & r_{12} & r_{13} & \dots & r_{1n} & r_{01} \\ r_{21} & 1 & r_{23} & \dots & r_{2n} & r_{02} \\ r_{31} & r_{32} & 1 & \dots & r_{3n} & r_{03} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r_{n1} & r_{n2} & r_{n3} & \dots & 1 & r_{0n} \\ r_{01} & r_{02} & r_{03} & \dots & r_{0n} & 1 \end{pmatrix}. \quad (6.67)$$

На відміну від матриці (6.48), розширення матриці кореляції $R_x = (r_{ij})_{n \times n}$ між потенційними предикторами відбулось шляхом додавання рядка і стовпця кореляцій між предиктантом і предикторами не на перших, як у матриці (6.48), а на останніх місцях.

В якості першого предиктора, як було зроблено у процедурі відбору предикторів за допомогою множинного коефіцієнта кореляції, приймається той k -й 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 (6.67) розраховується матриця частинних коефіцієнтів кореляції між

предиктантом і предикторами за умови, що вплив k -го предиктора, вже враховано. Ця матриця має вигляд:

$$R^{(1)} = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & r_{23.1} & r_{24.1} & \dots & r_{2n.1} & r_{02.1} \\ 0 & r_{32.1} & 1 & r_{34.1} & \dots & r_{3n.1} & r_{03.1} \\ 0 & r_{42.1} & r_{43.1} & 1 & \dots & r_{4n.1} & r_{04.1} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & r_{n2.1} & r_{n3.1} & r_{n4.1} & \dots & 1 & r_{0n.1} \\ 0 & r_{02.1} & r_{03.1} & r_{04.1} & \dots & r_{0n.1} & 1 \end{pmatrix}. \quad (6.68)$$

В матриці (6.68) всі елементи першого рядка і першого стовпця, окрім першого елемента, дорівнюють нулю.

Дійсно, наприклад, другий елемент першого рядка є частинним коефіцієнтом кореляції $r_{12.1}$. Якщо його розрахувати за формулою (6.47), в якій $y = x_1$, то отримаємо:

$$r_{12.1} = \frac{r_{12} - r_{11}r_{12}}{\sqrt{(1 - r_{12}^2)(1 - r_{11}^2)}}. \quad (6.69)$$

Оскільки $r_{11} = 1$, тоді приходимо до невизначеності вигляду $0/0$. Якщо здійснити граничний перехід в рівнянні (6.69) та застосувати *правило Лопіталя*, тоді будемо мати:

$$\begin{aligned} \lim_{r_{11} \rightarrow 1} r_{12.1} &= \lim_{r_{11} \rightarrow 1} \frac{\frac{\partial}{\partial r_{11}}(r_{12} - r_{11}r_{12})}{\frac{\partial}{\partial r_{11}} \left[\sqrt{(1 - r_{12}^2)(1 - r_{11}^2)} \right]} = \\ &= \lim_{r_{11} \rightarrow 1} \frac{-r_{12} \sqrt{(1 - r_{12}^2)(1 - r_{11}^2)}}{-2r_{11}(1 - r_{12}^2)} = 0. \end{aligned}$$

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 за абсолютною величиною. Предиктор, що відноситься до цього коефіцієнту кореляції, включається до складу статистично значущих предикторів. Йому приписується другий номер. 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 (6.68), тоді матриця $R^{(2)}$ приймає вигляд:

$$R^{(2)} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & r_{34.12} & \dots & r_{3n.12} & r_{03.12} \\ 0 & r_{43.12} & 1 & \dots & r_{4n.12} & r_{04.12} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & r_{n3.12} & r_{n4.12} & \dots & 1 & r_{0n.12} \\ 0 & r_{03.12} & r_{04.12} & \dots & r_{0n.12} & 1 \end{pmatrix}. \quad (6.70)$$

Третій предиктор визначається шляхом порівняння частинних коефіцієнтів кореляції між предиктантом і предикторами, які містяться в останньому стовпці (рядку) матриці (6.70) 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

$$t = \frac{|r_{0(s+1).12\dots s}|}{\sigma_r}. \quad (6.71)$$

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

6.1.3.3 «Просіювання» предикторів за методом покрокової регресії

Метод покрокової регресії отримав таку назву тому, що сукупність статистично значущих предикторів відбирається зі сукупності потенційних предикторів за допомогою деякої кількості послідовних кроків [21, 27].

1. Спочатку, на основі об'єму 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 кожним предиктором окремо:

$$R_{yx} = \begin{pmatrix} r_{yx_1} \\ r_{yx_2} \\ \vdots \\ r_{yx_i} \\ \vdots \\ r_{yx_k} \\ \vdots \\ r_{yx_n} \end{pmatrix} \quad (6.72)$$

за формулою

$$r_{yx_i} = \frac{\sum_{j=1}^m y_i x_{ij}}{m}, \quad (i = \overline{1, n}). \quad (6.73)$$

2. Аналіз координат вектора R_{yx} та визначення найбільшого за модулем коефіцієнта кореляції між предиктантом та предиктором. Нехай $|r_{yx_k}| = \max$. Тоді $X_k \rightarrow X_1$, тобто X_k визначається як перший предиктор із системи оптимальних предикторів. Він ставиться на перше місце, а ті предиктори, що залишилися, перенумеровуються.

3. Знаходиться оцінка рівняння регресії першого кроку предиктанта Y на перший предиктор

$$\hat{y}_j^{(1)} = a_1^{(1)} X_1 \quad (6.74)$$

(номер в дужках позначає номер кроку).

Коефіцієнт рівняння регресії першого кроку визначається за формулою

$$a_1^{(1)} = \frac{\sum_{j=1}^m y_j x_{1j}}{\sum_{j=1}^m x_{1j}^2}. \quad (6.75)$$

Далі розраховуються всі нев'язки першого кроку між значеннями предиктанта y_j та його значеннями за допомогою моделі (6.74):

$$\varepsilon_{1j} = y_j - \hat{y}_j^{(1)}, \quad \forall \quad j = \overline{1, m}. \quad (6.76)$$

Знаходять коефіцієнти кореляції $r_{\varepsilon_1 x_s}$ ($s = \overline{2, n}$) між нев'язками першого кроку та всіма предикторами, починаючи з предиктора за номером 2, за формулою:

$$r_{\varepsilon_1 x_s} = \frac{\sum_{j=1}^m \tilde{\varepsilon}_{1j} x_{sj}}{\sqrt{\sum_{j=1}^m \tilde{\varepsilon}_{1j}^2 \sum_{j=1}^m x_{sj}^2}}, \quad (6.78)$$

де

$\tilde{\varepsilon}_{1j}$ – центровані значення нев'язок.

Результат розрахунків – система коефіцієнтів кореляції

$$r_{\varepsilon_1 x_2}, r_{\varepsilon_1 x_3}, \dots, r_{\varepsilon_1 x_p}, \dots, r_{\varepsilon_1 x_n}, \quad (6.79)$$

які мають сенс частинних коефіцієнтів кореляції. Далі проводиться їх аналіз з метою вибору між ними найбільшого за модулем.

Нехай $|r_{\varepsilon_1 x_p}| = \max$, тоді $X_p \rightarrow X_2$, тобто X_2 вибирається як другий оптимальний предиктор. Інші $n-2$ предиктори перенумеровуються і

відбувається перехід до другого кроку.

4. За допомогою МНК будується рівняння регресії другого кроку:

$$\hat{y}_j^{(2)} = a_1^{(2)} X_1 + a_2^{(2)} X_2 \quad (6.80)$$

і розраховуються нев'язки другого кроку:

$$\varepsilon_{2j} = y_j - \hat{y}_j^{(2)}, \quad \forall j = \overline{1, m}. \quad (6.81)$$

Це означає, що зі значень предиктанта вилучаються впливи першого X_1 і другого X_2 предикторів. Після цього розраховуються коефіцієнти кореляції $r_{\varepsilon_2 x_s}$ ($s = \overline{3, n}$), аналіз яких дає підставу для визначення третього оптимального предиктора X_3 і т.д. Процедура триває до тих пір, доки отримані після деякого l -го кроку всі частинні коефіцієнти кореляції $r_{\varepsilon_{l+1} x_s}$ ($s = \overline{l+1, n}$) втрачають статистичну значущість. Гіпотеза H_0 про це перевіряється за допомогою критерію Стюдента

$$t = \frac{|r_{\varepsilon_{l+1} x_s}|}{\sigma_{r_{\varepsilon_{l+1} x_s}}}, \quad (6.82)$$

де

$$\sigma_{r_{\varepsilon_{l+1} x_s}} = \frac{1 - r_{\varepsilon_{l+1} x_s}^2}{\sqrt{m - 1}}. \quad (6.83)$$

Система нормальних рівнянь МНК для кожного k -го кроку, починаючи з другого, для визначення коефіцієнтів рівняння регресії

$$\hat{y}_j^{(k)} = a_1^{(k)} X_1 + a_2^{(k)} X_2 + \dots + a_{k-1}^{(k)} X_{k-1} + a_k^{(k)} X_k \quad (6.84)$$

має вигляд

$$\left\{ \begin{array}{l} a_1^{(k)} \sum_{j=1}^m x_{1j}^2 + a_2^{(k)} \sum_{j=1}^m x_{1j} x_{2j} + \dots a_{k-1}^{(k)} \sum_{j=1}^m x_{1j} x_{(k-1)j} + a_k^{(k)} \sum_{j=1}^m x_{1j} x_{kj} = \sum_{j=1}^m y_j x_{1j}; \\ a_1^{(k)} \sum_{j=1}^m x_{1j} x_{2j} + a_2^{(k)} \sum_{j=1}^m x_{2j}^2 + \dots a_{k-1}^{(k)} \sum_{j=1}^m x_{2j} x_{(k-1)j} + a_k^{(k)} \sum_{j=1}^m x_{2j} x_{kj} = \sum_{j=1}^m y_j x_{2j}; \\ \dots \dots \dots \dots \dots \dots \dots \\ a_1^{(k)} \sum_{j=1}^m x_{1j} x_{kj} + a_2^{(k)} \sum_{j=1}^m x_{2j} x_{kj} + \dots a_{k-1}^{(k)} \sum_{j=1}^m x_{kj} x_{(k-1)j} + a_k^{(k)} \sum_{j=1}^m x_{kj}^2 = \sum_{j=1}^m y_j x_{kj}; \end{array} \right. \quad (6.85)$$

(індекс k пробігає всі значення з послідовності $k: 2, 3, \dots, s, \dots, l$. Якщо $k = s$, то всі значення $X_{(s+1)l}$, $X_{(s+2)l}$, ... вважаються такими, що дорівнюють нулю.

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 оптимальних предикторів:

$$Y = a_1 X_1 + a_2 X_2 + \dots + a_l X_l. \quad (6.86)$$

Отже, при виконанні покрокової процедури, на кожному кроці, як допоміжний контрольний параметр, розраховується відповідний множинний коефіцієнт кореляції, який може використовуватися і для відбору оптимальних предикторів та для оцінки вірогідності прогностичної моделі.

6.1.4 Оцінка міри адекватності прогностичної моделі метеорологічного прогнозу

Використовуючи один з методів «просіювання» предикторів, створюється регресійна модель метеорологічного прогнозу на статистично значущих предикторах у вигляді лінійного множинного рівняння регресії.

Зрозуміло, що отримана модель взаємозв'язку між предиктантом та системою предикторів повинна бути адекватною тому процесу, який моделюється.

Можуть бути різні степені адекватності прогностичних моделей. Тому при моделюванні треба визначити кількісну міру цієї адекватності. Таку

кількісну міру називають критерієм якості або функцією цілі [13].

У регресійному аналізі велике значення має множинний коефіцієнт кореляції, оскільки він характеризує тісноту кореляційного зв'язку між предиктантом і системою предикторів X_1, X_2, \dots, X_n одночасно, тобто він може розглядатися як кількісна міра адекватності моделі метеорологічного прогнозу $Y = f(X_1, X_2, \dots, X_n)$ [27]. Чим його значення є більшим, тим тіснішим є зв'язок між предиктантом та системою відібраних предикторів і тим більш адекватною буде прогностична модель.

Оцінити міру адекватності прогностичної моделі метеорологічного прогнозу можливо шляхом перевірки статистичної гіпотези H_0 про те, що відносна залишкова дисперсія $\tilde{\sigma}_{y \bullet x_1 x_2 \dots x_n}^2$ незначуще відрізняється від дисперсії предиктанта σ_y^2 .

Перевірка гіпотези H_0 відбувається за допомогою критерію Фішера, котрий формується таким чином:

$$F = \frac{\sum_{j=1}^m (y_j - \bar{y})^2 / (m-1)}{\sum_{j=1}^m (y_j - \hat{y}_j)^2 / (m-k-1)}, \quad (6.87)$$

де

m – об'єм вибірок;

k – кількість предикторів, які включені в прогностичну модель (кількість оптимальних предикторів);

y_j і \hat{y}_j – відповідно фактичне та розрахункове значення предиктанта.

У рівнянні (6.87) у чисельнику розташована дисперсія предиктанта, а в знаменнику – залишкова дисперсія (дисперсія нев'язки; помилка моделі).

Гіпотеза H_0 не відхиляється, якщо $F < F_{кр}(\alpha, v_1, v_2)$ (для правосторонньої критичної області); $v_1 = m - 1$; $v_2 = m - k - 1$.

Вірогідність прогнозу на основі регресійної моделі можна також перевірити і за допомогою парного коефіцієнта кореляції між фактичним та розрахунковим значеннями предиктанта $r_{y_j \hat{y}_j}$. Зрозуміло, що модель тим

краще відповідає величині, що прогнозується, чим ближчим за модулем до одиниці є цей коефіцієнт кореляції.

6.2 Розв’язання типової задачі. Побудова лінійного рівняння множинної регресії

Побудувати лінійне рівняння множинної регресії, як моделі прогнозу опадів на ст.Одеса у лютому на центрованих та нормованих значеннях предикторів та предиктанта. На рівні значущості $\alpha = 0.05$ оцінити міру адекватності отриманої прогностичної моделі кліматичного прогнозу за допомогою парного коефіцієнта кореляції між фактичним та розрахунковим значеннями предиктанта.

В якості предикторів виступають часові ряди середнього місячного атмосферного тиску на двох станціях, що розташовані в Північній Атлантиці, зміна тиску на яких складає так зване Північно-Атлантичне коливання. Це ст. Рейк’явік (64° півн.ш., 22° зах.д., район Ісландської депресії) та ст. Понта-Дельгада ($37^\circ 44'$ півн.ш., $25^\circ 40'$ зах.д., район Азорського антициклону).

Враховуючи те, що взимку західний перенос у помірних широтах добре виражений, а Північно-Атлантичне коливання є тим фізичним механізмом, від якого залежить формування кліматичних показників на території України, було вирішено використати в якості предикторів саме атмосферний тиск на вказаних станціях.

Із 12-ти характеристик баричного поля Північної Атлантики було визначено 6 найбільш впливових факторів, від яких залежить, на наш погляд, формування атмосферних опадів на ст. Одеса у лютому.

В табл. 6.1 представлені вихідні дані для розв’язання поставленої задачі.

В якості 1-го, 2-го та 3-го предикторів виступають ряди атмосферного тиску на ст. Рейк’явік у листопаді (№1), грудні (№2), січні (№3), 4-го, 5-го та 6-го - ряди атмосферного тиску на ст. Понта-Дельгада у листопаді (№4), грудні (№5), січні (№6).

Оскільки прогноз опадів складається на лютий, то всі предиктори добиралися за попередній термін, враховуючи запізнення відгуків на 1-3 місяці на території України.

Реалізація алгоритму покрокової регресії проводилася на ПК за допомогою комп’ютерної програми «STEP» (Додаток В.4). Ці результати наведені в табл. 6.2.

Таблиця 6.1 – Статистичні сукупності потенційних предикторів та предиктанта

Рік	№ п/п	Атмосферний тиск, гПа, місяць						Одеса, опад, мм
		ст. Рейк'явік			ст. Понта-Дельгада			
		XI	XII	I	XI	XII	I	
1965	1	1017.7	997.5		1021.0	1024.0		
1966	2	1006.2	994.0	1009.2	1024.0	1026.0	1008.0	24
1967	3	1003.0	1008.7	1011.0	1023.0	1022.0	1019.0	88
1968	4	1004.5	1007.8	1001.5	1014.0	1020.0	1025.0	37
1969	5	1005.5	1001.5	1009.3	1019.0	1028.0	1010.0	135
1970	6	1003.5	1011.8	998.8	1021.0	1021.0	1006.0	48
1971	7	1008.2	997.9	999.8	1027.1	1023.1	1015.8	56
1972	8	1000.2	987.6	995.0	1023.7	1022.1	1020.8	5
1973	9	1010.7	1004.9	996.7	1019.8	1024.6	1021.7	58
1974	10	999.5	992.5	977.7	1021.3	1025.3	1021.0	11
1975	11	1000.5	1005.8	988.9	1023.2	1021.2	1023.4	5
1976	12	1000.0	1012.1	999.1	1023.0	1016.9	1024.8	11
1977	13	1002.6	999.6	1010.6	1017.3	1014.7	1018.6	52
1978	14	992.3	1005.9	996.9	1023.9	1011.7	1028.4	59
1979	15	995.5	991.2	1012.0	1021.0	1022.0	1014.0	23
1980	16	1014.0	997.9	1009.7	1015.0	1027.4	1017.8	4
1981	17	1005.4	1007.1	1005.0	1017.8	1018.3	1029.8	38
1982	18	993.6	991.9	1006.7	1023.3	1028.1	1017.7	33
1983	19	1010.2	997.3	986.1	1009.2	1020.8	1027.6	26
1984	20	998.0	993.0	989.1	1018.1	1023.8	1033.6	69
1985	21	1012.0	1007.2	1014.9	1015.0	1020.2	1014.7	66
1986	22	990.1	990.7	995.6	1025.7	1024.3	1032.2	55
1987	23	1005.2	1002.9	1013.6	1021.3	1012.6	1009.8	13
1988	24	1007.7	998.1	995.9	1013.5	1022.2	1024.8	15
1989	25	1008.4	1003.5	987.5	1017.0	1008.1	1027.4	10
1990	26	1009.5	994.7	981.0	1019.0	1023.0	1026.8	20
1991	27	995.6	1001.1	994.6	1022.0	1021.5	1022.0	24
1992	28	987.6	994.7	1004.0	1019.5	1023.7	1018.0	23
1993	29	991.7	993.0	984.2	1023.3	1029.9	1023.9	33
1994	30	997.9	990.5	995.2	1017.5	1022.8	1029.6	2
1995	31			992.0			1024.5	20

Таблиця 6.2 – Результати реалізації алгоритму покрокової регресії

Середні предикторів та предиктанта						
1002.56	999.41	998.72	1019.98	1021.64	1021.22	35.43
Середньоквадратичні відхили предикторів та предиктанта						
7.44	6.89	10.15	4.01	4.93	7.13	29.31
Матриця нормованих та центрованих значень предикторів та предиктанта						
1	2	3	4	5	6	y
2.04	-0.28	1.03	0.25	0.48	-1.85	-0.39
0.49	-0.79	1.21	1.00	0.88	-0.31	1.79
0.06	1.35	0.27	0.75	0.07	0.53	0.05
0.26	1.22	1.04	-1.49	-0.33	-1.57	3.40
0.40	0.30	0.01	-0.25	1.29	-2.13	0.43
0.13	1.80	0.11	0.25	-0.13	-0.76	0.70
0.76	-0.22	-0.37	1.78	0.30	-0.06	-1.04
-0.32	-1.71	-0.20	0.93	0.09	0.07	0.77
1.09	0.80	-2.07	-0.05	0.60	-0.03	-0.83
-0.41	-1.00	-0.97	0.33	0.74	0.31	-1.04
-0.28	0.93	0.04	0.80	-0.09	0.50	-0.83
-0.34	1.84	1.17	0.75	-0.96	-0.37	0.57
0.01	0.03	-0.18	-0.67	-1.41	1.01	0.80
-1.38	0.94	1.31	0.98	-2.02	-1.01	-0.42
-0.95	-1.19	1.08	0.25	0.07	-0.48	-1.07
1.54	-0.22	0.62	-1.24	1.17	1.20	0.09
0.38	1.12	0.79	-0.54	-0.68	-0.49	-0.08
-1.20	-1.09	-1.24	0.83	1.31	0.89	-0.32
1.03	-0.31	-0.95	-2.69	-0.17	1.74	1.15
-0.61	-0.93	1.59	-0.47	0.44	-0.91	1.04
1.27	1.13	-0.31	-1.24	-0.29	1.54	0.67
-1.68	-1.27	1.47	1.43	0.54	-1.60	-0.77
0.36	0.51	-0.28	0.33	-1.84	0.50	-0.70
0.69	-0.19	-1.11	-1.62	0.11	0.87	-0.87
0.79	0.59	-1.75	-0.74	-2.75	0.78	-0.53
0.93	-0.68	-0.41	-0.25	0.28	0.11	-0.39
-0.94	0.25	0.52	0.50	-0.03	-0.45	-0.42
-2.01	-0.68	-1.43	-0.12	0.42	0.38	-0.08
-1.46	-0.93	-0.35	0.83	1.68	1.17	-1.14
-0.63	-1.29	-0.66	-0.62	0.24	0.46	-0.53

Продовження табл. 6.2

Сумісна матриця R_x коеф. кореляції та коефіцієнтів значущості t_{ij}						R_{yx_i}
1.00	0.30	-0.13	-0.45	-0.08	0.11	0.188
1.80	1.00	0.10	-0.13	-0.49	-0.09	0.236
0.69	0.52	1.00	0.25	-0.01	-0.59	0.332
3.03	0.69	1.46	1.00	0.14	-0.33	-0.384
0.41	3.42	0.05	0.75	1.00	-0.07	-0.060
0.58	0.51	4.93	2.02	0.36	1.00	-0.189
1.05	1.35	2.01	2.43	0.33	1.06	1.00
КРОК 1-й Вільний член $A[0] = -1.0153739162E-10$ Коефіцієнти регресії: $A[1] = -0.384$ Коеф. значущості $t_a = 2.201$ Множинний коеф. кореляції $R_{y \bullet x_4} = 0.384$						Предиктор 4
КРОК 2-й Вільний член $A[0] = -1.0169853055E-10$ Коефіцієнти регресії: $A[1] = -0.501$ Коеф. значущості $t_a = 3.275$ $A[2] = 0.460$ Коеф. значущості $t_a = 3.006$ Множинний коеф. кореляції $R_{y \bullet x_4 x_3} = 0.587$						Предиктор 4 Предиктор 3
КРОК 3-й Вільний член $A[0] = -1.2021576358E-10$ Коефіцієнти регресії: $A[1] = -0.479$ Коеф. значущості $t_a = 3.177$ $A[2] = 0.441$ Коеф. значущості $t_a = 2.926$ $A[3] = 0.134$ Коеф. значущості $t_a = 0.889$ Множинний коеф. кореляції $R_{y \bullet x_4 x_3 x_2} = 0.602$						Предиктор 4 Предиктор 3 Предиктор 2
КРОК 4-й Вільний член $A[0] = -1.4530330937E-10$ Коефіцієнти регресії: $A[1] = -0.487$ Коеф. значущості $t_a = 3.244$ $A[2] = 0.440$ Коеф. значущості $t_a = 2.931$ $A[3] = 0.181$ Коеф. значущості $t_a = 1.209$ $A[4] = 0.099$ Коеф. значущості $t_a = 0.660$ Множинний коеф. кореляції $R_{y \bullet x_4 x_3 x_2 x_5} = 0.608$						Предиктор 4 Предиктор 3 Предиктор 2 Предиктор 5

<p>КРОК 5-й</p> <p>Вільний член $A[0] = -1.6885573405E-10$</p> <p>Коефіцієнти регресії:</p> <p>$A[1] = -0.508$ Коеф. значущості $t_a = 3.410$</p> <p>$A[2] = 0.381$ Коеф. значущості $t_a = 2.554$</p> <p>$A[3] = 0.168$ Коеф. значущості $t_a = 1.129$</p> <p>$A[4] = 0.088$ Коеф. значущості $t_a = 0.588$</p> <p>$A[5] = -0.110$ Коеф. значущості $t_a = 0.741$</p> <p>Множинний коеф. кореляції $R_{y \bullet x_4 x_3 x_2 x_5 x_6} = 0.614$</p>	<p>Предиктор 4</p> <p>Предиктор 3</p> <p>Предиктор 2</p> <p>Предиктор 5</p> <p>Предиктор 6</p>
<p>КРОК 6-й</p> <p>Вільний член $A[0] = -1.8221144925E-10$</p> <p>Коефіцієнти регресії:</p> <p>$A[1] = -0.523$ Коеф. значущості $t_a = 3.510$</p> <p>$A[2] = 0.378$ Коеф. значущості $t_a = 2.540$</p> <p>$A[3] = 0.179$ Коеф. значущості $t_a = 1.203$</p> <p>$A[4] = 0.092$ Коеф. значущості $t_a = 0.619$</p> <p>$A[5] = -0.112$ Коеф. значущості $t_a = 0.749$</p> <p>$A[6] = -0.034$ Коеф. значущості $t_a = 0.231$</p> <p>Множинний коеф. кореляції $R_{y \bullet x_4 x_3 x_2 x_5 x_6 x_1} = 0.615$</p>	<p>Предиктор 4</p> <p>Предиктор 3</p> <p>Предиктор 2</p> <p>Предиктор 5</p> <p>Предиктор 6</p> <p>Предиктор 1</p>

Відповідно до порядку виконання роботи, були реалізовані всі етапи: від підготовки вихідної інформації до розрахунків необхідних параметрів задачі.

Аналіз фактичних значень критерію Стюдента вказує на те, що між деякими предикторами зафіксовано статистично значущий зв'язок і побудувати модель на статистично незалежних предикторах неможливо. За визначеною методикою критичне значення критерію Стюдента за умов рівня значущості $\alpha = 0.05$ і числа степенів вільності $\nu = m - 1 = 30 - 1 = 29$, дорівнює: $t_{кр}(\alpha, \nu) = t_{кр}(0.05, 29) = 2.05$.

Отже виходить, що для чотирьох парних коефіцієнтів кореляції, що виражають лінійний кореляційний зв'язок між окремими предикторами, $t > t_{кр}(\alpha, \nu)$, а це означає, що гіпотеза H_0 про статистичну незначущість відхиляється.

За алгоритмом покрокової регресії розрахункова частина розпочинається з аналізу вектора парних коефіцієнтів кореляції між предиктантом та переліченими потенційними предикторами:

$$R_{yx} = \begin{pmatrix} 0,188 \\ 0,236 \\ 0,332 \\ -0,384 \\ -0,060 \\ -0,189 \end{pmatrix}.$$

Аналіз координат цього вектора вказує на те, що найбільшим за модулем є парний коефіцієнт кореляції між предиктантом та 4-тим предиктором. Він ставиться на перше місце ($x_4 \rightarrow x_1$), а ті, що залишилися, перенумеровуються. Множинний коефіцієнт кореляції першого кроку дорівнює $R_{y \bullet x_4}^{(1)} = 0,384$.

- Рівняння регресії першого кроку має вигляд:

$$\hat{y}_j^{(1)} = -0,384x_4.$$

Враховуючи алгоритм «просіювання» предикторів за методом покрокової регресії, проаналізуємо наступні кроки ітераційного процесу. На другому кроці множинний коефіцієнт кореляції зростає і вже набуває значення $R_{y \bullet x_4 x_3}^{(2)} = 0,587$, що вказує на суттєвий вплив і 3-го предиктора на предиктант.

- Рівняння другого кроку має вигляд:

$$\hat{y}_j^{(2)} = -0,501x_4 + 0,460x_3.$$

Аналіз результатів ітераційної процедури покрокової регресії (табл. 6.2) вказує на те, що на 5-6-ому кроках ітераційний процес усталюється.

Дійсно, невеликі різниці між коефіцієнтами регресії 5-го та 6-го кроків виявляються у третьому знаку, коефіцієнти Стюдента, на основі яких проводилася перевірка гіпотези про статистичну значущість коефіцієнтів моделі на кожному кроці, змінюються мало і вони є високими, порівняно з $t_{кр}(0.05, 29) = 2.05$.

Задовільну адекватність показує прогностична модель з предикторами, для яких коефіцієнти Стюдента $t_a > 2,05$.

Множинний коефіцієнт кореляції майже не змінюється, тобто на 5-6-ому кроках настає його насичення (рис. 6.1).

Отже, враховуючи всі вищеперелічені умови, як модель прогнозу опадів на ст. Одеса у лютому буде побудована на двох оптимальних предикторах, а параметри моделі відповідають 6-ому кроку ітераційного процесу.

Таким чином, ми отримали таку прогностичну модель для опадів на ст. Одеса у лютому:

$$\hat{y}_j^{(6)} = 0,378x_3 - 0,523x_4.$$

В рівнянні регресії предиктори та предиктант є центрованими та нормованими на середній квадратичний відхил.

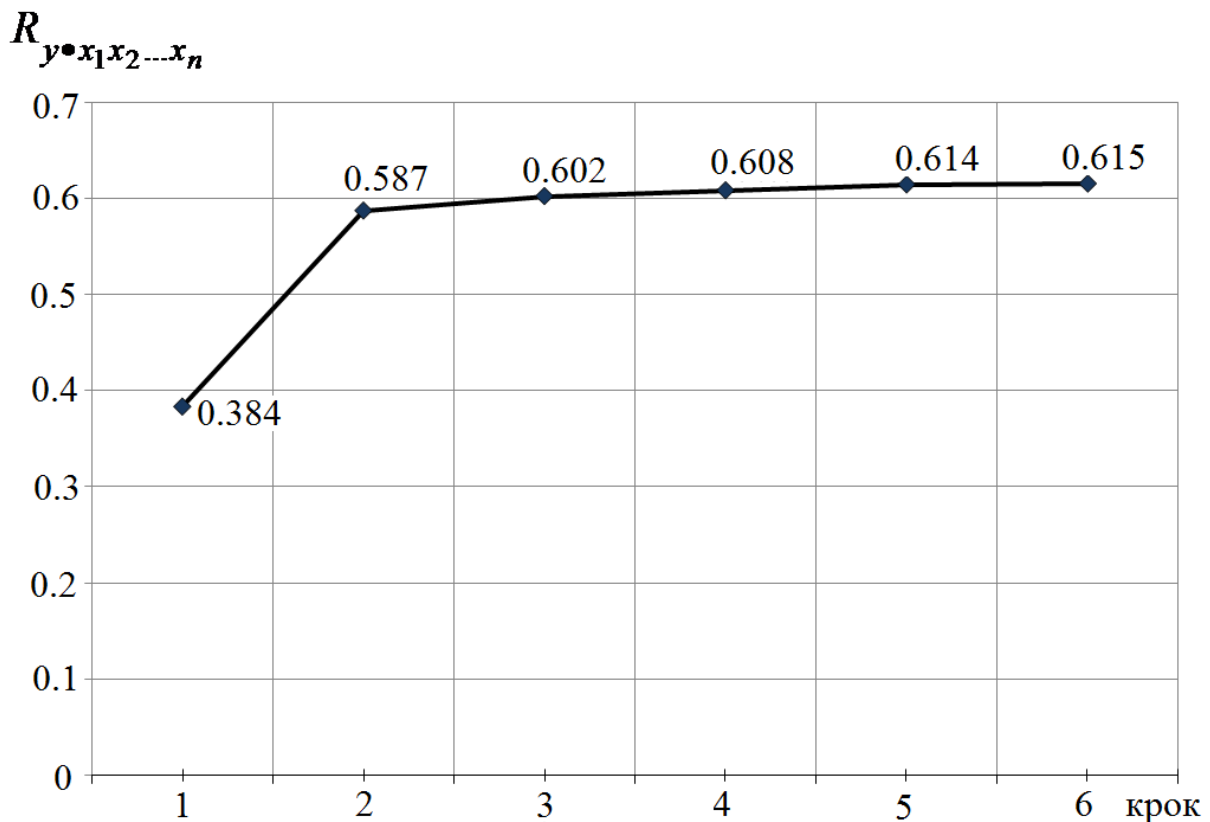


Рис. 6.1 – Зміна множинного коефіцієнта кореляції

Значення множинного коефіцієнта кореляції $R_{y \bullet x_1 x_2 \dots x_6} = 0,615$ свідчить про те, що адекватність моделі є задовільною. Удосконалюємося в цьому ще й шляхом перевірки статистичної гіпотези про значущість парного коефіцієнта кореляції [13] між фактичним та розрахунковим значеннями предиктанта (табл. 6.3).

Таблиця 6.3 – Статистичні сукупності фактичного (y_j) та модельного ($\hat{y}_j^{(6)}$) значень предиктанта

Предиктант	1	2	3	4	5	6	7	8	9	10
y_j	-0,39	1,79	0,05	3,40	0,43	0,70	-1,04	0,77	-0,83	-1,04
$\hat{y}_j^{(6)}$	0,26	-0,06	-0,29	1,17	0,13	-0,09	-1,07	-0,56	-0,76	-0,54

Предиктант	11	12	13	14	15	16	17	18	19	20
y_j	-0,83	0,57	0,80	-0,42	-1,07	0,09	-0,08	-0,32	1,15	1,04
$\hat{y}_j^{(6)}$	-0,41	0,05	0,28	-0,02	0,28	0,88	0,58	-0,90	1,05	0,85

Предиктант	21	22	23	24	25	26	27	28	29	30
y_j	0,67	-0,77	-0,70	-0,87	-0,53	-0,39	-0,42	-0,08	-1,14	-0,53
$\hat{y}_j^{(6)}$	0,53	-0,20	-0,28	0,43	-0,27	-0,02	-0,06	-0,48	-0,56	0,07

Розрахунки $r_{y\hat{y}}$ показали, що він дорівнює:

$$r_{y\hat{y}} = 0,58.$$

Оскільки об'єми вибірок є невеликими ($m=30$), то перевірку гіпотези будемо виконувати таким чином [13].

Гіпотеза H_0 : $r_{y\hat{y}} = 0,58$ на рівні значущості $\alpha = 0.05$ є статистично незначущим.

Гіпотеза H_1 : Альтернативна.

Гіпотеза H_0 перевіряється за допомогою z-перетворення Фішера:

$$\hat{z} = \frac{1}{2} \ln \frac{1+r_{y\hat{y}}}{1-r_{y\hat{y}}} = \frac{1}{2} \ln \frac{1+0.58}{1-0.58} = 0.66,$$

$$\sigma_z = \frac{1}{\sqrt{m-3}} = \frac{1}{\sqrt{30-3}} = \frac{1}{\sqrt{27}} = 0.19.$$

Фактичне значення критерію Стюдента треба порівняти з $t_{кр}(\alpha, \nu)$

$$t = \frac{\left| \hat{z} \right|}{\sigma_z} = \frac{0.66}{0.19} = 3.49, \quad \text{а} \quad t_{кр}(0.05, 29) = 2.05.$$

$t > t_{кр}(\alpha, \nu)$ – таке співвідношення вказує на те, що \hat{z} -перетворення з ймовірністю 95% є статистично значущим. Якщо \hat{z} -перетворення є статистично значущим, то й коефіцієнт кореляції $r_{y\hat{y}} = 0,58$ з ймовірністю 95% є статистично значущим.

Висновок: лінійне рівняння множинної регресії, як моделі прогнозу опадів у лютому на ст. Одеса, побудоване на двох оптимальних предикторах (із 6 потенційних) і воно має вигляд

$$\hat{y}^{(6)} = 0,378x_3 - 0,523x_4.$$

Опади на ст. Одеса у лютому залежать від атмосферного тиску на ст. Рейк'явік у січні та від атмосферного тиску на ст. Понта-Дельгада у листопаді. Вірогідність побудованої моделі складає 95%. Уточнення моделі проводилося на кожному кроці на основі перевірки статистичної гіпотези про значущість отриманих коефіцієнтів регресії.

6.3 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 містить?

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?

6.4 Завдання для закріплення теми

1. Отримати у викладача варіант задачі (Додаток Б) для виконання практичної роботи та комп'ютерну програму «STEP» (призначення програми та особливості користування дивись у Додатку В).
2. Реалізувати на ПК алгоритм розв'язання задачі покрокової регресії та отримати:
 - середні значення та середні квадратичні відхилення 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. Зробити висновок щодо роботи у цілому.

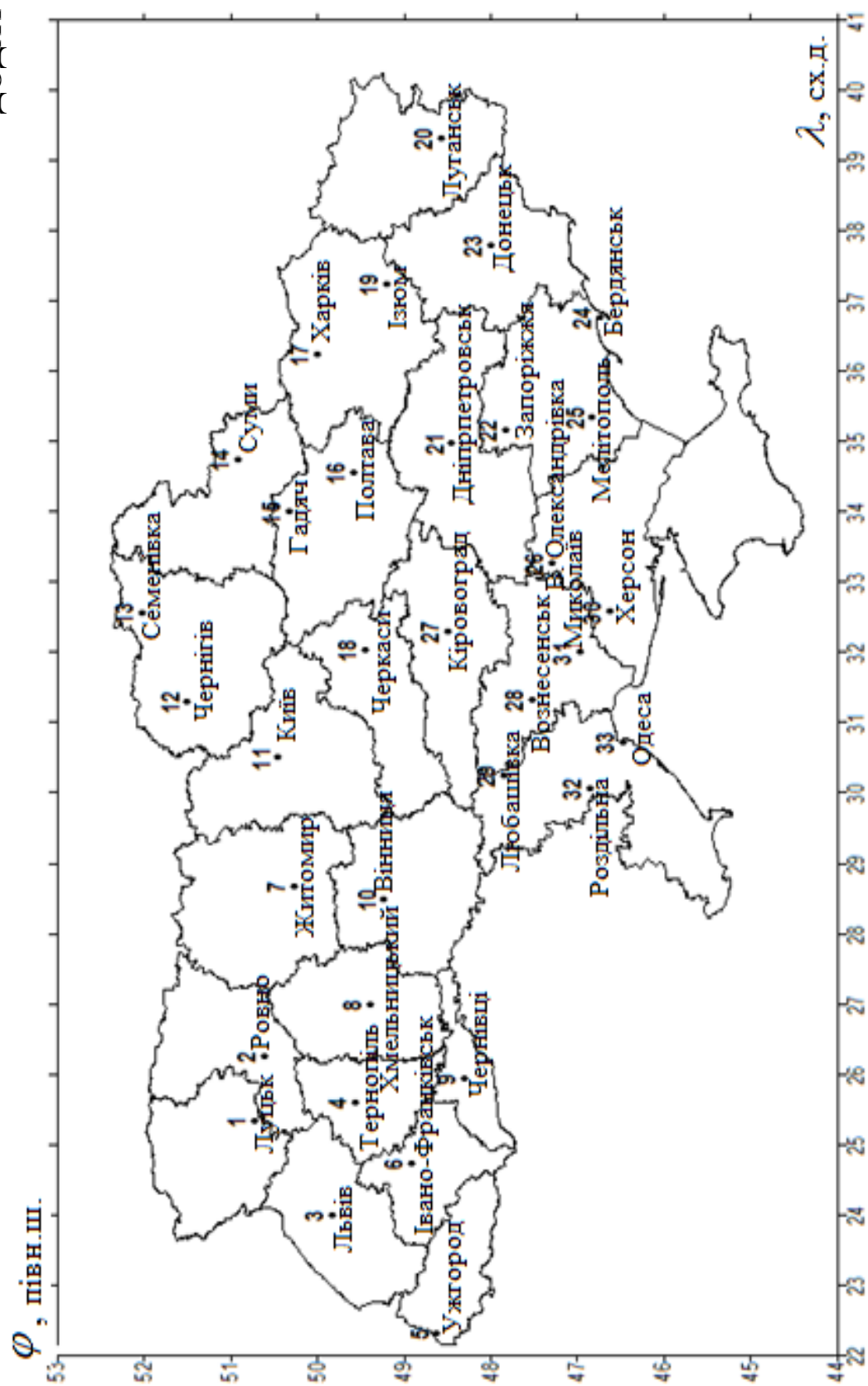


Рис. А.1 – Карта України

ДОДАТОК Б

Вихідні дані для виконання практичних вправ

Таблиця Б.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 табл. Б.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 табл. Б.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 Б.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 табл. Б.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 табл. Б.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 Б.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 табл. Б.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 табл. Б.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	Вознесеньк	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 Б.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 табл. Б.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	Вознесенськ	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 табл. Б.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 Б.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 табл. Б.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 табл. Б.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	102	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 Б.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 табл. Б.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 табл. Б.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	Вознесеньк	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 Б.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 табл. Б.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 табл. Б.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 Б.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 табл. Б.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 табл. Б.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 Б.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 табл. Б.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 табл. Б.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 Б.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 табл. Б.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 табл. Б.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 Б.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 табл. Б.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 табл. Б.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 Б.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 табл. Б.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 табл. Б.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

Таблиця Б.13 – Середня місячна температура повітря (травень)

Рік	Одеса	Київ	Харків	Львів	Вінниця	Донецьк
1965	13.2	12.7	13.4	10.5	12.2	13.7
1966	15.5	16.4	16.9	14.2	15.0	15.4
1967	15.8	17.5	18.5	13.7	15.3	18.4
1968	17.4	15.1	16.7	13.3	14.9	17.9
1969	14.4	14.8	14.5	15.0	14.6	14.6
1970	14.4	15.0	16.0	11.9	13.7	15.5
1971	15.9	16.1	15.8	15.0	15.2	15.5
1972	16.1	16.0	16.3	13.9	14.8	16.4
1973	14.5	14.4	15.3	12.9	13.7	15.1
1974	13.3	12.6	12.8	11.1	12.0	13.5
1975	17.4	19.0	18.9	15.0	16.8	18.5
1976	13.7	12.6	12.9	12.2	12.2	13.4
1977	14.6	15.2	15.9	13.3	14.0	15.2
1978	13.5	12.6	12.8	11.3	11.7	12.4
1979	18.2	17.6	19.6	14.6	15.8	20.0
1980	12.7	11.3	12.5	9.7	11.0	13.4
1981	14.6	16.1	15.7	13.3	14.3	14.9
1982	15.5	15.1	14.2	13.8	14.8	14.7
1983	17.2	17.9	17.2	15.5	16.1	17.4
1984	15.1	16.7	18.7	13.4	14.5	17.9
1985	15.8	17.0	17.9	14.9	15.7	17.8
1986	16.5	16.7	15.2	14.7	14.9	14.9
1987	12.5	14.2	15.4	11.7	13.2	14.6
1988	15.0	15.2	14.8	13.7	14.3	14.6
1989	15.3	15.2	14.5	13.6	13.8	14.1
1990	14.9	14.6	13.6	12.9	13.7	13.9
1991	13.8	12.8	13.7	10.5	11.7	13.7
1992	14.7	13.6	13.6	11.9	12.4	13.6
1993	15.2	16.5	15.8	16.0	15.6	15.1
1994	15.6	13.1	13.1	12.6	13.4	13.6
1995	14.1	14.0	15.8	12.1	12.6	15.4
1996	17.5	18.6	18.9	16.1	17.6	18.6
1997	16.6	16.1	16.2	14.0	15.2	16.3
1998	15.2	15.2	15.9	13.1	13.4	15.4
1999	14.0	12.8	12.8	12.1	12.1	12.1
2000	16.7	15.4	14.0	14.7	14.9	13.7
2001	14.8	14.2	13.8	14.1	13.3	13.6
2002	16.9	16.5	15.5	16.5	15.9	15.6
2003	18.4	19.4	18.8	16.9	18.6	19.1
2004	14.2	13.2	13.8	11.9	12.7	14.3
2005	16.9	16.4	17.9	13.0	14.7	18.2

Таблиця Б.14 – Середня місячна температура повітря (червень)

Рік	Одеса	Київ	Харків	Львів	Вінниця	Донецьк
1965	19.9	17.6	18.4	16.3	16.9	19.5
1966	17.4	16.9	17.4	15.9	15.8	17.4
1967	19.1	18.2	18.5	16.4	17.3	18.0
1968	19.8	20.1	20.0	18.3	19.0	19.9
1969	19.2	17.3	18.6	15.6	16.4	19.3
1970	18.4	17.2	17.6	16.0	16.5	17.5
1971	18.9	18.1	19.1	15.7	16.8	18.7
1972	21.3	19.7	22.5	16.8	18.3	23.0
1973	18.6	17.9	18.5	15.2	16.3	17.7
1974	18.6	17.0	17.9	14.1	16.1	18.1
1975	21.9	20.7	23.3	16.5	18.7	23.1
1976	18.4	16.2	16.6	14.9	15.6	16.7
1977	18.2	17.2	17.3	16.4	16.9	17.6
1978	18.4	16.6	16.4	14.9	16.0	16.6
1979	21.2	21.5	21.3	19.0	19.9	21.0
1980	18.1	17.1	17.9	14.8	16.0	18.2
1981	21.4	21.1	22.8	17.4	19.6	22.3
1982	18.7	16.9	16.5	15.8	16.2	16.8
1983	19.6	18.2	18.1	16.0	16.7	18.4
1984	17.8	15.7	18.0	13.9	14.5	18.6
1985	18.0	16.7	17.4	14.0	15.7	17.6
1986	20.7	19.3	20.5	16.3	17.7	18.6
1987	19.3	18.2	18.7	16.5	17.4	19.9
1988	19.3	18.3	19.6	15.5	16.8	19.2
1989	19.0	19.4	20.6	15.3	17.0	19.7
1990	19.0	16.9	16.3	15.8	16.3	17.0
1991	20.0	18.5	20.7	16.0	17.4	20.8
1992	18.9	18.7	18.8	17.6	17.4	18.5
1993	18.7	16.7	16.5	15.4	15.8	16.9
1994	18.8	16.3	16.4	15.6	15.6	16.5
1995	21.1	20.1	21.7	16.9	18.2	21.6
1996	20.5	18.8	18.9	16.8	17.9	19.0
1997	19.7	18.3	19.5	16.7	17.2	19.4
1998	21.1	19.8	22.0	17.6	18.6	21.6
1999	22.8	22.6	22.8	18.7	20.8	22.2
2000	19.8	17.9	17.8	16.9	17.2	17.9
2001	18.2	16.7	16.9	14.7	15.7	17.1
2002	20.4	18.4	18.9	17.0	17.0	19.2
2003	20.5	18.0	17.3	17.4	17.7	17.5
2004	19.0	17.7	17.3	16.2	16.5	17.0
2005	19.3	17.3	17.6	16.3	16.6	18.5

Таблиця Б.15 – Середньомісячні значення індексу
Північно-Атлантичного коливання (ПАК)

№ п/п	Рік	Місяць				
		I	II	III	IV	V
1	1965	-0.12	-1.55	-1.51	0.72	-0.62
2	1966	-1.74	-1.39	0.56	-0.75	0.22
3	1967	-0.89	0.19	1.51	0.18	-0.99
4	1968	0.13	-1.29	0.4	-1.08	-1.76
5	1969	-0.83	-1.55	-1.56	1.53	0.55
6	1970	-1.50	0.64	-0.96	-1.30	1.14
7	1971	-1.13	0.24	-0.84	-0.24	0.50
8	1972	0.27	0.32	0.72	-0.22	0.95
9	1973	0.04	0.85	0.30	-0.54	-0.44
10	1974	1.34	-0.14	-0.03	0.51	-0.24
11	1975	0.58	-0.62	-0.61	-1.60	-0.52
12	1976	-0.25	0.93	0.75	0.26	0.96
13	1977	-1.04	-0.49	-0.81	0.65	-0.86
14	1978	0.66	-2.20	0.70	-1.17	1.08
15	1979	-1.38	-0.67	0.78	-1.71	-1.03
16	1980	-0.75	0.05	-0.31	1.29	-1.50
17	1981	0.37	0.92	-1.19	0.36	0.20
18	1982	-0.89	1.15	1.15	0.10	-0.53
19	1983	1.59	-0.53	0.95	-0.85	-0.07
20	1984	1.66	0.72	-0.37	-0.28	0.54
21	1985	-1.61	-0.49	0.20	0.32	-0.49
22	1986	1.11	-1.00	1.71	-0.59	0.85
23	1987	-1.15	-0.73	0.14	2.00	0.98
24	1988	1.02	0.76	-0.17	-1.17	0.63
25	1989	1.17	2.00	1.85	0.28	1.38
26	1990	1.04	1.41	1.46	2.00	-1.53
27	1991	0.86	1.04	-0.2	0.29	0.08
28	1992	-0.13	1.07	0.87	1.86	2.63
29	1993	1.60	0.50	0.67	0.97	-0.78
30	1994	1.04	0.46	1.26	1.14	-0.57
31	1995	0.93	1.14	1.25	-0.85	-1.49
32	1996	-0.12	-0.07	-0.24	-0.17	-1.06
33	1997	-0.49	1.70	1.46	-1.02	-0.28
34	1998	0.39	-0.11	0.87	-0.68	-1.32
35	1999	0.77	0.29	0.23	-0.95	0.92
36	2000	0.60	1.70	0.77	-0.03	1.58
37	2001	0.25	0.45	-1.26	0.00	-0.02
38	2002	0.44	1.10	0.69	1.18	-0.22
39	2003	0.16	0.62	0.32	-0.18	0.01
40	2004	-0.29	-0.14	1.02	1.15	0.19
41	2005	1.52	-0.06	-1.83	-0.30	-1.25

Таблиця Б.16 – Середньомісячні значення індексу
Північно-Морського-Каспійського коливання (ПКК)

№ п/п	Рік	Місяць				
		I	II	III	IV	V
1	1965	-0.82	0.86	-0.33	0.62	-0.75
2	1966	-0.78	-1.08	-0.16	-0.20	0.23
3	1967	-0.08	0.55	-0.09	0.39	-1.30
4	1968	0.20	-0.63	-0.09	0.05	-1.63
5	1969	0.57	-0.76	0.02	-0.22	-0.25
6	1970	-0.48	-1.65	-1.72	-2.09	0.66
7	1971	-0.99	1.08	-0.83	0.78	0.37
8	1972	1.04	-0.18	0.68	-0.74	-0.67
9	1973	1.69	-0.47	1.37	-0.59	0.04
10	1974	0.84	-0.54	-0.32	2.03	-0.55
11	1975	-0.11	2.00	-0.35	-1.49	-0.33
12	1976	-0.04	1.44	0.16	0.56	-0.19
13	1977	-0.02	-1.49	-0.50	-1.80	-0.23
14	1978	-0.43	-1.04	-1.31	0.65	1.21
15	1979	-1.39	-0.23	-1.80	-0.38	-1.46
16	1980	0.29	0.61	-0.55	0.71	-0.23
17	1981	-0.26	0.34	-0.60	1.56	0.30
18	1982	0.49	1.33	0.14	-0.03	-0.76
19	1983	0.69	0.04	0.05	-1.24	-1.50
20	1984	-1.70	0.33	-0.30	1.22	-0.46
21	1985	-0.87	1.20	-0.28	-0.68	-0.16
22	1986	-1.36	-0.39	-0.20	-2.06	0.40
23	1987	0.07	-0.36	-0.24	1.77	-1.06
24	1988	-0.51	-1.21	-1.17	-0.06	1.03
25	1989	2.30	0.18	-0.73	-0.80	1.70
26	1990	0.55	-0.72	0.94	0.48	1.11
27	1991	0.90	-0.21	0.31	-0.34	1.26
28	1992	2.39	1.27	0.07	0.05	2.21
29	1993	0.65	1.80	1.15	0.71	1.07
30	1994	-0.96	0.65	-0.78	-0.52	-0.01
31	1995	-0.90	-1.28	-1.68	-0.10	-0.59
32	1996	0.43	-0.72	0.44	0.93	-1.20
33	1997	1.71	-0.02	1.47	0.61	-0.19
34	1998	0.71	1.48	0.74	-1.61	0.68
35	1999	-0.86	-0.54	-0.40	0.21	1.15
36	2000	0.65	-0.22	0.48	-0.70	1.15
37	2001	-0.49	0.04	-1.85	-1.17	1.13
38	2002	0.67	-0.87	-0.06	1.36	1.25
39	2003	-0.48	1.10	1.69	0.94	-0.05
40	2004	-0.94	0.29	-0.18	0.66	-0.10
41	2005	0.17	0.08	0.29	-0.33	-1.10

Пояснення до комп'ютерних програм

Додаток В.1 – 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 -й метеорологічній станції.

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

Приклад формування файлу вихідних даних місячної кількості опадів для 5 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

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

Додаток В.2 – Відомості про комп'ютерну програму «EOF»
(автор к.геогр.н. Верлан В.А.)

Програма «EOF» дозволяє реалізувати алгоритм повної проблеми власних значень і отримати власні значення та відповідні їм власні вектори матриці кореляції. Також розраховуються: середнє значення, середній квадратичний відхил випадкової величини на *i*-й метеорологічній станції (*i*-му 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 вихідної інформації.

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

Приклад формування файла вихідних даних місячної кількості опадів для 6 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:



Додаток В.3 – Відомості про комп'ютерну програму «Cla2003»
(автор к.геогр.н. Серга Е.М.)

Комп'ютерна програма «Cla2003» 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 за кластерами та номери векторів вихідної вибірки, що входять в певні кластери.

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

Вигляд файла 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 містяться в кожному вихідному векторі (статистична сукупність);

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

Додаток В.4 – Відомості про комп'ютерну програму «STEP»
(автор к.геогр.н. Верлан В.А.)

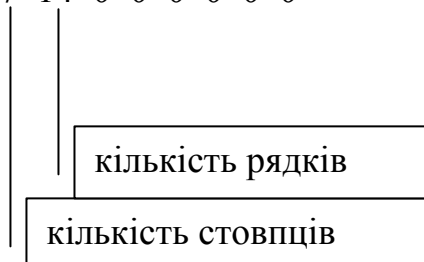
Комп'ютерна програма «STEP» дозволяє реалізувати «просіювання» предикторів за методом покрокової регресії на основі статистичних сукупностей потенційних предикторів $X_i (i = \overline{1, n})$, визначених при розв'язанні фізичної задачі, а також сукупності предиктанта Y такого ж об'єму m .

Програма передбачає формування файлу даних тільки в одному варіанті – спочатку в рядку наводиться обов'язкова інформація, а у стовпчиках (кількість яких $n+1$) – одна за одною статистичні сукупності потенційних предикторів $X_i (i = \overline{1, n})$, і в останньому – предиктанта Y .

Зразок файлу вихідних даних до програми «STEP»:

7	14	0	0	0	0	0	0
1017.7	997.5	1009.2	1021.0	1024.0	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 предикторів та одного предиктанта з об'ємами вибірок $m = 14$. Останній стовпчик (7-й) – предиктант.

ДОДАТОК Г

*Значення критерію Стьюдента для рівня
значущості α і числа степенів вільності ν*

ν	Двостороння критична область, α							
	0.2	0.1	0.05	0.02	0.01	0.005	0.002	0.001
6	1.44	1.94	2.45	3.14	3.71	4.32	5.21	5.96
7	1.41	1.89	2.36	3.00	3.50	4.03	4.79	5.41
8	1.40	1.86	2.31	2.90	3.36	3.83	4.50	5.04
9	1.38	1.83	2.26	2.82	3.25	3.69	4.30	4.78
10	1.37	1.81	2.23	2.76	3.17	3.58	4.14	4.59
11	1.36	1.80	2.20	2.72	3.11	3.50	4.02	4.44
12	1.36	1.78	2.20	2.68	3.05	3.43	3.93	4.32
13	1.35	1.77	2.16	2.65	3.01	3.37	3.85	4.22
14	1.34	1.76	2.14	2.62	2.98	3.33	3.79	4.14
15	1.34	1.75	2.13	2.60	2.95	3.29	3.73	4.07
16	1.34	1.75	2.12	2.58	2.92	3.25	3.69	4.02
17	1.33	1.74	2.11	2.57	2.90	3.22	3.65	3.97
18	1.33	1.73	2.10	2.55	2.88	3.20	3.61	3.92
19	1.33	1.73	2.09	2.54	2.86	3.17	3.58	3.88
20	1.33	1.72	2.09	2.53	2.85	3.15	3.55	3.85
21	1.32	1.72	2.08	2.52	2.83	3.14	3.53	3.82
22	1.32	1.72	2.07	2.51	2.82	3.12	3.51	3.79
23	1.32	1.71	2.07	2.50	2.81	3.10	3.48	3.77
24	1.32	1.71	2.06	2.49	2.90	3.09	3.47	3.75
25	1.32	1.71	2.06	2.49	2.79	3.08	3.45	3.73
26	1.32	1.71	2.06	2.48	2.78	3.07	3.44	3.71
27	1.31	1.70	2.05	2.47	2.77	3.06	3.42	3.69
28	1.31	1.70	2.05	2.47	2.76	3.05	3.41	3.67
29	1.31	1.70	2.05	2.46	2.76	3.04	3.40	3.66
30	1.31	1.70	2.04	2.46	2.75	3.03	3.39	3.65
40	1.30	1.68	2.02	2.42	2.70	2.97	3.31	3.55
60	1.30	1.67	2.00	2.39	2.66	2.91	3.23	3.46
120	1.29	1.66	1.98	2.36	2.62	2.85	3.16	3.37
∞	1.28	1.64	1.96	2.33	2.58	2.81	3.09	3.29

ДОДАТОК Д

Значення критерію Фішера F

для рівня значущості 0.05

 ν_1 – число степенів вільності для більшої дисперсії ν_2 – число степенів вільності для меншої дисперсії

ν_2	ν_1							
	15	20	24	30	40	60	120	∞
10	2.845	2.774	2.737	2.700	2.661	2.621	2.580	2.538
11	2.179	2.646	2.609	2.571	2.531	2.490	2.448	2.405
12	2.617	2.544	2.506	2.466	2.426	2.384	2.341	2.296
13	2.533	2.459	2.420	2.380	2.339	2.297	2.252	2.206
14	2.463	2.388	2.349	2.308	2.266	2.223	2.178	2.131
15	2.404	2.328	2.288	2.247	2.204	2.160	2.114	2.066
16	2.352	2.276	2.235	2.194	2.151	2.106	2.059	2.010
17	2.308	2.230	2.190	2.148	2.104	2.058	2.010	1.960
18	2.269	2.191	2.150	2.107	2.063	2.017	1.968	1.917
19	2.234	2.156	2.114	2.071	2.026	1.980	1.930	1.878
20	2.203	2.124	2.083	2.039	1.994	1.946	1.896	1.843
21	2.176	2.096	2.054	2.010	1.965	1.917	1.866	1.812
22	2.151	2.071	2.028	1.984	1.938	1.890	1.838	1.783
23	2.128	2.048	2.005	1.961	1.914	1.865	1.813	1.757
24	2.108	2.027	1.984	1.939	1.892	1.842	1.790	1.733
25	2.089	2.008	1.964	1.919	1.872	1.822	1.768	1.711
26	2.072	1.990	1.946	1.901	1.853	1.803	1.749	1.691
27	2.056	1.974	1.930	1.884	1.836	1.785	1.731	1.672
28	2.041	1.959	1.915	1.869	1.820	1.769	1.714	1.654
29	2.028	1.945	1.901	1.854	1.806	1.754	1.698	1.638
30	2.015	1.932	1.987	1.841	1.792	1.740	1.684	1.622
40	1.925	1.839	1.793	1.744	1.693	1.637	1.577	1.509
60	1.836	1.748	1.700	1.649	1.594	1.534	1.467	1.389
120	1.751	1.659	1.608	1.554	1.495	1.429	1.352	1.254
∞	1.666	1.571	1.517	1.459	1.394	1.318	1.221	1.000

Значення величини r для значень z від 0.00 до 2.99

z	0	1	2	3	4	5	6	7	8	9
0.0	0.000	0.010	0.020	0.030	0.040	0.050	0.060	0.070	0.080	0.090
0.1	0.100	0.110	0.119	0.129	0.139	0.149	0.159	0.168	0.178	0.188
0.2	0.197	0.207	0.217	0.226	0.236	0.245	0.254	0.264	0.273	0.282
0.3	0.291	0.300	0.310	0.319	0.328	0.336	0.345	0.354	0.363	0.371
0.4	0.380	0.389	0.397	0.405	0.414	0.422	0.430	0.438	0.446	0.454
0.5	0.462	0.470	0.478	0.485	0.493	0.501	0.508	0.515	0.523	0.530
0.6	0.537	0.544	0.551	0.558	0.565	0.572	0.578	0.585	0.592	0.598
0.7	0.604	0.611	0.617	0.623	0.629	0.635	0.641	0.647	0.653	0.658
0.8	0.664	0.670	0.675	0.681	0.686	0.691	0.696	0.701	0.706	0.711
0.9	0.716	0.721	0.726	0.731	0.735	0.740	0.744	0.749	0.753	0.757
1.0	0.762	0.766	0.770	0.774	0.778	0.782	0.786	0.790	0.793	0.697
1.1	0.801	0.804	0.808	0.811	0.814	0.818	0.821	0.824	0.828	0.831
1.2	0.834	0.837	0.840	0.843	0.846	0.848	0.851	0.854	0.857	0.859
1.3	0.862	0.864	0.867	0.869	0.872	0.874	0.876	0.879	0.881	0.883
1.4	0.885	0.888	0.890	0.892	0.894	0.896	0.898	0.900	0.902	0.903
1.5	0.905	0.907	0.909	0.910	0.912	0.914	0.915	0.917	0.919	0.920
1.6	0.922	0.923	0.925	0.926	0.928	0.929	0.930	0.932	0.933	0.934
1.7	0.935	0.937	0.940	0.939	0.940	0.941	0.943	0.944	0.945	0.946

Продовження додатка Е

z	0	1	2	3	4	5	6	7	8	9
1.8	0.947	0.948	0.949	0.950	0.951	0.952	0.953	0.954	0.955	0.955
1.9	0.956	0.957	0.958	0.959	0.960	0.960	0.961	0.962	0.963	0.963
2.0	0.964	0.965	0.965	0.966	0.967	0.967	0.968	0.969	0.969	0.970
2.1	0.970	0.971	0.972	0.972	0.973	0.937	0.974	0.974	0.975	0.975
2.2	0.976	0.976	0.977	0.977	0.978	0.978	0.979	0.979	0.979	0.980
2.3	0.980	0.981	0.981	0.981	0.982	0.982	0.982	0.983	0.983	0.983
2.4	0.984	0.984	0.984	0.985	0.985	0.985	0.986	0.986	0.986	0.986
2.5	0.987	0.987	0.987	0.987	0.988	0.988	0.988	0.988	0.989	0.989
2.6	0.989	0.989	0.989	0.990	0.990	0.990	0.990	0.990	0.991	0.991
2.7	0.991	0.991	0.991	0.992	0.992	0.992	0.992	0.992	0.992	0.993
2.8	0.993	0.993	0.993	0.993	0.993	0.993	0.994	0.994	0.994	0.994
2.9	0.994	0.994	0.994	0.994	0.994	0.995	0.995	0.995	0.995	0.995

Значення величини \hat{z} для значень r_{xy} від 0.00 до 0.99

r	0	1	2	3	4	5	6	7	8	9
0.0	0.000	0.010	0.020	0.030	0.040	0.050	0.060	0.070	0.080	0.090
0.1	0.100	0.111	0.121	0.131	0.141	0.151	0.161	0.172	0.182	0.192
0.2	0.203	0.213	0.224	0.234	0.245	0.255	0.266	0.277	0.288	0.297
0.3	0.310	0.321	0.332	0.343	0.354	0.365	0.377	0.388	0.400	0.412
0.4	0.424	0.436	0.448	0.460	0.472	0.485	0.497	0.510	0.523	0.536
0.5	0.549	0.563	0.576	0.590	0.604	0.618	0.633	0.648	0.663	0.678
0.6	0.693	0.709	0.725	0.741	0.758	0.775	0.793	0.811	0.829	0.848
0.7	0.867	0.887	0.908	0.929	0.951	0.973	0.996	1.020	1.045	1.071
0.8	1.099	1.127	1.157	1.188	1.221	1.256	1.293	1.333	1.376	1.422
0.9	1.472	1.528	1.589	1.658	1.738	1.832	1.946	2.092	2.298	2.647

ВИКОРИСТАНА ЛІТЕРАТУРА

1. Андерсон Т.В. Введение в многомерный статистический анализ. – М.:Физматгиз, 1963. – 500 с.
2. Багров Н.А. Аналитическое представление последовательности метеорологических полей посредством естественных ортогональных функций // Труды ЦИП, 1959. Вып. 74. – С.133-138.
3. Багров Н.А. Преобразование и отбор предсказателей в корреляционном анализе. – Л.: Труды ГМЦ, 1970. Вып. 64. – С. 3-23.
4. Вапник В.Н., Червоненкис А.Я. Теория распознавания образов. – М.: Наука, 1974. – 252 с.
5. Вентцель Е.С., Овчаров Л.А. Теория вероятностей. – М.: Наука, 1988. – 480 с.
6. Гандин Л.С., Каган Р.Л. Статистические методы интерпретации метеорологических данных. – Л.: Гидрометеиздат, 1976. – 360 с.
7. Гончарова Л.Д., Школьный Е.П. Сжатие информации о статистической структуре составляющих скоростей ветра над Северной Америкой. // Межвед. науч. сб. Украины. – Одесса. – 1995. – Вып. 32. – С. 14-17.
8. Гончарова Л.Д., Решетченко С.І. Вплив Північної Атлантики на температуру повітря, опади, приземний тиск на лівоберіжній Україні. // Український гідрометеорологічний журнал. – 2010. – Вип. 7. – С. 45-52.
9. Гончарова Л.Д., Решетченко С.І. Особливості статистичної структури полів температури повітря та опадів у другій половині ХХ століття на території лівобережної України. // Вісник Одеського державного екологічного університету. – 2010. – Вип. 9. – С. 103-113.
10. Гончарова Л.Д. Факторный анализ статистической структуры вертикальных профилей скорости ветра в тропо-стратосфере // Метеорология, климатология и гидрология: Межвед. науч. сб. Украины. – Одесса. – 1999. – Вып. 39. – С. 64-69.
11. Гончарова Л.Д. Воздушные течения тропосферы и стратосферы северного полушария. Монография. – Одесса: ТЭС, 2014. – 298 с.
12. Гончарова Л.Д. Дослідження статистичної структури вертикальних профілів температури повітря в тропосфері та стратосфері // Метеорологія, кліматологія та гідрологія: Міжвід. наук. зб. України. – Одеса. – 2002. – Вип. 45. – С. 31-36.
13. Гончарова Л.Д., Школьный Е.П. Методи обробки та аналізу гідрометеорологічної інформації (збірник задач і вправ). Навчальний посібник. – Одеса: Екологія, 2007. – 460 с.
14. Казакевич Д.И. Основы теории случайных функций и ее применение в гидрометеорологии. – А.: Гидрометиздат, 1971. – 268 с.

15. Лоули Д., Максвелл А. Факторный анализ как статистический метод. – М.: Мир, 1967. – 144 с.
16. Мещерская А.В., Репинская Р.П., Юдин М.И. Естественные ортогональные составляющие метеорологических полей. – Л.: Гидрометеиздат, 1970. – 205 с.
17. Многомерные статистические модели климата. – Л.: Гидрометеиздат, 1989. – 183 с.
18. Пановский Г.А., Брайер Г.В. Статистические методы в метеорологии. Пер. с англ. – Л.: Гидрометиздат, 1972. – 209 с.
19. Райзин Дж.Вэн. Классификация и кластер. – М.: Мир, 1980. – 389 с.
20. Себестиан Г.С. Процессы принятия решений при распознавании образов. – К.: Техника, 1965. – 310 с.
21. Себер Дж. Линейный регрессионный анализ. – М.: Мир, 1980. – 456 с.
22. Серга Э.Н. Универсальный адаптивный итерационный метод кластерного анализа // Метеорологія, кліматологія та гідрологія: Міжвід. наук. зб. України. – Одеса. – 2003. – Вип. 47. – С. 83-89.
23. Ту Дж., Гонсалес Р. Принципы распознавания образов. – М.: Мир, 1978. – 208 с.
24. Хедли Дж. Линейная алгебра. Пер. с англ. Ю.И. Соркина и А.А. Дмитриева. – М.: Высшая школа, 1966. – 208 с.
25. Шефтель З.Г. Теорія ймовірності. – К.: Вища школа, 1994. – 192 с.
26. Школьный Е.П., Волконская Н.К. Обобщенный фактор температурно-влажностного режима // Метеорология, климатология и гидрология: Межвед. науч. сб. Украины. – Одесса. – 1982. – Вып. 18. – С. 3-16.
27. Школьный Є.П., Лосєва І.Д., Гончарова Л.Д. Обробка та аналіз гідрометеорологічної інформації. Підручник. – К.: Міносвіти України, 1999. – 600 с.

ПРЕДМЕТНИЙ ПОКАЖЧИК

А	К
Аксіоматика метричного простору – 7	Кластер – 88
	Коваріація – 13
Б	Коефіцієнт кореляції
	– множинний – 110
Базис евклідового простору – 17,45	– частинний – 111
	Компоненти
	– головні – 44, 56
В	– ортогональні – 44
Вага фактора – 73	Кореляція – 17
Вектор	Критерій –
– ортогональних компонент – 52	– кластеризації – 92
– модифікований – 56	– Стьюдента – 116, 117, 121, 124, 131, 132, 135, 182
– «ситуації» – 88	– Фішера – 39, 114, 118, 126, 134, 183
Власне значення (власне число) – 46	
Власні вектори	М
– ортонормовані – 50	
– ортогональні – 49	
	Матриця кореляції
Д	– розширена – 111
«Далекі зв'язки» – 28, 31	Метрика – 6
Дисперсія	Мінор – 111
– поля сумарна – 54	Міра
– відтвореності – 116	– близькості – 89
	– схожості – 92
Е	
Евклідова відстань – 7	Н
Еталон кластера – 89	
	Наддовгостроковий прогноз – 28
З	Некорельованість – 17
Задача	Нетривіальний розв'язок – 47
– стиску інформації – 44, 56	Норма вектора – 7
– фільтрації інформації – 44, 57, 59	
«Залишки» – 73	

О

- «Образ»
 - репрезентативний – 88
- Операція
 - транспонування – 9
 - центрування – 12
 - обернення матриці – 15
- Ортогональність – 17
- Ортогональні функції
 - емпіричні – 44
 - природні – 44

П

- Параметризація факторів – 43, 45
- Перетворення
 - розтягу – 46
 - стиску – 46
- Поле
 - метеорологічне – 8
 - гідрологічне – 8
 - кореляційне – 18
 - однорідне – 33
 - ізотропне – 33
 - квазіоднорідне – 33
 - квазіізотропне – 33
- Поліс кореляції – 18
- Предиктант – 100
- Предиктор
 - потенційний – 68, 101
 - оптимальний – 101, 117
- «Просіювання»
предикторів – 43, 101
- Простір
 - багатовимірний арифметичний – 6
 - метричний арифметичний – 6
 - евклідовий – 7

Р

- Рівняння повної проблеми
власних значень – 45, 51
- «Розв’язувальне правило» – 91

С

- Система точок
 - нерегулярна – 8
 - регулярна – 8
 - упорядкована – 8

Ф

- Фактори
 - узагальнені – 72
- Функція
 - правдоподібності – 76
 - відстані – 89, 90
 - коваріційна – 25
 - кореляційна – 26
 - автоковаріаційна – 26
 - автокореляційна – 26
 - структурна – 26

Ц

- Центр
 - кореляції – 18, 35
 - кластера – 89

Ч

- Число Махаланобіса – 92

ІМЕННИЙ ПОКАЖЧИК

В

Верлан В.А. – 179, 181

Г

Гаусс – 106

Гончарова Л.Д. – 31, 32, 80

Е

Ерміт – 44

К

Коші-Буняковський – 17

Л

Лаггер – 44

Лежандр – 44

М

Махаланобіс – 92

П

Пірсон – 72

С

Серга Е.М. – 180

Спірман – 72

Стьюдент – 116, 117, 121, 124,
131, 132, 135, 182

Ф

Фішер – 39, 114, 118, 126,
134, 183

Х

Хотеллінг – 72

Ч

Чебишев – 44

Я

Якобі – 48

ЗМІСТ

ПЕРЕДМОВА.....	3
1 ОСНОВНІ ЗАСАДИ БАГАТОВИМІРНОГО СТАТИСТИЧНОГО АНАЛІЗУ.....	6
1.1 Базові теоретичні положення.....	6
<i>1.1.1 Елементи лінійної алгебри.....</i>	<i>6</i>
<i>1.1.2 Метеорологічні поля як об'єкти багатовимірного статистичного аналізу.....</i>	<i>7</i>
<i>1.1.3 Матричне представлення метеорологічних полів....</i>	<i>11</i>
<i>1.1.3.1 Матриця коваріацій та алгоритм її розрахунку.....</i>	<i>11</i>
<i>1.1.3.2 Матриця кореляцій та алгоритм її розрахунку.....</i>	<i>15</i>
1.2 Розв'язання типової задачі. Структура матриць коваріацій та кореляцій.....	19
1.3 Запитання для самоперевірки.....	23
1.4 Завдання для закріплення теми.....	24
2 КОРЕЛЯЦІЙНИЙ АНАЛІЗ МЕТЕОРОЛОГІЧНИХ ПОЛІВ....	25
2.1 Базові теоретичні положення.....	25
<i>2.1.1 Основні ідеї кореляційного аналізу.....</i>	<i>25</i>
<i>2.1.2 Напрями використання кореляційного аналізу в гідрометеорології.....</i>	<i>27</i>
<i>2.1.3 Однорідність та ізотропність випадкових полів....</i>	<i>32</i>
2.2 Розв'язання типової задачі. Дослідження однорідності та ізотропності метеорологічних полів.....	35
2.3 Запитання для самоперевірки.....	41
2.4 Завдання для закріплення теми.....	42

3 КОМПОНЕНТНИЙ АНАЛІЗ МЕТЕОРОЛОГІЧНИХ ПОЛІВ...	43
3.1 Базові теоретичні положення.....	43
<i>3.1.1 Напрями використання компонентного аналізу в гідрометеорології.....</i>	<i>43</i>
<i>3.1.2 Власні значення та власні вектори матриці коваріації (кореляції).....</i>	<i>45</i>
<i>3.1.3 Ортогональні компоненти випадкових метеорологічних об'єктів.....</i>	<i>51</i>
<i>3.1.4 Головні компоненти метеорологічних полів. Задачі стиску та фільтрації вихідної інформації.....</i>	<i>55</i>
3.2 Розв'язання типової задачі. Стиск та фільтрація вихідної інформації за допомогою компонентного аналізу.....	59
3.3 Запитання для самоперевірки.....	70
3.4 Завдання для закріплення теми.....	70
4 ФАКТОРНИЙ АНАЛІЗ.....	72
4.1 Базові теоретичні положення.....	72
<i>4.1.1 Основні засади факторного аналізу.....</i>	<i>72</i>
<i>4.1.2 Оцінки вагів узагальнених факторів та дисперсій залишків.....</i>	<i>75</i>
4.2 Приклади застосування факторного аналізу до гідрометеорологічної інформації.....	80
4.3 Запитання для самоперевірки.....	87
5 КЛАСИФІКАЦІЯ І КЛАСТЕР.....	88
5.1 Базові теоретичні положення.....	88
<i>5.1.1 Основи кластерного аналізу.....</i>	<i>88</i>
<i>5.1.2 Універсальний адаптивний ітераційний метод кластерного аналізу («УАІМКА»).....</i>	<i>94</i>

5.2	Розв'язання типової задачі. Кластеризація гідрометеорологічної інформації.....	96
5.3	Запитання для самоперевірки.....	99
5.4	Завдання для закріплення теми.....	99
6	РЕГРЕСІЙНІ МОДЕЛІ МЕТЕОРОЛОГІЧНОГО ПРОГНОЗУ.....	100
6.1	Базові теоретичні положення.....	100
6.1.1	<i>Прогностична модель лінійної множинної регресії.....</i>	<i>100</i>
6.1.1.1	<i>Постановка задачі.....</i>	<i>100</i>
6.1.1.2	<i>Структура моделі лінійної множинної регресії.....</i>	<i>101</i>
6.1.1.3	<i>Множинний коефіцієнт кореляції.....</i>	<i>106</i>
6.1.1.4	<i>Частинний коефіцієнт кореляції.....</i>	<i>110</i>
6.1.2	<i>Перевірка гіпотези про статистичну значущість коефіцієнтів регресійної моделі.....</i>	<i>114</i>
6.1.3	<i>Методи відбору статистично значущих предикторів.....</i>	<i>117</i>
6.1.3.1	<i>Відбір статистично значущих предикторів на основі множинного коефіцієнта кореляції.....</i>	<i>118</i>
6.1.3.2	<i>«Просіювання» предикторів за методом «включення».....</i>	<i>119</i>
6.1.3.3	<i>«Просіювання» предикторів за методом покрокової регресії.....</i>	<i>122</i>
6.1.4	<i>Оцінка міри адекватності прогностичної моделі метеорологічного прогнозу.....</i>	<i>125</i>
6.2	Розв'язання типової задачі. Побудова лінійного рівняння множинної регресії.....	127
6.3	Запитання для самоперевірки.....	135
6.4	Завдання для закріплення теми.....	136

ДОДАТОК А – Карта України.....	137
ДОДАТОК Б – Вихідні дані для виконання практичних вправ.....	138
ДОДАТОК В – Пояснення до комп'ютерних програм.....	178
ДОДАТОК Г – Значення критерію Ст'юдента.....	182
ДОДАТОК Д – Значення критерію Фішера.....	183
ДОДАТОК Е – Значення величини r_{xy} для значень z	184
ДОДАТОК Ж – Значення величини z для значень r_{xy}	186
ВИКОРИСТАНА ЛІТЕРАТУРА	187
ПРЕДМЕТНИЙ ПОКАЖЧИК	189
ІМЕННИЙ ПОКАЖЧИК	191

Навчальне видання

Гончарова Людмила Дмитрівна

**МЕТОДИ БАГАТОВИМІРНОГО
СТАТИСТИЧНОГО АНАЛІЗУ
МЕТЕОРОЛОГІЧНИХ ПОЛІВ ТА
АТМОСФЕРНИХ ПРОЦЕСІВ**

Навчальний посібник

Підп. до друку
Умовн. друк. арк.

Формат 60×84/16
Тираж

Папір
Зам. №

Надруковано з готового оригінал-макета

Одеський державний екологічний університет
65016, Одеса, вул.Львівська, 15
