

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Одеський державний екологічний університет

Н. С. Лобода

**МЕТОДИ
СТАТИСТИЧНОГО АНАЛІЗУ
У ГІДРОЛОГІЧНИХ РОЗРАХУНКАХ
І ПРОГНОЗАХ**

Навчальний посібник

Рекомендовано
Міністерством освіти і науки України
як навчальний посібник
для студентів вищих навчальних закладів,
які навчаються за напрямом «гідрометеорологія»

Одеса
«Екологія»
2010

ББК 26.22
Л 68
УДК 556.06

Рецензія:

д-р геогр. наук, проф. **О. О. Світличний**;
канд. техн. наук, доц. **С. В. Мельник**;
канд. техн. наук, доц. **В. Г. Явкин**

Гриф надано Міністерством освіти і науки України
(лист № 1/11-477 від 04.02.2010 р.)

Лобода Н. С.

Л 68 Методи статистичного аналізу у гідрологічних розрахунках і
прогнозах: Навчальний посібник / Н. С. Лобода. – Одеса :
Екологія, 2010. – 184 с.

ISBN 978-966-8470-63-3

В навчальному посібнику викладено технічні основи сучасних методів статистичного аналізу характеристик стоку, які використовуються при гідрологічних розрахунках та прогнозуванні. Значну увагу приділено методам просторово-часових узагальнень гідрологічних характеристик та обґрунтування вибору способу узагальнення залежно від об'єму вихідної інформації. У посібнику розглянуто та наведено приклади розрахунків та прогнозів за такими методами як сумісний аналіз даних, методи факторного аналізу та головних компонент, методики парної регресії, дискримінантний аналіз і стохастична фронтальність.

Посібник розраховано на студентів, магістрів, аспірантів гідрологічних, гідромеліоративних та гідротехнічних спеціальностей вищих навчальних закладів.

ББК 26.22
УДК 556.06

ISBN 978-966-8470-63-3

Одеський державний
екологічний університет, 2010

ЗМІСТ

ЗМІСТ	3
ВСТУП	5
1. ПРОБЛЕМИ ТА ЗАДАЧІ МЕТОДІВ ГЕОГРАФО-ГІДРОЛОГІЧНИХ УЗАГАЛЬНЕНЬ	7
2. ФАКТОРНИЙ АНАЛІЗ ПРИ ВИРІШЕННІ ЗАДАЧ ГІДРОЛОГІЧНИХ РОЗРАХУНКІВ	9
2.1. Теоретичні основи методу факторного аналізу	9
2.2. Застосування методу факторного аналізу до районування за синхронністю коливань стоку	12
2.3. Приклади застосування факторного аналізу у гідрологічних розрахунках	
3. МЕТОД ГОЛОВНИХ КОМПОНЕНТІВ У ГІДРОЛОГІЧНИХ РОЗРАХУНКАХ	18
3.1. Теоретичні основи методу головних компонентів	18
3.2. Застосування методу головних компонентів до аналізу синхронності коливань стоку	21
3.3. Вирішення задач фільтрації та відновлення гідрологічної інформації	25
4. МЕТОД СУМІСНОГО АНАЛІЗУ ПРОСТОРОВОЇ ДИСПЕРСІЇ ГІДРОЛОГІЧНИХ ХАРАКТЕРИСТИК	30
4.1. Теоретичні основи методу	30
4.2. Приклади застосування методу сумісного аналізу даних	34
5. РЕГРЕСІЙНІ МОДЕЛІ У ГІДРОЛОГІЧНИХ РОЗРАХУНКАХ	40
5.1 Основні положення регресійного аналізу	45
5.2. Оцінка адекватності регресійної моделі за складовими дисперсії випадкової величини	40

5.3. Визначення частинних коефіцієнтів кореляції	49
5.4. Приклад аналізу розрахунків за моделлю множинної лінійної регресії з покроковим добором предикторів	53

6. ДИСКРИМІНАНТНИЙ АНАЛІЗ ЯК МЕТОД ПРИЙНЯТТЯ АЛЬТЕРНАТИВНИХ РІШЕНЬ ПРИ ГІДРОЛОГІЧНОМУ РАЙОНУВАННІ ТА ПРОГНОЗУВАННІ

6.1. Схема побудови розв'язувального правила	60
6.2. Побудова розв'язувального правила на основі багатовимірного нормального розподілу	65
6.3 Приклади застосування дискримінантної функції у гідрологічних розрахунках та прогнозах	

7. САМОПОДІБНІСТЬ (ФРАКТАЛИ) У ПРОСТОРОВОМУ РОЗПОДІЛІ СТАТИСТИЧНИХ ПАРАМЕТРІВ	69
7.1. Поняття про фрактали	69
7.2. Визначення фрактальних розмірностей за просторовою структурною функцією	71
7.3. Визначення фрактальних розмірностей для часових рядів	74
Література	75
Додаток А	77
Додаток Б	78

ВСТУП

Особливістю гідрологічних явищ є їх обумовленість складними і різноманітними процесами, які проходять впродовж коротких й довготривалих інтервалів часу та відчувають вплив багатьох випадкових чинників. При вивченні гідрологічних об'єктів їхні дослідження виконуються у межах вибірових даних, які отримуються у результаті гідрометричних спостережень. На основі граничних теорем теорії імовірностей прийнято, що дані спостережень мають випадковий характер [33]. Урахувати вплив кожного з природних та антропогенних чинників на формування стоку практично неможливо. Вдається установити лише головні тенденції, які завжди проявляються на фоні випадкових флуктуацій. Якщо ж прийняти, що вибірки характеристик стоку, отримані за даними спостережень, є випадковими величинами, то при дослідженні гідрологічних явищ можна широко використовувати методи математичної статистики [7,33,38]. Первинний статистичний аналіз гідрологічних даних полягає у визначенні оцінок статистичних параметрів та установленні закону розподілу ймовірностей досліджуваної характеристики стоку [14,30,32,37]. Якщо ураховувати наявність внутрішньорядних зв'язків у рядах спостережень, то до статистичних досліджень рядів стоку має сенс залучити теорію випадкових процесів [7,33], де хронологічний ряд тієї чи іншої характеристики стоку розглядається як реалізація стаціонарного ергодичного випадкового процесу, яка описується стохастичною моделлю у вигляді простого ланцюга Маркова [6,16,23].

Оскільки формування річного стоку відбувається в результаті взаємодії різних гідрометеорологічних умов, то сукупність усіх процесів, які розглядаються, утворює гідрометеорологічну систему. Гідрометеорологічна система – це багаторівнева конструкція із взаємозв'язаними елементами. При розгляді системи вирішуються одночасно задачі аналізу і синтезу. Задачі аналізу застосовуються при розгляді окремих елементів системи та їх математичному описі.

Задачі синтезу передбачають об'єднання часткових математичних моделей для відтворення усієї гідрометеорологічної системи. Для створення математичних моделей необхідно мати інформацію про основні просторові та часові закономірності коливань стоку.

При використанні даних спостережень за річковим стоком у гідрологічних розрахунках надійна інформація може бути отримана у випадку, коли тривалість рядів спостережень становить 50-100 років [30,37]. Однак добре вивчені у гідрологічному відношенні водозбори – це рідкість, більшість рядів гідрологічних спостережень значно коротша. Ситуація ускладнюється тим, що водний режим багатьох річок трансформується в результаті водогосподарської діяльності та змін глобального клімату, що переводить такі річки до розряду недостатньо вивчених. Компенсація недостатності даних спостережень відбувається за рахунок просторово-часових узагальнень стоку [20,31]. При таких узагальненнях часто використовують методи багатомірного статистичного аналізу, до яких відносяться метод головних компонентів, факторного аналізу, дискримінантний аналіз та методи множинної регресії. Певний науковий інтерес при виконанні просторово-часових узагальнень гідрологічних величин становлять методи аналізу просторових дисперсій характеристик стоку та їх параметрів [16,17], які використовуються для обґрунтування вибору способу просторового узагальнення, а також аналіз просторових кореляційних та структурних функцій, що відображають основні закономірності просторового розподілу та дозволяють зробити висновки щодо просторової однорідності гідрологічних полів [20,31,38]. Значні перспективи для розробки методів просторово-часової інтерполяції та екстраполяції гідрологічних характеристик на базі вивчення їхньої мінливості на різних просторово-часових масштабах являють методи стохастичної теорії фракталів [18,19,39]. На основі сучасних методів математичної статистики вирішуються різноманітні практичні задачі гідрометеорології, такі як вивчення статистичної структури гідрометеорологічних полів, виявлення зв'язків між атмосферними та гідрологічними процесами, стискання та наступний аналіз гідрометеорологічної інформації, класифікація, районування та прогноз характеристик стоку у різні фази водного режиму.

У сучасний період методи математичної статистики інтенсивно розвиваються. Використання методів багатовимірної статистики передбачає звертання до системного аналізу розглядуваного гідрометеорологічного явища, головних його складових та зв'язків, прийняття рішень про характер установлених закономірностей. Програмно-алгоритмічне забезпечення такого аналізу має відношення до методів штучного інтелекту.

У навчальному посібнику розглядаються найбільш уживані у гідрологічних розрахунках й прогнозах та теоретично досить добре розроблені методи багатовимірного статистичного аналізу даних, наводяться приклади розрахунків. Передбачається, що студенти знайомі з основами первинної статистичної обробки даних. Окремі положення теорії матриць та теорії кореляцій, що використовуються у посібнику, наведені у додатках. Зміст посібника відповідає дисципліні ***“Методи просторового узагальнення гідрологічної інформації”***, головна задача якої полягає в засвоєнні основних положень багатовимірного статистичного аналізу та вивченні їх особливостей при застосуванні у розрахунках та прогнозах стоку.

Більшість розділів підручника може використовуватись студентами та магістрами факультету охорони природного середовища, у тому числі за спеціалізацією “гідроекологія” при вивченні вибіркової дисципліни “Методи математичної статистики в гідроекології”.

Зміст підручника може знайти широке використання у науково-дослідній роботі аспірантів, а також у практичній діяльності фахівців –гідрологів та гідроекологів.

Аналіз просторової та часової мінливості характеристик водного режиму необхідний для наукового обґрунтування роботи водогосподарських систем та вирішення проблем водного господарства, пов'язаних із змінами водних ресурсів в результаті антропогенної діяльності.

Значний внесок у розвиток статистичних методів при вирішенні спеціалізованих практичних задач гідрології внесли А.В.Рождєственський, та В.Л. Складенко (Російський гідрологічний інститут, Санкт-Петербург, Росія); С.М.Крицький та М.Ф. Менкель, Д.Я. Раткович, М.В. Болгов (Інститут водних проблем, Москва, Росія).

При підготуванні цього навчального посібника був використаний досвід викладання методів статистичного аналізу гідрологічних даних Санкт-Петербурзького державного гідрометеорологічного університету, інституту водних проблем Російської академії наук, Російського гідрологічного інституту (м.Санкт-Петербург) та багатолітній особистий досвід автора та співробітників кафедри гідрології суші в Одеському державному екологічному університеті.

Автор виражає подяку професору, д.техн.н, **Євгену Павловичу Школьному** (Одеський Державний Екологічний Університет, Україна) за його успішну та багатоплідну працю у області розвитку та застосування методів математичної статистики й теорії випадкових процесів при вирішенні наукових та практичних задач у гідрометеорології.

Щира подяка виражається інженеру кафедри гідрології суші ОДЕКУ Арестовій Ользі Василівні та завідувачу редакційним відділом Соколенко Олені Денісовні за корисну співпрацю при підготовці рукопису посібника.

1. ПРОБЛЕМИ ТА ЗАДАЧІ МЕТОДІВ ГЕОГРАФО-ГІДРОЛОГІЧНИХ УЗАГАЛЬНЕНЬ

Методи географо-гідрологічних узагальнень використовують при недостатній гідрологічній вивченості розрахункового водозбору або за умов відсутності даних спостережень [16,37]. Географо-гідрологічні дослідження спрямовані на характеристику усереднених (на різному просторово-часовому рівні) умов формування стоку. Для оцінки статистичних параметрів стоку невивчених, з погляду вимірювань стоку, водозборів розробляються карти ізоліній досліджуваних характеристик стоку або виконується їх районування. Для урахування впливу підстильної поверхні будуються залежності, які зв'язують характеристики річкового стоку з кількісними показниками чинників підстильної поверхні. Чинники або фактори формування стоку традиційно діляться на зональні, азональні та інтразональні (місцеві) [4]. Пов'язані з кліматом зональні чинники обумовлюють плавну й безперервну зміну характеристик стоку, просторове узагальнення яких представляється, зазвичай, у вигляді карт ізоліній. Інтразональні та азональні чинники спричиняють дискретність просторового розподілу характеристик стоку. При просторовому узагальненні ця особливість просторової мінливості стокових величин знаходить своє відображення в районуванні території, тобто виділенні ділянок, в межах яких можна прийняти єдине значення розрахункового параметра. Азональні чинники пов'язані з розміром, формою та структурою окремих водозборів й не залежать від географічного положення водозбору. Вплив азональних чинників усувається за допомогою введення поправочних коефіцієнтів до порайонних або знятих з карти ізоліній характеристик стоку. Як правило, для визначення таких поправочних коефіцієнтів установлюються їх регіональні залежності від морфометричних характеристик водозборів.

Основи методу географо-гідрологічних узагальнень закладені В.Г.Глушковым й одержали свій подальший розвиток в роботах

А.М.Бєфані, П.С. Кузіна та В.І. Бабкіна, С.М.Кріцкого та М.Ф.Менкеля, А.В.Рождественського, Г.П. Калініна, А.В. Хрістофорова та інш.

Перші схеми районування території України наводяться в роботах В.Докучаєва, Л. Берга, які були опубліковані в кінці XIX – на початку XX сторіч. На сучасному етапі досліджень просторово-часових узагальнень характеристик стоку слід відзначити роботи А. Марініча, А. Ланька, П.Шищенко, Р. Міллера.

Основні задачі методу географічного узагальнення можна сформулювати таким чином.

1. З'ясування доцільності географічних узагальнень.
2. Установлення необхідного ступеня географічного узагальнення.
3. Вибір способу географічного узагальнення.
4. Оцінка меж географічного узагальнення.
5. Пошук доцільних та обґрунтованих форм опису просторового розподілу характеристик річкового стоку (проблема картографування або районування).

Вибір методу картографування характеристик стоку пов'язаний з проблемою співвідношення між впливом зональних та азональних чинників. У тих випадках, коли роль зональних чинників переважає, застосовується принцип географічної інтерполяції й будується карта ізоліній. При використанні районування передбачається, що в межах виділеного району спостерігається відносна однорідність формувальних чинників стоку. При гідрологічному районуванні слід урахувати масштаб районування, оскільки такі таксономічні одиниці, як фізико-географічний пояс, країна, зона, провінція, підзона не завжди знаходять своє відображення в масштабах гідрологічних районів. Річ у тому, що гідрологічне районування найчастіше спирається на переважаючий вплив одного чинника формування стоку, тоді як при фізико-географічному районуванні переважає принцип комплексності або ландшафтно – генетичний принцип.

У теперішній час широко використовуються нові статистичні показники, ефективні способи аналізу та синтезу явищ, оцінки надійності й стійкості географічних узагальнень.

Першим кроком географічного узагальнення є аналіз вихідних даних та вибір репрезентативних рядів стоку [23,26]. На другому

етапі виконується аналіз закономірностей коливань стоку та виділяються райони з синхронними коливаннями стоку [24,25]. Надалі на підставі виділених районів виконується приведення коротких рядів стоку до довгого періоду спостережень за методом аналогії та аналізуються закономірності просторового розподілу стоку у кожному з районів. При переважанні впливу географічних чинників на просторовий розподіл характеристик стоку виконується картування, або розробляються розрахункові залежності досліджуваних характеристик від установлених чинників за регресійними моделями. Якщо характеристика стоку визначається за даними спостережень з малою достовірністю, то розглядається можливість її районування, тобто осереднення у межах розглядуваної території. При проведенні границь гідрологічних районів використовуються класифікаційні методи, які дозволяють вирішити проблему належності приграничних водозборів до того чи іншого класу.

2. ФАКТОРНИЙ АНАЛІЗ ПРИ ВИРІШЕННІ ЗАДАЧ ГІДРОЛОГІЧНИХ РОЗРАХУНКІВ

2.1 Теоретичні основи методу факторного аналізу

В факторному аналізі висувається гіпотеза про те, що *дані спостережень є лише непрямыми характеристиками явища, яке вивчається, і це явище можна описати за допомогою невеликої кількості деяких параметрів або властивостей. Такі теоретичні параметри або властивості називаються факторами. Загальний фактор – невимірювана (гіпотетична) скрита величина, яка ураховує кореляцію щонайменше між двома спостереженими змінними* [12,27,36,38].

Фактори є однаковими для всіх розглядуваних гідрометеорологічних величин, але входять в кожну з них із своєю вагою. Вони не повністю описують вихідні змінні. Залишається частина інформації, яку називають залишками. *Позитивна якість методу факторного аналізу полягає в тому, що безліч корельованих змінних описується набагато меншим числом факторів.*

Задача факторного аналізу - представити дані спостережень у вигляді лінійних комбінацій факторів:

$$x_j = \sum_{p=1}^k l_{jp} f_p + v_j, \quad (j=1, m) \quad (2.1)$$

де x_j - центрована початкова змінна;

m - кількість змінних;

k - число факторів ($k \ll m$);

p - номер фактора;

l_{jp} - навантаження j -тої змінної на p -тий фактор або

навантаження p -того фактора на j -ту змінну (факторна вага);

v_j - незалежні залишки (частина даних, яка не описується кінцевим числом факторів).

Якщо у рівності (2.1) розгорнути суму, то прийдемо до системи рівнянь виду:

де $j = 1, 2, \dots, m$.

$$V = \begin{bmatrix} v_1 \\ v_2 \\ \dots \\ v_m \end{bmatrix}; \quad (2.5)$$

$$L = \begin{bmatrix} l_{11} & l_{12} & \dots & l_{1k} \\ l_{21} & l_{22} & \dots & l_{2k} \\ \dots & \dots & \dots & \dots \\ l_{m1} & l_{m2} & \dots & l_{mk} \end{bmatrix}. \quad (2.6)$$

Всі вхідні величини x_j виражаються через однакові випадкові величини $f_p (p=\overline{1,k})$, але з різними ваговими коефіцієнтами l_{jp} .

Матрична форма рівняння (2.1) має вигляд

$$X = LF + V, \quad (2.7)$$

де X - матриця центрованих вихідних величин;

L - матриця факторних навантажень;

F - матриця факторів;

V - матриця незалежних залишків.

Матриця коваріацій знаходиться як

$$K = M[X \cdot X'] , \quad (2.8)$$

де X' - транспонована матриця центрованих вихідних величин (додаток Б).

Виражаючи X через (2.7), можна прийти до такого матричного виразу [38]

$$K = LL' + D, \quad (2.9)$$

де K - матриця коваріацій;

L - матриця факторних навантажень;

L' - транспонована матриця факторних навантажень L ;

D - діагональна матриця (додаток А), яка складається з дисперсій e_j незалежних залишків й має вигляд

$$D = \begin{bmatrix} e_1 & 0 & \dots & 0 \\ 0 & e_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & e_m \end{bmatrix} \quad (2.10)$$

Таким чином, коваріаційну матрицю можна виразити через матрицю вагових навантажень на фактори та діагональну матрицю дисперсій залишків. Фактори ураховують зв'язок між змінними, тобто вони представляють структуру коваріаційної або кореляційної матриці в термінах факторної моделі. Проте самі фактори представляються некорельованими (ортогональними).

Елементи матриці коваріацій можуть бути представлені через факторні ваги та залишкові дисперсії наступним чином

$$K_{jj} = \sum_{p=1}^k l_{jp}^2 + e_j, \quad \text{коли } j = i; \quad (2.11)$$

$$K_{ji} = \sum_{p=1}^k l_{jp} l_{ip}, \quad \text{коли } j \neq i. \quad (2.12)$$

Оскільки в матриці коваріацій на діагоналі розташовані дисперсії змінних, то можна зробити висновок, що квадрати факторних навантажень l_{ip}^2 є частками дисперсій змінних, які описуються відповідними факторами.

Коли вибіркова кореляційна матриця відома, рівняння (2.11) дає можливість знайти дисперсії залишків після розрахування вагових коефіцієнтів факторів:

$$e_j = K_{jj} - \sum_{p=1}^k l_{jp}^2. \quad (2.13)$$

Сума квадратів факторних навантажень по всіх виділених факторах може бути записаною таким чином

$$h_j^2 = \sum_{p=1}^k l_{jp}^2 . \quad (2.14)$$

Отримана величина h_j^2 визначає повноту відображення j -тої змінної в усіх факторах f_p й має назву сумарного внеску факторів.

Повний внесок S_p (у відсотках) фактора у сумарну дисперсію змінних визначається виразом

$$S_p = \frac{\sum_{j=1}^m l_{pj}^2}{m} 100\%, \quad (2.15)$$

де m - кількість розглядуваних змінних.

Загальний внесок всіх виділених факторів в сумарну дисперсію досліджуваних змінних дорівнює

$$S = \sum_{p=1}^k S_p . \quad (2.16)$$

Величина S має назву міри факторизації. Як правило, при практичному застосуванні до розгляду приймається така кількість факторів, яка описує понад 70% сумарної дисперсії змінних.

Оцінити факторні ваги (факторне навантаження) можна за методом найбільшої правдоподібності [27]. Але цей метод потребує великих обчислень. На практиці найбільш поширений апроксимаційний метод визначення факторних навантажень, який має назву центроїдного метода [12,38]. Визначення самих факторів відбувається як пошук лінійної функції фактора f_r від вихідної змінної x_j (Томсон, 1951).

2.2 Застосування методу факторного аналізу до районування за синхронністю коливань стоку

Синхронними називають коливання стоку річок, на яких спостерігається однаковий хід водності протягом всього інтервалу часу, а **асинхронними** - коливання стоку, які мають протилежний хід водності. Під синфазністю та асинфазністю стоку розуміють однаковий або протилежний хід коливань не на всьому розглядуваному інтервалі часу, а по періодах водності (по групах багатоводних та маловодних років).

Кількісною мірою синхронності коливань стоку є коефіцієнт кореляції між двома рядами. Коливання вважаються синхронними, якщо коефіцієнт кореляції перевищує 0,7 [30], й синфазними, коли коефіцієнт кореляції менший 0,7, але більший 0,4. Якщо коефіцієнт кореляції менший 0,4 коливання вважаються несинфазними.

При аналізі синхронності коливань стоку використовується геометрична інтерпретація факторів (Q- модифікація факторного аналізу), яка описана у роботах В.А. Жука та В.М. Євстигнєєва [10,11]. За цим методом було виконане районування України за синхронністю коливань річного стоку [21]. При графічних побудованнях координатні осі розглядаються як фактори. Кожний водозбір може бути представлений як вектор, проведений з початку координат у точку з координатами, які відповідають навантаженню кожного фактора на цей водозбір. Кут між двома векторами - водозборами є мірою кореляції стоку між цими водозборами. У випадку, коли перших два фактора описують більше 70% загальної дисперсії вихідних даних, на графіку, осі якого являють собою два фактори, проводять вектори з початку координат у точку з координатами, які відповідають факторним навантаженням. Довжина такого вектора розраховується за виразом

$$d_j = \sqrt{l_{j1}^2 + l_{j2}^2}, \quad (2.17)$$

де l_{j1} і l_{j2} - вагові коефіцієнти (факторні навантаження) першого та другого факторів на j - тий водозбір.

Величина d ототожнюється з h . Вона визначає повноту відображення j -го ряду спостережень першими двома факторами, а косинус кута між i -тим та k -тим вектором є коефіцієнтом кореляції між ними

$$r_{i,k} = \cos \varphi_{i,k}. \quad (2.18)$$

Таким чином, про ступінь зв'язку між рядами стоку можна судити по угрупованнях точок, які утворюються на площині [21]. Як міра схожості в даному випадку використовується міра відстані: чим ближче розташовані точки на графіку й менше косинус кута між ними, тим ближче значення коефіцієнта кореляції до одиниці.

З метою виконання одночасного аналізу трьох ефективних факторів рекомендується представляти навантаження не в декартових, а в полярних координатах [11]

$$\theta_j = \arcsin \frac{l_{j3}}{d_j} \quad (2.19)$$

$$\lambda_j = \arcsin \frac{l_{j2}}{\sqrt{l_{j1}^2 + l_{j2}^2}}, \quad (2.20)$$

де

$$d_j = \sqrt{l_{1j}^2 + l_{2j}^2 + l_{3j}^2}, \quad (2.21)$$

де θ та λ - полярні (сферичні) координати (в градусах або радіанах), які визначають положення пересікання j -тим вектором одиничної сфери.

Величина d_j оцінює внесок усіх трьох факторів у формування стоку j -того водозбору, включеного в аналіз. Якщо розглядається не матриця коваріацій, а кореляцій, де діагональні елементи дорівнюють 1 (додаток А), то близькість d_j до одиниці указує на те,

що дисперсія даної змінної значною мірою пояснюється першими трьома факторами.

Таким чином, *графічна інтерпретація факторів дозволяє спростити аналіз інформації, яка міститься в кореляційній або коваріаційній матриці*. Компактне трактування кореляційної матриці досягається при розгляді кожного угруповання водозборів. При наближенні кута між векторами, спрямованими з початку координат до центрів угруповань, до 90° , виділені угруповання розглядаються як райони з асинхронними коливаннями стоку. Чим менший кут між угрупованнями, тим тісніший зв'язок між стоком розглядуваних річок. Усередині угруповань (районів) також можна виділяти окремі групи точок, які по їх територіальному розташуванню й особливостям формування стоку можна інтерпретувати як підрайони.

2.3 Приклади застосування факторного аналізу у гідрологічних розрахунках

Приклад 1.

Завдання. Виділити райони з синхронними коливаннями стоку, використовуючи Q -модифікацію факторного аналізу у басейні р. Сіверський Донець - м.Лисичанськ (рис.2.1).

Вихідна інформація. Розглянуто 12 водозборів з рядами спостережень за річним стоком довжиною у 42 роки (1961-2002 рр.). Після застосування факторного аналізу встановлено, що перший фактор описує 45,3% вихідних даних, другий – 34,4%, третій – 20,4% . Таким чином, при залученні до аналізу тільки перших двох факторів міра факторизації становить 79,7%, а перших трьох – 88,6% (Н.С. Лобода, О.В. Бабаєва, 2008).

Розв'язок задачі. Графічні побудовання (рис. 2.2, рис. 2.3) дозволяють зробити висновок, що коливання стоку несинхронні, оскільки кут між центрами виділених угруповань на рис.2.2 становить близько 70° ($\cos 70^\circ = 0,34$). Район I охоплює верхню частину водозбору річки Сіверський Донець, включаючи р. Оскіл. Річки району I беруть початок у межах Середньоруської височини,

у лісостеповій зоні. До району II віднесені річки Донецького Кряжу. Проміжне положення між цими районами займають лівобережні притоки степової зони (номери водозборів - 36,40,41).

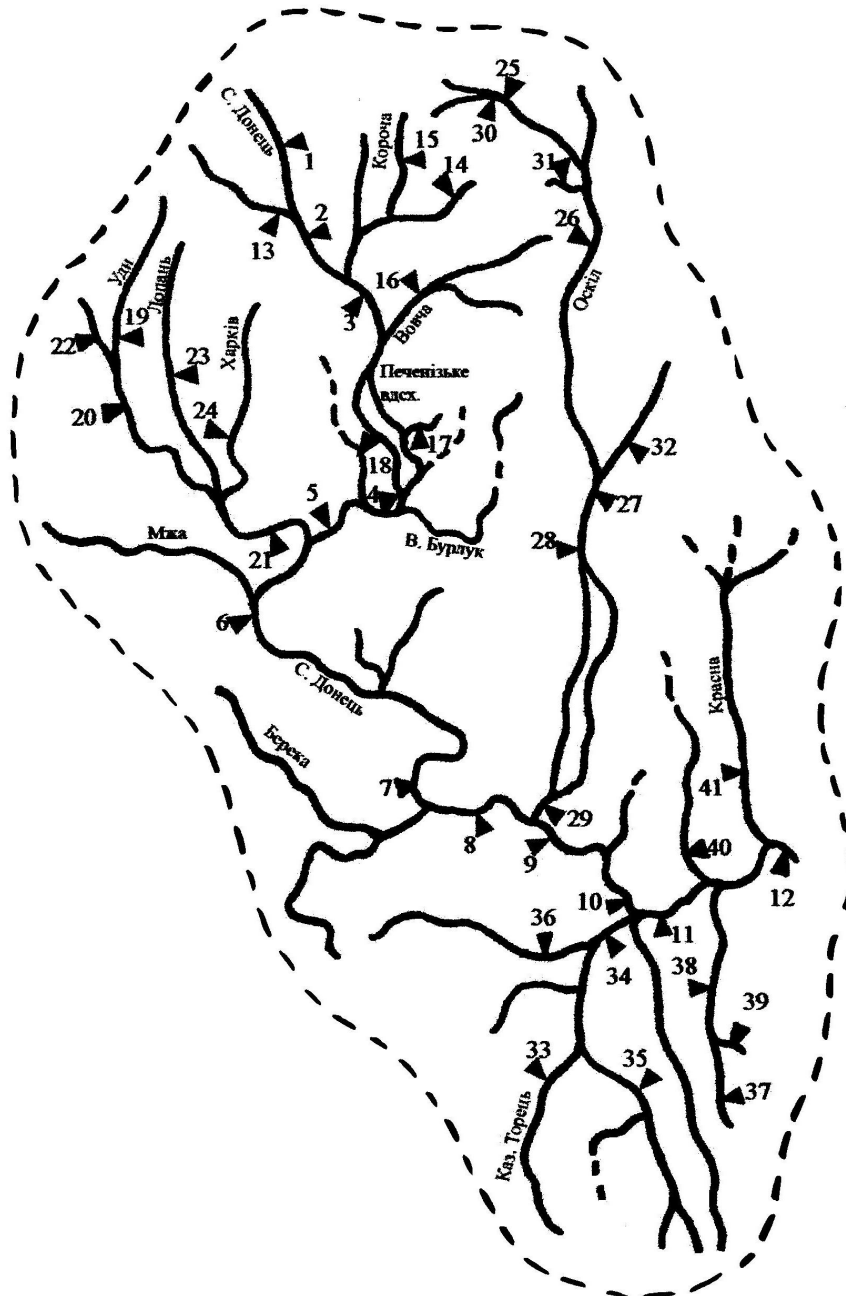


Рисунок 2.1 - Карта-схема гідрологічних постів в басейні р. Сіверський Донець –м.Лисичанськ

Таблиця 2.1 – Факторні ваги та коефіцієнти інформативності
для перших двох та трьох факторів

№ поста	Річка - пункт	l_{1j}	l_{2j}	d_j	l_{3j}	d_j	θ_j	λ_j
5	Сів. Донець – м.Чугуїв	0.20	-0.78	0.81	-0.48	0.88	-0.54	-1.32
16	Вовча – м.Вовчанськ	0.12	-0.90	0.91	-0.20	0.86	-0.22	-1.44
21	Уди – с.Безлюдівка	0.08	-0.30	0.31	-0.89	0.89	-1.24	-1.32
23	Лопань – с.Казаха Лопань	0.39	-0.44	0.58	-0.73	0.87	-0.90	-0.83
28	Червоний Оскіл – м. Куп'янськ	0.29	-0.84	0.89	-0.31	0.88	-0.33	-1.24
33	Казенний Торець – м. Райське	0.90	-0.19	0.92	-0.25	0.91	-0.26	-0.21
34	Казенний Торець – м.Слав'янськ	0.92	-0.14	0.93	-0.18	0.90	-0.19	-0.15
35	Кривий Торець – с.Олексіїво-Дружківка	0.71	0.08	0.70	-0.53	0.78	-0.64	0.11
36	Сухий Торець – с.Черкаське	0.78	-0.53	0,94	0.12	0.90	0.12	-0.60
38	Бахмутка – м.Яма (Сіверськ)	0.86	-0.33	0.92	-0.28	0.92	-0.28	-0.37
40	Жеребець - с.Торське	0.78	-0.59	0.98	-0.01	0.96	-0.01	-0.64
41	Красна – с.Краснопопівка	0.64	-0.67	0,93	-0.11	0.87	-0.12	-0.80

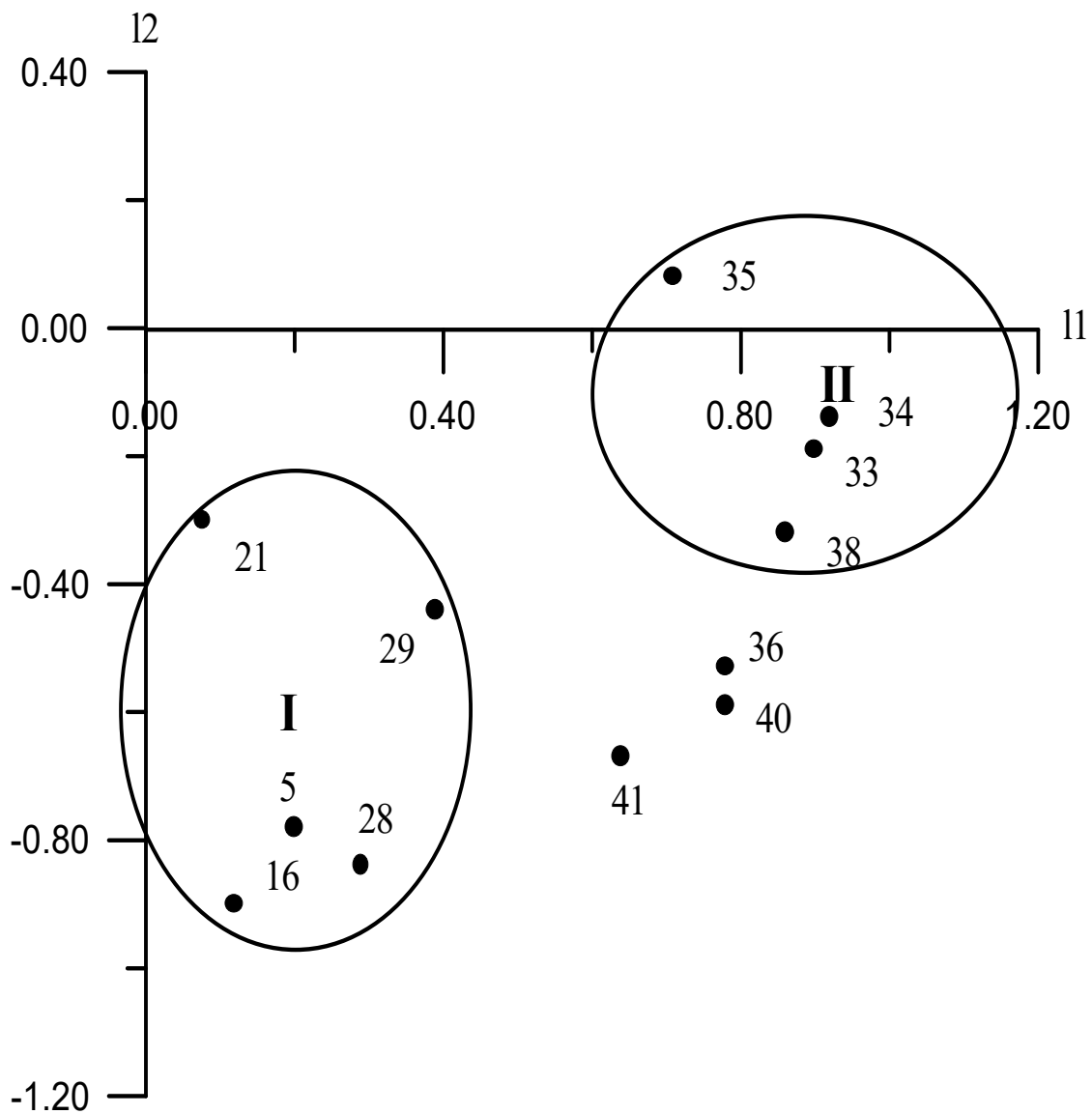


Рисунок 2.2 – Виділення угруповань з синхронними коливаннями стоку на основі двох факторів

I – район Середньоруської височини;
II – район Донецького Кряжу

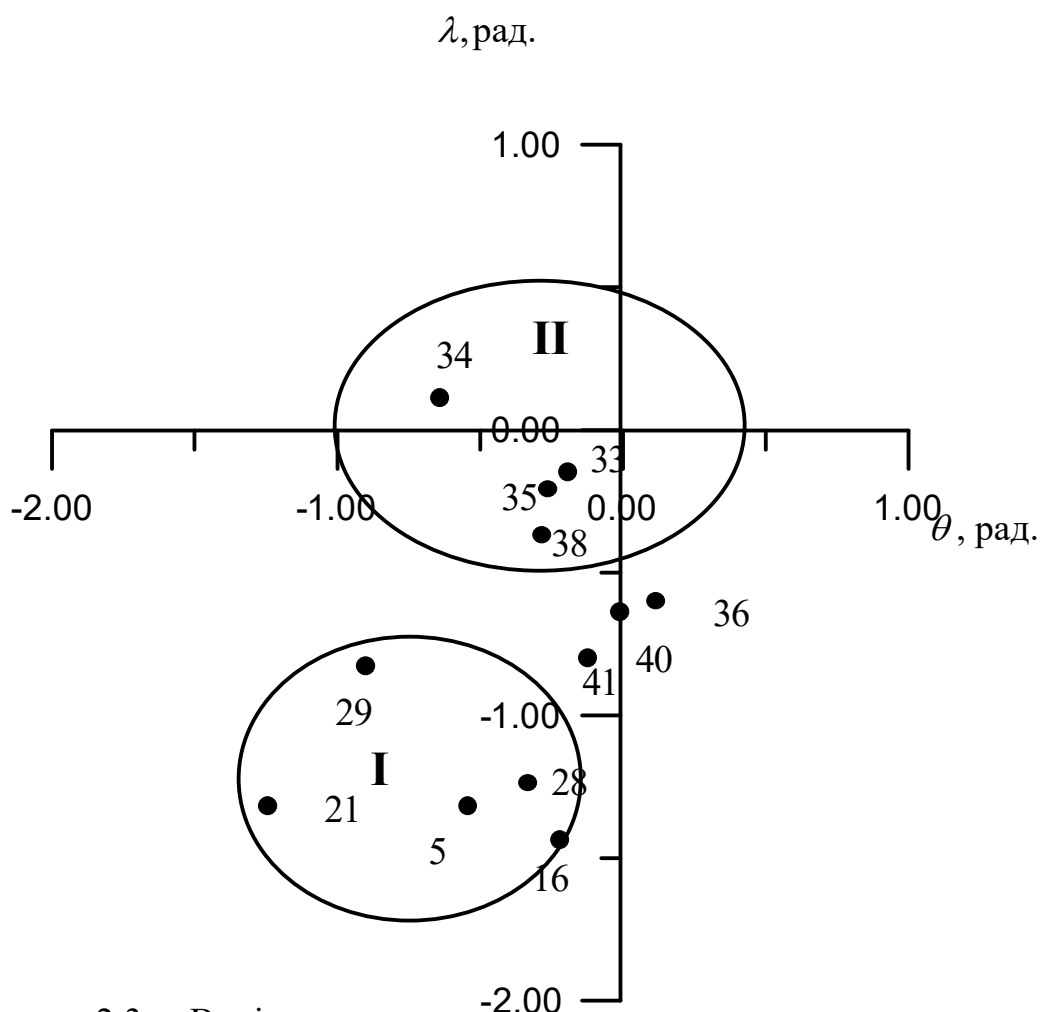


Рисунок 2.3 – Виділення угруповань з синхронними коливаннями стоку на основі трьох факторів

I – район Середньоруської височини;
II – район Донецького Кряжу

Достовірність отриманих результатів підтверджується таким чином. Осереднений коефіцієнт кореляції у межах усього водозбору р.Сіверський Донець - м. Лисичанськ дорівнює 0,34. У той же час осереднений коефіцієнт кореляції становить 0,71 у межах району I та 0,76 –у межах району II.

Відмінність у характері коливань річного стоку районів I та II можна пояснити як особливостями природних умов формування стоку, так і значним значного впливом скидання шахтних вод у поверхневі водотоки Донбасу.

Результати районування можна використовувати при подальшому аналізі й узагальненні даних по стоку, приймаючи до

уваги той факт, що стік річок Донбасу є суттєво перетвореним водогосподарською діяльністю.

Приклад 2.

Завдання. Виявити головні чинники, які обумовлюють просторовий розподіл середньо багаторічних величин річного стоку на основі застосування факторного аналізу до чинників формування річного стоку в басейні р. Сіверський Донець – м.Лисичанськ.

Вихідна інформація. Розглянуто 38 водозборів басейну р.Сіверський Донець із такими змінними: середньобагаторічний шар стоку $\bar{Y}$, логарифм площі водозбору $\lg(F+1)$, середня висота водозбору $H_{сер}$, середній ухил річки $I_{сер}$, середньозважений ухил водозбору $I_{сер.зв}$, залісеність f_l . В результаті розрахунків

отримана кореляційна матриця такого вигляду

Таблиця 2.2. - Коефіцієнти кореляції між нормами річного стоку та гідрографічними і морфологічними показниками

	$\bar{Y}$	$\lg(F+1)$	$H_{сер}$	$I_{сер}$	$I_{сер.зв}$	f_l
$\bar{Y}$	1,00	-0,28	0,65	0,11	0,11	-0,16
$\lg(F+1)$	-0,28	1,00	-0,56	-0,73	-0,70	0,17
$H_{сер}$	0,65	-0,56	1,00	0,30	0,31	-0,19
$I_{сер}$	0,11	-0,73	0,30	1,00	0,93	-0,17
$I_{сер.зв}$	0,11	-0,70	0,31	0,93	1,00	-0,23
f_l	-0,16	0,17	-0,19	-0,17	-0,23	1,00

Розв'язок задачі. До аналізу кореляційної матриці був залучений метод факторного аналізу та установлені факторні ваги й залишки (табл. 2.3).

Перший гіпотетичний фактор описує 42,8% сумарної дисперсії вихідних даних та обумовлює наявність зв'язків між площею та ухилами річок і водозборів.

Другий гіпотетичний фактор описує 30,9% вихідної інформації та обумовлює наявність зв'язків між нормою стоку $\bar{Y}$ ($l_{2\bar{Y}} = 0,906$) та середньою висотою $H_{сер}$ ($l_{2H_{сер}} = 0,871$).

Оскільки l_{2f_l} має від'ємний знак ($l_{2f_l} = - 0.309$), можна зробити висновок про протилежний у порівнянні із середньою висотою $H_{сер}$ вплив залісеності f_l на середньо багаторічну величину річного стоку $\bar{Y}$.

Таким чином, у якості головного чинника розрахункових моделей середньо багаторічної величини річного стоку можна використовувати середню висоту водозборів – $H_{сер}$.

Таблиця 2.3 – Факторні ваги та залишки

Змінна	l_{1j}	l_{2j}	v_j
$\bar{Y}$, мм	0,0341	0,906	-0,00140
$lg(F + 1)$	0,792	-0,399	-0,00134
$H_{сер}$, М	-0,251	0,871	-0,00140
$I_{сер}$, ‰	-0,959	0,0886	-0,00158
$I_{сер.зв.}$, ‰	-0,956	0,101	-0,00158
f_l , %	0,211	-0,309	-0,00240

Згідно з (2.1) величину річного стоку можна виразити через перші два фактори наступним чином

$$t_{\bar{Y},j} = 0,0341f_1 + 0,906f_2 - 0.00140, \quad (2.22)$$

де

$$t_{\bar{Y},j} = \frac{\bar{Y}_j - \bar{Y}_{сер}}{\sigma_{\bar{Y}}}, \quad (2.23)$$

причому $\bar{Y}_{сер}$ – осереднена у межах розглядуваної території величина $\bar{Y}$.

Контрольні запитання

1. У чому полягає задача факторного аналізу.
2. У чому полягає мета факторного аналізу.
3. Яка величина називається мірою факторизації.
4. Як визначається кількість факторів, достатніх для опису вихідної інформації.
5. Геометрична інтерпретація факторів при дослідженні синхронності коливань стоку.

3. МЕТОД АНАЛІЗУ ПРОСТОРОВОЇ ДИСПЕРСІЇ ГІДРОЛОГІЧНИХ ХАРАКТЕРИСТИК (МЕТОД СУМІСНОГО АНАЛІЗУ ДАНИХ)

3.1 Теоретичні основи методу

У гідрологічних розрахунках найчастіше використовується математична модель стоку, яка описує його випадкову природу. Такого роду моделі називаються *стохастичними* [5,16,23,32,33]. Вони включають до себе ряд гіпотез, які дозволяють звести розрахунки стоку до статистичної оцінки декількох параметрів моделі: середнього арифметичного значення $\bar{x}$ досліджуваної гідрологічної величини, коефіцієнту варіації C_v , коефіцієнту асиметрії C_s , коефіцієнту автокореляції $r(1)$.

Навіть при довгих рядах спостережень оцінки окремих статистичних параметрів визначаються з великою похибкою, тобто є статистично незначущими. До числа таких параметрів відносяться насамперед коефіцієнти автокореляції $r(1)$ й асиметрії C_s , а також розрахункове відношення C_s / C_v .

Обмеженість у часі наявних спостережень по більшості рядів стоку річок України породжує статистичну нестійкість цих параметрів, що може бути ефективно компенсоване за рахунок додаткової інформації про просторові закономірності розподілу розглядуваних характеристик річкового стоку. Для підвищення надійності оцінок статистичних параметрів за вибірковими даними рекомендується виконувати їхнє просторове узагальнення [37]. За допомогою методу, запропонованого С.М. Крицьким та М.Ф. Менкелем [16,17], можна обґрунтувати характер цього узагальнення. Цей метод отримав назву "методу сумісного аналізу даних". Суть методу зводиться до визначення географічної й випадкової складових загальної просторової дисперсії розглядуваного статистичного параметра A :

$$\sigma_H^2 = \sigma_G^2 + \sigma_B^2, \quad (3.1)$$

де σ_{II}^2 - повна просторова дисперсія розглядуваного параметра А;
 $\sigma_{Г}^2$ - географічна складова просторової дисперсії розглядуваного параметра А;
 σ_B^2 - випадкова складова просторової дисперсії розглядуваного параметра А.
 При цьому повна просторова дисперсія параметра А оцінюється за формулою

$$\sigma_{II}^2 = \frac{\sum_{j=1}^k (A_j - A_{CEP})^2}{k-1}, \quad (3.2)$$

де k - число об'єктів (водозборів), об'єднаних в одну групу;
 j - порядковий номер розглядуваного об'єкту (водозбору);
 A_j - індивідуальна оцінка параметра (оцінка, виконана по ряду спостережень на окремому j -тому водозборі);
 A_{CEP} - осереднена в межах виділеної групи оцінка параметра А.

Випадкова складова просторової дисперсії параметра визначається як осереднена за групою виділених об'єктів дисперсія індивідуальної оцінки параметра

$$\sigma_B^2 = \frac{\sum_{j=1}^k \sigma_{Aj}^2}{k}, \quad (3.3)$$

де σ_{Aj} - середнє квадратичне відхилення індивідуальної оцінки параметра A_j (додаток В).

Географічна складова знаходиться за допомогою зворотного розрахунку з (3.1):

$$\sigma_{Г}^2 = \sigma_{II}^2 - \sigma_B^2. \quad (3.4)$$

Якщо виконується умова

$$\frac{\sigma_B^2}{\sigma_{II}^2} > \frac{\sigma_I^2}{\sigma_{II}^2}, \quad (3.5)$$

то можна зробити висновок, що просторовий розподіл досліджуваного параметра в більшій мірі визначається випадковими властивостями поєднаних вибірок і в меншій - зміною фізико-географічних умов формування стоку на території. Таким чином, при виконанні умови (3.5) приймається рішення, що вибіркові оцінки параметрів можуть бути осереднені в межах досліджуваної території.

Необхідно підкреслити, що якість об'єднання тим вища, чим менший внесок географічної складової у повну просторову дисперсію параметра. Географічна складова є, власне кажучи, оцінкою статистичної неоднорідності вихідного матеріалу. Якість об'єднання, як правило, є високою, коли внесок випадкової складової у просторову дисперсію більше або дорівнює 70%, тобто

$$\frac{\sigma_B^2}{\sigma_{II}^2} \cdot 100 \geq 70\%. \text{ Коли похибки оцінки вибіркових параметрів дуже}$$

великі, географічна складова дисперсії, обчислена зворотним розрахунком за (3.4), може приймати негативні значення. У цьому випадку внесок випадкової складової у повну просторову дисперсію параметра може бути прийнятий рівним 100% , а географічної – 0,00%.

Середнє квадратичне відхилення осередненої у просторі оцінки статистичного параметра розраховується за співвідношенням

$$\sigma_{CEP} = \sqrt{\frac{\sigma_B^2}{k} + \sigma_I^2} . \quad (3.6)$$

Величина σ_{CEP} , поряд з умовою (3.5), також є критерієм якості об'єднання. Осереднена оцінка параметра визнається статистично достовірною, коли виконується умова

$$A_{CEP} > 2\sigma_{CEP} . \quad (3.7)$$

У ході послідовного об'єднання параметрів можна виявити ряди, статистичні властивості яких відрізняються від властивостей об'єднуваної сукупності: у міру збільшення числа поєднуваних об'єктів k при незначному зростанні географічної складової σ_G^2 дисперсія осередненого у межах поєднуваної сукупності параметра $\sigma_{СЕР}^2$ відповідно до виразу (3.6) повинна зменшуватися. “Сплеск” в убутній функції $\sigma_{СЕР}^2 = \varphi(k)$ свідчить про те, що досліджуваний параметр k -того ряду стоку значно відрізняється від осередненої у просторі оцінки та оцінок відповідного параметру інших рядів. Для таких рядів у наступних розрахунках рекомендується використовувати не осереднену, а уточнену оцінку параметра, за винятком випадків, коли досліджуваний ряд є статистично неоднорідним унаслідок водогосподарських перетворень або відноситься до іншого району, якому притаманні інші властивості.

Для оцінки якості розрахунків поряд з виразом (3.6) також використовуються так звані допустимі відносні середні квадратичні відхилення $\varepsilon_{ДОП.}$ оцінки параметра A за вибірковими даними. Якщо $\varepsilon_A \leq \varepsilon_{ДОП.}$, то вибіркове значення параметра приймається до розрахунку. Величина ε_A визначається за формулою

$$\varepsilon_A = \frac{\sigma_A}{A} \cdot 100\% , \quad (3.8)$$

де σ_A - середнє квадратичне відхилення оцінки параметра A .

Величини ε_A та $\varepsilon_{ДОП.}$ використовуються у гідрологічних розрахунках для оцінювання якості об'єднання таких статистичних параметрів як середнє арифметичне значення та коефіцієнт варіації. Для статистичних параметрів, що розраховуються за даними спостережень з великим середньоквадратичним відхиленням, осереднена в межах поєднуваної сукупності оцінка є більш достовірною, ніж індивідуальна. Осереднені у межах статистично однорідних районів оцінки статистичних параметрів рекомендуються до використання при побудові стохастичних моделей, а також при описуванні статистичних розподілів характеристик стоку тих водозборів, на яких спостереження за

стоком відсутні, або суттєво порушені водогосподарською діяльністю.

Уточнена по сукупності розглянутих об'єктів оцінка статистичного параметра розраховується на основі виразу

$$A'_j = \frac{A_j \sigma_{CER}^2 + A_{CER} \sigma_j^2}{\sigma_{CER}^2 + \sigma_j^2}, \quad (3.9)$$

де A'_j - уточнена оцінка індивідуального значення параметра A_j з урахуванням інформації, яка увійшла в поєднувану сукупність;

A_j - вихідне значення параметра по j -тому розглянутому об'єкту (водозбору);

σ_j^2 - дисперсія оцінки параметра A_j по j -тому розглянутому об'єкту (водозбору);

A_{CER} - осереднена в межах виділеної групи об'єктів оцінка параметра A ;

σ_{CER}^2 - дисперсія осередненої у межах виділеної групи об'єктів оцінки статистичного параметра.

Середнє квадратичне відхилення уточненого значення параметра визначається на основі наступної формули, отриманої за методом статистичних випробувань[16,17]

$$\sigma'_j = \frac{\sigma_j \sigma_{CER}}{\sqrt{\sigma_j^2 + \sigma_{CER}^2}}, \quad (3.10)$$

де σ'_j - середнє квадратичне відхилення уточненої оцінки параметра

A'_j по j -тому розглянутому об'єкту (водозбору);

σ_j - вихідне середнє квадратичне відхилення оцінки параметра A_j по j -тому розглянутому об'єкту (водозбору).

Метод С.М. Крицького та М.Ф. Менкеля дозволяє вирішувати багато задач географічного узагальнення. Обґрунтування вибору

способу узагальнення статистичних параметрів річного стоку та районування території України для параметрів $C_V, C_S / C_V, r(1)$ наведено у роботах Н.С.Лободи [21]. Наприклад, задача вибору способу географічного узагальнення може бути вирішена при розгляді умови (3.5). Якщо умова виконується, то як спосіб географічного узагальнення вибирається районування, тобто розглядувана характеристика осереднюється у межах виділеної території, якщо не виконується – представляється у вигляді карти ізоліній. **Визначення меж географічного узагальнення може спиратися на перевірку виконання умови (3.7) та аналіз залежності $\sigma_{CER}^2 = \varphi(k)$. Зростання географічної складової повної просторової дисперсії параметра, а, отже, й зростання σ_{CER} , буде тим інтенсивнішим, чим ширше межі просторового узагальнення.**

Слід зазначити, що перед застосуванням методу бажано провести попередній аналіз вихідної інформації, використовуючи для виділення початкових угруповань вже існуючі географічні узагальнення, наприклад фізико-географічне районування або районування за синхронністю коливань стоку.

На першому етапі узагальнень може бути прийнята гіпотеза про те, що не тільки крива розподілу, а й статистичні параметри усіх розглядуваних рядів належать до однієї генеральної сукупності. Надалі для окремих параметрів межі установлених статистично однорідних районів можуть розширюватися. Чим більший вплив підстильної поверхні на формування стоку, тим, як правило, менші просторові масштаби виділених районів.

4.2 Вплив кореляційних зв'язків між рядами стоку на результати просторових узагальнень за методом сумісного аналізу

Наявність кореляційних зв'язків між стоком близько розташованих річок обумовлює наявність зв'язків між статистичними параметрами стоку цих річок, визначеними за даними спостережень. Згідно з дослідженнями Є.Г. Блохінова (1970) кореляційний зв'язок між середніми арифметичними значеннями випадкових величин X та Y , статистичний розподіл яких

описується трьохпараметричним гамма-розподілом [17,21], дорівнює величині коефіцієнту кореляції між рядами випадкових величин X та Y , тобто

$$r_{\bar{x},\bar{y}} = r_{X,Y}; \quad (3.11)$$

справедливий вираз

$$r_{C_{Vx},C_{Vy}} = r_{X,Y}^2; \quad (3.12)$$

а для коефіцієнтів асиметрії C_S та відношення C_S / C_V

$$r_{C_{Sx},C_{Sy}} = r_{X,Y}^3, \quad (3.13)$$

$$r_{C_S / C_{Vx}, C_S / C_{Vy}} = r_{X,Y}^3, \quad (3.14)$$

де $r_{\bar{x},\bar{y}}$; $r_{C_{Vx},C_{Vy}}$; $r_{C_{Sx},C_{Sy}}$; $r_{C_S / C_{Vx}, C_S / C_{Vy}}$ - коефіцієнти кореляції, які характеризують тісноту зв'язку між статистичними параметрами двох рядів випадкових величин X та Y ;

$r_{X,Y}$ - коефіцієнт кореляції між двома рядами стоку X та Y .

Наявність зв'язку між стоком річок зменшує об'єм незалежної інформації, що обумовлює зниження випадкової складової просторової дисперсії параметру. Урахування впливу кореляції між стоком річок, які входять у одне угруповання, відбувається за формулою

$$\sigma_B^{2'} = \sigma_B^2 (1 - r_{ACEP}), \quad (3.15)$$

де $\sigma_B^{2'}$ - уточнена випадкова складова просторової дисперсії параметру A з урахуванням впливу кореляції між рядами;

σ_B^2 - вихідне значення випадкової складової просторової дисперсії параметра A ;

r_{ACEP} - середнє для угруповання значення коефіцієнту кореляції між параметрами A .

3.3 Приклади застосування методу сумісного аналізу даних у гідрологічних розрахунках

Завдання 1. Проаналізувати результати об'єднання за методом сумісного аналізу середньобагаторічних величин річного стоку у басейні р.Уссурі [24].

Вихідні дані. Результати застосування методу сумісного аналізу наведені у таблиці 3.1.

Таблиця 3.1 – Результати застосування методу сумісного аналізу даних до обґрунтування способу просторового узагальнення норм річного стоку річок в басейні р. Уссурі

Район	Середнє значення параметру $\bar{q}$, л/с·км ²	Дисперсія			ε_{CEP} , %
		повна	випадкова складова	географічна складова	
Північний	12,3	9,80	0,52	9,28	24,8
		100%	5,00%	95,0%	
Центральний	11,2	4,99	0,36	4,63	19,2
		100%	7,00%	93,0%	
Південний	8,94	2,55	0,587	1,97	15,8
		100%	23,0%	77,0%	
Водозбір р.Уссурі	10,3	6,39	0,48	5,91	34,0
		100%	8,00%	92,0%	

Виконання завдання. Попередні угруповання водозборів у райони (Північний, Центральний, Південний) виділені згідно з

результатами районування за синхронністю коливань стоку. Особливості коливань стоку у басейні р.Уссурі обумовлюються різним характером мусонних процесів на півночі та півдні водозбору. У північній частині переважає дія мусону помірних широт, у південній – субтропічних. Виділення Центрального району обумовлено особливостями гідрогеологічної структури басейну р.Уссурі. Для усіх трьох виділених районів та водозбору у цілому географічна складова просторової дисперсії значно перевищує випадкову, а відносна похибка осередненого параметру перевищує допустиму (10%), що свідчить про необхідність картування норм річного стоку, або пошуку рівнянь парної або множинної регресії, які б описували географічні закономірності розподілу норм річного стоку у межах водозбору р.Уссурі.

Завдання 2. Проаналізувати результати об'єднання за методом сумісного аналізу статистичних параметрів (коефіцієнту варіації, коефіцієнту автокореляції $r(1)$, відношення C_S/C_V) в басейні р. Дністер.

Вихідні дані. Результати об'єднання наведені у таблиці 3.2, 3.3, 3.4.

Виконання завдання. У межах басейну р.Дністер при обґрунтуванні способу узагальнення статистичних параметрів $C_V, r(1), C_S/C_V$, які визначаються з малим ступенем достовірності попередньо були виділені 4 райони за особливостями фізико-географічних умов. До першого входить гірський Дністер (правобережні притоки), до другого – передгірський Дністер (лівобережні притоки), до третього – річки Північної та Центральної Молдови, які характеризуються підвищеним підземним живленням річок за рахунок розвантаження карстових вод. Четвертий район утворюють річки Причорноморської низовини з посушливим кліматом та незначною часткою підземного живлення (5% від річного стоку). Просторова зміна коефіцієнтів варіації обумовлена характером зміни загальної зволоженості території: значення коефіцієнтів варіації збільшуються в міру переходу з зон надлишкового та достатнього зволоження (райони 1 і 2) в зони недостатнього зволоження (райони 3 та 4). Але на просторовий розподіл коефіцієнтів варіації впливає такий фактор підстильної поверхні, як гідрогеологічна структура водозбору й пов'язане з нею

підземне живлення. По перше, у межах водозбору р.Дністер існують карстові зони, які впливають на багаторічну мінливість стоку. По – друге, коефіцієнти варіації річок з площею меншою другої критичної можуть суттєво відрізнятись від коефіцієнтів варіації великих річок [4]. У зв'язку з цим, для об'єднання за методом сумісного аналізу бажано використовувати дані по водозборах з площею більшою 1000км^2 (друга критична площа для більшості розглядуваних річок). Певний інтерес являє собою район 1 (Карпати та Передкарпаття), де географічна й випадкова складові близькі одна до одної. У таких випадках при практичному застосуванні для водозборів із короткими рядами спостережень рекомендується використовувати не осереднені, а уточнені за формулою (3.9) значення коефіцієнтів варіації. Як впливає з табл.3.2, коефіцієнти варіації верхньої частини водозбору р.Дністер змінюються незначно в зоні надлишкового зволоження й достатнього зволоження, унаслідок чого райони 1 і 2 можливо об'єднати в один. Для району 3 (Середній Дністер) географічна складова стає від'ємною, у зв'язку з чим приймається рівною 0, у той час як випадкова складова дорівнює 100%. Для нижньої частини водозбору р.Дністер через близькість значень випадкової та географічної складових також бажано використовувати уточнені оцінки параметру C_v . Відносна похибка визначення осередненого значення параметру для усіх районів менша за допустиму (15%).

У гірській частині р.Дністер (зона надлишкового зволоження) та Причорномор'ї (зона недостатнього зволоження), де внесок підземного живлення у формування річного стоку слабо виражений, коефіцієнти автокореляції річного стоку близькі до нуля.

Таблиця 3.2 – Результати застосування методу сумісного аналізу даних до обґрунтування способу просторового узагальнення коефіцієнтів варіації річного стоку річок в басейні р.Дністер

Район	Середнє значення параметр u_{Cv}	Дисперсія			$\varepsilon_{Cv}, \%$	Внесок підземного живлення, %
		повна	випадкова складова	географічна складова		
Карпати та Прикарпаття	0,32	0,00283 100%	0,00164 58%	0,00119 42%	11,1	30
Лівобережний Дністер до впадіння р.Марківка	0,31	0,00215 100%	0,00206 96%	0,00009 4%	4,70	54
Середній Дністер (Північна та Центральна Молдова)	0,56	0,00512 100%	0,0116 100%	-0,00649 0%	4,39	25
Нижній Дністер (Причорномор'я)	0,75	0,0470 100%	0,0239 51%	0,0231 49%	12,4	5

Просторовий розподіл коефіцієнтів автокореляції визначається в більшій мірі не географічною зональністю, а внеском підземного живлення у формування стоку (табл. 3.3).

Таблиця 3.3 – Результати застосування методу сумісного аналізу даних до обґрунтування способу просторового узагальнення коефіцієнтів автокореляції річного стоку річок в басейні р. Дністер

Район	Середнє значення коефіцієнта автокореляції $r(1)$	Дисперсія			Середнє квадратичне відхилення $\sigma_{r(1)}$
		повна	випадкова складова	географічна складова	
Карпати та Прикарпаття	0,149	0,0140	0,0381 100%	-0,0241 0%	0,035
Лівобережний Дністер до р.Марківка	0,479	0,0273	0,0194 71%	0,00799 29%	0,096
Середній р.Дністер (Північна та Центральна Молдова)	0,499	0,0149	0,00234 100%	-0,00878 0%	0,036
Нижній р.Дністер (Причорномор'я)	0,001	0,0680	0,111 100%	-0,00431 0%	0,096

У межах Волино-Подільського артезіанського басейну (район 2), а також Північної та Центральної Молдови (район 3), де за рахунок близького залягання водоносних горизонтів й наявності карстових вод підземне живлення істотне, відзначаються високі

значення коефіцієнтів автокореляції, які наближаються до 0,5. Для оцінки якості об'єднання коефіцієнтів автокореляції $r(1)$ та відношення C_s / C_v використовувалася нерівність (3.7).

Слід зазначити, що за часів колишнього СРСР для території України було рекомендовано використовувати коефіцієнт автокореляції, який дорівнює 0,22. Але по результатах, наведених у табл.3.3, можна зробити висновок, що діапазон значень $r(1)$ набагато ширший. Так, у межах України на основі методу сумісного аналізу для коефіцієнта автокореляції річного стоку виділено 7 районів, а для коефіцієнта асиметрії – 10 [21].

Що стосується відношення C_s / C_v , то в зв'язку з великими похибками розрахунку коефіцієнта асиметрії за даними, перевірка гіпотези про можливість районування виконувалося в межах районів, виділених для коефіцієнта варіації (табл.3.2). У гірській зоні басейну р.Дністер відношення C_s / C_v можна прийняти рівним 2, на лівобережжі р.Дністер – 3, а в середньому плині 1,5. У нижній частині басейну р.Дністер відношення C_s / C_v рекомендується приймати рівним 1,7 (табл.3.4).

Завдання 3. Виконати уточнення коефіцієнтів варіації річного стоку річок водозбору р.Сіверський Донець за результатами об'єднання даних.

Вихідні дані. Результати об'єднання коефіцієнту варіації річного стоку, виконані за методом сумісного аналізу даних, дозволили виділити два райони. Для Північного району випадкова та географічна складові повної дисперсії параметру наближаються до 50% і становлять відповідно, 53 та 47 вісотків, при цьому $\sigma_{CP} = 0,047$; $C_{V_{CEP}} = 0,37$.

Виконання завдання. У випадку, коли географічна та випадкова складові майже дорівнюють одна одній, можна виконати уточнення параметру C_v з використанням формули (3.9). Результати розрахунків наведені у табл.3.5.

Таблиця 3.4 – Результати застосування методу сумісного аналізу даних до обґрунтування способу просторового узагальнення відношення C_s/C_v річного стоку річок в басейні р.Дністер

Район	Середнє значення C_s / C_v	Дисперсія			σ_{C_s / C_v}
		повна	випадкова складова	географічна складова	
Карпати та Прикарпаття	2,0	2,01 100%	1,86 93%	0,146 7%	0,350
Лівобережні притоки р. Дністер до р.Марківка	3,0	0,400 100%	1,72 100%	-1,32 0%	0,456
Середній Дністер (Північна та Центральна Молдова)	1,5	0,398 100%	0,873 100%	-0,474 0%	0,616
Нижній Дністер (Причорноморська низовина)	1,7	2,29 100%	1,41 62%	0,880 38%	1,01

Таблиця 3.5 - Уточнені значення коефіцієнту варіації річного стоку
Північного підрайону р.Сіверський Донець з використанням
результатів об'єднання за методом сумісного аналізу

Річка-пост	n	C_V	σ_{C_V}	$\sigma_{C_V}^2$	C_V'	σ_{C_V}'	ε_{C_V}' , %
С.Донець- с.Кисільово	15	0.40	0.0712	0.005	0.38	0.0393	10.3
Болховець- м.Белгород	27	0.38	0.0540	0.003	0.37	0.0350	9.46
Нежеголь- с.Велико - троїцьке	17	0.43	0.0770	0.0039	0.38	0.0400	10.5
Короча- с.Короча	15	0.49	0.0930	0.0087	0.39	0.0410	10.4
Бабка – с.П'ятницьке	21	0.39	0.0627	0.0040	0.37	0.0460	11.5
Уди – м.Золочів	12	0.46	0.0892	0.0080	0.40	0.0460	11.5
Рогозянка- с. Велика Рогозянка	13	0.46	0.0869	0.0080	0.39	0.0113	10.6
Оскіл – с. Нінівка	24	0.30	0.0444	0.00197	0.33	0.0322	9.68
Халань- с.Рос.Халань	12	0.49	0.1030	0.0106	0.39	0.0427	10.9
Валуй- с. Валуйки	25	0.56	0.0862	0.0744	0.41	0.0472	9.96

Контрольні запитання

1. Записати загальну просторову дисперсію параметра А через складові.
2. Визначення випадкової складової просторової дисперсії параметра А.

3. Визначення географічної складової просторової дисперсії параметра A .
4. При яких співвідношеннях між географічною та випадковою складовою приймається рішення про можливість картування параметра A .
5. При яких співвідношеннях між географічною та випадковою складовою приймається рішення про можливість районування параметра A .
6. При яких співвідношеннях між географічною та випадковою складовою приймається рішення про уточнення значень параметра A по об'єктах, які увійшли до об'єднуваної виборки.

4. МЕТОД ГОЛОВНИХ КОМПОНЕНТІВ

4.1 Теоретичні положення

Метод головних компонентів (природних ортогональних функцій або ПОФ) являє собою один з методів лінійного перетворення інформації, який полягає в лінійному ортогональному перетворенні полів вхідних величин у базисі власних векторів матриці кореляцій або коваріацій [2,15]. Інакше кажучи, на основі матриць коваріації або кореляції визначається система ортогональних, лінійно незалежних функцій, які прийнято називати власними векторами. Вони відповідають системі незалежних випадкових величин, іменованих власними значеннями або власними числами матриці коваріацій або кореляцій. Пошук власних векторів та власних значень досягається шляхом розв'язання матричних рівнянь виду

$$R_X U_i - \lambda_i U_i = 0, \quad (4.1)$$

$$K_X W_i - \lambda_i W_i = 0, \quad (4.2)$$

де R_X - матриця коефіцієнтів кореляції розміром $m \times m$ (m відповідає числу розглянутих об'єктів);

K_X - матриця коваріацій розміром $m \times m$;

U_i - власний вектор матриці кореляцій;

W_i - власний вектор матриці коваріацій;

λ_i - відповідне власному вектору власне значення.

Матричні рівняння виду (4.1) та (4.2) мають назву матричних рівнянь повної проблеми власних значень.

Власний вектор – це вектор U_i або W_i , який при даному лінійному перетворенні (4.1) або (4.2) не змінює свого напрямку, а тільки помножується на скаляр λ_i .

Власне значення матриці – це число λ_i , для якого існує ненульовий вектор U_i , який має властивість $R_X U_i = \lambda_i U_i$ або $K_X U_i = \lambda_i W_i$.

Власні вектори матриць кореляцій та коваріацій є **ортонормованими** й мають таку властивість [38]

$$U_i' U_j = \delta_{ij}, \quad (4.3)$$

де δ_{ij} - символ *Кронекера*, при цьому

$$\delta_{ij} = \begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases} \quad (4.4)$$

Сукупність ортонормованих власних векторів утворює ортогональну матрицю, яка має такі властивості

$$W'W = E, \quad (4.5)$$

де W' - транспонована матриця W розміром $m \times m$;

E - одинична матриця.

Матриця R_X має m коренів або m власних чисел λ , які є дійсними, позитивними й простими. Для знаходження m власних векторів, що відповідають m власним числам, необхідне розв'язання m систем лінійних рівнянь. Процедура розрахунку здійснюється, як правило, за допомогою ітераційних методів, серед яких найпоширенішим є метод Якобі [38].

Сукупність власних векторів утворює базис, у якому проводиться розкладання полів вихідних даних

$$U' \cdot \varphi_i = Z_i; \quad (4.6)$$

$$W' \cdot \Delta X_i = Z_i, \quad (4.7)$$

де U' - транспонована матриця U розміром $m \times m$;

W' - транспонована матриця W розміром $m \times m$;
 φ_i - i -ий випадковий вектор (поле) центрованих та нормованих вихідних даних, що підлягає розкладанню;
 ΔX_i - i -й випадковий вектор (поле) центрованих вихідних даних, що підлягає розкладанню;
 Z_i - вектор головних компонент, який є результатом лінійного перетворення поля вихідних даних відповідним власним вектором.

Рівності (4.6) та (4.7) означають, що вхідне поле розкладене на m незалежних компонент в базисі власних векторів матриць кореляцій та коваріацій, відповідно.

Складові вектора Z_i для p - тої компоненти розкладання визначаються таким чином:

$$z_{ip} = \sum_{k=1}^m U_{ki} \varphi_{kp} ; p = \overline{1, n} ; i = \overline{1, m} ; \quad (4.8)$$

або при використанні центрованих величин $\Delta X_i = x_i - \bar{x}$

$$z_{ip} = \sum_{k=1}^m W_{ki} \Delta X_{kp} ; p = \overline{1, n} ; i = \overline{1, m} \quad (4.9)$$

де z_{ip} - складові p - тої компоненти розкладання;

U_{pk} - вагові коефіцієнти, що відображують внесок k - того об'єкта в кожну p - ту компоненту (або внесок p - тої компоненти в k - тий об'єкт) при використанні кореляційної матриці;

W_{pk} - вагові коефіцієнти, що відображують внесок k - того об'єкта в кожну p - ту компоненту (або внесок p - тої компоненти в k - тий об'єкт) при використанні коваріаційної матриці;

ΔX_{ik} - i -й випадковий вектор (поле) центрованих вихідних даних, що підлягає розкладанню;

φ_{ik} - i -й випадковий вектор (поле) центрованих та нормованих вихідних даних, що підлягає розкладанню.

Центровані та нормовані дані розраховуються наступним чином

$$\varphi_{ik} = \frac{x_{ik} - \bar{x}_k}{\sigma_k}, \quad (4.10)$$

де x_{ik} - кожне значення i - того ряду спостережень при $i = \overline{1, n}$;
 $\bar{x}_k$ - середнє арифметичне значення k - того ряду спостережень;
 σ_k - середнє квадратичне відхилення k - того ряду спостережень.

Значення U_{pk} або W_{pk} змінюються в просторі при переході від об'єкта до об'єкта, але не залежать від часу. **Система функцій U_{pk} часто представляється як функція координат (x_k, y_k) для k - того об'єкта й має назву “базисної функції”.**

Оскільки базис власних векторів є ортогональним, то компоненти z_{ip} вектора Z_i є лінійно незалежними, а в статистичному значенні некорельованими.

Отримані компоненти мають наступну властивість: кожне p -е власне значення λ_p матриці кореляцій є дисперсією p -ї головної компоненти σ_{Zp}^2

$$\lambda_p = \sigma_{Zp}^2, \quad (4.11)$$

Тоді сума дисперсій m компонент дорівнює сумі власних чисел матриці й, отже, дорівнює сліду матриці кореляцій

$$\sum_{p=1}^m \sigma_{Zp}^2 = \sum_{p=1}^m \lambda_p = t_R R_x = m, \quad (4.12)$$

де t_R - слід матриці кореляцій (**слідом матриці кореляцій називається сума елементів, розташованих на головній діагоналі матриці**).

Якщо розкладання за ПОФ виконувати в базисі власних векторів матриці коваріацій, то

$$\sum_{p=1}^m \sigma_{Zp}^2 = \sum_{p=1}^m \lambda_p = t_R K_x = \sum_{p=1}^m \sigma_{Xp}^2. \quad (4.13)$$

Сума власних значень матриці кореляцій є, як це було показано раніше у (4.11), сумою дисперсій головних компонентів. Ця сума, отримана для матриці коваріацій, дорівнює сумарній дисперсії поля. Остання розподіляється таким чином, що найбільша її частина являє собою дисперсію першої головної компоненти, яка у свою чергу, згідно з (4.11), є першим власним значенням.

Сума дисперсій головних компонентів матриці коваріацій дорівнює сумі дисперсій вихідних рядів, тобто

$$\boxed{\sum_{p=1}^m \sigma_{Zp}^2 = \sum_{j=1}^m \sigma_{Xj}^2}. \quad (4.14)$$

Таке подання дозволяє більш наочно зрозуміти суть методу головних компонентів, тому що ці некорельовані лінійні комбінації вихідних змінних в (4.8) та (4.9) містять у собі всю дисперсію, укладену в m змінних вхідного масиву даних. Декілька перших власних чисел ($\lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_m$) вичерпують основну частину сумарної дисперсії поля, тому при аналізі результатів розкладання особлива увага приділяється першим власним значенням й відповідним їм компонентам. Оскільки великомасштабні процеси характеризуються великою дисперсією, те справедливе припущення, що саме вони відображені у перших компонентах.

Якщо розташувати власні числа матриці в убутному порядку, то перше власне число буде являти собою величину дисперсії, яка відповідає першій компоненті, друге власне число - величину дисперсії, що відповідає другій компоненті й т.д. Через те, що при використанні кореляційної матриці сума власних чисел дорівнює числу розглянутих змінних m , те розділивши кожне власне число на

m або $\sum_{j=1}^m \lambda_j$, можна одержати частку від сумарної дисперсії, що

описується кожною з розглянутих компонентів.

Частку істотної інформації із всієї сумарної інформації про поле оцінюється за допомогою співвідношення

$$S = \frac{\sum_{k=1}^p \lambda_k}{\sum_{k=1}^m \lambda_s}, \quad (4.15)$$

де чисельник дорівнює сумі дисперсій, що приходить на p перших головних компонент, а знаменник дорівнює сумарній дисперсії поля. Це дозволяє говорити, що **відношення (4.15) показує внесок p перших компонент у загальну дисперсію вихідних даних**. Задаючись значенням S (наприклад, $S = 0,70-0,80$), можна встановити число перших компонент, які варто урахувати, щоб скоротити обсяг вхідної інформації й зберегти при цьому її основний зміст. **Скорочення вхідної інформації при збереженні її смислу має назву “стиску інформації”**. При використанні методу головних компонент вхідна інформація не тільки заміщається малим числом статистичних функцій, але й зберігає в цих функціях їх фізичний зміст. Отримані в результаті розкладання функції є відображенням реальних фізичних процесів, які обумовлюють просторово-часовий розподіл досліджуваних гідрометеорологічних величин.

4.2 Просторове узагальнення вагових коефіцієнтів головних компонент

Установити фізичний зміст компонент можливо на основі пошуку зв'язків статистичної структури гідрологічних полів у вигляді складових базисних функцій із стокоформувальними чинниками [15].

Для Південно-Західної України Н.С. Лободою [21,41] установлені такі зв'язки між ваговими коефіцієнтами перших компонент розкладання полів річного стоку та стокоформувальними чинниками

$$W_{1i} = -0,000884\bar{E}_{mi} + 0,000361\bar{X}_i + 0,660; \quad R = 0,884 \quad (4.16)$$

$$W_{2i} = -0,00150\bar{E}_{mi} - 0,00290\bar{X}_i + 3,014; \quad R = 0,734 \quad (4.17)$$

$$W_{3i} = 0,00180\bar{E}_{mi} + 1,450; \quad r = -0,788, \quad (4.18)$$

де W_{pi} - вагові коефіцієнти, що відображують внесок p - тої компоненти в формування стоку i - того об'єкта) при використанні коваріаційної матриці;

$\bar{E}_{mi}$ - річна норма максимально можливого випаровування для i - того об'єкта;

$\bar{X}_i$ - річна норма опадів для i - того об'єкта;

r - коефіцієнт парної кореляції;

R - коефіцієнт множинної кореляції;.

Рівняння (4.16) – (4.17) отримані на основі методу множинної регресії, опис якого наданий у розділі 5. Зміст отриманих рівнянь можна інтерпретувати наступним чином: перші три компоненти розкладання полів річного стоку Південно-Західної України за ПОФ є математичним відображенням кліматичних процесів, які відбуваються на розглядуваній території.

Фізичне навантаження вагових коефіцієнтів залежить від просторових масштабів поєднаних рядів та характеру розглядуваної території. Наприклад, пошук зв'язків між ваговими навантаженнями на перші компоненти розкладання та показниками кліматичних чинників й чинників підстильної поверхні для Українських Карпат показав, що найбільш оптимальним предиктором при розрахунках річного стоку з водозборів Українських Карпат є висота місцевості. Саме висота місцевості має найбільш тісний кореляційний зв'язок із значеннями вагових коефіцієнтів $W_{1i}(x, y)$ першої, найбільш інформаційної, складової. Висота місцевості у даному випадку є інтегральним показником впливу кліматичних факторів, насамперед, опадів та випаровування, на формування річного стоку гірських районів. Вагові коефіцієнти $W_{2i}(x, y)$ другої компоненти розкладання мають тісний зв'язок з логарифмом площі водозборів $\lg(F)$, а третьої компоненти - із

залісеністю (f_L). Отримані результати свідчать про те, що фізичні процеси, представлені другою та третьою компонентами, обумовлюють мезомасштабні зміни характеристик стоку у просторі, які можуть бути пов'язані із азональними $\lg(F)$ та інтразональними f_L чинниками.

4.3 Застосування методу головних компонент до аналізу синхронності коливань стоку

Дослідження, які були виконані Н.П. Смирновим та В.Л.Скляренко (1986) по аналізу просторової структури багаторічних коливань стоку та районування територій за характером цих коливань, дозволили установити зв'язок між першими базисними функціями розкладання полів стоку за ПОФ та відповідними базисними функціями кліматичних характеристик (температури повітря, тиск, опади). Це дозволило зробити висновки про те, що структура полів кліматичних характеристик та полів стоку визначається одними й тими ж фізичними процесами, якими є атмосферні процеси. При аналізі річного стоку першій компоненті розкладання гідрометеорологічних полів за ПОФ відповідають глобальні атмосферні процеси, другій та третій – процеси синоптичного масштабу. Як правило, перші компоненти розкладання полів гідрометеорологічних величин є математичним представленням впливу на їх формування процесів атмосферної циркуляції. Чим більший порядок компоненти, тим менший масштаб процесу, який вона відображує. Так, водогосподарські перетворення у межах водозборів є процесами мезо- та мікромасштабу, отже їх вплив на формування полів стоку проявляється на рівні четвертої або п'ятої компонент (Лобода,2005).

Як правило перший, найбільш значущий атмосферний процес, якій відповідає першій компоненті розкладання за ПОФ, приводить до одночасного зменшення або збільшення стоку на всіх річках, які розглядаються. Вагове навантаження на першу компоненту зберігає знак, що може розглядатися як показник синфазності коливань стоку.

Зміна знаку базисної функції показує, що наслідки впливу другого за значимістю атмосферного процесу на формування стоку по обидві сторони від нульової ізолінії коливаються у протифазі. Щоб зрозуміти, як атмосферні процеси впливають на клімат, розглянемо такий атмосферний процес як Північне Атлантичне Коливання (ПАК). ПАК одночасно викликає сухість клімату південно-східної Європи та підвищену вологість у Скандинавії у роки, коли індекс (кількісна характеристика) цього коливання має позитивне значення, й, навпаки, коли індекс ПАК приймає від'ємні значення, у південно-східній Європі спостерігається підвищена зволоженість, а у Скандинавії – підвищена сухість. Оскільки Північне Атлантичне Коливання є проявом процесу атмосферної циркуляції і його коливання викликають зміни клімату, то воно може бути причиною відповідних змін у коливаннях стоку.

4.4. Вирішення задачі фільтрації гідрологічної інформації

Гідрометеорологічні поля формуються під дією атмосферних процесів різного масштабу: макропроцесів, пов'язаних із планетарними рухами мас повітря; процесів, обумовлених атмосферною циркуляцією (синоптичними процесами); процесів мезомасштабу, а також дрібномасштабних флуктуацій, пов'язаних з погрішностями вимірів та недоліками первинної обробки результатів спостережень. При необхідності виключити вплив малоінформативних фізичних процесів на формування гідрометеорологічних полів використовують математичний апарат, вже викладений в підрозділі 4.1, роблячи зворотний перехід від компонентів до значень досліджуваної величини.

Будь-який елемент матриці вихідних центрованих та нормованих значень φ_{ij} (на i -тім розглянутім об'єкті в j -тий момент часу) може бути розрахований, якщо проблему власних векторів вирішено

$$\varphi_{ij} = \sum_{k=1}^m U_{ki} z_{kj}, \text{ при } i = \overline{1, m}; j = \overline{1, n}, \quad (4.19)$$

де φ_{ij} - складові j -того випадкового вектора (поля) центрованих та нормованих вихідних даних, яке підлягало розкладанню за природними ортогональними функціями;

U_{ki} - вагові коефіцієнти, які відображають внесок i -того об'єкта в кожну k -ту компоненту або складові власних векторів матриці кореляцій;

z_{kj} - складові k -тої компоненти розкладання;

m - число об'єктів;

n - довжина вихідних рядів.

Складові вектора-рядка матриці компонентів Z
 $[z_{k1} \ z_{k2} \ \cdot \ \cdot \ z_{kp} \ \cdot \ \cdot \ \cdot \ z_{kn}]$ **можуть бути представлені як функції часу і мають назву амплітудних функцій. Амплітудні функції є загальними для всіх об'єктів, вони не залежать від координат і є функціями часу**

$$z_{kj} = f(t) = z_k(t). \quad (4.20)$$

Щоб мати краще уявлення про амплітудну функцію, розглянемо хронологічний графік першої амплітудної функції розкладання полів річного стоку річок України, а також близько розташованих річок Центральної та Південної Європи (рис.4.1). Цей хронологічний графік відображає характер коливання стоку річок, який формується під впливом найбільш великомасштабного фізичного процесу.

Для інтерпретації фізичного змісту перших компонент розкладання виконують пошук зв'язків між першими амплітудними функціями та індексами процесів атмосферної циркуляції. Наприклад, встановлено, що перша амплітудна функція полів річного стоку Центральної та Південної Європи зв'язана з індексом атмосферної циркуляції, який характеризує Арктичне коливання (рис.4.2). Цей індекс визначається за полями тиску Північної Півкулі (20° - 90° півн.ш.) у зимовий сезон (грудень-березень).

Розв'язання задачі фільтрації вихідної інформації полягає у розрахунках досліджуваної величини на основі тільки перших компонент розкладання поля за ПОФ. Як вже відмічалось, перші

компоненти містять у собі основну частку всієї вихідної інформації. Отримане зворотним розрахунком поле вихідних даних є згладженим, оскільки відображає головні особливості просторово-часового розподілу досліджуваної величини.

При розгляді тільки перших компонент розкладання, в яких утримується основна частина інформації вихідних полів, вираз (4.19) перетворюється до виду

$$\tilde{\varphi}_{ij} = \sum_{k=1}^p U_{ki} z_{kj}, \text{ при } i = \overline{1, m}; j = \overline{1, n} \quad (4.21)$$

де $\tilde{\varphi}_{ij}$ - відфільтровані або згладжені значення центрованих та нормованих величин стоку;

p - число перших компонент.

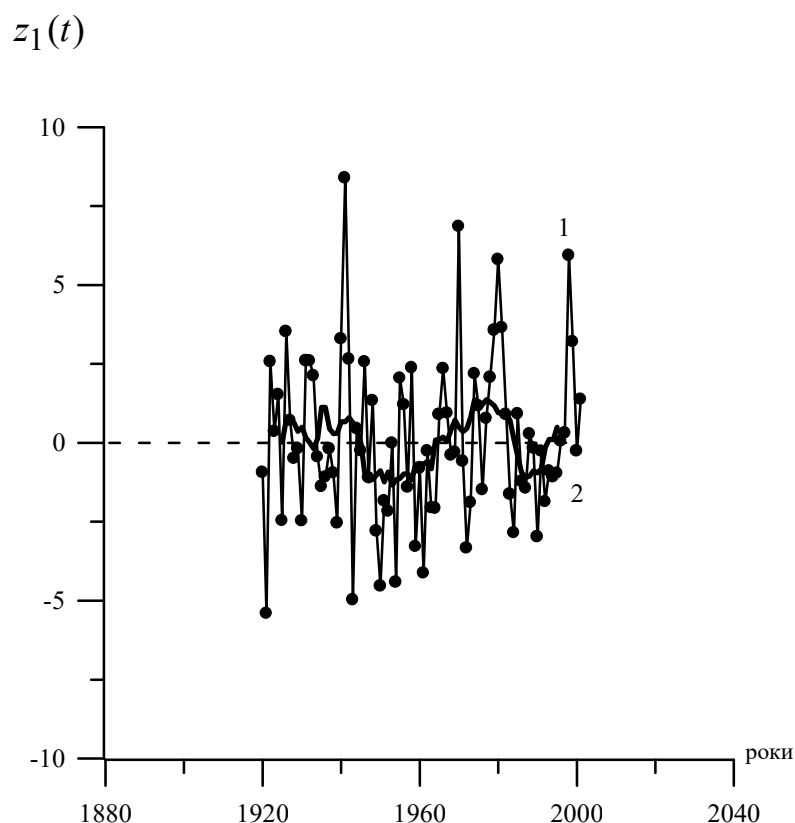


Рис. 4.1 Перша амплітудна функція коливань полів річного стоку річок України та Центральної й Південної Європи

1- хронологічний хід;

2- згладжений за 11-річками ряд

Індекс Арктичного коливання

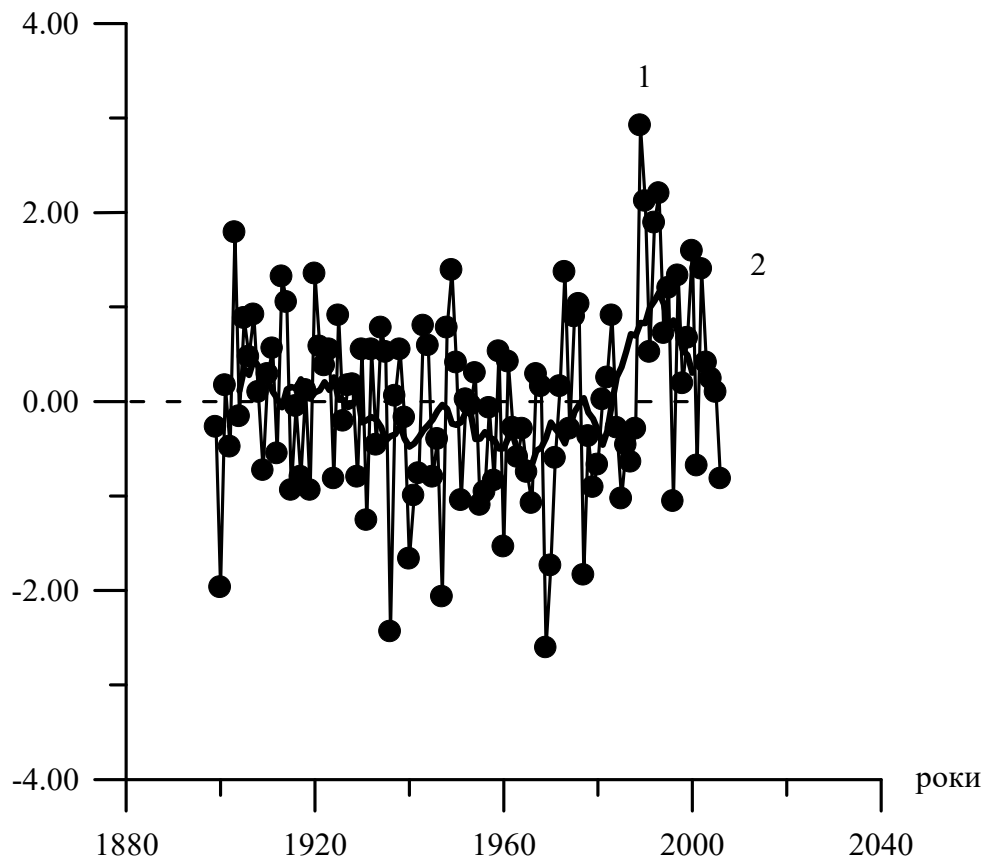


Рис.3.2. – Індекс Арктичного коливання (грудень-березень)

1-хронологічний хід;
2- згладжений за 11-річками ряд

Матрична форма запису (4.21) має вигляд

$$\tilde{\varphi}_i = U \cdot \tilde{Z}_i, \quad (4.22)$$

де $\tilde{\varphi}_i$ - відфільтрований вектор або поле центрованих та нормованих значень вихідної величини;

U - матриця власних векторів;

$\tilde{Z}_i$ - відфільтрований вектор компонент.

Перехід до величини $\tilde{x}_{ij}$ здійснюється за виразом

$$\tilde{x}_{ij} = \bar{x}_i + \sigma_i \sum_{k=1}^p U_{ki} \tilde{z}_{kj} ; \text{ при } i = \overline{1, m} ; j = \overline{1, n} \quad (4.23)$$

або

$$\tilde{x}_{ij} = \bar{x}_i + \sum_{k=1}^p w_{ki} z_{kj} ; \text{ при } i = \overline{1, m} ; j = \overline{1, n} \quad (4.24)$$

де $\bar{x}_i$ - середнє арифметичне значення вихідного ряду;

σ_i - середнє квадратичне відхилення вихідного ряду;

φ_{ki} - вагові коефіцієнти k -тої компоненти розкладання матриці коваріацій за ПОФ.

w_{ki} - вагові коефіцієнти k -тої компоненти розкладання матриці коваріацій за ПОФ.

При розв'язуванні задачі фільтрації вихідну інформацію треба звільнити від дрібномасштабних збурень. Цієї мети можна досягти, якщо всі складові k -тої компоненти від $p+1$ до m прирівняти до нуля, тоді k -а компонента представляється у вигляді

$$\tilde{Z}_j = \begin{pmatrix} Z_{1j} \\ Z_{2j} \\ Z_{3j} \\ \cdot \\ \cdot \\ Z_{pj} \\ 0 \\ \cdot \\ \cdot \\ 0 \end{pmatrix} \quad (4.25)$$

Щоб отримати матрицю “фільтрованих” вихідних даних, представлених у центрованому виді, необхідно виконати скалярне множення матриць за виразом наступного виду

$$W\tilde{Z} = \Delta\tilde{X}, \quad (4.26)$$

або

$$U\tilde{Z} = \tilde{\varphi}, \quad (4.27)$$

де $\tilde{Z}$ - матриця, у якій виконано фільтрацію компонент згідно із (4.25);

$\Delta\tilde{X}$ - матриця згладжених даних, представлених у вигляді центрованих величин;

$\tilde{\varphi}$ - матриця згладжених даних, представлених у вигляді центрованих та нормованих величин.

Перехід до величини згладженої величини $\tilde{x}_{ij}$ виконується за виразом

$$\tilde{x}_{ij} = \bar{x}_i + \Delta\tilde{X}_{ij}, \quad (4.28)$$

де $\bar{x}$ - середня арифметична величина ряду i .

Прикладне застосування фільтрації рядів дозволяє представити процес стоку у вигляді згладженого хронологічного ряду, який відображає властивості тільки тих компонент, а, отже, й відповідних їм фізичних процесів різних масштабів, які були застосовані при фільтрації. При цьому використовуються амплітудні функції (компоненти) і вагові коефіцієнти (значення власних векторів матриці кореляцій або коваріацій) перших складових розкладання, а також середні арифметичні значення.

Рівняння (4.28) було використане автором для розрахунків рядів природного (непорушеного водогосподарською діяльністю) річного стоку Причорноморської низовини (Лобода Н.С., 2005). Норма річного кліматичного (розрахованого за рівнянням водно-теплогового балансу) стоку розглядалася як середня багаторічна

величина природного стоку $\bar{x}_i$, а вагові коефіцієнти перших трьох компонент розкладання визначалися за показниками кліматичних факторів (4.16-4.18) – максимально можливого випаровування та опадів. Запропонований підхід дозволяє відновлювати хронологічні ряди природного стоку невивчених річок за рівнянням (4.28). При цьому використовуються амплітудні функції z_{kj} перших компонент, які є загальними для усіх водозборів розглядуваного регіону та вагові коефіцієнти розкладання W_{ki} , які визначаються за вказаними регресійними рівняннями в залежності від кліматичних факторів й не підлягають впливу водогосподарських перетворень (рис.4.3).

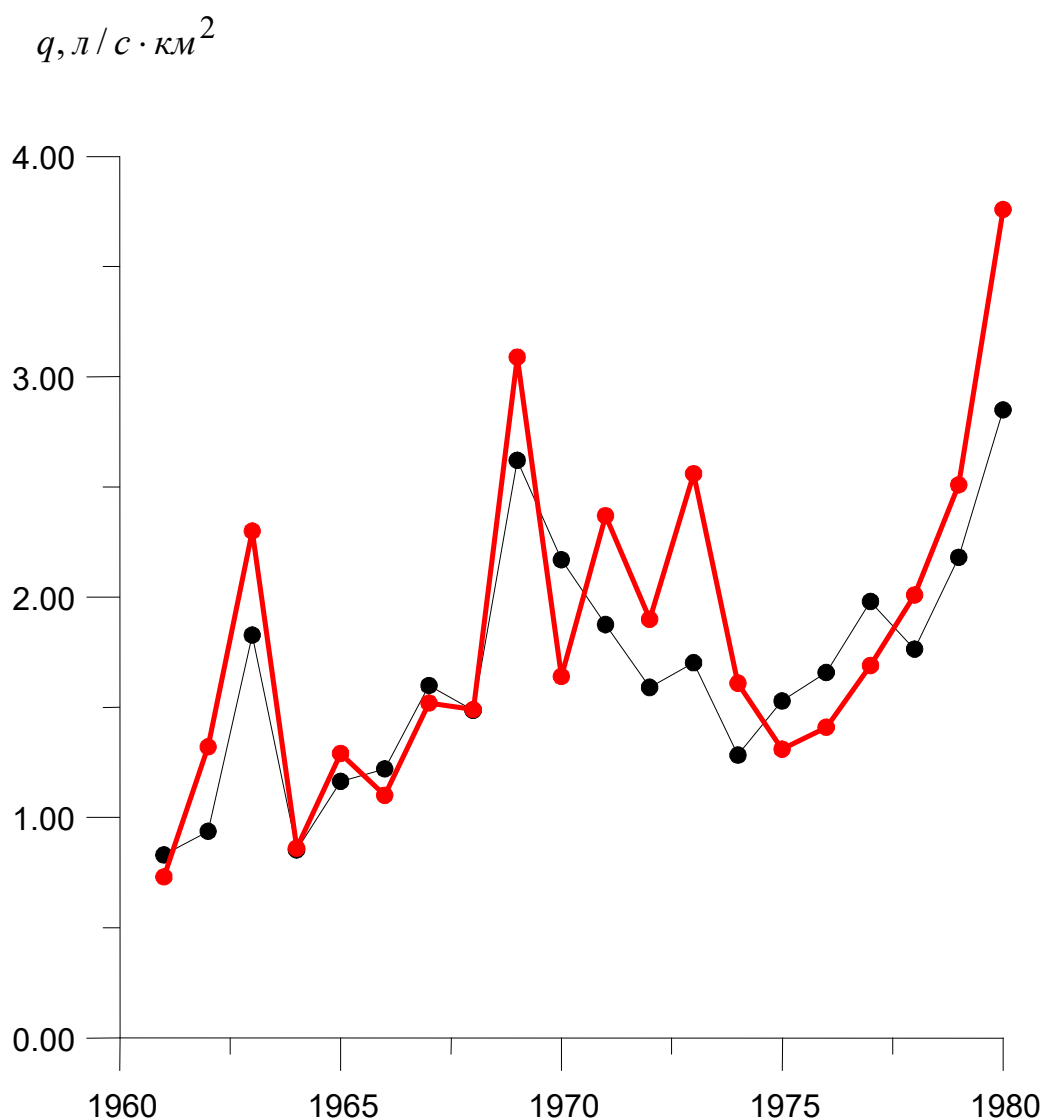


Рисунок 4.3. Порівняння спостереженого (нижня крива) та відновленого (верхня крива) за трьома компонентами ряду річного стоку р.Когільник

4.5 Використання методу головних компонент у гідрологічних прогнозах

У гідрологічних прогнозах метод головних компонент знайшов своє застосування при прогнозах дат льодових явищ, прогнозах водності річок та рівневого режиму [1,5].

Розглянемо приклад прогнозу середньомісячних витрат води літньо-осіннього періоду для півночі Європейської території колишнього СРСР (Н.Ф.Бефані, С.В.Борщ, 1982). Величина стоку за цей період формується за рахунок підземного припливу та надходження дощових вод до русла річки, що може бути представленим рівнянням водного балансу за інтервал часу Δt

$$\sum_{t_0}^{t_0+\Delta t} Q \Delta t = W_{t_0} + \sum_{t_0}^{t_0+\Delta t} Q_{\text{підз}} \Delta t + \sum_{t_0}^{t_0+\Delta t} Q_{\text{дощ}} \Delta t; \quad \Delta t > \tau_{\text{макс}}, \quad (4.29)$$

де $\tau_{\text{макс}}$ - максимальний час руслового добігання;

W_{t_0} - величина руслових запасів на момент часу t_0 ;

$\sum_{t_0}^{t_0+\Delta t} Q_{\text{підз}} \Delta t$ - об'єм підземного припливу за інтервал часу Δt ;

$\sum_{t_0}^{t_0+\Delta t} Q_{\text{дощ}} \Delta t$ - об'єм дощового припливу за інтервал часу Δt .

Непрямою характеристикою загального зволоження басейну, а також запасів води у русловій мережі та горизонтах підземних вод може бути середньомісячна витрата за місяць, який передує даті випуску прогнозу. У зоні надлишкового зволоження значна роль у формуванні літньо-осіннього стоку належить дощовій складовій, яку оцінюють через стокоформувальні опади, які також підлягають прогнозу. При довгострокових прогнозах опадів як головний предиктор використовують дані полів приземного тиску. Отже, внесок дощової складової у формування полів середньомісячного стоку літньо-осіннього періоду може бути представленим через вагові коефіцієнти розкладання полів аномалій середньомісячного тиску.

При розробці прогностичної залежності у якості головних чинників були використані:

- поля аномалій середньомісячного приземного тиску по 84 пунктам Північної півкулі за період травень-березень;
- значення середньомісячних витрат води за місяць, який передує даті випуску прогнозу по 30-ти створах;
- суми опадів за місяць, який передує даті випуску прогнозу по 13 метеостанціям півночі Європейської території колишнього СРСР;
- суми опадів за декаду, яка передує даті випуску прогнозу.

Період розглянутих спостережень становив 25 років – з 1950 по 1975 рр. Прогнозна величина – витрати за серпень місяць. Усі перелічені поля даних були підлеглі розкладанню за ПОФ. Перші п'ять членів розкладання описують 77-96% дисперсії вихідних даних.

Прогнозуються перші п'ять членів амплітудної функції полів середньомісячних витрат за серпень місяць:
 $z_{1,Q}; z_{2,Q}; z_{3,Q}; z_{4,Q}; z_{5,Q}$.

Прогноз виконується за значеннями амплітудних функцій гідрометеорологічних полів, які сформовані із даних, що спостерігалися до дати випуску прогнозу. Наприклад, перше значення амплітудної функції стоку за серпень місяць залежить від $z_{1,Q_{дек}}$ - першого значення амплітудної функції, отриманого при розкладанні полів стоку за третю декаду липня; від $z_{1,X_{липень}}$ - першого значення амплітудної функції, отриманої при розкладанні полів опадів за липень й т.д. Коли $z_{1,Q}; z_{2,Q}; z_{3,Q}; z_{4,Q}; z_{5,Q}$ визначені, можна скористуватися матрицею-стовпцем (4.25). Перехід до ΔQ або до $\Delta \tilde{\varphi}_Q$ здійснюється шляхом використання матричних рівнянь (4.26) або (4.27). Матриці W та U відомі, вони складаються з так званих базисних функцій, які визначаються за даними спостережень і не залежать від часу, а змінюються у просторі. Для кожного створу вагові коефіцієнти W або U є відомими.

Критерій якості s/σ такої методики становить 0,40-0,75 для різних місяців літньо-осіннього періоду.

4.6 Приклади застосування методу головних компонент у гідрологічних розрахунках

Приклад 1

Завдання. Визначити характеристики статистичної структури полів річного стоку на основі метода головних компонент.

Вихідні дані: розглянуто 14 рядів річного стоку тривалістю 82 роки (з 1920 по 2001гг.): р. Дніпро-м. Речиця (Білорусь); р.Сож – м.Гомель (Білорусь); р. Десна –г. Чернігів; р.Прип'ять – г. Мозир (Білорусь); р.Дунай – м. Рені; р. Дністер –с. Заліщики; р.Південний Буг – с. Олександрівка; р. Сіверський Донець – м. Лисичанськ; р. Дунай –м.Братислава (Словаччина); р. Тиса – м. Польшар (Угорщина); р.Мюреш – м.Арад (Румунія); р. Сан – м. Радомишль (Польща); р. Західний Буг - м.Вишків (Польща);р.Дон – ст.Раздорська (Росія).

Отримані результати. В результаті розрахунків, виконаних за допомогою ПЕОМ, були отримані власні значення кореляційної матриці, власні вектори, амплітудні функції.

Аналіз результатів. Перші власні значення матриці кореляцій дорівнюють: $\lambda_1 = 6,72$; $\lambda_2 = 2,54$; $\lambda_3 = 1,51$; $\lambda_4 = 0,74$. Внесок першої компоненти у загальну дисперсію вихідних даних становить 48%, другої – 22%, третьої -11%. Загальний внесок перших трьох компонент становить 77%. Це означає, що перших трьох компонент, які описують реальні фізичні процеси, достатньо для аналізу статистичної структури полів річного стоку.

Вагові коефіцієнти першої компоненти мають однаковий знак (табл.4.1). Це свідчить про однаковий напрям впливу найбільш масштабного фізичного процесу на формування полів річного стоку розглядуваної території. Найбільше значення коефіцієнту U_{1i} знаходиться в басейні р. Дунай, тобто самий масштабний фізичний процес над розглядуваною територією найсуттєвіше впливає на формування стоку р.Дунай та її приток Тиса й Мюреш.

Вагові коефіцієнти U_{2i} другої компоненти змінюють свій знак. Це означає, що другий по значущості фізичний процес неоднаково впливає на формування річного стоку різних водозборів.

За знаком U_{2i} можна виділити дві групи водозборів (табл. 4.1), які утворюють два райони у просторі – Західний та Східний. Західний район включає до себе річки Західної та Південної Європи, а також такі річки України, як Дністер, Дунай у нижній течії, Тиса. Східний район утворюють рр.Дніпро з притоками, Південний Буг, Сіверський Донець.

Таблиця 4.1. - Вагові коефіцієнти перших трьох компонент полів річного стоку
(використана матриця кореляцій річного стоку, період спостережень - з 1920 по 2001 роки)

Річка-пост	Вагові коефіцієнти		
	$U1$	$U2$	$U3$
р.Дніпро-м. Речиця	0,2213	-0,4020	-0,3102
р.Сож –м.Гомель	0,2266	-0,4172	-0,2865
р.Десна-м.Чернігів	0,2324	-0,3989	-0,1443
р.Прип'ять-м.Мозир	0,2958	-0,0934	-0,2648
р.Дунай – м. Рені	0,3102	0,1792	0,1687
р.Дністер-м.Заліщики	0,2947	0,2432	-0,0391
р.Південний Буг-п.Олександрівка	0,2780	-0,0352	0,1837
р.Сів.Донец-м.Лисичанськ	0,1837	-0,2880	0,5664
р.Дунай-м.Братислава	0,2373	0,2334	0,1494
р.Тиса – м.Польгар	0,3120	0,1968	0,0823
р.Мюреш-м.Арад	0,3156	0,1901	0,0300
р.Сан – м.Радомишль	0,2909	0,2314	-0,1409
р.Західний Буг – м.Вишків	0,3053	0,1798	-0,2305
р.Дон-ст.Раздонська	0,1832	-0,3253	0,4932

Вагові коефіцієнти третьої компоненти також змінюють знак, що дозволяє виділити за знаком Північний та Південний райони. До Північного району відносяться рр. Дніпро (до м.Речиця), Сож, Десна, Прип'ять, Сан, Західний Буг. У Південному районі знаходяться рр.Дунай, Південний Буг, нижня течія р.Дон. Річка

Дністер (до м.Заліщики) знаходиться біля лінії розділу районів ($U_2 = 0$) й відноситься до Північного району.

Таким чином, можна зробити висновок, що другий за своєю значущістю фізичний процес обумовлює різницю в формуванні полів річного стоку річок Західної та Східної України, а третій – різницю між формуванням стоку півночі й півдня України.

Приклад 2

Завдання. За даними розкладання полів річного стоку по ПОФ виділити райони з синхронними коливаннями стоку.

Вихідні дані. Аналізується кореляційна матриця, побудована для 20-ти рядів річного стоку басейну р.Уссурі за період сумісних спостережень, починаючи з 1960 року й закінчуючи 1986 роком, тобто довжина розглянутих рядів становить 27 років. Басейн р.Уссурі знаходиться у зоні дії повітряних течій східно-азиатської мусоної системи [38]. У північній частині переважає дія мусона помірних широт, у південній – субтропічних. Зона мусону помірних широт розташовується між 45° пн.ш. та 65° пн.ш.

Отримані результати. Внесок перших компонент у загальну дисперсію вихідних даних дорівнює: 63% для першої компоненти, 20% - для другої та 5% - для третьої (Лобода Н.С., Нгуєн Ву Ань, 2006). Внесок перших двох компонент становить 83%, а перших трьох- 88% [24]. Отже, перших двох компонент достатньо для описування статистичної структури усього поля даних.

Аналіз результатів. Оскільки для виявлення просторових закономірностей зміни базисних функцій перших компонент розкладання за природними ортогональними функціями необхідні дані про координати центрів тяжіння водозборів, досліджуваний водозбір був вкритий координатною “сіткою” й положення центрів тяжіння було представлене у вигляді умовних координат (рис.4.4). Знак вагових коефіцієнтів першої компоненти розкладання не змінюється (табл.4.2), що інтерпретується як однорідний вплив найбільш масштабного фізичного процесу на формування полів річного стоку.

Зазвичай, перша компонента розкладання розглядається як статистичне відображення дії атмосферних процесів глобального масштабу. Вагові коефіцієнти другої компоненти змінюють знак, що у літературних джерелах (Д.Л.Смірнов, В.Л. Складенко,1986) інтерпретується як існування різниці у закономірностях коливань характеристик стоку на різних ділянках території, обумовленої процесами синоптичного масштабу [35].

Інакше кажучи, другий за значущістю фізичний процес обумовлює прояв несинхронності у коливаннях характеристик стоку. **Нульова ізолінія другої базисної функції розглядається як лінія розділу між районами з синхронними чи синфазними коливаннями стоку** і поділяє територію водозбору р. Уссурі навпіл (рис.4.5). Різницю у коливаннях стоку між північчю та півднем можна пояснити тим, що південна частина басейну р.Уссурі знаходиться у області дії субтропічних мусонів, а північна - в області впливу мусонів помірних широт. Тобто другий за значимістю фізичний процес інтерпретується як атмосферний процес синоптичного масштабу.

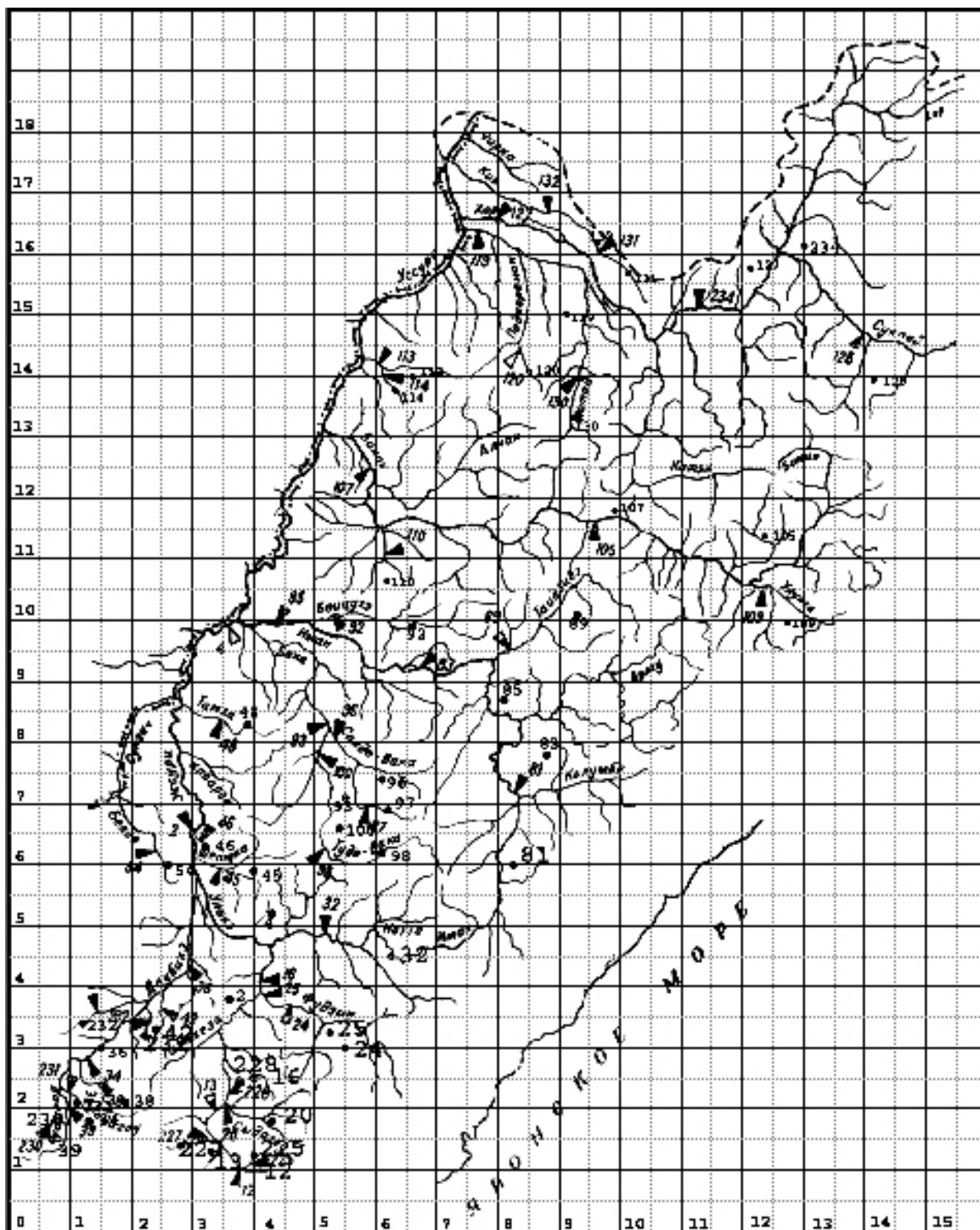


Рисунок 4.4 - Координатна сітка, гідрологічні пости та центри
тяжіння водозборів для басейну р.Уссурі

**Таблиця 4.2- Значення перших базисних функцій розкладання
полів річного стоку у басейні р.Уссурі**

Номер за картою	Річка –пост	U_1	U_2	U_3
2	р.Уссурі-пос.Кировський	0,2582	0,2184	0,0398
12	р.Улахе-с.Березняки	0,2587	0,2592	0,4140
20	р.Сидагоу-с.Извілінка	0,2208	0,2684	0,3047
25	р.Фудзін-с.Уборка	0,2566	0,2193	0,1200
36	р.Даубіхе-с.Яковлівка	0,2266	0,2933	-0,0678
42	р.Хоніхеза- с.Варфоломєєвка	0,2465	0,2414	0,0768
45	р.Шетуха-с.Криловка	0,2359	0,1001	-0,3672
48	р.Тамга-с.Тамга	0,2446	0,0027	-0,3074
83	р.Іман-с.Картун	0,2317	-0,0645	-0,1489
85	р. Іман-пос. Вагутон	0,2406	-0,0745	-0,1916
89	р.Сібічі-с. Сібічі	0,2078	-0,2371	-0,1818
93	р.Вака-с.Ракітне	0,1819	0,0402	-0,2741
98	р.Тудо-Вака-с.Аріадне	0,2067	0,0836	-0,3054
107	р.Бікін-ст.Звєньєва	0,2290	-0,2414	0,0145
114	р.Горбун-с.Пушкіно	0,1943	-0,3282	0,1602
119	р.Подхорьонок- с.Дормідонтівка	0,1914	-0,3282	0,1762
120	р.Правий Подхорьонок – лзу. Медвежий Ключ	0,2159	-0,3268	0,1797
128	р.Сукпай-мет.ст.Сукпай	0,2337	-0,2188	0,2627
131	р.Кія-с.Марусіно	0,2038	-0,3472	0,1932
204	р.Каменка-пос. Каменський	0,1548	0,2030	0,1436

Правильність прийнятих рішень щодо районування за синхронністю підтверджується підрахунками осередненого у

просторі коефіцієнта кореляції, який для усього басейну р.Уссурі дорівнює 0,52, для району 1 – 0,77 та для району 2 – 0,79.

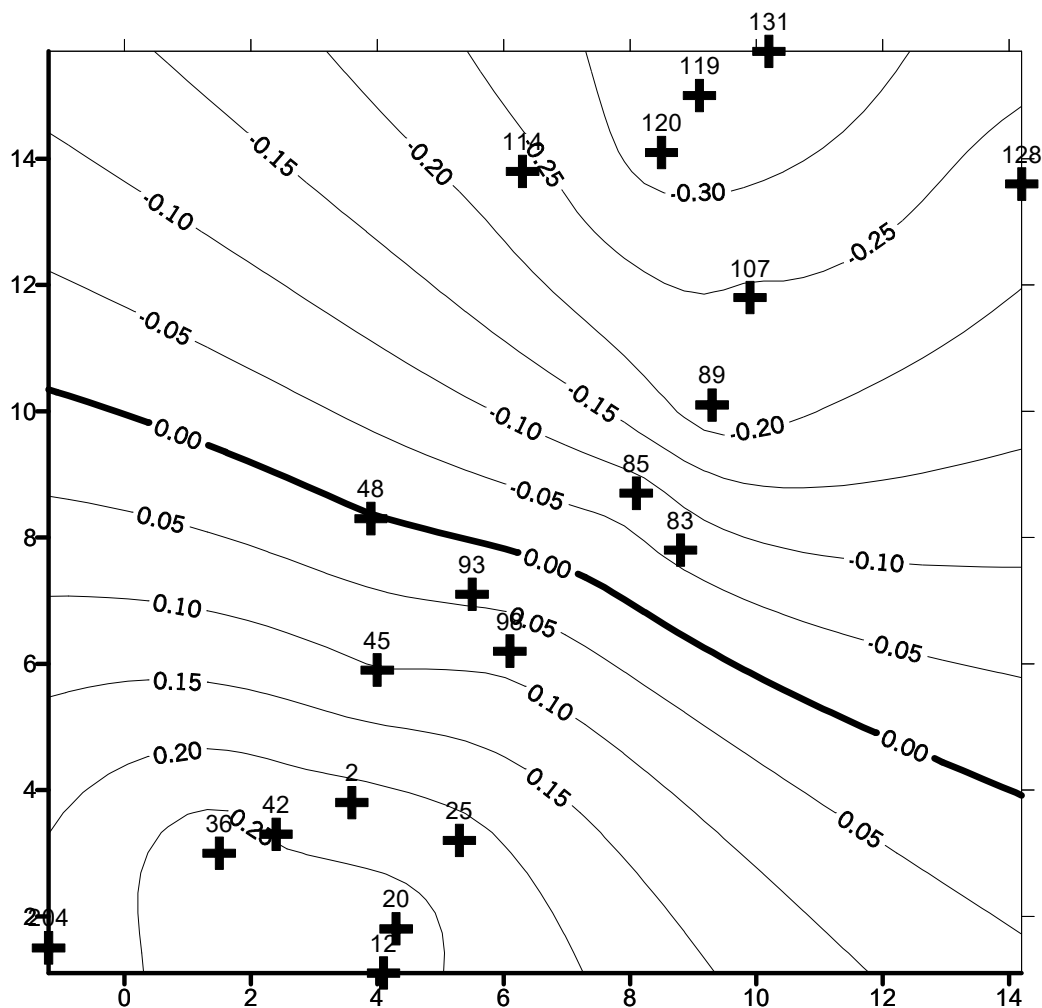


Рисунок 4.5 - Схема розподілу U_2 у басейні р.Уссурі
(+ - центри тяжіння водозборів)

Приклад 3.

Завдання. Виконати фільтрацію рядів стоку за першими компонентами

Вихідні дані - результати розкладання полів річного стоку за природними ортогональними функціями. Розглянуто матрицю коваріацій річного стоку для 9 гідрологічних станцій. Тривалість спостережень для кожного ряду дорівнює n .

Отримані результати. Установлено, що перші дві компоненти розкладання описують 80% вихідної інформації.

Задача: Отримати відфільтровані дані для четвертого ряду спостережень.

Матриця базисних функцій виглядає наступним чином

$$W = \begin{pmatrix} W_{11} & W_{12} & \cdot & \cdot & \cdot & W_{1p} & \cdot & \cdot & \cdot & W_{19} \\ W_{21} & W_{22} & \cdot & \cdot & \cdot & W_{2p2} & \cdot & \cdot & \cdot & W_{29} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ W_{p1} & W_{p2} & \cdot & \cdot & \cdot & W_{pp} & \cdot & \cdot & \cdot & W_{p9} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ W_{91} & W_{92} & \cdot & \cdot & \cdot & W_{9p} & \cdot & \cdot & \cdot & W_{99} \end{pmatrix}. \quad (4.30)$$

Матриця амплітудних функцій має вигляд

$$Z = \begin{pmatrix} z_{11} & z_{12} & \cdot & \cdot & \cdot & z_{1p} & \cdot & \cdot & \cdot & z_{1n} \\ z_{21} & z_{22} & \cdot & \cdot & \cdot & z_{2p2} & \cdot & \cdot & \cdot & z_{2n} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ z_{p1} & z_{p2} & \cdot & \cdot & \cdot & z_{pp} & \cdot & \cdot & \cdot & z_{pn} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ z_{91} & z_{92} & \cdot & \cdot & \cdot & z_{9p} & \cdot & \cdot & \cdot & z_{9n} \end{pmatrix}. \quad (4.31)$$

Щоб отримати матрицю “відфільтрованих” вихідних даних, представлених у центрованому виді, необхідно виконати скалярне множення матриць за виразом (4.26). Оскільки у даному прикладі головна інформація міститься тільки у двох перших компонентах, для отримання згладжених рядів амплітудні функції, починаючи з 3-ої і закінчуючи 9-тою, представляємо, як нульові згідно з (4.25)

$$\begin{aligned}
& \begin{bmatrix} W_{11} & W_{12} & W_{13} & W_{14} & W_{15} & W_{16} & W_{17} & W_{18} & W_{19} \\ W_{21} & W_{22} & W_{23} & W_{24} & W_{25} & W_{26} & W_{27} & W_{28} & W_{29} \\ W_{31} & W_{32} & W_{33} & W_{34} & W_{35} & W_{36} & W_{37} & W_{38} & W_{39} \\ W_{41} & W_{42} & W_{43} & W_{44} & W_{45} & W_{46} & W_{47} & W_{48} & W_{49} \\ W_{51} & W_{52} & W_{53} & W_{54} & W_{55} & W_{56} & W_{57} & W_{58} & W_{59} \\ W_{61} & W_{62} & W_{63} & W_{64} & W_{65} & W_{66} & W_{67} & W_{68} & W_{69} \\ W_{71} & W_{71} & W_{72} & W_{73} & W_{74} & W_{75} & W_{76} & W_{77} & W_{78} \\ W_{81} & W_{82} & W_{83} & W_{84} & W_{85} & W_{86} & W_{87} & W_{88} & W_{89} \\ W_{91} & W_{92} & W_{93} & W_{94} & W_{95} & W_{96} & W_{97} & W_{98} & W_{99} \end{bmatrix} \cdot \\
& \cdot \begin{bmatrix} z_{11} & z_{12} & z_{13} & \dots & z_{1n} \\ z_{21} & z_{22} & z_{23} & \dots & z_{2n} \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix} = \\
& = \begin{bmatrix} \Delta\tilde{X}_{11} & \Delta\tilde{X}_{12} & \Delta\tilde{X}_{13} & \dots & \Delta\tilde{X}_{1n} \\ \Delta\tilde{X}_{21} & \Delta\tilde{X}_{22} & \Delta\tilde{X}_{23} & \dots & \Delta\tilde{X}_{2n} \\ \Delta\tilde{X}_{31} & \Delta\tilde{X}_{32} & \Delta\tilde{X}_{32} & \dots & \Delta\tilde{X}_{3n} \\ \Delta\tilde{X}_{41} & \Delta\tilde{X}_{42} & \Delta\tilde{X}_{43} & \dots & \Delta\tilde{X}_{4n} \\ \Delta\tilde{X}_{51} & \Delta\tilde{X}_{52} & \Delta\tilde{X}_{53} & \dots & \Delta\tilde{X}_{5n} \\ \Delta\tilde{X}_{61} & \Delta\tilde{X}_{62} & \Delta\tilde{X}_{63} & \dots & \Delta\tilde{X}_{6n} \\ \Delta\tilde{X}_{71} & \Delta\tilde{X}_{72} & \Delta\tilde{X}_{73} & \dots & \Delta\tilde{X}_{7n} \\ \Delta\tilde{X}_{81} & \Delta\tilde{X}_{82} & \Delta\tilde{X}_{83} & \dots & \Delta\tilde{X}_{8n} \\ \Delta\tilde{X}_{91} & \Delta\tilde{X}_{92} & \Delta\tilde{X}_{93} & \dots & \Delta\tilde{X}_{9n} \end{bmatrix}, \tag{4.32}
\end{aligned}$$

звідки

$$\begin{aligned}
& \Delta\tilde{X}_{41} = W_{41}z_{11} + W_{42}z_{12} + W_{43} \cdot 0 + W_{45} \cdot 0 + W_{46} \cdot 0 + W_{47} \cdot 0 + \\
& + W_{48} \cdot 0 + W_{49} \cdot 0 = W_{41}z_{11} + W_{42}z_{12}
\end{aligned} \tag{4.33}$$

або

$$\Delta \tilde{X}_{42} = W_{41} z_{12} + W_{42} z_{22}; \quad (4.34)$$

$$\Delta \tilde{X}_{43} = W_{41} z_{13} + W_{42} z_{23}; \quad (4.35)$$

$$\Delta \tilde{X}_{44} = W_{41} z_{14} + W_{42} z_{24}; \quad (4.36)$$

$$\Delta \tilde{X}_{45} = W_{41} z_{15} + W_{42} z_{25}; \quad (4.37)$$

$$\Delta \tilde{X}_{46} = W_{41} z_{16} + W_{42} z_{26}; \quad (4.38)$$

$$\Delta \tilde{X}_{47} = W_{41} z_{17} + W_{42} z_{27}; \quad (4.39)$$

$$\Delta \tilde{X}_{48} = W_{41} z_{18} + W_{42} z_{28}; \quad (4.40)$$

$$\Delta \tilde{X}_{49} = W_{41} z_{19} + W_{42} z_{29}. \quad (4.41)$$

Перехід до величини згладженої величини $\tilde{x}_{ij}$ виконується за виразом (4.28).

Контрольні запитання

1. Визначення поняття “власне значення” матриці.
2. Визначення поняття “власний вектор” матриці.
3. Записати матричне рівняння повної проблеми власних значень.
4. Які задачі гідрології можуть бути вирішені на основі методу головних компонент.
5. Як визначається кількість компонент, які у достатній мірі описують вхідну інформацію.
6. У чому полягає розв’язання задачі “стискання” вхідної інформації.
7. У чому полягає розв’язання задачі “фільтрації” вхідної інформації.

5 РЕГРЕСІЙНІ МОДЕЛІ У ГІДРОЛОГІЧНИХ РОЗРАХУНКАХ

5.1 Основні положення регресійного аналізу

Модель лінійної парної регресії описує зв'язок генеральних сукупностей залежних випадкових величин X та Y [9]. Задача користувача полягає в тому, щоб за обмеженими даними спостережень (вибірками), зробити висновки про характер зв'язку в цілому [33]. У загальному випадку рівняння лінійної парної регресії є рівнянням умовного математичного сподівання випадкової величини Y , залежної від випадкової величини X :

$$m_{y/x} = m_y + r_{xy} \frac{\sigma_y}{\sigma_x} (x - m_x), \quad (5.1)$$

або рівняння умовного математичного сподівання випадкової величини X , залежної від Y :

$$m_{x/y} = m_x + r_{xy} \frac{\sigma_x}{\sigma_y} (y - m_y), \quad (5.2)$$

де $m_{y/x}, m_{x/y}$ - умовні математичні сподівання Y по X та X по Y , відповідно;

r_{xy} - коефіцієнт кореляції;

σ_y, σ_x - середні квадратичні відхилення випадкових величин Y та X , відповідно;

m_y, m_x - безумовні математичні сподівання випадкових величин Y та X , відповідно.

Для вибірок (рядів спостережень) рівняння (5.1) представляється у вигляді

$$\tilde{y}_i = \tilde{y}(x_i) = \hat{m}_{y/x} = ax_i + b, \quad (5.3)$$

де x_i - дискретні значення випадкової величини X ;

y_i - дискретні значення випадкової величини Y ;

$\tilde{y}_i$ - значення випадкової величини Y , розраховані за рівнянням регресії;

a, b - параметри рівняння.

Оцінки параметрів, які входять в рівняння лінійної парної регресії, розраховуються на основі методу найменших квадратів [28,33]. Це метод обробки емпіричного матеріалу, основна вимога якого полягає в тому, щоб сума квадратів відхилень даних спостережень від лінії регресії була найменшою, тобто

$$\Delta = \sum_{i=1}^n [y_i - \tilde{y}(x_i)]^2 = \min, \quad (5.4)$$

де n - довжина вибірки.

Відповідно до методу найменших квадратів a та b повинні бути такими, щоб сума Δ досягала свого мінімуму. Вимога екстремуму означає, що частинні похідні від Δ , узяті по a та b , дорівнюють нулю

$$\frac{\partial \Delta(a,b)}{\partial a} = \frac{\partial \left[\left(\sum_{i=1}^n y_i - ax_i - b \right)^2 \right]}{\partial a} = 0 ; \quad (5.5)$$

$$\frac{\partial \Delta(a,b)}{\partial b} = \frac{\partial \left[\left(\sum_{i=1}^n y_i - ax_i - b \right)^2 \right]}{\partial b} = 0 \quad (5.6)$$

Вирішуючи рівняння (5.5) та (5.6) відносно a та b , одержуємо

$$a = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2}; \quad (5.7)$$

$$b = \bar{y} - a\bar{x}, \quad (5.8)$$

де $\bar{y}, \bar{x}$ - середні арифметичні значення.

Чисельник дробу, який знаходиться в правій частині рівняння (5.7) є оцінкою коваріації (коваріаційного моменту) $\hat{K}_{xy}$, розрахованого за дискретною вибіркою завдовжки n

$$\hat{K}_{xy} = cov(x, y) = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}), \quad (5.9)$$

а знаменник – оцінкою дисперсії випадкової величини X

$$\hat{\sigma}_x^2 = S_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \overline{x^2} - \bar{x}^2, \quad (5.10)$$

де S_x - оцінка середнього квадратичного відхилення σ_x випадкової величини X .

Оцінка коефіцієнта кореляції, який відображає тісноту лінійного зв'язку між рядами спостережень, що представляють собою спостережені сукупності випадкових величин Y та X , записується у вигляді

$$\hat{r}_{xy} = r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 (y_i - \bar{y})^2}}, \quad (5.11)$$

Оцінка параметра a рівняння лінійної парної регресії виражається через коефіцієнт кореляції та середнє квадратичне відхилення випадкових величин Y та X , розрахованих за даними спостережень й позначеними як S_y та S_x

$$a = r \frac{S_y}{S_x}. \quad (5.12)$$

Математична модель множинної лінійної регресії представляється рівнянням виду

$$\tilde{y}_i - \bar{y} = b_1(x_{1i} - \bar{x}_1) + b_2(x_{2i} - \bar{x}_2) + b_3(x_{3i} - \bar{x}_3) \dots + b_k(x_{ki} - \bar{x}_k) \quad (5.13)$$

де $\tilde{y}_i - \bar{y}$ - центровані значення залежної величини (предиктанта);
 $x_{ji} - \bar{x}_j$ - центровані значення j -того аргументу (предиктора);
 $b_1, b_2, b_3, \dots, b_k$ - коефіцієнти рівняння множинної лінійної регресії;
 k - число предикторів.

Ідентифікація структури і параметрів рівняння множинної лінійної регресії виконується, виходячи з принципу найменших квадратів за (1.4), як і для випадку парної лінійної регресії. Результуючі формули для розрахунку коефіцієнтів рівняння множинної лінійної регресії за даними спостережень мають вид [30]

$$b_j = \frac{\sigma}{\sigma_j} \frac{D_{0j}}{D_{00}}, \quad (5.14)$$

де σ - оцінка середнього квадратичного відхилення досліджуваної характеристики y ;

σ_j - оцінка середнього квадратичного відхилення j -того предиктора;

D_{0j} - мінор визначника розширеної матриці коефіцієнтів кореляції, у якого викреслений перший рядок та стовпець, що відповідає змінній j , вказаній в мінорі;

D_{00} - мінор визначника розширеної матриці коефіцієнтів кореляції, у якого викреслений перший рядок і перший стовпець.

Елементами початкового визначника є коефіцієнти парної кореляції r_{ij} між предикторами i та j , а також коефіцієнти парної кореляції r_{0j} між предиктантом і предикторами. При цьому визначник розширеної матриці кореляцій другого порядку записується у вигляді

$$D = \begin{vmatrix} 1 & r_{01} & r_{02} \\ r_{10} & 1 & r_{12} \\ r_{20} & r_{21} & 1 \end{vmatrix}, \quad (5.15)$$

а мінори цього визначника записуються таким чином

$$D_{00} = \begin{vmatrix} 1 & r_{12} \\ r_{21} & 1 \end{vmatrix}, \quad (5.16)$$

$$D_{01} = \begin{vmatrix} r_{10} & r_{12} \\ r_{20} & r_{21} \end{vmatrix}, \quad (5.17)$$

$$D_{02} = \begin{vmatrix} r_{10} & 1 \\ r_{20} & r_{21} \end{vmatrix} \quad (5.18)$$

У записах виду (5.15-5.18) r_{0j} - коефіцієнт кореляції між предиктантом (0) та предиктором (j).

Рівняння лінійної множинної регресії для двох предикторів буде мати вигляд

$$\tilde{y}_i - \bar{y} = b_1(x_{1i} - \bar{x}_1) + b_2(x_{2i} - \bar{x}_2), \quad (5.19)$$

де

$$b_1 = \frac{\sigma}{\sigma_1} \frac{D_{01}}{D_{00}}; \quad (5.20)$$

$$b_2 = \frac{\sigma}{\sigma_2} \frac{D_{02}}{D_{00}}. \quad (5.21)$$

Рівняння (5.19) може бути записане й таким чином

$$\tilde{y}_i = b_1 x_{1i} + b_2 x_{2i} + b_0, \quad (5.22)$$

де

$$b_0 = \bar{y} - b_1 \bar{x}_1 - b_2 \bar{x}_2, \quad (5.23)$$

причому $\bar{y}$ - середнє арифметичне значення предиктанта;

$\bar{x}_1$ та $\bar{x}_2$ - середні арифметичні значення предикторів X_1 та X_2 .

Середнє квадратичне відхилення $\sigma_{\tilde{y}}$ спостережених даних від обчислених за рівнянням множинної лінійної регресії може бути визначене за наступною залежністю

$$\sigma_{\tilde{y}} = \sigma \sqrt{1 - R^2}, \quad (5.24)$$

де R - коефіцієнт множинної лінійної кореляції, який обчислюється за рівнянням

$$R = \sqrt{1 - \frac{D}{D_{00}}}, \quad (5.25)$$

причому D - визначник розширеної матриці коефіцієнтів кореляції.

Якщо парні коефіцієнти кореляції, які характеризують лінійний зв'язок між двома залежними випадковими величинами, змінюються від -1 до 1 , то повний коефіцієнт множинної кореляції рівняння множинної регресії змінюється від 0 до 1 .

Лінійна залежність відсутня при $r = 0$ та $R = 0$. У разі функціональної залежності $R = 1,0$. Чим більше коефіцієнт множинної кореляції, тим більшою мірою адекватності характеризується модель множинної регресії.

Оцінити міру адекватності можна і іншим шляхом, наприклад, шляхом перевірки статистичної гіпотези про те, що залишкова дисперсія (дисперсія вхідних даних, яка не описується рівнянням регресії) незначущо відрізняється від дисперсії предиктанта. Якщо така гіпотеза приймається, то прогноз (розрахунок) по моделі не відрізняється від випадкового [38].

5.2 Оцінка адекватності регресійної моделі за складовими дисперсії випадкової величини

За вибірковими даними повна або загальна дисперсія змінної Y може бути розрахована за формулою [38]

$$\sigma_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n - 1}. \quad (5.26)$$

В основі формули (5.26) лежить відхилення спостереженої величини y_i від середнього арифметичного значення (рис. 5.1).

Запишемо $y - \bar{y}$ у вигляді складових

$$y_i - \bar{y} = (y_i - \tilde{y}_i) + (\tilde{y}_i - \bar{y}), \text{ або } y_i - \bar{y} = (\tilde{y}_i - \bar{y}) + (y_i - \tilde{y}) \quad (5.27)$$

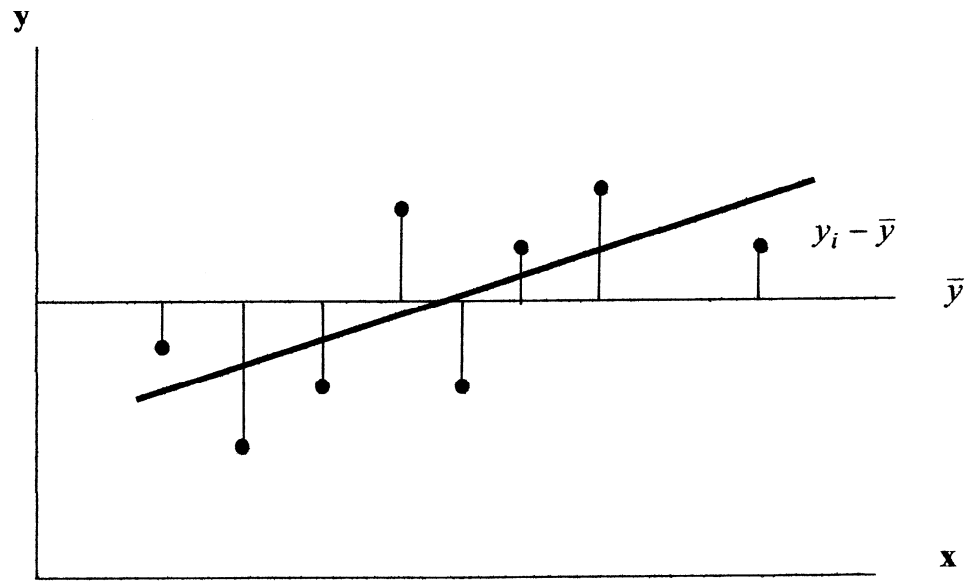


Рис.5.1. – Ілюстрація розсіювання спостережених значень y_i від $\bar{y}$

Тобто відхилення $(y_i - \bar{y})$ складається з відхилення значень $\tilde{y}_i$, обчислених за регресійним рівнянням, від середнього $\bar{y}$ та з відхилення розрахованих значень $\tilde{y}_i$ від спостережених y_i .

Обидві частини рівняння возведемо у квадрат

$$(y_i - \bar{y})^2 = [(\tilde{y}_i - \bar{y}) + (y_i - \tilde{y}_i)]^2. \quad (5.28)$$

Після підсумовування відхилень $(y_i - \bar{y})$ отримаємо

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\tilde{y}_i - \bar{y})^2 + 2 \sum_{i=1}^n (\tilde{y}_i - \bar{y})(y_i - \tilde{y}_i) + \sum_{i=1}^n (y_i - \tilde{y}_i)^2 \quad (5.29)$$

Складова $2 \sum_{i=1}^n (\tilde{y}_i - \bar{y})(y_i - \tilde{y}_i)$ дорівнює нулю у випадку, коли $(\tilde{y}_i - \bar{y})(y_i - \tilde{y}_i)$ некорельовані, що справедливо для нормально розподілених величин [33].

Отже, можна записати, що

$$\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1} = \frac{\sum_{i=1}^n (\tilde{y}_i - \bar{y})^2}{n-1} + \frac{\sum_{i=1}^n (y_i - \tilde{y}_i)^2}{n-1} \quad (5.30)$$

Рівняння (5.30) представляє собою суму дисперсій

$$\sigma_y^2 = \sigma_p^2 + \sigma_{зал}^2 \quad (5.31)$$

Величина σ_p^2 має назву поясненої дисперсії, оскільки вона показує, яка частина загальної дисперсії обумовлена залежністю Y від X .

Величина $\sigma_{зал}^2$ показує ту частину дисперсії величини Y , яка не описується залежністю Y від X і має назву залишкової.

Відхилення лінії регресії $(\tilde{y}_i)$ від $\bar{y}$ графічно представлене на рис.5.2.

Величина $(y_i - \tilde{y}_i)$ характеризує розсіювання точок, які відповідають даним спостережень, від значень, розрахованих за рівнянням лінійної регресії (рис. 5.3).

Якщо розрахунковою моделлю є рівняння лінійної парної регресії, то гіпотеза про адекватність обраної моделі перевіряється за критерієм Фішера (додаток Д), який формується таким чином

$$F = \frac{\sum_{i=1}^n (y_i - \bar{y})^2 / n - 1}{\sum_{i=1}^n (y_i - \tilde{y}_i)^2 / n - 1} = \frac{\sigma_y^2}{\sigma_{\text{зал}}^2}. \quad (5.32)$$

Гіпотеза H_0 про те, що залишкова дисперсія незначуще відрізняється від загальної дисперсії, не відхиляється, коли

$$F < F_{\text{кр}}(\alpha, \nu_1, \nu_2), \quad (5.33)$$

де $\nu_1 = n - 1$; $\nu_2 = n - 2$; α - заданий рівень значущості.

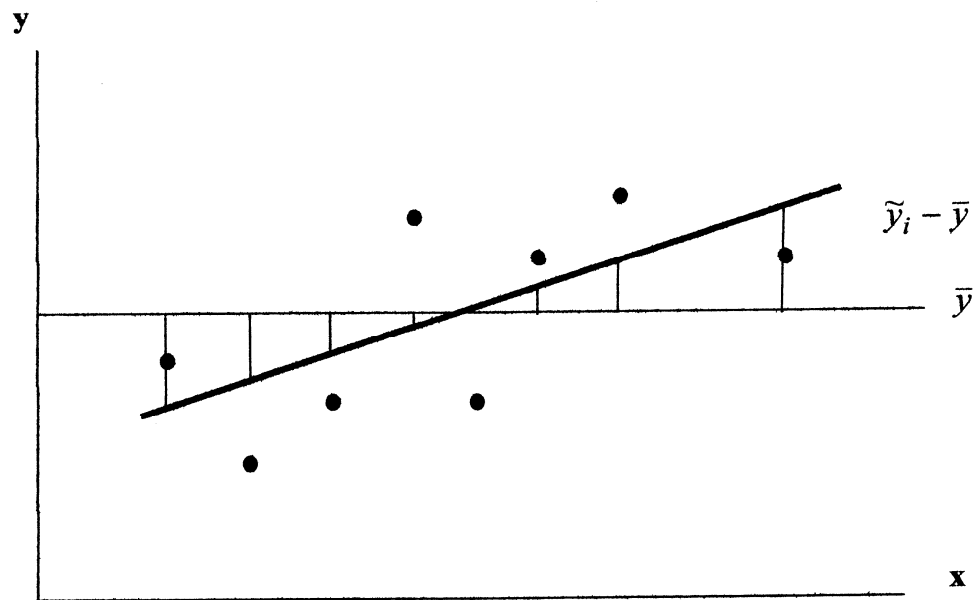


Рис. 5.2 – Ілюстрація відхилення лінії регресії від $\bar{y}$

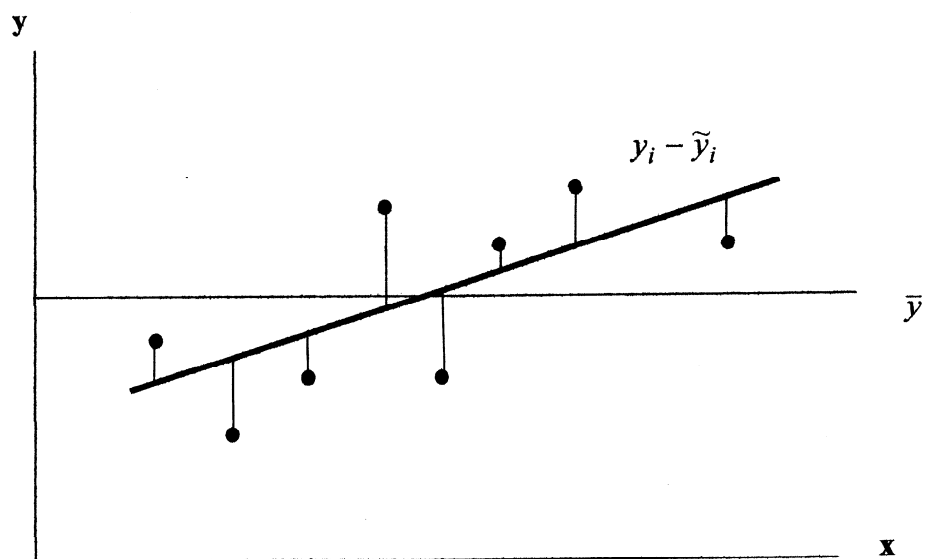


Рис. 5.3 – Ілюстрація відхилення спостережених даних від лінії регресії

Коли ж мова йде про множинну регресію, то залишкова дисперсія - це та частина дисперсії вихідної величини Y , яка не описується залежністю предиктанта Y від предикторів $X_1, X_2, \dots, X_k$. За аналогією з (5.32) перевірка гіпотези про адекватність моделі множинної лінійної регресії даним спостережень здійснюється за критерієм Фішера, який у даному випадку визначається за формулою

$$F = \frac{\sum_{i=1}^n (y_i - \bar{y})^2 / (n-1)}{\sum_{i=1}^{n-k} (y_i - \tilde{y}_i)^2 / (n-k-1)}, \quad (5.34)$$

де n - об'єм вибірок;

k - кількість предикторів;

$y_i, \tilde{y}_i$ - фактичні і розрахункові величини.

Нульова гіпотеза не відкидається, коли $F < F_{KP}(\alpha, \nu_1, \nu_2)$, де $\nu_1 = n-1$, а $\nu_2 = n-k-1$, α - заданий рівень значущості.

Певну проблему при побудові моделі множинної лінійної регресії складає вибір оптимальних предикторів, які відображають вплив основних чинників формування стоку.

Збільшення числа предикторів далеко не завжди приводить до кращих результатів, оскільки при збільшенні числа предикторів збільшується порядок матриці кореляції. Більш того, серед потенційних предикторів існує багато таких, які тісно зв'язані між собою, у зв'язку з чим матриця кореляції може бути погано обумовленою.

Звичайно це приводить до значних помилок при оцінках коефіцієнтів регресії і, отже, до погіршення якості розрахункової моделі. Щоб уникнути проблем такого роду з числа потенційних предикторів вибирають тільки ті, які є статистично значущими.

Вибір оптимальних предикторів називають операцією “просіювання” вихідних даних.

Просіювання може відбуватися за допомогою частинних коефіцієнтів кореляції.

5.3 Визначення частинних коефіцієнтів кореляції

Проведемо визначення частинного коефіцієнта кореляції й розглянемо алгоритм його оцінки. Припустимо, що на випадкову величину Y впливають дві випадкові величини X_1 і X_2 . Частинним коефіцієнтом кореляції між випадковими величинами Y і X_1 ($r_{yx_1 \bullet x_2}$) називають коефіцієнт кореляції між ними при умові, що вплив другої випадкової величини X_2 на Y вже є врахованим. Таким же чином визначається частинний коефіцієнт кореляції $r_{yx_2 \bullet x_1}$ [38].

Будемо вважати, що нам відома матриця кореляцій

$$R_x = \begin{vmatrix} 1 & r_{x_1 x_2} \\ r_{x_2 x_1} & 1 \end{vmatrix} \quad (5.35)$$

і вектор парних кореляцій між Y і X_1 та Y і X_2

$$R_{yx} = \begin{vmatrix} r_{yx_1} \\ r_{yx_2} \end{vmatrix}. \quad (5.36)$$

На їх основі сформуємо розширену матрицю кореляцій такого виду

$$\tilde{R} = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} \\ r_{yx_1} & 1 & r_{x_1x_2} \\ r_{yx_2} & r_{x_1x_2} & 1 \end{vmatrix}. \quad (5.37)$$

Як очевидно, вона утворюється з матриці R_x шляхом додавання до неї рядка та стовпця, що складаються з координат вектора R_{yx} . На основі матриці (5.37) розрахуємо мінори $|R_x|, D_{yx_i}, D_{yx_i}^-$ ($i=1,2$). Мінори 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 становиться перший стовпець і, після цього, викреслюють перший рядок і перший стовпець. Очевидно:

$$D_{yx_1} = r_{yx_1} - r_{yx_2} r_{x_1x_2}, \quad (5.38)$$

$$D_{yx_2} = r_{yx_2} - r_{yx_1} r_{x_1x_2} \quad (5.39)$$

Означення мінора $D_{yx_1}^-$ має такий сенс: це мінор визначника $|\tilde{R}|$, який не утримує парної кореляції r_{yx_1} . Очевидно, його можливо отримати, якщо із мінора $|\tilde{R}|$ викреслити рядок і стовпець, що

утримують парну кореляцію r_{yx_1} . Таким чином, мінор $D_{yx_1}^-$ набуде вигляду

$$D_{yx_1}^- = \begin{vmatrix} 1 & r_{yx_2} \\ r_{yx_2} & 1 \end{vmatrix} = 1 - r_{yx_2}^2, \quad (5.40)$$

$$D_{yx_2}^- = \begin{vmatrix} 1 & r_{yx_1} \\ r_{yx_1} & 1 \end{vmatrix} = 1 - r_{yx_1}^2. \quad (5.41)$$

Частинні коефіцієнти кореляції визначаються таким чином:

$$r_{yx_1 \bullet x_2} = \frac{D_{yx_1}}{\sqrt{|R_x| D_{yx_1}^-}} = \frac{r_{yx_1} - r_{yx_2} r_{x_1 x_2}}{\sqrt{(1 - r_{x_1 x_2}^2)(1 - r_{yx_2}^2)}}, \quad (5.42)$$

$$r_{yx_2 \bullet x_1} = \frac{D_{yx_2}}{\sqrt{|R_x| D_{yx_2}^-}} = \frac{r_{yx_2} - r_{yx_1} r_{x_1 x_2}}{\sqrt{(1 - r_{x_1 x_2}^2)(1 - r_{yx_1}^2)}} \quad (5.43)$$

Виникає питання, яка суттєва інформація утримується в частинних коефіцієнтах кореляції? Щоб відповісти на нього, розглянемо такий приклад.

Нехай коефіцієнти парної кореляції між випадковими величинами мають значення: $r_{yx_1} = 0.72$; $r_{yx_2} = 0.91$; $r_{x_1 x_2} = 0.84$. 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 дійсно чинить вплив на величину Y , треба розрахувати частинні коефіцієнти кореляції за допомогою формул (5.42) і (5.43).

Розрахунки дають такі їх значення: $r_{yx_1 \bullet x_2} = -0.05$; $r_{yx_2 \bullet x_1} = 0.75$. Таким чином ясно, що в дійсності на випадкову величину Y чинить вплив випадкова величина X_2 [38].

Кореляційний зв'язок Y з величиною X_1 , якщо урахувати її зв'язок з величиною X_2 , є не тільки незначним, але навіть має обернений характер.

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

$$\tilde{R} = \begin{vmatrix} 1 & r_{yx_1} & r_{yx_2} & r_{yx_3} & \dots & r_{yx_k} & \dots & r_{yx_n} \\ r_{yx_1} & 1 & r_{x_1x_2} & r_{x_1x_3} & \dots & r_{x_1x_k} & \dots & r_{x_1x_n} \\ r_{yx_2} & r_{x_2x_1} & 1 & r_{x_2x_3} & \dots & r_{x_2x_k} & \dots & r_{x_2x_n} \\ r_{yx_3} & r_{x_3x_1} & r_{x_3x_2} & 1 & \dots & r_{x_3x_k} & \dots & r_{x_3x_n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ r_{yx_k} & r_{x_kx_1} & r_{x_kx_2} & r_{x_kx_3} & \dots & 1 & \dots & r_{x_kx_n} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ r_{yx_n} & r_{x_nx_1} & r_{x_nx_2} & r_{x_nx_3} & \dots & r_{x_nx_k} & \dots & 1 \end{vmatrix} \quad (5.44)$$

і на її основі визначити мінори $|R_x|$, D_{yx_k} , $D_{yx_k}^-$ ($k=1,2,\dots,n$)

$$r_{yx_k \bullet x_1, x_2, \dots, x_{k-1}, x_{k+1}, \dots, x_n} = \frac{D_{yx_k}}{\sqrt{|R_x|} D_{yx_k}^-}, \quad (5.45)$$

де D_{yx_k} - міnor розширеної матриці R_{yx} , який складається таким чином: стовпець, на першому місці в якому розташовується кореляційний коефіцієнт r_{yx_k} , переставляється на перше місце, а на його місце ставиться перший стовпець, після чого викреслюються перший рядок і перший стовпець з матриці кореляцій;

$D_{yx_k}^-$ - мінор розширеної матриці R_{yx} , з якої викреслюється рядок і стовпець, що містять парний коефіцієнт кореляції r_{yx_k} , іншими словами, виключається парна кореляція r_{yx_k} ;

$|R_x|$ - визначник матриці, що містить тільки коефіцієнти кореляцій між предикторами.

Вибір оптимальних предикторів виконується за наступною схемою.

1. З числа предикторів вибирається той, який має найтісніший зв'язок з предиктантом і йому привласнюється номер 1.

2. Розраховується матриця частинних коефіцієнтів кореляції, за умови, що вплив першого предиктора вже враховано.

3. З числа частинних коефіцієнтів кореляції, які характеризують зв'язок між предиктантом і предикторами за умови, що вплив першого предиктора вже враховано, вибирається найбільший за абсолютною величиною (номер 2) і знов розраховується матриця частинних коефіцієнтів кореляції, але вже з урахуванням впливу перших двох предикторів.

Процедура повторюється до тих пір, поки на деякому $k + 1$ етапі всі частинні коефіцієнти кореляції не втрачають статистичну значущість.

Гіпотеза про статистичну незначущість параметра перевіряється за допомогою критерію Стюдента (додаток Ж). Вона не спростовується, якщо $t < t_{kp}(\alpha, \nu)$, де $\nu = m - k$.

5.4 Приклад аналізу розрахунків за моделлю множинної лінійної регресії з покроковим добором предикторів

Завдання. Необхідно добрати оптимальні предиктори та отримати розрахункове рівняння множинної лінійної регресії для визначення середньобагаторічної величини $\bar{q}$ річного стоку невивчених у гідрологічному відношенні водозборів басейну р. Уссурі.

Вихідна інформація. Розглянуто 40 водозборів у басейні р. Уссурі (табл. 5.1). До розрахунків залучений пакет статистичних програм „Microstat”.

Як потенційні предиктори розглядаються логарифм площі водозборів $\lg(F + 1)$; норма річних опадів $\bar{X}$; середня висота водозборів $H_{сер}$; заболоченість $f_{\bar{o}}$; залісеність $f_{л}$; умовна довгота λ ; умовна широта φ .

Таблиця 5.1 – Вихідні дані застосовані для річок водозбору р. Уссурі.

N ПП	$\bar{q}$, л/с·км ²	$\lg(F + 1)$	$\bar{X}$, мм	$H_{сер}$, м	$f_{\bar{o}}$, %	$f_{л}$, %	λ , см	φ , см
88	17.2	3.44	813	679	-	100	9.3	10.1
45	8.64	3.03	823	260	19	81	4.0	5.9
46	11.4	1.93	848	340	8	92	3.2	6.3
48	8.72	2.69	764	176	12	78	3.9	8.3
81	10.6	3.83	886	670	0.5	100	8.2	6.0
83	12.3	4.27	881	685	0.5	100	8.8	7.8
85	12.1	4.36	871	601	5	95	8.1	8.7
92	10.3	3.28	823	205	17	83	6.6	9.9
93	10.3	3.67	831	373	4	96	5.5	7.1
96	11.1	3.25	977	445	-	94	6.1	7.4
97	13.8	2.63	1005	591	-	100	6.2	6.9
98	12.5	3.07	1180	568	-	100	6.1	6.2
100	9.74	3.41	914	403	2	98	5.4	6.6
105	12.4	4.12	956	790	0.5	100	12.4	11.4
107	11.4	4.33	919	560	4	93	9.9	11.8
1	9.34	3.71	924	879	-	91	3.6	3.8
2	8.95	4.39	802	435	5	96	4.3	5.2
12	10.8	2.73	904	879	-	91	4.1	1.1
13	9.84	3.24	843	729	-	89	3.3	1.3
16	9.19	3.97	832	558	1	91	4.0	2.5
20	1.6	3.06	907	811	-	98	4.3	1.8
24	9.61	3.06	907	811	-	98	5.4	3.0
25	9.22	3.53	852	578	0.5	94	5.3	3.2
34	9.33	3.39	843	552	-	94	1.3	1.8
36	7.79	3.71	810	402	6	82	1.5	3.0

Продовження таблиці 5.1

N ПП	$\bar{q}$, л/с·км ²	$lg(F+1)$	$\bar{X}$, мм	$H_{сер}$, м	$f_{\bar{q}}$, %	f_{λ} , %	λ , см	φ , см
38	9.78	2.88	892	573	-	97	1.9	2.1
39	9.94	2.97	931	635	-	99	0.7	1.5
41	7.2	2.79	739	235	6	76	1.2	3.4
227	10.7	2.35	907	810	-	100	2.8	1.4
228	8.01	1.26	860	619	-	100	3.7	2.3
230	8.12	1.97	854	599	-	100	0.6	1.6
231	9.08	2.13	853	591	-	98	1.1	2.1
232	6.87	1.54	809	416	0.5	98	1.2	3.4
113	10.1	2.86	889	264	8	88	6.6	13.9
114	8.36	2.16	817	218	7	88	6.3	13.8
119	8.73	3.37	816	189	17	81	9.1	15.0
127	15.9	4.39	1048	629	0.5	94	12.2	15.7
128	13.3	3.49	1022	948	-	97	14.2	13.6
130	13.3	3.05	1040	350	-	99	9.3	13.3
131	11.3	2.70	905	216	10	81	10.2	15.7

Розв'язок задачі. На першому етапі розраховуються коефіцієнти лінійної парної кореляції між предиктантом та усіма предикторами. Обирається предиктор, який має найбільш тісний зв'язок з предиктантом ($\bar{q}$). У розглянутому випадку таким предиктором визнається умовна довгота λ , яка характеризується тісним кореляційним зв'язком з середньою багаторічною величиною стоку - $r_{\bar{q}\lambda} = 0,692$.

Отримане рівняння лінійної парної регресії має наступний вигляд

$$\begin{aligned}\bar{q} &= 0.440\lambda + 8.03; \\ \sigma_{\bar{q}} &= 1.61\end{aligned}\tag{5.46}$$

Підраховуються повна дисперсія предиктанта й регресійна та залишкова складові повної дисперсії

$$\sigma_{\bar{q}}^2 = \frac{\sum_{i=1}^n (q_i - \bar{q})^2}{n-1} = \frac{\sum_{i=1}^{40} (q_i - \bar{q})^2}{39} = \frac{189}{39} = 4.85 \quad (5.47)$$

$$\sigma_p^2 = \frac{\sum_{i=1}^n (\tilde{q}_i - \bar{q})^2}{n-1} = \frac{90.4}{39} = 2.32 \quad (5.48)$$

$$\sigma_{\text{зал}}^2 = \frac{\sum_{i=1}^n (q_i - \tilde{q}_i)^2}{n-1} = \frac{98.5}{39} = 2.51 \quad (5.49)$$

Підраховується критерій Фішера за (5.34)

$$F = \frac{189/39}{98.5/38} = \frac{4.85}{2.59} = 1.87 \quad (5.50)$$

$$\nu = 39; \nu = 38; F_{kp} = 1.7 \quad (\text{додаток Д}) \quad F > F_{kp}$$

Отже нульова гіпотеза про те, що залишкова дисперсія незначуще відрізняється від загальної відкидається.

Надалі розраховуються частинні коефіцієнти кореляції між предиктантом $\bar{q}$ та предикторами, які не увійшли до рівняння (5.46):

$$\begin{aligned}
r_{\bar{q},lg(F+1) \bullet \lambda} &= 0.16; \\
r_{\bar{q},\bar{x} \bullet \lambda} &= 0.44; \\
r_{\bar{q},H \bullet \lambda} &= 0.40; \\
r_{\bar{q},f_{\delta} \bullet \lambda} &= 0.45; \\
r_{\bar{q},f_{\lambda} \bullet \lambda} &= 0.46; \\
r_{\bar{q},\varphi \bullet \lambda} &= 0.26
\end{aligned}
\tag{5.51}$$

До складу оптимальних предикторів додається f_{λ} (залісеність). Цей предиктор має найбільш тісний зв'язок з $\bar{q}$, за умови, що вплив довготи урахований ($r_{\bar{q},f_{\lambda} \bullet \lambda} = 0.46$).

Рівняння множинної лінійної кореляції, яке описує зв'язок між нормою річного стоку та двома обраними предикторами має вигляд

$$\begin{aligned}
\bar{q} &= 0.105f_{\lambda} + 0.420\lambda - 1.66; \\
\sigma_{\bar{q}} &= 1.44; \quad R = 0.768
\end{aligned}
\tag{5.52}$$

Коефіцієнт множинної кореляції зріс, у порівнянні з коефіцієнтом лінійної парної кореляції від 0.692 до 0.768.

Регресійна складова, що показує, яка частина дисперсії величини $\bar{q}$ описується її залежністю від λ та f_{λ} , збільшується й становить 2.94, а залишкова складова зменшується до 2.09.

Критерій Фішера приймає значення

$$F = \frac{189/39}{77.3/37} = \frac{4.85}{2.09} = 2.32
\tag{5.53}$$

При $\nu_1 = 39$ та $\nu_2 = 37$ $F_{kp} = 1.8$.

Оскільки $F > F_{kp}$, розрахунки по моделі будуть відрізнятися від випадкових.

На наступному етапі знов обчислюються частинні коефіцієнти кореляції між предиктантом $\bar{q}$ та предикторами, які не увійшли до рівняння (5.52):

$$\begin{aligned}
r_{\bar{q},lg(F+1)\bullet\lambda,f_n} &= 0.25; \\
r_{\bar{q},\bar{X}\bullet\lambda,f_n} &= 0.27; \\
r_{\bar{q},H\bullet\lambda,f_n} &= 0.13; \\
r_{\bar{q},f_\delta\bullet\lambda,f_n} &= 0.14; \\
r_{\bar{q},\varphi\bullet\lambda,f_n} &= 0.00
\end{aligned}
\tag{5.54}$$

Найбільший частинний коефіцієнт кореляції за умови, що вплив довготи та залісеності урахований, установлений між нормою річного стоку ($\bar{q}$) та нормою опадів ($\bar{X}$) й дорівнює 0,27.

Оскільки величина частинного коефіцієнта кореляції $r_{\bar{q},\bar{x}\bullet\lambda,f_n}$ має невисоке значення, то необхідно перевірити його значущість за формулами [21]

$$t = \frac{r_{xy}}{\sqrt{\sigma_{r_{xy}}^2}}, \tag{5.55}$$

$$\sigma_{r_{xy}} = \frac{1 - r_{xy}^2}{\sqrt{n - 1}}, \tag{5.56}$$

де t - критерій Стьюдента.

$$\sigma_{r_{\bar{q},\bar{x}\bullet\lambda,f_n}} = \frac{1 - 0.27^2}{\sqrt{40 - 1}} = \frac{1 - 0.0729}{6.24} = \frac{0.927}{6.24} = 0.149 \tag{5.57}$$

$$t = \frac{0.27}{0.149} = 0.85 \tag{5.58}$$

Критичне значення t для рівня значущості 0.05 та числа степенів свободи $40-1=39$ дорівнює 3 (додаток Ж).

Отже, при $t < t_{kp}$ нульова гіпотеза про те, що $r_{x,y}$ не відрізняється від нуля ($H_0 : r_{xy} = 0$) не відхиляється, тобто частинний коефіцієнт кореляції $r_{\bar{q}, \bar{x} \bullet \lambda, f_{\lambda}} = 0.25$ не розглядається як значущий і подальшого добору предикторів не відбувається. Слід зазначити, що у більшості пакетів статистичних програм передбачається перевірка статистичної значущості усіх частинних коефіцієнтів кореляції.

Після виконання перевірки частинних коефіцієнтів кореляції на їх значущість, можна зробити висновок, що як розрахункове рівняння може бути прийняте рівняння (5.52), а оптимальними предикторами є довгота та залісеність.

На основі отриманого розрахункового рівняння виконуються перевірні розрахунки, які зводяться у таблицю.

На основі результатів розрахунків обчислюється відносна похибка розрахунків $\delta_i = \frac{(y_i - \tilde{y}_i)}{y_i} \cdot 100\%$ та устанавлюється середнє арифметичне значення абсолютних величин цієї похибки (табл.5.2).

Таблиця 5.2 – Оцінка похибки перевірних розрахунків

№	$\bar{q}_i$	$\tilde{q}_i$	$\Delta\bar{q} = \bar{q}_i - \tilde{q}_i$	$\delta_i = \frac{\Delta\bar{q}_i}{\bar{q}_i} \cdot 100\%$
1	17.20	12.76	4.44	25.80
2	8.64	8.53	0.11	1.23
3	11.40	9.35	2.04	17.80
4	8.72	8.17	0.54	6.20
5	10.60	12.29	-1.69	-15.90
6	12.30	12.55	-0.24	-1.95
7	12.10	11.72	0.37	3.05
8	10.30	9.83	0.46	4.50
9	10.30	10.74	-0.44	-4.27
10	11.10	10.78	0.31	2.88
11	13.80	11.45	2.34	16.90
12	12.50	11.41	1.08	8.64

продовження таблиці 5.2

13	9.74	10.90	-1.16	-11.90
14	12.40	14.06	-1.66	-13.38
15	11.40	12.27	-0.87	-7.63
16	9.34	9.41	-0.07	-0.75
17	8.95	10.23	-1.28	-14.30
18	10.80	9.62	1.17	10.83
19	9.84	9.08	0.76	7.72
20	9.19	9.58	-0.39	-4.24
21	11.60	10.44	1.15	9.90
22	9.61	10.90	-1.29	-13.40
23	9.22	10.44	-1.22	-13.20
24	9.33	8.76	0.56	6.00
25	7.79	7.58	0.20	2.56
26	9.78	9.33	0.44	4.59
27	9.94	9.03	0.90	9.05
28	7.20	6.83	0.36	5.13
29	10.70	10.02	0.67	6.26
30	8.01	10.40	-2.39	-29.80
31	8.12	9.10	-0.98	-12.06
32	9.08	9.10	-0.02	-0.22
33	6.87	9.14	-2.27	-33.04
34	10.10	10.36	-0.26	-2.57
35	8.36	10.23	-1.87	-22.36
36	8.73	10.67	-1.94	-22.22
37	15.90	13.34	2.55	16.03

$$|\delta|_{сер} = 10.08$$

Контрольні запитання

1. Записати вираз рівняння лінійної парної регресії та формули для визначення коефіцієнтів цього рівняння.
2. Записати вираз рівняння лінійної множинної регресії та вирази для визначення коефіцієнтів цього рівняння.
3. Записати вираз для розрахунку парного коефіцієнта кореляції.

4. Записати вираз для розрахунку коефіцієнта множинної кореляції.
5. Сутність методу найменших квадратів.
6. Граничні значення коефіцієнту кореляції.
7. Граничні значення коефіцієнту множинної кореляції.
8. Визначення частинного коефіцієнту кореляції.
9. Етапи добору оптимальних предикторів за допомогою частинного коефіцієнта кореляції.

6. ДИСКРИМІНАНТНИЙ АНАЛІЗ ЯК МЕТОД ПРИЙНЯТТЯ АЛЬТЕРНАТИВНИХ РІШЕНЬ ПРИ ГІДРОЛОГІЧНОМУ РАЙОНУВАННІ ТА ПРОГНОЗУВАННІ

6.1 Схема побудови розв'язувального правила

Дискримінантний аналіз - статистичний метод, який дозволяє вивчати різницю між двома або більше групами об'єктів по декількох змінних одночасно.

У гідрології часто виникає задача щодо віднесення того чи іншого об'єкта або спостереження до одного з відомих класів. Такими класами можуть бути гідрологічні райони, до яких має бути віднесений той чи інший об'єкт [21]. В практиці гідрологічного прогнозування часто виникає потреба скласти прогноз здійснення або нездійснення того чи іншого гідрологічного явища [26]. Такий прогноз називають **альтернативним**. Поставлені задачі називаються задачами прийняття альтернативних рішень, або задачами класифікації. Питання про віднесення об'єкта до тієї чи іншої сукупності (групи або класу) вирішується шляхом порівняння ознак, характерних для кожної із розглядуваних сукупностей, із ознаками самого розглядуваного об'єкта [38].

У теорії розпізнавання образів задача класифікації формулюється таким чином: на основі відомостей про окремих представників різних класів навчаючої системи із характерними для них ознаками (предикторами) необхідно знайти розв'язувальне правило, за яким той чи інший об'єкт може бути віднесеним до одного з класів. Сформульована задача є типовою задачею **розпізнавання образів**. Суть її полягає у тому, що, по-перше, необхідно розділити весь простір образів на два підпростори. У одному підпросторі розглядуване явище відбувається, а у другому – ні. По-друге, треба побудувати правило, за допомогою якого можна віднести образ, який підлягає розпізнаванню, до того чи іншого з підпросторів.

Нехай ми маємо множину V векторів-предикторів (образів), які складають простір зображень R_V . Припустимо, що цей простір розділяється на два підпростори R_{V1} та R_{V2} . У першому з них розташовується множина V_1 образів X , при яких явище відбувається, а у другому - множина V_2 образів X , коли явище не відбувається. Треба побудувати поверхню, яка б розділяла підпростори R_{V1} та R_{V2} . Наведемо для пояснення прості приклади.

Нехай простір буде двовимірним $R_V = R_V(x_1, x_2)$ (рис.6.1). Тоді ми маємо на площині (x_1, x_2) лінію $x_1 = f(x_2)$, що відділяє підпростір R_{V1} від підпростору R_{V2} . Досить простим буде й випадок трьохвимірного простору: $R_V = R_V(x_1, x_2, x_3)$ (рис.6.2). У цьому випадку підпростори R_{V1} та R_{V2} розділяє деяка поверхня у трьохвимірному просторі, рівняння якої має вид $x_3 = \varphi(x_1, x_2)$. Більш складні умови виникають, коли розглядаються образи із багатовимірному простору $R_V = R_V(x_1, x_2, \dots, x_n)$. Поверхня, яка розділяє цей простір на підпростори R_{V1} та R_{V2} називається *розділювальною гіперповерхнею*, а її рівняння має вигляд:

$$F(x_1, x_2, \dots, x_n) = 0. \quad (6.1)$$

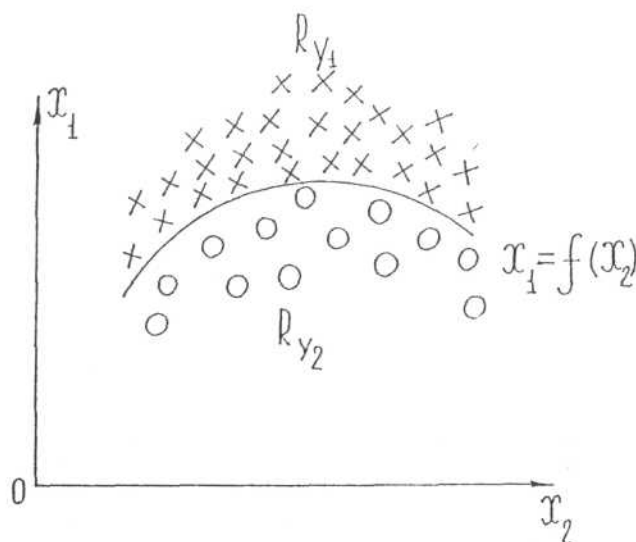


Рисунок 6.1 - Образи та розділювальна функція у двовимірному просторі[38]

Надалі отримується правило, за допомогою якого приймається рішення віднести вектор X , що підлягає розпізнаванню, до підпростору R_{V1} або підпростору R_{V2} . Це правило називають **розв'язувальним правилом**. Якщо у відповідності із побудованим правилом приймається рішення, що $X \in R_{V1}$, то явище прогнозується, коли ж приймається рішення, що $X \in R_{V2}$, то явище не прогнозується. Етап, який складається з побудови розділювальної гіперповерхні та розв'язувального правила, носить назву **етапа навчання**.

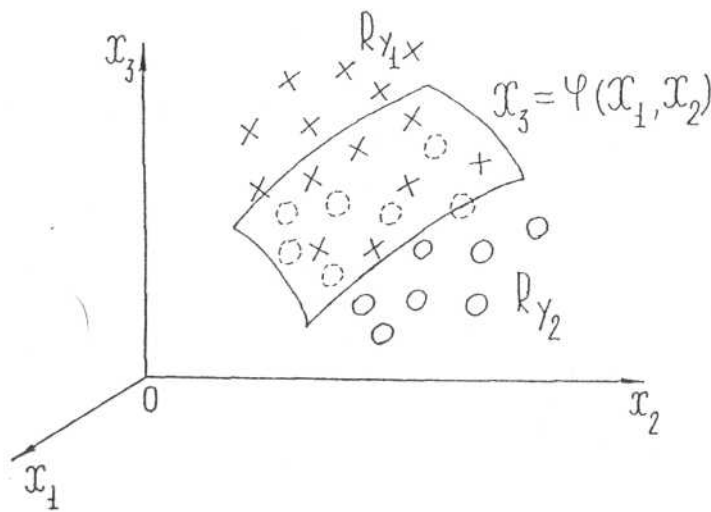


Рисунок 6.2.- Образи та розділювальна поверхня в тривимірному просторі [38]

Прийняття рішення про належність вектора X до підпросторів R_{V1} або R_{V2} називають **етапом розпізнавання**. Множина V векторів-предикторів, на основі якої реалізуються перелічені етапи, називається навчаючою сукупністю. Крім неї, створюється ще й перевірна сукупність, яка використовується для перевірки адекватності (відповідності) моделі альтернативного прогнозу фактичним даним. Саме розв'язувальне правило може бути представлене у вигляді деякої математичної функції, яку назовемо дискримінантною.

Функції, які забезпечують можливість віднесення об'єкта, який підлягає класифікації, до однієї з виділених груп, називають дискримінантними.

Застосування дискримінантного аналізу, з однієї сторони, передбачає принципову роздільність класів, а з іншого боку, допускає їхнє часткове “перекривання” (рис.6.3). Таким чином, рішення про віднесення явища або об'єкту до того чи іншого класу приймається з тією чи іншою долею похибки. “Перекривання класів” виражається у тому, що інтервали числових значень ознак у об'єктів, які належать до різних класів, перекриваються. Це утруднює віднесення об'єкта або явища до визначеного класу, якщо їхня кількісна ознака лежить у області перекриття. Наприклад, виділені два гідрологічних райони, які відрізняються один від одного кількісними характеристиками стоку. Але розподіл стокових величин у просторі підпорядковується закону географічної зональності. Отже, існують “приграничні” водозбори, на яких розглядувані характеристики можуть попадати у область “перекриття”. Коли класифікація виконується за однією з ознак, то водозбір буде віднесений до першого класу (району А), якщо ознака x буде меншою деякої величини x_0 , й до другого класу (району В), якщо $x > x_0$. Неправильні рішення позначимо через α_a та α_b . Величина x_0 може бути серединою інтервалу перекриття (x'_0) або абсцисою точки перетину (x''_0) кривих щільності розподілу ознаки x . Проте більш доцільно вибрати таку точку x_0^* , для якої буде виконуватись $\alpha_a = \alpha_b$. За цієї умови максимальна похибка (найбільша з двох) приймає мінімальне значення. Таким чином, коли розглядається одна ознака, то задача зводиться до побудови точки, яка розділяє дві сукупності. Результат класифікації визначається відносно цієї точки. Як уже відмічалось, при використанні двох ознак розділ сукупностей відбувається відносно кривої, трьох – відносно поверхні, і т.д.

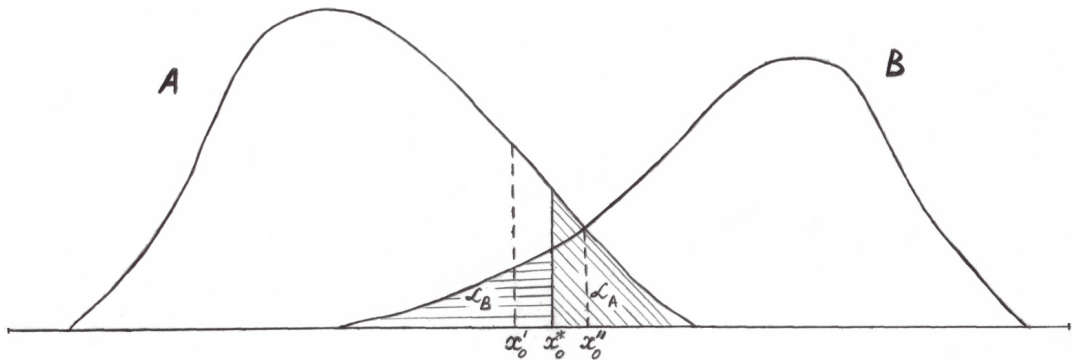


Рисунок 6.3 – Вирішальні правила при розмежуванні
одновимірних сукупностей

При вивченні двох ознак отримуємо двовимірний розподіл (рис.6.4). Проекції двовимірного розподілу на площини zx та zy відповідають щільності розподілу кожної із ознак. Видно, що урахування при класифікації тільки однієї з ознак не дозволяє повністю розділити сукупності через наявність перекриття. Якщо ж використовувати обидві ознаки, то розділення може бути з успіхом проведене за допомогою прямої MN .

Розглянемо основні ідеї теорії розпізнавання образів [36,38]. Позначимо через H_1 гіпотезу, згідно з якою образ $X \in V_1$. Альтернативною буде гіпотеза H_2 , згідно з якою, $X \in V_2$. Задача розпізнавання полягає у тому, щоб знайти правило, яке дозволяє обґрунтовано прийняти гіпотезу H_1 або H_2 . Всіляка процедура перевірки гіпотез передбачає, що приймаючи те чи інше рішення, ми можемо припустити помилку 1-го чи 2-го роду. Нагадаємо, що помилку 1-го роду ми припускаємо, коли відкидаємо правильну гіпотезу. Помилка 2-го роду пов'язана з прийняттям невірної гіпотези. Помилку другого роду також називають “похибкою хибної тривоги”.

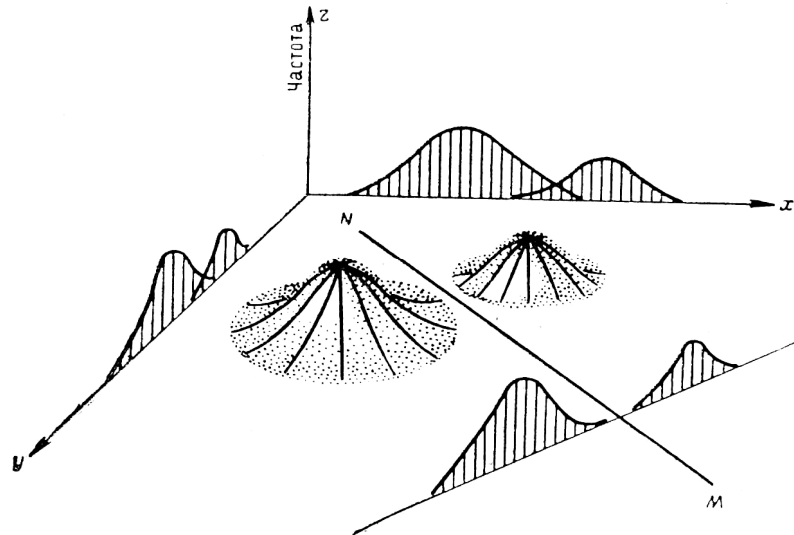


Рисунок 6.4 – Лінійне розмежування двовимірних сукупностей

Будемо вважати, що відомими є умовні ймовірності класів V_1 й V_2

$$P(x_1, x_2, \dots, x_n / V_1) \quad (6.2)$$

та

$$P(x_1, x_2, \dots, x_n / V_2). \quad (6.3)$$

Позначимо ймовірність помилки 1-го роду через P_a , а 2-го роду - через P_b . Знаючи ймовірності (6.2) та (6.3), а також апіорні ймовірності $P(V_1)$ та $P(V_2)$ класів V_1 і V_2 , можна розрахувати ймовірності помилок 1-го та 2-го роду [38].

Розв'язувальне правило або дискримінантна функція будується на основі функції подібності

$$\lambda(x_1, x_2, \dots, x_n) = \frac{P(x_1, x_2, \dots, x_n / V_1)}{P(x_1, x_2, \dots, x_n / V_2)}. \quad (6.4)$$

Величину вигляду

$$\frac{\delta_b P(V_2)}{\delta_a P(V_1)} = \theta \quad (6.5)$$

називають **порогом**.

Таким чином, можна прийти до такого розв'язувального правила:

$$\text{вектор } X \in V_1, \text{ якщо } \lambda(x_1, x_2, \dots, x_n) > \theta; \quad (6.6)$$

$$\text{вектор } X \in V_2, \text{ якщо } \lambda(x_1, x_2, \dots, x_n) < \theta. \quad (6.7)$$

Якщо є підстави вважати, що $\delta_a = \delta_b$ та $P(V_1) = P(V_2)$, то $\theta = 1$ й розв'язувальне правило набуває вигляду:

$$\text{вектор } X \in V_1, \text{ якщо } \lambda(x_1, x_2, \dots, x_n) > 1, \quad (6.8)$$

$$\text{вектор } X \in V_2, \text{ якщо } \lambda(x_1, x_2, \dots, x_n) < 1. \quad (6.9)$$

У загальному випадку розв'язувальне правило базується на нерівності вигляду

$$\frac{P(x_1, x_2, \dots, x_n / V_1)}{P(x_1, x_2, \dots, x_n / V_2)} > \frac{\delta_b P(V_2)}{\delta_a P(V_1)}, \quad (6.9)$$

яка може бути представленою у логарифмічному виді

$$\ln \frac{P(x_1, x_2, \dots, x_n / V_1)}{P(x_1, x_2, \dots, x_n / V_2)} > \ln \frac{\delta_b P(V_2)}{\delta_a P(V_1)} \quad (6.10)$$

Функцію вигляду

$$F(x_1, x_2, \dots, x_n) = \ln P(x_1, x_2, \dots, x_n / V_1) - \ln P(x_1, x_2, \dots, x_n / V_2) + \ln \frac{\delta_a P(V_1)}{\delta_b P(V_2)} \quad (6.11)$$

називають **дискримінантною функцією**.

Якщо використовується дискримінантна функція, то розв'язувальне правило записується таким чином:

$$X \in V_1, \text{ якщо } F(x_1, x_2, \dots, x_n) > 0; \quad (6.12)$$

$$X \in V_2, \text{ якщо } F(x_1, x_2, \dots, x_n) < 0; \quad (6.13)$$

Ясно, що рівняння $F(x_1, x_2, \dots, x_n) = 0$ є рівнянням розділювальної поверхні для підпросторів R_{V_1} та R_{V_2} .

Методи, які базуються на теорії прийняття статистичних рішень, мають такі обмеження: для їхньої реалізації необхідно знати щільності умовних розподілів образів у класах V_1 та V_2 . Ці закони розподілів є багатовимірними, й отримання їх аналітичною виду на основі множини векторів-предикторів класів V_1 і V_2 , є дуже складною задачею. Тому вважають, що вид законів розподілу є відомим. У такому разі задача зводиться до необхідності на основі вибірок векторів-предикторів отримати оцінки параметрів їхніх законів розподілу. Така процедура носить назву відновлення закону розподілу. При практичних реалізаціях цього підходу найбільш часто приймається, що класи векторів-предикторів підпорядковуються умовним нормальним законам розподілу. У дійсності ці припущення строго не виконуються. Дуже добре відомо, що для багатьох гідрометеорологічних величин, які виступають у ролі предикторів, вимога підкорення нормальному закону розподілу не виконується. Але, як показує досвід, це не вносить суттєвих похибок у результати розрахунків, якщо більшість предикторів має одномодальний розподіл [38].

6.2 Побудова розв'язувального правила на основі багатовимірного нормального закону розподілу

Будемо вважати, що вектори-предиктори вигляду

$$X_j = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \dots \\ x_{nj} \end{pmatrix} \quad (6.14)$$

підпорядковуються багатовимірному нормальному розподілу. Параметри нормального закону розподілу системи n випадкових величин, як відомо, можуть бути представлені [7] у вигляді векторів математичних сподівань

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \dots \\ \mu_n \end{pmatrix}, \quad (6.15)$$

та векторів дисперсій

$$D = \begin{pmatrix} D_1 \\ D_2 \\ \dots \\ D_n \end{pmatrix}, \quad (6.16)$$

і матриці коваріацій розміром $n \times n$. Тому щільності умовних законів розподілу для класів V_1 і V_2 мають вигляд:

$$P(x_1, x_2, \dots, x_n / V_1) = \frac{1}{(2\pi)^{n/2} |K_1|^{1/2}} \exp \left[-\frac{1}{2} (X - \mu_1)' K_1^{-1} (X - \mu_1) \right], \quad (6.17)$$

$$P(x_1, x_2, \dots, x_n / V_2) = \frac{1}{(2\pi)^{n/2} |K_2|^{1/2}} \exp \left[-\frac{1}{2} (X - \mu_2)' K_2^{-1} (X - \mu_2) \right], \quad (6.18)$$

де $P(x_1, x_2, \dots, x_n / V_1)$ – умовна ймовірність класу V_1 ;

$P(x_1, x_2, \dots, x_n / V_2)$ – умовна ймовірність класу V_2 ;

$P(V_1)$ – апіорна ймовірність класу V_1 ;

$P(V_2)$ – апіорна ймовірність класу V_2 ;

V_1 та V_2 – множини векторів-предикторів;

δ_a - ціна помилки 1-го роду;

δ_b – ціна помилки 2-го роду;

X – матриця вихідних даних;

μ_1, μ_2 – вектори математичних сподівань 1-го та 2-го класів;

K_1, K_2 – матриці коваріацій 1-го та 2-го класів;

n – кількість векторів-предикторів.

Після підставлення (6.16) та (6.17) до дискримінантної функції виду (6.11) прийдемо до такого рівняння:

$$F(x) = -\frac{n}{2} \ln 2\pi - \frac{1}{2} \ln |K_1| - \frac{1}{2} (X - \mu_1)' K_1^{-1} (X - \mu_1) + \frac{n}{2} \ln 2\pi + \frac{1}{2} \ln |K_2| + \frac{1}{2} (X - \mu_2)' K_2^{-1} (X - \mu_2) + \ln \frac{\delta_a P(V_1)}{\delta_b P(V_2)}. \quad (6.19)$$

Після скорочень дискримінантна функція (6.19) набуде вигляду

$$F(x) = \frac{1}{2} \left[(X - \mu_2)' K_2^{-1} (X - \mu_2) - (X - \mu_1)' K_1^{-1} (X - \mu_1) + \ln \frac{|K_2|}{|K_1|} \right] + \ln \frac{P(V_1)}{P(V_2)} \quad (6.20)$$

де K_1^{-1}, K_2^{-1} - обернені коваріаційні матриці.

У (6.20) приймається умова, що ціни помилок першого й другого роду однакові ($\delta_a = \delta_b$).

Дискримінантна функція $F(x)$, яка описується формулою (6.20), має назву **квадратичної дискримінантної функції**. Така назва пов'язується з тим, що перші два члени у квадратних дужках являють собою квадратичні форми, тобто багаточлени степеня не більше другого.

Для визначення дискримінантної функції за (6.20) необхідно виконати операції обернення коваріаційних матриць K_1 та K_2 . Ця операція може привести до негативних наслідків, коли матриці коваріацій є погано обумовленими. У такому разі похибки, що містяться в коваріаціях, можуть привести до великих похибок

коефіцієнтів квадратичних форм й, таким чином, до помилок на етапі розпізнавання. У ряді випадків ці похибки можуть бути більшими ніж ті похибки, які виникають при прийнятті умови:

$$K_1 = K_2 = K. \quad (6.21)$$

Прийняття умови (6.21) означає, що матриці коваріацій двох класів однакові. Частіш за все приймається умова

$$K = \frac{K_1 + K_2}{2}. \quad (6.22)$$

Якщо прийняти умову (6.22) та вважати, що $P(V_1) = P(V_2)$, то з (6.20) отримаємо:

$$\begin{aligned} F(x) &= \frac{1}{2} \left[(X - \mu_2)' K^{-1} (X - \mu_2) - (X - \mu_1)' K^{-1} (X - \mu_1) \right] = \\ &= (\mu_1 - \mu_2)' K^{-1} X + \frac{1}{2} \left[\mu_2' K^{-1} \mu_2 - \mu_1' K^{-1} \mu_1 \right] \end{aligned} \quad (6.23)$$

Дискримінантна функція вигляду (6.23) називається **лінійною дискримінантною функцією**.

Використання дискримінантного аналізу значно спрощується, якщо є підстави вважати коваріації рівними нулю. Таке припущення можна зробити, коли всі недіагональні елементи матриць коваріацій класів V_1 та V_2 значно менші від діагональних й ними можна знехтувати.

Тоді коваріаційна матриця K набуде вигляду

$$K = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_n^2 \end{pmatrix}, \quad (6.24)$$

де $\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2$ - дисперсії.

Обернена матриця K^{-1} розраховується наступним чином

$$K^{-1} = \begin{pmatrix} \frac{1}{\sigma_1^2} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_2^2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \dots \frac{1}{\sigma_n^2} \end{pmatrix}. \quad (6.25)$$

У такому разі операція обернення матриць коваріацій значно спрощується. При цьому спрощується й процедура розрахунків коефіцієнтів квадратичних форм у квадратичній дискримінантній функції (6.20).

Коли виконується умова (6.21) й, крім того, можна вважати коваріаційну матрицю діагональною, значно спрощується вигляд лінійної дискримінантної функції (6.23). У цьому випадку вона дорівнює:

$$F(x) = \sum_{i=1}^n \frac{\mu_{1i} - \mu_{2i}}{\sigma_i^2} x_i + \sum_{i=1}^n \frac{\mu_{2i}^2 - \mu_{1i}^2}{\sigma_i^2} \quad (6.26)$$

Звичайно, при практичному використанні розглянутих дискримінантних функцій замість складових векторів математичних сподівань і дисперсій предикторів використовують їхні статистичні оцінки, тобто середні арифметичні значення та вибіркві дисперсії предикторів.

Критерієм якості проведення розділяючої поверхні є число Махаланобіса, яке характеризує відстань між центрами виділених класів. При використанні лінійного виду дискримінантної функції число Махаланобіса визначається за матричним рівнянням вигляду

$$\Delta = (\mu_1 - \mu_2)' K^{-1} (\mu_1 - \mu_2), \quad (6.27)$$

а при використанні дискримінантної функції спрощеного виду (6.26) число Махаланобіса представляється у такій формі

$$\Delta = \sum_{i=1}^n (\mu_{1i} - \mu_{2i})^2 / \sigma_i^2. \quad (6.28)$$

Чим більше число Махаланобіса, тим менше ймовірність похибки класифікації. При $\Delta=11$ ймовірність похибки класифікації досягає 5% [38], через це до розрахунків та прогнозів залучають таке розв'язувальне правило, для якого $\Delta > 11$.

Значущість різниці між двома класами можна також перевірити за допомогою критерію Фішера F . Величина F розраховується таким чином

$$F = \frac{m_1 m_2}{m_1 + m_2} \cdot \frac{m_1 + m_2 - n - 1}{m_1 + m_2 - 2} \cdot \frac{\sum_{i=1}^n a_i (\bar{x}_{1i} - \bar{x}_{2i})^2}{n}, \quad (6.29),$$

де m_1 та m_2 кількість розглянутих об'єктів у кожній групі;
 a_i - коефіцієнти лінійної спрощеної дискримінантної функції;
 $\bar{x}_1$ та $\bar{x}_2$ - середні арифметичні значення для i - тої ознаки.

Коли різниця між середніми арифметичними значеннями є суттєвою, розрахована величина F перевищує критичне значення $F > F_{kp, \alpha}(\nu_1, \nu_2)$, де α - рівень значущості, а ν_1, ν_2 - число ступенів свободи ($\nu_1 = n, \nu_2 = m_1 + m_2 - 1$).

6.3 Приклади застосування дискримінантної функції у гідрологічних розрахунках та прогнозах

Приклад 1

Завдання. Виконати ідентифікацію (розпізнавання) окремих водозборів лівобережної частини басейну р.Дністер за характеристиками водозборів на основі дискримінантного аналізу із застосуванням лінійної дискримінантної функції.

Вихідна інформація. Розглянуто 35 водозборів лівобережжя р.Дністер з такими ознаками: середня висота водозборів $H_{сер}$, уклін водозборів I , заболоченість $f_{\bar{o}}$, залісеність $f_{л}$.

Розв'язок задачі.

Розділення водозборів на два райони (класи) було виконано на основі аналізу даних спостережень за річним стоком річок лівобережжя р.Дністер: більшість статистичних параметрів річок Поділля, які впадають у р.Дністер зліва значно відрізняється від статистичних параметрів річного стоку інших лівобережних приток. Така різниця може бути поясненою різними умовами формування стоку, насамперед, чинниками підстильної поверхні, які можуть бути представленими такими ознаками як середня висота водозборів $H_{сер}$, уклін I , заболоченість $f_{\bar{o}}$, залісеність $f_{л}$. На цій основі були сформовані навчаючі виборки (класи). До першого класу (район А) увійшли 17 водозборів, до другого (район В) – 18 водозборів.

Коваріаційні матриці виділених районів мають такий вигляд.

Коваріаційна матриця класу А

	$H_{сер}$	I	$f_{\bar{o}}$	$f_{л}$
$H_{сер}$	428	401	-19.0	134
I	401	913	-35.1	131
$f_{\bar{o}}$	19.0	-35.1	4.81	-2.4
$f_{л}$	135	131	-2.40	49.6

Коваріаційна матриця класу В

	$H_{сер}$	I	$f_{\bar{o}}$	$f_{л}$
$H_{сер}$	505	777	-65.2	196
I	777	315	-11.6	458
$f_{\bar{o}}$	65.2	-11.6	3.08	-0.41
$f_{л}$	196	458	-0.41	29.0

Виходячи з аналізу коваріаційних матриць, можемо прийняти умову (6.20), тобто $K = K_1 = K_2$, причому

$$K = \frac{K_1 + K_2}{2}. \quad (6.29)$$

Таким чином, до виведення розв'язувального правила можна застосувати лінійну дискримінантну функцію.

Після використання пакету статистичних програм та проведення розрахунків за даними навчаючих виборок, отримуємо лінійну дискримінантну функцію виду

$$F(x) = -9.52 - 0.001H_{сер} + 0.01I + 1.14f_{\bar{o}} + 0.37f_{л}. \quad (6.30)$$

Число Махаланобіса дорівнює $\Delta = 11.1$, тобто ймовірність похибки класифікації менша 5%.

Таким чином, можна зробити висновок, що отримане розв'язувальне правило (6.30) можна застосувати до ідентифікації (розпізнання) водозбору за ознаками. Наприклад, розглядається водозбір р.Стрвяж – с.Луки, який не увійшов до навчальної вибірки даних й для якого треба здійснити ідентифікацію. Необхідно визначити приналежність цього водозбору до класу А чи класу В, тобто вирішити задачу розпізнавання образу.

Ознаки цього водозбору такі: $H_{сер} = 500\text{м}$; $I = 156\text{проміле}$; $f_{\bar{o}} = 0,5\%$; $f_{л} = 44\%$. Отримана згідно із (6.29) величина $F(x) = 8.39$. Отже, $F(x) > 0$ й можна зробити висновок, що водозбір р.Стрвяж – с.Луки належить до району А.

Дискримінантна функція може бути також застосованою до уточнення меж виділених районів. Наприклад, коли необхідно з'ясувати належність “приграничного” водозбору до першого чи другого гідрологічного району та правильно провести границю.

Приклад 2

Завдання. Побудувати розв’язувальне правило для альтернативного прогнозу появи льодових явищ на ділянці р.Дністер – м.Могильов – Подільський та р.Дністер – с.Сороки. (Альтернативним називається прогноз, якій випускається за принципом: “явище буде спостерігатися” або “явище не буде спостерігатися”).

Вихідна інформація. Розглянуті матеріали про осінньо-зимовий сезон на р. Дністер за 1966-2000р. Предикторами, які обумовлюють появу плавучого льоду, є:

$\sum \theta$ – сума від’ємних температур повітря від дати переходу температури повітря через 0°C до дати утворення плавучого льоду у розглядуваних створах (означена сума характеризує в непрямий спосіб витрати тепла водної маси);

ϑ_0 - температура води у створах Могильов – Подільський та Сороки перед появою плавучого льоду, яка має непрямим чином описувати теплоємність водної маси;

H_1 – рівень води в день переходу температури повітря через 0°C ;

H_2 – рівень води в день появи плавучого льоду у створах.

Розв’язок задачі.

На практиці для прогнозування дат появи льодових явищ зазвичай застосовуються фізико – статистичні залежності, які базуються на установленні стохастичних зв’язків між необхідною для утворення льоду сумою від’ємних температур повітря та показником теплового стану водної маси, яким є температура води [22,34]. Теоретичною основою прогнозів появи льодових явищ є метод теплового балансу граничної поверхні «вода – повітря» [1,5], рівняння якого записується наступним чином

$$A + B = 0, \quad (6.31)$$

де A – потік тепла від водної маси до поверхні поділу “вода – повітря”;

B – результуюча теплових потоків на водній поверхні.

Прогноз появи плавучого льоду можна випускати при виконанні умови, яка представляється у вигляді

$$\vartheta \leq -\frac{B}{\gamma}, \quad (6.32)$$

де ϑ - температура водної поверхні;

γ - коефіцієнт віддачі тепла від водного об'єкта до поверхні поділу “вода – повітря”.

Льодоутворення починається тоді, коли температура водної поверхні знижується до температури замерзання, а тепловіддача водною поверхнею тепла у атмосферу B перевищує приплив тепла A з водної товщі. Для прогнозу появи плавучого льоду використовують залежність вигляду [5]

$$\sum \Theta^- = f(V_0), \quad (6.33)$$

де $\sum \Theta^-$ - сума від'ємних температур повітря, яка підраховується від дати переходу температур повітря через 0°C до дати утворення плавучого льоду; при цьому використовуються дані метеорологічної станції, розташованої поблизу розглядуваної ділянки річки;

V_0 - початкова температура води напередодні переходу температури повітря від позитивних до від'ємних значень або за попередній інтервал часу.

Дата появи льодових явищ відповідає добі, на яку накопичується мінімально необхідна сума від'ємних температур повітря. Застосування графічних побудовань обмежує можливості прогнозу. По перше, використовується мала кількість показників, які визначають розглядуване явище. По друге, розділювальна лінія, яка відокремлює випадки утворення або неутворення льодових явищ, проводиться на графічній площині суб'єктивно, як нижня огинаюча випадків, коли плавучий лід спостерігався.

На рис. 6.5 показаний класичний вигляд прогностичної методики, яка використовується для прогнозу появи льодових явищ. Вище лінії розділу розташовується підпростір R_{V1} , нижче лінії розділення – підпростір R_{V2} . Лінію розділення застосовують для прогнозу наступним чином. Для заданої температури води V_0 за розділювальною лінією визначають $\sum |\Theta|_{\min}$, тобто мінімально

необхідну для утворення льодових явищ суму від'ємних температур повітря. Надалі підраховується сума від'ємних температур повітря від дати переходу через 0°C до дати випуску прогнозу з використанням даних метеорологічного прогнозу. Якщо $\sum|\Theta| < \sum|\Theta|_{\min}$, то появи плавучого льоду не буде. Якщо $\sum|\Theta| \geq \sum|\Theta|_{\min}$, то поява льоду прогнозується на дату накопичення $\sum|\Theta|_{\min}$.

Щоб використати лінійну дискримінанту функцію (6.23) при складанні альтернативного прогнозу появи льодових явищ на річці, треба здійснити "етап навчання".

Навчаюча сукупність, використовується для того, щоб знайти статистичні оцінки векторів математичних сподівань класів μ_1 і μ_2 , якими є вектори середніх значень $\bar{X}_1$ і $\bar{X}_2$ предикторів, а також коваріаційних матриць $\hat{K}_1$ і $\hat{K}_2$. Клас має бути сформований таким чином, щоб кожне спостереження належало тільки до одного класу. Побудовою лінійної дискримінантної функції завершується етап навчання.

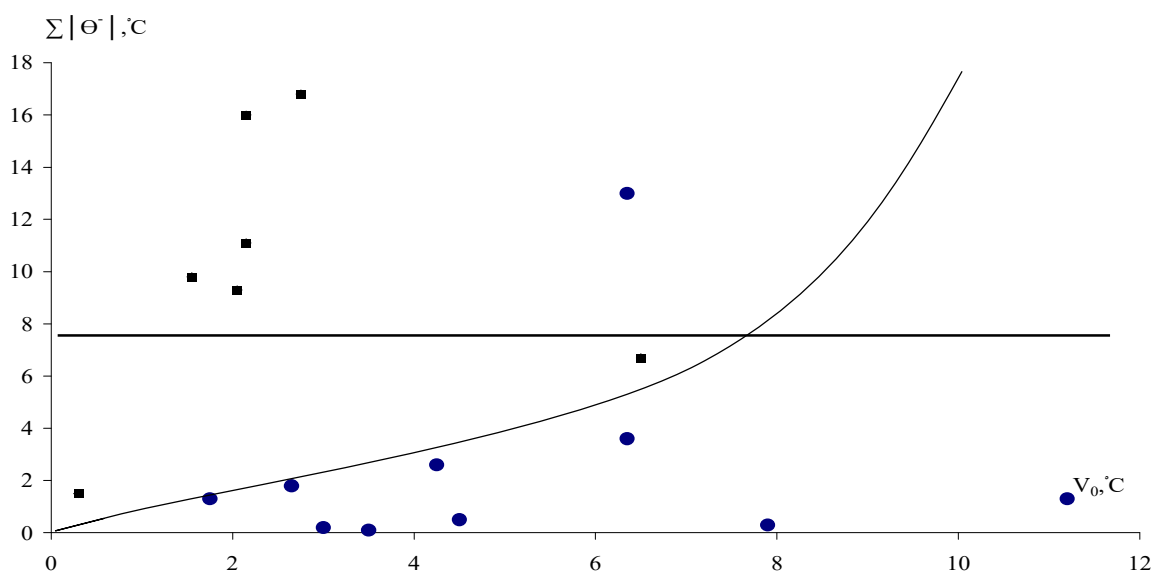


Рисунок 6.5 – Графік залежності $\sum|\Theta| = f(V_0)$ для р. Дністер – м.Могильов-Подільський

● – випадки, коли плавучий лід не утворюється; ▲ – випадки, коли плавучий лід утворюється.

За даними спостережень були утворені два класи або дві навчальні вибірки. Першому класу відповідають випадки, коли перехід температури повітря через 0°C супроводжувався утворенням плавучого льоду, а другому класу – випадки, коли не супроводжувався. Певна кількість випадків, що не увійшли до навчальних класів, була залучена для утворення третьої групи, за якою виконувалась адаптація прогнозової моделі.

На першому етапі як предиктори були використані тіж предиктори, що й при прогнозах за фізико – статистичними залежностями, тобто температура води V_0 та $\sum|\Theta^-|$.

Для створу р. Дністер – м. Могильов – Подільський була отримана дискримінантна функція виду [26]

$$F(x) = 1.78\sum|\Theta^-| - 0.69V_0 - 9.73, \quad (6.34)$$

$$\Delta = 21.6;$$

а для створу р. Дністер – м. Сороки

$$F(x) = 0.12\sum|\Theta^-| - 0.55V_0 + 0.41 \quad (6.35)$$

$$\Delta = 5.38$$

На другому етапі до числа предикторів були додані такі як H_1 та H_2 , де H_1 – рівень води в день переходу температури повітря через 0°C , H_2 – рівень води в день появи плавучого льоду. Результати розрахунків показали, що використання як предикторів H_1 та H_2 не впливає на кінцевий результат. Це відбувається внаслідок того, що рівні змінюються у часі незначно. Використання значень H_1 у прогнозах сприяє збільшенню завчасності прогнозу.

Отримана дискримінантна функції для створу р. Дністер – м.Могильов – Подільський має такий вигляд

$$F(x) = 3.43\sum|\Theta^-| - 0.77V_0 - 0.12H_1 + 0.31; \quad (6.36)$$

$$\Delta = 32.1,$$

а для створу р. Дністер – Сороки

$$F(x) = 0.16 \sum |\Theta^-| - 0.56V_0 - 0.02H_1 + 4.71; \quad (6.37)$$

$$\Delta = 24.8,$$

де Δ - число Махаланобіса.

Зростання числа Махаланобіса свідчить про покращення прогностичної методики за рахунок введення додаткового предиктора H_1 .

Після побудови прогностичної методики переходять до **етапу розпізнавання**, тобто прогнозу. Значення дискримінантної функції залежать від значень характеристик, які входять до рівняння дискримінантної функції й називаються **дискримінантними змінними**.

Введення у прогностичну методику додаткового предиктора H_1 значно покращило якість та ефективність прогностичної методики. Оцінка якості та ефективності прогностичної методики при прогнозах дат наступу явища оцінюється через забезпеченість допустимої похибки прогнозів

$$P = \frac{n - m}{n} * 100\% , \quad (6.38)$$

де P – забезпеченість прогнозу;

n – загальна кількість прогнозів;

m – кількість прогнозів, які не виправдались.

В короткострокових прогнозах дат появи льодових явищ допустима похибка приймається в залежності від завчасності метеорологічного прогнозу. Допустима похибка дорівнює половині завчасності метеорологічного прогнозу.

Виконані перевірні прогнози для р. Дністер - м. Могильов-Подільський та для р. Дністер – м. Сороки показали, що забезпеченість перевірних прогнозів зростає від 85% до 93% при застосуванні дискримінантної функції.

Таким чином, застосування дискримінантної функції до альтернативних прогнозів дозволяє уникнути суб'єктивності при побудові прогностичної методики та підвищити якість прогнозів.

Контрольні запитання

1. Який прогноз називається альтернативним.
2. У чому полягає задача розпізнавання образів.
3. Що таке розділювальна крива та розділювальна поверхня.
4. З якою метою будується дискримінантна функція.
5. Етапи побудови дискримінантної функції.
6. Критерій якості розрахункової або прогностичної методики, побудованої у вигляді дискримінантної функції.

7. ФРАКТАЛИ У ГІДРОЛОГІЇ

7.1 Поняття про фрактали

Термін “фрактал” був введений Мандельбротом [39], співробітником дослідницького центру імені Томаса Дж. Уотсона корпорації ІВМ в Йорктаун-Хейтсі (шт. Нью-Йорк) і походить від латинського слова *fractus*, що означає “ламати, розбивати”. Під фрактальними об’єктами розуміють, насамперед, об’єкти з сильно покращеною, негладкою формою. Головною відзнакою фрактального об’єкта від усіх інших є перевищення відповідної йому розмірності Хаусдорфа-Безиковича над його топологічною розмірністю [29]. Топологічна розмірність визначається наступним чином. Об’єкт має розмірність $d_T = k$, якщо може бути розділений на дві незв’язні частини шляхом виключення з нього $(k - 1)$ - мірної множини точок. Початкова розмірність нульова, вона, наприклад, відповідає фігурі із скінченного числа точок. Топологічна розмірність $d_T = 1$ характеризує криву, $d_T = 2$ - поверхню, $d_T = 3$ - шар. Топологічна розмірність може бути тільки цілим числом. Розмірність об’єкта за Хаусдорфом у випадку кривої визначається за формулою

$$d = k - \alpha, \quad (7.1)$$

$$\alpha = (k - 1)\alpha', \quad (7.2)$$

де k - розмірність простору;

α' - швидкість убуття Фур’є спектра.

Об’єкт буде фрактальним, коли має місце строга нерівність

$$d > d_T. \quad (7.3)$$

Для визначення фрактальної розмірності кривої можна підійти і з іншого боку. Для цього необхідно розглянути процедуру визначення довжини кривої як суми ланок апроксимуючої її ломаної. Якщо крива фрактальна, то сума довжин відрізків (η) ломаної при $\eta \rightarrow 0$ буде наближатися до безкінечності і складати величину порядку $\eta^{\alpha'-1}$ для плоскої кривої та $\eta^{2\alpha'-2}$ - для просторової. Вираз для довжини ломаної записується у вигляді

$$L_\eta = \lambda \eta^{1-d}, \quad (7.4)$$

або

$$\lambda = L_0^d, \quad (7.5)$$

де L_0 - відстань по прямій між двома граничними точками за умови, що фрактальність розповсюджується до масштаба L_0 ;

λ - величина з дробовою розмірністю, яка є мірою подовжності та викривленості граничної безперервної кривої і називається *мірою Хаусдорфа*.

Багато які структури володіють фундаментальною властивістю масштабної регулярності, відомої як інваріантність по відношенню до масштабу або властивість “самоподібності”. Іншими словами, *якщо розглядати об'єкти в різному масштабі, то постійно виявляються одні й ті ж фундаментальні елементи (фрактали)*. Важливо підкреслити, що спочатку фрактали застосовувалися як своєрідна мова геометрії. Проте, на відміну від звичних об'єктів евклидової геометрії (пряма лінія, коло й таке інше) фрактали не можуть бути безпосередньо спостереженими. Фрактали виражаються не в первинних геометричних формах, а в алгоритмах, наборах математичних процедур. Саме їх пошук і обґрунтування є центральною задачею сучасної теорії фракталів.

Багато які природні процеси, не дивлячись на те, що вони підкорюються певним детерміністичним законам, на достатньо великих часових інтервалах є непередбачуваними й проявляють схожі закономірності у варіаціях в різних часових масштабах подібно тому, як об'єкти, які характеризуються інваріантністю (самоподібністю) у просторових масштабах, проявляють схожі структурні закономірності в просторі.

Незалежно від природи або методу побудови у всіх фракталів є одна важлива загальна властивість: ступінь складності їхньої структури може бути виміряна певним характеристичним числом - фрактальною розмірністю. Таким чином, *фрактал є математичним об'єктом, що має дробову розмірність на відміну від традиційних математичних фігур цілої розмірності*. Визначення фрактальної розмірності може відбуватися різними методами. За ідеєю Мандельброта визначити фрактальну розмірність можна шляхом “підрахунку квадратів”. Уявімо об'єкт складної форми, який “вкритий” квадратами. Одна частина квадратів буде містити у собі елементи розглядуваного об'єкту, а друга – ні, тобто у другій частини квадрати будуть пустими (незайнятими об'єктом). Кількість непустих квадратів N залежить від форми об'єкту та від розмірів

квадрату E . Постулюється, що величина N пропорційна $1/E^d$ (чим дрібніша сітка, тим більше непустих квадратів). Показник степеня d і є розмірністю фрактального об'єкта.

Цей випадок може бути описаний у загальному математичному вигляді. Якщо множина розбивається на N підмножин, кожна з яких в n раз менше всієї множини, то фрактальна розмірність дорівнює

$$d = \frac{\ln N}{\ln n}, \quad (7.6)$$

або

$$N = n^d. \quad (7.7)$$

Розмірність d , визначену за (7.7) ще називають розмірністю самоподібності, де N - кількість об'єктів, подібних даному, але з розмірами в n разів меншими за даний об'єкт. Коефіцієнт подібності дорівнює $1/n$.

Якщо поблизу точки фрактальної кривої виділити невелику область, то частина кривої, що попала до цієї області, буде подібною до вихідної кривої, тобто буде мати місце подібність кривої своїй частині. Виражається це у тому, що кількість відрізків ломаної з довжиною ланки η_1 , які розміщуються між сусідніми вершинами ломаної з довжиною ланки η_2 ($\eta_2 < \eta_1$), буде залежати від відношення $\frac{\eta_2}{\eta_1}$, але не від η_2 та η_1 , розглянутих окремо одна від

одної. Фрактальна розмірність або розмірність самоподібності згідно із (7.4) та (7.6) розраховується наступним чином

$$d = \frac{\ln\left(\frac{L_\eta}{\eta}\right)}{\ln\left(\frac{L_0}{\eta}\right)}, \quad (7.8)$$

де L_η - довжина кривої, яка описується відрізками довжиною η ;

L_0 - границя поширення фрактальних розмірностей;

d - фрактальна розмірність.

Властивість самоподібності є однією з фундаментальних властивостей фрактальних об'єктів. Друга їх важлива властивість – однорідність. Для фрактальної кривої, наприклад, ця властивість проявляється у тому, що всі її ділянки між сусідніми вершинами апроксимуючої ломаної породжують однакову кількість відрізків другої, більш дрібномасштабної ломаної. Для природних фрактальних об'єктів властивості однорідності та самоподібності можна визначити у статистичному змісті.

Фрактальні властивості об'єктів можуть проявлятися не на всій області масштабів, характерних для об'єкта, а лише у певному діапазоні $L_1 \ll l \ll L_2$. Величина L_1 залежить від масштабу складових елементів об'єкту. Величина L_2 відповідає довжині радіуса кореляції [36]. На масштабах $l > L_2$ властивість самоподібності зникне. Об'єкти, фрактальні властивості яких спостерігаються на малих масштабах й не проявляються на великих, часто називають однорідними фракталами. L_1 і L_2 можна визначати як внутрішній та зовнішній масштаби фрактальності. З виразу (7.8) можна отримати рівняння, яке за змістом відповідає рівнянню (7.4)

$$L_\eta = L_0^d \eta^{1-d} = \frac{L_0^d}{\eta^{d-1}}. \quad (7.9)$$

Фрактальні розмірності можуть бути визначені на основі статистичного підходу з використанням (7.4). У першому випадку будуються залежності $\lg L_\eta = f(\lg \eta)$, для яких тангенс кута нахилу буде дорівнювати $(1-d)$. У другому випадку будуються згідно із (7.4) та (7.5) залежності вигляду $\lg L_{\eta_0} = f(\lg L_0)$, де тангенс кута нахилу дорівнює d . У першому випадку незалежною змінною є довжина ланки апроксимуючої ломаної η , у другому - L_0 при фіксованому η_0 . Величина η_0 повинна бути близькою до внутрішнього масштабу фрактальності.

Проте, в більшості випадків масштабні множники неоднорідні, тобто у фракталів є цілий спектр скейлінгів (scaling – масштабування). Такі фрактали називаються мультифракталами й характеризуються спектром фрактальних розмірностей.

Фрактали надають можливість надзвичайно компактного способу описування об'єктів та процесів. Фрактальний підхід до опису різних гідрологічних величин останнім часом активно розвивається в прикладній гідрології [17,19,21]. Ідея масштабної

самоподібності топографії земної поверхні, що була вперше показана в роботах Мандельбротта, знайшла свій подальший розвиток в роботах, де розглядається гідравліко-геометрична подібність річкових систем [29].

Серед російських й українських вчених, які вивчають мультифрактальні властивості гідрометеорологічних об'єктів слід віднести А.Г.Бернадського (1990), який досліджував турбулентні структури, С.С. Іванова (1994), який розглядав фрактальні властивості глобального рельєфу. Сучасний погляд на генезис фрактальних розмірностей турбулентних пульсацій в атмосфері, представлений ученими В.Д. Русовим та О.В. Глушковим (2003) .

Зміна величини d може бути використана як основа районування або типізації річкових систем, а сам масштабний множник d може використовуватися при побудові моделей гідрологічних процесів як чинник.

7.2 Фрактальна геометрія у гідрології

Фрактальні властивості виявлені при розгляді таких гідрографічних характеристик як довжина берегової лінії, площа водної поверхні та об'єм водної маси, характерні глибини, планові розміри водойм та водотоків. До основних гідрографічних характеристик водотоків відносяться довжина, уклон та звивистість. З перерахованих характеристик найбільш яскраво фрактальні властивості проявляються при розгляді берегової лінії водойм. Структура берегової лінії обумовлена трьома самостійними групами факторів: геологічними, гідрографічними, гідродинамічними. Останні пов'язані із структурою течії та хвильовою діяльністю у прибережних зонах, що визначає процеси наміву та розмиву берегів. У практичній гідрології при вимірюванні довжини берегової лінії або довжини звивистих річок за допомогою топографічних карт, шукана довжина апроксимується ламаними з вершинами, які лежать на вимірюваних кривих, за допомогою однакових відрізків довжиною η . Довжина η визначається в залежності від масштабу карти. Відомо, що L_η залежить від масштабу карти та величини відрізка η , причому визначена довжина L_η зростає із зменшенням

η . Передбачалося, що при малих значеннях η величина L_η повинна наближатися до кінцевого значення L , яке і слід приймати як істинну довжину. Однак, на практиці при достатньо великій звивистості досягнути цього не вдалося. Навпаки, величина $L_\eta = N \cdot \eta$, де N - кількість вимірювань за допомогою відрізка η при $\eta \rightarrow 0$ необмежено зростає. Цей ефект був вперше установлений англійським фізиком Річардсоном, який намагався установити довжину узбережжя о.Британія методом наближення замкнутої ламаної лінії. При цьому величина d згідно із (7.4) була установлена рівною 1.24, а для узбережжя Австралії – 1.13. Наведені дані свідчать про фрактальну структуру берегової лінії, показником покрайності якої може бути величина d . Нижня границя розмірності берегової лінії дорівнює 1, верхня – 2. Якщо $d \rightarrow 1$, то берегова лінія є гладкою, коли ж $d \rightarrow 2$, то ступінь її покрайності зростає.

Мандельбротом (1967) була також звернута увага на важливу властивість ламаної, яка описує довжину берегової лінії

$$N = L \cdot \eta^d, \quad (7.10)$$

де N - кількість вимірювань;

L - вимірювана довжина.

Тобто при постійному d величина N залежить лише від співвідношення L та η , а не від L та η розглянутих по окремо, отже, ламана є самоподібною. Окремі ділянки берегової лінії або річки значно відрізняються ступенем звивистості, й, отже, і величини d повинні розрізнятися. Однак, можна обмежитися властивістю локальної самоподібності на невеликих ділянках і знаходити загальну довжину як суму довжин при різних d , або ж знаходити усереднене значення d .

Як показали дослідження, величина d для річок також може змінюватися від 1 до 2 (Л.С. Кучмент, 2008).

Граничні масштаби фрактальної моделі довжини річки відповідають ширині річки $B(L_1)$ та ширині дна її долини $B_0(L_2)$. У кількісному відношенні ці масштаби можуть відрізнятися на декілька порядків.

Як відомо, коефіцієнт звивистості річки визначається за такою формулою:

$$K_{36} = \frac{L}{AB}, \quad (7.11)$$

де L - довжина річки з урахуванням її звивистості;

AB - довжина прямої, яка поєднує витік да гирло річки.

Для короткої ділянки річки, подовженість якої не перевищує зовнішній масштаб фрактальності (тобто має порядок ширини дна долини річки), коефіцієнт звивистості можна записати у вигляді

$$K_{36} = \frac{L_\eta}{L_0} = \left(\frac{L_0}{\eta} \right)^{d-1}, \quad L_0 = L \sim B_0 \quad (7.12)$$

L - відстань по прямій між створами, які обмежують ділянку.

Для ділянки річки подовженістю більшою за B_0 ($L \gg B_0$), наприклад, при розгляді русла річки від витoku до гирла, використовуються такі позначення визначення внутрішнього L_1 та зовнішнього L_2 масштабів фрактальності

$$L_1 = \alpha B, \quad L_2 = cB_0, \quad (7.13)$$

де α, c - постійні.

У такому випадку коефіцієнт звивистості набуває вигляду

$$K_{36} = \frac{L_{p\eta}}{L_0} = \frac{nL_\eta}{nL_2} = \frac{L_2^d \eta^{1-d}}{L_2} = \left(\frac{cB_0}{\eta} \right)^{d-1}, \quad (7.14)$$

де $L_{p\eta}$ - довжина річки, яка визначається за методом наближення ломаної з довжиною ланки η ($\alpha B < \eta < cB_0$);

L_η - довжина русла на короткій ділянці ($L = cB_0$).

Якщо $\eta = \alpha B$, рівняння (7.14) набуде вигляду

$$K_{36} = \left(\frac{cB_0}{\alpha B} \right)^{d-1} . \quad (7.15)$$

За допомогою теорії фракталів стало можливим пояснити давно відомі закономірності будови річкової мережі. Рисунок річкової мережі має основні типи: радіальний відцентровий або радіальний доцентровий; деревоподібний; перистий; решетчатий; паралельний або субпаралельний; концентричний. Для достатньо великих річкових систем само подібність досить легко простежується: будова кожної з приток подібна будові головної річки.

Річкову систему можна представити сукупність самоподібних індивідуальних потоків (рис.7.1с) - кривої Коша (рис.7.1а) та деревоподібної кривої (рис.7.1б), або як ієрархічну (порядкову) модель, у якій річкова система розглядається як сукупність відрізків (притоків) різної довжини без урахування фрактальності індивідуальних водотоків (рис.7.2). Р.С. Хортон вніс пропозицію розглядати річкову мережу як відкриту деревоподібну систему, що складається з ієрархії різних підсистем. Самою дрібною одиницею вважається елементарний нерозгалужений водотік, якому надається перший порядок. Річки другого рівня ієрархії приймають до себе тільки притоки першого порядку, річки третього порядку утворюються при злитті річок другого порядку й т.д.

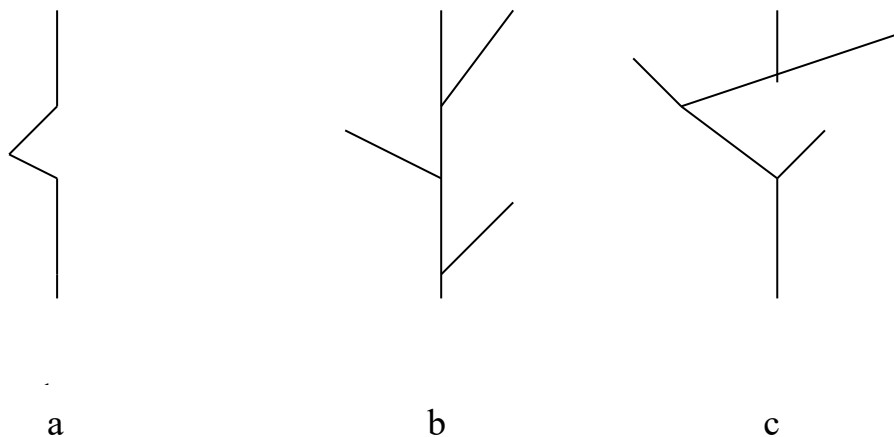


Рисунок 7.1 – Конструювання самоподібної фрактальної річкової мережі із самоподібними індивідуальними потоками

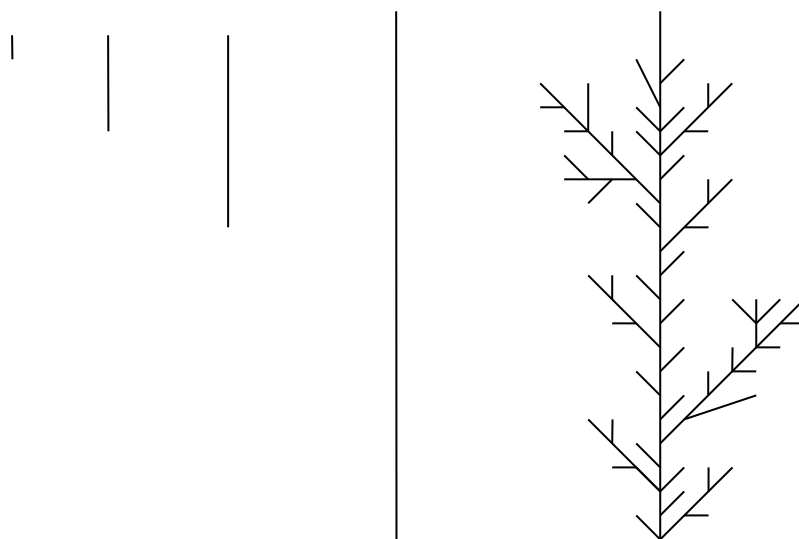


Рисунок 7.2 – Приклад порядкової фрактальної моделі річкової мережі

М. А. Ржаніцин вніс корективи у метод Р. Є. Хортон, щодо можливостей підрахунку річок більш високих порядків.

Розглядаючи річкову мережу як відкриту деревоподібну систему (рис.7.2), що складається з приток різних порядків, і, використовуючи як топологічний параметр порядок водотоку n (порядок може визначатися різними способами), Р.Є. Хортон (1943) сформулював ряд закономірностей, які стали основою сучасної гідрографічної науки.

Перша із закономірностей носить назву закономірності кількості приток: в кожній річковій системі співвідношення між кількістю приток суміжних порядків є величина постійна, тобто:

$$\sigma_0 = \frac{N_{n-1}}{N_n}, \quad (7.16)$$

де σ_0 називається коефіцієнтом біфуркації;

N_{n-1}, N_{n-2} - кількість приток суміжних порядків.

Отже, коефіцієнт біфуркації є, одночасно, константою R_B .

Друга закономірність визначає таке положення: співвідношення між довжинами приток річок суміжних порядків залишається в середньому постійним:

$$\lambda_0 = \frac{L_n}{L_{n-1}}. \quad (7.17)$$

Третя закономірність полягає у тому, що площі водозборів приток суміжних порядків також знаходяться в певному співвідношенні:

$$\varphi_0 = \frac{F_n}{F_{n-1}}, \quad (7.18)$$

де φ_0 - коефіцієнт площі;

F_n та F_{n-1} - площі водозборів приток суміжних порядків n та $n-1$.

Четверта закономірність Хортон встановлює співвідношення між середніми уклонами річок суміжних порядків:

$$I_0 = \frac{I_{n-1}}{I_n}, \quad (7.19)$$

де I_0 - коефіцієнт уклонів приток I_{n-1} і I_n суміжних порядків $n-1$ та n .

Пізніше закономірності, установлені Р.Є. Хортоном, були пояснені фрактальною поведінкою річкової мережі. Вперше фрактальна поведінка річкової мережі була виявлена Хаком (1957) у вигляді степеневі залежності довжини річки від площі та описана Мандельбротом (1977)

$$L_r^{1/d} \sim F^{0.5}, \quad (7.20)$$

де L_r - довжина річки уздовж продольної осі.

Установлені закономірності Р.Є. Хортон увійшли до формул визначення фрактальної розмірності річкової мережі, де

використовуються константи, які одночасно є коефіцієнтами Р.Є. Хортон

$$R_B = \frac{N_{n-1}}{N_n} = \sigma_0; \quad (7.21)$$

$$R_L = \frac{L_n}{L_{n-1}} = \lambda_0; \quad (7.22)$$

$$R_A = \frac{F_n}{F_{n-1}} = \varphi_0; \quad (7.23)$$

$$R_I = \frac{I_{n-1}}{I_n} = I_0, \quad (7.24)$$

де N_n, L_n, I_n, F_n - кількість приток n -го порядку та їх середня довжина, уклон та площа;

$N_{n-1}, L_{n-1}, I_{n-1}, F_{n-1}$ - кількість приток $n-1$ -го порядку та їх середня довжина, уклон та площа.

Константи R_B, R_L, R_A, R_I можуть визначатися за даними про притоки різних суміжних порядків. Фрактальна розмірність розглянутої річкової мережі визначається за формулою як тангенс кута нахилу залежності $\ln R_B = f(\ln R_L)$ або $\ln R_B = f(\ln R_A)$

$$d = \frac{\ln R_B}{\ln R_L}; \quad (7.25)$$

або

$$d = \frac{\ln R_B}{\ln R_L} = 2 \frac{\ln R_B}{\ln R_A}. \quad (7.26)$$

Фрактальна розмірність річкової мережі, визначена на основі ієрархічної моделі за формулами різних авторів може дещо відрізнятися. Справа у тому, що поведінка річкової мережі є наближено самоподібною (самоафінною).

Наступні відношення були використані для кількісної оцінки само афінних властивостей річкової мережі

$$l \sim L'^{\nu_1}; \quad (7.27)$$

$$\omega \sim L'^{\nu_2}, \quad (7.28)$$

де l та ω - характеристики подовжніх та поперечних розмірів мережі;

L' - повна довжина річкової мережі;

ν_1 та ν_2 - показники, що близькі до 2/3 та 1/3, відповідно.

Коли $\nu_1 = \nu_2 = \frac{1}{d}$, річкова мережа характеризується фрактальною розмірністю d . Коли $\nu_1 \neq \nu_2$ - поведінка річкової мережі є самоафінною.

Якщо урахувати, що повна довжина річкової мережі L' зв'язана із площею водозбору, то можна записати, що

$$l \sim F^{\nu_1}; \quad (7.29)$$

$$\omega \sim F^{\nu_2}, \quad (7.30)$$

де F - площа водозбору.

Сума показників $\nu_1 + \nu_2$ визначає так звану лакунарну розмірність $D_G = \frac{2}{\nu_1 + \nu_2}$ (для випадку самоподібності $D_G = D$), а відношення ν_2/ν_1 інтерпретується як експонента Хурста H_0 , тобто один з показників фрактальної поведінки об'єктів, який розраховується при застосуванні стохастичного підходу до описування просторової або часової мінливості досліджуваної характеристики (детальний опис H_0 надається у наступному підрозділі).

З (7.29) та (7.30) походять співвідношення, які характеризують фрактальну геометрію річкової мережі:

$$l \sim F^{\nu_1 / (\nu_2 + \nu_1)} ; \quad (7.31)$$

$$\omega \sim l^{\nu_2 / \nu_1} ; \quad (7.32)$$

$$\omega \sim F^{\nu_2 / (\nu_1 + \nu_2)} ; \quad (7.33)$$

$$L' \sim F^{1 / (\nu_1 + \nu_2)} . \quad (7.34)$$

Добуток $l\omega$ у цих відношеннях розглядається як площа водозбору, тобто

$$F \sim l\omega . \quad (7.35)$$

Відношення (7.31) можна представити як інтерпретацію закону Хака (7.20)

$$L \sim F^\beta , \quad (7.36)$$

де L - довжина головної річки.

Якщо представити довжину головної річки через відрізки l близькі за властивостями до каналів, то можна записати

$$l \sim L^{\nu_3} , \quad (7.37)$$

де ν_3 - експонента, що близька до 1.

Використовуючи (7.31) та (7.37), отримаємо

$$\beta = \nu_1 / (\nu_1 + \nu_2) / \nu_3 . \quad (7.38)$$

Оскільки $\nu_3 = 1$, то

$$\beta = \nu_1 / (\nu_1 + \nu_2) = 1 / (1 + H_0) \quad (7.39)$$

Іноді у гідрології застосовується фрактальна розмірність виду ε

$$L' \sim F^{\varepsilon}, \quad (7.40)$$

$$\text{де} \quad \varepsilon = 1/(\nu_1 + \nu_2) = D_G / 2. \quad (7.41)$$

Для визначення показників фрактальності річкової мережі доцільно використовувати графіки $L = f(F)$ та $L' = f(F)$.

Описані фрактальні властивості установлені шляхом аналізу та побудови статистичних залежностей на основі фактичних даних про гідрографічні характеристики річок та річкових систем. Такий підхід до установлення фрактальних розмірностей носить назву статистичного.

Наприклад, фрактальні параметри річок Молдови такі: $F=142 - 7760 \text{ км}^2$; $\nu_1=0.62$; $\nu_2=0.43$; $D_G=1.90$ та $H=0.69$ (Нікора, 1988).

7.3 Стохастична фрактальність у просторових та часових рядах спостережень

Важливо відзначити, що фрактальні розмірності несуть в собі певну інформацію про просторово-часовий розподіл досліджуваних величин, подібно статистичним параметрам, які широко застосовуються в гідрології.

Якщо часові ряди стаціонарні, то найпростішим способом їх масштабування є стандартний спектральний аналіз, на основі якого отримується енергетичний спектр $E(f)$ - спектральна щільність випадкової функції, яка залежить від частоти спектру f . Для стаціонарних часових рядів за наявності внутрішньорядних кореляційних зв'язків повинна існувати залежність вигляду

$$E(f) \sim f^{-\beta}, \quad (7.42)$$

де f - частота.

Як відомо, автокореляційна функція стоку стаціонарного ряду стоку набуває вигляду

$$r(s) \sim s^{-\gamma}, \quad (7.43)$$

де s - крок в часі.

При цьому показник степеня β для функції нормованої спектральної щільності у (7.42) пов'язаний із показником степеня відповідної автокореляційної функції $r(s)$ наступним співвідношенням

$$\gamma = 1 - \beta; \quad (7.44)$$

При цьому γ та β можуть відігравати роль масштабуючих множників (одним з видів представлення фрактальних властивостей часових рядів).

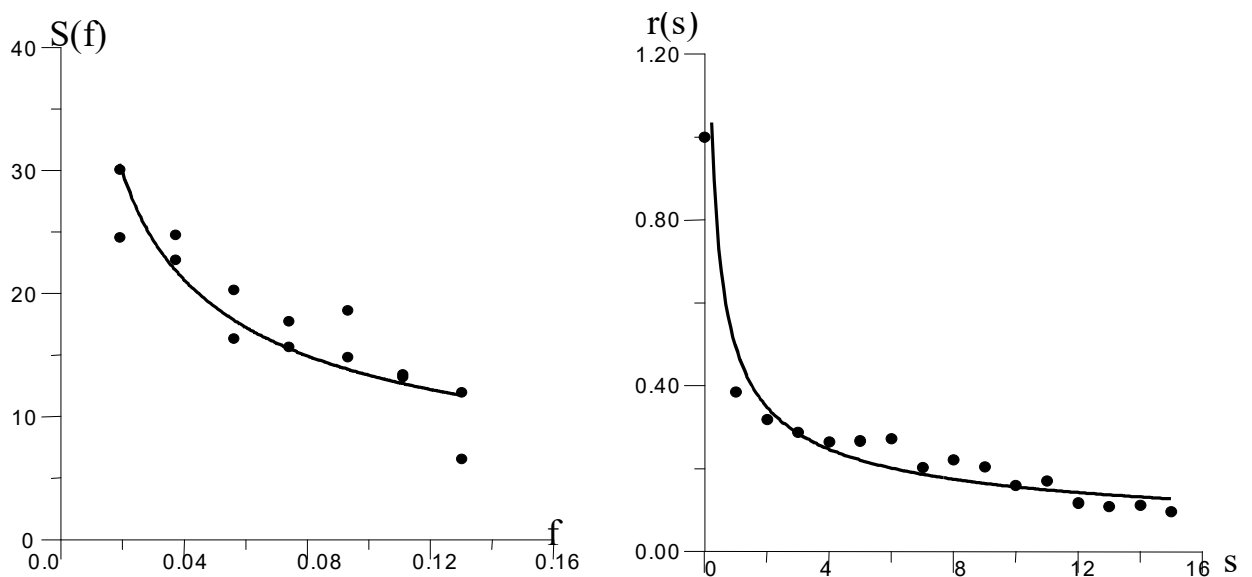


Рисунок 7.3 – Спектральна та автокореляційна функції для ряду річного стока

р.Дерекойка – г.Ялта ($\gamma = 0,498$, $\beta = 0,497$)

Більш ніж півстоліття назад вчений Хурст показав, що в рядах річного стоку різних річок виявляється певна степенева залежність, яка указує на існування властивостей самоподібності в хронологічному ході коливань стоку.

У загальному випадку установлення властивостей самоподібності передбачає наявність степеневі залежності між статистичним моментом $F_q(s)$ порядку q та масштабом s [20,40]

$$F_q(s) = s^{h(q)}, \quad (7.45)$$

де $h(q)$ - масштабуючий степе́нь Рені або фрактальна розмірність, яку по відношенню до часових рядів називають також узагальненим показником Хурста (друге позначення – H).

Якщо $h(q)$ не залежить від q (для усіх q значення $h(q)$ постійні), то ця обставина розглядається як властивість монофрактальності. Для монофрактальних об'єктів $h(q) = H_0$, показник Хурста. Зазвичай вивчається флуктуаційна функція другого порядку (так званий другий статистичний момент) виду

$$F_2(s) = s^H. \quad (7.46)$$

Пошук фрактальної розмірності відбувається шляхом побудови емпіричної залежності $F_2(s)$ від s , де H - степеневий показник кривої, який визначається як тангенс кута нахилу після логарифмування обох осей.

7.4 Способи визначення фрактальних розмірностей за структурною функцією

В сучасній стохастичній гідрології масштабування та пов'язане з ним встановлення фрактальних розмірностей розподілу більшості гідрометеорологічних величин може бути отримано в результаті дослідження часової або просторової варіації величини, яка вивчається в часі або просторі, з кроком s . Під варіацією розуміють зміну досліджуваної характеристики на деякому часовому інтервалі

або заданій відстані. Для характеристики варіації у математичній статистиці найчастіше вживаються такі характеристики як дисперсія, середнє квадратичне відхилення та коефіцієнт варіації. Дисперсія є другим центральним статистичним моментом. Просторова або часова дисперсія випадкової функції має назву **структурної функції** $M(s)$. Квадратний корінь із структурної функції є характеристикою варіації (середнє квадратичне відхилення) й являє собою флуктуаційну функцію другого порядку $F_2(s)$

$$\sqrt{M(s)} = F_2(s) \sim s^H, \quad (7.47)$$

або

$$M(s) \sim s^{2H}, \quad (7.48)$$

де H розглядається як фрактальна розмірність.

В статистичній гідрології структурна функція $M(s)$ застосовувалася при дослідженні стаціонарних ергодичних процесів, реалізації яких і представляють ряди стоку, чия стаціонарність носить локальний характер і зберігається на порівняно невеликих інтервалах зміни аргументу[33]. Структурну функцію визначають як математичне сподівання квадрата різниці перетинів випадкової функції. Якщо реалізація ергодичного випадкового процесу задана в дискретних точках вимірювання, то структурна функція може бути представленою у вигляді

$$M(s) = \frac{1}{(n-s)} \sum_{i=1}^{n-s} (x_i - x_{i+s})^2 = 2[\sigma_x^2 - R(s)], \quad (7.49)$$

де n - число вимірювань;

σ_x^2 -дисперсія;

$R(s)$ -автоковаріаційна функція.

Структурну та автоковаріаційну функції можна нормувати шляхом ділення на величину дисперсії. Тоді між нормованою структурною

та кореляційною (нормованою коваріаційною) функціями виникає відношення виду

$$\frac{M(s)}{2\sigma_x^2} = [1 - r(s)], \quad (7.50)$$

де $r(s)$ - автокореляційна функція.

Для стаціонарних однорідних процесів структурну функцію можна виразити через s та показник $\alpha = H$, наведений у (7.2)

$$M(s) \sim s^{2\alpha}. \quad (7.51)$$

Величину α легко визначити за графіком $\lg M(s) = f(\lg s)$, а фрактальну розмірність d за (7.1).

Ураховуючи існування зв'язку між нормованою структурною та кореляційною функціями (7.50), а також відношення (7.51), можна записати

$$r(s) = 1 - As^{2\alpha}, \quad (7.52)$$

де A - коефіцієнт.

У свою чергу спектральна щільність кривої, яка має структурну функцію типу (7.50), описується відношенням

$$S(k_1) \sim k_1^{-m}, \quad m = 2\alpha + 1, \quad (7.53)$$

де k_1 - хвильове число.

Якщо $1 < m < 3$, то спектральна крива є фрактальною.

Виходячи з (7.53), розмірність d визначається з використанням (7.1) через нахил спектральної кривої у логарифмічних координатах.

7.5 Фрактальність при вивченні турбулентності руслових потоків

Поняття про зовнішні та внутрішні масштаби при вивченні фрактальних властивостей гідрологічних об'єктів можна легко засвоїти, розглядаючи турбулентний рух води у руслах річок [9].

У основі підходу до описування турбулентних структур потоку лежить уявлення А.Н. Колмогорова про те, що турбулентний потік є сукупністю накладених на усереднений рух води збурень (вихорів) різних масштабів. Масштаби великих вихорів мають порядок довжини, яка відповідає шляху перемішування, вони співномірні з лінійними розмірами потоків. Ці вихорі розпадаються на вихорі менших порядків. Найменші вихорі характеризуються малими амплітудами й більшою частотою. Максимальний масштаб вихорів (зовнішній масштаб турбулентності) визначається розмірами потоку, мінімальний (внутрішній) масштаб турбулентності визначається в'язкістю рідини.

Можна застосувати гіпотезу “замороженої турбулентності”, згідно з якою індивідуальні зміни турбулентних збурень, що переносяться потоком, незначні у порівнянні з властивостями усередненого потоку, й, отже, допустимо прийняти, що вихори переміщуються із середньою швидкістю потоку. Тоді зовнішній масштаб турбулентних збурень визначається за формулою

$$l = \bar{u} \tau_0, \quad (7.54)$$

де l - зовнішній масштаб турбулентності;

$\bar{u}$ - усереднена швидкість;

τ_0 - час нульової кореляції (час, при якому кореляційна функція дорівнює нулю).

Для визначення τ_0 використовують нормовану кореляційну функцію, яка розраховується за даними про швидкість потоку

$$R(\tau) = \frac{1}{\sigma_u^2 (n - \tau)} \sum_i^{n-\tau} u_i u_{i+\tau}, \quad (7.55)$$

де n - довжина ряду спостережень;

$u'_i = u_i - \bar{u}$ - пульсація подовжньої складової швидкості у момент i ;

$u'_{i+\tau} = u_{i+\tau} - \bar{u}$ - пульсація подовжньої складової швидкості у момент $i + \tau$.

Часова нормована кореляційна функція оцінює ступінь зв'язку між швидкостями у різні моменти часу. Коли через живий переріз річки проходить єдиний великомасштабний вихор, між виміряними швидкостями буде існувати кореляційний зв'язок. Коли через живий переріз річки буде проходити інший вихор, то зв'язок між попередніми і наступними швидкостями перерветься і значення кореляційної функції буде дорівнювати нулю.

Використовуючи графік $R(\tau) = \psi(\tau)$, можна визначити τ_0 як відстань від початку координат до першого перетину кривої $R(\tau)$ осі τ (рис.7.4). Визначена величина τ_0 називається *радіусом кореляції*.

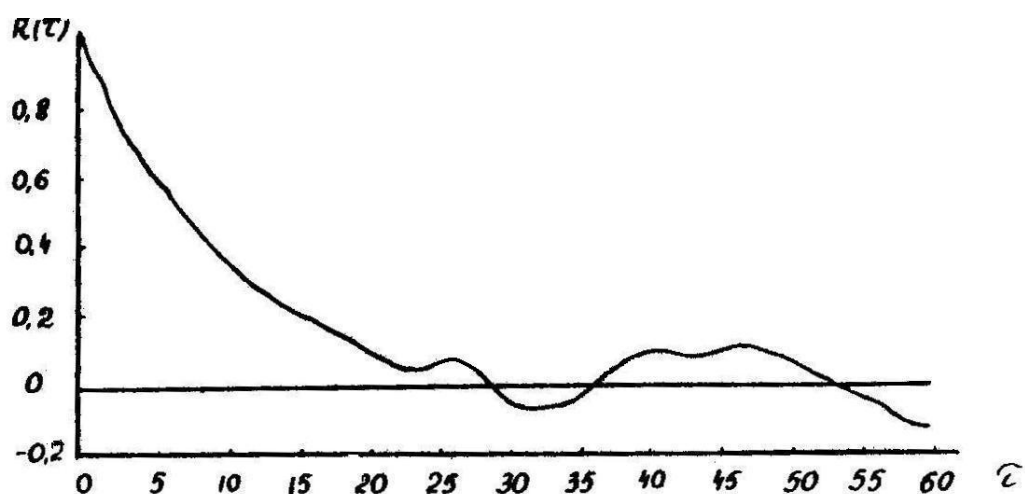


Рисунок – 7.4 Нормована часова кореляційна функція

За даними вимірювань поздовжньої складової швидкості на р.Турунчук були обчислені та побудовані графіки осереднених емпіричних часових структурних функцій, які показували залежність величин структурної функції від інтервалу часу між вимірюваннями для різних глибин (Д.І.Грінвальд,1974). Розрахунки структурних часових функцій виконувались за такою формулою

$$M(\tau) = \frac{1}{(n-s)} \sum_{i=1}^{n-\tau} \left(u'_i - u'_{i+\tau} \right)^2 = 2 \left[\sigma_u^2 - r(\tau) \right]. \quad (7.56)$$

Установлено, що структурна функція зростала до певного часу $\tau_{\max}$ (рис.7.5). При $\tau > \tau_{\max}$ структурна функція мало змінюється, досягаючи стану “насичення”. Зміна структурної функції $\tau < \tau_{\max}$ добре описується степеневим законом вигляду

$$M(\tau) = A \tau^{n_2}, \quad (7.57)$$

де $M(\tau)$ -часова структурна функція;

A - коефіцієнт, який характеризує дисипацію енергії;

n_2 - показник ступеня, який можна розглядати як один з видів представлення фрактальної розмірності.

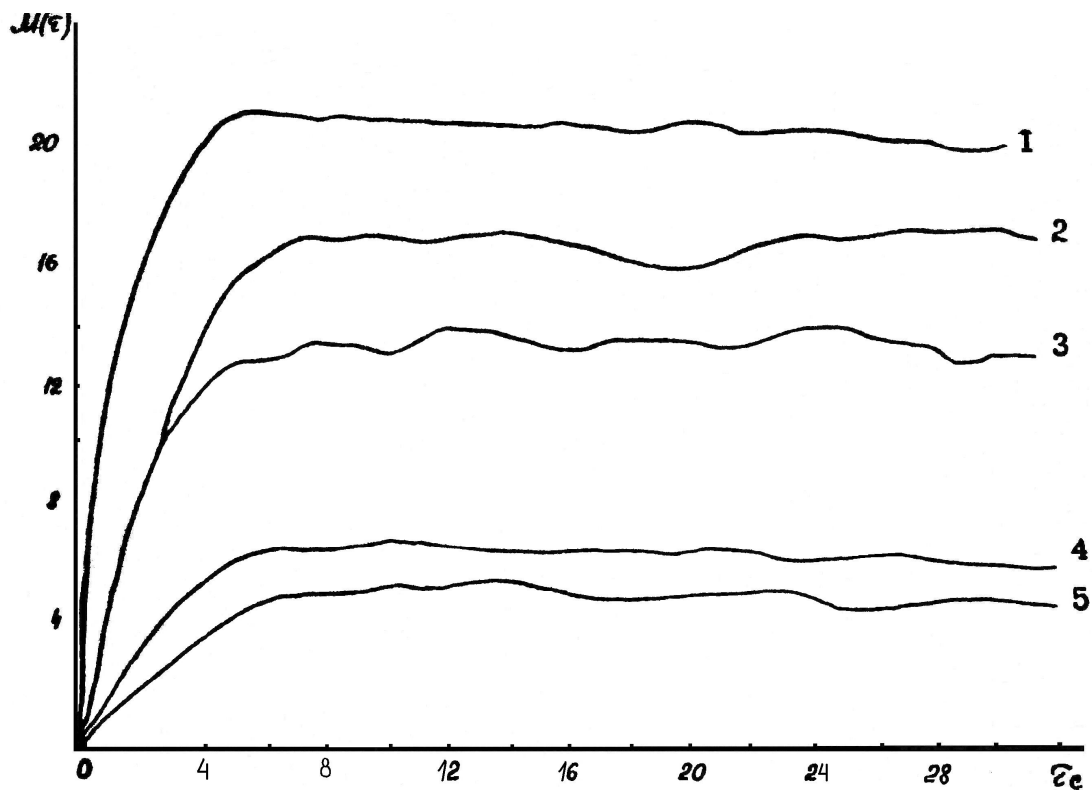


Рисунок – 7.5 Часові структурні функції:
1 – дно; 2 – 0.8H; 3 – 0.6H; 4 – 0.2H; 5 - поверхня

Якщо відомий вид структурної функції, то за (7.51) можна установити показник α . Функція (7.57), представлена у логарифмічних координатах, добре апроксимується прямою (рис.7.6). Показник степеня $n_2 = 2\alpha = 2/3$, Тоді $\alpha = 1/3$. Тоді згідно з (7.1) фрактальна розмірність руслових потоків $d = k - \alpha$ дорівнює $5/3$, якщо розмірність простору $k = 2$.

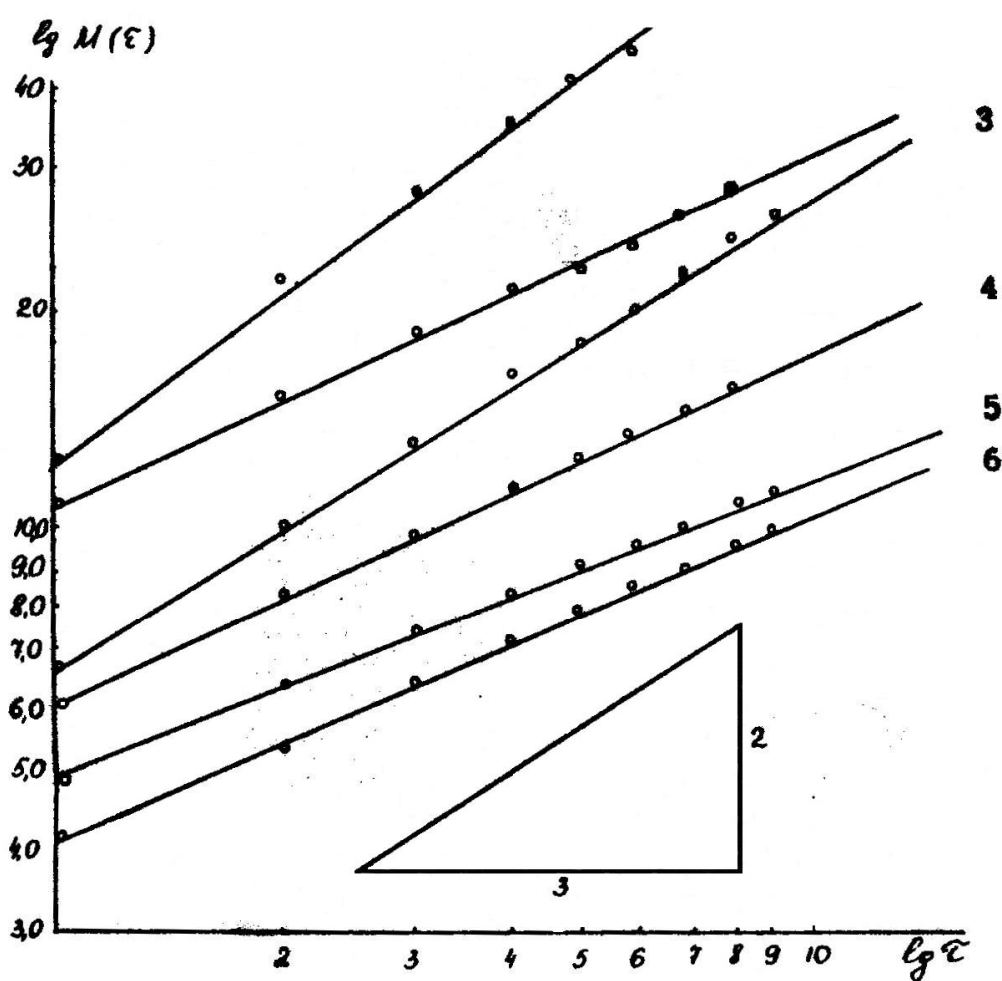


Рисунок 7.6 – Залежність структурної функції від часу у логарифмічних координатах:
1 – дно; 2 – $0.9H$; 3 – $0.8H$; 4 – $0.6H$; 5 – $0.2H$; 6 – поверхня

Структурна функція швидкості потоку може бути не тільки часовою, а і просторовою

$$M_u(\Delta x) = \overline{[u(x) - u(x + \Delta x)]^2}, \quad (7.58)$$

де $u(x)$ - швидкість потоку у точці x ;

$u(x + \Delta x)$ - швидкість потоку у точці $(x + \Delta x)$.

Лінія зверху означає операцію усереднення квадрату різниць швидкостей на відстані Δx .

Завдяки численним дослідженням було доведено, що для руслових потоків виконується універсальний закон $2/3$ ($\alpha=2/3$), установлений Колмогоровим – Обуховим

$$M_u(\Delta x) = A\Delta x^{2/3}. \quad (7.59)$$

А.Г. Бершадським [3] установлено, що для великомасштабних турбулентних структур, фрактальна розмірність d дорівнює $4/3$. Прикладом великомасштабних турбулентних структур можуть бути великомасштабні скупчення матерії в океані, атмосфері, галактиці.

7.7 Метод варіацій Марка та Аронсона, застосований при вивченні фрактальних властивостей рельєфу Землі

Використання структурної функції постановлено в основу методу варіацій Марка та Аронсона [13], розробленого для вивчення наближено самоподібних (самоафінних) об'єктів. Суть методу полягає в дослідженні масштабної поведінки просторової варіації досліджуваної величини. Просторова варіація V може бути представленою структурною функцією виду

$$M(s) = \langle (Z_i - Z_j)^2 \rangle, \quad (7.60)$$

де Z_i, Z_j - значення досліджуваної величини в точках i та j , а трикутні дужки означають усереднювання по всьому ансамблю точок.

Слід зазначити, що формула (7.60) містить у собі таку ж саму інформацію, що і формула (7.58), але записана у іншому вигляді.

Для проведення розрахунків був обраний екваторіальний пояс Землі. Третина поясу охоплює глибоководну область Тихого океану та фрагменти Східно-Тихоокеанського підняття та перетинає Атлантичний та Індійський океани. У межах виділеного поясу була виділена квадратна сітка з кроком 111 км. Було використано 144400 точок, які створили більш 100 млн. пар. Для кожної з пар урахувалася варіація висоти та відстань між точками. Для центру кожної заданої відстані було встановлено осереднене значення $M(s)$ та побудований графік виду $\lg M(s) = f(\lg s)$, та отримане значення $H = \alpha = 0.62$ (рис.7.7).

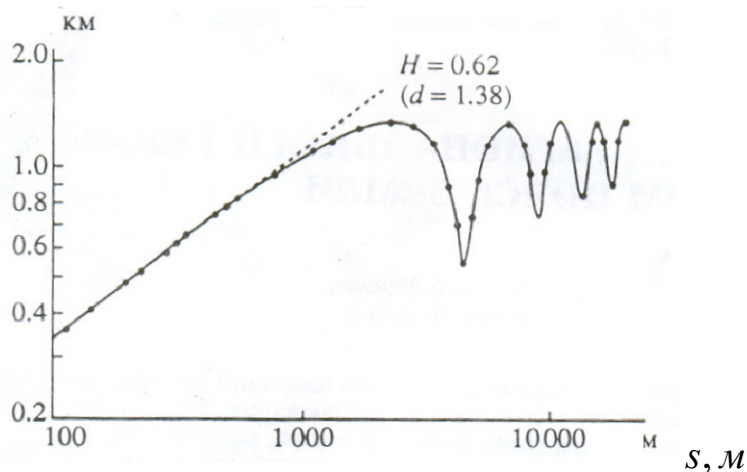


Рисунок 7.7 – Залежність $\sqrt{M(s)}$ від відстані між точками екваторіального поясу Землі

Фрактальна розмірність $d = k - \alpha = 2 - H = 1.38$, через те, що розмірність простору $k = E = 2$, де E - має назву евклідової розмірності. Величина $d = 1.38$ є оцінкою фрактальної розмірності глобального рельєфу.

7.8 Метод варіації Марка та Аронсона при вивченні фрактальних властивостей просторового розподілу статистичних параметрів річного стоку України

Метод варіацій Марка та Аронсона був застосований до виявлення масштабної інваріантності в просторовому розподілі статистичних параметрів річного стоку [19]. Розглянемо просторовий розподіл середніх багаторічних значень річного стоку $\bar{q}$ для правобережної України (79 гідрологічних постів). Просторова варіація оцінювалася як просторова структурна функція, представлена значеннями величини $\bar{q}$ в кожній точці простору

$$M(\Delta L) = \frac{\sum (A_i - A_j)^2}{m - 1} \quad (7.61)$$

де $M(\Delta L)$ - значення просторової структурної функції на відрізку ΔL , віднесене до його центру;

A_i, A_j - досліджувані характеристики в точках i та j (в даному прикладі $A_i = \bar{q}_i$);

m - число пар розглядуваних значень, які потрапили у відрізок ΔL .

В межах України була задана координатна сітка у вигляді квадратів із сторонами 75 км. Відстані між центрами тяжіння виділених водозборів визначалися за допомогою їх умовних координат, які розраховувалися за теоремою Піфагора

$$L_{i,j} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2} \quad (7.62)$$

де L - відстань між об'єктами, яка оцінюється для кожної пари об'єктів i, j ;

x_i, y_i та x_j, y_j - умовні координати просторового положення об'єктів i, j , відповідно.

З метою обчислення просторової структурної функції були задані сегменти, що не перекриваються (градації) ΔL . На основі

матриці відстаней вибиралися пари водозборів, які потрапляють в задану градацію ΔL . Для кожної з пар водозборів визначалася різниця $(A_i - A_j)$ й для кожної градації – відповідне значення структурної функції, розрахованої за (7.61). Результати розрахунків представлені у виді графіків (рис.7.8).

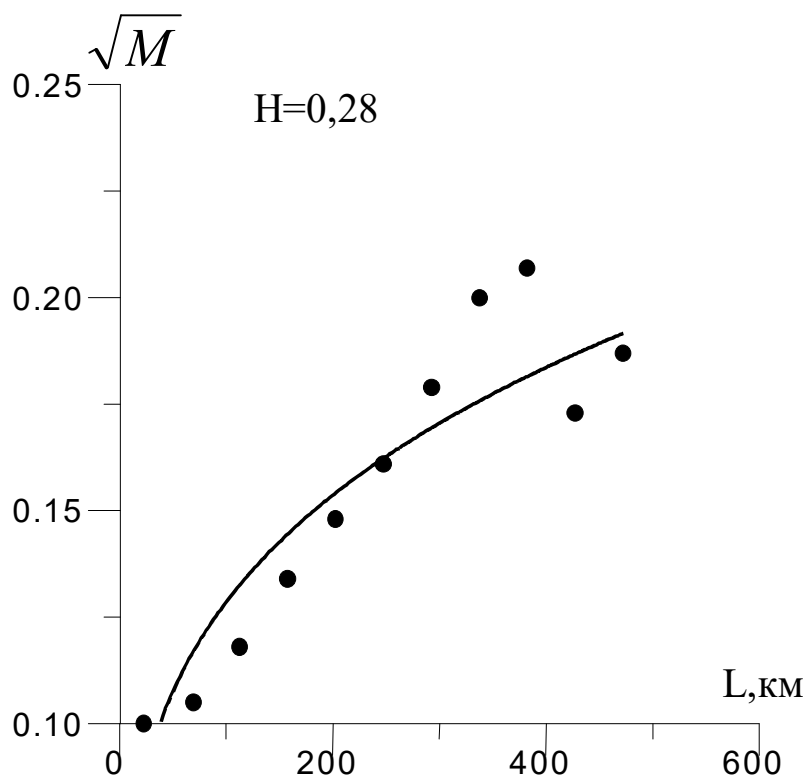


Рисунок 7.8 – Залежність значень флуктуаційної функції другого порядку від відстаней між центрами тяжіння водозборів

З рисунка видно, що при $L \leq 400 \text{ км}$ залежність $\sqrt{M(s)}$ від відстані носить степеневий характер з показником ступеня, який визначався на основі подвійного логарифмування осей. При $L \leq 400 \text{ км}$, шукане значення $H = \alpha = 0,28$. Тоді, згідно із (7.1), фрактальна розмірність дорівнює $d = k - \alpha = 2,00 - 0,28 = 1,72$. При $L > 400 \text{ км}$ структурна функція $M(s)$ й відповідна їй флуктуаційна функція $F_2(s) = \sqrt{M(s)}$ досягають стану насичення, що

розглядається як ознака відсутності кореляційних зв'язків між статистичними параметрами.

Аналогічним чином були установлені властивості інваріантності у просторовому розподілі коефіцієнта варіації C_v річного стоку, для якого $H = \alpha$ дорівнює 0,37, а фрактальна розмірність $d = 1,63$. На відміну від середньоарифметичного значення просторова скорельованість значень C_v простежується тільки на відстані $L \leq 200 \text{ км}$.

7.9 Фрактальні властивості рядів річного стоку

Другий статистичний момент (флуктуаційна функція) рядів річного стоку визначається за формулою

$$\sqrt{M(s)} = F_2(s) = \left\{ \frac{1}{N_s} \sum_{v=1}^{N_s} [z_{vs} - z_{v-1,s}]^2 \right\}^{\frac{1}{2}} \quad (7.63)$$

де v - порядковий номер виділених відрізків, що не перекриваються, розміром s ;

$N_s = \text{int}(\frac{N}{s})$ - число значень досліджуваної величини, які потрапляють в кожний відрізок розміром s ;
різниця $z_{vs} - z_{v-1,s}$ є приростом z_n на кожному відрізку завдовжки s .

Величина z_n є функцією послідовного прирощування відхилень значень річного стоку від середнього $\bar{x}$ і має назву “профіль стоку”

$$z_n = \sum_{i=1}^{i=n} \varphi_i \quad n = 1, 2, \dots, N, \quad (7.64)$$

де $\varphi_i = x_i - \bar{x}$ - центрована величина стоку.

Профіль стоку має такий самий фізичний зміст, що і різницева інтегральна крива, на основі якої можна зробити висновок про зміну фаз водності у коливаннях стоку. Якщо переважають позитивні

відхилення величин річного стоку від середнього арифметичного значення, то функція z_n буде зростати, коли ж переважають негативні відхилення, то функція z_n буде убувати.

Флуктуаційна функція, яка використовується у фрактальному аналізі часових рядів, по суті, виконує розбиття множини N на підмножини N_s . На основі цього розкладання робляться висновки про існування ознак масштабної самоподібності. З метою встановлення таких ознак розглядаються залежності узагальненої функції $F_2(s)$ від s , яка має вигляд степеневі функції [19], схожої на приведену на рис.7.4.

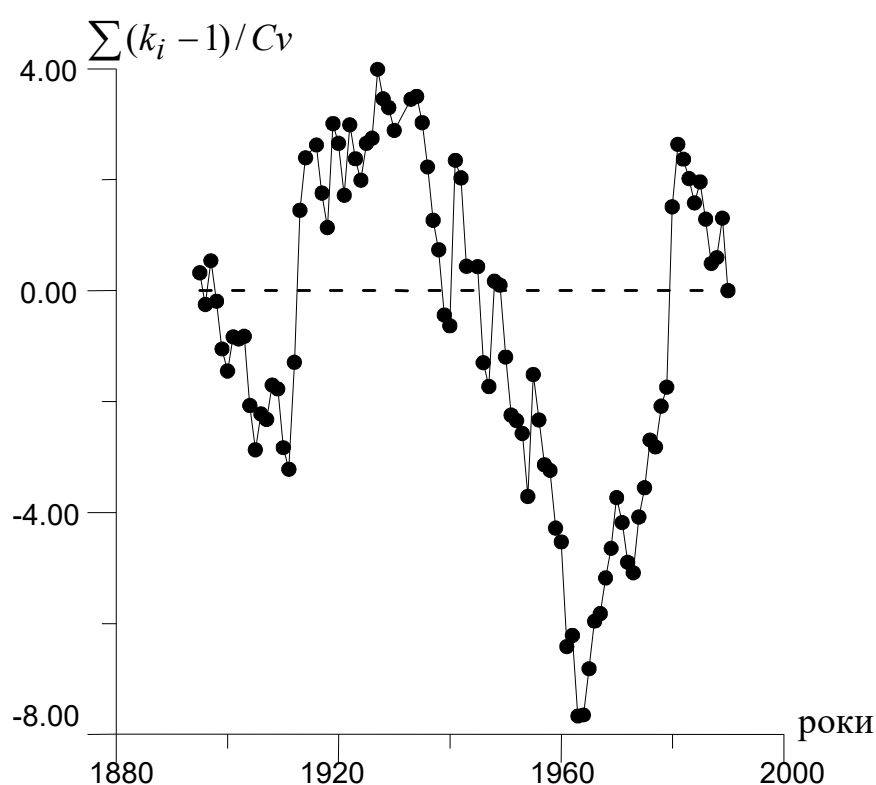


Рис. 7.9 –Різницева інтегральна крива річного стоку
р.Дністер – м. Галич

Установлено, що експонента γ (7.43) для стаціонарних ергодичних рядів річного стоку змінюється від 0,1 до 0,9. На великих часових масштабах величина $\alpha = H$ менша 1. Експонента Хурста H річного стоку змінюється від 0,55 до 0,95, фрактальна розмірність d змінюється, відповідно від 1,45 до 1,05. Середня величина показника Хурста становить 0,75. Якщо $H=1$, це є

свідченням самоподібності. Якщо $H < 1$, то мова йде про “самоафінність”. Величина H пов’язана з характеристиками автокореляційних та спектральних функцій наступним рівнянням

$$H = 1 - \gamma/2 = (1 + \beta)/2. \quad (7.65)$$

Таким чином, можна зробити висновок, що районування за особливостями коливань стоку рядів може відбуватися на підставі установлених фрактальних розмірностей, подібно до того, як відбувається районування за коефіцієнтами автокореляції.

7.10 Приклади визначення фрактальної розмірності річкової мережі факторного аналізу у гідрологічних розрахунках

Завдання. Визначити фрактальну розмірність d за моделлю ієрархічної (порядкової) річкової структури.

Вихідна інформація. Розглядається річкова мережа р.Дністер. Порядок кожного притоку визначався наближеним методом Н.О.Ржаніцина (табл.7.1).

Таблиця 7.1 - Довжина річок та їх порядок за Н.О. Ржаніциним

Довжина річки L_n , км	Порядок річки n	Довжина річки L_n , км	Порядок річки n	Довжина річки L_n , км	Порядок річки n
< 1	1	11 – 24	6	266 – 510	11
1,1 - 2,0	2	25 – 44	7	511 – 880	12
2,1 - 3,5	3	45 – 80	8	881 – 1570	13
3,6 - 6,5	4	81 – 147	9	1571 – 2800	14
6,6 - 10	5	148 - 265	10	2801 - 4620	15

Розв’язок задачі. Фрактальна розмірність річкової мережі за ієрархічним методом визначається на основі співвідношення (7.26)

$$d = \frac{\ln R_B}{\ln R_L} = 2 \frac{\ln R_B}{\ln R_A},$$

де R_B, R_L, R_A топологічні коефіцієнти Хортон, які визначаються за (7.21), (7.22), (7.23).

Спочатку розраховуються значення R_B, R_L, R_A для приток суміжних порядків (табл.7.2, табл.7.3), після чого будуються залежності вигляду $\ln R_B = f(\ln R_L)$ - (рис.7.10) та $\ln R_B = f(\ln R_A)$ - (рис 7.11).

Таблиця 7.2 - Вихідні дані для розрахунку фрактальної розмірності річкової мережі р. Дністер

Порядок притоки	Кількість приток N	Довжина Ln, км	Площа, км ²
IV	1	3,6	159
V	30	9,95	53,6
VI	234	16,2	75,2
VII	67	32,4	192
VIII	32	58,6	609
IX	14	100,9	1181
X	7	209,7	3256

Таблиця 7.3 – Визначення топологічних коефіцієнтів р.Дністер

Порядок притоку	RL	Ln RL	RB	Ln RB	RA	Ln RA
V	1,63	0,489	0,362	-1,020	1,40	0,336
VI	2	0,693	0,614	-0,488	2,56	0,940
VII	1,81	0,593	0,5	-0,693	3,16	1,15
VIII	1,72	0,542	0,553	-0,592	1,94	0,663
IX	2,08	0,732	0,581	-0,543	2,76	1,02

Тангенс кута нахилу лінії на графіках $\ln R_B = f(\ln R_L)$ та $\ln R_B = f(\ln R_A)$ і є шуканою фрактальною розмірністю d .

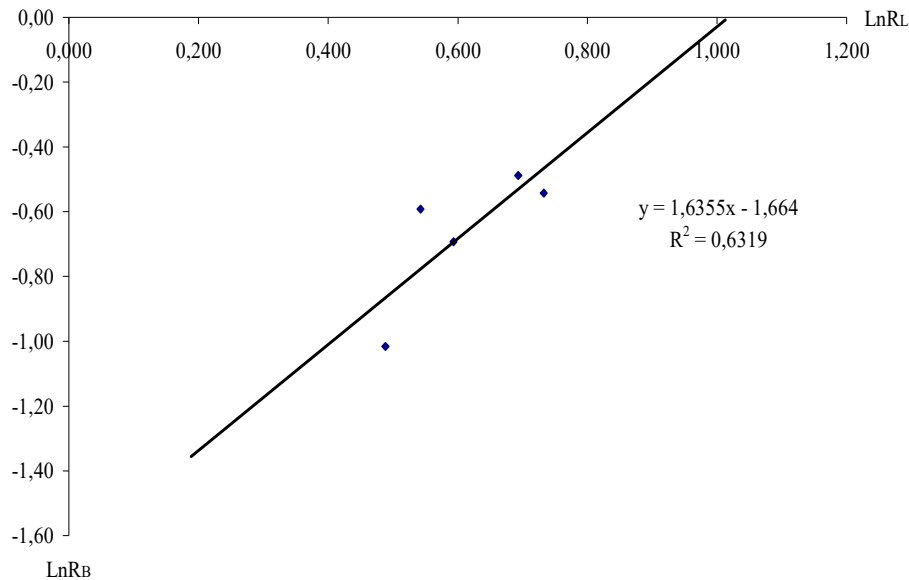


Рисунок 7.10 - Визначення фрактальної розмірності р.Дністер на основі топологічних коефіцієнтів R_B, R_L

Мандельброт, який вперше ввів поняття фракталів у науку, припустив, що фрактальна розмірність для річкової мережі має наближатися до 2. Як видно з рисунків, фрактальна розмірність у першому випадку дорівнює 1,64, а у другому – 1,44.

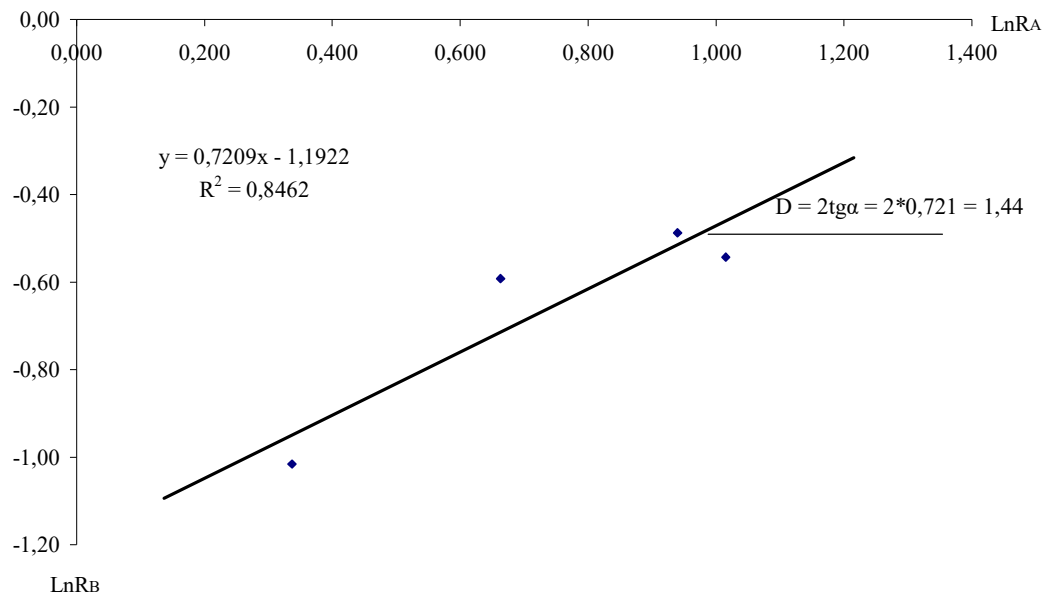


Рисунок 7.11 - Визначення фрактальної розмірності р.Дністер на основі топологічних коефіцієнтів R_A, R_L

Контрольні запитання

- 1.Топологічна розмірність кривої, поверхні та шару.
- 2.Визначення розмірності фрактального об'єкта шляхом підрахунку квадратів за Мандельбротом.
- 3.Чому дорівнює коефіцієнт подібності, коли вихідна множина розбивається на N множин в n разів менше вихідної.
- 4.Визначити поняття зовнішній та внутрішній масштаби фрактальності.
- 5.Записати вираз для визначення структурної функції.
- 6.Записати вираз для визначення флуктуаційної функції.
- 7.Як змінюється експонента Хурста для рядів річного стоку.

8 КОРЕЛЯЦІЙНІ ФУНКЦІЇ

8.1 Поняття про випадкову функцію

На практиці доводиться оперувати величинами, які змінюються у процесі випробувань. Такі величини отримали назву випадкових функцій. Вивченням подібних випадкових явищ, в яких випадковість набуває форми процесу, займається спеціалізований розділ теорії ймовірностей – теорія випадкових функцій або стохастичних процесів.

Випадкова функція – це функція, значення якої встановлюються за допомогою випробувань і можуть бути різними в залежності від ходу випробувань. **Випадковою функцією називається функція, яка в разі випробувань може набрати того чи іншого конкретного вигляду, наперед невідомо, якого саме. Конкретний вигляд, якого набуває випадкова функція в результаті випробувань, називається реалізацією випадкової функції.**

Зустрічаються випадкові функції, які залежать не від одного аргументу, а від декількох. У гідрологічних розрахунках найчастіше розглядаються функції тільки одного аргументу. Якщо аргументом випадкової функції є час, то для її позначення використовується термін «випадковий процес».

Випадкову функцію вивчають аналітичним способом, заснованим на визначенні багатовимірного закону її розподілу, і статистичним, заснованим на визначенні тільки числових характеристик такої функції. Перший спосіб застосовують у теоретичних дослідженнях. Другий спосіб широко використовують для вирішення різних прикладних задач у теорії випадкових функцій.

Розглянемо випадкову функцію $X(t)$, над якою проведено n незалежних випробувань і отримано n реалізацій. Кожна реалізація є звичайною, тобто не випадковою функцією.

Сімейство реалізацій функції, що відповідає ймовірнісному процесу з безперервною зміною функції в залежності від аргументу, наведене на рис.8.1.

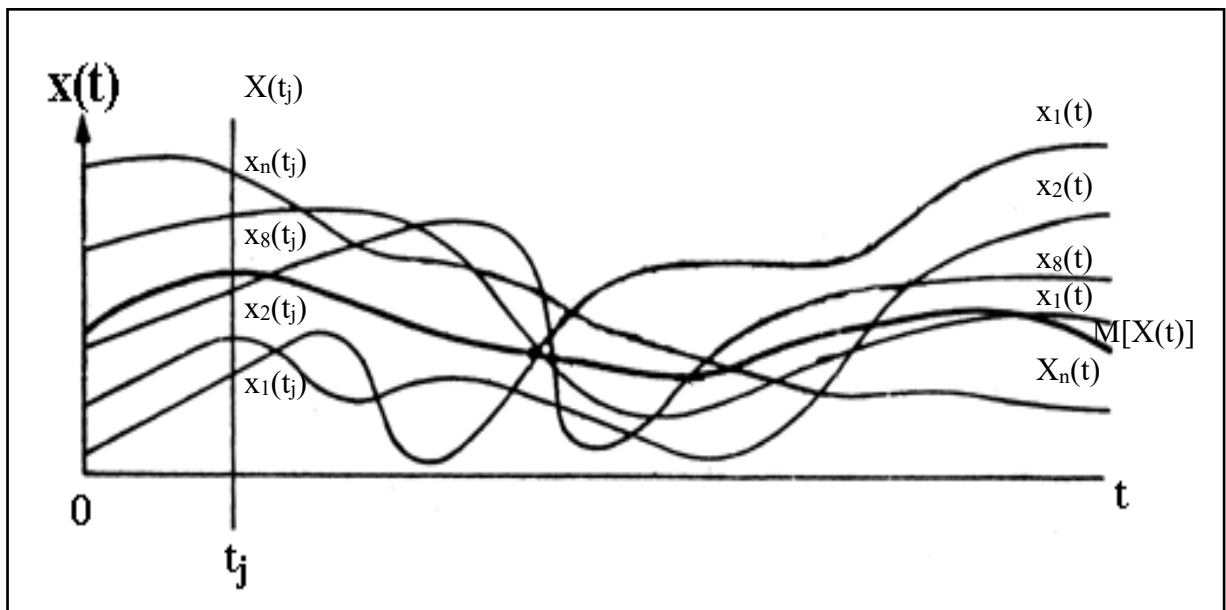


Рисунок 8.1 - Сімейство реалізацій випадкової функції

Реалізації випадкової функції можна отримати за допомогою самописного приладу, осцилограми запису змін процесу в часі, кінозйомки, періодичної фотозйомки та інш.

8.2 Закон розподілу випадкової функції

Закон розподілу однієї випадкової величини є функція одного аргументу, закон розподілу системи двох величин – функція двох аргументів і т.п. Однак, для практичного використання функції розподілу багатьох аргументів застосовувати незручно. Зазвичай відмовляються від використання законів розподілу і розглядають тільки числові характеристики випадкових функцій.

Закон розподілу випадкової функції може представляти собою функцію безлічі аргументів, яку чисто формально можна записати в символічній формі.

Якщо зафіксувати деяке значення аргументу t_j , то при заданому значенні аргументу t_j (рис.8.1) випадкова функція перетворюється у випадкову величину $X(t_j)$, яка називається перетином випадкової функції, що відповідає моменту t_j . Значеннями цієї випадкової величини будуть значення $x_1(t_j), x_2(t_j), \dots, x_n(t_j)$, де n - кількість реалізацій.

Таким чином, випадкова функція суміщує у собі риси випадкової величини і функції. Якщо зафіксувати значення аргументу, то отримуємо випадкову величину. В результаті кожного випробування випадкова функція, у свою чергу, перетворюється у звичайну не випадкову функцію.

Розглянемо випадкову величину $X(t)$ - перетин випадкової функції в момент часу t . Ця випадкова величина має закон розподілу, що у загальному випадку залежить від t і позначається як $f(x, t)$. Функція $f(x, t)$ називається одновимірним законом розподілу випадкової функції $X(t)$. Функція $f(x, t)$ не є повною вичерпною характеристикою випадкової функції $X(t)$. Дійсно, ця функція характеризує тільки закон розподілу $X(t)$ для даного, хоча й довільного t ; вона не відповідає на запитання про взаємозв'язок випадкових величин $X(t)$ при різних t . З цього погляду більш повною характеристикою випадкової функції $X(t)$ є так званий двовимірний закон розподілу

$$f(x_1, x_2; t_1, t_2). \quad (8.1)$$

Це закон розподілу системи двох випадкових величин $X(t_1), X(t_2)$, тобто закон розподілу для двох довільно узятих перетинів випадкової функції $X(t)$. Однак і ця характеристика не є вичерпною, ще більш повним описом випадкової функції був би тривимірний закон:

$$f(x_1, x_2, x_3; t_1, t_2, t_3). \quad (8.2)$$

Теоретично можна необмежено збільшувати число аргументів, але оперувати ними вкрай незручно. У зв'язку з цим на практиці розглядають не закони розподілу, а найпростіші характеристики випадкових функцій, аналогічні числовим характеристикам випадкових величин.

8.8 Характеристики випадкових функцій та їх визначення

Основними числовими характеристиками випадкового процесу є: математичне сподівання, що визначає таку функцію, навколо якої групуються реалізації випадкової функції; дисперсія, що показує ступінь розсіювання цих реалізацій відносно математичного сподівання, кореляційна функція, що визначає

внутрішню структуру випадкового процесу і характеризує силу взаємного зв'язку між реалізаціями випадкової функції.

На відміну від числових характеристик випадкових величин, які представляють собою числа, характеристики випадкових функцій представляють собою у загальному випадку не числа, а функції.

Математичне сподівання випадкової величини визначається в такий спосіб. Розглянемо перетин випадкової функції $X(t)$ при фіксованому t . У цьому перетині ми маємо звичайну випадкову величину; визначимо її математичне сподівання. Очевидно, у загальному випадку воно залежить від t , тобто являє собою деяку функцію t :

$$m_x(t) = M[X(t)]. \quad (8.8)$$

Таким чином, математичним сподіванням випадкової функції $X(t)$ називається не випадкова функція $m_x(t)$, яка при кожному значенні аргументу t дорівнює математичному сподіванню відповідного перетину випадкової функції. На рис.8.1 більш тонкими лініями показані реалізації випадкової функції, жирною лінією – її математичне сподівання.

Дисперсією випадкової функції $X(t)$ називається не випадкова функція $D_x(t)$, значення якої для кожного t дорівнюють дисперсії відповідного перетину випадкової функції

$$D_x(t) = D[X(t)]. \quad (8.4)$$

Дисперсія $D_x(t)$ є невід'ємною функцією. Добуваючи з неї квадратний корінь, одержимо функцію $\sigma_x(t)$ - середнє квадратичне відхилення випадкової функції

$$\sigma_x(t) = \sigma[X(t)] = \sqrt{D[X(t)]}. \quad (8.5)$$

Досить часто користуються центрованою характеристикою випадкової функції

$$\delta[X(t)] = X(t) - M[X(t)] \quad (8.6)$$

Очевидно, що

$$M\delta[X(t)] = 0. \quad (8.7)$$

Нормована випадкова функція представляється у вигляді

$$X_0(t) = \delta[X(t)] / \sigma[X(t)], \quad (8.8)$$

для якої справедливі рівняння

$$M[X_0(t)] = 0; \quad D[X_0(t)] = 1. \quad (8.9)$$

Дисперсія характеризує розсіювання реалізацій випадкової функції відносно функції математичного сподівання. Не виключена можливість однакої дисперсії для всіх значень аргументу t , тобто $D[X(t)] = D(X)$. Якщо така дисперсія дорівнює нулю, то можна вважати, що випадкова функція $X(t) = M[X(t)]$ має ймовірність появи, яка дорівнює одиниці.

Бувають випадки, коли математичні сподівання випадкових функцій однакові, а їхні дисперсії різні. Наприклад, на рис. 8.2 показані реалізації випадкових функцій $X(t), Y(t), Z(t)$, що мають однакові математичні сподівання, але різні дисперсії.

Математичне сподівання і дисперсія визначають тільки смугу можливих реалізацій випадкової функції, але не поведінку реалізацій усередині такої смуги.

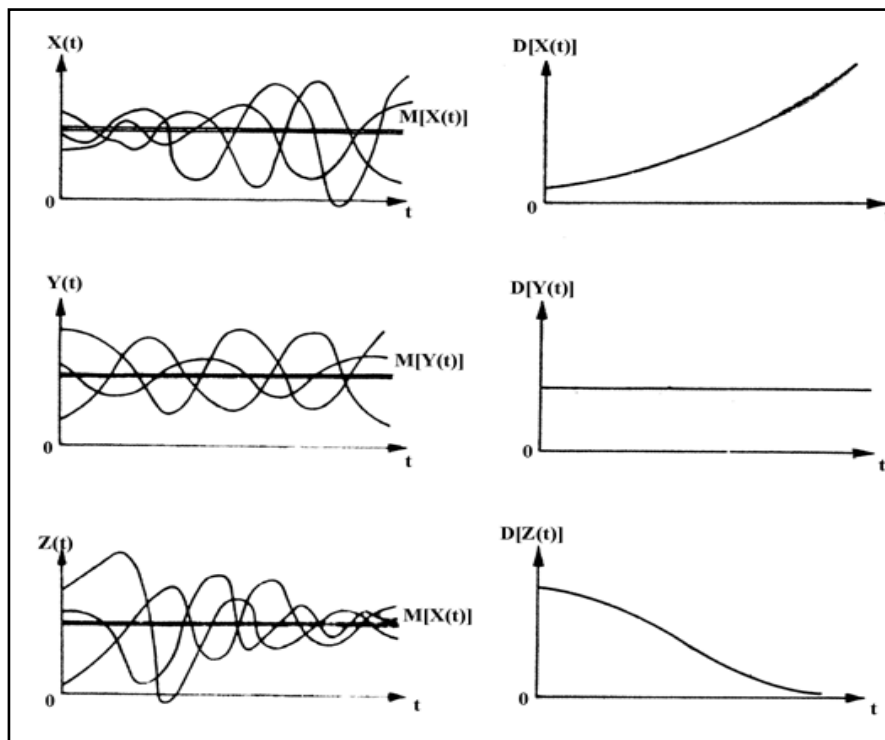


Рисунок 8.2 - Зміна дисперсії у часі для різних випадкових процесів

На рис. 8.3 наведені сімейства реалізацій (а) випадкових функцій, скедастичні криві (б), що показують зміну дисперсій, і кореляційні еліпси (в), які характеризують взаємний зв'язок випадкових ординат по перетинах t_1 і t_2 . Зв'язок між випадковими величинами $X(t_i)$ і $X(t_j)$ може бути виражений за допомогою кореляційного моменту чи коефіцієнта кореляції. Ступінь залежності між перетинами випадкової функції у різних аргументах t , характеризується коваріаційною або кореляційною функціями.

Коваріаційним моментом випадкової функції називається не випадкова функція двох аргументів K_x , яка для кожної пари перетинів випадкової функції відповідних аргументів t_i і t_j дорівнює кореляційному моменту

$$K_x = K[X(t_i), X(t_j)] = M\{\delta[X(t_i)] \cdot \delta[X(t_j)]\}. \quad (8.10)$$

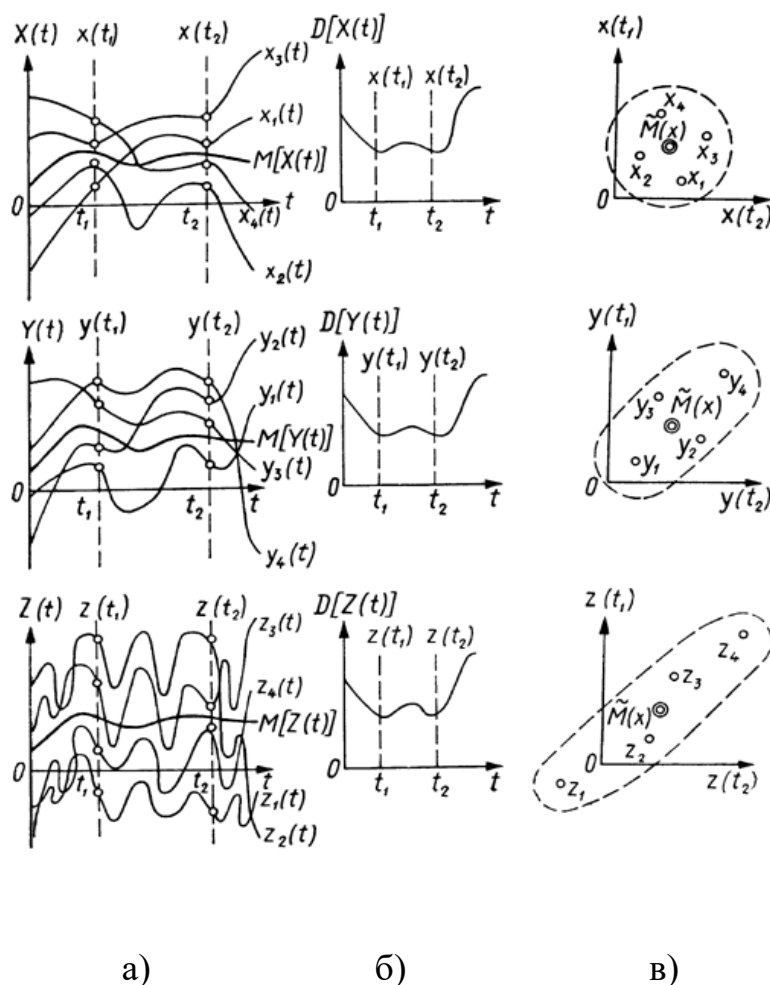


Рисунок 8.3 - Випадкові функції, їх дисперсії та еліпси розсіювання

Розглянемо два випадкових процеси $X_1(t)$ і $X_2(t)$ з однаковими математичними сподіваннями та дисперсіями (рис. 8.4, 8.5).

Коваріаційна функція випадкової функції $X_1(t)$ повільно убуває в міру збільшення проміжку $t = t'$. Навпаки, коваріаційна функція випадкової функції $X_2(t)$ убуває швидко. Ці особливості випадкових процесів характеризуються саме коваріаційними моментами.

На практиці також використовують кореляційну функцію, яка являє собою нормований кореляційний момент

$$r_x = r[X(t_i), X(t_j)] = \frac{K[X(t_i), X(t_j)]}{\sigma(t_i) \cdot \sigma(t_j)}. \quad (8.11)$$

Кореляційну функцію випадкового процесу часто називають **автокореляційною**, що означає кореляцію процесу із самим собою.

При $t_i = t_j = t$, одержимо

$$K[X(t), X(t)] = M[\delta^2(t)] = D_x(t), \quad (8.12)$$

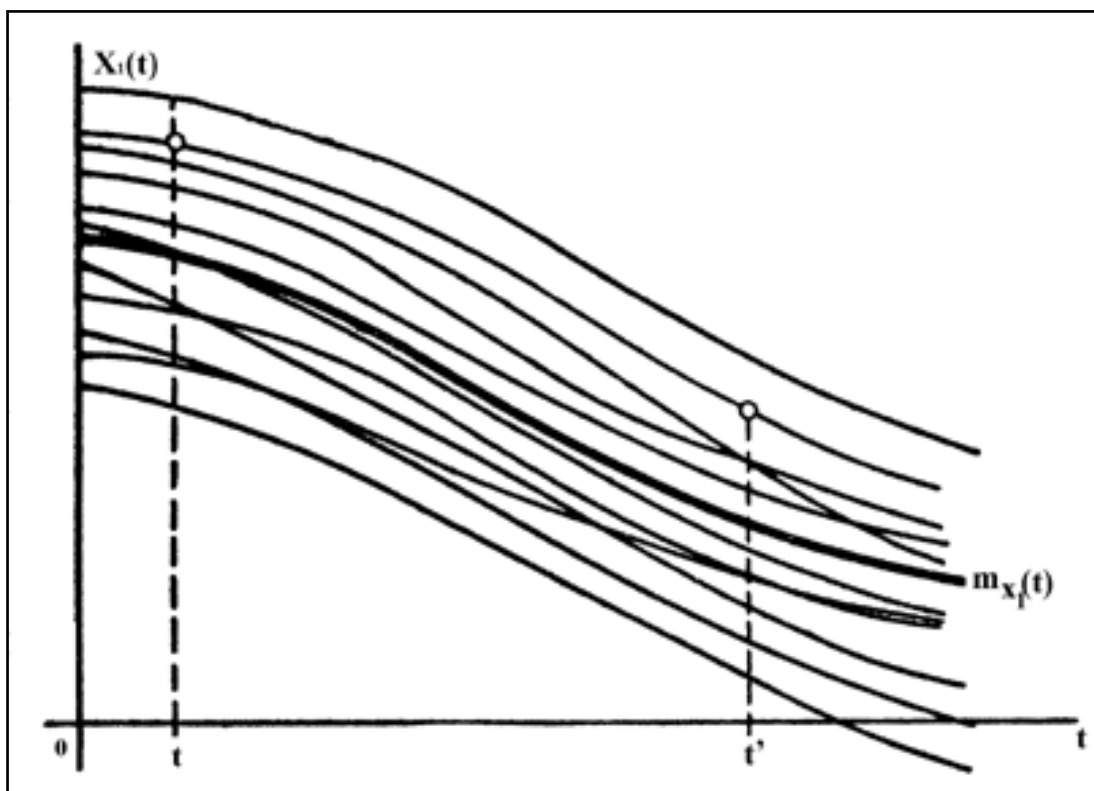


Рисунок 8.4 - Реалізації випадкової функції $X_1(t)$

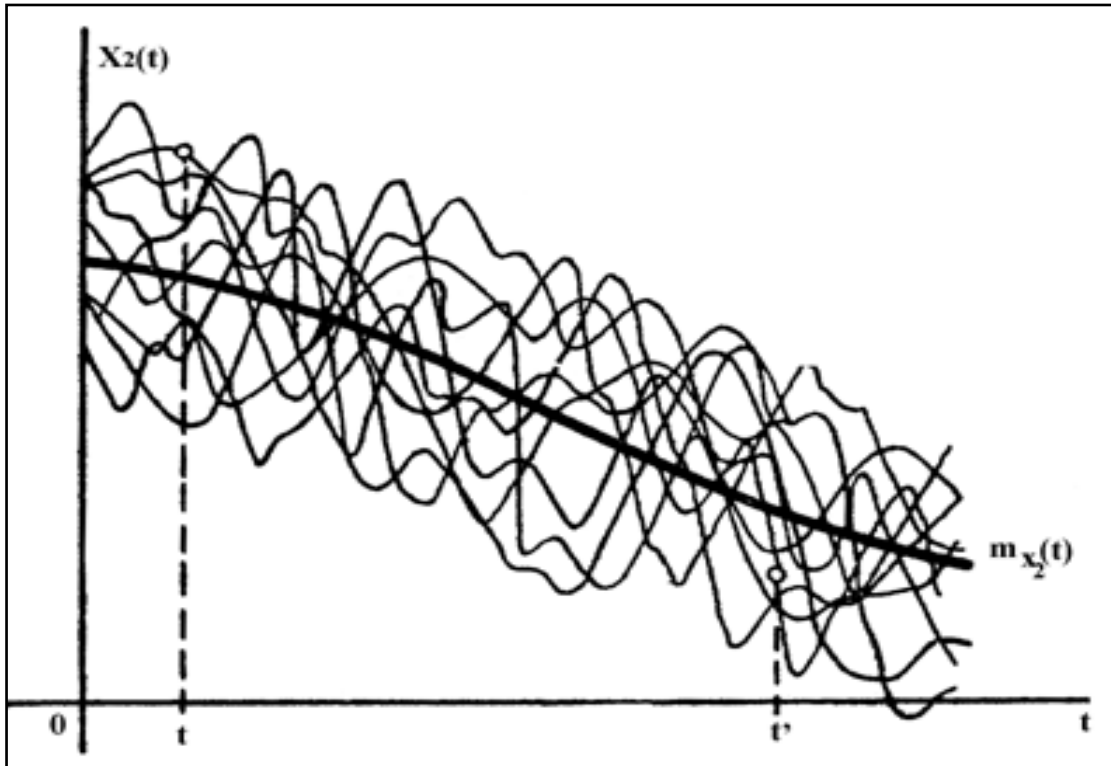


Рисунок 8.5 - Реалізації випадкової функції $X_2(t)$

тобто при $t_i = t_j$ коваріаційна функція стає дисперсією випадкової функції, а кореляційна функція дорівнює 1.

Таким чином, при $t_i = t_j$ необхідність в дисперсії, як в окремій характеристиці випадкової функції, відпадає: як основною характеристикою випадкової функції достатньо розглядати її математичне сподівання і кореляційну функцію.

Коваріаційний момент двох випадкових величин $X(t)$ і $X(t')$ не залежить від послідовності, при якій ці величини розглядаються, отже, коваріаційна функція симетрична відносно своїх аргументів, тобто не змінюється при переміні аргументів місцями

$$K_x(t, t') = K_x(t', t). \quad (8.14)$$

Інтервал $(t_i - t_j)$ називають «запізнюванням» чи «зсувом».

Очевидно, що значення коваріаційної і автокореляційної функцій залежать від інтервалу між t_i і t_j . Вони зменшуються із збільшенням цього інтервалу. При одних і тих же математичних сподіваннях і дисперсіях дві випадкові функції можуть мати різні автокореляційні функції. Кореляційна функція може бути залежною не від окремих значень t_i і t_j , а від різностей $t_i - t_j$. Якщо при

одному із значень t_i чи t_j випадкова функція $X(t)$ стає не випадковою величиною, то $K[X(t_i), X(t_j)] = 0$ при будь-якому значенні іншого аргументу.

Коваріаційна (кореляційна) функція симетрична щодо своїх аргументів. Вона зображується в тривимірній системі прямокутних координат t_i, t_j і $K[X(t_i), X(t_j)]$ у вигляді поверхні, симетричної щодо вертикальної площини Q , яка проходить через бісектрису кута $t_i O t_j$ (рис.8.6). Ця властивість коваріаційної функції відповідає властивості коваріаційної матриці системи випадкових величин (табл.8.1), які приблизно виражають випадкову функцію. По головній діагоналі цієї матриці розташовуються дисперсії. Симетрично щодо неї розміщуються кореляційні моменти (коваріації), які задовольняють умову $K[X(t_i), X(t_j)] = K[X(t_j), X(t_i)]$. У прикладному варіанті теорії випадкових функцій оперують числовими характеристиками, визначеними за спостереженими даними (з випробувань).

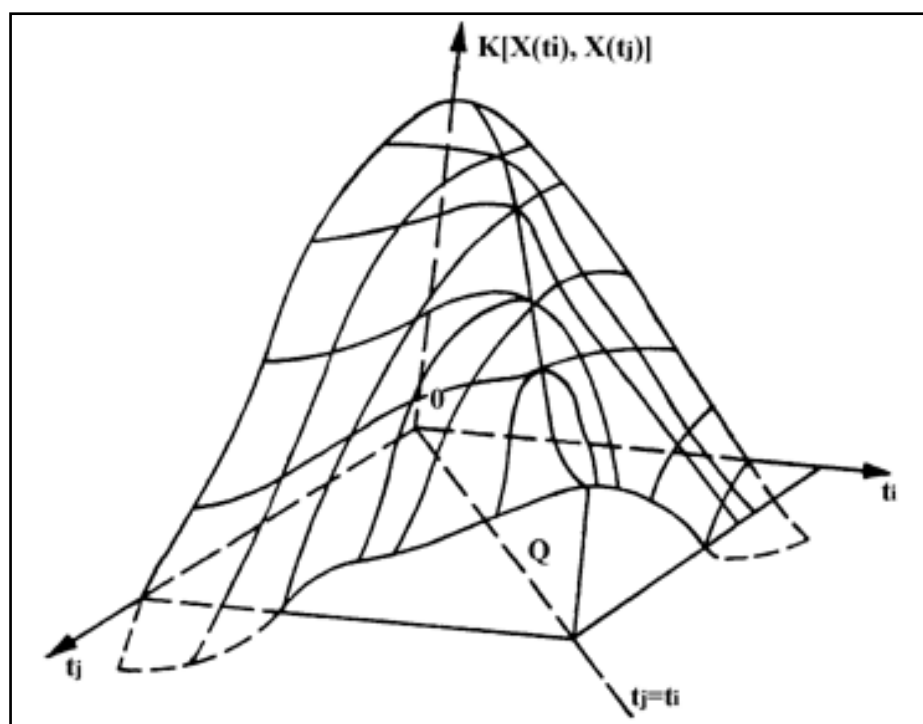


Рис. 8.6 - Коваріаційна функція у вигляді поверхні

Таблиця 8.1 – Коваріаційна матриця системи випадкових величин

t_i	t_j					
	t_1	t_2	...	t_j	...	t_n
t_1	D_1	K_{12}	...	K_{1j}	...	K_{1n}
t_2	K_{21}	D_2	...	K_{2j}	...	K_{2n}
...	...	...	...	...	...	...
t_i	K_{i1}	K_{i2}	...	D_j	...	K_{in}
...	...	...	...	...	...	...
t_n	K_{n1}	K_{n2}	...	K_{nj}	...	D_n

Нехай над випадковою функцією $X(t)$ зроблено n незалежних випробувань (спостережень), у результаті отримано n реалізацій випадкової функції. Якщо число реалізацій дорівнює m , то для кожного перетину t_i можна обчислити емпіричні (статистичні) оцінки характеристик випадкової функції

Якщо число реалізацій дорівнює m , то для кожного перетину t_i можна обчислити емпіричні (статистичні) оцінки характеристик випадкової функції $X(t)$. Так, оцінкою математичного сподівання в перетині t_i є просте середнє арифметичне

$$\bar{x}(t_i) = \frac{\sum_{j=1}^m x(t_i)}{m}, \quad (8.15)$$

оцінкою дисперсії -

$$\hat{\sigma}_x(t_i) = S_X^2(t_i) = \frac{1}{m} \sum_{j=1}^m \{ \delta[X(t_i)] \} = \frac{\sum_{j=1}^m [x(t_i) - \bar{x}(t_i)]^2}{m-1} \quad (8.16)$$

Для кореляційних і коваріаційних функцій розрахунок виконується по наступних формулах

$$\hat{r}_x(t_i, t_k) = \frac{\sum_{j=1}^m [x(t_i) - \bar{x}(t_i)] \cdot [x(t_k) - \bar{x}(t_k)]}{(m-1)S_x(t_i)S_x(t_k)} ; \quad (8.17)$$

$$\hat{K}_x(t_i, t_k) = \frac{\sum_{j=1}^m [x(t_i) - \bar{x}(t_i)] \cdot [x(t_k) - \bar{x}(t_k)]}{(m-1)}. \quad (8.18)$$

Після того, як характеристики обчислені, можна побудувати залежності середньої арифметичних $\bar{x}$ і дисперсії S_x^2 від часу. Функції двох аргументів $\hat{r}_x(t_i, t_k)$ і $\hat{K}_x(t_i, t_k)$ також відтворюються в прямокутній сітці точок. За необхідністю всі ці функції апроксимуються аналітичними виразами.

8.4 Поняття про стаціонарність випадкового процесу.

Ергодичність стаціонарних випадкових процесів

Випадкові процеси, що протікають у часі приблизно однорідно, називають стаціонарними. Вони мають вигляд безперервних коливань відносно деякого середнього значення. Реалізації таких процесів знаходяться ніби-то у стані статистичної рівноваги. З часом середня амплітуда і характер коливань істотно не змінюються.

У строгому розумінні слова випадковий процес називають стаціонарним, якщо всі його закони розподілу не змінюються при додаванні до всіх значень аргументу одного й того ж числа, тобто якщо усі вони залежать тільки від взаємного розташування значень аргументу одного й того ж числа або якщо усі вони залежать тільки від взаємного розташування значень аргументу, але не від самих цих значень.

Таким чином, випадковий процес $X(t)$ є стаціонарним, якщо при будь-якому n і будь-якому t_0 будуть виконуватися рівності

$$f_n(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = f_n(x_1, x_2, \dots, x_n; t_1 + t_0, t_2 + t_0, \dots, t_n + t_0) \quad (8.19)$$

Отже, щільності розподілу інваріантні стосовно зсуву початку відліку аргументу t .

Зокрема, для одновимірної щільності розподілу $f_1(x, t)$ стаціонарного випадкового процесу, беручи $t_0 = -t$, отримуємо

$$f_1(x, t) = f_1(x; t - t) = f_1(x; 0) = f_1(x). \quad (8.20)$$

Таким чином, одновимірна щільність розподілу стаціонарного процесу не залежить від t , вона є однією і тією ж для усіх перетинів випадкового процесу. Двовимірна щільність розподілу при $t_0 = -t_1$ представляється у вигляді

$$f_2(x_1, x_2; t_1, t_2) = f_2(x_1, x_2; 0; t_2 - t_1) = f_2(x_1, x_2; t_2 - t_1) = f_2(x_1, x_2; \tau), \quad (8.21)$$

тобто двовимірна щільність розподілу залежить не від двох аргументів t_1, t_2 , а тільки від одного аргументу – їхньої різниці $\tau = t_2 - t_1$. Звідси для стаціонарного випадкового процесу, згідно з (8.20), одержуємо

$$m_x(t) = \int_{-\infty}^{\infty} x f_1(x) dx = m_x = \text{const}, \quad (8.22)$$

тобто математичне сподівання стаціонарного випадкового процесу не залежить від аргументу t і є постійною величиною.

Згідно з (8.21) вираз для кореляційного моменту буде наступний

$$K_x(t_1, t_2) = \iint_{-\infty}^{\infty} (x_1 - m_x)(x_2 - m_x) \cdot f_2(x_1, x_2; \tau) dx_1 dx_2 = K_x(\tau). \quad (8.28)$$

Таким чином, коваріаційна функція стаціонарного випадкового процесу є функцією тільки одного аргументу $\tau = t_2 - t_1$. Умови (8.22) і (8.28) виконуються для будь-якого стаціонарного процесу. Однак вони не є достатніми для визнання стаціонарності, тобто їхнє виконання не гарантує виконання умов (8.19). Таке розуміння стаціонарності випадкових функцій називають стаціонарністю у вузькому розумінні.

У прикладній теорії стаціонарних випадкових функцій користуються поняттям про стаціонарність у широкому розумінні. За звичай припускають, що стаціонарний випадковий процес повинний задовольняти три умови:

$$M[X(t_s)] = \text{const}, D[X(t_s)] = \text{const}; K_X[X(t_s), (t_k)] = K_\tau. \quad (8.24)$$

Таким чином, випадкова функція називається *стаціонарною*, якщо усі її статистичні характеристики не залежать від t , точніше, не змінюються при будь-якому зсуві аргументів, від яких вони залежать.

Однак не всі ці умови рівноцінні. Не можна вважати сталість математичного сподівання істотною вимогою до стаціонарної випадкової функції. При переході до центрованої випадкової функції відразу ж виконується рівність $M\{\delta[X(t)]\} = 0$, тобто перша умова (8.24). Отже, у випадку змінного математичного сподівання $M[X(t)]$ досить перейти до центрування для того, щоб виконати одну з вимог до стаціонарного випадкового процесу.

Виконання другої умови (8.24) є частинним випадком третього. Для стаціонарного випадкового процесу характерна незалежність значень коваріаційної функції від положення відрізка τ на осі t . Коваріаційна функція повинна залежати тільки від довжини проміжку τ (8.28). Покладемо $t_s = t; t_k = t + \tau$. Якщо $\tau = 0$, то згідно (8.12)

$$D_x(t) = K_X[X(t), X(t + 0)] = K_X(\tau = 0) = \text{const}. \quad (8.25)$$

Іншими словами, у випадку стаціонарного випадкового процесу його дисперсія є не що інше, як значення коваріаційної функції при $\tau = 0$, тобто не є функцією аргументу. Слід зазначити, що з властивості симетричності коваріаційної і кореляційної функцій

$$K_x(\tau) = K_x(-\tau); \quad (8.26)$$

$$r_x(\tau) = r_x(-\tau), \quad (8.27)$$

витікає незалежність цих функцій від знака величини τ , тобто можна вважати $\tau = |t_2 - t_1|$. Тому коваріаційну і кореляційну функції знаходять тільки для позитивного аргументу.

Найчастіше зустрічаються стаціонарні випадкові процеси, коваріаційні функції яких апроксимуються функціями наступних типів (рис. 8.7)

$$K_X(\tau) = \sigma^2 e^{-a|\tau|}, \quad a > 0 \quad (8.28)$$

$$K_X(\tau) = \sigma^2 e^{-a\tau^2}, \quad a > 0 \quad (8.29)$$

$$K_X(\tau) = \sigma^2 e^{-a|\tau|} \cdot \cos \beta\tau, \quad a > 0 \quad (8.80)$$

$$K_X(\tau) = \sigma^2 e^{-a\tau^2} \cdot \cos \beta\tau, \quad a > 0 \quad (8.81)$$

$$K_X(\tau) = \sigma^2 e^{-a|\tau|} \left(\cos \beta\tau + \frac{\alpha}{\beta} \sin \beta|\tau| \right), \quad a > 0, \beta > 0 \quad (8.82)$$

$$K_X(\tau) = \begin{cases} \sigma \left(1 - \frac{\tau}{\tau_0} \right) & \text{при } |\tau| \leq \tau_0 \\ 0 & \text{при } |\tau| > \tau_0 \end{cases} \quad (8.83)$$

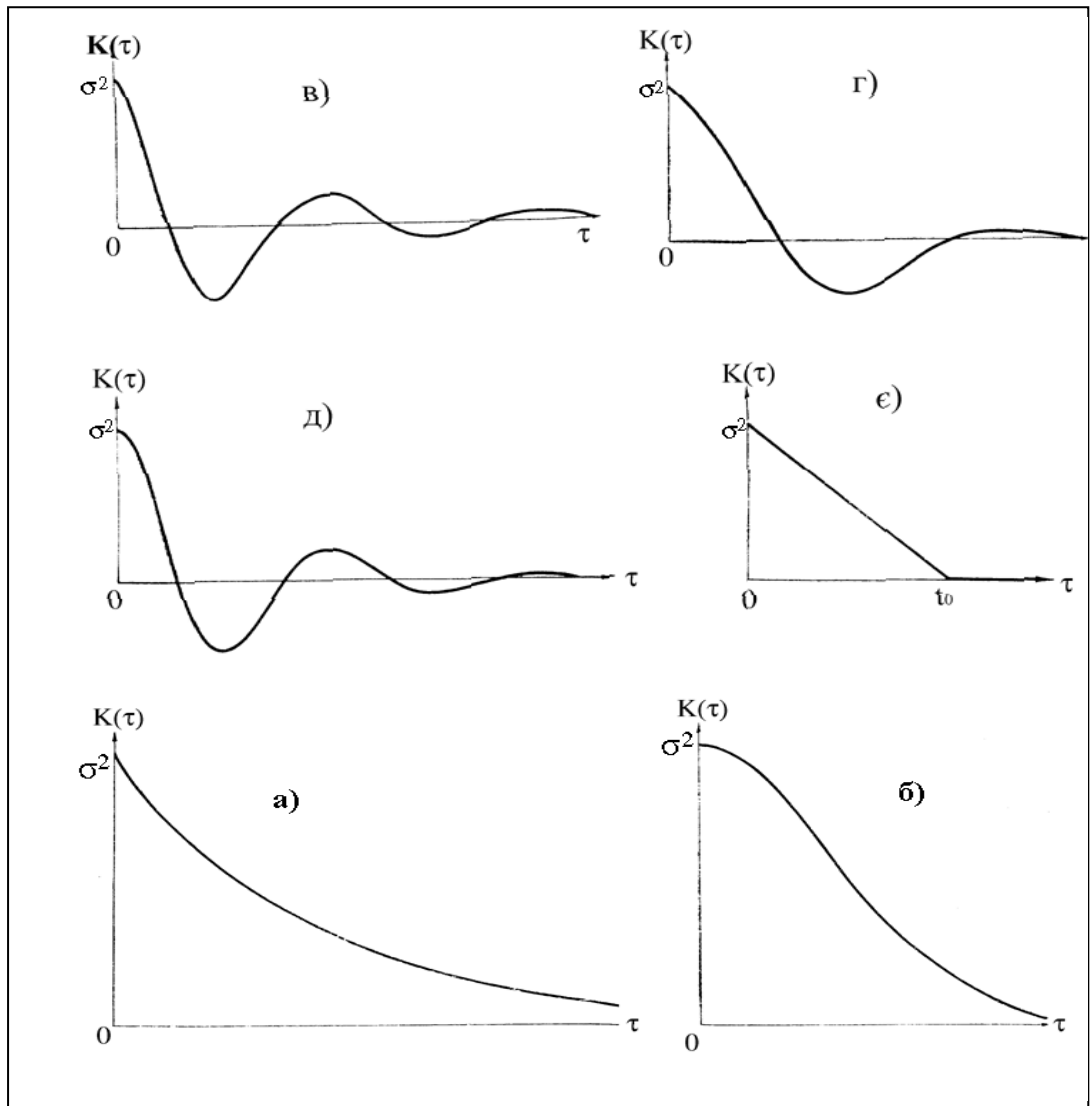


Рисунок 8.7 - Приклади коваріаційних функцій стаціонарних процесів

На рис. 8.7 наведені графіки коваріаційних функцій тільки для $\tau > 0$, у силу парності цих функцій при $\tau < 0$ їм будуть відповідати криві, симетричні щодо осі ординат. З рис. 8.7 видно, що значення кореляційних функцій убувають з ростом τ , тобто кореляційний зв'язок між різними перетинами випадкових функцій убуває зі збільшенням інтервалу часу між ними. Криві на рис.8.7.а, 8.7.б мають вигляд гармонійних коливань з амплітудою, яка зменшується у часі. Вигляд цих кривих указує на наявність періодичності в структурі випадкової функції.

Одержання від'ємних значень $K_X(\tau)$ означає прояв негативних зв'язків між перетинами випадкової функції. Для всіх наведених випадків коваріаційна, а, отже, і кореляційна функція прагне до нуля при прагненні τ до нескінченності. Ця властивість виконується практично для всіх гідрометеорологічних рядів. Коли в структурі випадкової функції, як постійний доданок, є деяка випадкова величина, при $\tau \rightarrow \infty$ коваріаційна функція буде прагнути не до 0, а до дисперсії D цієї величини (рис.8.8). Як правило, вигляд автокореляційної функції, наведений на рис.8.8, характерний для випадкових процесів, які знаходяться під впливом монотонно зростаючих антропогенних перетворень.

Переважає більшість стаціонарних випадкових функцій властива ергодичність. Це відноситься до випадкових функцій, що характеризуються стаціонарністю у вузькому і широкому розумінні.

Властивість ергодичності полягає у тому, що кожна окрема реалізація стаціонарної випадкової функції є повноправним представником усієї сукупності можливих реалізацій. Одна реалізація достатньої довжини може замінити при статистичній обробці всі реалізації в цілому. Статистичні характеристики кожної реалізації є одними й тими ж для усіх реалізацій.

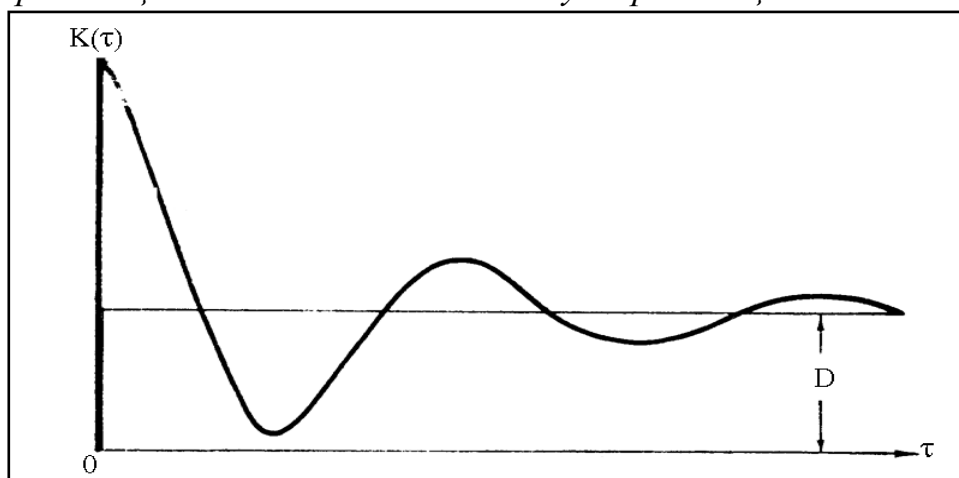


Рисунок 8.8 - Кореляційна функція при наявності постійного впливу на випадковий процес

Розглянемо дві стаціонарні випадкові функції $X_1(t)$ і $X_2(t)$, наведені на рис.8.9. Кожна з реалізацій випадкової функції $X_1(t)$ має одні й тіж самі характерні ознаки: середнє значення, навколо якого коливаються реалізації; розмах коливань. Якщо вибрати довільно одну з таких

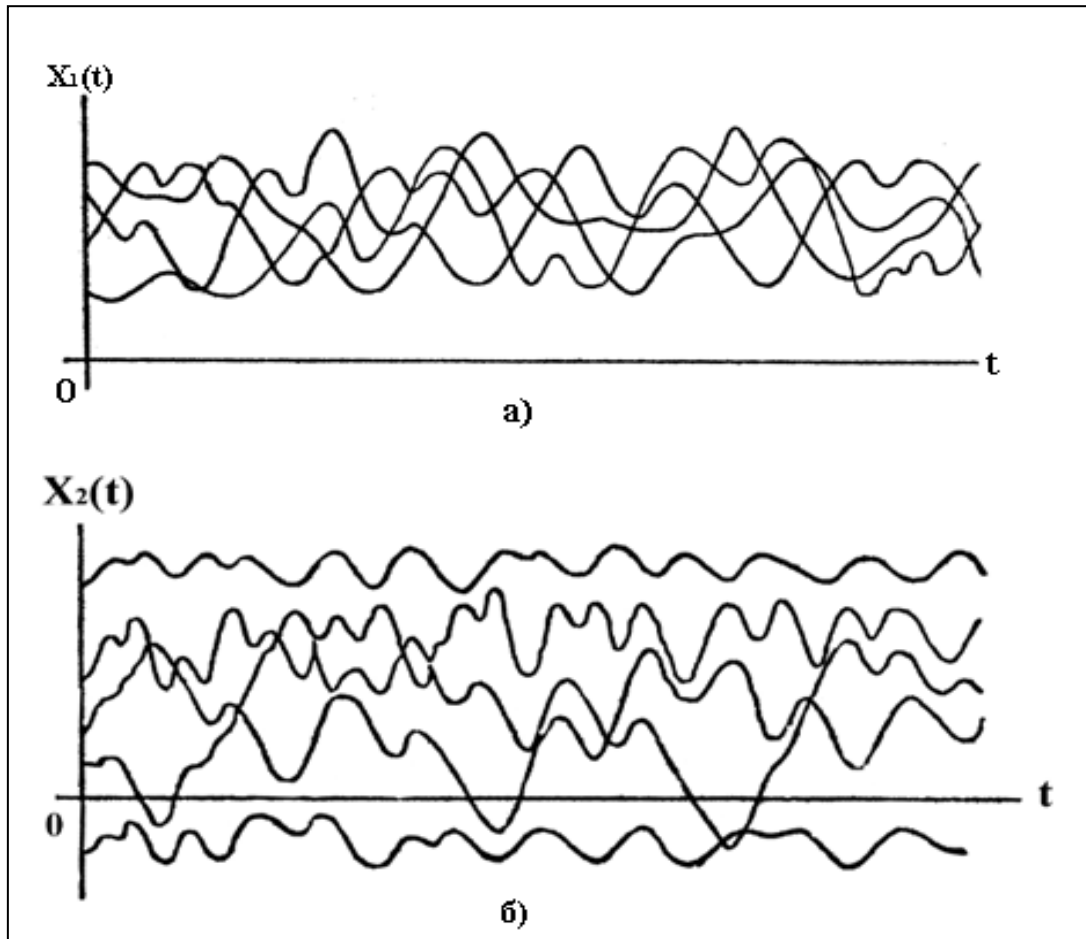


Рисунок 8.9 - Приклади стаціонарних випадкових процесів

реалізацій, то при досить довгій її тривалості T по ній можна одержати повне уявлення про випадкову функцію в цілому.

Осереднюючи значення цієї реалізації за часом, одержують оцінку математичного сподівання:

$$\hat{m}_x \approx \frac{1}{T} \int_0^T x(t) dt, \quad (8.84)$$

автокореляційної функції:

$$\hat{K}_x(\tau) \approx \frac{1}{T-\tau} \int_0^{T-\tau} [x(t) - \hat{m}_x] \cdot [x(t+\tau) - \hat{m}_x] dt. \quad (8.85)$$

Якщо розбити інтервал запису випадкової функції на n рівних частин довжиною Δt і позначити середини отриманих ділянок $t_1, t_2, \dots, t_n$ (рис.8.10), то інтеграли (8.84) і (8.85) можна представити у вигляді:

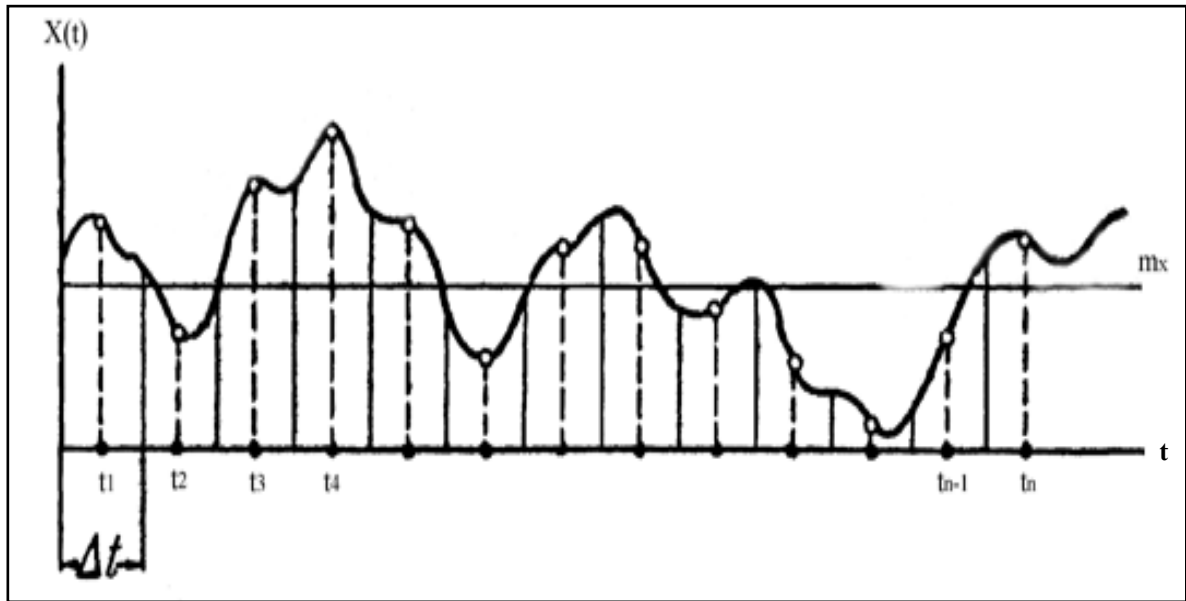


Рисунок 8.10 - Запис випадкової функції, яка розбивається на n різних частин

$$\hat{m}_x = \frac{1}{n} \sum_{i=1}^n x(t_i) = \bar{x}; \quad (8.86)$$

$$\hat{K}_x(\tau) = \frac{1}{n-\tau} \sum_{i=1}^{n-\tau} [x(t_i) - \bar{x}] \cdot [x(t_i + \tau) - \bar{x}], \quad (8.87)$$

де $\tau = m\Delta t$, причому $m = 0, 1, 2, \dots$

У подальшому викладенні для простоти запису знак $\wedge$, який означає, що мова йде про оцінку характеристики за вибірковими даними, прибирається, а t_i представляється як i

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i; \quad (8.88)$$

$$K_x(\tau) = \frac{1}{n - \tau} \sum_{i=1}^{n-\tau} (x_i - \bar{x})(x_{i+\tau} - \bar{x}). \quad (8.89)$$

Загальний вигляд функції $K_x(\tau)$ відтворюється по окремих точках. Для того, щоб математичне сподівання і кореляційна функція були визначені з задовільною точністю, потрібно, щоб число n було досить велике (порядку сотні, а в деяких випадках навіть і більше). Вибір довжини елементарної ділянки Δt визначається характером зміни випадкової функції. Якщо випадкова величина змінюється порівняно плавно, ділянки Δt можна вибирати більшими, ніж коли функція робить різкі і часті коливання. Чим більш високочастотний склад мають коливання, що утворюють випадкову функцію, тим частіше потрібно розміщувати опорні точки при обробці. Орієнтовно можна вибирати елементарні ділянки Δt так, щоб на повний період самої високочастотної гармоніки в складі випадкової функції припадало порядку 5-10 опорних точок.

Значення величини m задаються послідовно, аж до таких m , при яких кореляційна функція практично дорівнює нулю чи починає робити невеликі нерегулярні коливання біля нуля.

8.5 Автокореляційна функція, коефіцієнт автокореляції

У гідрологічних розрахунках річковий стік розглядається як ергодичний процес з річним періодом. Така гіпотеза зручна й необхідна, оскільки по кожному гідрологічному створу є лише одна реалізація процесу стоку, яка представлена гідрологічними спостереженнями. Проте точного математичного доказу справедливості цієї гіпотези немає. Очевидно, що прийнята гіпотеза буде повністю знехтувана, якщо вимірювати час геологічними масштабами. В межах проміжків, що охоплюються періодом гідрологічних спостережень у минулому і перспективного планування в майбутньому за умови незначних антропогенних навантажень, порушення ергодичності стоку маловірогідні.

Кореляційний аналіз є одним з традиційних методів в дослідженнях зв'язків між послідовними членами рядів річного стоку. Проте значне випадкове розсіювання вибірових оцінок коефіцієнтів кореляції при невеликій тривалості вихідних рядів і невеликих числових значеннях $r_x(\tau)$ викликає природне утруднення в інтерпретації результатів розрахунку.

Ординати автокореляційної функції за вибіровими даними обчислюються з урахуванням зміщеності за наступною формулою:

$$r_x(\tau) = \frac{\sum_{i=1}^{n-\tau} (x_i - \bar{x})(x_{i+\tau} - \bar{x})}{\hat{\sigma}_x^2 (n - \tau - 1)}, \quad (8.40)$$

де n - число членів вихідного ряду спостережень;

τ - зсув ряду по відношенню до самого себе при підрахунку коефіцієнта кореляції (запізнювання);

x_i - член ряду від x_i до $x_{n-\tau}$;

$x_{i+\tau}$ - члени ряду від $x_{i+\tau}$ до x_n ;

$\bar{x}, \hat{\sigma}_x$ - відповідно середнє і стандарт вихідного ряду.

В рекомендується враховувати зміну середнього і стандарту вибірки з ростом запізнювання автокореляційної функції. При цьому (8.40) набуває вигляду:

$$r_x(\tau) = \frac{\sum_{i=1}^{n-\tau} (x_i - \bar{x}_i)(x_{i+\tau} - \bar{x}_{i+\tau})}{\sigma_i \sigma_{i+\tau} (n - \tau - 1)}, \quad (8.41)$$

де $x_i, \bar{x}_{i+1}, \sigma_i$ і $\sigma_{i+\tau}$ - середні і стандарт відповідних відрізків вихідного ряду.

На думку більшості дослідників, вираз перетворює вибіркове значення $r_x(\tau)$. Суть перетворення полягає у тому, що лінія регресії проходить при цьому через центр тяжіння не всієї сукупності точок, а тільки точок, які безпосередньо використовуються для підрахунку коефіцієнта кореляції: перетворення посилюється із зростанням запізнювання, причому чим більше T , тим сильніше відрізняються середнє і стандарт вихідного ряду від відповідних значень тієї його частини, яка при зсуві використовується для підрахунку коефіцієнта кореляції.

Характеру автокореляційних функцій рядів річного стоку присвячена велика кількість досліджень як у колишньому СРСР, так і за межами країни. Найбільш суттєві роботи А.Ш.Резніковського, О.І.Чеботарьова і А.В.Рождественського, Д.Я.Ратковича, М.В. Болгова. На цей же час встановлена наявність позитивного зв'язку між річним стоком суміжних років. Загальноприйнятого погляду на значущість внутрішньорядних зв'язків в послідовностях річного стоку поки немає. Навіть при відносно тривалих рядах стоку (50-60 років) недостовірні як окремі ординати автокореляційної функції, так і її контури. Тому при розрахунках річного стоку його ряди розглядаються як простий ланцюг Маркова, тобто враховуються

кореляційні зв'язки тільки між стоком суміжних років. *Послідовність випадкових випробувань, які мають таку властивість, що ймовірність результатів наступного випробування залежить тільки від результатів безпосередньо попередніх випробувань, називаються простим ланцюгом Маркова.* Таким чином, з усіх ординат автокореляційної функції береться до уваги тільки перша, яка відповідає $\tau = 1$.

Коефіцієнт кореляції між стоком суміжних років обчислюють згідно з виразом (8.41), виходячи з умови, що часовий зсув становить одиницю ($\tau = 1$):

$$r(\tau = 1) = r(1) = \frac{\sum_{i=1}^{n-1} (x_i - \bar{x})(x_{i+1} - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2}. \quad (8.42)$$

Середня квадратична похибка емпіричного коефіцієнта автокореляції визначається з рівняння:

$$\sigma_{r(1)} = \frac{1 - r(1)^2}{\sqrt{n-1}}. \quad (8.43)$$

Розрахункове значення коефіцієнта кореляції вважається значущим, якщо виконується умова

$$r(1) \geq 2\sigma_{r(1)}. \quad (8.44)$$

Внутрішньорядні зв'язки у рядах річного стоку зменшують обсяги незалежної інформації. В той же час методи розрахунків статистичних параметрів широко припускають незалежність членів вихідної вибірки. Для усунення впливу внутрішньорядних зв'язків між суміжними членами рядів річного стоку на точність розрахунків статистичних параметрів до коефіцієнтів варіації та асиметрії вносяться поправки

$$C_v = (a_1 + a_2 / n) + (a_3 + a_4 / n) \tilde{C}_v + (a_5 + a_6 / n) \tilde{C}_v^2; \quad (8.45)$$

$$C_s = (b_1 + b_2 / n) + (b_3 + b_4 / n) \tilde{C}_s + (b_5 + b_6 / n) \tilde{C}_s^2, \quad (8.46)$$

де $a_1, \dots, a_6; b_1, \dots, b_6$ - коефіцієнти, які визначаються по табл.8.2 і 8.8;

$\tilde{C}_v, \tilde{C}_s$ - оцінки коефіцієнтів варіації і асиметрії, установлені без урахування внутрішньорядних зв'язків

$$\tilde{C}_v = \sqrt{\frac{\sum_{i=1}^n (k_i - 1)^2}{n - 1}} ; \quad (8.47)$$

$$\tilde{C}_s = \frac{n}{(n-1)(n-2)} \frac{\sum_{i=1}^n (k_i - 1)^3}{\tilde{C}_v^3} ; \quad (8.48)$$

де $k_i = x_i / \bar{x}$.

Таблиця 8.2 – Коефіцієнти a у формулі (8.45)

C_s / C_v	$r(1)$	a_1	a_2	a_3	a_5	a_6	a_7
1	2	8	4	5	6	7	8
2.00	0.00	0.00	0.19	0.99	-0.88	0.01	1.54
	0.80	0.00	0.22	0.99	-0.41	0.01	1.51
	0.50	0.00	0.18	0.98	0.41	0.02	1.47
8.00	0.00	0.00	0.69	0.98	-4.84	0.01	6.78
	0.80	0.00	1.15	1.02	-7.58	-0.04	12.88
	0.50	0.00	1.75	1.00	-11.79	-0.05	21.18
4.00	0.00	0.00	1.86	1.02	-9.68	-0.05	15.55
	0.80	-0.02	2.61	1.18	-19.85	-0.22	84.15
	0.50	-0.02	8.47	1.18	-29.71	-0.41	58.08

Таблиця 8.8 – Коефіцієнти b у формулі (8.46)

$r(1)$	b_1	b_2	b_3	b_4	b_5	b_6
0.00	0.08	2.00	0.92	-5.09	0.08	8.10
0.80	0.08	1.77	0.98	-8.45	0.08	8.08
0.50	0.08	1.68	0.92	-0.97	0.08	7.94

Середні квадратичні похибки вибірових параметрів обчислюються таким чином:

$$\sigma_x = \frac{\sigma_x}{\sqrt{n}} \sqrt{\frac{1+r(1)}{1-r(1)}}; \quad (8.49)$$

$$\sigma_{c_v} = \frac{C_v}{n + 4C_v^2} \sqrt{\frac{n(1 + C_v^2)}{2}} \left(1 + \frac{3C_v r(1)^2}{1 - r(1)} \right). \quad (8.50)$$

8.6 Приклад розрахунків

8.6.1 Розрахувати коефіцієнт автокореляції та уточнити статистичні параметри річного стоку з водозбору р. Дерекойка – м. Ялта за 89-ти річний період спостережень

Вихідні дані та результати обчислень оформлюють у вигляді табл.8.4.

- середнє багаторічне значення річного стоку $\bar{q} = 10.5$ л/с км²;
- середньоквадратичне відхилення $\sigma_q = 4.2$ л/с км²;
- коефіцієнт варіації річного стоку $\hat{C}_v = 0.40$;
- коефіцієнт асиметрії річного стоку $C_s = 1.16$;
- співвідношення $\hat{C}_s / \hat{C}_v = 2.89 \approx 3.00$
- середня квадратична похибка розрахунку параметра $\bar{q}$:
 $\sigma_{\bar{q}} = 0.67$ л/с · км²;

Коефіцієнт автокореляції $r(1)$ при зсуві $\tau = 1$, розрахований по (8.42), дорівнює 0.38, а середньоквадратична похибка його $\sigma_{r(1)}$ становить 0.152. Умова (8.45) виконується, тобто $0.880 > 2 \times 0.152$, через що значення параметра $r(1)$ слід визнати значущим. Використовуючи формули (8.45) та (8.46), а також табл. 8.2 і 8.8, визначаємо параметри $\hat{C}_v$ і $\hat{C}_s$ з урахуванням внутрішньорядних зв'язків. Уточнені значення параметрів такі: $\hat{C}_v = 0.40$, $\hat{C}_s = 1.39$, $\hat{C}_s / \hat{C}_v = 3.43$, $\sigma_q = 0.67$, $\sigma_{C_v} = 0.049$. Як показують результати розрахунків, вплив внутрішньорядних зв'язків особливо виражений при розрахунках коефіцієнта асиметрії та співвідношення $\hat{C}_s / \hat{C}_v$.

Таблиця 8.4 – Обчислення коефіцієнту автокореляції річного стоку з водозбору р.Дерекойка - м.Ялта

№ пп	q_i л/скм ²	q_{i+1} л/скм ²	$q_i - \bar{q}$	$(q_i - \bar{q})^2$	$q_{i+1} - \bar{q}$	$(q_i - \bar{q}) \times$ $(q_{i+1} - \bar{q})$
1	2	8	4	5	6	7
1	11.8		0.80	0.64		
2	10.8	11.8	-0.20	0.04	0.80	-0.16
8	9.46	10.8	-1.04	1.08	-0.20	0.21
4	8.45	9.46	-2.05	4.20	-1.04	2.18
5	7.44	8.45	-8.06	9.86	-2.05	6.27
6	18.9	7.44	8.40	11.6	-8.06	-10.4
7	14.5	18.9	4.00	16.0	8.40	18.6
8	7.85	14.5	-2.65	7.02	4.00	-10.6
9	9.66	7.85	-0.84	0.71	-2.65	2.28
10	18.8	9.66	2.80	7.84	-0.84	-2.85
11	4.48	18.8	-6.07	86.8	2.80	-17.0
12	7.85	4.48	-2.65	7.02	-6.07	16.1
18	9.05	7.85	-1.45	2.10	-2.65	8.84
14	8.25	9.05	-2.25	5.06	1.45	8.26
15	9.26	8.25	-1.24	1.54	-2.25	2.79
16	11.5	9.26	1.00	1.00	-1.24	-1.24
17	8.05	11.5	-2.45	6.00	1.00	-2.45
18	8.25	8.05	-2.25	5.06	-2.45	5.51
19	18.5	8.25	8.00	9.00	-2.25	-6.75
20	10.5	18.5	0.00	0.00	8.00	0.00
21	17.5	10.5	7.00	49.0	0.00	0.00
22	20.1	17.5	9.60	92.2	7.00	67.2
28	10.1	20.1	-0.40	0.16	9.60	-8.84
24	10.9	10.1	0.40	0.16	-0.40	-0.16
25	8.82	10.9	-6.68	44.6	0.40	-2.67
26	4.00	8.82	-6.50	42.2	-6.60	48.4
27	10.5	4.00	0.00	0.00	-6.50	0.00
28	7.24	10.5	-8.26	10.6	0.00	0.00

Продовження таблиці 8.4

№ пп	q_i л/скм ²	q_{i+1} л/скм ²	$q_i - \bar{q}$	$(q_i - \bar{q})^2$	$q_{i+1} - \bar{q}$	$(q_i - \bar{q}) \times$ $(q_{i+1} - \bar{q})$
1	2	8	4	5	6	7
29	9.26	7.24	-1.24	1.54	-8.26	4.04
80	8.85	9.26	-1.65	2.72	-1.24	1.55
81	6.84	8.85	-8.66	18.4	-8.66	6.04
82	5.68	6.84	-4.87	28.7	-4.87	17.8
88	11.5	5.68	1.00	1.00	1.00	-4.87
84	12.9	11.5	2.40	5.76	2.40	2.40
85	24.7	12.9	14.2	202	14.2	8.40
86	15.9	24.7	5.40	29.2	5.40	76.7
87	14.7	15.9	4.20	17.6	4.20	22.7
88	8.45	14.7	-2.05	4.20	-2.05	-8.61
89	9.86	8.45	-0.64	0.41	-0.64	1.81
Сума				672		262

Таблиця 3.2 – Коефіцієнти a у формулі (3.45)

C_s / C_v	$r(1)$	a_1	a_2	a_3	a_5	a_6	a_7
1	2	3	4	5	6	7	8
2.00	0.00	0.00	0.19	0.99	-0.88	0.01	1.54
	0.30	0.00	0.22	0.99	-0.41	0.01	1.51
	0.50	0.00	0.18	0.98	0.41	0.02	1.47
3.00	0.00	0.00	0.69	0.98	-4.34	0.01	6.78
	0.30	0.00	1.15	1.02	-7.53	-0.04	12.38
	0.50	0.00	1.75	1.00	-11.79	-0.05	21.13
4.00	0.00	0.00	1.36	1.02	-9.68	-0.05	15.55
	0.30	-0.02	2.61	1.13	-19.85	-0.22	34.15
	0.50	-0.02	3.47	1.18	-29.71	-0.41	58.08

Таблиця 3.3 – Коефіцієнти b у формулі (3.46)

$r(1)$	b_1	b_2	b_3	b_4	b_5	b_6
0.00	0.03	2.00	0.92	-5.09	0.03	8.10
0.30	0.03	1.77	0.93	-3.45	0.03	8.03
0.50	0.03	1.63	0.92	-0.97	0.03	7.94

Таблиця 3.4 – Обчислення коефіцієнту автокореляції річного стоку з водозбору р.Дерекойка - м.Ялта

№ пп	q_i л/скм ²	q_{i+1} л/скм ²	$q_i - \bar{q}$	$(q_i - \bar{q})^2$	$q_{i+1} - \bar{q}$	$(q_i - \bar{q}) \times$ $(q_{i+1} - \bar{q})$
1	2	3	4	5	6	7
1	11.3		0.80	0.64		
2	10.3	11.3	-0.20	0.04	0.80	-0.16
3	9.46	10.3	-1.04	1.08	-0.20	0.21
4	8.45	9.46	-2.05	4.20	-1.04	2.13
5	7.44	8.45	-3.06	9.36	-2.05	6.27
6	13.9	7.44	3.40	11.6	-3.06	-10.4
7	14.5	13.9	4.00	16.0	3.40	13.6
8	7.85	14.5	-2.65	7.02	4.00	-10.6
9	9.66	7.85	-0.84	0.71	-2.65	2.23
10	13.3	9.66	2.80	7.84	-0.84	-2.35
11	4.43	13.3	-6.07	36.8	2.80	-17.0
12	7.85	4.43	-2.65	7.02	-6.07	16.1
13	9.05	7.85	-1.45	2.10	-2.65	3.84
14	8.25	9.05	-2.25	5.06	1.45	3.26
15	9.26	8.25	-1.24	1.54	-2.25	2.79
16	11.5	9.26	1.00	1.00	-1.24	-1.24
17	8.05	11.5	-2.45	6.00	1.00	-2.45
18	8.25	8.05	-2.25	5.06	-2.45	5.51
19	13.5	8.25	3.00	9.00	-2.25	-6.75
20	10.5	13.5	0.00	0.00	3.00	0.00
21	17.5	10.5	7.00	49.0	0.00	0.00
22	20.1	17.5	9.60	92.2	7.00	67.2
23	10.1	20.1	-0.40	0.16	9.60	-3.84
24	10.9	10.1	0.40	0.16	-0.40	-0.16
25	3.82	10.9	-6.68	44.6	0.40	-2.67
26	4.00	3.82	-6.50	42.2	-6.60	43.4
27	10.5	4.00	0.00	0.00	-6.50	0.00
28	7.24	10.5	-3.26	10.6	0.00	0.00

Продовження таблиці 3.4

№ пп	q_i л/скм ²	q_{i+1} л/скм ²	$q_i - \bar{q}$	$(q_i - \bar{q})^2$	$q_{i+1} - \bar{q}$	$(q_i - \bar{q}) \times$ $(q_{i+1} - \bar{q})$
1	2	3	4	5	6	7
29	9.26	7.24	-1.24	1.54	-3.26	4.04
30	8.85	9.26	-1.65	2.72	-1.24	1.55
31	6.84	8.85	-3.66	13.4	-3.66	6.04
32	5.63	6.84	-4.87	23.7	-4.87	17.8
33	11.5	5.63	1.00	1.00	1.00	-4.87
34	12.9	11.5	2.40	5.76	2.40	2.40
35	24.7	12.9	14.2	202	14.2	3.40
36	15.9	24.7	5.40	29.2	5.40	76.7
37	14.7	15.9	4.20	17.6	4.20	22.7
38	8.45	14.7	-2.05	4.20	-2.05	-8.61
39	9.86	8.45	-0.64	0.41	-0.64	1.31
Сума				672		262

Література

1. Аполлов Б.А., Калинин Г.П., Комаров В.Д. Курс гидрологических прогнозов. – Л.: Гидрометеиздат, 1984.
2. Багров Н.А. Аналитическое представление последовательности метеорологических полей посредством естественных ортогональных составляющих // Труды ЦИП.- 1959. - Вып. 74. - С. 133-138.
3. Бершадский А.Г. Крупномасштабные фрактальные структуры в лабораторной турбулентности, океане и астрофизике // Успехи физических наук. – 1990. – Т. 160. – Вып. 12. – С. 189-194.
4. Бефани А.Н., Мельничук О.П. Расчет нормы стока временных водотоков и горных Украинских Карпат // Труды УкрНИГМИ. - Л.: Гидрометеиздат. - 1967. - вып. 69. - С. 105 -131.
5. Бефани Н.Ф., Калинин Г.П. Упражнения и методические разработки по гидрологическим прогнозам. – Л.: Гидрометеиздат, - 1983. – 390с.
6. Болгов М.В., Мишон В.М., Сенцова Н.И. Современные проблемы оценки водных ресурсов и водообеспечения. – М.: Наука, 2005, 318 с.
7. Вентцель Е.С. Теория вероятностей. – М. ФМ, 1962, 264 с.
8. Вишневський В.І., Косовець О.О. Гідрологічні характеристики річок України.- К.: - Ніка-Центр. – 2003.-324 с.
9. Гринвальд Д.И. Турбулентность русловых потоков. – Л.: Гидрометеиздат, 1974, 166с.
10. Жук В.А., Евстигнеев В.М. Исследование синхронности колебаний годового стока отдельных регионов приемами факторного анализа: Труды ВНИИГМИ-МЦД. - М.: Гидрометиздат. С. 78-91с.
11. Евстигнеев В.М. Речной сток и гидрологические расчеты. – М.: Изд-во МГУ, 1990. – 304 с.
12. Иберла К. Факторный анализ: Пер. с англ. - М.: Статистика, 1980. - 397с.
13. Иванов С.С. Определение фрактальной размерности глобального рельефа // Океанология. – 1994. – Т. 34, №1. – С. 102-106.

14. Исследования и расчеты речного стока / Под ред. В.Д.Быкова. - М.: Изд-во МГУ, 1981. -.228 с.
15. Карасев И.Ф., Савельева Л.Н. Разложение гидрологических полей на естественные ортогональные составляющие и расчет слоев стока весеннего половодья неизученных рек // Моделирование и прогнозы гидрологических процессов. – Л.: РГГМИ, 1992. – Вып.113. – С.76-84.
16. Крицкий С.Н., Менкель М.Ф. Гидрологические основы управления речным стоком.- М. Наука,1981. - 235с.
17. Крицкий С.Н., Менкель М.Ф. Гидрологические основы управления водохозяйственными системами.- М. Наука,1982. – 271 с.
18. Кучмент Л.С. Речной сток (генезис, моделирование, предвычисление). – М.: Институт водных проблем РАН, 2008. – 394с.
19. Лобода Н.С. Установление свойств самоподобия в пространственном распределении статистических параметров годового стока рек Украины // Гідрологія, гідрохімія і гідроекологія. – Київ. – Обрії. – 2004. – Т.6. – С.116-123.
20. Лобода Н.С., Горобец Т.В. Фрактальные свойства в многолетних колебаниях годового стока рек Украины // Вісник Одеського державного екологічного університету. – Вип.1. - К:КНТ. –2005. - С. 174 -182.
21. Лобода Н.С. Расчеты и обобщения характеристик годового стока рек Украины в условиях антропогенного влияния: Монография. – Одесса: Экология, 2005. – 208 с.
22. Лобода Н.С. Гідрологічні прогнози (конспект лекцій). – Одеса.: 2003. – 138с.
23. Лобода Н.С., Гопченко Є.Д. Стохастичні моделі у гідрологічних розрахунках. - Навчальний посібник.– Одеса: Екологія, 2006. – 200 с.
24. Лобода Н.С., Нгуен Ву Ань Статистическая структура полей годового стока в бассейне р.Уссури и стокоформирующие факторы // Украинский гидрометеорологический журнал. - №1. – 2006. – С.170-176.
25. Лобода Н.С. Ландшафтна різноманітність та районування характеристик стоку Українських Карпат // Науковий Вісник Чернівецького університету. – Вип.305. , Географія. – 2006. - С. 12-19.

- 26.Лобода Н.С., Сіренко А.М. Використання дискримінантної функції в короткострокових прогнозах появи плаваючого льоду (на прикладі річки Дністер на ділянці м.Могильов – Подільський-м.Сороки) // Вісник Одеського державного екологічного університету. – Вип.5. - Одеса:ТЕС. –2008. - С. 168-174.
- 27.Лоули Д., Максвелл А. Факторный анализ как статистический метод.- М.: Мир, 1967. -144с.
- 28.Мостеллер Ф., Тьюки Дж. Анализ данных и регрессия. - М.: Финансы и статистика. - 1982. - 120 с.
- 29.Никора В.И. Фрактальные свойства некоторых гидрологических объектов. – Кишинев,1988- 43с.
- 30.Пособие по определению расчетных гидрологических характеристик. - Л.: Гидрометеиздат,- 1984. - 447с.
- 31.Пространственно-временные колебания стока рек СССР / Под ред.А.В. Рождественского. – Л.: Гидрометеиздат,1988. – 376 с.
- 32.Раткович Д.Я. Многолетние колебания речного стока. - Л.: Гидрометеиздат. - 1976. - 255 с.
- 33.Рождественский А.В.,Чеботарев А.И. Статистические методы в гидрологии. - Л.: Гидрометеиздат, 1974. - 424с.
- 34.Руководство по гидрологическим прогнозам. – Вып. 3. Прогноз ледовых явлений на реках и водохранилищах. – Л.: Гидрометеиздат. – 1989. – 168с.
- 35.Смирнов Н.П., Складенко В.Л. Методы многомерного статистического анализа в гидрологических исследованиях. - Л.: Ленингр. ун-т, 1986. -192 с.
- 36.Факторный, дискриминантный и кластерный анализ. Пер. с англ. /Дж.-О.Ким,Ч.У.Мьюллер,У.Р. Клекка и др.- Финансы и статистика, 1989, - 215с.
- 37.Христофоров А.В. Надежность расчетов речного стока. – М.: Изд.-во МГУ. – 1993. – 168 с.
- 38.Школьный Є.П., Лоева І.Д., Гончарова Л.Д. Обробка та аналіз гідрометеорологічної інформації: навчальний підручник. - К.: Міносвіти України, 1999. - 600 с.
- 39.Kantelhardt Jan W., Stephan A. Zschiegner, Eva Koscielny-Bunde, Shlomo Havlin, Armin Bunde, H. Eugene Stanley. Multifractal

- detrended fluctuation analysis of nonstationary time series // Physika A. –316.- 2002. –P.87-114.
- 40.Mandelbrot B.B. The Fractal Geometry of Nature / Ed. Freeman W.N. and Co. New-York,1983. -469p.
- 41.Loboda N., Glushkov A., Khohlov V. Using meteorological data for reconsturction of annual runoff series over an ungauged area: Empirical orthogonal function approach to Moldova-Southwest Ukraine region // Atmospheric Research. – 2005. - Vol 77/1-4. – P. 100-113.

ДОДАТОК А

ОСНОВНІ ВІДОМОСТІ З ТЕОРІЇ МАТРИЦЬ

Матрицею називається система елементів, розташованих у визначеному порядку і утворюючих таблицю. Якщо в цій таблиці m рядків та n стовбців, а її елементи позначаються як a_{ij} , де i - номер рядка, а j - номер стовпця, на перетині яких знаходиться елемент a_{ij} , то матриця записується у наступному вигляді

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix} \quad (\text{A.1})$$

Ця матриця може бути представленою також таким чином $A = [a_{ij}]$, ($i = 1, 2, 3, \dots, m$; $j = 1, 2, 3, \dots, n$). Наведена матриця має розмір $m \times n$.

Якщо кількість рядків m не дорівнює кількості стовбців n , то матриця зветься **прямокутною**.

Якщо в матриці тільки один стовпець, тобто $n=1$, то така матриця називається матрицею-стовпцем і має наступний вигляд

$$A = \begin{bmatrix} a_{11} \\ a_{21} \\ a_{31} \\ \dots \\ a_{m1} \end{bmatrix} \quad (\text{A.2})$$

Матриця – стовпець також називається **вектором – стовпцем**. Якщо в матриці тільки один рядок, тобто $m=1$, то вона називається матрицею-рядком або **вектором – рядком**.

$$A = [a_{11} \ a_{12} \ a_{13} \ \dots \ a_{1n}] \quad (\text{A.3})$$

Якщо в матриці поміняти місцями рядки та стовбці, то отримаємо нову матрицю, яка у порівнянні з матрицею A називається транспонованою й позначається як A' .

Наприклад, якщо

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix}, \quad (\text{A.4})$$

то транспонована матриця A' буде записуватися у виді

$$A' = \begin{bmatrix} a_{11} & a_{21} & a_{31} \\ a_{12} & a_{22} & a_{32} \\ a_{13} & a_{23} & a_{33} \\ a_{14} & a_{24} & a_{34} \end{bmatrix} \quad (\text{A.5})$$

Матриця A має розмір 3×4 , а транспонована матриця 4×3 . Коли в матриці стільки ж рядків, скільки і стовпців, то така матриця зветься **квадратною**. Наприклад,

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{bmatrix} \quad (\text{A.6})$$

Наведена квадратна матриця має розмір $n \times n$.

Квадратна матриця, в якій всі елементи, що розташовані поза головною діагоналлю, дорівнюють нулю, зветься **діагональною**. Наприклад,

$$A = \begin{bmatrix} a_{11} & 0 & 0 & \dots & 0 \\ 0 & a_{22} & 0 & \dots & 0 \\ 0 & 0 & a_{33} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & a_{nn} \end{bmatrix} \quad (\text{A.7})$$

В діагональній матриці всі елементи з індексами, які не дорівнюють одне одному, дорівнюють нулю.

Діагональна матриця, всі елементи якої, що розташовані на головній діагоналі, дорівнюють одне одному, називається *скалярною*. Наприклад,

$$B = \begin{bmatrix} a_{11} & 0 & 0 & \dots & 0 \\ 0 & a_{11} & 0 & \dots & 0 \\ 0 & 0 & a_{11} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & a_{11} \end{bmatrix} \quad (\text{A.8})$$

Одинична матриця – це діагональна матриця, у якій всі елементи головної діагоналі дорівнюють одиниці. Вона позначається символом E і має такий вигляд:

$$E_n = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} \quad (\text{A.9})$$

Індекс n у символі E показує порядок одиночної матриці.

Верхня трикутна матриця – це квадратна матриця, в якій елементи, розташовані нижче головної діагоналі, дорівнюють нулю. Наприклад,

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & a_{22} & a_{23} & \dots & a_{2n} \\ 0 & 0 & a_{33} & \dots & a_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & a_{nn} \end{bmatrix} \quad (\text{A.10})$$

Нижня трикутна матриця – це квадратна матриця, в якій всі елементи, розташовані вище головної діагоналі, дорівнюють нулю. Наприклад,

$$A = \begin{bmatrix} a_{11} & 0 & 0 & 0 & 0 \\ a_{21} & a_{22} & 0 & 0 & 0 \\ a_{31} & a_{32} & a_{33} & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & a_{n3} & \dots & a_{nn} \end{bmatrix} \quad (\text{A.11})$$

Симетричною матрицею називається квадратна матриця, в якій всі елементи, розташовані симетрично до головної діагоналі, дорівнюють одне одному. Для симетричної матриці має місце

рівність $a_{ij} = a_{ji}$. Наприклад, $\begin{bmatrix} 1 & 4 & 2 \\ 4 & 3 & 6 \\ 2 & 6 & 5 \end{bmatrix}$.

Визначником квадратної матриці називається визначник, складений з елементів матриці, розташованих таким самим чином, як і в матриці. Визначник матриці A позначається символом $|A|$.

Якщо визначник матриці дорівнює нулю, то вона називається **виродженою**, або **сингулярною**.

У прямокутній матриці можна викреслити декілька рядків та декілька стовбців таким чином, щоб елементи, які залишились невикресленими, утворили квадратну матрицю порядку k . Визначник такої матриці називається **мінором** вихідної матриці. Якщо взяли матрицю розміром 3×4 , то викреслюючи один будь-

який стовпець можна отримати чотири квадратні матриці 3-го порядку. Викреслюванням двох стовпців та одного рядка можна отримати 18 матриць другого порядку.

Кількість матриць порядку k ($k \leq \min(s, r)$), яке можна отримати з матриці розміром $s \times r$, підраховується за формулою

$$n = C_s^k C_r^k, \quad (\text{A.12})$$

де C_s^k та C_r^k є кількість комбінацій з s елементів по k у кожній та з r елементів по k у кожній. Кількість комбінацій з n елементів по m підраховується за формулою

$$C_n^m = \frac{n!}{m!(n-m)!} \quad (n \geq m) \quad (\text{A.13})$$

Найбільший порядок r мінору матриці, визначник якої не дорівнює нулю, називається **рангом** матриці.

Ранг діагональної квадратної матриці дорівнює її порядку, зрозуміло, за умови, що жоден з її діагональних елементів не дорівнює нулю.

Дві матриці A і B називаються **рівними**, якщо:

- обидві мають один і той же розмір, тобто кількість рядків матриці A дорівнює числу рядків матриці B , а кількість стовбців матриці A дорівнює кількості стовбців матриці B ;
- відповідні елементи цих матриць рівні між собою. (Під відповідними елементами слід розуміти елементи з одними й тими ж індексами).

Таким чином, якщо матриця A має вигляд

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix}, \quad (\text{A.14})$$

а матриця B -

(A.15)

И $a_{ij} = b_{ij} \quad (i = 1, 2, \dots, m; \quad j = 1, 2, \dots, n)$, то $A = B$.

Припустимо, що n величин $x_1, x_2, x_3, \dots, x_n$ пов'язані з m величинами $y_1, y_2, y_3, \dots, y_m$ такими залежностями

(A.16)

За цими формулами величини $y_1, y_2, y_3, \dots, y_m$ виражаються лінійно й однорідно через n інших величин $x_1, x_2, x_3, \dots, x_n$.

Відбувається перетворення величин $x_1, x_2, x_3, \dots, x_n$ у величини $y_1, y_2, y_3, \dots, y_m$. Таке перетворення має назву **лінійного** й скорочено записується таким чином

(A.17)

Коефіцієнти лінійного перетворення можуть бути записані у вигляді матриці

(A.18)

яка називається **матрицею лінійного перетворення** величин x_j ($j=1,2,\dots,n$) у величини y_i ($i=1,2,\dots,m$).

Сумою двох прямокутних матриць A та B рівних розмірів $m \times n$ називається матриця C того ж розміру, елементи якої C_{ij} дорівнюють сумі відповідних елементів матриць A та B

$$A + B = C, \quad (\text{A.19})$$

$$C = \begin{bmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \dots & \dots & \dots & \dots \\ c_{m1} & c_{m2} & \dots & c_{mn} \end{bmatrix}, \quad (\text{A.20})$$

де $c_{ij} = a_{ij} + b_{ij}$ ($i=1,2,\dots,m; j=1,2,\dots,n$).

Різниця матриць A та B , які мають однаковий порядок, є матриця того ж порядку D , елементи якої дорівнюють різниці відповідних елементів матриць A й B , тобто

$$d_{ij} = a_{ij} - b_{ij} \quad (\text{A.21})$$

Сума матриць підкорюється наступним законам (сполучному або асоціативному та переставному або комутативному):

$$\begin{aligned} A + B &= B + A \\ (A + B) + C &= A + (B + C) \end{aligned} \quad (\text{A.22})$$

Щоб помножити матрицю A на число α , потрібно помножити на це число кожен елемент матриці A . Нова матриця записується таким чином

$$\alpha A = \begin{bmatrix} \alpha a_{11} & \alpha a_{12} & \alpha a_{13} & \dots & \alpha a_{1n} \\ \alpha a_{21} & \alpha a_{22} & \alpha a_{23} & \dots & \alpha a_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ \alpha a_{m1} & \alpha a_{m2} & \alpha a_{m3} & \dots & \alpha a_{mn} \end{bmatrix} \quad (\text{A.23})$$

Добуток BA двох матриць можливо отримати у тому й тільки у тому випадку, коли кількість стовбців матриці B дорівнює кількості рядків матриці A . Коли розмір матриці B дорівнює $s \times m$, а розмір матриці A дорівнює $m \times n$, то матриця BA має розмір $s \times n$.

У символічному виді це записується так

$$(s \times m)(m \times n) = s \times n \quad (\text{A.24})$$

Якщо $C = BA$, то кожний елемент c_{ij} матриці C дорівнює сумі добутків елементів i -того рядка матриці B на відповідні елементи j -того стовпця матриці A . Наприклад, елемент C_{14} представляє собою суму добутків першого рядка матриці B на четвертий стовбець матриці A , тобто

$$c_{14} = b_{11}a_{14} + b_{12}a_{24} + b_{13}a_{34} + \dots + b_{1m}a_{m4} \quad (\text{A.25})$$

У загальному випадку

$$c_{ij} = b_{i1}a_{1j} + b_{i2}a_{2j} + b_{i3}a_{3j} + \dots + b_{im}a_{mj} = \sum_{k=1}^m b_{ik} \cdot a_{kj} \quad (\text{A.26})$$

Добуток двох матриць не має властивості комутативності

$$BA \neq AB \quad (\text{A.27})$$

та асоціативності

$$A(BC) \neq (AB)C. \quad (\text{A.28})$$

ДОДАТОК Б

Матриці коваріацій і кореляцій

Розв'язок чисельних задач сучасної гідрометеорології потребує знань про статистичну структуру метеорологічних полів, таких як, наприклад, поля опадів, температур, стоку, тиску та вологості повітря, швидкості вітру, тощо. Сукупність m гідрометеорологічних полів, які відносяться до визначених термінів спостереження, можна, як відзначалося вище, зобразити матрицею порядку $n \times m$ вигляду [9,38]

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2m} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{i1} & x_{i2} & \dots & x_{ij} & \dots & x_{im} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{ni} & \dots & x_{nm} \end{bmatrix} \quad (\text{Б.1})$$

Таким же чином можна скласти інформацію, наприклад, про m вертикальних профілів метеорологічної величини, яка вимірюється на n рівнях атмосфери або про множину m векторів-предикторів у випадку, коли вирішується проблема побудови статистичної моделі гідрометеорологічного прогнозу. Можна навести приклади й інших задач. Щоб не перелічувати кожний раз ці задачі, будемо у подальшому інколи іменувати поля, вертикальні профілі гідрологічних величин, вектори-предиктори й т.п. гідрометеорологічними об'єктами.

Матриця (Б.1) утримує великий об'єм інформації. У матриці (Б.1) концентрується інформація про n об'єктів. Рядки матриці являють собою, як вже зазначалося, часові ряди відповідної гідрометеорологічної величини. Таке матричне зображення гідрометеорологічних об'єктів є дуже раціональним, оскільки дає можливість побудувати прості алгоритми дослідження їх статистичної структури. Найбільш важлива інформація про

статистичну структуру гідрометеорологічних об'єктів міститься у матриці коваріацій. Алгоритм побудови цієї матриці складається з декількох етапів. Знайдемо, насамперед, вектор середніх значень

$$\bar{X} = \begin{bmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \dots \\ \bar{x}_i \\ \dots \\ \bar{x}_n \end{bmatrix}, \quad (\text{Б.2})$$

де

$$\bar{x}_i = \frac{1}{m} \sum_{j=1}^m x_{ij} \quad (i = \overline{1, m}) . \quad (\text{Б.3})$$

Підкреслимо ще раз, що коли йдеться про поля гідрометеорологічних величин, то вектор (Б.2) є осереднене поле гідрометеорологічної величини.

На основі матриці (Б.1) та вектора (Б.2) визначимо відповідну матрицю центрованих елементів. Для цього від елементів кожного рядка матриці (Б.1) віднімемо відповідне середнє арифметичне значення. У такому випадку отримаємо:

$$\Delta X = \begin{bmatrix} \Delta x_{11} & \Delta x_{12} & \dots & \Delta x_{1j} & \dots & \Delta x_{1m} \\ \Delta x_{21} & \Delta x_{22} & \dots & \Delta x_{2j} & \dots & \Delta x_{2m} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \Delta x_{i1} & \Delta x_{i2} & \dots & \Delta x_{ij} & \dots & \Delta x_{im} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \Delta x_{n1} & \Delta x_{n2} & \dots & \Delta x_{ni} & \dots & \Delta x_{nm} \end{bmatrix}, \quad (\text{Б.4})$$

де

$$\Delta x_{ij} = x_{ij} - \bar{x}_i \quad (\text{Б.5})$$

Операція, що проведена над матрицею (Б.1), називається, як зазначалось вище, операцією центрування. Матриця коваріацій K_x визначається за таким матричним рівнянням:

$$K_x = \frac{1}{m} \Delta X \Delta X' , \quad (\text{Б.6})$$

де $\Delta X'$ - транспонована матриця центрованих величин. Покажемо справедливість цього твердження. За цією метою розглянемо більш просту матрицю

$$\Delta X = \begin{bmatrix} \Delta x_{11} & \Delta x_{12} & \dots & \Delta x_{1j} & \dots & \Delta x_{1m} \\ \Delta x_{21} & \Delta x_{22} & \dots & \Delta x_{2j} & \dots & \Delta x_{2m} \end{bmatrix} \quad (\text{Б.7})$$

Для такої матриці рівність (Б.6) в координатній формі прийме вигляд:

$$K_x = \frac{1}{m} \begin{bmatrix} \Delta x_{11} & \Delta x_{12} & \dots & \Delta x_{1j} & \dots & \Delta x_{1m} \\ \Delta x_{21} & \Delta x_{22} & \dots & \Delta x_{2j} & \dots & \Delta x_{2m} \end{bmatrix} \times \begin{bmatrix} \Delta x_{11} & \Delta x_{21} \\ \Delta x_{12} & \Delta x_{22} \\ \dots & \dots \\ \Delta x_{1j} & \Delta x_{2j} \\ \dots & \dots \\ \Delta x_{1m} & \Delta x_{2m} \end{bmatrix} = \quad (\text{Б.8})$$

$$= \begin{bmatrix} \frac{1}{m} \sum_{j=1}^m \Delta x_{1j}^2 & \frac{1}{m} \sum_{j=1}^m \Delta x_{1j} \Delta x_{2j} \\ \frac{1}{m} \sum_{j=1}^m \Delta x_{2j} \Delta x_{1j} & \frac{1}{m} \sum_{j=1}^m \Delta x_{2j}^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & K_{12} \\ K_{21} & \sigma_2^2 \end{bmatrix}.$$

Узагальнюючи проведені обчислювання, отримаємо:

$$K_x = \begin{bmatrix} \sigma_1^2 & K_{12} & \dots & K_{1j} & \dots & K_{1n} \\ K_{21} & \sigma_2^2 & \dots & K_{2j} & \dots & K_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ K_{i1} & K_{i2} & \dots & K_{ij} & \dots & K_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ K_{n1} & K_{n2} & \dots & K_{nj} & \dots & \sigma_n^2 \end{bmatrix} \quad (\text{Б.9})$$

Елемен

ти матриці (Б.9) розраховуються за формулами:

$$\sigma_i^2 = \frac{1}{m} \sum_{s=1}^m \Delta x_{is}^2, \quad (\text{Б.10})$$

$$K_{ij} = \frac{1}{m} \sum_{s=1}^m \Delta x_{is} \Delta x_{js} \quad (\text{Б.11})$$

Отже, як впливає з формул (Б.10) і (Б.11), на головній діагоналі матриці (Б.9) розташовуються дисперсії досліджуваної гідрометеорологічної величини σ_j . Порядковий номер дисперсії на діагоналі відповідає номеру гідрометеорологічної станції, якщо йдеться про гідрометеорологічні поля, або номеру предиктора, якщо досліджуються статистичні особливості предикторів при побудові моделі прогнозу. Інші елементи матриці (Б.9) – відповідні коваріації. Матриця коваріації має такі властивості:

її елементи є дійсними числами;

вона є симетричною;

матриця коваріацій є додатно визначеною.

З останньої властивості випливає, що $|K_x| > 0$. (Прямими дужками позначається визначник). До речі, матриця коваріацій, поряд з вектором математичних сподівань m_x , відіграє роль параметра щільності імовірностей багатовимірного нормального розподілу

$$f(X) = \frac{1}{(2\pi)^{n/2} |K_x|^{1/2}} \times \exp \left\{ -\frac{1}{2} (X - m_x) K_x^{-1} (X - m_x) \right\} \quad (\text{Б.12})$$

Маючи матрицю коваріацій, можна легко сформувати діагональну матрицю σ середніх квадратичних відхилень. Вона має вид:

$$\sigma = \begin{bmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_n \end{bmatrix} \quad (\text{Б.13})$$

Обернена матриця від діагональної матриці визначається таким чином

$$\sigma^{-1} = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_2} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{1}{\sigma_n} \end{bmatrix} \quad (\text{Б.14})$$

Якщо помножити ліворуч та праворуч матрицю коваріацій K_x на матрицю (Б.14) отримаємо матрицю кореляцій R_x

$$R_x = \sigma^{-1} K_x \sigma^{-1} \quad (\text{Б.15})$$

Матриця кореляцій порядку m має наступний вигляд

$$R_x = \begin{bmatrix} 1 & r_{12} & \dots & r_{1j} & \dots & r_{1n} \\ r_{21} & 1 & \dots & r_{2j} & \dots & r_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r_{i1} & r_{12} & \dots & r_{ij} & \dots & r_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r_{n1} & r_{n2} & \dots & r_{nj} & \dots & 1 \end{bmatrix}, \quad (\text{Б.16})$$

де r_{ij} - коефіцієнт кореляції, який характеризує тісноту лінійного зв'язку між двома змінними (i та j).

Коефіцієнт кореляції пов'язаний із коваріацією K_{ij} наступним співвідношенням

$$r_{ij} = \frac{K_{ij}}{\sigma_i \sigma_j} \quad (\text{Б.17})$$

При наявності рядів спостережень довжиною m , вибіркове значення коефіцієнту кореляції між двома змінними x_j та x_k розраховується таким чином

$$r_{jk} = \frac{\sum_{i=1}^m (x_{ij} - \bar{x}_j)(x_{ik} - \bar{x}_k)}{(m-1)\sigma_j \sigma_k}, \quad (\text{Б.18})$$

де x_{ij} - значення j -того ряду спостережень довжиною m ;

x_{ik} - значення k -того ряду спостережень довжиною m ;

$\bar{x}_j, \bar{x}_k$ - середні арифметичні значення для ряду j та ряду k ;

σ_j, σ_k - середні квадратичні відхилення для ряду j та ряду k ;

m - кількість спостережень.

При $k = j$ $r_{jj} = 1$, оскільки коваріація у такому випадку є дисперсією вихідного ряду j .

Абсолютне значення коефіцієнту кореляції змінюється від 0 до 1 ($0 \leq |r_{ij}| \leq 1$). Знак "-" означає існування оберненої залежності між двома змінними. Якщо $r_{ij} = 0$, то лінійного зв'язку між змінними не існує (величини не корельовані), якщо $r_{ij} = 1$, то існує функціональний зв'язок.

Матриця кореляцій має властивості аналогічні властивостям матриці коваріації, тобто вона дійсна симетрична і додатно визначена.

ДОДАТОК В

Випадкові похибки визначення статистичних параметрів за даними спостережень

Метод статистичних випробувань

Мірою точності обчислення статистичного параметра є середнє квадратичне відхилення або випадкова похибка оцінки цього параметра.

Формули для обчислення випадкових похибок визначення статистичних параметрів отримуються на основі методу статистичних випробувань або методу Монте – Карло [21]. Цей метод дозволяє моделювати процеси, на розвиток яких суттєво впливає випадковий фактор.

Моделювання виконується наступним чином.

1. Задається ряд випадкових чисел δ_i (наприклад, за спеціально розробленою таблицею випадкових чисел, де числа змінюються в інтервалі від 0 до 1, тобто $0 < \delta_i < 1$).
2. Отримані випадкові числа приймаються за забезпеченості, тобто $\delta_i = P_i$.
3. Величина стоку заданої забезпеченості визначається за вибраним законом розподілу

$$P(x) = \int_x^{\infty} f(x) dx. \quad (B.1)$$

З виразу (B.1) або за таблицею вибраного закону розподілу встановлюється шукане x . При цьому статистичні параметри розподілу задаються рівними відповідним параметрам генеральної сукупності.

Розглянемо випадок, коли як теоретичний закон розподілу вибраний закон розподілу Пірсона III. Це трипараметричний закон розподілу випадкової величини, для практичного застосування якого необхідно задати три параметри – математичне сподівання m_x , коефіцієнт варіації C_v , коефіцієнт асиметрії C_s . Припустимо, що ці параметри задані у вигляді конкретних чисел.

Як відомо, закон Пірсона III представляється у вигляді чисел Форстера Φ_P

$$\Phi_P = \frac{x_P - m_x}{\sigma_x}, \quad (\text{B.2})$$

звідки

$$x_P = m_x + \Phi_P \sigma_x, \quad (\text{B.3})$$

де P - забезпеченість.

За виразом (B.3) можна генерувати штучні ряди довжиною в декілька тисяч членів.

4. Після генерації довготривалого ряду величин стоку різної забезпеченості, який розглядається як генеральна сукупність, відбувається його розбиття на вибірки меншої довжини. Наприклад, якщо ряд довжиною 5000 членів розбити на вибірки довжиною в 100 елементів, то буде отримано 50 вибірок, а ,отже, і 50 значень оцінок статистичних параметрів.

5. Середня квадратична похибка визначення середнього арифметичного значення при довжині виборки $n=100$ буде визначатися за формулою

$$\sigma_{\bar{x}, n=100} = \sqrt{\frac{\sum_{i=1}^m (\bar{x}_j - m_x)^2}{m-1}}, \quad (\text{B.4})$$

$\sigma_{\bar{x}, n=100}$ - середня квадратична похибка визначення середнього арифметичного значення при довжині вибірки $n=100$;

$\bar{x}_j$ - середнє арифметичне, розраховане за j -тою вибіркою;

m – число вибірок.

Середнє квадратичне відхилення коефіцієнтів варіації, визначене по 200 вибірках довжиною в 20 елементів, розраховується таким чином

$$\sigma_{Cv, n=20} = \sqrt{\frac{\sum_{i=1}^m (Cv_j - Cv)^2}{m-1}}, \quad (\text{B.5})$$

де - $C_{V,n=20}$ - середня квадратична похибка визначення коефіцієнта варіації при довжині вибірки $n = 20$;

Cv_j - значення коефіцієнта варіації, розраховане за j -тою вибіркою;

m – число вибірок.

Середньоквадратичні відхилення статистичних параметрів при їхньому визначенні за вибірковими даними або так звані випадкові похибки можна розраховувати для вибірок різної довжини і у подальшому установити математичне співвідношення між похибками та довжиною вибірок $\sigma_{Cv} = f(n)$, або між похибками та середньоквадратичним відхиленням вихідних вибірок $\sigma_{Cv} = f(Cv)$. Саме методом статистичних випробувань і були отримані формули для розрахунків випадкових похибок визначення статистичних параметрів C_S та C_V за обмеженими у часі даними [16,17,33].

***Точність визначення статистичних параметрів,
обчислених за методом найбільшої правдоподібності***

Що стосується оцінки математичного сподівання, то її середньоквадратичне відхилення розраховується за формулою, розробленою для величин, які підлягають нормальному закону розподілу, тобто формула може бути отримана аналітичним шляхом, дотримуючись при цьому припущення, що нормальний закон розподілу вибірових середніх зберігається і для вибірок, які відхиляються від нормального розподілу

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}. \quad (B.6)$$

Похибки σ_{Cv} коефіцієнтів варіації C_V , розрахованих за методом моментів, обчислюються за формулою СН – 435-72

$$\sigma_{Cv} = Cv \sqrt{(1 + C_v^2) / 2n} \quad (B.7)$$

або за формулою Є. Г. Блохінова

$$\sigma_{C_v} = \frac{C_v}{n + 4C_v^2} \sqrt{\frac{n}{2}(1 + C_v^2)}. \quad (\text{B.8})$$

Середня квадратична похибка коефіцієнта асиметрії σ_{C_s} визначається за теоретичною формулою С.М.Крицького та М.Ф.Менкеля

$$\sigma_{C_s} = \sqrt{\frac{6}{n}(1 + 6C_v^2 + 5C_v^4)}, \quad (\text{B.9})$$

або формулою А.Ш. Резніковського

$$\sigma_{C_s} = \frac{1}{C_v} \sqrt{\frac{6}{n}}. \quad (\text{B.10})$$

Для визначення якості розрахунків статистичних параметрів за даними спостережень необхідно ввести критерій якості, а саме – допустиму похибку розрахунків. Допустиму похибку краще представляти у частках або відсотках від значення параметра, точність визначення якого оцінюється. Отже, і самі випадкові похибки зручніше представляти у відсотках, тобто

$$\varepsilon_{\bar{x}} = \frac{\sigma_{\bar{x}}}{\bar{x}} = \frac{\sigma_x}{\bar{x}\sqrt{n}} = \frac{C_v}{\sqrt{n}} \cdot 100\%. \quad (\text{B.11})$$

Із виразу (B.11) видно, що відносна випадкова похибка визначення середнього арифметичного значення пропорційна коефіцієнту C_v та зворотна числу членів вибірки n . Із збільшенням числа спостережень відносна похибка $\varepsilon_{\bar{x}}$ зменшується.

Для середніх арифметичних значень допустимою похибкою є 10%. Якщо фактична похибка розрахунку, визначена за формулою (B.11), перевищує 10%, то вихідний ряд вважається коротким, а визначений статистичний параметр (у даному випадку – середнє арифметичне) потребує уточнення. Уточнення може бути виконано

із залученням даних по річці-аналогу більшої довжини. Методи уточнення параметрів розрахункового ряду з використанням даних по річці – аналогу називаються методами приведення статистичних параметрів коротких рядів до довгого періоду.

Відносна випадкова похибка коефіцієнта варіації визначається таким чином

$$\varepsilon_{C_v} = \frac{\sigma_{C_v}}{C_v} \cdot 100\% . \quad (B.12)$$

Допустимою похибкою визначення коефіцієнта варіації є величина, що дорівнює 15%.

Відносна похибка визначення коефіцієнта асиметрії розраховується за виразом

$$\varepsilon_{C_s} = \frac{\sigma_{C_s}}{C_s} \cdot 100\% . \quad (B.13)$$

При практичному застосуванні формул (B.7) – (B.13) використовуються значення коефіцієнтів варіації, розрахованих за вибіркою.

З формул (B.6) – (B.11) можна знайти довжину рядів, при яких випадкова похибка дорівнює заданій величині. Так, при $C_v = 0.2-1.0$ та заданій похибці $\varepsilon_{C_v} = 15\%$ необхідні ряди спостережень довжиною 300- 350 років. Зазначимо, що гідрологічні ряди найчастіше мають довжину 40-50 років і лише окремі з них досягають довжини в 100 років й більше. У зв'язку з великою похибкою визначення за даними спостережень коефіцієнт C_s нормують за співвідношенням C_s/C_v у межах окремих районів.

***Точність визначення статистичних параметрів,
обчислених за методом найбільшої правдоподібності***

Середня квадратична похибка обчислення n -річних середніх арифметичних значень стокових рядів $\sigma_{\bar{x}}$ може бути обчислена за тією ж формулою, що й в методі моментів

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}. \quad (\text{B.14})$$

Середня квадратична похибка розрахунку коефіцієнта варіації C_V , отримана на основі статистичних випробувань, для умови $C_S/C_V = 2$ становить

$$\sigma_{C_V} = \frac{C_V}{\sqrt{2n}} \sqrt{\frac{3}{3 + C_V^2}}. \quad (\text{B.15})$$

Якщо $C_S / C_V \neq 2.0$, стандарт вибіркової оцінки C_V , визначеної методом найбільшої правдоподібності, корегується за допомогою поправкового коефіцієнта

$$K_{C_V} = \frac{\varphi_{C_V} (\text{при } C_V = mC_V)}{\varphi_{C_V} (\text{при } C_S = 2C_V)}, \quad (\text{B.16})$$

який знімається з графіка залежності K_{C_V} для вибірових оцінок C_V та співвідношення C_S/C_V (рис.В.1). Середнє квадратичне відхилення для C_S/C_V , визначеного за методом найбільшої правдоподібності, устанавлюється за зв'язком σ_{C_S/C_V} і C_V (рис.В.2). Обчислення відносних середньоквадратичних відхилень статистичних параметрів відбувається за формулами (В.7 – В.13).

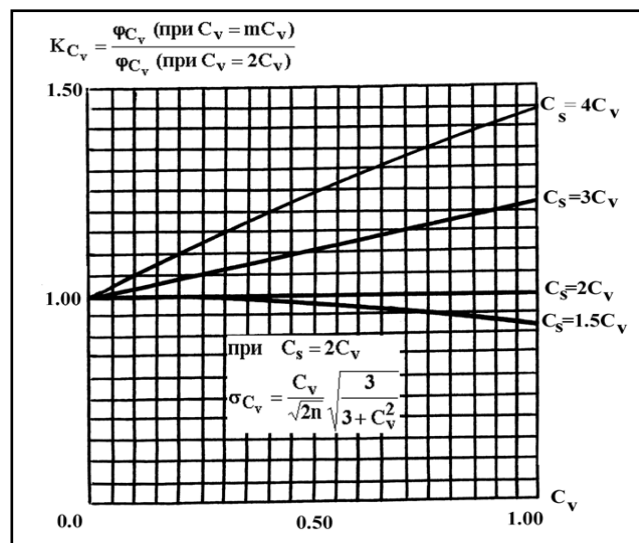


Рисунок В.1 – Графік для визначення поправкового коефіцієнта K_{C_V}

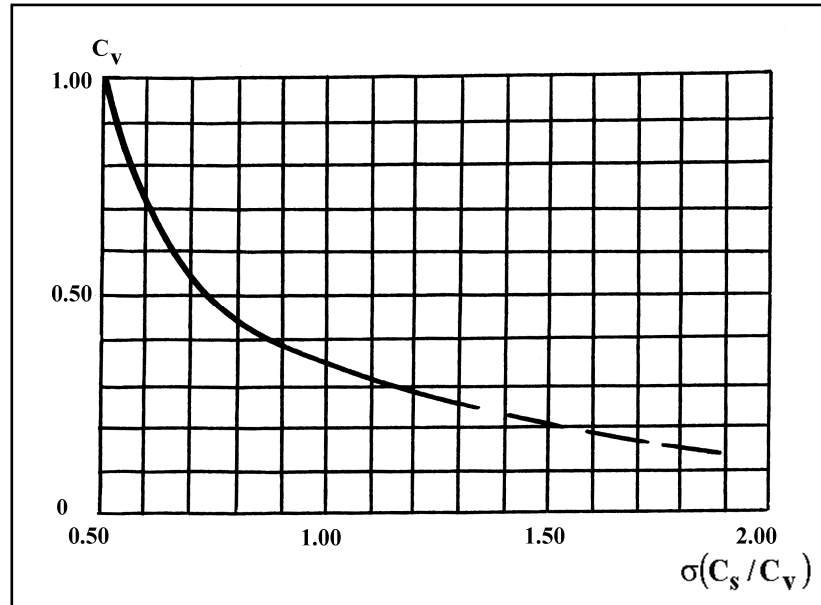


Рисунок В.2 - Графік для визначення σ_{C_s / C_v}

Додаток Д

ν_2	ν_1								
	12	14	16	20	24	30	40	50	75
10	2.9	2.9	2.8	2.8	2.7	2.7	2.7	2.6	2.6
11	2.8	2.7	2.7	2.7	2.6	2.6	2.5	2.5	2.5
12	2.7	2.6	2.6	2.5	2.5	2.5	2.4	2.4	2.4
13	2.6	2.6	2.5	2.5	2.4	2.4	2.3	2.3	2.3
14	2.5	2.5	2.4	2.4	2.4	2.3	2.3	2.2	2.2
15	2.5	2.4	2.3	2.3	2.3	2.3	2.2	2.2	2.2
16	2.4	2.4	2.3	2.3	2.2	2.2	2.2	2.1	2.1
17	2.4	2.3	2.3	2.3	2.2	2.2	2.1	2.1	2.0
18	2.3	2.3	2.3	2.3	2.2	2.1	2.1	2.0	2.0
19	2.3	2.3	2.2	2.2	2.1	2.1	2.0	2.0	2.0
20	2.3	2.2	2.2	2.1	2.1	2.0	2.0	2.0	1.9
21	2.3	2.2	2.2	2.1	2.1	2.0	2.0	1.9	1.9
22	2.2	2.2	2.1	2.1	2.0	2.0	1.9	1.9	1.9
23	2.2	2.1	2.1	2.0	2.0	2.0	1.9	1.9	1.8
24	2.2	2.1	2.1	2.0	2.0	1.9	1.9	1.9	1.8
25	2.2	2.1	2.1	2.0	2.0	1.9	1.9	1.8	1.8
26	2.2	2.1	2.1	2.0	2.0	1.9	1.9	1.8	1.8
27	2.1	2.1	2.0	2.0	1.9	1.9	1.8	1.8	1.8
28	2.1	2.1	2.0	2.0	1.9	1.9	1.8	1.8	1.8
29	2.1	2.1	2.0	1.9	1.9	1.9	1.8	1.8	1.7
30	2.1	2.0	2.0	1.9	1.9	1.8	1.8	1.8	1.7
32	2.1	2.0	2.0	1.9	1.9	1.8	1.8	1.7	1.7
34	2.1	2.0	2.0	1.9	1.8	1.8	1.7	1.7	1.7
36	2.0	2.0	1.9	1.9	1.8	1.8	1.7	1.7	1.7
38	2.0	2.0	1.9	1.9	1.8	1.8	1.7	1.7	1.6
40	2.0	2.0	1.9	1.8	1.8	1.7	1.7	1.7	1.6
42	2.0	1.9	1.9	1.8	1.8	1.7	1.7	1.6	1.6
44	2.0	1.9	1.9	1.8	1.8	1.7	1.7	1.6	1.6
46	2.0	1.9	1.9	1.8	1.8	1.7	1.7	1.6	1.6
48	2.0	1.9	1.9	1.8	1.7	1.7	1.6	1.6	1.6
50	2.0	1.9	1.9	1.8	1.7	1.7	1.6	1.6	1.6
100	1.9	1.8	1.8	1.7	1.6	1.6	1.5	1.5	1.4

Продовження додатку Д

ν_2	ν_1								
	12	14	16	20	24	30	40	50	75
200	1.8	1.7	1.7	1.6	1.6	1.5	1.7	1.4	1.4
∞	1.8	1.6	1.6	1.6	1.5	1.5	1.4	1.4	1.3

Значення критерію Фішера F для рівня значущості 0.05

ν_1 - число ступенів свободи для більшої дисперсії

ν_2 - число ступенів свободи для меншої дисперсії

Додаток Ж

Значення критерію Ст'юдента при рівні значущості 0.05

Число ступенів свободи	Рівень значущості 0.05	Число ступенів свободи	Рівень значущості 0.05
1	12.70	18	2.10
2	4.30	19	2.09
3	3.18	20	2.09
4	2.78	21	2.08
5	2.57	22	2.07
6	2.45	23	2.07
7	2.36	24	2.06
8	2.31	25	2.06
9	2.26	26	2.06
10	2.23	27	2.05
11	2.20	28	2.05
12	2.18	29	2.05
13	2.16	30	2.04
14	2.14	40	2.02
15	2.13	60	2.00
16	2.12	120	1.98
17	2.11		

ІМЕННИЙ ПОКАЖЧИК

А

Аронсон – (7)

Б

Бабаєва О.В. (2)

Бабкін В.І. -8

Берг Л. -8

Бершадський А.Г. (7)7

Бефані А.М. – 8

Бефані Н.Ф. – (4)

Блохінов Є.Г. (3)

Болгов М.В. -7, (8)

Борщ С.В. – (4)

Г

Глушков В.Г. –

Глушков О.В. – (7)

Д

Докучаєв В.

Є

Євстигнєєв В.М.

І

Іванов С.С. (7)

Ж

Жук В.А. (2)

К

Калінін Г.П. -8

Колмогоров А.М. (7)

Крицький С.М. – 7, (3), (3)

Кузін П.С. – 8

Л

Ланько А. – 8

Лобода Н.С. (2), (3), (4), (4), (4)

М

Мандельброт (7), (7)
 Марініч А. -8
 Марк – (7)
 Менкель М.Ф. -7, (3), (3),
 Міллер Р. -8

Н

Нгуєн Ву Ань – (4)

Р

Раткович Д.А. -7, (8)
 Резниковський А.Ш. –(8)
 Рождественський А.В. – 7,8, (8)
 Русов В.Д. (7)

С

Скляренко В.Л., (4), (4)
 Смірнов Н.П.-7, (4), (4)

Т

Том сон

Х

Хортон Р.Є.
 Христофоров А.В.

Ч

Чеботарьов А.І. (8)

Ш

Школьний Є.П. -7
 Шищенко П. -8

ПРЕДМЕТНИЙ ПОКАЖЧИК

А

Амплітудні функції
Автоковаріаційна функція
Автокореляційна функція

Б

Базисні функції

В

Вектор-стовпець
Вектор-рядок
Визначник квадратної матриці
Випадкова складова просторової дисперсії
Випадкова функція
Власний вектор
Власне значення матриці

Г

Географічна складова просторової дисперсії
Геометрична інтерпретація факторів

Д

Дискримінантний аналіз
Дискримінантна функція
Дискримінантна функція квадратична
Дискримінантна функція лінійна

Е

Ергодичність

З

Загальна дисперсія змінної
Забезпеченість допустимої похибки прогнозів

Задача фільтрації вихідної інформації
Залишкова дисперсія змінної

К

Коваріаційний момент
Кореляційний момент
Коефіцієнт множинної лінійної кореляції

Коефіцієнт кореляції

Критерій Стюдента

Критерій Фішера

Л

Лінійна парна регресія

Лінійне перетворення величин

М

Матриця

Матриця діагональна

Матриця квадратна

Матриця коваріацій

Матриця кореляцій

Матриця одинична

Матриця симетрична

Матриця скалярна

Матриця сингулярна

Метод варіацій Марка і Аронсона

Метод Монте-Карло

Метод головних компонентів

Метод найменших квадратів

Метод статистичних випробувань

Міnor матриці

Міра точності обчислення параметра

Множинна лінійна регресія

Монофрактальність

Мультифрактальність

О

Ортонормовані вектори

П

Повна складова просторової дисперсії

Пояснена дисперсія змінної

Предиктант

Предиктор

Природні ортогональні функції

Простий ланцюг Маркова

Профіль стоку

Р

Ранг матриці
Рівні матриці
Розпізнавання образів
Розв'язувальне правило

С

Самоподібність
Середня квадратична похибка
Синхронні коливання стоку
Синфазні коливання стоку
Слід матриці
Стаціонарна випадкова функція
Структурна функція
Сумісний аналіз даних

У

Умовне математичне сподівання

Ч

Частинний коефіцієнт кореляції
Число Махаланобіса

Ф

Факторний аналіз
Факторні навантаження
Факторів сумарний внесок
Факторизації міра
Флуктуаційна функція
Фрактал
Фрактальна розмірність