

**НАЦИОНАЛЬНАЯ АКАДЕМИЯ НАУК БЕЛАРУСИ  
МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ**

**Государственное научное учреждение  
«Институт проблем использования природных ресурсов и экологии»**

**Государственное научное учреждение  
«Отдел проблем Полесья»**

**Учреждение образования  
«Брестский государственный технический университет»**

**В. Ф. Логинов, А. А. Волчек, П. В. Шведовский**

**ПРАКТИКА ПРИМЕНЕНИЯ  
СТАТИСТИЧЕСКИХ МЕТОДОВ ПРИ  
АНАЛИЗЕ И ПРОГНОЗЕ ПРИРОДНЫХ  
ПРОЦЕССОВ**

**Брест 2004**

УДК 502.54.06:712+556.013(62-50)

ББК 20.18

Л 68

Рекомендовано к изданию Ученым Советом Института проблем использования природных ресурсов и экологии Национальной академии наук Беларуси

Рецензенты:

Доктор геолого-минералогических наук, профессор БГПУ им. М. Танка  
**М. Г. Ясовеев**

Кандидат физико-математических наук, старший научный сотрудник  
ИПИПРиЭ **В. С. Микуцкий**

**Логинов В. Ф., Волчек А. А., Шведовский П. В.**

**Л 68** Практика применения статистических методов при анализе и прогнозе природных процессов. Брест: Изд-во БГТУ, 2004. – 301 с, ил., табл.

**ISBN 985-6584-93-0**

В книге изложены теоретические основы и практика применения статистических методов при анализе и прогнозе природных процессов.

Наиболее продвинутыми являются вопросы статистического анализа в гидрометеорологии, поэтому в книге в подавляющем числе случаев используются примеры обработки гидрометеорологических данных.

Книга рассчитана на специалистов в области гидрологии и климатологии, природопользования, а также студентов и аспирантов соответствующих специальностей.

УДК 502.54.06:712+556.013(62-50)

ББК 20.18

© В. Ф. Логинов, 2004

© А. В. Волчек, 2004

© П. В. Шведовский, 2004

© Из-во БГТУ, 2004

**ISBN 985-6584-93-0**

*«Судьба людей и их воля не поддается расчетам. Однако математика может смягчить воздействие природы на человеческие судьбы и усилить эффект ответных реакций человечества»*

Д. Ван. Данцинг

## ПРЕДИСЛОВИЕ

Увеличение скорости протекания многих природных процессов в последние годы привело к нарушению природного равновесия. Это во многом обусловило неопределенность состояния природной среды и соответственно стратегию и тактику взаимодействия природы и человека.

Успешное решение этих проблем требует наличия высококвалифицированных специалистов, знакомых с основами и принципами оптимальных методов рационального природопользования и природообустройства.

По этим вопросам написаны сотни статей, книг и их поток увеличивается с каждым годом.

При решении проблем природопользования обычно сталкиваются с вопросами сбора, обработки, контроля и распространения данных.

Однако в нашей стране литература по анализу временных рядов, описывающих и характеризующих природные (гидрологические, климатические и др.) процессы в 70-80-ые годы прошлого столетия не только несколько устарела, но и стала библиографической редкостью. В то же самое время в области природопользования и природообустройства работает большое количество специалистов, не имеющих специальной математической подготовки.

Данная книга задумана как практическое пособие, в котором основной математический аппарат по мере возможности максимально упрощен, без большой потери математической строгости.

Авторы полагают, что эта работа поможет более широкому использованию математических методов в решении проблем взаимодействия человека с окружающей средой и с благодарностью примут все замечания по существу изложенной в книге системы применения статистических методов при анализе и прогнозе природных процессов.

## 1. ОСНОВНЫЕ ХАРАКТЕРИСТИКИ ГИДРОМЕТЕОРОЛОГИЧЕСКОЙ ИНФОРМАЦИИ И ЕЕ УПОРЯДОЧЕНИЕ ПО ПРОСТРАНСТВЕННЫМ И ВРЕМЕННЫМ ПРИЗНАКАМ

Гидрометеорологические данные – это значения физических величин, характеризующих состояние атмосферы и гидросферы.

В практике используются следующие типы данных:

*I* – числовые, целые (баллы облачности, метеорологической видимости, коды типов погоды);

*R* – числовые, действительные, необязательно целые (температура, давление и другие величины);

*C* – символные или текстовые (название пунктов наблюдений и измеренных параметров, сведения о формах облаков, типы погоды – текстуральные);

*B* – логические (сведения о наличии явлений).

Следует отметить, что в атмосфере наблюдается очень большая изменчивость практически всех процессов. А. С. Монин выделяет в спектре периодов колебаний девять интервалов.

1. Микрометеорологические колебания с периодами от долей секунды до минут.

2. Мезометеорологические колебания, соответствующие периодам от минут до часов.

3. Синоптические колебания с периодами от нескольких часов до нескольких дней.

4. Глобальные колебания с периодом от недели до месяцев.

5. Сезонные колебания – колебания с годовым периодом и их гармониками.

6. Междугодичные колебания.

7. Внутривековые колебания.

8. Междувековые колебания.

9. Ледниковые периоды плейстоцена.

Эта сводка масштабов показывает, насколько должен быть широк диапазон требований к точности, пространственному и временному разрешению наблюдений и длительности периодов, за которые необходимо собрать сведения об атмосфере.



К системам получения данных обычно предъявляются следующие требования:

- точность измерений;
- пространственное и временное разрешение;
- репрезентативность измерений;
- время сбора данных;
- полнота;
- достоверность измерений;
- стабильность (однородность измерений);
- надежность функционирования;
- экономичность функционирования.

Для определения требований к данным необходимо взять в расчет следующие обстоятельства.

1. Величина изменчивости зависит от пространственно-временных масштабов. Например, сезонная изменчивость больше межгодовой, а межгодовая для ряда характеристик по величине существенно больше изменчивости между соседними десятилетиями.

2. Существует определенная связь между пространственными и временными масштабами осреднения.

3. Имеется тесная связь между некоторыми характеристиками климатической системы. Это позволяет использовать такую связь для осуществления контроля и установления требований к точности данных.

4. Точность определения параметров наблюдательными системами, которые обеспечивают минимальную точность для единичных измерений (при большой случайной ошибке средней), можно увеличить, используя большее количество измерений.

5. Количество данных может быть оценено априори для различных регионов посредством оптимальной интерполяции метеорологических полей, если известна их автокорреляционная или ковариационная функция.

В зависимости от цели использования данных требования будут различаться. Основные цели это:

- верификация и развитие моделей;
- долговременный мониторинг медленно изменяющихся параметров, включая воздействующие на систему факторы;

- изучение специфических процессов (микротурбулентность) параметризации и прикладных задач (например, порывистость ветра).

Что касается точности исходных данных, то она зависит от ошибок измерительных приборов, ошибок метода измерений, ошибок интерполяции или осреднения во времени или пространстве. Ошибки в данные могут быть внесены в процессе передачи сообщений, обработки или хранения. Ошибки можно классифицировать следующим образом:

- по происхождению: аппаратная (инерция, неисправность), методическая, погрешность отсчета, погрешность градуировки;
- по характеру изменений: статическая, динамическая, дополнительная (отклонение от условий эксплуатации);
- по связи с величиной: аддитивная (не зависящая от истинного значения величины  $X$ ), мультипликативная (коррелирующая с истинным значением величины  $X$ ), погрешность нуля.

Если сигнал задан значением  $X^1$ , а ошибка  $\Delta$ , то истинное значение  $X$  связано с измеренным значением  $X^1$  при аддитивной ошибке соотношением  $X^1 = X + \Delta$ .

В общем случае мультипликативную ошибку можно рассмотреть как аддитивную, если записать  $X^1 = Xv_0 + \Delta$ :

$$\Delta = X(v - v_0) = X \cdot \xi. \quad (1.1)$$

При сложной системе измерений могут присутствовать оба типа ошибок, т. е.  $X^1 = Xv + \Delta$ .

На практике меру степени неопределенности состояния определяют энтропией. В качестве объекта, о котором передается информация, рассмотрим физическую величину  $X$ , которая может оказаться в том или ином состоянии, т. е. систему, которой заведомо присуща какая-то степень неопределенности.

Степень неопределенности определяется как число ее возможных состояний, так и вероятностями состояний. Например, сравним монету и игральную кость (2 и 6 состояний). Неопределенность второй системы больше. Другой пример. Устройство может быть в двух состояниях: 1) исправно (99%); 2) отказывает (1%). Такая система обладает меньшей неопределенностью, чем, допустим, монета, хотя и там и здесь – два состояния.

Очевидно, сведения, полученные в системе, будут тем ценнее и содержательнее, чем больше была неопределенность системы до получения этих сведений.

В качестве меры априорной неопределенности системы в теории информации применяется специальная характеристика, называемая энтропией.

*Энтропия системы*  $H(X)$  – это сумма произведений вероятностей различных состояний системы на логарифмы этих вероятностей, взятая с обратным знаком:

$$H(X) = - \sum_{i=1}^n P_i \cdot \log P_i. \quad (1.2)$$

Знак «-» перед суммой поставлен для того, чтобы энтропия была положительной (числа  $P_i$  меньше единицы и их логарифмы отрицательны).

Энтропия обращается в нуль, когда одно из состояний системы достоверно, а другие невозможны. При заданном числе состояний она обращается в максимум, когда эти состояния равновероятны.

Энтропия обладает свойством аддитивности, т. е. когда несколько независимых систем объединяются в одну, их энтропии складываются. Логарифм в формуле (1.2) может быть взят при любом основании,  $a > 1$ . Перемена основания равносильна простому умножению энтропии на постоянное число, а выбор основания равносильно выбору определенной единицы измерения энтропии. Если за основание выбрано число 10, то говорят о «десятичных» единицах энтропии, если 2 – то о «двоичных» единицах. На практике пользуются логарифмами при основании 2 и измеряют энтропию в «двоичных» единицах.

При выборе 2 в качестве основания логарифмов за единицу измерения энтропии принимают энтропию простейшей системы  $X$ , которая имеет два равновероятных состояния:

$X_1$	$X_1$	$X_2$
$P_1$	1/2	1/2

Определенная таким образом единица энтропии называется «двоичной единицей», а иногда – *bit* (binary digit – двоичный знак). Это энтро-

пия одного разряда двоичного числа, если оно с одинаковой вероятностью может быть «0» или «1».

*Байт* – минимальная единица количества информации, содержащаяся в 8 двоичных разрядах (битах). Байт – единица памяти, равная 8 битам.

Каждому алфавитному, цифровому и специальному символу соответствует своя кодовая комбинация. Из байтов составляют более крупные единицы информации – слова, тексты.

Если система имеет  $n$  равновероятных состояний, то

$$H(X) = -n \cdot \frac{1}{n} \cdot \log \frac{1}{n} = -\log 1 + \log n = \log n, \quad (1.3)$$

т. е. энтропия системы с равновероятными состояниями равна логарифму числа состояний. Например, для системы с восемью состояниями  $H(X) = \log 8 = 3$ .

Энтропия сложной системы  $(X, Y)$  равна сумме произведений вероятностей возможных ее состояний на их логарифмы с обратным знаком:

$$H(X, Y) = -\sum_{i=1}^n \sum_{j=1}^m P_{ij} \cdot \log P_{ij}. \quad (1.4)$$

При объединении двух независимых систем их энтропии складываются:

$$H(X, Y) = H(X) + H(Y). \quad (1.5)$$

Чтобы найти энтропию системы, составленной из зависимых элементов, используют понятие условной энтропии. Обозначив  $P(y_j / x_i)$  условную вероятность того, что система  $Y$  примет состояние  $y_j$  при условии, если система  $X$  находится в состоянии  $x_i$ :

$$P(y_j / x_i) = P(Y \sim y_j / X \sim x_i). \quad (1.6)$$

Условная энтропия системы  $Y$  при условии, что система  $X$  находится в состоянии  $x_i$ , может быть определена по формуле:

$$H(Y / x_i) = -\sum_{j=1}^m P(y_j / x_i) \cdot \log P(y_j / x_i). \quad (1.7)$$

Средняя (полная) условная энтропия системы  $Y$  относительно  $X$ :

$$H(Y / X) = -\sum_{i=1}^n \sum_{j=1}^m P_{ij} \cdot \log P(y_j / x_i). \quad (1.8)$$

Рассмотрим на примере выявление неопределенностей. Пусть из многолетних наблюдений известно, что вероятность того, что 15 июня будет идти дождь, равна 0,4, а дождя не будет – 0,6 ( $\alpha_1$ ); 15 ноября вероятность дождя 0,65, вероятность снега 0,15, а отсутствие осадков – 0,2 ( $\alpha_2$ ).

В какой из перечисленных дней погоду в рассматриваемом пункте следует считать неопределенной?

Опыт $\alpha_1$	Исходя из опыта	Дождь	Отсутствие осадков
	Вероятность	0,4	0,6
Опыт $\alpha_2$	Исходя из опыта	Дождь	Отсутствие осадков
	Вероятность	0,65	0,2

Энтропии двух опытов равны:

$$H(\alpha_1) = -0,4 \log 0,4 - 0,6 \log 0,6 \approx 0,97 \text{ бита};$$

$$H(\alpha_2) = -0,65 \log 0,65 - 0,15 \log 0,15 - 0,2 \log 0,2 \approx 1,28 \text{ бита}.$$

Следовательно, погоду 15 ноября следует считать более неопределенной, чем 15 июня.

## **2. СИСТЕМЫ СБОРА И МЕТОДЫ (ТЕХНОЛОГИИ) ОБРАБОТКИ ГИДРОМЕТЕОРОЛОГИЧЕСКИХ ДАННЫХ**

### **2.1. Общие принципы**

Системы сбора гидрометеорологических данных делятся на оперативные, неоперативные и системы сбора данных специальных экспериментов.

Глобальной и беспрецедентной по объему сбора гидрометеорологических данных является Всемирная служба погоды (ВСП).

Всемирная служба погоды охватывает все страны и континенты, моря и океаны, а также космическое пространство. Она представляет собой согласованную систему пунктов и средств наблюдений, а также систему взаимосвязанных центров обработки. Центры обработки в структуре ВСП по своим функциям делятся на мировые, региональные и территориальные.

Рассмотрим элементы ВСП.

**I. Глобальная система наблюдений (ГСН).** Основу наземного компонента ГСН создают около 9900 синоптических станций на суше и около 6700 морских станций, которые через каждые 3 ч осуществляют наблюдения приземных метеорологических характеристик и явлений погоды. Свыше 1000 дрейфующих буев и 300 заякоренных буев проводят метеорологические наблюдения в приземном слое атмосферы над океанами. Около 1000 аэрологических станций дают вертикальные профили геопотенциала, температуры, влажности воздуха, скорости и направления ветра. Пять геостационарных и пять полярно-орбитальных спутников обеспечивают метеорологические наблюдения за состоянием глобальной атмосферы.

Основу наземной сети составляют наземные и морские, автоматические и обслуживаемые персоналом синоптические станции. Они проводят наблюдения за температурой и влажностью воздуха, температурой почвы (воды), атмосферным давлением, скоростью и направлением ветра, дальностью горизонтальной видимости, формой, количеством облаков и за высотой их нижней границы, видом, количеством и интенсивностью осадков, атмосферными явлениями. Наблюдения на синоптических станциях, входящих в международную сеть, проводятся синхронно в основ-

ные (00, 06, 12, 18 ч гринвичского среднего времени) и дополнительные (03, 09, 15, 21 ч) синоптические сроки. Погрешность измерения температуры находится в пределах 0,1 – 0,3 °С, давление – 0,1 – 0,2 гПа; средние скорость и направление ветра – 0,5 м/с и 5°.

Наземная система наблюдений включает обсерватории, станции и посты наблюдений за состоянием природной среды. Как правило, они ведут наблюдения по согласованной методике. Наиболее развитой системой наземных наблюдений является система гидрометеорологических наблюдений, координируемая ВМО. Эта система включает десятки тысяч обсерваторий, станций и постов, которые осуществляют наблюдения за состоянием воздушной и водной оболочек Земли, а также за температурным и влажностным режимами почв.

При этом стационарные исследования проводятся в заповедниках, стационарах и полигонах различных ведомств. Их преимущества – полнота и непрерывный характер, а недостаток – локальность и выборочность. Экспедиционные же исследования существенно расширяют рамки стационарных изменений, дают дополнительную информацию.

Круглосуточные восьмисрочные метеорологические наблюдения на территории Республики Беларусь, непрерывные наблюдения за опасными и стихийными гидрометеорологическими явлениями и передачу информации в Гидрометцентр в настоящее время осуществляют синоптические станции Минской области – Березино, Борисов, Вилейка, Воложин, Любань, Марьина Горка, Минск, Нарочь, Слуцк, Столбцы; Брестской области – Барановичи, Брест, Высокое, Ганцевичи, Дрогичин, Ивацевичи, Пинск, Полесская, Пружаны; Гродненской области – Волковыск, Гродно, Лида, Новогрудок, Ошмяны, Щучин; Могилевской области – Бобруйск, Горки, Кличев, Костюковичи, Могилев, Славгород; Гомельской области – Брагин, Василевичи, Гомель, Житковичи, Жлобин, Лельчицы, Мозырь, Октябрь, Чечерск; Витебской области – Верхнедвинск, Витебск, Докшицы, Езерище, Лепель, Лынтупы, Орша, Полоцк, Толочин, Шарковщина.

Сеть гидрометеорологических наблюдений Беларуси включает сеть станций Всемирной метеорологической организации (ВМО) (35), станции МАГАТЭ (8), станции, передающие информацию по РОСС (10) и региональному обмену (8), станции, передающие информацию в Росгид



дромет (28), передающие телеграммы «Климат» в ВМО (6) и Росгидромет (12), используемые в региональных целях (8).

Сводки с результатами синоптических наблюдений передают в международных кодах SYNOP (для станций на суше), SHIP (для судовых морских станций) и BUOY (для станций на дрейфующих и заякоренных буях).

*Аэрологические станции* проводят вертикальное зондирование атмосферы с помощью радиозондов, позволяющих получать систематические массовые данные о давлении, температуре, влажности, скорости и направлении ветра до высот 30 км. Радиозондирование атмосферы в силу меньшей временной изменчивости метеорологических величин и свободной атмосфере проводят реже. Основными сроками наблюдений на сети аэрологических станций являются 00 и 12 ч гринвичского среднего времени, а дополнительными – 06 и 18 ч.

В настоящее время в ведении Гидрометцентра Беларуси имеются два аэрологических комплекса – «АВК-1» в Гомеле и новый российский комплекс «Бриз», созданный на основе современных информационных технологий.

**2. Глобальная система телесвязи (ГСТ).** Состоит из устройств и средств, необходимых для быстрого сбора и распределения требуемых данных наблюдений и обработанной информации.

**3. Глобальная система обработки данных (ГСОД).** Состоит из метеорологических центров, оснащенных средствами обработки данных для оперативного использования, хранения и поиска данных для неоперативного использования.

Глобальная система обработки данных базируется на трех мировых метеорологических центрах (Вашингтон, Москва, Мельбурн), на 23 региональных метеорологических центрах и на национальных метеорологических центрах (рис. 2.1).

Всего на территории СНГ работают четыре региональных гидрометеорологических центра – в Москве, Новосибирске, Ташкенте и Хабаровске. Московский центр одновременно выполняет роль и мирового центра гидрометеорологических данных в системе, созданной ВМО.

Важнейшим звеном системы автоматизированной обработки данных являются *региональные гидрометеорологические центры* (РГМЦ).



Каждый из них ответственен за сбор и обработку всей гидрометеорологической информации по определенной территории (региону).

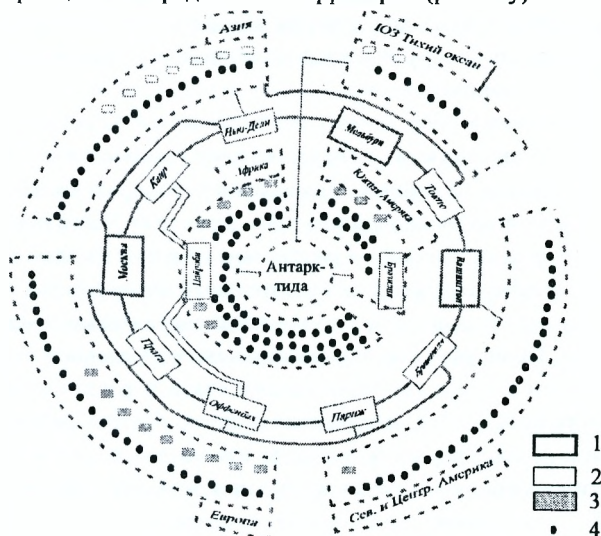


Рис. 2.1. Диаграмма центров обработки данных главной магистральной цепи телесвязи: 1 – мировые метеорологические центры; 2 – региональные центры на главной магистральной цепи и ее ответвлениях; 3 – остальные региональные центры, изображенные в своих региональных ассоциациях; 4 – национальные метеорологические центры (число символов отвечает количеству центров обработки)

Наряду с информацией своего района (региона) каждый региональный центр собирает данные по некоторым сопредельным территориям, а также такую глобальную информацию, как спутниковая. Региональные центры осуществляют взаимный обмен некоторыми видами данных.

*Территориальные центры* осуществляют основную часть первичной обработки наблюдений с последующей передачей в региональные центры лишь некоторых обобщенных характеристик. Они руководят работой зональных гидрометеорологических обсерваторий, которые предназначены для проведения расширенного комплекса метеорологических наблюдений. Зональные обсерватории непосредственно участвуют в ус-

тановке и внедрении технических средств получения и передачи информации на сети станций, обеспечивают их бесперебойную работу.

Регламентами ВСП для центров каждого типа определены обязательные функции по сбору и передаче данных, их обработке и распространению обработанных данных.

Неоперативный сбор данных осуществляется национальными центрами и по соглашению между государствами производится обмен накопленными комплектами данных. Неоперативно собираются данные, которые не передаются по глобальной системе телесвязи. Часть передаваемых по системе телесвязи данных по различным обстоятельствам не доходит до адресата, поэтому создается дополнительный поток данных (например, пересылка по почте) для полного укомплектования накапливаемых массивов данных. Таким образом массив наполняется в среднем на 30%.

Кроме того, существующие каналы связи и регламенты их функционирования не для всех собираемых данных обеспечивают необходимую точность и достоверность, которые нужны для подготовки справочников по гидрометеорологическому режиму или для оценки изменений климата. В этом случае организуется неоперативный сбор требуемых данных.

Задачи исследований Земли подталкивают ученых к организации и проведению специальных международных исследовательских программ, одной из основных задач которых является получение и сбор максимального больших комплектов данных по геофизическим процессам.

В настоящее время выполняются следующие программы и эксперименты:

- 1) всемирная климатическая программа (WCP);
- 2) эксперимент по изучению циркуляции вод Мирового океана (WOCE)
- 3) эксперимент по изучению Тропического океана и глобальной атмосферы (TOGA);
- 4) стратосферные процессы и их роль в формировании климата (SPARC);
- 5) программа по изучению климатической изменчивости и предсказуемости (CLIVAR);

б) эксперимент по изучению энергетического и гидрологических циклов в глобальном масштабе (GEWEX).

Сегодня на основе современных космических технологий создана измерительная аппаратура, способная проводить измерения атмосферных характеристик с необходимой точностью и пространственным разрешением.

При этом геостационарные спутники, занимая фиксированное географическое положение над земной поверхностью, проводят наблюдения каждые 20 – 30 мин, что позволяет непрерывно наблюдать за определенными секторами земной атмосферы и подстилающей поверхности.

Благодаря короткому интервалу времени между последовательными снимками геостационарные спутники дают ценную информацию для мониторинга и прогноза неблагоприятных и опасных гидрометеорологических явлений.

В Беларуси широко используется система сбора данных дистанционного зондирования Земли из космоса через российский космический комплекс обзорного мониторинга Земли и детального мониторинга территории.

Основные технические характеристики снимков, получаемых с космических аппаратов подсистемы обзорного мониторинга, представлены в табл. 2.1.

Основные характеристики фотографических систем для детального мониторинга территорий приведены в табл. 2.2.

Космические снимки дают возможность определять площадь, периметр, количество, частоту встречаемости, длину, ширину разнообразных природных и антропогенных объектов. Косвенные дешифровочные признаки в отдельных случаях позволяют получать также информацию о состоянии земельных ресурсов, запасов кормов, плодородии почв, качества земель и др. При этом наибольший эффект достигается при использовании их в комплексе с традиционными источниками: картографическими, фондовыми, статистическими литературными.

Трансформированные и фотометрические преобразованные растровые изображения космической фотосъемки (КФС) являются также и основой для создания карты современного состояния территории (ландшафтов).

Таблица 2.1

## Технические характеристики космической аппаратуры

ИСЗ «Метсор-3»						
Аппаратура	сканирующая ТВ-аппаратура с системой за поминания данных на борту для режима глобального обзора	сканирующая ТВ-аппаратура для режима передачи данных	ИК-радиометр для режима глобального обзора и передачи данных	радиационно-метрический комплекс		
Спектральный диапазон	0,5-0,8 мкм	0,5-0,8 мкм	10,5-12,5 мкм	0,17-600 МэВ		
Ширина полосы захвата при высоте 900 км	3100км	2600км	3100км	-		
Разрешение, км	0,7x1,4	1x2	3x3	-		
Режим работы	Запись информации на бортовое запоминающее устройство с последующей передачей на наземные станции приема; передача информации в масштабе реального времени.					
ИСЗ «Океан-О» (серийный номер «Космос-1 766»)						
Аппаратура	РЛС БО	сканирующий СВЧ-радиометр РМ-08	МСУ-М	МСУ-С	МСУ-СК	несканирующий трассовый СВЧ-поляриметр
Спектральный диапазон	3,2 см	0,8см	0,5-0,6 мкм 0,6-0,7 мкм 0,7-0,8 мкм 0,8-1,1 мкм	0,6-0,7 мкм 0,8-1,1 мкм	0,8-1,1 мкм	3см
Разрешение, км	1200x1500	15000x20000	1900	370	500	12° вдоль орбиты, 6° перпендикулярно орбите
Полоса обзора, км	450	550	1900	1100	1150	-

продолжение табл. 2.1

ИСЗ «Ресурс-О» (серийный номер «Космос - 1 989»)			
Аппаратура	оптико-электронный сканер высокого разрешения с плоскостной разверткой МСУ-Э	5-канальное сканирующее устройство среднего разрешения МСУ-СК	2-канальное сканирующее устройство среднего разрешения МСУ-С
Спектральный диапазон	0,5 – 0,6 0,6 – 0,7 0,8 – 0,9	0,5 – 0,6 0,6 – 0,7 0,7 – 0,8 0,8 – 1,1 10,4 – 12,6	0,5 – 0,7 0,7 – 1,0
Разрешение, км	45x30	170,600	250
Полоса обзора, км	45 в зоне обзора 600-700 км (80 в режиме работы одновременно двух приборов)	600	1400

Материалы КФС позволяют создавать различные карты масштабов от 1:5000 до 1:1000000. Цифровые карты масштабов 1:100000 – 1:1 000000 обеспечивают региональный уровень территориальных ГИС, а масштабов от 1:5000 до 1:50000 – локальный.

В зависимости от площади территории и ее хозяйственной освоенности картографическая база территориальных ГИС регионального уровня имеет различное информационное наполнение. Используя контурную основу карты современного состояния территории и ее содержательную часть, создаются согласованные констатирующие цифровые тематические карты:

- гидрографии;
- генетических типов и форм рельефа;
- литологии;
- растительного покрова;
- уровня и минерализации грунтовых вод;
- современного использования земель.

Таблица 2.2

### Характеристика фотографических систем для детального мониторинга территорий

Характеристики	«Ресурс-Ф1»		«Ресурс-Ф2»	ТКК	
	Рабочий диапазон высоты фотографирования (Н), км	180-400		200-450	200-250
Наклон орбиты, град	82,3		82,3	65(71)	
Тип камеры	КФА-1000 (2 камеры)	КАТЭ-200 (3 камеры)	МК-4	ТК-350	КВР-1000
Фокусное расстояние, мм	1000	200	300	350	1000
Размер кадра, мм	300x300	180x180	180x180	300x450	180x720
Число диапазонов	1	3	4 (из 6)	1	1
Спектральный диапазон, нм: Для черно-белых пленок	570-760	500-600 600-700 700-900	460-505 515-565 635-690 810-900	580-720	580-720
Для цветной пленки Для СПЗ пленки	— 560-810	— —	400-700 580-800	— —	— —
Масштаб снимков	1:1000Н	1:5000Н	1:3333Н	1:2857Н	1:1000Н
Разрешение на местности, м: На черно-белой пленке На СПЗ, цветной	4-6 5-8	15-30 —	6-12 10-15	8-12 —	2-3 3-4
Перекрытие, %	60	60	60	80	10

Кроме того, в картографическую базу ГИС входят и цифровые карты, основу которых составляют тематические и топографические карты, создаваемые дистанционными и традиционными методами.

В результате анализа цифровых тематических карт с помощью ГИС-технологий создаются производные карты:

- зон затопления половодьями рек;
- устойчивости ландшафтов к антропогенному воздействию;
- рекомендуемых природоохранных мероприятий;
- функционального зонирования территорий.

Создаваемая картографическая база является исходным материалом при мониторинге территорий для фиксации последующих изменений природной среды и проявления результатов хозяйственной деятельности.

Ведение метеорологических наблюдений на территории Беларуси имеет достаточно давнюю историю. Первые эпизодические наблюдения в Беларуси проводились еще в начале XIX столетия (1804 г. – Могилев, 1810 г. – Витебск, 1834 г. – Брест, 1841 г. – Горки, 1849 г. – Минск). Реально имеющиеся в настоящее время данные наблюдений имеют, однако, более позднее начало. По температуре воздуха доступны данные в основном с 1881 г., по осадкам – с 1890 г. К тому же ряды наблюдений имеют большое количество пропусков в отдельные периоды времени. Так, для такой длиннорядной станции, как Василевичи, данные наблюдений температуры и осадков имеют пропуски за отдельные дни в 1883, 1900, 1935 и 1937 гг., а также большие пропуски за 1918 – 1923 и 1941 – 1944 гг. Похожая ситуация существует и по другим длиннорядным станциям. Кроме того, с течением времени менялось количество и расположение метеорологических станций. В результате при сравнении рядов наблюдений или их осреднении по времени и по пространству возникает проблема неоднородности данных.

Начиная с первой половины XX столетия, сеть метеостанций отличается достаточной густотой и равномерностью (данные по метеостанциям на территории Беларуси приведены в табл. 2.3 и на рис. 2.2).

Таблица 2.3

### Перечень действующих станций по областям

Станция	Начало работы станции, год	Высота станции над уровнем моря, м
1	2	3
Витебская область		
Верхнедвинск	1906	132



продолжение табл. 2.3

1	2	3
Езерище	1956	172
Полоцк	1901	132
Щарковщина	1940	130
Витебск	1886	174
Лынтупы	1945	208
Докшицы	1950	193
Лепель	1934	173
Сенно	1937	173
Орша	1895	186
Толочин	1937	208
Минская область		
Вилейка	1940	163
Борисов	1894	188
Воложин	1946	229
Минск	1887	222
Березино	1937	153
Столбцы	1966	172
Марьино Горка	1894	175
Слуцк	1934	159
Любань	1986	140
Гродненская область		
Ошмяны	1962	211
Лида	1926	154
Гродно	1894	133
Щучин	1962	177
Новогрудок	1929	278
Волковыск	1940	180
Могилевская область		
Горки	1881	200
Могилев	1887	192
Кличев	1958	154
Славгород	1938	171
Костюковичи	1937	168
Бобруйск	1901	156
Брестская область		
Барановичи	1940	192
Ганцевичи	1944	157
Ивацевичи	1940	153
Пружаны	1929	162



продолжение табл. 2.3

1	2	3
Высокос	1951	163
Дрогичин	1984	150
Полесская	1947	133
Брест	1888	142
Пинск	1881	140
<b>Гомельская область</b>		
Жлобин	1931	140
Чечерск	1938	153
Октябрь	1958	141
Гомель	1927	125
Василевичи	1881	139
Житковичи	1929	136
Мозырь	1944	185
Лельчицы	1935	141
Брагин	1945	115

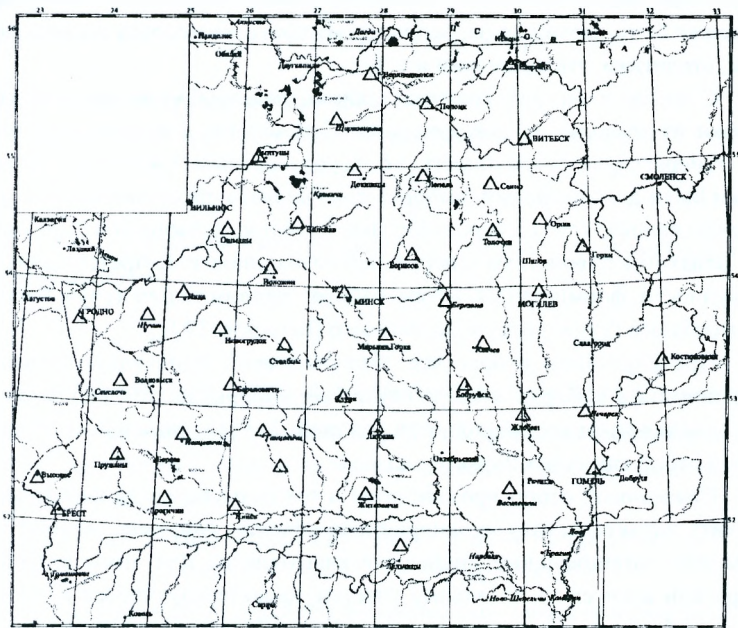


Рис. 2.2. Схема расположения метеостанций на территории Беларуси

## 2.2. Теоретические основы и методы статистической обработки и анализа данных

Так как абсолютно все природные процессы характеризуются тем, что их будущее течение принципиально нельзя предсказать однозначно, а можно определить только вероятность события, то важнейшей задачей статистической обработки является установление (выявление) ряда статистических параметров, которые в комплексе достаточно полно характеризуют свойства исследуемой генеральной совокупности, т.е. всех возможных данных о процессах и явлениях, например, связанных с функционированием геосистем, объектов или ландшафтных комплексов. При этом число членов в генеральной совокупности может быть как конечным, так и бесконечным.

Например, конечным числом членов генеральной совокупности являются все реки, по которым исследуется максимальный сток воды весеннего половодья. Бесконечным числом членов генеральной совокупности может быть расход воды, величина которого колеблется по месяцам, годам, столетиям, тысячелетиям и т. д.

С целью экономии времени анализа, целесообразно формировать выборки из *генеральной совокупности*. С помощью выборки легко оценить генеральную совокупность по вероятностным свойствам. Чтобы оценки были достоверными, выборка должна быть представительной (репрезентативной), т. е. ее вероятностные свойства должны совпадать или быть близкими к свойствам генеральной совокупности. Репрезентативные совокупности формируются путем отбора: рандомизации (случайного), направленного (типичного) и смешанного [Бочаров, 1971].

При случайном отборе все объекты имеют одинаковую возможность попасть в выборку. Иногда случайная выборка не отвечает задачам исследования из-за неоднородности условий. Тогда производится направленный отбор типичных характеристик.

Смешанный отбор производится в тех случаях, когда необходимо получить характеристику неоднородного объекта, например, водные ресурсы или потенциальные почвенные влагозапасы. Для этого водосбор или регион делится на локальные участки, характеризующиеся однородными свойствами, например, типами почв. Для каждого выделенного

участка производится случайный отбор проб, и полученные величины объединяются в одну выборку.

Наиболее полной характеристикой случайной величины (как дискретной, так и непрерывной) является закон (функция) ее распределения ( $f(x)$ ), т. е. соотношение, устанавливающее связь между возможными значениями случайной величины ( $x$ ) и соответствующими им вероятностями ( $p(x)$ ).

Однако при решении некоторых прогнозно-практических задач чаще всего прибегают не к законам распределения, а к числовым (статистическим) характеристикам случайной величины, которая в сжатой форме выражает наиболее существенные особенности ее распределения.

Важнейшими числовыми (статистическими) характеристиками являются: *математическое ожидание* ( $\mu$ ) и *дисперсия* ( $\sigma^2$ ) случайной величины ( $x$ ). *Математическое ожидание* – значение переменной, вокруг которого сгруппирована выборка, определяется интегралом

$$\mu = \int_{-\infty}^{\infty} x \cdot \varphi(x) \cdot dx, \quad (2.1)$$

где  $\varphi(x)$  – плотность распределения величины  $x$ .

Для дискретной конечнозначной случайной величины

$$\mu = \sum_{i=1}^n p_i \cdot x_i, \quad (2.2)$$

где  $x_i$  и  $p_i$  – отдельные значения и соответствующие им вероятности случайной величины

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad i=1, n, \quad (2.3)$$

Следовательно, понятие математического ожидания совпадает с понятием среднего арифметического, ( $\bar{x}$ ) представляет эмпирическое или выборочное среднее. При этом математическое ожидание имеет следующие основные свойства:

$$\left. \begin{aligned} \mu(c) = c; \mu(x+c) = \mu(x) + c; \mu(x+y) = \mu(x) + \mu(y); \\ \mu(c \cdot x) = c \cdot \mu(x); \mu(x \cdot y) = \mu(x) \cdot \mu(y). \end{aligned} \right\} (2.4)$$

*Дисперсия* ( $\sigma^2$ ) случайной величины это – мера «разброса» или «разброса» выборки:

для непрерывной случайной величины  $x_i$  –

$$\sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 \cdot \varphi(x) \cdot dx ; \quad (2.5)$$

для конечнозначной случайной величины, имеющей  $N$  – значений, –

$$\sigma^2 = \sum_{i=1}^n p_i \cdot (x_i - \bar{x})^2 ; \quad (2.6)$$

*дисперсия выборочной совокупности (выборочная дисперсия)*, состоящей из  $n$  – значений случайной величины, вычисляется по формуле

$$\bar{S}^2 = \frac{1}{n-1} \cdot \sum_{i=1}^n (x_i - \bar{x})^2 . \quad (2.7)$$

Дисперсия является удобной естественной мерой рассеивания случайной величины, поскольку в равной степени учитывает отклонения отдельных величин от средней и одновременно усредняет их по всем результатам.

При этом дисперсия характеризуется следующими свойствами:

$$\left. \begin{aligned} \sigma^2(c) &= 0; \quad \sigma^2(x+c) = \sigma^2(x); \quad \sigma^2(c \cdot x) = c^2 \cdot \sigma^2(x); \\ \sigma^2(-x) &= \sigma^2(x); \quad \sigma^2(x+y) = \sigma^2(x) + \sigma^2(y). \end{aligned} \right\} \quad (2.8)$$

Выборочное среднеквадратическое отклонение может быть найдено по формуле:

$$\bar{S} = \sqrt{\bar{S}^2} . \quad (2.9)$$

Из других выборочных характеристик чаще всего используется *коэффициент вариации*  $v = S / \bar{x}$ , являющийся мерой относительной изменчивости наблюдаемой случайной величины.

Кроме того, часто используются *статистические моменты третьего и четвертого порядков*:

$$\left. \begin{aligned} m_3 &= \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^3 ; \\ m_4 &= \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^4 . \end{aligned} \right\} ; \quad (2.10)$$

Третий центральный момент служит характеристикой асимметрии, а четвертый – характеристикой «крутости», т.е. островершинности распределения.

Однако все сказанное выше относится к *равноточным* измерениям и наблюдениям, т. е. к измерениям, которые содержат только случайную погрешность, подчиняющуюся закону нормального распределения вероятностей. В практике обычно пользуются результатами измерений различного происхождения, в т. ч. выполненными разными инструментами и разноточными методами. Результаты таких измерений называются *неравноточными*. В этом случае при статистической обработке данных необходимо использовать метод получения средневзвешенных величин из имеющихся в наличии серий наблюдений [Тейлор, 1985].

**В качестве примера** выполним статистический анализ величин атмосферных осадков в контексте изучения режима их выпадения на территории Беларуси. Для этого располагаем в алфавитном порядке все метеостанции и, используя таблицу «Случайные числа», выбираем 100 метеопунктов с соответствующими им годовыми нормами атмосферных осадков. Начав с любой колонки случайных чисел и двигаясь по столбцам сверху-вниз (снизу-вверх), выписываем те первые (последние) цифры четырехзначного числа, которые по величине укладываются в общее число метеопунктов и будут представлять номера метеопунктов, включаемых в формируемую выборку (табл. 2.4).

Таблица 2.4

**Выборка годовых норм атмосферных осадков на территории  
Беларуси, мм**

736	719	788	719	735	720	695	813	814	739
731	777	787	746	733	785	799	763	786	781
753	824	736	792	763	772	753	721	769	699
722	782	716	734	800	760	756	731	707	686
706	695	737	768	784	758	658	684	676	744
841	784	756	763	750	751	727	812	731	828
747	731	690	836	749	773	625	774	726	679
753	667	713	663	758	768	655	708	703	661
656	630	658	638	677	673	676	706	716	737
741	758	696	684	747	642	671	739	674	624

Составив выборку годовых норм осадков для региона, определим основные характеристики данной выборки (табл. 2.5).

Таблица 2.5

**Выборочные характеристики распределения атмосферных осадков на территории Беларуси**

$\bar{x}$	$S^2$	$\bar{S}^2$	$S$	$\bar{S}$	$m_1$	$m_2$	$m_3$	$m_4$	$v$
732,2	2446,6	2471,3	49,5	49,7	0	2446,6	-15068,6	15007490	0,07

Таким образом, величина среднего по территории региона слоя осадков в средний многолетний год (норма) составляет 732,2 мм при коэффициенте вариации (изменчивости), равном 0,07 или 7%.

Все случайные величины, представленные рядом количественных показателей, образуют статистическую (выборочную) совокупность, члены которой иногда называются *вариантой* или *датой*. Для обработки вариант составляется вариационный ряд, в котором варианты располагают в возрастающем или убывающем порядке. Те варианты, которые резко отличаются от вариант статистической совокупности и вызывают сомнения (*артифакт*), исключаются из статистической совокупности.

Вариационный ряд может быть представлен графически в виде полигона (кривая распределения частот) или гистограммы. При построении вариационной кривой по оси абсцисс откладываются значения вариант или средин классов, по оси ординат – частоты. Величина классового интервала  $i$ , которая зависит от принятого числа классов  $k$  и объема выборки  $n$ , определяется как:

$$i = (x_{max} - x_{min}) / k. \quad (2.11)$$

Число классов, в зависимости от выборки, определяется как:

$$k \approx 1 + 3,32 \cdot \lg(n). \quad (2.12)$$

Можно определить класс выборки по объему:

$n$	30...50	51...100	101...400	401...1000	1001...2000
$k$	4-6	6-8	8-9	9-11	11-12

Величина классового интервала должна быть одинаковой на протяжении всего вариационного ряда. Границы классов выбираются такими, чтобы каждая варианта могла быть отнесена только к одному классу. Первый и последний классы могут быть неполными. Границы классов желательно выбирать так, чтобы крайние варианты  $x_{max}$ ;  $x_{min}$ , по возможности, оказались ближе к середине интервала своего класса.



При построении гистограммы по оси абсцисс откладываются границы классов, а число вариант каждого класса обозначается высотой или площадью соответствующего прямоугольника. При сравнении изменчивости одинаковых условий или признаков, полученные вариационные кривые распределения частот наносятся на один график. Группировка вариант в классы для сравниваемых выборок должна быть одинаковой. Если объем выборок одинаков, все частоты должны быть выражены в процентах от объема выборки по каждой совокупности отдельно.

Гистограмма и полигон распределений являются графическим отображением частот, которые, в свою очередь, представляют собой оценки плотностей вероятностей. Кумулятивная линия представляет собой график накопленных частот и оценивает функцию распределения  $F(x)$  в точке  $x$ .

Рассмотрим технологию построения полигона на примере данных табл. 2.4. Разбивка данных на классы приведена в табл.2.6.

Таблица 2.6

**Разбивка массива исходных данных на классы, вычисление частот**

N п/п	Классы (величины атмосферных осадков, мм)	Средины интервалов	Абсолютные частоты	Относительные частоты	Относительные накопленные частоты
1	610-640	625	4	0,04	0,04
2	641-670	655	9	0,09	0,13
3	671-700	685	15	0,15	0,28
4	701-730	715	21	0,21	0,49
5	731-760	745	26	0,26	0,75
6	761-790	775	16	0,16	0,91
7	791-820	805	7	0,07	0,98
8	821-850	835	2	0,02	1,00

 $\Sigma 100$ 

Если распределение случайной величины подчиняется нормальному закону хотя бы приближенно, оно может быть описано как

$$Y_i = \alpha \cdot \exp(-b \cdot x^2), \quad (2.13)$$

то такое распределение является нормальным (гауссовым).

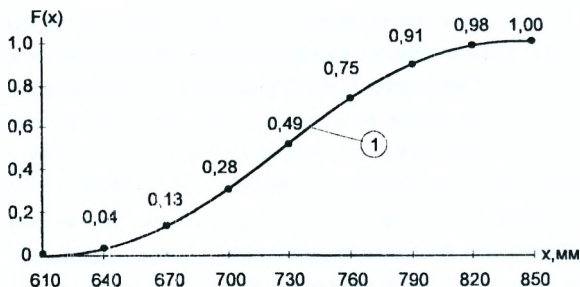


Рис. 2.3. Кумулятивная кривая (1) распределения вариационного ряда норм атмосферных осадков

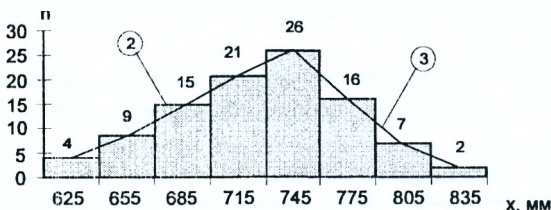


Рис. 2.4. Графическое представление распределения вариационного ряда норм осадков: 2 – гистограмма; 3 – полигон

Оно обладает очень важными свойствами:

– ордината  $y$ , которая характеризует высоту кривой для каждой точки оси абсцисс, представляет собой плотность вероятности некоторого значения переменной  $x$  и определяется по следующей формуле

$$y_i = f(x_i) = \frac{1}{\sigma\sqrt{2\cdot\pi}} \exp(-0,5(\frac{x-\mu}{\sigma})^2), (-\infty < x_i < +\infty, \sigma > 0), (2.14)$$

где  $\sigma$  – среднеквадратическое отклонение теоретического распределения;  
 $\mu$  – математическое ожидание теоретического распределения;

– кривая нормального распределения симметрична относительно среднего значения, следовательно, среднее, мода и медиана совпадают;

– максимум ординаты кривой определяется как  $y_{max} = 1/\sigma \sqrt{2\pi}$ , что, при  $\sigma=1$ , составляет примерно 0,4; если  $x \rightarrow \pm\infty$ , то  $y \rightarrow 0$ , т. е. очень большие и очень малые значения переменной  $x$  маловероятны;



– примерно 68,3% всех случаев наблюдений лежит в площади, отсекаемой перпендикулярами к оси  $0x$ ,  $\mu \pm \sigma$ ; соответственно, в пределах от  $-2\sigma$  до  $+2\sigma$  находятся 95,5% вариант, в пределах  $-3\sigma$  до  $+3\sigma$  – 99,7% (рис. 2.5).

Однако распределения множества природных процессов зачастую характеризуются ярко выраженной асимметрией, которая может быть описана следующими функциями:

$$x'_i = \lg(x_i \pm \alpha); \quad x'_i = 1/x_i; \quad x'_i = (x_i)^{-0,5}; \quad x'_i = (x_i)^\alpha. \quad (2.15)$$

Закон нормального распределения проявляется только при  $n > 20$  – 30. В практике часто приходится иметь дело с ограниченным числом данных и базировать выводы на малых выборках, для которых результаты обычно близки и редко появляются большие отклонения. Следовательно, в этих случаях полагаться на критерии нормального распределения в своих выводах нельзя [Крамер, 1975].

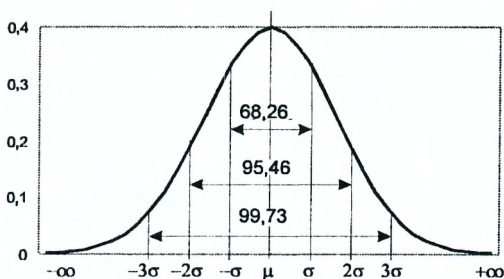


Рис. 2.5. Процент случаев наблюдений (площадь), ограниченный кривой нормального распределения, для различных значений  $\sigma$ .

Наиболее распространено в природопользовании логнормальное распределение, полученное из предположения, что нормально распределенной случайной величиной является не  $x$ , а  $\ln x$ . Естественно, что логарифмически нормальное распределение вероятностей может иметь только такая случайная величина  $x$ , которая принципиально не может принимать отрицательных значений (например, расход воды в реке). Плотность

логарифмически нормального распределения для  $x < 0$  равна нулю, а для  $x > 0$  она имеет вид

$$f(x) = \frac{1}{\gamma \cdot \sqrt{2 \cdot \pi}} \cdot \exp\left(-\frac{(\ln x - \ln b)^2}{2 \cdot \gamma^2}\right), \quad (2.16)$$

где  $\ln b$  и  $\gamma^2$  — суть математическое ожидание и дисперсия случайной величины  $\ln x$ .

Параметры  $b$  и  $\gamma^2$  нетрудно выразить через математическое ожидание  $\mu$  и дисперсию  $\sigma^2$  случайной величины  $x$ :

$$b = \frac{\mu^{1.5}}{(\sigma^2 + \mu^2)^{\frac{2}{3}}}; \quad \gamma^2 = \ln\left(\frac{(\sigma^2 + \mu^2)^{\frac{2}{3}}}{\mu}\right); \quad (2.17)$$

Логарифмически нормальное распределение, в отличие от нормального, имеет несимметричный график плотности вероятности (рис. 2.6).

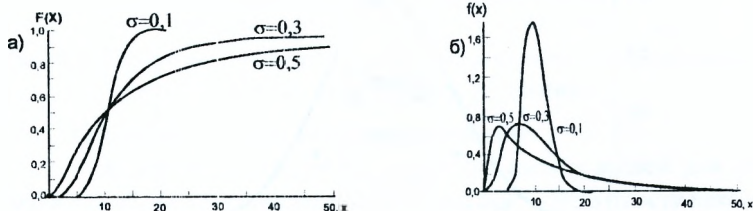


Рис. 2.6. Логарифмически нормальное распределение вероятностей в дифференциальных (а) и интегральных (б) видах.

Еще большое практическое значение имеет  $t$ -распределение, получившее название *распределение Стьюдента*, характеризуемое для выборочных средних равенством

$$t = (\bar{x}_i - \mu) / \frac{\bar{S}}{\sqrt{n}}. \quad (2.18)$$

Числитель формулы (2.18) означает отклонение выборочной средней от средней всей совокупности  $\mu$ , а знаменатель является показателем оценки величины  $\sigma / \sqrt{n} = \sigma_u$  или стандартной ошибки средней генераль-

ной совокупности. Таким образом, величина  $t$  измеряется отклонением выборочной средней  $\bar{x}$  от средней совокупности  $\mu$ , выраженным в долях ошибки выборки  $\bar{S}$ , принятой за единицу.

Максимумы частот нормального и  $t$ -распределения совпадают, но форма кривой  $t$ -распределения зависит от числа степеней свободы.

Распределение Стьюдента имеет важное значение при работе с малыми выборками и позволяет определить доверительный интервал, накрывающий среднюю совокупности  $\mu$ , а также проверить ту или иную гипотезу относительно генеральной совокупности. При этом нет необходимости знать параметры совокупности  $\mu$  и  $\sigma$ , достаточно иметь их оценки  $\bar{x}$  и  $\chi_{\phi}^2 < \chi_{\tau}^2$  для определенного объема выборки.

Если из нормально распределенной совокупности взять две независимые выборки объемом  $n_1$  и  $n_2$  и подсчитать дисперсии  $\bar{s}_1^2$  и  $\bar{s}_2^2$  со степенями свободы  $\nu_1 = n_1 - 1$  и  $\nu_2 = n_2 - 1$ , то можно определить отношение дисперсий

$$F = \bar{S}_1^2 / \bar{S}_2^2, \quad F \geq 1, \quad (2.19)$$

представляющее собой  $F$ -распределение Фишера.

Для оценки степени соответствия эмпирических данных определенным теоретическим предпосылкам или нулевой гипотезе  $H_0$  обычно используют распределения хи-квадрат ( $\chi^2$ ) описываемое равенством

$$\chi^2 = \sum_{i=1}^n \left( \frac{f_i - F_i}{F_i} \right)^2, \quad (2.20)$$

где  $f_i$  – фактические и  $F_i$  – гипотетические частоты численности объектов выборки.

Гипотеза опровергается – если  $\chi_{\phi}^2 \geq \chi_{\tau}^2$ , и не опровергается – если  $\chi_{\phi}^2 < \chi_{\tau}^2$ . Когда фактические и теоретически ожидаемые частоты полностью совпадают, тогда  $\chi^2 = 0$ .

Распределение  $\chi^2$ , так же, как и  $t$ -распределение, частной случайной  $F$  – распределения, при  $\nu_1 = \nu$  и  $\nu_2 = \infty$ , описывается зависимостью

$$F(\nu_1, \nu_2 = \infty) = \chi^2 / \nu. \quad (2.21)$$

Когда событие имеет очень малую вероятность свершения, например, небольшое число раз на 1000 или 10000 обычных явлений, тогда распределение случайной величины следует определенному закону редких событий (Пуассона)

$$P_{x_i} = \frac{\alpha^{x_i} \exp(-\alpha)}{x_i!}, \quad (2.22)$$

где  $P_{x_i}$  – вероятность значения  $x_i$ ;  $x_i$  – число редких событий, происходящих в каждой большой группе ( $x_i=0, 1, 2, 3$ , и т. д.);  $\alpha$  – среднее число редких событий на каждую большую группу;  $x_i!$  – факториал.

Основными свойствами пуассоновских процессов являются:

- вероятности не изменяются со временем, так что соответствующая последовательность не имеет тренда;
- вероятность наступления двух или более событий пренебрежимо мало;
- вероятность наступления события в интервале  $t$  совершенно не зависит, что произошло до момента  $t$ .

Однако следует отметить, что, как правило, вероятность геофизических процессов изменяется во времени, а поэтому применение закона Пуассона мало оправдано.

Распределение Пуассона определяется двумя параметрами – средней и дисперсией этого распределения. Графически распределение представляет асимметричную кривую, при этом асимметрия тем больше, чем меньше вероятность события.

Отсюда следует, что все теоретические распределения можно построить только на основании одной выборочной средней. Как видим, любые генеральные совокупности описываются некоторыми постоянными числовыми характеристиками распределения, оценка которых базируется на случайных выборках.

Для среднего значения  $\mu$  нормального распределения широко используются следующие оценки  $T_n$ :

- $x_j$  – первый элемент выборки;
- $(x_{max} + x_{min})/2$  – среднее арифметическое максимального и минимального элементов выборки;

- $M_0$  – мода, которая, при нормальном распределении, равна математическому ожиданию  $\mu$ ;
- $M_e$  – медиану, которая, при нормальном распределении, также равна  $\mu$ ;
- $\bar{x}$  – среднее арифметическое.

Для того, чтобы установить, какая из оценок лучше, надо знать основные свойства (виды) оценок. *Несмещенной* называется оценка  $T_n$ , среднее значение которой равно оцениваемому параметру  $\Theta$

$$\bar{T}_n = \mu . \quad (2.23)$$

Если это условие не выполняется, то оценку называют смещенной, при этом смещение вычисляется как разность  $\bar{T}_n - \Theta$ . Следует отметить, что кроме несмещенности оценки, существуют требования *состоятельности* и *эффективности*. Оценка параметра считается состоятельной, по мере роста  $n$  – числа наблюдений  $n \rightarrow N$  – в случае конечной генеральной совокупности объема  $N$ ; при  $n \rightarrow \infty$  – в случае бесконечной генеральной совокупности, она (оценка) стремится к оцениваемому теоретическому значению параметра (например, для дисперсии  $\lim_{n \rightarrow \infty} S^2(n) = \sigma$ ). Следует четко разделять понятия состоятельности и несмещенности. Рассмотрим для примера две оценки для дисперсии:

$$S^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 ; \quad (2.24)$$

$$\bar{S}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 . \quad (2.25)$$

Эти оценки состоятельны, но вторая является несмещенной, первая – смещенной, так как в ней содержится систематическая отрицательная погрешность –  $\sigma^2/n$ , поскольку математическое ожидание  $\left( \bar{S}^2 = (\sigma^2 - \sigma^2)/n \right)$ , с ростом  $n$  монотонно убывает. Из этого следует,

что требование несмещенности особенно важно при малом количестве данных наблюдений. Оценка параметра называется *эффективной*, если среди прочих оценок того же параметра она обладает наименьшей дис-

персией. Если среднее оценивать как  $(x_{max} + x_{min})/2$  или как  $\left(\frac{1}{n} \sum_{i=1}^n x_i\right)$ , то эти оценки обладают свойствами состоятельности и несмещенности, но при этом второй способ оценки является состоятельным, несмещенным и эффективным, а первый – только состоятельным и несмещенным.

Для оценки очень важна точность  $\Delta$ , которая определяется доверительной вероятностью (надежностью)  $\gamma$

$$\gamma = P(T_n - \Delta \leq \Theta \leq T_n + \Delta). \quad (2.26)$$

Точность оценки  $\Delta$  фактически определяет длину доверительного интервала  $2\Delta$ . Доверительная вероятность задается обычно значением, близким к единице, например, 0,95; 0,97; 0,99 и т.д. Доверительная вероятность ( $\gamma$ ), точность оценки  $\Delta$  и объем выборки  $n$  связаны между собой. Если определены две величины, то тем самым будет определена и третья.

Рассмотрим определение доверительного интервала для среднего значения  $\mu$  нормального распределения. Так как  $\sigma^2$  неизвестна, то непосредственно воспользоваться нормальным распределением нельзя. Учитывая, что случайная величина

$$t = \frac{\bar{x} - \mu}{S} \sqrt{n} \quad (2.27)$$

имеет распределение Стьюдента ( $t$ -распределение) с числом степеней свободы  $n - 1$ , то для получения интервальной оценки необходимо выполнение условия

$$\gamma = P\left(\left|\frac{\bar{x} - \mu}{S} \sqrt{n}\right| \leq t_\gamma\right). \quad (2.28)$$

Величина  $t_\gamma$  определяется по таблицам распределения Стьюдента. На основании условия  $P(|x| > t_\gamma) = \gamma$  определяется  $t_\gamma$ , но в данном случае, имеем противоположное неравенство, значит, необходимо использовать условие

$$\gamma = P\left(\bar{x} - t_\gamma \frac{\bar{S}}{\sqrt{n}} \leq \mu \leq \bar{x} + t_\gamma \frac{\bar{S}}{\sqrt{n}}\right), \quad (2.29)$$

где доверительный интервал указан в скобках и

$$t_{\gamma} \frac{\bar{S}}{\sqrt{n}} = \Delta . \quad (2.30)$$

Соответственно доверительный интервал для дисперсии нормального распределения  $\sigma^2$  определен следующим условием

$$\gamma = P \left( \frac{(n-1) \cdot \bar{S}^2}{u_2} \leq \sigma^2 \leq \frac{(n-1) \cdot \bar{S}^2}{u_1} \right), \quad (2.31)$$

где в скобках задан доверительный интервал для дисперсии  $\sigma^2$ . Доверительный интервал для среднего квадратического (стандартного) отклонения  $\sigma$  определяется следующим условием

$$\bar{S} \cdot \sqrt{n-1} / \sqrt{u_2} \leq \sigma \leq \bar{S} \cdot \sqrt{n-1} / \sqrt{u_1} . \quad (2.32)$$

Все это касается выборок с фиксированным объемом  $n$ .

По закону больших чисел предпочтение отдается выборкам с большим объемом. Но обычно большой объем выборки требует и больших затрат для ее получения и обработки. Поэтому, на практике, целесообразно использовать тот минимальный объем, который позволяет получить удовлетворительные результаты. При вычислении доверительных интервалов среднего значения нормального распределения, можно определить объем выборки

$$n = u_{\gamma}^2 \cdot \sigma^2 / \Delta^2 . \quad (2.33)$$

Таким образом, объем выборки  $n$  прямо пропорционален  $\sigma^2$  и  $u_{\gamma}^2$  (при этом последнее значение зависит от « $\gamma$ » и обратно пропорционален  $\Delta^2$ ). По этим данным можно задать значения вероятности и длину интервала  $\Delta$ . Если необходимо получить интервал с большей доверительной вероятностью  $-\gamma$  (вместе с тем увеличивается и « $u_{\gamma}$ »), то следует увеличить объем выборки  $n$ , как и при необходимости укоротить этот интервал. Если  $\sigma$  неизвестно, то выборки объема  $n$  имеется зависимость

$$n = t_{\gamma}^2 \cdot \bar{S}^2 / \Delta^2 . \quad (2.34)$$

Определим доверительные интервалы для математического ожидания и дисперсии, а также объем выборки, чтобы получить названные характеристики с заданной точностью, используя данные табл. 2.4. Выбо-



рочные характеристики распределения норм атмосферных осадков на территории региона представлены в табл. 2.5.

Из  $t$ -распределения Стьюдента, при  $n-1=99$  – степенях свободы и доверительной вероятности  $\gamma=0,95$ , найдем  $t_\gamma=1,9944$ . Используя формулу (2.30), имеем  $\Delta=9,91$  и математическое ожидание запишется, как  $\mu=732,2 \pm 9,9$  мм, при  $722,3 \leq \mu \leq 742,1$  мм. Таким образом, с вероятностью 95% математическое ожидание  $\mu$  находится в диапазоне  $722,3 \leq \mu \leq 742,1$  мм. Из таблицы  $\chi^2$ -распределения  $u_1=82,36$  и  $u_2=135,8$ . Используя формулу (2.32), имеем

$$\frac{2471,3(100-1)}{135,8} \leq \sigma^2 \leq \frac{2471,3(100-1)}{82,36} \quad 1801,6 \leq \sigma^2 \leq 2970,6 \text{ или} \\ 42,4 \leq \sigma \leq 54,5.$$

Таким образом, с вероятностью 95% дисперсия  $\sigma$  лежит в диапазоне  $42,4 \leq \sigma \leq 54,5$ . Для определения объема выборки, воспользуемся формулой (2.33). При  $\gamma=95$  и  $n=100$ , по таблицам  $t$ -распределения имеем  $t_\gamma=1,98$  и, тогда, приняв  $\Delta=10$  мм, имеем  $n=1,98^2 \frac{2446,6}{10^2} = 95,9 \approx 96 < 100$ , а при  $\Delta=9$  мм -  $n=1,98^2 \frac{2446,6}{9^2} = 118 > 100$ .

Таким образом, чтобы получить математическое ожидание с точностью до 10 мм слоя атмосферных осадков при 95%-ной вероятности достаточно 96 членов ряда, а с точностью до 9 мм – уже 118 членов ряда.

### 2.3. Особенности проверки статистических гипотез о видах и свойствах распределения

Гипотезы о неизвестном параметре  $\Theta$  распределения могут быть простые и сложные.

Простая гипотеза утверждает, что параметр  $\Theta$  имеет одно конкретное значение  $\Theta=\Theta_0$ , а сложная гипотеза утверждает, что параметр  $\Theta$  имеет значение из совокупности значений  $\Theta < \Theta_0$ ,  $\Theta > \Theta_0$ ,  $\Theta \neq \Theta_0$ .

Проверка простой статистической гипотезы состоит из следующих этапов: 1) определение гипотез  $H_0$  и  $H_1$ ; 2) выбор статистики и задание уровня значимости  $\alpha$ ; 3) определение (по таблицам) критической области



по уровню значимости  $\alpha$  и по альтернативной гипотезе  $H_1$ ; 4) вычисление по выборке значения статистики; 5) сравнение значения статистики с критической областью; 6) принятие решения: если значение статистики не входит в критическую область, то принимается гипотеза  $H_0$  и отвергается гипотеза  $H_1$ ; если значение статистики входит в критическую область, то отвергается гипотеза  $H_0$  и принимается гипотеза  $H_1$ .

Итак, если в результате проверки статистической гипотезы принята гипотеза  $H_1$ , то можно считать ее доказанной; а если – гипотеза  $H_0$ , то признали, что она не противоречит результатам наблюдений.

Проверка же сложной статистической гипотезы состоит из следующих этапов:

а) гипотеза о среднем значении нормального распределения при известном  $\sigma$ .

Предположим, что генеральная совокупность имеет нормальное распределение  $x \in N(\mu, \sigma)$ , где  $\sigma$  – известно. При уровне значимости  $\alpha$  нужно проверить гипотезу  $H_0 : \mu = \mu_0$ . В качестве альтернативной можно использовать одну из следующих гипотез  $H_1 : \mu < \mu_0$ ,  $H_1 : \mu > \mu_0$  или  $H_1 : \mu \neq \mu_0$ . В качестве статистики воспользуемся случайной величиной

$$Z = \frac{\bar{x} - \mu_0}{\sigma} \sqrt{n}, \quad (2.35)$$

которая при истинной гипотезе  $H_0$  имеет нормальное распределение  $Z \in N(0, 1)$ . Критическая область определяется с помощью таблицы функции нормального распределения. Если альтернативная гипотеза имеет вид  $H_1 : \mu < \mu_0$ , то используется левосторонняя критическая область (рис. 2.7), которая удовлетворяет следующему условию  $P(z < -z_\alpha) = \Phi(-z_\alpha) = \alpha$ .

Так как таблицы составлены только для положительных значений аргумента, поэтому из таблиц находится  $z_\alpha$ , с учетом того, что  $\Phi(z_\alpha) = 1 - \alpha$  и  $z < -z_\alpha$ .

Если альтернативная гипотеза имеет вид ( $H_1 : \mu > \mu_0$ ), то используется правосторонняя критическая область (рис. 2.8), которая удовлетворяет условию  $P(z > z_\alpha) = \alpha$ .

Из таблиц нормального распределения получается значение, с учетом того, что  $P(z < z_\alpha) = \Phi(z_\alpha) = 1 - \alpha$  и  $z > z_\alpha$ .

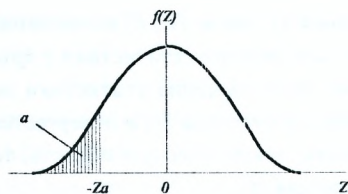


Рис. 2.7. Левосторонняя критическая область

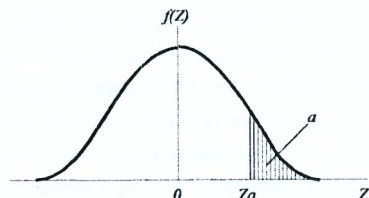


Рис. 2.8. Правосторонняя критическая область

И наконец, при альтернативной гипотезе  $H_1: \mu \neq \mu_0$ , используется двусторонняя критическая область (рис. 2.9), удовлетворяющая условию  $P(|z| > z_\alpha) = \alpha$ .

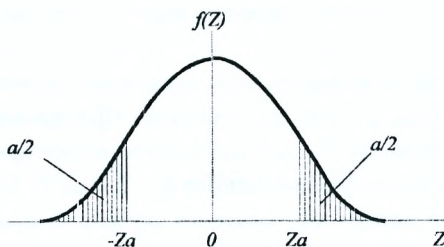


Рис. 2.9. Двусторонняя критическая область

Отсюда —  $P(z < z_\alpha) = P(z > z_\alpha) = \alpha/2$ ;  $\Phi(z_\alpha) = 1 - \alpha/2$  и  $|z| > z_\alpha$ .

Поэтому для вычисления статистики с помощью формулы (2.35) нужно только по выборке найти среднее арифметическое  $\bar{x}$ ;

б) гипотеза о среднем значении нормального распределения при неизвестном  $\sigma$ .

Предположения те же, что и в предыдущем случае, но только  $\sigma$  неизвестно. В этом случае, в качестве статистики используется случайная величина

$$t = \frac{\bar{x} - \mu_0}{S} \sqrt{n}, \quad (2.36)$$

которая, если верна гипотеза  $H_0$ , имеет  $t$ -распределение Стьюдента с числом степеней свободы  $(n - 1)$ , где  $n$  — объем выборки.

Критические области определяются так же, как и в предыдущем пункте, но использование таблицы  $t$ -распределения Стьюдента проще, так как она составлена именно для этих целей. При нахождении левосторонней или правосторонней критических областей используется верхняя головка таблицы, а для двусторонней – нижняя. Перед вычислением по формуле (2.36) значения статистики нужно по выборке вычислить  $\bar{x}$  и  $\bar{S}$ ;

в) *гипотеза о дисперсии нормального распределения*

Предположим, что генеральная совокупность имеет нормальное распределение –  $x \in N(\mu, \sigma)$ , где параметр  $\sigma$  неизвестен. Требуется при уровне значимости  $\alpha$  проверить гипотезу  $H_0: \sigma^2 = \sigma_0^2$ . В качестве статистики используется случайная величина

$$\chi^2 = (n-1) \cdot \bar{S}^2 / \sigma_0^2. \quad (2.37)$$

Если гипотеза  $H_0$  верна, то случайная величина  $\chi^2$  имеет  $\chi^2$  – распределение Пирсона с числом степеней свободы  $(n-1)$ . Критическая область определяется в зависимости от альтернативной гипотезы  $H_1$  по специальным таблицам  $\chi^2$  – распределения. Если альтернативная гипотеза имеет вид  $H_1: \sigma^2 < \sigma_0^2$ , то используется левосторонняя критическая область, удовлетворяющая условие  $P(\chi^2 < x_\alpha) = \alpha$ .

Так как таблица  $\chi^2$  – распределения составлена в соответствии с противоположным условием, поэтому, для нахождения из таблицы  $\chi_\alpha$ , используется условие  $P(\chi^2 > x_\alpha) = 1 - \alpha$ .

При альтернативной гипотезе  $H_1: \sigma^2 > \sigma_0^2$  находится правосторонняя критическая область, исходя из условия  $P(\chi^2 > x_\alpha) = \alpha$ , по которому  $\chi_\alpha$  можно найти непосредственно из таблицы.

При альтернативной гипотезе  $H_1: \sigma^2 \neq \sigma_0^2$  находится двусторонняя критическая область согласно условию  $P((\chi^2 < x'_\alpha) \cup (\chi^2 < x''_\alpha)) = \alpha$ .

Обычно принимается симметричная по вероятности критическая область, удовлетворяющая условию  $P((\chi^2 < x'_\alpha) = P(\chi^2 \geq x''_\alpha)) = \alpha / 2$ .

При этом, можно сразу из таблицы найти  $x''_\alpha$ , а для получения  $x'_\alpha$  следует сделать преобразование

$$P(\chi^2 > x_\alpha) = 1 - \alpha / 2. \quad (2.38)$$

По выборке нужно вычислить несмещенную оценку дисперсии генеральной совокупности  $\bar{S}^2$ , а затем по формуле (2.37) – найти значение статистики  $\chi^2$ ;

г) *гипотеза о равенстве двух средних значений*

Предположим, что две генеральные совокупности имеют нормальное распределение –  $X_1 \in N(\mu_1, \sigma_1)$ ,  $X_2 \in N(\mu_2, \sigma_2)$ , при этом стандартные отклонения  $\sigma_1$  и  $\sigma_2$ , неизвестны, но должны быть равными. Сначала нужно проверить гипотезу о равенстве дисперсий. Из обеих генеральных совокупностей сделаны независимые выборки с параметрами –  $(n_1, \bar{x}_1, \bar{S}_1)$ ;  $(n_2, \bar{x}_2, \bar{S}_2)$  соответственно. Обозначим разность средних через  $\delta = \mu_1 - \mu_2$ . Зафиксировав уровень значимости  $\alpha$ , проверим гипотезу  $H_0: \delta = \delta_0$ , используя статистику

$$t = \frac{\bar{x}_1 - \bar{x}_2 - \delta_0}{\sqrt{\frac{(n_1 - 1)\bar{S}_1^2 + (n_2 - 1)\bar{S}_2^2}{n_1 + n_2 - 2} \cdot \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}. \quad (2.39)$$

Если гипотеза  $H_0$  верна, то случайная величина имеет  $t$ -распределение с числом степеней свободы  $(n_1 + n_2 - 2)$ . Обычно,  $\delta_0 = 0$ , т. е. проверяется гипотеза о равенстве средних значений генеральных совокупностей. Критическая область определяется в зависимости от вида альтернативной гипотезы  $H_0(\delta < \delta_0, \delta > \delta_0$  или  $\delta \neq \delta_0)$ ;

д) *гипотеза о равенстве двух дисперсий*

Предположим, как и в предыдущем случае, что заданы две генеральные совокупности  $X_1$  и  $X_2$  имеют нормальные распределения –  $X_1 \in N(\mu_1, \sigma_1)$  и  $X_2 \in N(\mu_2, \sigma_2)$ . Из этих генеральных совокупностей образованы независимые выборки с параметрами –  $(n_1, \bar{x}_1, \bar{S}_1)$  и  $(n_2, \bar{x}_2, \bar{S}_2)$  соответственно. Требуется при уровне значимости  $\alpha$  проверить гипотезу  $H_0: \sigma_1^2 = \sigma_2^2$  при альтернативной гипотезе  $H_0: \sigma_1^2 > \sigma_2^2$ . Обычно здесь другие альтернативные гипотезы не используются. Предполагая, что  $\bar{S}_1^2 > \bar{S}_2^2$ , принимаем в качестве статистики величину

$$F = \bar{S}_1^2 / \bar{S}_2^2. \quad (2.40)$$

Если гипотеза  $H_0$  верна, то случайная величина  $F$  имеет  $F$ -распределение Фишера с числом степеней свободы  $(n_1 - 1)$  и  $(n_2 - 1)$ . Критическая область только правосторонняя; определяется условием  $P(F > f_\alpha) = \alpha$ .

Величины  $f_\alpha$  находятся из таблиц  $F$ -распределения, которые зависят от трех величин: уровня значимости  $\alpha$  и степеней свободы  $\nu_1$  и  $\nu_2$ , поэтому, таблицы составлены отдельно для каждого значения  $\alpha$ . В таблицах число степеней свободы большей дисперсии  $\nu_1$  находится в верхней их части;

### е) $\chi^2$ – критерий согласия

Рассмотрим, как можно проверить гипотезу о распределении генеральной совокупности  $X$ . Пусть генеральная совокупность имеет какое-то неизвестное распределение. Сделаем выборку из генеральной совокупности. На основании выборки составим гипотезу о конкретном распределении генеральной совокупности, выраженной через функцию распределения –  $F(x)$ . Это распределение является теоретическим. По выборке можно найти эмпирическую функцию распределения –  $F^*(x)$ . Гипотеза  $H_0$  о распределении генеральной совокупности принимается тогда, когда эмпирическое распределение хорошо согласуется с теоретическим. Полного совпадения, конечно, ожидать не стоит. При использовании  $\chi^2$  – критерия согласия вся область изменения генеральной совокупности  $X$  делится на  $k$  – интервалов, которые могут быть различной длины. По выборке составляется вариационный ряд по этим же интервалам. Если в некотором интервале частота  $n_i$  слишком мала (меньше 5), то этот интервал объединяется с соседним интервалом. При дискретной генеральной совокупности интервал может содержать только одно значение генеральной совокупности. По выборке вычисляются оценки параметров теоретического распределения. Тем самым, теоретическое распределение будет полностью определено. Теперь по теоретическому распределению вычисляются вероятности  $p_i$  того, что случайная величина  $x$  принимает значение из  $i$ -го интервала, при этом,  $\sum_{i=1}^k p_i = 1$ . Затем, находятся теоретические частоты  $m_i = n_i \cdot p_i$ . Гипотеза  $H_0$  верна, если теоретические и эмпирические частоты

$m_i$  и  $n_i$  достаточно мало отличаются друг от друга. Для проверки гипотезы  $H_0$  используется следующая статистика

$$Q^2 = \sum_{i=1}^k \frac{(n_i - m_i)^2}{m_i} \quad (2.41)$$

Случайная величина  $Q^2$  имеет  $\chi^2$ -распределение с числом степеней свободы  $(k - v - 1)$ , где  $k$  – количество интервалов,  $v$  – количество параметров теоретического распределения, оценки которых вычисляются по выборке. Чем больше  $Q^2$ , тем хуже согласованы теоретическое и эмпирическое распределения. При достаточно большом значении  $Q^2$  нужно отвергнуть гипотезу  $H_0$ . Поэтому, используется только правосторонняя критическая область;

ж) *отсев грубых погрешностей*

При выборке небольшого объема ( $n \leq 25$ ), можно воспользоваться методом вычисления максимального отклонения

$$\tau' = \frac{|x_{i \cdot \max(\min)} - \bar{X}|}{\sqrt{(n-1)/n \cdot S}} \quad (2.42)$$

где  $x_{i \cdot \max(\min)}$  – крайний (наибольший или наименьший) элемент выборки, по которой подсчитывались  $\bar{x}$  и  $\bar{S}$ ;  $\tau_{1-p}$  – табличное значение статистики  $\tau$ , вычисленной при доверительной вероятности  $q=1-P$ . Таким образом, для выделения аномального значения вычисляется

$$\tau = |x_{i \cdot \max(\min)} - \bar{X}| / \bar{S} \quad (2.43)$$

которое, затем, сравнивается с табличным значением ( $\tau_{1-p}$ ) и  $\tau \leq \tau_{1-p}$ .

Если неравенство соблюдается, то результат наблюдения не отсеивается, если не соблюдается, – он исключается. После исключения того или иного результата наблюдения (нескольких результатов), характеристики эмпирического распределения должны быть пересчитаны по данным сокращенной выборки. Квантили распределения статистики  $t$ , при уровнях значимости –  $P=0,10$ ;  $P=0,05$ ;  $P=0,01$ , или, соответственно доверительной вероятности  $(1-P)=q=0,90$ ;  $0,95$ ;  $0,975$ ;  $0,99$  даны в таблице  $t$ -распределения Стьюдента. На практике обычно используется уровень значимости –  $P=0,05$  (с 95%-ой доверительной вероятностью). Проверку отсева можно повторить и для следующего по абсолютной величине мак-



симального относительного отклонения, но предварительно необходимо пересчитать  $\bar{x}$  и  $\bar{S}$  для выборки нового объема  $(n-1)$ . Рассмотрим другой метод отсева грубых погрешностей для малой выборки. В этом случае, вычисляется

$$\tau' = \frac{|x_{i-\max(\min)} - \bar{X}|}{\sqrt{(n-1)/n \cdot \bar{S}}}, \quad (2.44)$$

и полученный результат сравнивается с критическим табличным значением  $t$ -распределения Стьюдента при следующих  $n$  и  $(1 - P)$ .

Отсев грубых погрешностей для больших выборок лучше всего производить с использованием  $t$ -распределения Стьюдента. Известно, что критическое значение « $\tau_p$ » ( $p$ -процентная точка нормированного выборочного отклонения) выражается через  $(t_{p;n-2})$  – критическое значение распределения Стьюдента ( $t$  – распределение).

$$\tau_{(p;n)} = \frac{t_{(p;n-2)}\sqrt{n-1}}{\sqrt{n-2 + (t_{(p;n-2)})^2}}. \quad (2.45)$$

Рекомендуемый метод отсева грубых погрешностей удобен еще тем, что максимальные относительные отклонения в процессе вычисления могут быть разделены на группы: 1)  $\tau \leq \tau(5\%;n)$ ; 2)  $\tau(5\%;n) < \tau < \tau(0,1\%;n)$ ; 3)  $\tau > \tau(0,1\%;n)$ . Результаты наблюдений, попавшие в первую группу, нельзя отсеивать ни в коем случае, из второй группы - можно отсеять, если в пользу этой процедуры имеются еще и другие соображения экспериментатора, из третьей группы, по-видимому, результаты отсеиваются всегда.

Проверим выборку из табл. 2.4 на наличие в средних многолетних годовых величинах атмосферных осадков грубых погрешностей.

Применяем следующую процедуру отсева грубых погрешностей измерения атмосферных осадков для больших выборок:

1) из табл. 2.4 выбираем результат наблюдения, имеющий наибольшее отклонение, –  $(x_i - \bar{x}) = 841 - 732,2 = 109,2$  мм;

2) по (2.43) вычисляем  $\tau$  и  $\tau = 2,2$ ;



3) находим процентные точки  $t$ -распределения Стьюдента ( $t_{p;n-2}$ ) -  $t_{5\%;98}=1,6615$ ;  $t_{0,1\%;98}=3,1750$ ;

4) по формуле (2.45) вычисляем соответствующие точки  $\tau_{5\%;98}$  и  $\tau_{0,1\%;98}$ :  $\tau_{(5\%;98)} = 1,647$ ;  $\tau_{(0,1\%;98)} = 3,039$ .

Значение  $\tau=2,20$  находится между табличными критическими значениями:  $1,647 \leq 2,20 \leq 3,039$ .

В этом случае, от отсева, выделяющегося наблюдения, лучше воздержаться;

б) проверим гипотезу нормального распределения на выборке средних многолетних годовых величин осадков (табл. 2.4).

Рассмотрим пять основных методик проверки гипотезы нормальности распределения: по среднему абсолютному отклонению (CAO), по размаху варьирования  $Rv$ , по показателям асимметрии и эксцесса, по  $\chi^2$ -критерию и по критерию Колмогорова-Смирнова (К-С-критерию).

Методика проверки нормальности распределения по показателям асимметрии и эксцесса хорошо иллюстрирует использование моментов, а также удобна при проведении расчетов на ЭВМ. Проверка по К-С – критерию проводится только в отдельных случаях. Для практического применения рекомендуются, в основном, две методики: по размаху варьирования и по  $\chi^2$ -критерию, причем, первая служит для быстрой «прикладной» проверки, а вторая – для основательной проверки нормальности распределения.

Для небольших выборок ( $n < 120$ ) можно использовать среднее абсолютное отклонение (CAO), выражаемое формулой

$$CAO = \sum |x_i - \bar{x}| / n . \quad (2.46)$$

Для выборки, имеющей приближенно нормальный закон распределения, должно быть справедливо выражение

$$|CAO / \bar{S} - 0,7979| < 0,4\sqrt{n} . \quad (2.47)$$

При соотношении,  $|41,0/49,7 - 0,7979| < 0,4\sqrt{100}$ , гипотеза нормальности распределения выборки данных, приведенных в табл. 2.4, принимается.

Быстрая проверка гипотезы нормальности распределения выполняется с помощью размаха варьирования  $Rv$  (при  $3 < n < 100$ ). Рассчитывается отношение  $R/\bar{S}$  и сопоставляется с его критическими верхними и нижними границами. Если  $R/\bar{S}$  меньше нижней или больше верхней границы, то нормального распределения нет. Особенно важно, чтобы это условие соблюдалось при  $P=0,10$  (10%-ый уровень значимости). В нашем случае,  $R/\bar{S} = 4,37$ . При  $n=100$  и  $p=0,10$ , нижняя и верхняя границы соответственно равны 4,44 и 5,65, т. е. условие не выполняется и, следовательно, с вероятностью  $P=90\%$  можем отвергнуть гипотезу о нормальности распределения.

Некоторое представление о близости эмпирического распределения к нормальному дает анализ показателей асимметрии и эксцесса.

Показатель асимметрии можно определить по формуле

$$g_1 = \frac{m_3}{m_2^{1,5}} \text{ и } q_1 = -0.12 \neq 0 ; \quad (2.48)$$

следовательно, *некоторая асимметрия имеет место.*

В качестве показателя эксцесса принимается величина

$$g_2 = \frac{m_4}{m_2^2} - 3 \quad \text{и} \quad q_2 = -0.49 \neq 0 \quad (2.49)$$

как видно, *эксцесс также имеет место.*

Несмещенные оценки для показателей асимметрии и эксцесса определяются с использованием соответствующих формул и

$$G_1 = -0,122 ; G_2 = -0,453 .$$

При проверке гипотезы нормальности распределения следует также вычислить среднеквадратические отклонения для показателей асимметрии и эксцесса и  $S_{G_1} = 0,24 ; S_{G_2} = 0,48$ .

Если выполняются условия:

$$|G_1| \leq 3S_{G_1} ; |G_2| \leq 5S_{G_2} , \quad (2.50)$$

то гипотеза нормальности исследуемого распределения может быть принята.

В нашем случае,  $|-0,122| \leq 3 \cdot 0,24$  и  $|-0,45| \leq 5 \cdot 0,48$ , следовательно, условия (2.50) выполняются и гипотеза нормальности распределения может быть принята.

Рассмотрим методику проверки гипотезы нормальности распределения по  $\chi^2$ -критерию. Применение критерия  $\chi^2$  предполагает также использование свойств, так называемого, стандартного нормального распределения. Уравнение кривой стандартного нормального распределения имеет вид

$$y = f(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right) \approx 0,4 \exp\left(-\frac{z^2}{2}\right), \quad (2.51)$$

где  $z = (x_i - \mu)/\sigma$  – значения ординат кривой стандартного нормального распределения.

Расчеты выполняются в табличной форме (табл. 2.7) с использованием данных табл. 2.6. Имеем:  $\bar{x} = 727,61$ ;  $\bar{S} = 48,1$ ;  $k' = 62,4$ ;

Из табл. 2.7 видно, что критерий  $\hat{\chi}^2 = 3,857$ . Число степеней свободы  $\nu = 3$ , так как оценивается два параметра:  $\bar{x}$  и  $\bar{S}$

Находим значение  $-\chi^2_{(3,104)} = 6,251$ .

Таким образом, гипотеза о том, что наблюдаемые частоты распределены нормально, принимается на 10%-ном уровне.

Данные табл. 2.7 позволяют проверить гипотезы нормальности распределения с помощью критерия согласия Колмогорова-Смирнова (К-С – критерия). Для этого вычисляется

$$\hat{D} = \max |F_{n_i} - F_{m_i}| / n, \quad (2.52)$$

где  $F_{n_i}$  – накопленная наблюдаемая частота;  $F_{m_i}$  – накопленная ожидаемая частота.

Результаты вычисления К-С – критерия приведены в табл. 2.8. Величины  $n_i$  и  $m_i$  получены накоплением их частот. Затем, выбрано максимальное значение  $F_{n_i} - F_{m_i}$  и по нему определен критерий согласия Колмогорова-Смирнова  $\hat{D}$ . Полученное значение сравнивается с критиче-

ским, взятым из стандартных таблиц  $D \leq D^T$ . В нашем случае,  $D_{(100;1;10)} = 0,121 > 0,025 = D$ , т. е. можно сделать вывод: гипотеза нормального распределения на достаточно «жестком» 10%-ом уровне принимается.

Таким образом, четыре теста из пяти не отвергают гипотезу нормального распределения, следовательно, можно сделать заключение, что вариационный ряд годовых атмосферных осадков (табл. 3.2) подчиняется нормальному закону распределения вероятностей.  $\hat{D} = 2,461/100 = 0,025$ .

Рассмотрим, различимы ли средние величины годовых атмосферных осадков для различных частей региона, при уровне значимости  $\alpha = 0,05$ . Рассчитав среднюю величину осадков и дисперсию для различных склонов соответственно имеем:  $x_4 = 741,1$  мм;  $\bar{S}_4^2 = 1272,1$ ;  $n_4 = 39$  и  $x_6 = 737,5$  мм;  $\bar{S}_3^2 = 1940,7$ ;  $n = 30$ .

Сначала проверим гипотезу о равенстве дисперсии, т. е.  $H_0: \sigma_4^2 = \sigma_6^2$ , при  $H_1: \sigma_4^2 > \sigma_6^2$ . Статистика  $F$  (формула 3.56) имеет  $F$ -распределение с числом степеней свободы  $(n_4 - 1)$  и  $(n_6 - 1)$ . Найдем правостороннюю критическую область согласно условию  $P(F > f_{\alpha}) = \alpha$ ; при  $F$ -распределении имеем  $P(F > f_n) = 0,05$ ;  $f_n = 1,85$ ;  $F > 1,85$ .

Таблица 2.7

 Результаты вычисления  $\chi^2$ 

№ класса	Середина интервалов ( $x_i$ )	Частоты ( $n_i$ )	$x_i^2$	$n_i x_i$	$n_i x_i^2$	$x_i - \bar{x}$
1	2	3	4	5	6	7
1	625	4	390625	2500	1562500	-102,6
2	655	9	429025	5895	3861225	-72,6
3	685	15	469225	10275	7038375	-42,6
4	715	21	511225	15015	10735725	-12,6
5	745	26	555025	19370	14430650	17,4
6	775	16	600625	12400	9610000	47,4
7	805	7	648025	5635	4536175	77,4
8	835	2	697225	1670	1394450	107,4
		$\Sigma 100$		72760	53169096	

→ продолжение таблицы 2.7

$\frac{ x_i - \bar{x} }{s} = z$	Ординаты $f(z)$	$f(z)k'$	$m_i$	$n_i - m_i$	$(n_i - m_i)^2$	$\frac{(n_i - m_i)^2}{m_i}$
8	9	10	11	12	13	14
2,133	0,0413	2,577				
1,509	0,1276	7,962	10,539	2,461	6,057	0,575
0,886	0,2685	16,754	16,754	-1,754	3,077	0,184
0,262	0,3857	24,068	24,068	-3,068	9,413	0,391
0,362	0,3739	23,331	23,331	2,669	7,124	2,669
0,985	0,2444	15,251	15,251	0,749	0,561	0,037
1,609	0,1092	6,814	8,886	0,114	0,013	0,001
2,233	0,0332	2,072				
						Σ 3,857

Таблица 2.8

## Результаты вычисления К-С - критерия

$n_i$	4	9	15	21	26	16	7	2
$m_i$	2,577	7,962	16,754	24,068	23,331	15,251	6,814	2,072
$F_{n_i}$	4	13	28	49	75	91	98	100
$F_{m_i}$	2,577	10,539	27,293	51,361	74,692	89,943	96,757	98,829
$F_{n_i} - F_{m_i}$	1,423	2,461	0,707	2,361	0,308	1,057	1,243	1,171

По формуле (2.39) вычисляем значение статистики и  $F = 1,53$ .

Это значение не принадлежит критической области, поэтому нет оснований отвергать гипотезу ( $H_0: \sigma_4^2 = \sigma_6^2$ ), т. е. можно считать дисперсии генеральных совокупностей равными. Проверим гипотезу о равенстве средних значений генеральных совокупностей, т. е.  $H_0: \delta = 0$ , при  $H_1: \delta < 0$ .

Статистика « $t$ » (формула 2.39) имеет  $t$ -распределение с числом степеней свободы  $(n_4 + n_6 - 2)$ . По альтернативной гипотезе  $H_1$ , используя условие  $P(t < t_\alpha) = 0,05$  и таблицу  $t$ -распределения, найдем левостороннюю критическую область  $t_\alpha = -1,67$ ;  $t < -1,67$ .

По формуле (2.39) вычисляем значение статистики и  $t = -0,37$ .

Полученное значение не принадлежит критической области, поэтому оснований отвергать гипотезу  $H_0: \delta = 0$  нет, т. е. считаем средние значения генеральных совокупностей равными.

При уровне значимости  $\alpha = 0,05$  проверим гипотезу о среднем значении  $H_0: \mu = 750$ , при  $H_1: \mu < 750$  (табл. 2.5).

Статистика « $t$ » имеет  $t$ -распределение с числом степеней свободы --  $(n-1=100-1=99)$ . При альтернативной гипотезе  $H_1$ , найдем левостороннюю критическую область по условию  $P(t < t_{\alpha} = 0,05)$ .

Из  $t$ -распределение Стьюдента получаем  $t_{\alpha=0,05} = -1,6602$ .

Значение статистики вычисляем по формуле (2.35) и  $t = -3,5$ .

Значение статистики принадлежит критической области. Следовательно, отвергаем гипотезу  $H_0$  и принимаем альтернативную гипотезу  $H_1: \mu < 750$ .

При уровне значимости  $\alpha = 0,05$ , проверим гипотезу о дисперсии, т. е.  $H_0: \sigma^2 = 2450$ , при  $H: \sigma^2 > 2450$ .

Статистика « $\chi^2$ » (формула 2.37) имеет  $\chi^2$  - распределение с числом степеней свободы  $(n-1=100-1=99)$ . По альтернативной гипотезе ( $H_1$ ) найдем правостороннюю критическую область, используя условие (2.37) и таблицу  $\chi^2$ -распределения Пирсона. При этом имеем:  $P(\chi^2 > \chi_{\alpha} = 0,05; \chi_{\alpha} = 123,3; \chi^2 > 123,3$ .

По формуле (2.36) вычислим значение статистики и:  $\chi^2 = 99,9$ .

Это значение принадлежит критической области, поэтому гипотеза  $H_0: \sigma^2 = 2450$  не будет отвергнута.

#### 2.4. Алгоритмы и методы контроля информации

В гидрометеорологии разработано и внедрено в практику множество методов контроля как оперативной, так и режимной информации. Все методы можно объединить в две группы: синтаксический контроль и логический контроль. Одновременно эти группы являются и двумя этапами контроля. На первом этапе выявления ошибок используются синтаксические методы контроля, на втором – логические.

Первая группа методов контроля существует для проверки структуры (синтаксиса) построения сообщения. Каждое сообщение строится по строго определенным правилам (схемам), которые и берутся за основу методов синтаксического контроля. Синтаксический контроль обычно осуществляется в ЭВМ одновременно с раскодированием информации. В числе алгоритмов, по которым проводится проверка, используются:

- проверка каждого символа сообщения и исключение запрещенных комбинаций элементов. Например, форматом представления некоторых видов гидрометеорологических данных предусматривается использовать только цифровые символы кода; наличие букв в таком случае будет указывать на ошибку в сообщении;

- проверка наличия тех элементов, которые должны быть в сообщении в обязательном порядке. Это относится, например, к прозрачной части сообщения;

- проверка знака числа у тех элементов наблюдения, для которых этот знак строго определен. Например, дата и срок наблюдений, видимость, давление и другие элементы должны быть всегда положительными величинами;

- контроль формата представления величины (количество разрядов символов, занятых тем или иным элементом сообщения, не должно превышать определенного числа). Такой алгоритм используется в тех случаях, если длина групп в сообщении не является величиной постоянной;

- анализ точности записи числа путем проверки положения запятой в значениях, перфорация которых осуществляется в естественном виде (т. е. со знаком запятой);

- анализ сочетаний отдельных элементов или групп сообщения. Например, если общее количество облачности равно нулю, то должны быть равны нулю значения облачности верхнего и среднего ярусов;

- проверка правильности общего построения сообщения: общего количества групп, наличия признаков начала и конца в сокращении и т. д.

Здесь перечислены лишь некоторые алгоритмы синтаксического контроля гидрометеорологической информации.

Уже на этапе синтаксического контроля могут быть выявлены отдельные случайные ошибки или грубые промахи. Однако после этого этапа контроля часть ошибок в информации может еще сохраниться. По



этому наряду с методами синтаксического контроля используются методы логического контроля. В основу последних положены различные закономерности изменения того или иного элемента наблюдений во времени или пространстве, взаимосвязи отдельных элементов и их комплексов.

Одним из приемов, который используется для контроля, является установление для каждого элемента некоторых экстремальных (максимальных и минимальных) границ его изменения. Эти границы могут устанавливаться по известным физическим представлениям или по результатам многолетних наблюдений.

Для таких элементов, как температура воздуха, пределы изменений устанавливаются по фактическим наблюдениям. Например, если за длительный период работы метеорологической станции температура воздуха не была ниже  $-45^{\circ}\text{C}$  и не поднималась выше  $+35^{\circ}\text{C}$ , то эти значения могут быть приняты в качестве предельных (пределов) при контроле текущей информации.

В общем случае алгоритм контроля величины любого элемента ( $X_i$ ) на пределы может быть представлен следующим выражением:

$$X_{\min} \leq X_i \leq X_{\max}$$

где  $X_{\min}$ ,  $X_{\max}$  – минимальный и максимальный пределы для данной величины.

Эти пределы задаются в программах контроля информации на ЭВМ в виде констант (постоянных значений). Такие константы могут задаваться одинаковыми для всех пунктов наблюдения и для всех месяцев или сезонов года, но можно задавать их и более точно, выбирая для каждого месяца или сезона, а также для каждого пункта свое значение. В таком случае количество констант в программе увеличивается и несколько усложняется программа контроля, так как каждый раз в наборе констант необходимо отыскивать нужное значение, соответствующее данному месяцу и конкретному посту.

Алгоритм контроля на пределы нашел широкое применение при анализе результатов гидрометеорологических наблюдений. Он может использоваться при контроле как оперативной, так и режимной информации. Метод контроля на пределы по своему характеру близок к синтаксическому контролю и является первым этапом логического контроля.

Алгоритмы логического контроля оперативной информации строятся в основном на сравнении между собой значений элементов, наблюдаемых одновременно во многих соседних пунктах. Сюда относятся так называемые методы объективного анализа. На данном этапе проверяется согласованность наблюдений соседних станций. Этот прием называется «горизонтальным контролем» в отличие от «вертикального контроля», при котором используется аэрологическая информация только одного пункта наблюдения. В качестве одного из основных алгоритмов при этом применяется метод оптимальной интерполяции. Методика горизонтального контроля основывается на следующем.

С помощью метода оптимальной интерполяции по данным о значении элемента на окружающих станциях вычисляют значение его на проверяемой станции. Затем находят разность  $\Delta\Delta\text{ЛАГ} = X_p - X_n$  между рассчитанным ( $X_p$ ) и наблюдаемым ( $X_n$ ) значениями элемента. Если разность окажется большой, то можно высказать предположение о наличии ошибки в исходной информации данной станции. Допустимое значение разности определяется экспериментальным путем по результатам предыдущих наблюдений и закладывается в программу контроля. При оптимальной интерполяции значение элемента в контролируемой точке (принимая его за неизвестное) вычисляется сразу по нескольким соседним станциям, показания которых берутся с некоторыми весовыми коэффициентами, устанавливаемыми на основе многолетних данных. Эти коэффициенты выбираются таким образом, чтобы средняя квадратическая ошибка интерполяции была минимальной.

Наличие значительных расхождений между вычисленным и наблюдаемым значениями элемента в точке еще не указывает на местоположение ошибки, которая может находиться как в данных контролируемой станции, так и в показаниях соседних станций. Для определения ее местоположения из расчета последовательно исключают каждую из окружающих точек. Если в результате такого исключения разность становится допустимой, можно предположить, что ошибочны данные как раз этой точки из окружения. Если же при последовательном исключении всех точек разность не уменьшилась до требуемого значения контролируемой станции, то она предполагается ошибочной и подлежит исправлению. Исправление осуществляется путем замены ошибочного значения

на значение элемента, вычисленное методом оптимальной интерполяции. В этом большое преимущество данного метода перед теми, которые не позволяют автоматически исправить ошибочное значение.

Описанный выше алгоритм позволяет довольно легко обнаружить одиночные станции с ошибочными данными наблюдений. Однако если грубые ошибки содержатся одновременно в данных наблюдений на нескольких станциях, расположенных близко друг от друга, то в отыскивании ошибок возникают определенные трудности. В этом случае ошибочные значения, используемые одновременно в обработке, могут взаимно уменьшать разность между вычисленным и наблюдаемым значениями контролируемого элемента и тем самым делают невозможным их поиск. Поэтому в гидрометеорологии прибегают к повторному горизонтальному контролю, а также к предварительному устранению грубых ошибок с помощью так называемого «вертикального контроля».

Методы вертикального контроля наиболее применимы для тех видов информации, которые включают в себя набор сведений об одних и тех же элементах за несколько сроков наблюдений или по нескольким высотам (уровням) в одной и той же точке (станции). К таким видам относится, например, аэрологическая информация, структура сообщения которой содержит сведения о температуре, давлении, влажности воздуха и других характеристиках на различных высотах от поверхности земли в пункте наблюдения. Другим примером информации, пригодной для вертикального контроля, являются все виды режимных данных, так как они включают большое количество сведений об одних и тех же элементах (от срока к сроку) в пункте наблюдения.

При вертикальном контроле, как и при других видах контроля, используются проверки на соответствие контролируемой величины различного рода физическим представлениям. Например, при контроле аэрологических данных некоторые из элементов сообщения (геопотенциал изобарических поверхностей и температура воздуха) должны соответствовать уравнению статики (статический контроль). При этом учитывается изменение температуры воздуха с высотой. В ряде случаев статический контроль позволяет не только обнаружить, но и исправить ошибочные данные или восполнить отсутствующие.

При контроле согласованности данных по высоте (или по глубине, если речь идет об океанологической и гидрологической информации) можно обнаруживать случайные ошибки разных типов, а именно: ошибки в наблюдениях, расчете некоторых элементов, кодировании и передаче данных. Ошибки, связанные с неисправностью измерительной аппаратуры или с неправильным снятием показаний приборов, таким путем непосредственно не обнаруживаются, так как при них в данных не возникает несоответствия между значениями элемента на различных высотах (глубинах). Подобные ошибки могут быть выявлены лишь с помощью приемов горизонтального контроля, в том числе и путем объективного анализа гидрометеорологических полей.

Другим примером метода вертикального контроля служит проверка согласованности данных по срокам наблюдений. Последовательность операций здесь примерно та же, что и при проверке согласованности данных по высоте (глубине). Согласно этому алгоритму анализируется значение элемента в пункте наблюдений от срока к сроку. При этом анализ может осуществляться несколькими способами, из которых наиболее распространенным является интерполяционный.

При *интерполяционном способе* проверки согласованности наблюдений от срока к сроку каждое из них последовательно исключается из ряда и восстанавливается путем интерполяции между наблюдениями в смежные сроки. После такого восстановления будем иметь за один и тот же срок два значения элемента – наблюдаемое ( $X_n$ ) и вычисленное ( $X_b$ ). Далее определяется разность между этими двумя значениями:

$$\Delta X = X_n - X_b,$$

по величине которой можно судить о достоверности контролируемого элемента наблюдения ( $X_n$ ). Суждение о достоверности, сомнительности или ошибочности наблюдения принимается по схеме:

если  $|\Delta x| \leq c_1$ , то – достоверное значение;

если  $c_1 < |\Delta x| \leq c_2$ , то  $x_n$  – сомнительное значение;

если  $|\Delta x| \geq c_2$ , то  $x_n$  – ошибочное значение.

Здесь  $c_1$  – некоторая числовая константа, за которую при достоверных значениях  $x_n$  не должно выходить абсолютное значение  $\Delta x$ ;  $c_2$  – другая константа, выход за которую абсолютного значения  $\Delta x$  происходит только в тех случаях, когда значение  $x_n$  явно ошибочно.

Очевидно, что если величина  $\Delta x$  попадает в диапазон между  $c_1$  и  $c_2$ , то соответствующее ей контролируемое значение  $x_n$  должно быть отнесено к разряду сомнительных, так как абсолютное значение  $\Delta x$ , с одной стороны, уже слишком велико, чтобы значение  $x_n$  могло быть безусловно принято за достоверное, но, с другой стороны, оно еще недостаточно велико, чтобы контролируемое значение можно было отнести к разряду ошибочных. Для значений  $x_n$ , отнесенных к сомнительным, применяются дополнительные приемы анализа в целях отнесения их либо к достоверным, либо к ошибочным. Эти приемы основаны, как правило, на использовании некоторой дополнительной информации данного или других сообщений.

Если сомнительное значение удалось отнести к разряду достоверных, то далее над ним осуществляются все необходимые вычислительные операции. Если значение признано ошибочным, оно может быть автоматически исправлено машиной. Чем меньше абсолютная разность между наблюдаемым и соответствующим ему вычисленным (восстановленным) значением, тем больше вероятность того, что значение контролируемого элемента является достоверным.

Константы  $c_1$  и  $c_2$  заблаговременно вводятся в ЭВМ и непосредственно включаются в состав машинных программ. Следует, однако, сказать, что при определении численного значения величин  $c_1$  и  $c_2$  для каждого конкретного элемента гидрометеорологических наблюдений встречаются серьезные трудности. Действительно, если назначить  $c_1$  и  $c_2$  слишком малыми, то очень большое количество наблюдений будет относиться машиной к сомнительным или ошибочным. При этом возникает необходимость ручного дополнительного анализа или даже исключения (бракования) правильных данных.

Некоторое распространение в программах автоматизированного контроля гидрометеорологической информации получили так называемые статистические методы. Суть их заключается в том, что результаты наблюдений за определенный промежуток времени (например, за месяц) рассматриваются как случайный стационарный ряд, к которому применимы определенные статистические законы. Эти законы устанавливают определенные соотношения между значениями элементов числового ряда и частотой их появления. Статистические методы позволяют отыскивать



лишь грубые промахи. Недостатком этих методов является то, что они не позволяют исправить ошибочное значение. Для исправления ошибки приходится пользоваться другими методами.

Систематические ошибки, которые обычно не достигают больших значений, практически невозможно контролировать описанными выше способами. Контроль таких ошибок основан на том свойстве, что их влияние сказывается тем больше, чем за более длительный период времени анализируется информация. Проникая в данные каждого наблюдения с одним и тем же знаком, ошибки суммируются при суммировании наблюдаемых значений. Такое суммирование, как известно, производится при подсчете средних значений (за месяц, год). Именно эти средние значения и подвергаются дополнительному контролю на предмет выявления систематических ошибок. При этом используются методы горизонтального контроля. Средние значения контролируемой станции сравниваются с такими же данными той из соседних станций, по которой в предыдущие годы существовала устойчивая связь.

Большинство алгоритмов контроля дает положительный результат лишь в том случае, если количество ошибок в информации невелико, и если они имеют случайный характер. Однако и в этом случае часть ошибок в информации не будет обнаружена. К ним относятся ошибки, которые находятся в пределах естественных колебаний элемента. Для получения лучших результатов в автоматизированном контроле обычно используют сочетание описанных выше методов.

### **2.5. Особенности автоматизированной обработки гидрометеорологической информации**

Автоматизация обработки данных – это переработка информации в целях решения серий последовательных задач, в которых результаты решения одной задачи являются сигналом к автоматическому переходу для решения другой, в том числе и к формированию соответствующих документов, выражающих результаты решения.

Обработка данных от момента получения до момента выдачи выходной продукции потребителю проходит ряд технологических этапов:

- получение данных в результате измерений (наблюдений) в пунктах наблюдений;

- подготовка данных для обработки на ЭВМ;
- первичная обработка в целях формирования фонда данных длительного хранения;
- длительное хранение данных;
- обработка данных длительного хранения.

В ходе этого процесса данные группируются, контролируются, переносятся с одного носителя на другой, подвергаются перестройке структуры и т. д.

Данные, полученные после завершения этапа обработки, называются данными соответствующего уровня. С каждым уровнем связываются две подсистемы: подсистема формирования данных этого уровня (этапы технологического процесса) и подсистема обслуживания потребителей на основе этого уровня.

Обслуживание может вестись с различной степенью автоматизации в зависимости от уровня данных. Неавтоматизированное обслуживание ведется в основном за счет первичных данных наблюдений, не переданных в архив, а также за счет информации, хранящейся в виде листовых материалов и справочников.

Автоматизированное обслуживание предполагает представление информации в более разнообразных формах: технический носитель информации, графический материал, передача таблиц по каналам связи.

Многие системы обработки гидрометданных имеют фиксированный ритм функционирования, определяемый расписанием работы систем наблюдений и каналов передачи данных, расписанием подготовки выходной продукции (обработанных данных, прогнозов, справочных пособий).

Системы, участвующие в подготовке прогностической информации, составляют оперативную гидрометеорологическую систему, а системы, обеспечивающие накопление данных для хранения и подготовки выходной справочных материалов о гидрометеорологическом режиме, – режимную гидрометеорологическую систему.

Основной задачей оперативной гидрометеорологической системы является выпуск следующей продукции.

1. Прогноз погоды, гидрологического режима и других, связанных с ним явлений.



2. Информация о текущем состоянии погоды и гидрометеорологического режима.

Основное требование, которое должно быть выполнено при проектировании и разработке таких систем, – это скорость получения выходной продукции. С этой целью задается время отсечения, т. е. время, в течение которого после очередного срока наблюдений осуществляются сбор и накопление информации. Таким образом, чтобы обеспечить своевременность поступления информации, в жертву приносится полнота, а часто в определенной мере и качество собираемых комплектов данных.

Выходная продукция систем оперативной обработки: таблицы, графики, карты, алфавитно-цифровые сообщения в каналах связи.

Автоматизированная обработка оперативной гидрометеорологической информации включает несколько этапов: ввод с канала связи и запоминание в ЭВМ исходной информации, декодирование и опознание сообщений, контроль и исправление данных, вычисление прогнозируемых величин, выдача результатов расчета, архивация. Иногда этапы работ, предшествующие непосредственно вычислению прогнозируемых величин, объединяют под общим названием «первичная обработка».

Вторичная обработка данных включает в себя расчеты фактических и будущих значений метеоэлементов в узлах регулярной сетки или других обобщенных характеристик метеорологических полей. Важнейшим этапом обработки оперативной информации является формирование полученных результатов для включения их в так называемую информационно-поисковую систему, основное назначение которой состоит в обмене информацией с другими центрами или в выдаче ее из ЭВМ для последующего анализа специалистом.

В настоящее время сбор и распространение оперативной гидрометеорологической информации осуществляются системой радиометеорологических центров (РМЦ) различных категорий. Эти центры действуют в соответствии с регламентом ВМО.

Сводки, принимаемые с каналов связи, содержат данные наблюдения или иную гидрометеорологическую информацию (например, данные различных прогнозов, штормовые предупреждения и др.). Каждый вид сводок имеет определенную структуру и код. Вид сводки обычно указы-

вается в ее заголовке либо с помощью специальных названий, либо специальными индексами.

В сводках различаются следующие части: символы начала связи; заголовок сводки; текст сводки; символы окончания сводки.

Поступающие на вход ЭВМ закодированные гидрометеорологические данные сортируются по видам наблюдений. Далее информация переводится в машинный код. На этом же этапе выявляются, анализируются и исправляются некоторые погрешности поступившей информации. Такими погрешностями бывают слияние двух соседних групп, разделение группы на две части и т. д. При этом используются приемы синтаксического контроля. Информация, переведенная в машинный код и частично исправленная, комплектуется компактно и записывается на внешнюю память ЭВМ.

В ходе дальнейшей обработки сообщений восстанавливаются наблюдавшиеся значения метеозлементов по их кодовым значениям. Телеграммы конкретного вида данных в общем потоке информации обнаруживают по отличительным группам заголовка.

Одним из существенных элементов системы обработки оперативной гидрометеорологической информации является так называемый словарь или каталог пунктов наблюдений. В простейшем случае *словарь* — это хранимый в памяти машины список условных номеров (индексов) пунктов наблюдений. Наряду с этими номерами словарь может включать и другую информацию о станции, необходимую для обработки данных. Сюда относятся географические координаты пункта наблюдения, высота его над уровнем моря, сведения о единицах измерения различных гидрометеорологических элементов. В словарь включаются также дополнительные сведения, касающиеся последующей обработки: тип информации данного пункта, порядок дальнейшей обработки (в частности, использовать ли данные по этой станции при анализе карты того или иного масштаба или в иных видах обработки), место во внешней памяти, куда следует после обработки записывать исходную информацию, и др.

После поиска и декодирования телеграмм одного вида наблюдений осуществляется автоматизированный контроль по алгоритмам, описанным ниже. На этом заканчивается первичная обработка оперативной информации. Информацию, полученную в результате этой обработки, мож-

но использовать для различных целей. Однако в ряде случаев требуется дальнейшая обработка, а полученные данные оказываются лишь промежуточными результатами в сложном комплексе расчетов. На следующем этапе решаются такие основные задачи, как:

- 1) подготовка исходных данных для численного прогноза;
- 2) построение карт гидрометеорологических элементов;
- 3) преобразование информации к виду, удобному для длительного хранения, в целях использования ее для получения различных справочных сведений.

При численном прогнозе исходными материалами служат не данные наблюдений неравномерно расположенной сети станций, а данные в точках («узлах») правильной («регулярной») сетки. Такая сетка может быть, например, прямоугольной, треугольной, географической (узлы совпадают с пересечениями меридианов и параллелей) и др. Следовательно, необходимым этапом подготовки данных для численного прогноза является расчет значений гидрометеорологических элементов в узлах регулярной сетки по данным наблюдений на станциях. Наряду с использованием для целей численного прогноза значения гидрометеорологических элементов в узлах регулярной сетки могут служить для вычисления диагностических полей различных расчетных элементов (вихря скорости, вертикальной компоненты скорости и т. п.). Процесс определения значений гидрометеорологического элемента в узлах регулярной сетки получил название объективного численного анализа.

Следующий этап машинной обработки информации – построение карт гидрометеорологических элементов (например, карт барической топографии, карт приземного давления и др.). Важным этапом подготовки таких карт является расчет положения изолиний элементов (изогипс, изобар, изотерм и т. п.).

Исходная информация после обработки преобразуется к виду, удобному для длительного хранения в целях использования ее для получения различных справочных сведений и режимной обработки. В процессе преобразования информация уплотняется и записывается на технические носители.

В задачу объективного анализа входит контроль за правильностью наблюдений, а также согласование между собой данных о различных ме-

теорологических элементах или же данных – различных наблюдательных систем. В более широком плане численный анализ может включать в себя обнаружение фронтов, проведение изолиний и т. д. Однако главная задача численного анализа – получение значений метеорологических элементов в узлах сетки.

Программа объективного анализа оперативной гидрометеорологической информации на ЭВМ состоит из подпрограмм:

- 1) ввода;
- 2) контроля исходной информации;
- 3) оптимальной интерполяции;
- 4) вывода результатов анализа или стыковки с программой прогноза.

Блок поиска станций устанавливает, в каком квадрате располагается анализируемый узел, выясняет, получены ли данные наблюдений со станций, расположенных в этом квадрате, и в каком участке оперативной памяти ЭВМ эти данные хранятся. Если в квадрате, охватывающем узел, имеются данные менее чем с восемью станциями, то аналогичным образом проводится осмотр ближайших квадратов и извлечение соответствующей информации.

Для целей поиска информации поверхность географической карты разбивается на квадраты. В памяти ЭВМ формируется «справочная таблица», показывающая, на каком участке памяти располагается информация, относящаяся к каждому квадрату. Если в данном квадрате станций меньше, чем требуется для интерполяции, то осуществляется поиск данных в соседних квадратах.

В блоке оптимальной интерполяции вычисляются координаты обрабатываемого пункта (узла регулярной сетки или проверяемой станции), и с помощью блока поиска станций отбирается информация с нескольких (обычно с восьми) ближайших станций. Затем рассчитываются расстояния влияющих станций до обрабатываемого пункта (узла) и расстояния между влияющими станциями. Расстояния ( $r_{ij}$ ) вычисляются по формуле

$$r_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}, \quad (2.53)$$

где  $x_i, y_i$  и  $x_j, y_j$  – координаты  $i$ -й и  $j$ -й станций в декартовой системе.

По вычисленным расстояниям определяется значение весовых коэффициентов  $P_i$  для каждой станции. Эти коэффициенты зависят от расстояния и статистической структуры поля, которая устанавливается по результатам многолетних наблюдений. В конечном итоге с помощью коэффициентов  $P$ , обеспечивается интерполяция значений элемента в заданную точку с минимальной погрешностью. После того как эти коэффициенты для каждой из станций определены, искомое значение элемента ( $\bar{f}$ ) в заданной точке можно определить по формуле

$$f = \sum_{i=1}^n P_i f_i, \quad (2.54)$$

где  $f_i$  - значения элемента на  $j$ -й станции, а  $n$  - число окружающих станций.

Вычисленные по этой формуле значения элементов в узлах регулярной сетки используются для автоматического построения различных диагностических карт гидрометеорологических элементов.

Построение карт с помощью ЭВМ основано на следующих принципах:

- изображение карты путем нанесения цифр или иных символов с использованием технических устройств;
- вычерчивание изолиний карты, а также нанесение цифр и других символов на бланки карты с помощью принтера.

К наиболее распространенным в практике способам машинной выдачи карт в настоящее время относится способ изображения карт с помощью принтеров; применяются также геоинформационные системы и компьютерная графика. Такими картами можно пользоваться как для обычной работы прогнозиста, так и для последующей передачи по каналам связи.

Блок-схема автоматизированной обработки оперативной гидрометеорологической информации представлена на рис. 2.10.

При численном прогнозе исходными материалами служат не данные наблюдений неравномерно расположенной сети станций, а данные в узлах регулярной сетки, полученные в результате объективного анализа.



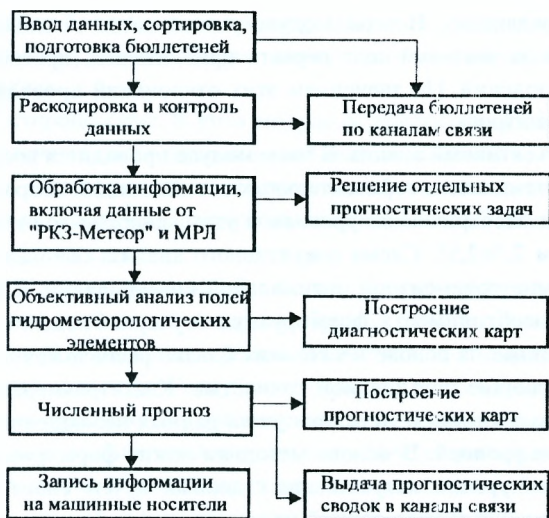


Рис.2.10. Блок-схема автоматизированной обработки оперативной гидрометеорологической информации

Увеличение объема информации, необходимой для анализа состояния атмосферы, дают модели численного прогноза погоды, способные с большой точностью экстраполировать во времени метеорологические поля на основе данных наблюдений за предшествующий срок. Процесс отображения на заданную сетку реального состояния атмосферы путем слияния данных наблюдений с прогностическим решением гидродинамической модели принято называть усвоением данных или четырехмерным объективным анализом. При этом схемы и методы прогноза зависят от вида наблюдений и типа прогноза.

Система четырехмерного усвоения данных о состоянии атмосферы включает в себя циклическое (через каждые 6 ч) выполнение следующей последовательности основных модулей.

1. Модуль первичной обработки данных. На этапе подготовки набора с данными наблюдений производится выборка из базы данных и накопление метеорологических сообщений о наблюдениях за состоянием атмосферы за интервал времени  $\pm 3$  ч от срока усвоения.

2. Преданализ. В этом модуле вычисляются отклонения данных наблюдений от значений поля первого приближения (прогноза на 6 ч) в точках наблюдений. По значениям этих отклонений проводится предварительный контроль.

3. Объективный анализ. В этом модуле проводится восстановление полей геопотенциала и горизонтальных составляющих скорости ветра на стандартных изобарических уровнях в узлы широтно-долготной сетки с разрешением  $2,5 \times 2,5^\circ$ . Схема оперативного анализа основана на квазитрехмерной многоэлементной оптимальной интерполяции. В рамках этой схемы при необходимости формируются «супернаблюдения», т. е. данные, полученные на основе нескольких близко расположенных наблюдений и обладающие повышенной точностью. Квазитрехмерность определяется с использованием методики учета данных наблюдений с соседних по вертикали уровней. В основе методики лежит формирование на рассматриваемом уровне искусственных данных путем сноса отклонений фактических данных наблюдений от прогностических значений с соседних уровней с последующей двумерной оптимальной интерполяцией. По квазитрехмерной схеме проводится анализ метеорологических полей примерно в 7 раз быстрее по сравнению с трехмерной схемой без существенной потери точности.

4. Препроцессинг. В модуле препроцессинга осуществляется переход от переменных объективного анализа (геопотенциал и горизонтальные составляющие скорости ветра), заданных в узлах географической сетки на стандартных изобарических уровнях, к переменным гидродинамической модели (приземное давление, температура и горизонтальные составляющие скорости ветра), заданным в узлах гауссовой сетки на сигма-уровнях.

5. Инициализация. Из-за несбалансированности начальных полей ветра и геопотенциала в начале интегрирования гидродинамической модели возникают высокочастотные шумы. Для подавления высокочастотного шума применяются разные способы согласования (инициализации) начальных полей ветра и геопотенциала. К числу наиболее эффективных относится метод нелинейной инициализации по нормальным моделям.

6. Прогностическая модель. Спектральная модель атмосферы T40L15 обеспечивает гидродинамическую экстраполяцию метеорологи-



ческих полей на 6 ч вперед от предыдущего срока, восполняя отсутствие информации в тех районах, где нет данных наблюдений на момент объективного анализа.

7. Постпроцессинг. В этом модуле осуществляется обратный переход от переменных гидродинамической модели, заданных в узлах гауссовой сетки на сигма-уровнях, к переменным объективного анализа, заданным в узлах равномерной широтно-долготной сетки на стандартных изобарических уровнях.

8. Модуль пересчета ошибок полей первого приближения. Данный блок позволяет взвесить некоторым образом относительный вклад данных наблюдений и численного решения спектральной модели.

В Гидрометцентре Республики Беларусь используется прогностическая модель прогноза барического поля, основных метеозлементов и явлений погоды с заблаговременностью до 36 ч, разработанная в 80-е годы. Отсутствие в ее составе самых последних достижений в области моделирования атмосферных процессов компенсируется учетом местных условий. Основными модулями программы являются модуль объективного анализа синоптической информации (GOLNC), модуль высотного комплекса (ВКОМ), модули прогноза геопотенциала, вертикальных токов и траекторий (PGWT), оценки качества прогноза барического поля (OCENK), построения адвентивной стратификации (ADVES) и прогноза метеозлементов и явлений погоды (PREDIC).

## **2.6. Обработка режимной гидрометеорологической информации**

Обработка в целях создания фонда данных по гидрометеорологии и окружающей среде для обслуживания потребителей должна обеспечить накопление комплектов гидрометеорологических данных, прошедших более строгий контроль качества. При этом принимаются специальные меры для получения наиболее полных комплектов данных путем организации перезапроса сомнительных или не поступивших данных. Выходная продукция этих систем: массивы гидрометданных на технических носителях для длительного хранения, ежемесячники, ежегодники, которые содержат как данные первичных наблюдений, так и обобщения за месяц или год и справочные материалы спецприменения (еженедельные, ежедекад-

ные, ежемесячные обобщения для организаций). Особенности этих систем следующие:

- большие возможности применения типовых программ из-за меньших требований к эффективности;
- при организации контроля данных большее участие человека в технологии обработки;
- большой объем массивов, участвующих в обработке (большее время отсечения, междустанционный контроль, межэлементный контроль, возможность временного контроля).

Режимную обработку гидрометеорологических данных с фиксированным циклом иногда называют первичной обработкой по аналогии с соответствующим этапом в обработке данных автоматизированных систем управления (АСУ).

Основные задачи первичной обработки: контроль достоверности и полноты данных, преобразование данных к виду, удобному для дальнейшего использования; получение агрегированной информации (обобщение за декаду, месяц или по группе объектов, заданной территории), сбор сведений о функционировании системы первичной обработки.

Массивы гидрометданных на технических носителях для длительного хранения предназначены для многоцелевого использования. Поэтому к их организации предъявляются разнообразные, очень часто противоречивые требования, как по комплектации, так и способу упорядочения.

Обработка режимной гидрометеорологической информации осуществляется на основе регламентирующего документа: «Временные методические указания по машинной обработке и контролю данных гидрометеорологических наблюдений» (ВМУ). Он включает правила кодирования данных, порядок их сбора, методы контроля, обработки и хранения информации (рис. 2.11.).

Технология обработки включает несколько этапов: ввод и запоминание исходной информации в ЭВМ, раскодирование, контроль данных и исправление ошибок, вычисление различных характеристик гидрометеорологического режима, выдачу результатов расчета, комплектование информации для хранения. Здесь присутствуют многие из этапов, которые были названы в описании технологии обработки оперативной информа-

ции. Однако содержание таких совпадающих этапов иногда сильно различается. Обработке информации на ЭВМ предшествуют некоторые предварительные операции, которые заключаются в получении в центре обработки подготовленных к вводу в ЭВМ данных.

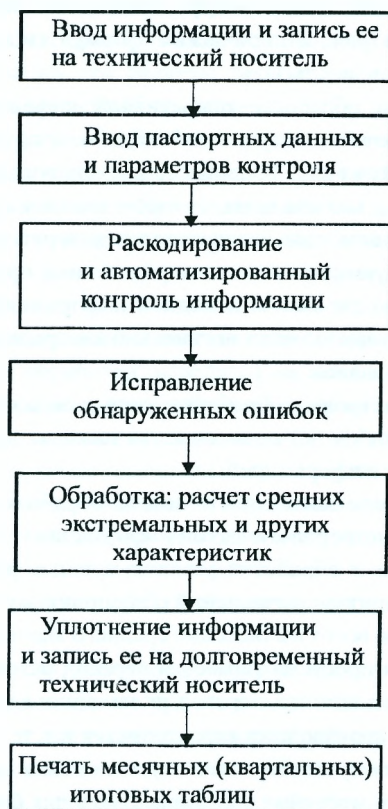


Рис. 2.11. Блок-схема первого этапа обработки режимной информации

Первым этапом автоматизированной обработки информации является ввод исходных данных в ЭВМ. Одновременно с результатами наблюдений в оперативную память вводятся так называемые паспортные

данные по каждому пункту, которые до начала обработки хранятся на отдельном техническом носителе. В состав паспортных данных входят обычно следующие сведения: географические координаты пункта, его название и название объекта, над которым осуществляются наблюдения, параметры контроля в виде набора предельных значений элемента или значений параметров эмпирических формул, используемых для контрольных расчетов, и т. п.

Как и при обработке оперативной информации, ЭВМ вначале должна раскодировать поступившие сведения и привести их к определенному «машинному» виду. В процессе раскодирования ЭВМ переводит в машинную форму все символы, которые заносились в соответствующем коде. Цифровые символы переводятся в десятичные цифры, буквенные символы используются в качестве определенных признаков информации.

После раскодирования исходной информации и перевода ее в машинные коды осуществляется логический контроль данных, методики которых уже описывались.

В итоге контроля и обработки данных по всем пунктам составляется ведомость ошибок, которая является важным документом для последующей работы с информацией.

Общая продолжительность цикла обработки месячного массива режимной гидрометеорологической информации – 2 – 2,5 месяца.

В программы обработки различных видов режимной информации включаются алгоритмы вычислений общепринятых характеристик. К ним относится прежде всего вычисление среднего значения элемента за сутки. В ряде программ предусматриваются перевод различных показаний приборов в значения элемента, перевод их из одних единиц измерения в другие, вычисление производных характеристик и т. п.

По средним суточным значениям элемента вычисляют средние декадные и средние месячные значения. Далее для большинства элементов гидрометеорологических наблюдений из ряда выбирают максимальные и минимальные значения. Одновременно с поиском максимальных и минимальных значений того или иного гидрометеорологического элемента определяют даты, к которым относятся эти найденные значения.

В процессе дальнейшей обработки данных некоторых видов наблюдений вычисляются другие статистические характеристики. Сюда от-

носятся прежде всего подсчет числа случаев того или иного явления и превышения определенных числовых пределов, а также распределение значений элемента по градациям. В ряде программ обработки предусматриваются вычисление средних квадратических отклонений ряда, расчет автокорреляционных и спектральных функций, проведение дисперсионного анализа и т. п. Однако для большинства видов режимной агрометеорологической информации эти расчеты производятся лишь для более длительных интервалов времени (больше месяца).

Некоторые из программ обработки режимной гидрометеорологической информации имеют дополнительные блоки для обработки данных, считываемых с лент самописцев. Эти данные обычно дополняют основную информацию и получаются в результате обработки лент, производимой на станциях вручную или с помощью полуавтоматизированных считывающих устройств.

Обработанная информация движется далее по двум направлениям. Во-первых, результаты обработки выводятся на печать через принтер и представляются в виде разного рода таблиц. Одновременно обработанная информация записывается во внешнюю память ЭВМ (технический носитель) для накопления за более длительные промежутки времени. Таблицы, содержащие результаты первичной обработки месячных массивов информации, печатаются на ЭВМ в нескольких экземплярах, часть которых из центров обработки направляют на станции и в обсерватории для практического использования и обслуживания народнохозяйственных организаций.

Полученные на ЭВМ таблицы проходят на станциях и в обсерваториях критический контроль. Если в процессе такого контроля будут обнаружены ошибки, то об этом извещается центр обработки, где в информацию вносятся соответствующие исправления. После того как вся информация за месяц собрана, проконтролирована, обработана и в необходимых случаях исправлена, приступают к получению с помощью ЭВМ таблиц ежемесячников. Эти таблицы поступают из отдела ЭВМ в группу выпуска для критического контроля, взаимной увязки данных соседних станций. Далее таблицы размножаются с помощью копировальной техники. Экземпляры таблиц используются для комплектования ежемесячни-



ков, которые направляются в различные организации для практического применения.

Как правило, месячные таблицы печатаются по установившейся форме, однако иногда в них включаются дополнительные, ранее не публиковавшиеся характеристики. Так, в ряде случаев в таблицах для наглядности приводится графическое изображение хода того или иного элемента. Кроме того, включаются параметры, которые часто запрашиваются различными организациями. В целом таблицы с режимной гидрометеорологической информацией, полученные в результате ее автоматизированной обработки, более приспособлены для удовлетворения практических запросов, поступающих от различных организаций.

При записи во внешнюю память информация предварительно преобразуется и уплотняется. При этом уплотнение непрерывно или учащенно регистрируемой (с помощью автоматизированных устройств) информации производится после дополнительного анализа. Вполне очевидно, что если используется устройство, регистрирующее данные довольно часто и с постоянным шагом во времени, то в обработку вовлекается некоторая избыточная информация. Ее включение в механизированный фонд данных приведет к увеличению объема этого фонда, возрастанию непроизводительных затрат на хранение, поиск и последующую обработку данных. Для исключения избыточности информации анализируется плавность хода элемента наблюдения во времени.

Каждая из программ обработки режимной информации обязательно предусматривает сжатие или упаковку информации перед ее записью во внешнюю память, т. е. архивацию.

Кодирование и уплотнение информации осуществляются по заранее разработанному макету, структура которого определяется свойствами самой информации, в частности, максимально возможными размерами (значениями) того или иного элемента наблюдения.

Под уплотнением информации понимается ее компактное представление путем уменьшения избыточности и использования рациональных методов кодирования.

Записью месячных (в некоторых случаях трехмесячных – квартальных) объемов информации на долговременный технический носитель заканчивается лишь первый этап обработки (см. рис. 3.2). На долговре-

менном техническом носителе режимная информация накапливается в течение года, по окончании которого осуществляется второй этап обработки информации. Основным содержанием второго этапа являются сортировка данных на технических носителях, создание годовых массивов информации на основе разрозненных месячных объемов наблюдений, а также расчеты и печать годовых таблиц с различными характеристиками гидрометеорологического режима. На этом же этапе осуществляется переписывание скомплектованных массивов информации на долговременный носитель, хотя по отдельным видам данных создание архива на долговременных носителях может производиться и ежемесячно.

Одновременно с комплектованием годового массива вычисляются различные характеристики гидрометеорологического режима в годовом разрезе: средние годовые, максимальные и минимальные, повторяемости по градациям и др. Эти величины вычисляются по тем же алгоритмам, которые использовались при обработке месячных массивов.

При комплектовании результатов обработки на долговременных технических носителях информация часто сортируется по степени обобщения на несколько массивов, хранящихся далее на отдельных долговременных технических носителях. В числе таких массивов могут быть:

- а) данные по срокам наблюдения;
- б) средние суточные характеристики;
- в) месячные выводы;
- г) годовые сведения.

Такое разделение информации необходимо при последующих расчетах, в которых не требуются первичные (срочные) наблюдения, позволяет вводить в ЭВМ лишь обобщенные характеристики и за счет этого экономить машинное время. Объем данных в каждом из перечисленных массивов быстро убывает в зависимости от степени обобщения.

Особенно выгодно раздельное хранение информации в виде нескольких массивов, скомплектованных по степени обобщения для многосрочных наблюдений.

Полученные с помощью ЭВМ материалы, как правило, не требуют какой-либо доработки вручную. Тем не менее таблицы и графики должны пройти хотя бы выборочный контроль и редактирование, осуществляемые специалистами-гидрометеорологами. Участие специалистов, хорошо



знающих местные особенности, в просмотре результатов машинной обработки наблюдений способствует повышению качества информации. Выдачей на печать годовых таблиц и графиков и созданием базы данных на технических носителях заканчивается второй этап обработки информации. Блок-схема приведена на рис. 2.12.

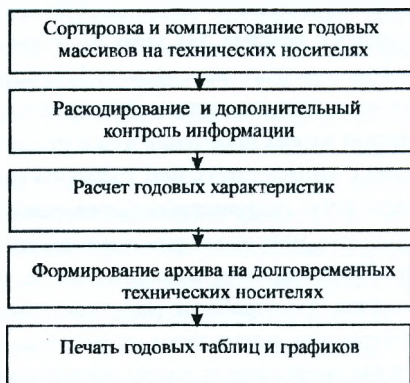


Рис. 2.12. Блок-схема второго этапа обработки режимной информации

### 3. ПРИНЦИПЫ ВЫЯВЛЕНИЯ И АНАЛИЗ ФУНКЦИОНАЛЬНЫХ СВЯЗЕЙ СЛУЧАЙНЫХ ВЕЛИЧИН

#### 3.1. Общие принципы выявления функциональных связей

При анализе и прогнозе природных процессов редко приходится иметь дело с функциональными связями, когда каждому значению одной величины соответствует строго определенное значение другой величины. Чаще встречаются такие соотношения между переменными, когда каждому значению признака  $x$  соответствует не одно, а множество возможных значений признака  $y$ , т. е. их распределение. Такие связи, обнаруживаемые лишь при массовом изучении признаков, в отличие от функциональных, называются *стохастическими* (вероятностными) или *корреляционными*.

Математический анализ связей, существующих между случайными величинами  $x$  и  $y$ , составляет содержание корреляционного анализа. Корреляционный анализ сводится, прежде всего, к измерению степени тесноты сопряженности между варьирующимися признаками, в качестве показателя которой наибольшее распространение получил линейный коэффициент корреляции. Корреляционный анализ включает в себя определение формы и направления существующей между  $x$  и  $y$  связи и др.

Однако корреляцию не следует отождествлять с причинностью. Например, минерализация воды понижается с севера на юг, в этом же направлении понижается содержание питательных веществ в почве. Между рассматриваемыми показателями может быть получена положительная достоверная зависимость. Однако степень минерализации воды не определяет оптимальное содержание питательных веществ в почвенном покрове. Иначе в ландшафтах пустынь плодородие было бы максимальным, так как здесь самая высокая минерализация воды, а это противоречит истине. Поэтому установление подобной связи при проведении теоретических исследований бессмысленно.

По форме корреляционная связь бывает *линейной* и *нелинейной*, по направлению – *прямой* и *обратной*, по величине – от 0 до  $\pm 1$ , по количеству коррелируемых признаков – *парной* и *множественной*.

*Парная корреляционная связь может быть:*

- параллельно-соотносительной (ассоциативной) – оба признака изменяются сопряженно, частично под действием общих причин и

следствий (приуроченность растительности и почв к формам рельефа);

- субпричинной – один фактор выступает как отдельная причина сопряженного изменения признака (связь урожайности с динамикой почвенных влагозапасов);
- взаимоупреждающей – причина и следствие, находясь в устойчивой взаимосвязи, последовательно влияют друг на друга (гидрометеорология и естественная увлажненность водосборов).

Если зависимость между признаками указывает на линейную корреляцию, то необходимо рассчитывать коэффициент корреляции ( $r$ ), который позволяет, с одной стороны, оценить тесноту связи переменных величин, с другой, – выяснить: какая доля изменений признака обусловлена влиянием основного фактора, какая – влиянием других факторов. При положительной зависимости величина коэффициента корреляции изменяется от 0 до +1, при отрицательной – от 0 до -1. Если  $r=0$ , то связь между признаками отсутствует. Принято считать, что при  $r<0,5$  корреляционная связь слабая, при  $r=0,5...0,7$  – средняя, при  $r=0,7...0,99$  – сильная [Брукс, Карузере, 1963].

Однако коэффициент корреляции приближенно характеризует тесноту связи между признаками. Поэтому иногда при высоком значении коэффициента корреляции и небольшом объеме выборки связь между признаками может быть слабой. Мерой корреляционной связи является коэффициент детерминации, определяемый по зависимости:

$$d_{xy}=r^2. \quad (3.1)$$

Коэффициент детерминации указывает на долю взаимной связи между признаками. Например, если  $r=0,5$ , то  $d_{xy}=0,25$ , т. е. 25% всех изменений одного признака связано с изменением другого. Отсюда следует, что, при  $r\geq 0,70$ , истинная взаимообусловленность признаков составляет около 50%.

Одна и та же величина коэффициента корреляции будет по-разному определять достоверность зависимости признаков для разных выборок. Например, при  $P=0,95$ , для  $n=5$ , достоверны значения  $r\geq 0,878$ , для  $n=20$ , – достоверной величиной будет  $r\geq 0,44$ , для  $n=100$  – достоверны значения  $r\geq 0,196$ .

При работе с малыми выборками используется следующая формула для расчета коэффициента корреляции:

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \cdot \sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (3.2)$$

где  $(x_i - \bar{x})$ ,  $(y_i - \bar{y})$  – отклонения значений индивидуальных вариантов  $x_i$ ;  $y_i$  от их средних значений  $\bar{x}$ ,  $\bar{y}$ .

Любой выборочный (эмпирический) коэффициент корреляции ( $r_{xy}=r$ ), являясь величиной случайной, может оказаться отличным от нуля даже при независимом варьировании признаков  $x$  и  $y$ . Отсюда возникает необходимость проверки надежности связи  $r_{xy}$ , тем более, что  $r_{xy}$  рассматривается в качестве оценки неизвестного (истинного) генерального параметра  $\rho_{xy}$ . При оценке достоверности вычисленного коэффициента корреляции ( $r_{\phi}$ ) необходимо с помощью таблицы коэффициентов корреляции сравнить его с табличным значением  $r_T$ , а также установить достоверность коэффициента корреляции через  $t$  – критерий Стьюдента.

Если  $r_{\phi} > r_T$ , то влияние фактора на признак достоверно; наоборот, если  $r_{\phi} < r_T$ , то коэффициент корреляции недостоверен и влияние фактора на признак несущественно.

При использовании  $t$  – критерия Стьюдента для доказательства достоверности  $r$ , вначале определяется стандартная (квадратическая) ошибка коэффициента корреляции по формуле:

$$\sigma_r = \sqrt{(1 - r^2)/(n - 2)}, \quad (3.3)$$

где  $n$  – число сопряженных пар в сравниваемых выборочных совокупностях.

Значение коэффициента корреляции записывается с учетом его ошибки:  $r \pm \sigma_r$ . Затем, проверяется нулевая гипотеза  $H_0$ , применительно к оценке генерального  $\rho_{xy}$  по величине эмпирического коэффициента корреляции ( $r_{xy}$ ), которая заключается в предположении, что между случайными величинами  $x$  и  $y$  корреляция отсутствует, т. е.  $H_0: \rho_{xy} = 0$ . Для проверки нулевой гипотезы используется  $t$  – критерий Стьюдента

$$t = |r| / \sigma_r. \quad (3.4)$$

Выборочная функция  $t$  при условии  $H_0$  удовлетворяет распределению Стьюдента с  $m = n - 2$  – степенями свободы. При уровне значимости  $\alpha$

и  $m$  – степенях свободы по таблице  $t$  – распределения Стьюдента можно найти значение предела  $t_{\alpha; m}$ . При  $t \geq t_{\alpha; m}$  гипотеза должна быть отвергнута. Это значит, что выборочный коэффициент корреляции  $r_{xy}$  существенно отличен от нуля. Тогда можно принять, что в генеральной совокупности  $\rho_{xy} \neq 0$ , т. е. случайные величины  $x$  и  $y$  не являются независимыми. При  $t < t_{\alpha; m}$  гипотеза  $H_0$  не отвергается, отклонение  $r_{xy}$  от нуля – чисто случайного характера,  $t_{\alpha; m}$  – двусторонний критерий. В случае одностороннего критерия, при уровне значимости  $\alpha$ , вычисленное значение  $t$  – критерия сравнивается с величиной  $t_{\alpha; m}$ .

Для проверки надежности  $r$ , вычисленных при достаточно больших  $n$  ( $n \geq 40$ ), используется соотношение В. И. Романовского:

$$|r| / \sigma_r > t_{\alpha; m}, \quad (3.5)$$

где  $\sigma_r = (1 - r^2) / \sqrt{n}$  рассматривается как ошибка вычисления  $r$ .

Этот критерий менее жесткий, чем предыдущий, но во многих случаях является достаточным.

В тех случаях, когда эмпирический коэффициент корреляции мал  $|r| \leq 0,5$ , возникает вопрос о том, не являются ли в действительности случайные величины некоррелированными, т. е. не объясняется ли наличие малого  $r$  случайными ошибками измерений (эксперимента). Фактически речь идет о проверке гипотезы  $H_0: \rho_{xy} = \rho_0 = 0$ .

Если для эмпирического значения  $r$  произведение  $|r| \cdot \sqrt{n-1}$  окажется больше некоторого критического значения (табл. 3.1), то с надежностью  $P$  можно утверждать, что  $r$  истинный  $\rho_0 \neq 0$ , т. е. гипотеза  $H_0$  отвергается.

По В. И. Романовскому, если  $|r| \cdot \sqrt{n-1} \geq 3$ , можно считать  $r$  значимым и связь реальной. Если  $|r| \cdot \sqrt{n-1} < 3$ , то  $r$  отличается от нуля с большой вероятностью лишь случайно.

Когда  $n$  достаточно велико и  $r$  не слишком близко к единице, распределение выборочных коэффициентов корреляции стремится к нормальному закону с центром распределения, равным  $\bar{r}$ , и средним квадратическим отклонением  $\sigma_r = (1 - r^2) / \sqrt{n-1}$ . При малом  $n$ , когда  $|r| \rightarrow 1$ , распределение выборочных коэффициентов корреляции все более укло-

няется от нормального. Коэффициент корреляции, рассчитанный по выборке конечного объема  $n$ , в среднем всегда меньше коэффициента корреляции генеральной совокупности [Пановский, Брайер, 1972].

Таблица 3.1

 Критические значения  $|r| \cdot \sqrt{n-1}$ 

$n$	$P=0,95$	$P=0,99$	$P=0,999$	$n$	$P=0,95$	$P=0,99$	$P=0,999$
10	1,89	2,29	2,62	25	1,94	2,47	3,03
11	1,90	2,32	2,68	30	1,94	2,49	3,07
12	1,91	2,35	2,73	35	1,95	2,50	3,10
13	1,91	2,37	2,77	40	1,95	2,51	3,13
14	1,92	2,39	2,81	50	1,95	2,527	3,160
15	1,92	2,40	2,84	60	1,953	2,536	3,184
16	1,93	2,41	2,87	70	1,954	2,541	3,198
17	1,93	2,42	2,90	80	1,955	2,546	3,209
18	1,93	2,43	2,92	90	1,956	2,550	3,219
19	1,93	2,44	2,94	100	1,956	2,553	3,226
20	1,94	2,45	2,96	$\infty$	1,960	2,576	3,291

При относительно малых объемах выборок ( $n < 50$ ) и, особенно при больших  $r$ , для оценки случайного рассеивания выборочных коэффициентов корреляции обычно используется преобразование Фишера, основанное, в свою очередь, на использовании специальной переменной  $Z_r$ , функционально связанной с  $r$  выражением:

$$Z_r = \frac{1}{2} \ln \frac{1+r}{1-r} = 1,1513 \cdot \lg \frac{1+r}{1-r} . \quad (3.6)$$

Величина  $Z_r$  распределена асимптотически нормально с дисперсией  $\sigma_{Z_r}$ :

$$\sigma_{Z_r} = 1/\sqrt{n-3} \quad (3.7)$$

и математическим ожиданием  $\bar{Z}_r$ ,

$$\bar{Z}_r = \frac{1}{2} \ln \frac{1+\rho_0}{1-\rho_0} + \frac{\rho_0}{2 \cdot (n-1)} . \quad (3.8)$$

Величина

$$U_r = (Z_r - \bar{Z}_r) / \sigma_{Z_r} \quad (3.9)$$

приближенно удовлетворяет нормированному нормальному закону распределения. В силу основного свойства величины  $Z_r$ , имеем:



$$P(-x \cdot \sigma_{Z_r} < Z_r - \bar{Z}_r < x \cdot \sigma_{Z_r}) = 2 \cdot \Phi(x) - 1, \quad (3.10)$$

где  $P$  – символ вероятности события, описываемого содержанием круглых скобок;  $\Phi(x)$  – функция нормированного нормального распределения:

$$\Phi(x) = \frac{1}{2 \cdot \sqrt{\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) \cdot dt. \quad (3.11)$$

Задавая определенное (достаточно большое) значение разности  $2 \cdot \Phi(x) - 1$  в формуле 3.10, по таблицам нормированного нормального распределения, найдем  $x$ , после чего можно найти границы доверительного интервала неизвестного  $\rho_0$ .

Нередко возникает необходимость сравнения двух коэффициентов корреляции  $r_1$  и  $r_2$ , полученных по выборкам  $n_1$  и  $n_2$  – пар наблюдений одних и тех же признаков  $x$  и  $y$ . Это достигается путем проверки нулевой гипотезы  $H_0: \rho_1 = \rho_2$ , которая гласит, что выборки взяты из двух совокупностей с одинаковыми коэффициентами корреляции. Для проверки нулевой гипотезы может быть использовано преобразование Фишера. Для оценки существенности расхождения между выборочными коэффициентами корреляции  $r_1$  и  $r_2$ , вычисляется отношение:

$$Z = |Z_1 - Z_2| / \sqrt{\sigma_{Z_1}^2 + \sigma_{Z_2}^2}, \quad (3.12)$$

где  $Z_i$  и  $\sigma_{Z_i}^2$  – определяются по формулам (3.6) и (3.7).

Выбрав по табл. 3.2 пороговое значение  $Z_\alpha$ , соответствующее уровню значимости  $\alpha$ , сравним его с величиной  $Z$ , найденной по формуле (3.12). Если  $|Z| < Z_\alpha$ , делается вывод, что гипотеза  $H_0$  не отвергается и, следовательно, расхождение между  $r_1$  и  $r_2$  можно считать случайным. Это также значит, что в  $(100 - \alpha)\%$  – случаях гипотеза  $H_0$  будет отвергнута.

В табл. 3.2, наряду с  $\alpha$ , показаны также значения  $P = 1 - \alpha$  – статистической достоверности гипотезы  $H_0$ . Разность  $Z_1 - Z_2$  имеет также нормальное распределение с дисперсией:

$$\sigma_{Z_1 - Z_2}^2 = \sigma_{Z_1}^2 + \sigma_{Z_2}^2 \quad (3.13)$$

и средней, равной нулю. Поэтому, вероятность  $P$  определяется как:



$$P(|Z_1 - Z_2|) \geq Z \cdot \sigma_Z = 2 - 2 \cdot \Phi(Z), \quad (3.14)$$

где  $Z$  – вычисляется по формуле (3.12);  $\Phi(Z) = \Phi(x)$  – функция нормированного распределения – по (3.14).

Таблица 3.2

**Значения  $Z\alpha$  для различных  $\alpha$  при двустороннем ограничении**

$\alpha$	0,001	0,027	0,010	0,0455	0,050
$P=1-\alpha$	0,999	0,9973	0,990	0,9545	0,950
$Z\alpha$	3,291	3,000	2,576	2,000	1,960

При значительной вероятности, расхождение между  $Z_1$  и  $Z_2$ , определяемое различием  $r_1$  и  $r_2$ , имеет случайный характер, а  $r_1$  и  $r_2$  несущественно отличаются друг от друга.

Важной особенностью применения  $Z$ -критерия Фишера является то, что в основу расчета коэффициентов корреляции  $r_1$  и  $r_2$  должна быть положена одна и та же группировка данных.

Например, имеется ряд наблюдений за годовым расходом воды в р. Птичь – с. Лучицы; необходимо подобрать реки с синхронными колебаниями водности. Исходя из комплексного анализа физико-географических, гидрографических характеристик, отбираем реки – кандидаты: Ясельда – с. Сенин, Оресса – с. Верхутино, Оресса – с. Андреевка. Гидрометрические данные по реке Птичь – с. Лучицы и по рекам – кандидатам приведены в табл. 3.3.

Таким образом, предпочтение отдается р. Оресса – с. Андреевка, так как для р. Птичь – с. Лучицы она является лучшим аналогом,  $r_{(П-Л)-(О-А)} = 0,95$  и колебания их водности наиболее синхронные (рис. 3.1). Коэффициент детерминации  $d_{(П-Л)-(О-А)} = 0,90$ , т. е. 90% всех колебаний этих рек взаимообусловлены. При этом стандартная ошибка коэффициента корреляции  $r_{(П-Л)-(О-А)}$ , рассчитанная по формуле (3.2), составляет:  $\sigma_r = 0,06$ , а значение коэффициента корреляции запишется как  $r = 0,95 \pm 0,06$ , т. е. он изменяется от 0,89 до 1,0.

Проверим нулевую гипотезу, т. е.  $H_0: \rho = 0$ , при  $H_1: \rho > 0$ , используя  $t$ -критерий Стьюдента (3.3)  $t = |0,95| / 0,06 = 15,83$ .

Таблица 3.3

Расходы воды рек, м<sup>3</sup>/с

№ п/п	Птичь – с. Лучицы	Ясельда – с. Сенин	Оресса – с. Верхутино	Оресса – с. Андреевка
1	46,6	11,5	2,52	12,8
2	48,8	17,6	2,43	16,3
3	33,1	19,0	1,68	10,4
4	27,3	14,6	2,07	9,85
5	49,3	21,4	3,06	13,8
6	24,1	9,91	1,64	9,14
7	41,3	17,4	2,38	13,4
8	17,6	4,0	0,64	6,52
9	40,8	14,2	2,22	16,3
10	54,5	14,8	3,06	20,0
11	42,6	18,3	2,61	16,2
12	90,8	38,1	2,58	36,0
13	41,9	12,4	1,97	15,2
14	36,2	14,9	1,49	13,5
15	55,2	18,5	3,76	20,4
16	38,2	12,2	2,54	14,1
17	35,7	9,74	2,37	13,4
18	49,1	13,5	3,14	18,8
19	48,7	20,1	3,30	19,7
20	49,3	22,6	4,15	20,3
21	50,5	23,2	3,84	20,0
22	42,1	15,2	2,93	16,9
23	73,2	37,6	4,39	26,5
24	57,4	29,6	3,40	20,9
25	32,3	12,0	2,51	13,7
26	36,9	15,7	3,02	15,7
27	44,4	31,7	3,43	19,2
28	60,8	29,6	3,57	23,1
$x = \frac{\sum_{i=1}^{28} (x_i)}{28}$	45,31	18,55	2,74	16,86
$\sum_{i=1}^{28} (x_i - x)^2$	5737,61	1854,75	19,52	917,15

Теоретическое значение  $t$ -критерия Стьюдента (при  $m=28-2=26$  – степенях свободы и  $\alpha=0,05$ ) будет  $-t^T_{(26;0,05)}=2,05$ . Так как  $15,83 > 2,05$  отвергаем нулевую гипотезу и принимаем альтернативную – коэффициент корреляции отличен от нуля. Для проверки нулевой гипотезы можно воспользоваться таблицей проверки коэффициента корреляции на значимость относительно нуля. В нашем случае, критическое значение коэффициента корреляции ( $r_{кр}=0,374$ ), что значительно меньше полученного. Следовательно, нулевая гипотеза отвергается. Выполним еще одну проверку нулевой гипотезы: определим величину  $-|r| \cdot \sqrt{n-1} = 0,95 \cdot \sqrt{28-1} = 4,9$ ; сопоставим эту величину с данными табл. 3.1.  $4,9 > 1,94$  (при  $P=0,95$ ),  $4,9 > 2,48$  (при  $P=0,99$ ) и  $4,9 > 3,03$  (при  $P=0,999$ ), убедившись, что нулевая гипотеза отвергается. Как было сказано выше, для малых выборок при значениях  $|r| \rightarrow 1$ , распределение выборочных коэффициентов корреляции заведомо отличается от нормального. Поэтому  $t$ -критерий Стьюдента становится ненадежным. Чтобы обойти это затруднение, воспользуемся преобразованием Фишера. В данном случае, согласно (3.6), (3.7), (3.8), соответственно, имеем: переменную –  $Z_r = 1,83$ ; дисперсию –  $\sigma_{Z_r} = 0,19$ ; математическое ожидание –  $\bar{Z}_r = 1,85$ ; величину –  $U_r = 0,11$ .

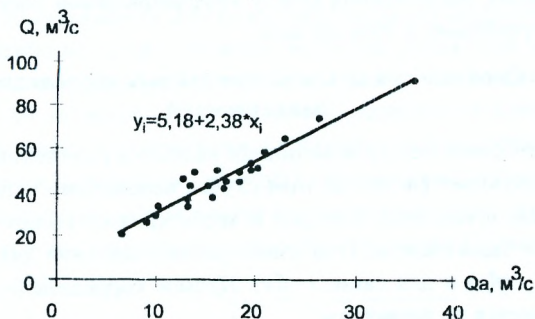


Рис. 3.1. Связь годового стока рек Птичь – с. Лучицы и Оресса – с. Андреевка.

Находим интервал изменения  $r$ , соответственно интервалу изменения  $Z$

$$Z_r \pm \sigma_{Z_r} = 1,83 \pm 0,19 = \left\{ \frac{2,02 - \text{верхняя граница}}{1,64 - \text{нижняя граница}} \right\}.$$

По верхней и нижней границам  $Z_r$  находим границы изменения « $r$ » (по специальным таблицам), откуда  $r=0,965$  – верхняя граница и  $r=0,928$  – нижняя граница, т. е.  $r = \left\{ \begin{array}{l} 0,965 \\ 0,928 \end{array} \right\}$ .

Сравним различимы ли между собой  $r_{(п-л)-(о-а)}=0,95$  и  $r_{(п-л)-(я-с)}=0,82$ . Проверим нулевую гипотезу  $H_0: \rho_1 = \rho_2$ , которая гласит, что выборки должны быть взяты из двух совокупностей с одинаковыми коэффициентами корреляции.

Для этого, используя формулы (3.6) и (3.7), вычислим следующие параметры:

$$Z_{r_{(п-л)-(о-а)}} = 1,83; \quad \sigma_{Z_{(п-л)-(о-а)}} = 0,19; \quad Z_{r_{(п-л)-(я-с)}} = 1,16;$$

$$\sigma_{Z_{(п-л)-(я-с)}} = 0,19.$$

Воспользовавшись формулой (3.12), имеем  $Z_{\alpha} = 2,49$ .

Выберем по табл. 3.2 пороговое значение  $Z_{\alpha}=1,96$ , при уровне значимости  $\alpha = 0,05$ . Так как  $2,49 > 1,96$ , – нулевую гипотезу отвергаем и, следовательно, расхождения между коэффициентами корреляции нельзя считать случайными в 95% случаев.

### 3.2. Особенности анализа нелинейных корреляционных зависимостей

Часто связи между величинами являются нелинейными, что существенно усложняет расчет. В этом случае необходимо использовать корреляционное отношение, которое и характеризует тесноту связи криволинейности зависимости. Оно имеет разные значения для корреляционной связи  $y=f(x)$  и для связи  $x=f(y)$ . Прямое корреляционное отношение  $y/x$  вычисляется по формуле:

$$\eta_{y/x} = \sqrt{k \frac{\sum_{j=1}^m \left( \frac{y_j}{x_j} - \bar{y} \right)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}}, \quad (3.15)$$

где  $y_j/x_j$  – точка эмпирической линии регрессии;  $k$  – число вариантов в частной группе;  $y_j/x_j - \bar{y}$  – отклонение точек эмпирической линии регрессии от

общего среднего по  $\bar{y}$ . Обратное корреляционное отношение  $x/y$  для малых выборок вычисляется по формуле:

$$\eta_{x/y} = k \cdot \frac{\sum_{j=1}^m \left( \frac{x_j}{y_j} - \bar{x} \right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad (3.16)$$

где  $x_j/y_j - \bar{x}$  – отклонение точек эмпирической регрессии от общей средней по  $\bar{x}$ ;  $x_i - \bar{x}$  – отклонение индивидуальных вариантов в выборке от общей средней по  $\bar{x}$ .

Для оценки достоверности полученных значений  $\eta$ , можно использовать  $t$ -критерий

$$t = \eta \cdot \sqrt{(n-2)/(1-\eta^2)}. \quad (3.17)$$

Нулевая гипотеза об отсутствии связи между признаками отвергается, если  $t \geq t_{\alpha, m}$ , где  $\alpha$  – уровень значимости;  $m = n - 2$  – число степеней свободы. Параметр  $t_{\alpha, m}$  определяется по таблицам  $t$ -распределения Стьюдента. Коэффициент корреляции характеризует только линейную связь, а корреляционное отношение – любую форму связи. При строго линейной связи  $\eta_{y/x} = \eta_{x/y} = |r|$ . При наличии нелинейной связи  $\eta_{y/x} \neq \eta_{x/y}$  и  $\eta \neq |r|$ . Следовательно, по разности между этими показателями можно судить о форме корреляционной зависимости между варьирующимися признаками.

Для решения вопроса – является ли исследуемая зависимость линейной или криволинейной – используется критерий криволинейности. Наиболее простой, но менее строгий, способ заключается в определении разности коэффициентов корреляции и корреляционного отношения; при этом, используется неравенство  $\eta^2 - r^2 \geq 0,1$ . Корреляция считается криволинейной, если полученный результат соответствует этому неравенству. Второй способ оценки степени криволинейности связан с применением  $t$ -критерия Стьюдента

$$t = \gamma \cdot \sqrt{\frac{n-2}{\gamma - \gamma^2 \cdot (2 - \eta^2 - r^2)}}, \quad (3.18)$$

где  $\gamma = \eta^2 - r^2$ .

Если при уровне значимости  $\alpha$  и  $m = n - 2$  – степенях свободы  $t < t_{\alpha, m}$ , то корреляция между признаками с  $P$ - $\alpha$  – надежностью оценивается как

прямолинейная. При  $t \geq t_{\alpha; m}$ , зависимость между признаками следует считать заметно отличающейся от прямолинейной.

Можно также использовать  $F$ -критерий (Фишера) для определения степени приближения криволинейной зависимости к прямолинейной:

$$F = \frac{(\eta^2 - r^2) \cdot (n - k_x)}{(1 - \eta^2) \cdot (k_x - 2)}, \quad (3.19)$$

где  $n$  – объем выборки;  $k_x$  – число групп по ряду  $x$ .

Связь можно практически принять за линейную, если  $F < F_T$  и определять показатели для прямолинейной корреляции и регрессии. Корреляция нелинейна, если  $F \geq F_T$ . Теоретические значения  $F_T$  берутся из специальных таблиц, для  $\nu_1 = k_x - 2$  и  $\nu_2 = n - 2$  – степеней свободы [Верещагин, Наумов и др., 1990].

Вычислим корреляционное отношение  $\eta_{y/x}$  и  $\eta_{x/y}$ , используя данные о средних многолетних годовых расходах воды малых рек одного из регионов  $Y$  ( $\text{м}^3/\text{с}$ ) и площадях водосборов ( $\text{км}^2$ ) (табл. 3.4).

Таблица 3.4

#### Значения расходов малых рек региона и площадей их водосборов

$Y$ – расход ( $\text{м}^3/\text{с}$ )	0,65	1,15	1,25	1,75	2,25	2,10	2,55	2,50	3,05
$X$ – площадь водосбора ( $\text{км}^2$ )	50	25	75	50	100	150	150	200	250
$Y$ – расход ( $\text{м}^3/\text{с}$ )	2,75	3,15	3,60	3,50	4,00	3,75	3,85	4,00	4,35
$X$ – площадь водосбора ( $\text{км}^2$ )	300	400	450	550	650	700	800	900	1000

Из графика видно, что кривая  $Q=f(F)$  близка к логарифмической, зависимость положительная (рис. 3.2).

На основании данных табл. 3.4, рассчитываем корреляционное отношение  $\eta_{y/x}$  между  $x$  и  $y$ . Выборки разделим на шесть частных групп, однако, дальнейшие операции проводим с  $y$ . Рассчитываем общее среднее  $\bar{y}$ . Точки эмпирической линии регрессии  $y/x$  представляют собой среднее арифметическое частных групп. Записываем отклонения точек эмпирической линии регрессии от общего среднего по  $y$ , возводим эти отклонения в квадрат и суммируем. Далее, вычисляем отклонения индивидуальных показателей  $y_i$  от общего среднего  $\bar{y}$  и суммируем (сумма



должна быть равна нулю или близка к нему). Каждое отклонение возводим в квадрат и суммируем. Результаты расчетов сводим в табл. 3.5.

Таблица 3.5

**Результаты вычисления прямого корреляционного отношения для невзвешенных рядов**

$Y_i$ (расход, м <sup>3</sup> /с)	$\Sigma y_i$ (по группам)	$y/x$	$y/x - \bar{y}$	$(y/x - \bar{y})^2$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
1 группа						
0,65					-2,14	4,5796
1,15	3,05	1,02	-1,77	3,1447	-1,64	2,6896
1,25					-1,54	2,3716
2 группа						
1,75					-1,04	1,0816
2,10	6,10	2,03	-0,76	0,5725	-0,69	0,4761
2,25					-0,54	0,2916
3 группа						
2,50					-0,29	0,0841
2,55	8,10	2,70	-0,09	0,0081	-0,24	0,0576
3,05					0,26	0,0676
4 группа						
2,75					-0,04	0,0016
3,15	9,50	3,17	0,38	0,1419	0,36	0,1229
3,60					0,81	0,6561
5 группа						
3,50					0,71	0,5041
4,00	11,25	3,75	0,96	0,9216	1,21	1,4641
3,75					0,96	0,9216
6 группа						
3,85					1,06	1,1236
4,00	12,20	4,07	1,28	1,6299	1,21	1,4641
4,35					1,56	2,4336
$\sum_{i=1}^{18} y_i = 50,2$	$\bar{y} = 2,79$		$\Sigma 0$	$\Sigma 6,4187$	$\Sigma -0,02 \approx 0$	$\Sigma 20,3978$

Полученные данные используем для расчета прямого корреляционного отношения по формуле (3.15):  $\eta_{y/x} = 0,972$ . Достоверность результатов определим по  $t$ -критерию Стьюдента, учитывая зависимость (3.17):  $t_{\eta_{y/x}} = 14,33$ . Поскольку ( $t_{\Phi} = 14,33 > t_T = 1,746$ ), при  $P=0,95$ , для

$v=16$ , значение корреляционного отношения следует признать доказанным; зависимость между расходами воды рек  $Y=Q$  и площадями водосборов  $X=F$  положительна и достоверна. Аналогично вычисляем обратное корреляционное отношение ( $\eta_{x/y}$ ) (табл. 3.6).

Таблица 3.6

**Результаты вычисления обратного корреляционного отношения для невзвешенных рядов**

$x_i$ (площадь, км <sup>2</sup> )	$\Sigma x_i$ (по группам)	$x/y$	$x/y - \bar{x}$	$(x/y - \bar{x})^2$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1 группа						
50					-328	107584
25	150	50	-328	107584	-353	124609
75					-303	91809
2 группа						
50					-328	107584
100	300	100	-278	77284	-278	77284
150					-228	51984
3 группа						
150					-228	51984
200	600	200	-178	31684	-178	31684
250					-128	16384
4 группа						
300					-78	6084
400	1150	383	5	25	22	484
450					72	5184
5 группа						
550					172	29584
650	1900	633	255	65025	272	73984
700					322	103684
6 группа						
800					422	178084
900	2700	900	522	272484	522	272484
1000					622	386884
$\sum_{i=1}^{18} x_i = 6800$ $\bar{x} = 378$			$\Sigma -2 \approx 0$	$\Sigma 554086$	$\Sigma -4 \approx 0$	$\Sigma 1717362$

Получим корреляционное отношение:  $\eta_{x/y} = 0,980$ , а также

$t$ -критерий Стьюдента:  $t_{\eta_{y/x}} = 20,10$ . Так как  $t_{\sigma} = 20,10 > t_T = 1,746$ , при

$P=0,95$  для  $\nu=16$ , то значение корреляционного отношения следует признать доказанным.

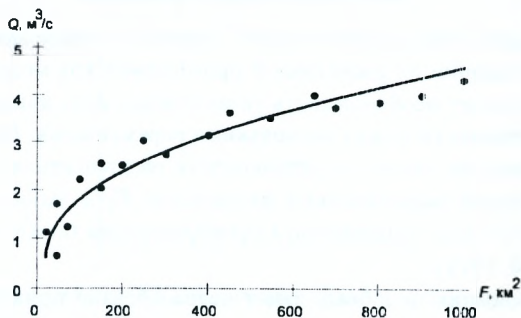


Рис. 3.2. Зависимость расходов воды малых рек региона от их площадей водосборов.

Оценим линейность связи между расходами воды рек и площадями водосборов (на основе данных табл. 3.4)  $r=0,90$ ;  $\eta_{y/x}=0,972$ ;  $\gamma=0,135$ ;  $t_{y/x}=1,495$ .

По  $t$ -распределению Стьюдента для  $\alpha=0,10$  и  $m=16$  находим  $t_{0,05,16}=1,337$ . Полученное значение  $t$  – критерия превышает  $(t_{\alpha, m})$ , при  $\alpha=0,10$  и  $m=16$ , и это позволяет заключить, что зависимость между расходами воды рек и площадями водосборов отличается от прямолинейной.

Оценим также линейность связи между площадями водосборов и расходами воды рек  $\gamma=0,154$ ;  $t_{\eta_{x/y}}=1,579$ . Полученное значение также превышает теоретическое значение  $t$  – критерия Стьюдента, при  $\alpha=0,10$  и  $m=16$ , и зависимость между этими признаками также отличается от линейной.

Проверим степень приближения криволинейной зависимости к прямолинейной, используя критерий Фишера (3.20):  $F=9,76$ . Используя таблицы  $F$ -распределения Фишера, на 5% уровне значимости для  $\nu_1=6-2=4$  и  $\nu_2=18-2=16$  – степеней получаем  $F_{(4,16,5\%)}^T=3,01$  и так как это значение меньше  $F_\phi$  – корреляция может быть признана нелинейной.

### 3.3. Особенности выявления корреляционных связей многофакторных процессов

При изучении многофакторных природных процессов и, в частности, при построении их расчетных и прогностических моделей достаточно часто возникает необходимость установления вида линейной корреляционной зависимости между несколькими переменными. Для примера, на рис. 3.3 приведена схема упорядоченности расположения переменных в анализе функций экологической надежности  $F_2$  и эколого-социальных последствий  $F_1$  при ландшафтном преобразовании больших территорий [Шведовский, 1998].

Для решения подобных задач целесообразно привлекать аппарат множественной корреляции. Сущность этого подхода состоит в распределении основных положений метода линейной корреляции двух переменных  $y$  на случай зависимости интересующей нас переменной, от произвольного числа аргументов  $x$ .

При множественном корреляционном анализе вычисляются два типа парных коэффициентов корреляции:

- $r_{yxj}$  – коэффициенты, определяющие тесноту связи между функцией отклика и одним из факторов  $x_j$ ;
- $r_{x_j x_m}$  – коэффициенты, показывающие тесноту связи между одним из факторов и фактором  $x_m$  ( $j, m = \overline{1, k}$ ), где  $k$  – число факторов.

Для вычисления коэффициентов корреляции используется формула (3.2). Значимость парных коэффициентов корреляции проверяется по стандартной методике. Если один из коэффициентов  $r_{x_j x_m}$  окажется равным 1, то факторы  $x_j$  и  $x_m$  функционально (невероятно) связаны между собой и тогда целесообразно один из них исключить из рассмотрения, причем, оставив тот фактор, у которого коэффициент  $r_{y, x_j}$  – больше.

После вычисления всех парных коэффициентов корреляции и исключения из рассмотрения того или иного фактора можно построить матрицу коэффициентов корреляции вида:

$$\begin{pmatrix}
 1 & r_{yx_1} & r_{yx_2} & \dots & r_{yx_j} & \dots & r_{yx_k} \\
 r_{x_1y} & 1 & r_{x_1x_2} & \dots & r_{x_1x_j} & \dots & r_{x_1x_k} \\
 r_{x_2y} & r_{x_2x_1} & 1 & \dots & r_{x_2x_j} & \dots & r_{x_2x_k} \\
 \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
 r_{x_jy} & r_{x_jx_1} & r_{x_jx_2} & \dots & 1 & \dots & r_{x_jx_k} \\
 \dots & \dots & \dots & \dots & \dots & \dots & \dots \\
 r_{x_ky} & r_{x_kx_1} & r_{x_kx_2} & \dots & r_{x_kx_j} & \dots & 1
 \end{pmatrix}. \quad (3.20)$$

Используя матрицу (3.20), можно вычислить частные коэффициенты корреляции, которые показывают степень влияния одного из факторов  $x_j$  на функцию отклика  $y$  при условии, что остальные факторы имеют постоянный уровень (закреплены на нем). Формула для вычисления частных коэффициентов корреляции следующая:

$$r_{y^*x_1, x_2, x_3, \dots, x_j, \dots, x_k} = \frac{D_{1j}}{\sqrt{D_{11} \cdot D_{jj}}}, \quad (3.21)$$

где  $D_{1j}$  – определитель матрицы, образованной из матрицы (3.20) вычеркиванием 1-й строки  $j$ -го столбца. Определители  $D_{11}$  и  $D_{jj}$  вычисляются аналогично, как и парные коэффициенты; частные коэффициенты корреляции изменяются от  $-1$  до  $+1$ .

Значимость и доверительный интервал для коэффициентов частной корреляции определяются как для коэффициентов парной корреляции, только число степеней свободы вычисляется по формуле:

$$\nu = n - k' - 2, \quad (3.22)$$

где  $k' = k - 1$  – порядок частного коэффициента парной корреляции.

Для изучения тесноты связи между функцией отклика  $y$  и несколькими факторами  $x_1, x_2, x_3, \dots, x_j, \dots, x_k$  используется коэффициент множественной корреляции  $R$ . Коэффициент множественной корреляции служит для оценки качества предсказания;  $R$  – всегда положителен и изменяется от 0 до 1. Чем больше  $R$ , тем лучше качество предсказаний данной моделью опытных данных.

Для вычисления коэффициента множественной корреляции используется матрица (3.20):

$$R = \sqrt{1 - D/D_{11}}, \quad (3.23)$$

где  $D$  – определитель матрицы (3.20);  $D_{11}$  – минор определителя, получаемый из матрицы (3.20) путем исключения первой строки и первого столбца матрицы.

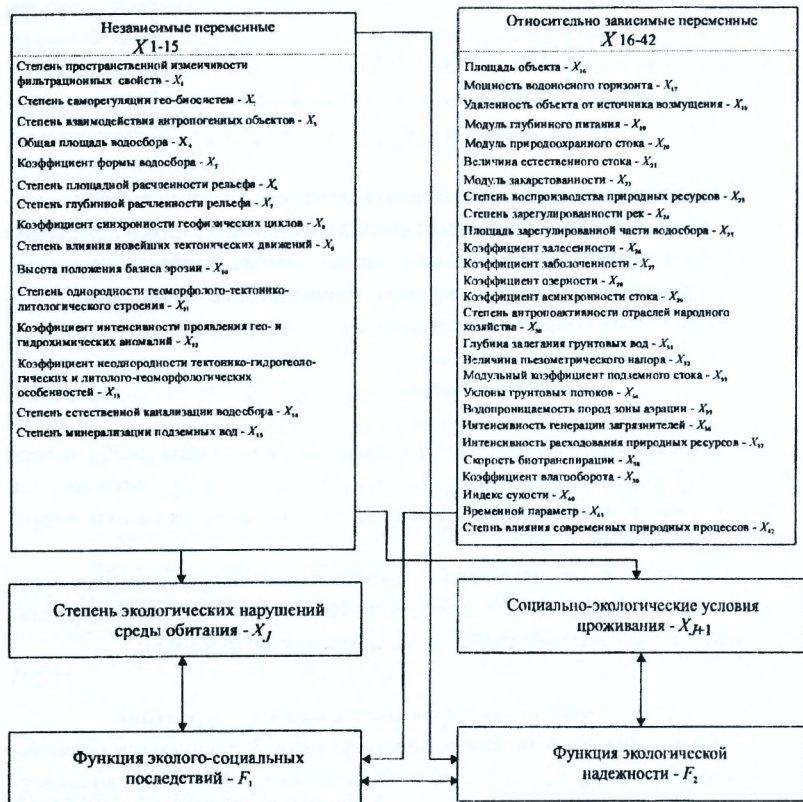


Рис. 3.3. Схема упорядоченности расположения переменных в анализе функций  $F_1$  и  $F_2$  при ландшафтном преобразовании территорий

Коэффициент множественной корреляции можно найти также по формуле:



$$R = \sqrt{1 - \bar{S}_{ост}^2 / \bar{S}_y^2}, \quad (3.24)$$

где  $\bar{S}_{ост}^2$  – остаточная дисперсия;  $\bar{S}_y^2$  – общая дисперсия. Или

$$R = \sqrt{1 - \frac{n-1}{n-k-1} (1-R')^2}, \quad (3.25)$$

$$\text{где } R' = \sqrt{1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}}; \quad (3.26)$$

$\hat{y}_i$  – вычисленное значение.

Значимость коэффициента множественной корреляции проверяется по  $t$  – критерию Стьюдента

$$t_R = R / \bar{S}_R \geq t_{(n-k-1)}^T, \quad (3.27)$$

где  $\bar{S}_R$  – среднеквадратическая погрешность коэффициента множественной корреляции, определяемая как:

$$\bar{S}_R = (1 - R^2) / \sqrt{n - k - 1}. \quad (3.28)$$

Значимость можно проверить также и по  $F$ -критерию:

$$F_R = \frac{R^2 (n - k - 1)}{(1 - R^2) \cdot (k - 1)}. \quad (3.29)$$

Полученное значение  $F_R$  сравнивается с табличным  $F_m$  при выбранном уровне значимости и числах степеней свободы –  $\nu_1 = n - k - 1$  и  $\nu_2 = k - 1$ . Нулевая гипотеза о равенстве нулю множественного коэффициента корреляции, в совокупности,  $H_0: R = 0$  принимается, если  $F_R < F_m$ , и отвергается – если  $F_R \geq F_m$ .

Величина  $R^2$  называется *множественным коэффициентом детерминации*. Она показывает, какая часть дисперсии функции отклика объясняется вариацией линейной комбинации выбранных факторов  $x_1, x_2, x_3, \dots, x_j, \dots, x_n$ .

Рассмотрим пример вычисления коэффициента множественной корреляции. При анализе синхронности колебания водности рек (табл. 3.3) получены следующие парные коэффициенты корреляции, которые сведены в табл. 3.7.

Таблица 3.7

## Матрица коэффициентов парной корреляции

$j \setminus m$	Птичь – с. Лучицы ( $y$ )	Ясельда – с. Сенин ( $x_1$ )	Оресса – с. Вер- хутино ( $x_2$ )	Оресса – с. Анд- реевка ( $x_3$ )
$y$	1	0,82	0,65	0,95
$x_1$	0,82	1	0,65	0,83
$x_2$	0,65	0,65	1	0,66
$x_3$	0,95	0,83	0,66	1

Так критическое значение коэффициентов парной корреляции, при уровне значимости  $P=0,95$  и  $\nu=26$ , будет  $r_{kp}=0,374$ , таким образом, все значения коэффициентов в табл. 3.7 значимы.

Необходимо выявить степень синхронности колебания годового стока рр. Птичь – с. Лучицы, Ясельда – с. Сенин, Оресса – с. Верхутино и Оресса – с. Андреевка, т. е. рассчитать коэффициент множественной корреляции. Для этого вычислим определитель  $D$  и его минор  $D_{11}$ . Имеем:  $D = 0,010$ ;  $D_{11} = 0,165$  и тогда  $R = 0,969$ .

Значимость коэффициента  $R$ , проверяем по формулам (3.27) и (3.28):  $\bar{S}_R = 0,012$ ,  $t_R = 80,8$ .

Теоретическое значение  $t$  – критерия Стьюдента определяется как  $t_{(23,5\%)}^T = 1,711$ , что значительно меньше  $t_R = 80,8$ . Вычислим критерий Фишера:  $F_R = 184,6$ .

Табличное значение  $F^m$ , при  $\nu_1=3$  и  $\nu_2=28-4=24$  – степенях свободы, будет  $F_{5\%}^T = 3,01$ , что меньше  $F_R = 184,6$ .

Таким образом, взаимосвязь между колебаниями водности рассматриваемых рек  $R=0,969$  значима на 5%-ом уровне:  $F_{5\%}^T < F_R$ . Судя по коэффициенту множественной детерминации  $R^2=0,969^2=0,939$ , вариация водности р. Птичь – с. Лучицы на 93,9 % связана с ее колебаниями других рек и только 6,1% вариации  $1-R^2$  вызваны индивидуальными особенностями самой р. Птичь – с. Лучицы.

Рассмотрим более сложную задачу использования корреляционных связей – восстановление и продление рядов. Следует отметить, что сегодня зачастую с этой целью балансовые методы, в основу которых

положено уравнение водного или водохозяйственного баланса, то есть прямое определение происшедших изменений в стоке в виде алгебраической суммы различных антропогенных воздействий (измеренных или оцененных): заборов или сбросов воды, регулирование стока водохранилищами, потерь на испарение, изменения поверхностного и подземного стока и т. д.

Наиболее полно это может быть сделано путем решения обратной задачи водохозяйственного баланса участков и бассейнов рек. Для этого в уравнении баланса естественные характеристики стока принимаются за искомую величину, определяемую расчетом при наличии независимо определенных данных по всем остальным элементам баланса.

К настоящему времени такой способ практически реализован при составлении отчетных водохозяйственных балансов по фактическим данным в процессе ведения государственного водного кадастра по разделу «использование вод». Способ применяется для восстановления естественного стока и контроля правильности данных при учете заборов и сбросов воды (безвозвратного потребления воды по отношению к речному стоку) и регулирования стока.

В условиях, когда в водохозяйственном балансе существенное значение имеет процесс руслового регулирования стока (колебания русловых запасов воды при изменениях уровня), что характерно, в первую очередь, для рек с широкими поймами и при восстановлении стока за периоды паводков (с большой интенсивностью колебаний стока), этим явлением пренебрегать нельзя, не рискуя потерей точности расчетов.

Прямое определение происшедших изменений в стоке может быть выполнено и в буквальном смысле, путем отыскания алгебраической суммы поправок к измеренному стоку, учитывающих известные (также измеренные или рассчитанные) количественные изменения стока под влиянием различных антропогенных воздействий, определенных автономно, т. е. в виде

$$\Delta Q_{\Sigma} = \sum_{S=1}^S \Delta Q_S, \quad (3.30)$$

где  $\Delta Q_S$  – происшедшее изменение стока под влиянием  $S$ -го воздействия.

В обоих случаях должны учитываться возможные зависимости в изменениях стока от сезона года, водности реки, влагозапасов водосбора,

циклов вегетации растений и т. д. Это относится, прежде всего, к изменениям, обусловленным преобразованием водосборов (осушение земель, агромелиоративные мероприятия, распашка новых территорий и др.). Несмотря на длительность и большое число работ по каждому из видов воздействий, количественные методы расчета, обусловленные ими, изменений, которые широко применялись бы на практике, до сих пор нет. Поэтому расчеты по восстановлению стока, равно как расчеты ресурсов речного стока проводятся без оценки указанных последствий, что является для ряда случаев существенным недостатком.

Отсюда зачастую целесообразно восстановление рядов осуществлять с помощью построения корреляционных связей между естественным стоком в данном створе и некоторыми выбранными индикаторами. В их число могут входить сток рек-аналогов, стокоформирующие факторы (осадки, снего- и влагозапасы водосбора), другие характеристики. Связи строятся за период с ненарушенным стоком в данном створе и затем распространяются на искомый период. Разница между измеренным и полученным по корреляционной связи стоком интегрально характеризует наиболее вероятное значение поправки  $\Delta \bar{Q}_\Sigma$ . Сами связи находятся путем последовательного испытания их аналитического вида (линейные, нелинейные), числа и состава индикаторов и отбора наилучшей связи по критерию минимума остаточной дисперсии. Последняя величина может использоваться для моделирования добавки к величине  $\Delta \bar{Q}_\Sigma$ , учитывающей стохастический характер связей по выражению:

$$\Delta Q_\Sigma = \Delta Q_\Sigma + \zeta(\Delta Q_\Sigma), \quad (3.31)$$

где  $\zeta(\Delta Q_\Sigma)$  – знакпеременная случайная поправка (распределенная по нормальному закону) с нулевым математическим ожиданием и среднеквадратическим отклонением, численно равным остаточной дисперсии корреляционной связи.

Не менее эффективно использование и комплексного подхода, т. е. восстановление стока балансовым методом с проверкой полученных данных по корреляционным связям, на статическую однородность и на соответствие показателей внутригодового распределения стока в естественных и нарушенных условиях. Этот метод реализован в программном комплексе «Гидролог» [Волчек, 1998]. Программа предназначена для

восстановления отсутствующих данных наблюдений за элементами гидрологического режима рек с использованием метода множественной линейной регрессии; в качестве аргументов привлекаются любые гидрологические характеристики. Кроме получения поточечно восстановленных данных, рассчитываются все возможные уравнения линейной регрессии, коэффициент множественной корреляции и его среднеквадратическая ошибка, а также даются оценки стандартных статистических параметров (оценки среднего значения стока, среднего квадратического отклонения, коэффициентов вариации, асимметрии и автокорреляции) как по существующим данным наблюдений, так и по данным восстановленных рядов.

Уравнение, по которому восстанавливаются ряды, представляется в виде:

$$M_{i0} = \sum_{j=1}^K A_j \cdot M_{ij} + A_{K+1}, \quad (3.32)$$

где  $M_{ij}$  – гидрологическая характеристика  $j$  – аналога за  $i$  – интервал времени;  $K$  – число эффективных аналогов.

Основные требования для выбора аналогов можно свести к следующему:

- аналоги должны располагаться в сходных с приводимым пунктом физико-географических и климатических условиях;
- аналоги должны располагаться вокруг приводимого пункта;
- число аналогов должно быть не менее 2 – 4;
- коэффициенты корреляции между приводимым рядом и рядами аналогов должны быть выше коэффициентов корреляции между рядами аналогов.

На предварительном этапе отбора к анализу привлекается до 10 аналогов и оценивается эффективность уравнений регрессии, образованных комбинациями этих аналогов.

Наиболее полно вопрос выбора эффективных аналогов рассмотрен в работе Г. А. Алексеева [Алексеев, 1971]. Выбор осуществляется по двум критериям:

- доли вклада в дисперсию приводимого ряда;
- точности определения коэффициентов уравнения регрессии.

Выбор одного пункта-аналога осуществляется по вариантам:

– выбирается пункт наблюдений, имеющий более продолжительный период, но меньший коэффициент корреляции со стоком приводимого ряда;

– выбирается пункт с менее продолжительными наблюдениями, но с более тесной связью между величинами стока рассматриваемых рядов.

Если оба варианта удовлетворяют условиям (3.34), то сначала выбираются ряды с наибольшими коэффициентами корреляции, а затем они дополняются рядами с меньшими коэффициентами корреляции, но с большим периодом наблюдений. В общей сложности для каждого восстанавливаемого ряда возможно привлечение до девяти предполагаемых аналогов.

Для решения данной задачи в основу положен так называемый «способ просеивания», методика которого подробно изложена в работе Е. П. Борисенкова и М. А. Романова [Борисенков, Романов, 1969]. Также использованы алгоритмы решения аналогичных задач, приведенных в работах [Исследование..., 1981; Рождественский, Чеботарев, 1975], по которым разработан следующий алгоритм.

1. Формирование матрицы исходных данных

$$\|M_{ij}\| \quad (j = \overline{1, K}; i = \overline{1, N}). \quad (3.33)$$

2. Вычисление значений коэффициентов парной корреляции

$$\|R_{0j}\| \quad (j = \overline{1, K}). \quad (3.34)$$

3. Ранжирование аналогов ( $R_{0j} > R_{0j+1}$ ).

4. Расчет начинается с первого аналога  $K = 1$  и  $R_{0j}^{max}$ .

5. Определение коэффициентов множественной корреляции ( $R$ ) и среднеквадратической ошибки ( $S$ ).

6. а) если  $K \geq 2$  определяется изменение величин  $S$  и  $R$

$$\left| \frac{S_{K+1}}{\sigma_0} - \frac{S_K}{\sigma_0} \right| > \varepsilon, \quad (3.35)$$

и

$$|R_{K+1} - R_K| > \beta, \quad (3.36)$$



где  $S_K$  и  $S_{K+1}$  – относительные ошибки при использовании  $K$  и  $K+1$  аналогов;  $\sigma_0$  – среднеквадратическое отклонение приводимого ряда;  $R_K$  и  $R_{K+1}$  – аналогичные коэффициенты множественной корреляции;  $R_K \varepsilon$  и  $\beta$  – заданные числа.

б) если ( $K=1$ , то  $K=K+1$  расчет переходит к пункту 5.

7. Если выполняются неравенства (3.35) и (3.36), то привлекается очередной аналог  $K=K+1$  вычисления повторяются с пункта 5. Этот циклический процесс прерывается в случае невыполнения неравенств (3.35) и (3.36). Тогда первые  $K$  аналогов считаются эффективными и для уравнения регрессии (2.30) вычисляются неизвестные коэффициенты

$$A_j = (-1)^{j+1} \cdot \frac{\sigma_0 \cdot D_{0j}}{\sigma_j \cdot D_{00}}, \quad (3.37)$$

где  $\sigma_j$  – среднеквадратическое отклонение  $j$ -го аналога;  $D_{0j}$  – минор определителя  $D_{00}$ .

Значения стока, рассчитанные по уравнениям регрессии, систематически преуменьшают размах колебаний рассматриваемой гидрологической характеристики по сравнению с данными наблюдений.

Систематическое преуменьшение дисперсий исключается путем дополнительного расчета годовых значений  $\bar{M}_i$  по формуле

$$M_{i0}' = \frac{M_{i0} - \bar{M}_{n'}}{R} + \bar{M}_{n'}, \quad (3.38)$$

где  $M_{i0}$  – годовые значения гидрологической характеристики, рассчитанные по уравнению регрессии;  $\bar{M}_{n'}$  – среднее значение гидрологической характеристики за совместный период наблюдений.

Если имеющийся фактический материал не соответствовал требованиям, предъявляемым при статистической обработке, то такой ряд гидрометеорологических наблюдений считался недостаточным. Недостаточность данных обычно связана с коротким периодом наблюдений, не отражающим основной характер колебаний элемента водного баланса во времени. Однако есть случаи, когда и относительно долгий период наблюдений не дает необходимой надежности при расчете соответствующих характеристик, если они испытывают большие колебания во времени. Размах временных колебаний характеристик элементов водного ба-

ланса, обусловленных естественными причинами, зависит от степени увлажненности территории: чем меньше средняя увлажненность территории, тем резче колебания по годам (и внутри года по сезонам и месяцам) и тем больший период наблюдений нужен для определения их основных закономерностей.

Период наблюдений считался недостаточным [Международное ... 1984; Пособие ..., 2000; Пособие ..., 1984], если:

1) среднеквадратические погрешности статистических параметров ряда ( $\sigma_Q$  и  $\sigma_{C_v}$ ) превышали допустимые пределы;

2) имеющийся ряд данных не содержал полные циклы водности и наиболее многоводные или маловодные годы или группы лет (в зависимости от рассматриваемой характеристики). Используемые ряды должны быть качественно однородными и случайными.

Для надлежащей оценки гидрометеорологические параметры или в целом ряды наблюдений приводились к длительному (достаточному, репрезентативному) периоду. Основой приведения явился метод аналогии, состоящий в следующем:

1) в подборе к расчетному (неизученному) объекту другого, гидрометеорологически изученного, находящегося в сходных с неизученным физико-географических условиях;

2) в распространении гидрометеорологических характеристик изученного объекта на расчетный объект с вводом поправок на неполную аналогию физико-географических факторов. Для расчетов используют обычно один-три пункта гидрометеорологических наблюдений, расположенных сходных физико-географических условиях. Их можно использовать последовательно (парная корреляция) или одновременно (множественная корреляция).

Приведение параметров к многолетнему периоду осуществляется последовательно по нескольким уравнениям регрессии в порядке убывания парного или множественного коэффициентов корреляции при соблюдении следующих требований [Международное ..., 1984]:

$$n' \geq 10; R \geq 0,7; r \geq 0,7; \frac{k}{\sigma_k} \geq 2; T_a \geq T, \quad (3.39)$$

где  $n'$  – число лет совместных наблюдений в приводимом пункте и пункте-аналоге;  $R$  – множественный коэффициент корреляции ( $r$  – парный

коэффициент корреляции при одном аналоге) между значениями стока в пункте приведения и пунктах-аналогах;  $k$  – коэффициент регрессии;  $\sigma_k$  – среднее квадратическое отклонение коэффициента регрессии;  $T_a$  – период наблюдений на реке-аналоге;  $T$  – период наблюдений на исследуемой реке.

Для этой цели все уравнения регрессии, удовлетворяющие условиям (3.34), располагаются в порядке убывания коэффициентов корреляции. Затем по уравнению регрессии, имеющему наибольший коэффициент парной (один аналог) корреляции, восстанавливаются годовые значения стока в приводимом пункте за весь период наблюдений в пунктах-аналогах. Далее используется уравнение регрессии, коэффициент корреляции для которого меньше предыдущего, но больше всех остальных. По данному уравнению восстанавливаются годовые значения стока, которые не были восстановлены ранее.

Процедура поэтапного удлинения ряда продолжается до тех пор, пока не будет восстановлено требуемое количество значений стока, при этом вновь привлекаемые уравнения регрессии должны удовлетворять условиям 3.39.

По восстановленному ряду расходов  $M_{10}$  совместно с данными наблюдений рассчитываются параметры распределения ( $\bar{M}, C_V, C_S$ ) и коэффициент автокорреляции  $r(1)$  между значениями стока смежных лет.

По этой методике выполнено восстановление рядов месячных величин речного стока, атмосферных осадков, дефицитов влажности воздуха и др.

Полученные результаты проверялись методом скользящего контроля. Он состоит в том, что после выбора эффективных аналогов для восстановления ряда за  $i$ -ый месяц фактические значения приводимого ряда и рядов – аналогов за этот месяц изымались и коэффициенты регрессии вычислялись без них. Аналогичная процедура повторялась при восстановлении ряда за  $i+1$  месяц и т. д. В результате сравнивался вычисленный и фактический ряды за весь период наблюдений.

После того как подобраны расчетные уравнения и восстановлены месячные величины, необходимо увязать полученные результаты в годовом разрезе. Так как годовые величины восстанавливаются с меньшей ошибкой, чем месячные, то они могут быть контролем при восстановле-

нии месячных величин. Если невязка  $\left| \sum_{i=1}^{12} \bar{M}_{mi} - \bar{M}_{\Gamma} \right| = \Delta \rightarrow 0$ , где  $\bar{M}_{mi}$  – восстановленная гидрологическая величина за  $i$ -ый месяц;  $\bar{M}_{\Gamma}$  – то же за год, меньше допустимой, то она распределяется пропорционально месячным величинам на каждый месяц.

Результаты расчетов показывают хорошую сходимость фактических и рассчитываемых рядов (рис. 3.4).

Для сопоставимости получаемых результатов все исследуемые ряды приведены к единому периоду наблюдений, равному 56 лет (1945 – 2000 гг.). Продолжительность репрезентативного периода установлена по разностям интегральным кривым с учетом многолетних колебаний исследуемой величины. Продление выполнено для годового, максимального весеннего и минимального летнего речного стока. В качестве примера приведены результаты продления для годового стока рек Брестской области (табл. 3.8).

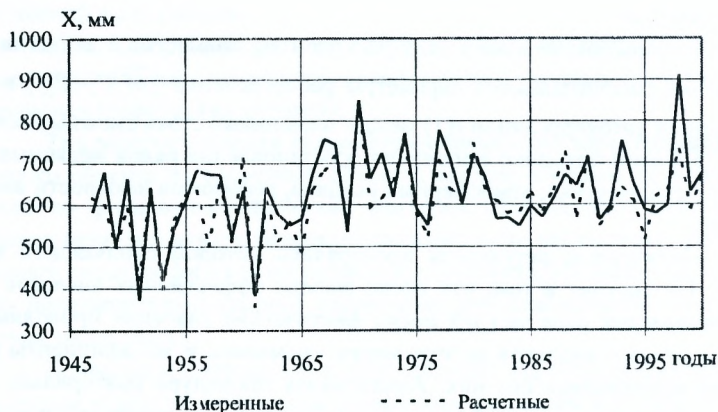


Рис. 3.4. Измеренные и расчетные значения годовых величин атмосферных осадков по метеостанции Ганцевичи (уравнение регрессии  $X_{28}=0,217 \cdot X_{27}+0,564 \cdot X_{31}+0,225 \cdot X_{29}+28,513$ ) (коэффициент корреляции  $r=0,93$ ).

Таблица 3.8

**Основные характеристики исходных (в числителе) и про-  
дленных (в знаменателе) рядов годового стока**

Реки – створ	Норма стока, м <sup>3</sup> /с	Коэффициенты		
		вариации	асимметрии	корреляции
Бобрик – с. Парохонск	5,87/6,13	0,37/0,36	0,17/0,30	0,15/-0,01
Горынь – пос. Горынь	77,2/74,7	0,34/0,29	0,99/1,29	0,12/-0,19
Горынь – пгт. Речица	111/102	0,31/0,33	0,43/0,59	0,09/0,16
Гривда – гпт. Ивацевичи	2,78/2,82	0,32/0,27	2,37/1,59	0,08/-0,05
Жабинка – с. Малая Жа- бинка	0,64/0,61	0,45/0,40	1,09/1,47	0,07/0,08
Жегулянка – с. Нехачево	1,13/0,98	0,31/0,40	1,84/0,98	-0,17/0,08
Каменка – пос. Мухавец	0,31/0,30	0,37/0,47	0,29/1,01	-0,09/0,04
кан. Винец – с. Рыгали	0,74/0,66	0,34/0,44	1,81/1,16	-0,20/0,00
Копаяювка – с. Черск	1,29/1,26	0,56/0,55	0,79/0,88	0,19/0,19
Лесная – с. Замосты	8,42/8,43	0,28/0,28	1,20/1,20	0,16/0,10
Лесная – с. Тюхиничи	<b>10,2/11,4</b>	<b>0,16/0,28</b>	1,38/1,22	<b>0,01/0,16</b>
Малорыга – г. Малорита	1,87/1,89	0,46/0,48	1,31/0,81	-0,04/-0,04
Меречанка – с. Ставок	0,50/0,50	0,32/0,28	<b>1,22/0,44</b>	0,30/0,15
Меречанка – с. Крассево	0,58/0,54	0,28/0,32	0,51/0,52	0,21/0,17
Мухавец – г. Брест	26,2/24,6	0,57/0,54	2,62/2,77	0,14/0,10
Мухавец – г. Пружаны	0,340/0,36	<b>0,41/0,33</b>	1,29/0,99	0,01/0,09
Мышанка – с. Березки	3,70/3,94	0,28/0,34	0,88/0,57	0,45/0,65
Неслуха – с. Рудск	1,40/1,36	0,31/0,40	<b>0,19/1,80</b>	-0,41/-0,02
Припять – с. Коробы	115/117	0,37/0,36	0,66/0,78	0,15/-0,09
Припять – пгт. Туров	263/268	0,34/0,34	0,82/0,77	0,06/0,03
Припять – г. Пинск	70,4/64,1	0,29/0,30	0,46/0,62	-0,54/-0,08
Пульва – г. Высокое	1,21/1,21	0,29/0,27	0,96/0,88	0,19/0,14
Рудавка – с. Рудня	0,71/0,67	0,42/0,37	1,44/1,30	0,08/0,03
Ружанка – г. Ружаны	<b>3,31/2,47</b>	1,11/0,97	1,39/2,69	0,43/0,46
Рыга – с. Малые Радваничи	4,16/4,13	0,47/0,44	0,97/1,03	0,04/0,05
Цна – с. Дятловичи	4,59/4,35	0,42/0,44	0,40/0,45	<b>-0,05/0,09</b>
Щара – с. Доманово	17,2/16,6	0,28/0,23	1,07/0,53	0,07/0,03
Щара – с. Залужье	4,12/4,01	0,24/0,28	0,94/1,36	-0,27/-0,12
Ясельда – г. Береза	4,80/4,80	0,35/0,35	0,74/0,74	-0,06/-0,06
Ясельда – с. Сепин	19,3/19,5	0,38/0,37	0,79/0,72	0,11/0,02
<b>Среднее</b>	<b>25,4/25,0</b>	<b>0,38/0,39</b>	<b>1,03/1,13</b>	<b>0,04/0,07</b>

Следует отметить, что наиболее перспективным для решения задач по продлению и восстановлению тепловоднобалансовых характеристик



[Волчек, Лукша, 2003] является использование нейросетевого моделирования, в основе которого лежит разработка компьютерных моделей, построенных по аналогии со структурой поведения реальных нейронов и, что самое ценное, они способны распознавать образы, обрабатывать их, а также обучаться. Совершенствование методов построения, использования и обучения нейронных сетей позволит перейти к практическому применению этих разработок.

Использование нейронных сетей позволит, наряду с повышением точности, автоматизировать процесс восстановления и продления временных рядов, исключить влияние субъективного фактора.

Естественные и восстановленные (квазиестественные) характеристики стока как гидрологическая основа (для суждения об его природных колебаниях) оценки водных ресурсов могут применяться:

- а) непосредственно рядами гидрометрических наблюдений;
- б) репрезентативными (расчетными) гидрологическими характеристиками;
- в) искусственными (смоделированными) гидрологическими рядами большой длительности.

### 3.4. Особенности выбора оптимального числа предсказателей

В силу того, что влияющие факторы  $x_2, x_3, \dots, x_n$ , как правило, зависимы, то в ряде задач нет необходимости привлекать их достаточно много. Поэтому при статистической обработке больших эмпирических данных, целесообразно предварительно произвести статистическую оценку и выбрать оптимальное число влияющих факторов, при достижении которого дальнейшее их увеличение не приведет к существенному уменьшению ошибки  $S_{1,2,3,\dots,n}$ . Такой метод получил название метода просеивания. Можно использовать два способа.

*Первый* способ заключается в последовательном вычислении значения множественного коэффициента корреляции и корреляционной матрицы парных коэффициентов корреляции и оценке оптимального числа предсказателей до нахождения коэффициентов уравнения регрессии.

Эффективность  $n$ -го предсказателя может быть оценена по разности  $R_n^2 - R_{n-1}^2 = \Delta R^2$  и тогда соответственно  $R_n^2 = R_{n-1}^2 - r_n^2 - r_n^2 \cdot R_{n-1}^2$



или  $\Delta R^2 = r_n^2 \cdot (1 - R_{n-1}^2)$ , где  $R_n$  – множественный коэффициент корреляции, полученный с учетом  $n$  влияющих факторов;  $R_{n-1}$  – множественный коэффициент корреляции с учетом  $n-1$  факторов;  $r_n$  – частный коэффициент корреляции между искомой величиной ( $x_1$ ) и ( $x_n$ ) при учете всех остальных ( $n-1$ ) факторов.

Нетрудно видеть, что при малых  $r_n$  не имеет смысла привлекать  $n$ -ый предсказатель, так как не только увеличивается объем вычислений, но и возможно проявление отрицательного эффекта из-за плохой обусловленности получаемой системы уравнений.

*Второй* способ заключается в последовательном переборе всех предсказателей до достижения «насыщения», обуславливающего факт, что при дальнейшем увеличении числа влияющих факторов средняя квадратическая ошибка заметно не уменьшается и соответственно не увеличивается коэффициент множественной корреляции.

Последовательность расчетных операций следующая:

- составляется корреляционная матрица парных коэффициентов корреляции при использовании  $n$  влияющих факторов;
- выбирается матрица коэффициентов, составленная с учетом наиболее значимых факторов, например,  $x_2$  и  $x_3$ ;
- по формулам

$$S_{1,2,3...n} = \sigma_2 \cdot \sqrt{\frac{D}{D_{11}}}; R_{1,2,3...n} = \sqrt{1 - \frac{D}{D_{11}}}, \quad (3.40)$$

где  $D = \begin{vmatrix} 1 & r_{12} & r_{13} & \dots & r_{1n} \\ r_{21} & 1 & r_{23} & \dots & r_{2n} \\ r_{31} & r_{32} & 1 & \dots & r_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ r_{n1} & r_{n2} & r_{n3} & \dots & 1 \end{vmatrix}$ ;  $D_{11}$  – соответствующий минор, получае-

мый в результате вычеркивания первой строки и первого столбца, рассчитываются  $S_{1,2,3...n}$  или  $R_{1,2,3...n}$  при этом –

$$D = \begin{vmatrix} 1 & r_{12} & r_{13} \\ r_{21} & 1 & r_{23} \\ r_{31} & r_{32} & 1 \end{vmatrix}; D_{11} = \begin{vmatrix} 1 & r_{23} \\ r_{32} & 1 \end{vmatrix}; \quad (3.41)$$

в) прибавляется предсказатель  $x_4$  и соответственно оценивается величина  $S_{1,2,3...n}$  или  $R_{1,2,3...n}$  при

$$D = \begin{vmatrix} 1 & r_{12} & r_{13} & r_{14} \\ r_{21} & 1 & r_{23} & r_{24} \\ r_{31} & r_{32} & 1 & r_{34} \\ r_{41} & r_{42} & r_{43} & 1 \end{vmatrix}; \quad D_{11} = \begin{vmatrix} 1 & r_{23} & r_{24} \\ r_{32} & 1 & r_{34} \\ r_{42} & r_{43} & 1 \end{vmatrix}; \quad (3.42)$$

г) если выигрыш незначительный, то  $x_4$  отбрасывается, а вместо него испытывается предсказатель  $x_5$ . В этом случае в определителе изменится последний столбец и последняя строка. Такие операции осуществляются до достижения «насыщения».

Иногда бывает целесообразно располагать влияющие факторы в такой последовательности, чтобы их место соответствовало бы их значимости. Для этой цели корреляционная матрица трансформируется таким образом, что на первом месте помещается тот предсказатель  $x_2$  для которого  $r_{12}$  наибольшее, затем следует  $x_3$  для которого имеет следующее по абсолютной величине  $r_{13}$  и т. д.

### 3.5. Оценка значимости и тесноты связи между качественными признаками

В практике применения корреляционного анализа нередко встречаются случаи, когда исследуемые признаки не поддаются измерению и не распределяются в вариационный ряд. Для них целесообразно использовать коэффициент ассоциации Юла или коэффициент сопряженности [Дэвис, 1990].

#### Методика использования коэффициента ассоциации

Если учитываемые признаки группируются в четырехклеточную корреляционную решетку (табл. 3.9), степень сопряженности между ними измеряется с помощью коэффициента ассоциации (тетрагорическим показателем связи):

$$r = \frac{a \cdot d - b \cdot c}{\sqrt{(a+b) \cdot (c+d) \cdot (a+c) \cdot (b+d)}}, \quad (3.43)$$

где  $a, b, c, d$  – частоты признаков в ячейках табл. 3.9.

Квадратный корень в формуле (3.43) всегда берется со знаком плюс. В табл. 3.9 символы  $\bar{A}$  и  $\bar{B}$  обозначают признаки, противоположные признакам  $A$  и  $B$ , или указывают на их отсутствие. Тогда  $a, b, c$  – частоты комбинаций  $AB, A\bar{B}, \bar{A}B, \bar{A}\bar{B}$ . Коэффициент ассоциации  $r$  всегда заключен между  $-1$  и  $+1$ . Когда  $r=0$ , признаки  $A$  и  $B$  не зависят друг от друга (тогда  $a/c=b/d$  и « $A$ » появляется относительно одинаково часто совместно с « $B$ » и « $\bar{B}$ »). Если  $r=-1$ , при появлении  $B$ , признак  $A$  не появляется, а при появлении  $B$  (событие  $\bar{B}$ ) – появляется. Когда  $r=+1$ , то  $A$  появляется и не появляется только одновременно с появлением и непоявлением  $B$ . Таким образом, при  $r=\pm 1$ , имеется полная прямая или обратная связь между  $A$  и  $B$ . Достоверность выборочного коэффициента ассоциации оценивается по его отношению к средней ошибке  $\sigma_r$ , определяемой по формуле, аналогичной формуле  $\sigma_r = (1-r^2)/\sqrt{n}$ . Нулевая гипотеза заключается в предположении, что связь между учитываемыми альтернативными признаками отсутствует. Гипотеза  $H_0$  отвергается, если  $(|r|/\sigma_r \cdot f \cdot t_{\alpha, m})$ , при заданных уровне значимости  $\alpha$  и  $m=n-2$  – степенях свободы. Для оценки достоверности выборочного коэффициента ассоциации можно воспользоваться также данными табл. 3.1. Если величина  $[|r| \cdot \sqrt{n-1}]$  превосходит указанные в таблице критические значения, для принятого уровня надежности  $P=1-\alpha$ , нулевая гипотеза отвергается;

Таблица 3.9

Таблица сопряженности категорий двух признаков

Признак	$B$	$\bar{B}$	Сумма
$A$	$a$	$b$	$a+b$
$\bar{A}$	$c$	$d$	$c+d$
Сумма	$a+c$	$b+d$	$n$

#### Методика использования коэффициента сопряженности

В некоторых случаях, данные качественные признаки подразделяются более, чем на два класса. Тогда мерой тесноты связи между ними служит коэффициент сопряженности Пирсона:

$$C\sqrt{\chi^2/(\chi^2+n)}, \quad (3.44)$$

где  $n$  – число наблюдений, а  $\chi^2$  – определяется как:

$$\chi^2 = \sum_{i=1}^k \frac{(m_i - m'_i)^2}{m_i}, \quad (3.45)$$

где  $m_i$  и  $m'_i$  – частоты комбинаций различных классов (градаций) коррелируемых признаков (фактические  $m_i$  и  $m'_i$  – при полном отсутствии связи между признаками);  $k$  – число категорий в таблице сопряженности коррелируемых признаков.

Коэффициент сопряженности  $C$  как мера тесноты связи между качественными признаками изменяется от 0 до размеров таблицы сопряженности, т. е.  $0 \leq C \leq C_{max}$ . Если число строк и столбцов в таблице сопряженности одинаково и равно  $S$ , то:

$$C_{max} = \sqrt{(S-1)/S}. \quad (3.46)$$

В частности, при  $S=3$ , коэффициент сопряженности, для случая идеальной связи, равен 0,817. Для того, чтобы верхнюю границу величины  $C$  привести к 1, вычисленное по формуле (3.46) значение  $C$  рекомендуется разделить на произведение значений  $C'$  (табл. 3.10), в зависимости от числа классов (градаций), на которые разбит каждый коррелируемый признак.

Таблица 3.10

### Поправка к $C$ на число классов $S$

$S$	$C'$	$S$	$C'$	$S$	$C'$
2	0,798	5	0,943	10	0,985
3	0,859	6	0,959	12	0,989
4	0,915	8	0,976	15	0,999

Значимость коэффициентов сопряженности  $C$  оценивается с помощью критерия  $\chi^2$ . Нулевая гипотеза заключается в предположении, что связь между качественными признаками отсутствует. Гипотеза  $H_0$  отвергается, если, вычисленное по формуле (3.45), значение  $\chi^2$  окажется больше некоторого  $\chi^2_{\alpha, \nu}$ , определяемого по специальным таблицам при заданном уровне значимости  $\alpha$  и  $\nu$  – степенях свободы. Распределение  $\chi^2$  непрерывно, а распределение частот – дискретно. Поэтому в результаты вычисления  $\chi^2$  по формуле (3.45), рекомендуется вводить поправки. Суть

их состоит в том, что из каждого возведенного в квадрат абсолютного значения разности  $|m_i - m_i'|$  вычитается 0,5.

Сходной, но не вполне сравнимой мерой сопряженности между качественными признаками, когда каждый из них делится более чем на два качественных класса, является также коэффициент сопряженности А. А. Чупрова  $T$ , вычисляемый без поправочного коэффициента  $C'$ :

$$T = \sqrt{\frac{\chi^2}{n \cdot (k-1)^{0,5} \cdot (m-1)^{0,5}}}, \quad (3.47)$$

где  $n$  – число наблюдений;  $k$  – число строк;  $m$  – число граф в таблице сопряженности.

Рассмотрим порядок вычисления корреляции между качественными признаками на гидрологических данных:

а) данные о расходах воды в р. Неман – г. Гродно за межень и год, в целом, в период с 1966 по 2000 гг. представлены в виде их обеспеченности, причем, разделены на два класса: «выше нормь» ( $P_2 < 50\%$ ,  $P_M < 50\%$ ), где «г» – индекс года, «м» – индекс межени, и «равно» и «ниже нормь» ( $P_2 \geq 50\%$ ,  $P_M \geq 50\%$ ).

Имеется ли синхронная связь между годовыми и меженными расходами р. Неман – г. Гродно?

Пользуясь данными табл. 3.11 и формулой (3.43), находим:  $r = 0,314$ , что указывает на наличие слабой прямой связи между годовым и меженным стоком р. Неман – г. Гродно за период с 1966 по 2000 гг. Оценим достоверность выборочного коэффициента ассоциации. Подставив числовое значение полученного  $r$  в формулу  $(|r| \cdot \sqrt{n-1})$ , находим  $(|0,314| \cdot \sqrt{35-1}) = 1,831$ . Эта величина меньше критического значения  $(1,831 < 1,950)$ , для  $P=0,95$  и  $n=35$  (табл. 3.1). Следовательно, при принятом уровне достоверности  $P=0,95$ , связь между рассматриваемыми признаками в генеральной совокупности носит лишь случайный характер, т. е. не является надежной;

б) рассмотрим случай, когда качественные признаки можно разделить более чем на два качественных класса.

Таблица 3.11

**Сопряженность годовых и меженных расходов  
р. Неман – г. Гродно (1966 - 2000 гг.)**

$P$	$P < 50\%$	$P \geq 50\%$	Сумма
$P_M < 50\%$	$a=11$	$b=6$	$a+b=17$
$P_M \geq 50\%$	$c=6$	$d=12$	$c+d=18$
Сумма	$a+c=17$	$b+d=18$	

Ежегодные обеспеченности стока р. Припять – с. Туров за межень  $P_c$  и обеспеченности мелиоративных норм за июль в зоне метеостанции г.п. Житковичи  $P_M$  за период с 1966 по 2000 гг. разделены на три градации: 1) «норма» ( $H$ ) – рассматриваемые величины лежат в диапазоне обеспеченностей – ( $33\% \leq P \leq 66\%$ ); 2) «выше нормы» ( $BH$ ) для стока –  $P < 33\%$  и для мелиоративных норм –  $P > 66\%$ ; 3) «ниже нормы» ( $HH$ ) для стока –  $P > 66\%$  и для мелиоративных норм –  $P < 33\%$ ; они сведены в таблицу сопряженности (табл. 3.12).

Таблица 3.12

**Сопряженность обеспеченностей стока р. Припять – ст. Туров за межень и обеспеченности мелиоративных норм в зоне г.п. Житковичи за июль (1966 – 2000 гг.)**

Обеспеченность	Обеспеченность мелиоративных норм			Сумма
	$HH$	$H$	$BH$	
$HH$	$m_1=2$	$m_2=5$	$m_3=3$	10
	$m_1'=3$	$m_2'=4$	$m_3'=3$	10
$H$	$m_4=3$	$m_5=4$	$m_6=6$	13
	$m_4'=3$	$m_5'=5$	$m_6'=4$	12
$BH$	$m_7=4$	$m_8=5$	$m_9=3$	12
	$m_7'=3$	$m_8'=6$	$m_9'=4$	13
Сумма	9	14	12	35
	9	15	11	(35)

Какова связь между обеспеченностью стока р. Припять – с. Туров в межень и мелиоративной нормой за июль месяц в зоне метеостанции г.п. Житковичи. Цифры в скобках (табл. 3.12) показывают частоты  $m_i'$ , если бы между рассматриваемыми характеристиками не было связи. Эти числа представляют собой произведения суммы столбца на сумму стро-



ки, деленные на общую сумму. Так, в нашем случае, для первой клетки имеем:  $mi^2 = (10 \cdot 9) / 35 = 3$ .

Пользуясь данными табл. 3.12 и рекомендациями по вычислению  $\chi^2$ , согласно (3.45), находим:  $\chi^2 = 0,99$ .

По формуле (3.42), при  $n=35$ , имеем:  $C = 0,165$ .

По табл. 3.10, находим поправку  $C^*$  к  $C$  на число классов:  $C^* = 0,737$ . Деление  $C = 0,165$  на  $C^* = 0,737$  дает исправленное значение коэффициента сопряженности:  $C^* = 0,224$ . Найденное значение коэффициента сопряженности  $C^* = 0,224$  указывает на определенную, хотя и не очень тесную, связь между обеспеченностями расходов воды в межень р. Припять – с. Туров и обеспеченностями мелиоративных норм за июль месяц для зоны метеопункта Житковичи. Однако при 4-х степенях свободы (табл. 3.12) и  $\alpha = 10\%$ , табличное значение  $\chi_{0,10;4}^2 = 7,78$  и превышает найденное  $\chi^2 = 0,99$ . Таким образом, нулевая гипотеза об отсутствии связи между рассматриваемыми признаками (в генеральной совокупности) при принятом уровне значимости не отвергается. Воспользуемся мерой сопряженности между рассматриваемыми качественными признаками (коэффициент сопряженности «Г» А. А. Чупрова), при  $m=k=3$ ,  $T = 0,12$ .

Значение величины  $T$  также указывает на наличие слабой связи между рассматриваемыми признаками. Отметим, что в некоторых случаях, особенно когда непараметрические критерии для определения связей имеют отрицательные значения, их интерпретация затруднительна. Абсолютные значения критериев могут изменяться в пределах от 0 до  $\pm 1$ . Чем ближе абсолютные значения к единице, тем теснее связь между исследуемыми признаками. При этом, абсолютные величины коэффициентов, соответствующие условию высокой связи, в случае корреляции больше двух признаков, как правило, ниже, чем в случае корреляции двух признаков.

Верхний предел коэффициентов сопряженности зависит от размеров таблицы сопряженности. Поэтому сравнивать можно только коэффициенты, полученные из таблиц с одинаковым числом столбцов и строк. Не принято сравнивать коэффициенты сопряженности с коэффициентами линейной корреляции, вычисленными более точно на основе параметрических критериев. Вместе с тем, коэффициенты ассоциации способны

измерять взаимосвязи, отличающиеся от линейных с произвольным распределением частот.

### 3.6. Корреляция числовой случайной переменной с дихотомической

Дихотомическая переменная, как известно, может принимать два исключающих друг друга значения (дождь или отсутствие дождя, туман или отсутствие тумана и т. д.).

Проанализируем связь между точкой росы и частотой туманов. Среднее значение точки росы для случаев с туманом  $\bar{x}_1$ , а без тумана –  $\bar{x}_2$ . Тогда

$$r_{bis} = \frac{\bar{x}_2 - \bar{x}_1}{\sigma} \cdot \frac{P \cdot q}{Z}, \quad (3.48)$$

где  $r_{bis}$  – коэффициент бисеральной корреляции;  $P$  и  $q$  – соответственно эмпирические вероятности наличия или отсутствия тумана;  $Z$  – ордината нормальной кривой, которая делит всю площадь под этой кривой на две части (допустим  $P=30\%$  с одной стороны и  $P=70\%$  с другой стороны).

Площадь заключенная между этой ординатой и средним значением, составляет 20%. Мету этой площади можно найти по специальным таблицам.

### 3.7. Особенности выявления тесноты нелинейных связей (информационно-энтропийный анализ)

Кроме корреляционных методов, которые позволяют определить тесноту связи и использующих значения членов сопоставляемых рядов без выявления их вероятностных характеристик, известны другие показатели, которые также являются количественной зонной мерой тесноты связей и основаны на использовании только вероятностных характеристик, связанных между собой систем. Последние имеют то преимущество, что они позволяют оценить тесноту нелинейных связей. Кроме того, на их величину не могут оказать большого влияния резко отклоняющиеся члены выборки, как это имеет место при вычислении корреляционных функций. Таким показателем тесноты связи является показатель сопряженности переменных Чупрова-Пирсона и аналогичное ему по смыслу

энтропическое или информационное отношение. Практическое использование энтропии было показано в разделе 1.2.

Однако не менее значимым и чрезвычайно важным является понятие информации  $\mathfrak{I}_x$ , характеризующее собой степень осведомленности о системе.  $\mathfrak{I}_x$  равна разности энтропии системы  $X$ , вычисленных до и после получения этой информации:

$$\mathfrak{I}_x = H(X)_0 - H(X). \quad (3.49)$$

Допустим, что необходимо определить одно из  $K=100$  возможных состояний системы  $Y$ .

Если о системе  $Y$  ничего неизвестно, то естественно предположить, что каждое из состояний является равновероятным. Степень нашего незнания может быть выражена энтропией системы  $Y$ , которая в данном случае будет максимальной:

$$H_{max} = -\sum_{\ell=1}^K P_{\ell} \cdot \lg P_{\ell} = -\sum_{\ell=1}^{100} \frac{1}{100} \cdot \lg \frac{1}{100} = -(\lg 1 - \lg 100), \quad (3.50)$$

или  $H_{max} = \lg n = \lg 100 = 6,643$  (двоичных единиц);

Если после этого станет известно, что все 100 возможных состояний системы  $Y$  равномерно группируются по 10 (т. е. весь диапазон колебаний разбивается на  $\ell=10$  разрядов  $y_{\ell} = y_1, y_2, \dots, y_{10}$ , чтобы в каждом из них оказалось по 10 состояний системы  $Y$ ), то для определения необходимого состояния системы достаточно будет найти одну из 10 групп. В данном случае стало кое-что известно (группировка по 10) и степень нашего незнания уменьшилась. Теперь она выражается энтропией:

$$\begin{aligned} H(Y)_1 &= -\sum_{\ell=1}^{K^2=10} P_{\ell} \cdot \lg P_{\ell} = -\sum_{\ell=1}^{K^2=10} \frac{1}{100} \cdot \lg \frac{1}{100} = \\ &= -\lg \frac{1}{10} = -(\lg 1 - \lg 10) = \lg 10 \end{aligned} \quad (3.51)$$

или  $H(Y)_1 = \lg 10 = 3,321$  (двоичных единиц).

Приращение информации, т. е. наша осведомленность о системе  $Y$  равняется:

$$\mathfrak{I}_1 = H_{max} - H(Y)_1 = 6,643 - 3,321 = 3,322 \text{ (двоичных единиц)}. \quad (3.52)$$

Если же система  $Y$  принимает состояние  $y_1$  в 22 случаях из 100, состояние  $y_2$  – в 10 случаях и т. д., как показано в табл. 3.13, т. е. возможные

состояния системы образуют неодинаковые группы, то осведомленность о системе по сравнению с предыдущим случаем увеличивается, и степень нашего незнания о ней будет характеризоваться энтропией:

$$H(Y)_2 = -(0,22 \cdot \lg 0,22 + 2 \cdot 0,10 \cdot \lg 0,10 + 2 \cdot 0,07 \cdot \lg 0,07 + 2 \cdot 0,06 \cdot \lg 0,06 + 0,08 \cdot \lg 0,08 + 2 \cdot 0,12 \cdot \lg 0,12)$$

Но и после этого имеет место определенная степень нашего незнания о системе  $Y$ . Для увеличения осведомленности о системе надо использовать связанную с ним систему  $X$  (если такая имеется). В этом случае можно подсчитать энтропию  $H(Y/X)$ , которая характеризует степень неопределенности системы  $Y$ , остающуюся после того, как состояние системы  $X$  полностью определилось, и называется полной условной энтропией системы  $Y$  относительно  $X$ .

Последняя определяется по формуле:

$$H(Y/X) = - \sum_{q=1}^K \sum_{\ell=1}^S P_{q\ell} P(Y_\ell / X_q) \cdot \lg P(Y_\ell / X_q), \quad (3.53)$$

где величины  $P_{q\ell}$ , характеризуют вероятности того, что при состоянии  $X_q$  система  $Y$  находится  $y_\ell$ ,  $\ell=1, 2, 3, \dots$  и  $q=1, 2, 3, 4$ .

Используя ранее приведенное соотношение:  $\eta(p) = -p \cdot \lg p$  последнюю формулу можно записать так:  $\eta(p(y_\ell / x_q)) = -p(y_\ell / x_q) \cdot \lg p(y_\ell / x_q)$ , тогда получим окончательно:

$$H(Y/X) = - \sum_{q=1}^K \sum_{\ell=1}^S P_{q\ell} \cdot \eta[P(y_\ell / X_q)] \quad (3.54)$$

Таблица 3.13

### Распределение состояний системы $Y$ и их вероятностей

$y_\ell$	$Y_1$	$Y_2$	$Y_3$	$Y_4$	$Y_5$	$Y_6$	$Y_7$	$Y_8$	$Y_9$	$Y_{10}$	$\Sigma$
Количество состояний	22	10	10	7	7	6	6	8	12	12	100
Вероятность	0,22	0,10	0,10	0,07	0,07	0,06	0,06	0,08	0,12	0,12	1,0

Кроме количественной оценки степени неопределенности системы, статистическая энтропия и информация могут быть использованы для оценки тесноты связей рассматриваемых систем.

Действительно уменьшение нашего незнания о системе происходит потому, что последовательно обнаруживаются связи элементов системы  $Y$  между собой (группировка членов), связь элементов этой системы с другой системой  $X$ .

Введем новый показатель – информационное отношение

$$J = \frac{J_{x \rightarrow y}}{H(Y)} = \frac{H(Y) - H(Y/X)}{H(Y)} = 1 - \frac{H(Y/X)}{H(Y)}, \quad (3.55)$$

Этот показатель может быть использован, как измеритель тесноты связи на разных этапах получения информации о системе  $Y$ , получаемую за счет того, что удаётся наблюдать связанную с нею другую систему  $X$ . В случае независимости систем  $H(Y) = H(Y/X)$ ,  $J_{x \rightarrow y} = 0$ , т. е. если к неизвестной системе  $Y$  добавить не связанную с ней систему  $X$ , то степень незнания системы  $Y$  не изменится.

Таким образом, информационное отношение является количественной мерой тесноты связи. Дополнительная информация о системе  $Y$  получается за счет названной связи. Без существования этой связи было бы невозможно уменьшить неопределенность системы  $Y$ .

При оценке тесноты внутрирядных связей геофизических процессов вместо  $X(i)$  необходимо рассмотреть систему  $X(i+\tau)$ , полученную сдвигом членов ряда  $X(i)$  на  $\tau$  членов (интервалов времени).

В этом случае энтропический показатель тесноты внутрирядной связи (информационное отношение) может быть представлен в виде:

$$J = \frac{J_{i \rightarrow i+\tau}}{H(X_{i+\tau})} = 1 - \frac{H(X_{i+\tau}/X_i)}{H(X_{i+\tau})}, \quad (3.56)$$

Принимая для сокращения  $H(X_{i+\tau}) = H(X_j) = H(j)'$  и  $H(X_i) = H(i)$  можно записать:

$$J = 1 - \frac{H(j/i)}{H(j)}. \quad (3.57)$$

Вычисление информационной функции производится по соответствующим таблицам частот, составленным по рядам натуральных наблюдений. При этом весь диапазон изменений  $X_i$  и  $X_{j=i+\tau}$  разбивается на «к» разрядов.

Вычисления энтропии и информационного отношения может быть основана и на принципе экстремальности энтропии. Например, при заданной корреляционной функции максимальной энтропией обладает Гауссов процесс. Его энтропия вычисляется по формуле:

$$H(X) = \lg\left((2 \cdot \pi \cdot \ell)^{n/2} \cdot \sqrt{D}\right) \quad (3.58)$$

где  $n$  – число членов реализации;  $D$  – определитель корреляционной матрицы.

Оценить точность корреляционной функции трудно, а это, в свою очередь, затрудняет оценку энтропии. Также неопределенной получается энтропия по относительным частотам, являющимся оценками вероятностей, так как последние зависят от объема выборки, кроме того, если процесс нестационарный, то частоты не всегда могут быть надежными оценками вероятностей.



#### 4. МЕТОДЫ МАТЕМАТИЧЕСКОГО ОПИСАНИЯ И ВЗАИМООЦЕНКИ ПАРАМЕТРОВ ВЫЯВЛЕННЫХ ЗАКОНОМЕРНОСТЕЙ

Математически описать выявленные закономерности и оценить одни параметры через другие, т. е. развить корреляционные связи, позволяет регрессионный анализ.

Составив и решив уравнение регрессии, можно произвести выравнивание эмпирических линий регрессии, т. е. выполнить моделирование исследуемого процесса (явления) путем подбора функции, график которой представляет собой теоретическую линию регрессии. Если подобная функция отражает сущность процесса или явления, то возможно прогнозирование значений признака за пределами сделанных наблюдений.

Подобно корреляции, регрессия может быть парной (простой) и множественной, по форме связи – линейной и нелинейной, по зависимости – односторонней (изменяется лишь один признак под влиянием другого) и двусторонней (взаимодействуя, изменяются оба признака).

Регрессия выражается несколькими способами: путем построения эмпирических линий, путем составления уравнения и, затем, – построения теоретических линий регрессии, а также с помощью коэффициента регрессии. Уравнение наиболее точно выражает зависимость между двумя переменными  $X$ ,  $Y$ , если корреляция между ними близка к единице.

*Уравнения регрессии могут быть составлены одним из следующих способов:*

- координат точек, с использованием двух-трех точек, расположенных на эмпирической линии (начале, середине и конце), для тех случаев, когда не требуется большая точности расчетов;
- наименьших квадратов, когда для составления уравнения регрессии привлекаются все сопряженные наблюдения и требуется повышенная точность расчетных величин.

Рассмотрим простые способы составления уравнений регрессии.

##### 4.1. Уравнение линейной регрессии с одним переменным фактором

Линейная модель имеет вид:

$$Y_i = \beta_0 + \beta_1 \cdot X_i + \epsilon_i. \quad (4.1)$$

где  $Y_i$  – зависимая переменная (предиктор, функция отклика);  $X_i$  – независимая переменная (входная переменная, фактор, регрессор);  $\beta_0$ ,  $\beta_1$  и  $\epsilon_i$  – постоянные коэффициенты.

В уравнении (4.1) величины  $\beta_0$ ,  $\beta_1$  и  $\epsilon_i$  неизвестны, причем, последнюю, на самом деле, будет трудно исследовать, поскольку она меняется в ходе наблюдений. Величины  $\beta_0$  и  $\beta_1$  остаются постоянными, но без точного изучения всех возможных сочетаний  $Y_i$  и  $X_i$ , ограничимся лишь их оценками  $b_0$  и  $b_1$ . В данной интерпретации запишем:

$$\hat{Y}_i = b_0 + b_1 \cdot X_i. \quad (4.2)$$

При этом –

$$b_1 = \frac{\sum_{i=1}^n X_i \cdot Y_i - ((\sum_{i=1}^n X_i) \cdot (\sum_{i=1}^n Y_i)) / n}{\sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2 / n} = \frac{\sum_{i=1}^n (X_i - \bar{X}) \cdot (Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}; \quad (4.3)$$

$$b_0 = \bar{Y} - b_1 \cdot \bar{X}.$$

Для статистического оценивания коэффициентов регрессии проверяется нуль-гипотеза  $H_0: \beta=0$ , т. е. условие, – отличается ли статистически значимо оценка коэффициента регрессии от нуля? Граница значимости устанавливается на основании распределения Стьюдента:

$$t = |b| / \bar{S}_b \geq t_{(n-2, P)}^T; \quad (4.4)$$

где  $\bar{S}_{b_1} = \bar{S}_{ост} / \sqrt{\sum_{i=1}^n (X_i - \bar{X})^2}$ ;  $\bar{S}_{b_0} = \bar{S}_{ост} / \sqrt{n}$ ;

$\bar{S}_{ост} = \sqrt{\sum_{i=1}^n (y_i - \hat{Y}_i)^2 / (n-2)}$  – остаточная дисперсия.

Погрешность предсказанного среднего значения «Y» (или « $Y_0$ ») при заданном « $X_0$ ») определяется значением соответствующей дисперсии:

$$\bar{S}_{y(x)}^{-2} = \bar{S}_{ост}^{-2} \cdot \left[ \frac{1}{n-2} + \frac{(X_i - \bar{X})^2}{\sum_{i=1}^n (X_i - \bar{X})^2} \right]. \quad (4.5)$$

Дисперсия  $\bar{S}_{y(x)}^{-2}$  характеризует рассеивание ординат выборочной линии регрессии относительно линии регрессии, соответствующей генеральной совокупности. Для проверки значимости уравнения регрессии, в

целом, с использованием  $F$ -критерия Фишера, общая дисперсия  $\overline{S}_Y^2$  сравнивается с остаточной дисперсией  $\overline{S}_{ост}^2$ , которая представляет собой показатель ошибки предсказания на базе уравнения регрессии результатов опытов; во сколько раз уравнение регрессии предсказывает результаты опытов лучше, чем среднее, видно из  $F$ -критерия:

$$F = \overline{S}_Y^2 / \overline{S}_{ост}^2 . \quad (4.6)$$

Искомое уравнение, при  $\alpha$  %-м уровне значимости, описывает результаты опытов лучше среднего  $\overline{Y}$  в  $F_{(v_1, v_2, P\%)}$  - раз (при условии  $F > F_{(v_1, v_2, P\%)}$ ) [Дрейпер, Смит, 1986].

Рассмотрим решение некоторых проблем, требующих использования регрессионного анализа:

а) по материалам 28-летних наблюдений за годовым стоком рр. Оресса – с. Андреевка ( $X$  – значение аргумента) и Птичь – с. Лучицы ( $Y$  – значение функции) необходимо построить уравнение регрессии (табл. 3.3).

1) По исходным данным определяются характеристики:

$$\overline{X} = 16,86, \text{ м}^3/\text{с}; \quad \overline{Y} = 45,31, \text{ м}^3/\text{с};$$

$$\sum_{i=1}^{28} (X_i - \overline{X}) \cdot (Y_i - \overline{Y}) = 2187,84; \quad \sum_{i=1}^{28} (X_i - \overline{X})^2 = 917,15;$$

$$\sum_{i=1}^{28} (Y_i - \overline{Y})^2 = 5737,61; \quad \overline{S}_Y^2 = 220,68; \quad \overline{S}_X^2 = 35,28 .$$

2) На основании полученных данных, по формулам (4.3), определяются

$$b_1 = 2,38; \quad b_0 = 5,18, \text{ при } r = 0,95 \pm 0,006 (0,89 - 1,00).$$

3) Следовательно, уравнение регрессии  $Y$  по  $X$  будет иметь вид –  $Y_i = 45,31 + 2,38 \cdot (X_i - 16,86)$  или  $Y_i = 5,18 + 2,38 \cdot X$ ,  $\text{м}^3/\text{с}$ , а уравнение регрессии  $X$  по  $Y$  вид –  $X_i = 0,38 \cdot Y_i - 0,42$ ,  $\text{м}^3/\text{с}$ .

4) Остаточная дисперсия определяется как  $\overline{S}_{ост}^2 = 19,14$ , а стандартные ошибки коэффициентов регрессии определяются по (4.4):  $\overline{S}_{b_0} = 0,84$ ;  $\overline{S}_{b_1} = 0,14$ .

5) Результаты расчетов параметров уравнения регрессии следующие:

$$b_0 \pm \bar{S}_{b_0} = 5,18 \pm 0,84 ; \quad b_1 \pm \bar{S}_{b_1} = 2,38 \pm 0,14 .$$

6) Стандартная ошибка ординаты уравнения регрессии определяется по (4.4). В частности,  $\bar{S}_{\hat{Y}(x)}^2 = 1,31, \text{ м}^3/\text{с}$  (при  $X_i=10$ );  $\bar{S}_{\hat{Y}(x)}^2 = 0,96, \text{ м}^3/\text{с}$  (при  $X_i=20$ );  $\bar{S}_{\hat{Y}(x)}^2 = 0,84 = \bar{S}_{b_0}^2$ , (при  $X_i = \bar{X} = 16,86, \text{ м}^3/\text{с}$ ) – как это и должно быть.

7) Значимость коэффициентов регрессии проверяется по  $t$ -критерию Стьюдента (4.4):  $t_{b_1} = 17,0$  ;  $t_{b_0} = 6,17$ .

Критическое значение  $t^T$  находится по специальным таблицам. При 5%-м уровне значимости и  $n=28$ , оно составляет  $t_{(28,5\%)}^T = 1,701$ . Следовательно, найденные значения коэффициентов регрессии статистически значимы.

8) Адекватность уравнения исходным данным проверяется по критерию Фишера (4.6) –  $F=11,53$ . Табличное значение  $F^T - F_{(27,26,5\%)} = 1,92$  – значительно меньше фактического, следовательно, полученное выше уравнение регрессии  $Y$  и  $X$  адекватно отражает исследуемую связь годовых расходов воды рр. Оресса – с. Андреевка ( $X$ ) и Птичь – с. Лучицы ( $Y$ );

#### 4.2. Нелинейная парная регрессия

*Нелинейная парная регрессия* сводится к получению заданной нелинейной зависимости  $Y(X)$ , приближающей совокупность чисел  $X_i$  и  $Y_i$  с наименьшей среднеквадратической погрешностью. Сведение нелинейной регрессии к линейной выполняется с помощью линеаризирующих преобразований  $Y_i$  и  $X_i$ . Используя метод наименьших квадратов и линеаризирующие преобразования, можно построить практически любые формы нелинейной парной связи [Мостеллер, Тьюки, 1982].

В табл. 4.1 приведены наиболее часто встречающиеся парные зависимости и линеаризирующие преобразования переменных. Качество предсказания результатов проверяется с помощью уравнения  $\hat{Y}'_i = b'_0 + b'_1 \cdot X'_i$ . После вычисления коэффициентов  $b'_0$  и  $b'_1$  по методу наименьших квадратов (как для парной линейной зависимости), выпол-

няются обратные преобразования [по  $b'_0$  и  $b'_1$  определяются  $b_0$  и  $b_1$ ], в соответствии с содержанием табл. 4.1. Очевидно, что парная зависимость может иметь разнообразную форму [Гельдерман, 1991].

Таблица 4.1

**Функции и линеаризующие преобразования**

N п/п	Функция	Линеаризующие преобразования			
		Преобразование переменных		Выражения для величин $b_0$ и $b_1$	
		$Y'$	$X'$	$b_0$	$b_1$
1	$Y=b_0+b_1 \cdot X$	$Y$	$X$	$b_0'$	$b_1'$
2	$Y=1/(b_0+b_1 \cdot X)$	$1/Y$	$X$	$b_0'$	$b_1'$
3	$Y=b_0+b_1/X$	$Y$	$1/X$	$b_0'$	$b_1'$
4	$Y=X/(b_0+b_1 \cdot X)$	$X/Y$	$X$	$b_0'$	$b_1'$
5	$Y=b_0 \cdot b_1^X$	$lgY$	$X$	$10^{b_0'}$	$10^{b_1'}$
6	$Y=b_0 + exp(b_1/X)$	$lnY$	$1/X$	$exp(b_0')$	$b_1'$
7	$Y=b_0 \cdot 10^{b_1 \cdot X}$	$lgY$	$X$	$10^{b_0'}$	$b_1'$
8	$Y=1/(b_0+b_1 \cdot exp(-X))$	$1/Y$	$exp(-X)$	$b_0'$	$b_1'$
9	$Y=b_0 \cdot X^{b_1}$	$lgY$	$lgX$	$10^{b_0'}$	$b_1'$
10	$Y=b_0+b_1 \cdot lgX$	$Y$	$lgX$	$b_0'$	$b_1'$
11	$Y=b_0+b_1 \cdot lnX$	$Y$	$lnX$	$b_0'$	$b_1'$
12	$Y=b_0/(b_1+X)$	$1/Y$	$X$	$1/b_1'$	$b_0' \cdot b_0$
13	$Y=b_0 \cdot X/(b_1+X)$	$1/Y$	$1/X$	$1/b_0'$	$b_0' \cdot b_1'$
14	$Y=b_0 \cdot exp(b_1/X)$	$lnY$	$1/X$	$exp(b_0')$	$b_1'$
15	$Y=b_0 \cdot 10^{b_1 X}$	$lgY$	$1/X$	$10^{b_0'}$	$b_1'$
16	$Y=b_0+b_1 \cdot X^n$	$Y$	$X^n$	$b_0'$	$b_1'$

До тех пор, пока не проверены все известные формы связи, исследователь не может быть уверен, что выбрана лучшая из них (с точки зрения точности обобщения результатов опытов); исключение – случай, когда «облако» точек имеет определенную и интерпретируемую форму. Кроме приведенных в табл. 4.1 уравнений, можно использовать квадратическую парную регрессию, которая имеет вид

$$Y_i = b_0 + b_1 \cdot X_i + b_2 \cdot X_i^2. \quad (4.7)$$

Коэффициенты регрессии квадратического уравнения  $b_0$ ,  $b_1$ ,  $b_2$  находятся аналогично с уравнением (4.7), при решении системы нормальных уравнений с тремя неизвестными:

$$\begin{cases} b_0 \cdot n + b_1 \cdot \sum_{i=1}^n X_i + b_2 \cdot \sum_{i=1}^n X_i^2 = \sum_{i=1}^n Y_i; \\ b_0 \cdot \sum_{i=1}^n X_i + b_1 \cdot \sum_{i=1}^n X_i^2 + b_2 \cdot \sum_{i=1}^n X_i^3 = \sum_{i=1}^n X_i Y_i; \\ b_0 \cdot \sum_{i=1}^n X_i^2 + b_1 \cdot \sum_{i=1}^n X_i^3 + b_2 \cdot \sum_{i=1}^n X_i^4 = \sum_{i=1}^n X_i^2 Y_i. \end{cases} \quad (4.8)$$

Нетрудно заметить, что аналогичная система уравнений может иметь место при получении уравнения парной зависимости любого порядка. Статистическая оценка коэффициентов регрессии и проверка значимости уравнения регрессии, в целом, осуществляются как для линейной регрессии.

Для примера, рассмотрим решение следующей проблемы:

а) по данным табл. 4.2 найти вид эмпирического уравнения регрессии, выражающего зависимость расхода воды реки от размера водосборной площади.

1) Из теоретических соображений, в качестве зависимости между площадью водосбора  $X$  и расходом воды реки  $Y$  по табл. 4.1 принимаем зависимость  $\hat{Y}_i = b_0 \cdot X_i^{b_1}$ , где  $b_0$  и  $b_1$  – некоторые постоянные, подлежащие определению.

2) Выполнив преобразование  $X' = \ln X$  и  $Y' = \ln Y$ ;  $b_0' = \ln b_0$  и  $b_1' = b_1$ , получим систему нормальных уравнений для определения параметров  $b_0$  и  $b_1$ :

$$\begin{cases} b_0' \cdot n + b_1' \cdot \sum_{i=1}^n X_i' = \sum_{i=1}^n Y_i'; \\ b_0' \cdot \sum_{i=1}^n X_i' + b_1' \cdot \sum_{i=1}^n X_i'^2 = \sum_{i=1}^n (Y_i' \cdot X_i'). \end{cases}$$

Неизвестные  $b_0'$  и  $b_1'$ , при этом, можно найти по формулам:



$$\left. \begin{aligned}
 b'_0 &= \frac{\sum_{i=1}^n (\ln Y_i) \cdot \sum_{i=1}^n (\ln X_i)^2 - \sum_{i=1}^n (\ln Y_i \cdot \ln X_i) \cdot \sum_{i=1}^n (\ln X_i)}{n \cdot \sum_{i=1}^n (\ln X_i^2) - (\sum_{i=1}^n (\ln X_i))^2}; \\
 b'_1 &= \frac{n \cdot \sum_{i=1}^n (\ln Y_i \cdot \ln X_i) - \sum_{i=1}^n (\ln X_i) \cdot \sum_{i=1}^n (\ln Y_i)}{n \cdot \sum_{i=1}^n (\ln X_i^2) - (\sum_{i=1}^n (\ln X_i))^2}.
 \end{aligned} \right\}$$

В табл. 4.2 показана последовательность определения констант  $b'_0$  и  $b'_1$  решаемого уравнения. Подставляя найденные (табл. 4.2) промежуточные значения в формулы, получаем параметры  $b'_0$  и  $b'_1$ :

$$\begin{cases}
 b'_0 = -1,3477; \\
 b'_1 = 0,4156.
 \end{cases}$$

Таблица 4.2

**Результаты расчет параметров  $b'_0$  и  $b'_1$** 

$N \text{ и } n$	$Y_i$	$X_i$	$\ln Y_i$	$\ln X_i$	$(\ln X_i)^2$	$\ln Y_i \ln X_i$	$\bar{Y}_i$	$\bar{Y}_i - Y_i$
1	0,65	50	-0,4308	3,9120	15,3039	-1,6852	1,320	0,670
2	1,15	25	0,1398	3,2189	10,3612	0,4499	0,990	-0,160
3	1,25	75	0,2231	4,3175	18,6407	0,9634	1,563	0,313
4	1,75	50	0,5596	3,9120	15,3039	2,1892	1,320	-0,430
5	2,25	100	0,8109	4,6052	21,2076	3,7345	1,761	-0,489
6	2,10	150	0,7419	5,0106	25,1065	3,7176	2,085	-0,015
7	2,55	150	0,9361	5,0106	25,1065	4,6904	2,085	-0,465
8	2,50	200	0,9163	5,2983	28,0722	4,8548	2,349	-0,151
9	3,05	250	1,1151	5,5215	30,4865	6,1572	2,578	-0,472
10	2,75	300	1,0116	5,7038	32,5331	5,7700	2,781	0,031
11	3,15	400	1,1474	5,9915	35,8977	6,8746	3,134	-0,016
12	3,60	450	1,2809	6,1092	37,3229	7,8255	3,291	-0,309
13	3,50	550	1,2528	6,3099	39,8151	7,9048	3,577	0,077
14	4,00	650	1,3863	6,4770	41,9512	8,9790	3,834	-0,166
15	3,75	700	1,3218	6,5511	42,9167	8,6589	3,954	0,204
16	3,85	800	1,3481	6,6846	44,6840	9,0113	4,180	0,330
17	4,00	900	1,3863	6,8024	46,2726	9,4301	4,390	0,390
18	4,35	1000	1,4702	6,9078	47,7171	10,1556	4,586	0,236
$\Sigma$	50,2	6800	16,6174	98,3439	558,6992	99,6817	-	-
	$\bar{y}=2,79$	$\bar{x}=378$						

3) Используя преобразование  $-1,3477 = \ln b_0$  или  $\exp(-1,3477)$ , т. е.  $b_0 = 0,2598$ , получим уравнение регрессии -

$$\hat{Y}_i = 0,2598 \cdot X_i^{0,4156}.$$

4) Подставляя в уравнение значения  $X$ , т. е. площадь водосбора ( $\text{км}^2$ ), находим расчетные величины расхода воды в реке  $Y_i$ ,  $\text{м}^3/\text{с}$  (табл. 4.2). Сравнение теоретических  $\hat{Y}_i$  и фактических  $Y_i$  расходов воды для одних и тех же водосборов свидетельствует о хорошем их согласии. Средняя квадратическая погрешность вычисления расхода воды в реке по теоретической зависимости (при  $n=18$ ;  $i=1, 2, \dots, n$ ) составляет

$$\delta = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (\hat{Y}_i - \bar{Y}_i)^2} = \pm 0,329, \text{ м}^3 / \text{с}.$$

Ошибка вычисления  $\delta = \pm 0,329$ ,  $\text{м}^3/\text{с}$  значительно меньше инструментальной погрешности измерения расхода воды в реке, что свидетельствует о хорошем согласии определяемого теоретически и замеряемого фактического расходов воды реки. Адекватность полученного уравнения проверим по критерию Фишера:  $F^2 = 12,02$ ; табличное значение  $F_{(17; 16,5\%)}^T = 2,398$  значительно меньше фактического, следовательно, искомое уравнение адекватно отражает исследуемую связь.

### 4.3. Линейная множественная регрессия

При решении задач оптимизации природопользования нередко количественное изменение изучаемого явления (*функции отклика*) зависит не от одной, а от нескольких причин (*факторов*), т. е. результатами наблюдений являются уже не два вектор - столбца ( $Y$ ) и ( $X$ ), как при проведении парного регрессионного анализа, а матрица результатов наблюдений:

$$\left\| \begin{array}{cccccc} Y_1 & X_{11} & X_{12} & \dots & X_{1j} & \dots & X_{1k} \\ Y_2 & X_{21} & X_{22} & \dots & X_{2j} & \dots & X_{2k} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ Y_i & X_{i1} & X_{i2} & \dots & X_{ij} & \dots & X_{ik} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ Y_n & X_{n1} & X_{n2} & \dots & X_{nj} & \dots & X_{nk} \end{array} \right\|, \quad (4.9)$$

где  $n$  – количество опытов;  $k$  – число факторов;  $X_{ij}$  – значение  $j$ -го фактора для  $i$ -го опыта;  $Y_i$  – значение функции отклика для  $i$ -го опыта. В задачу множественного регрессионного анализа входит построение уравнения плоскости (+1) – мерного пространства, от которого отклонения результатов наблюдений  $Y_i$  были бы минимальными. Следовательно, необходимо вычислить значения коэффициентов  $b_0, b_j$  в линейном полиноме

$$\hat{Y}_i = b_0 + \sum_{j=1}^k b_j \cdot X_{ij}, \quad (4.10)$$

что равносильно минимизации выражения

$$\sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n \left[ Y_i - (b_0 + \sum_{j=1}^k b_j \cdot X_{ij}) \right]^2, \quad (4.11)$$

где  $\hat{Y}_i$  – вычисленные (предсказываемые) выровненные значения исследуемой характеристики.

Для отыскания минимума можно использовать широко известный способ наименьших квадратов или способ детерминантов (определителей).

В этом случае общее выражение для коэффициентов регрессии  $b_j$  можно записать в форме

$$b_j = (-1)^{j+1} \cdot \frac{\sigma_y}{\sigma_{x_j}} \cdot \frac{D_{1j}}{D_{11}}, \quad (4.12)$$

где  $\sigma_y$  – среднее квадратическое отклонение зависимой переменной (функции);  $\sigma_{x_j}$  – среднее квадратическое отклонение независимой переменной;  $D_{1j}$  и  $D_{11}$  – миноры определителя  $D$ .

В результате проведенных операций полином первой степени (4.10) получается с известными коэффициентами  $b_0, b_j$ . Проверка значимости (качества предсказания) множественного уравнения регрессии, в принципе, мало отличается от соответствующей проверки парной зависимости. Остаточная дисперсия вычисляется по формуле

$$\bar{S}_{ост}^2 = \frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n - k - 1}, \quad (4.13)$$

и, затем, сравнивается с дисперсией среднего  $\left(\bar{S}_y^2\right)$  с помощью  $F$ -критерия Фишера, с числом степеней свободы в числителе  $\nu_1=(n-1)$ , в

знаменателе –  $v_2=(n-k-1)$ . Считается, что по уравнению (4.10) предсказываются результаты опытов лучше среднего, если  $F$  превышает или достигает границы значимости при выбранном ее уровне: обычно  $P=(1-q)=5\%$ .

Значимость коэффициентов регрессии  $b_0, b_j$  проверяется по  $t$ -критерию Стьюдента

$$t = |b_j| / \bar{S}_{b_j}, \quad (4.14)$$

в котором  $\bar{S}_{b_j}$  – погрешности коэффициентов регрессии, определяемые как

$$\bar{S}_{b_j} = \frac{\sigma_Y}{\sigma_{X_j}} \cdot \sqrt{\frac{1-R^2}{n-k} \cdot \frac{\Delta_{jj}}{D_{11}}}, \quad (4.15)$$

где  $\Delta_{jj}$  – минор определителя  $D$ , который получается из  $D$  вычеркиванием  $(j+1)$  строки столбца. Вычисленное значение  $t$  сравнивается с  $t^T$ , при числе степеней свободы  $v=(n-k-1)$ . Доверительный интервал для коэффициентов регрессии устанавливается по соотношению

$$b_j - t^T \cdot \bar{S}_{b_j} \leq \beta_j \leq b_j + t^T \cdot \bar{S}_{b_j}, \quad (4.16)$$

где  $\beta_j$  – соответствующее значение для коэффициентов регрессии в генеральной совокупности.

На примере приведения к многолетнему периоду величин годового стока р. Птичь – с. Лучицы, рассмотрим порядок применения линейной множественной регрессии.

Для этого необходимо построить уравнение множественной регрессии зависимости расхода р. Птичь – с. Лучицы от расходов рек – аналогов. В качестве аналогов, как и в других примерах, принимаем рр. Ясельда – с. Сенин, Оресса – с. Верхутино, Оресса – с. Андреевка. Парные коэффициенты корреляции приведены ранее.

1. Определитель  $D$  и его миноры равны:

$$D = 0,010; D_{yy} = 0,1676; D_{yx_1} = 0,0186; D_{yx_2} = -0,0006; D_{yx_3} = 0,1441.$$

2. Подсчет средних значений и средних квадратических отклонений исходных рядов дает следующие оценки:

$$\bar{Y} = 45,31, \quad \bar{X}_1 = 18,55, \quad \bar{X}_2 = 2,74, \quad \bar{X}_3 = 16,86$$

$$\sigma_y = 14,31, \quad \sigma_{x_1} = 8,14, \quad \sigma_{x_2} = 0,83, \quad \sigma_{x_3} = 0,1441.$$

3. Коэффициенты регрессии уравнения связи определяются по формуле (4.12):  $b_1 \approx 0,1951$ ;  $b_2 = 0,0668$ ;  $b_3 = 2,1511$ .

4. Уравнение регрессии, согласно исходным данным, получает вид  $Y - \bar{Y} = b_1 \cdot (X_1 - \bar{X}) + b_2 \cdot (X_2 - \bar{X}) + b_3 \cdot (X_3 - \bar{X})$ , или

$$Y - 45,31 = 0,1951 \cdot (X_1 - 18,55) + 0,0668 \cdot (X_2 - 2,74) + 2,1511 \cdot (X_3 - 16,86),$$

т. е.  $Y = 0,1951 \cdot X_1 + 0,0668 \cdot X_2 + 2,1511 \cdot X_3 + 5,24$ .

5. Погрешность полученной связи распределяется как

$$\sigma_Y = \sigma_Y \cdot \sqrt{1 - R^2} = 3,54.$$

Погрешности коэффициентов регрессии  $b_1, b_2, b_3$ , применительно к случаю четырех переменных, согласно формуле (4.15), будут равны:

$$\bar{S}_{b_1} = 0,1589; \bar{S}_{b_2} = 1,1758; \bar{S}_{b_3} = 0,2295.$$

6. Доверительный интервал для коэффициентов регрессии определяется по формуле (4.16) при значении  $t^T = 1,711$ , числе степеней свободы  $v = 28 - 3 - 1 = 24$ :  $-0,0768 \leq \beta_1 \leq 0,4670$ ;  $-1,9450 \leq \beta_2 \leq 2,0786$ ;  $1,7584 \leq \beta_3 \leq 2,5438$ .

7. Значимость коэффициентов регрессии  $b_1, b_2, b_3$  определяется с использованием критерия Стьюдента по формуле (4.14):

$$t_1 = 1,23 < t^T = 1,711; t_2 = 0,06 < t^T = 1,711; t_3 = 9,37 > t^T = 1,711.$$

Судя по критерию Стьюдента, коэффициенты регрессии  $b_1$  и  $b_2$  статистически незначимы, следовательно, они могут быть исключены из полученного уравнения (имеет место достаточно высокая корреляция между расходами рек – аналогов).

8. Для проверки качества предсказания результатов по полученному уравнению, по формуле (4.12) вычисляется остаточная дисперсия

$\bar{S}_{ост}^2 = 95,24$ , которая сравнивается с дисперсией среднего  $\bar{S}_Y = 204,78$  и с помощью  $F^m$ -критерия при числе степеней свободы  $(v_1 = 28 - 1 = 27)$ ,  $(v_2 = 28 - 3 - 1 = 24)$  и 5 %-ом уровне значимости, устанавливается соотношение  $F$  и  $F^m$

$$F = \frac{\bar{S}_Y}{\bar{S}_{ост}^2} = 2,15 > F^T = 1,98.$$

Полученное уравнение предсказывает годовые расходы воды р. Птичь – с. Лучицы в 2,15-раза лучше, чем их средняя величина.

В тех случаях, когда методами линейной множественной регрессии не удастся получить приемлемую математическую модель, прибегают к моделям нелинейной множественной регрессии.

#### 4.4. Нелинейная множественная регрессия

*Первый этап* нелинейного множественного анализа – получение так называемой квадратичной формы или модели второго порядка, имеющей вид:

$$Y_j = \beta_0 + \beta_1 \cdot X_{1i} + \beta_2 \cdot X_{2i} + \dots + \beta_j \cdot X_{ji} + \dots + \beta_k \cdot X_{ki} + \beta_{22} \cdot X_{2i}^2 + \dots + \beta_{jj} \cdot X_{ji}^2 + \dots + \beta_{kk} \cdot X_{ki}^2 + \beta_{12} \cdot X_{1i} \cdot X_{2i} + \beta_{13} \cdot X_{1i} \cdot X_{3i} + \dots + \epsilon_i. \quad (4.17)$$

Аналогичным способом можно получить и модель более высоких порядков, повышая степень уравнения до тех пор, пока уменьшается остаточная дисперсия  $\bar{S}_{ост}^2$ . Для определения оценок коэффициентов регрессии  $b_0, b_j, b_{jm}, b_{jj}$ , модель типа (4.17) приводится к линейному виду путем замены переменных:  $X'_{1i} = X_{1i}$ ;  $X'_{2i} = X_{2i}$ ; ...  $X'_{k+1,i} = X_{1i}^2$ ;  $X'_{k+2,i} = X_{2i}^2$ ; ... В случае неадекватности модели второго порядка, используется модель третьего порядка.

Рассмотрим основные приемы, используемые в практике нелинейного регрессионного анализа:

а) мультипликативная модель –

$$Y_i = \beta_0 \cdot X_{1i}^{\beta_1} \cdot X_{2i}^{\beta_2} \cdot \dots \cdot X_{ji}^{\beta_j} \cdot \dots \cdot X_{ki}^{\beta_k} \cdot \epsilon_i, \quad (4.18)$$

где  $\beta_0, \beta_1, \beta_2, \dots, \beta_j, \dots, \beta_k$  – неизвестные параметры;  $\epsilon_i$  – мультипликативная случайная ошибка. Логарифмирование уравнения (4.18) по основанию позволяет перевести модель в линейную форму

$$\ln Y_i = \ln \beta_0 + \beta_1 \cdot \ln X_{1i} + \beta_2 \cdot \ln X_{2i} + \dots + \beta_j \cdot \ln X_{ji} + \dots + \beta_k \cdot \ln X_{ki} + \ln \epsilon_i. \quad (4.19)$$

Далее, производится замена переменных:  $Y'_i = \ln Y_i$ ;  $\beta'_0 = \ln \beta_0$ ;  $X'_{1i} = \ln X_{1i}$ ;  $X'_{2i} = \ln X_{2i}$ , и т. д. Преобразованная модель (4.18), с заменой переменных, имеет форму уравнения (4.10), и поэтому возможно применение стандартных методов исследования линейной регрессии. В конце расчетов производится только одно обратное преобразование для получения величины  $\beta_0$ ;

б) «обратная» модель –



$$Y_i = \frac{1}{\beta_0 + \beta_1 \cdot X_{1i} + \beta_2 \cdot X_{2i} + \dots + \beta_j \cdot X_{ji} + \dots + \beta_k \cdot X_{ki} + \varepsilon_i} \quad (4.20)$$

Обращая обе части, получим

$$\frac{1}{Y_i} = \beta_0 + \beta_1 \cdot X_{1i} + \beta_2 \cdot X_{2i} + \dots + \beta_j \cdot X_{ji} + \dots + \beta_k \cdot X_{ki} + \varepsilon_i; \quad (4.21)$$

в) экспоненциальная модель –

$$Y_i = \exp(\beta_0 + \beta_1 \cdot X_{1i} + \beta_2 \cdot X_{2i} + \dots + \beta_j \cdot X_{ji} + \dots + \beta_k \cdot X_{ki}) \cdot \varepsilon_i; \quad (4.22)$$

логарифмируя обе части по натуральному основанию, получим

$$\ln Y_i = \beta_0 + \beta_1 \cdot X_{1i} + \beta_2 \cdot X_{2i} + \dots + \beta_j \cdot X_{ji} + \dots + \beta_k \cdot X_{ki} + \ln \varepsilon_i; \quad (4.23)$$

г) комбинированная модель –

$$Y_i = \frac{1}{1 + \exp(\beta_0 + \beta_1 \cdot X_{1i} + \beta_2 \cdot X_{2i} + \dots + \beta_j \cdot X_{ji} + \dots + \beta_k \cdot X_{ki} + \varepsilon_i)} \quad (4.24)$$

применяя обращение и вычитая единицу, затем, логарифмируя по натуральному основанию обе части уравнения, получаем –

$$\ln\left(\frac{1}{Y_i} - 1\right) = \beta_0 + \beta_1 \cdot X_{1i} + \beta_2 \cdot X_{2i} + \dots + \beta_j \cdot X_{ji} + \dots + \beta_k \cdot X_{ki} + \varepsilon_i. \quad (4.25)$$

Такое последовательное преобразование зависимой переменной используется для приведения сложной нелинейной модели к линейному виду. Более простые модели преобразований, приводятся ниже;

д) «обратное» преобразование –

$$Y_i = \beta_0 + \beta_1 \cdot \frac{1}{X_{1i}} + \beta_2 \cdot \frac{1}{X_{2i}} + \dots + \beta_j \cdot \frac{1}{X_{ji}} + \dots + \beta_k \cdot \frac{1}{X_{ki}} + \varepsilon_i; \quad (4.26)$$

е) логарифмическое преобразование –

$$Y_i = \beta_0 + \beta_1 \cdot \ln X_{1i} + \beta_2 \cdot \ln X_{2i} + \dots + \beta_j \cdot \ln X_{ji} + \dots + \beta_k \cdot \ln X_{ki} + \varepsilon_i; \quad (4.27)$$

ж) преобразование типа квадратного корня –

$$Y_i = \beta_0 + \beta_1 \cdot \sqrt{X_{1i}} + \beta_2 \cdot \sqrt{X_{2i}} + \dots + \beta_j \cdot \sqrt{X_{ji}} + \dots + \beta_k \cdot \sqrt{X_{ki}} + \varepsilon_i. \quad (4.28)$$

Мерой тесноты связи в нелинейной зависимости служит множественное корреляционное отношение, которое вычисляется по формуле (4.24), при использовании для вычисления  $\hat{Y}_i$  нелинейной формы уравнения. Отметим, что в одной и той же модели может содержаться несколько различных преобразований и поэтому выбор должен осуществляться на основе знаний о входящих в модель факторах. Цель преобразований такого рода состоит в том, чтобы получить для преобразованных

переменных более простую регрессионную модель, чем для исходных. Процедура множественного регрессионного анализа громоздка, трудоемка и изначально ориентируется на использование ЭВМ.

#### 4.5. Основные аспекты выбора оптимальной регрессионной модели

Теоретически – точность аппроксимации можно повысить одновременно с повышением степени полинома, однако, практически – для полиномов высоких степеней, при проведении матричных операций на ЭВМ накапливаются столь значительные погрешности округления, что решение становится невозможным. Поэтому, на практике обычно ограничиваются построением полинома второго порядка и проведением шагового регрессионного анализа, с включением или исключением переменных. Для выбора оптимальной формы регрессии было бы лучше всего вычислить коэффициенты всех возможных уравнений регрессии, а затем выбрать «лучшее» уравнение по минимальной остаточной дисперсии или значению коэффициента корреляции. Однако такое решение задачи не всегда реально. Например, общее число уравнений при 10 определяющих факторах составляет  $2^{10}=1024$ . Если сюда добавить и нелинейные формы уравнений, то чисто технически решение такой задачи проблематично. Возникают трудности, также, при выборке из нескольких уравнений с одинаковыми статистически значимыми критериями, когда необходимо принять во внимание следующие моменты: 1) могут быть основательные теоретические причины, которые заставляют нас поверить в то, что регрессионное уравнение имеет определенную математическую форму; 2) регрессионное уравнение должно обеспечить наибольшее приближение к наблюдаемым значениям (тогда уравнение может быть с достаточной уверенностью использовано в прогнозных целях); 3) регрессионное уравнение должно быть как можно более простым, но при этом адекватно описывать исследуемые явления.

При формальных методах отбора факторов для сравнения их важности имеются пять показателей: 1) коэффициенты парной корреляции между функцией отклика и изучаемым фактором; 2) частные коэффициенты корреляции; 3)  $t$ -критерии Стьюдента; 4) остаточные дисперсии  $\bar{S}_{ост.}^2$ ; 5) частные  $F$ -критерии Фишера.

Решающее значение при выборе того или иного фактора имеют природа явления или даже интуитивные представления о важности этих факторов.

Алгоритм множественного регрессионного анализа методом включения факторов (переменных) следующий:

- 1) вычислить частные коэффициенты корреляции;
- 2) ранжировать факторы по значению коэффициентов корреляции

$$r_{yx_j} > r_{yx_{j+1}};$$

3) построить линейную модель с одним фактором, у которого частный коэффициент корреляции наибольший;

- 4) вычислить  $R$ ,  $R^2$  и  $\bar{S}_{ост}^2$ ;

5) добавить следующий фактор, у которого коэффициент корреляции наибольший из оставшихся и построить регрессионную модель типа (4.10);

6) вычислить критерии для оценки качества предсказания результатов опыта на базе модели  $R$ ,  $R^2$  и  $\bar{S}_{ост}^2$ ;

- 7) вычислить выражения –

$$\left| \frac{S_{j+1}}{\sigma_y} - \frac{S_j}{\sigma_y} \right| > \varepsilon \text{ и } |R_{j+1} - R_j| > \beta; \quad (4.29)$$

в которых  $S_j$  и  $S_{j+1}$  – относительные ошибки при использовании  $j$  и  $j+1$  – факторов;  $\sigma_y$  – среднеквадратическое отклонение исследуемого ряда;  $R_j$  и  $R_{j+1}$  – соответственно, коэффициенты множественной корреляции;  $\varepsilon$  и  $\beta$  – заданные пороги чувствительности модели;

8) если условия (4.27) выполняются, то привлекается очередной фактор и расчет продолжается с пункта 5, а если не выполняются и первые из  $j$  – факторов значимы, модель считается построенной.

*Алгоритм множественного регрессионного анализа методом исключения факторов* (переменных) основан на принципах, изложенных выше (для метода включения переменных):

1) уравнение регрессии расширяется сразу до полной квадратичной, при возможности, – до полной кубической формы;

2) исключение начинается с фактора, имеющего наименьший  $t$ -критерий Стьюдента;

3) после исключения очередного фактора, для нового уравнения регрессии, вычисляются: множественный коэффициент корреляции, остаточная дисперсия, и т.д.;

4) для прекращения исключения факторов используются следующие подходы:

- исключение прекращается, начиная с момента увеличения остаточной дисперсии;
- то же, исходя из условия назначения уровня значимости ( $\alpha, \%$ ) при вычислении  $t$ -критерия Стьюдента для последнего оставляемого фактора.

Рассмотрим практические аспекты оптимизации регрессионной модели на примере выявления теплоэнергетических ресурсов процесса испарения.

Теплоэнергетические ресурсы процесса испарения для любого расчетного промежутка времени определяется как:

$$LE_{mi} = R_i^+ + P_i^+ \pm \Delta B_i - \Delta E_{0i}, \quad (4.30)$$

где  $E_{mi}$  – эквивалент теплоэнергетических ресурсов – максимально возможное испарение, м;  $L$  – скрытая теплота испарения, Дж/м<sup>3</sup>;  $R_i^+$  – положительная составляющая радиационного баланса, Дж/м<sup>2</sup>;  $P_i^+$  – положительная составляющая турбулентного теплообмена, Дж/м<sup>2</sup>;  $\Delta B_i$  – изменение теплозапасов деятельного слоя почвы, Дж/м<sup>2</sup>;  $\Delta_i$  – расход тепла на таяние снега, льда, прогревание почвы, Дж/м<sup>2</sup>.

Из-за ограниченности материалов наблюдений за радиационным режимом, турбулентным и почвенным теплообменами использовать уравнение (4.30) в практических расчетах затруднительно. Исследования З. И. Пивоваровой [Пивоварова, 1966], показали, что оптимальное расстояние между актинометрическими станциями должно быть не более 100 км. При существующей сети актинометрических станций данное условие не выполняется. В связи с этим есть необходимость поиска связи элементов радиационного баланса с широко измеряемыми климатическими характеристиками.

В настоящее время имеется ряд зависимостей для определения годовых норм  $E_0$  в условиях Беларуси [Голченко, 1976; Марчук, 1982] по сумме температур воздуха выше 10°C. Нами же найдено приближенное

решение уравнения (4.30) для месячных интервалов времени [Волчек, 1986]. На основе анализа имеющихся режимных данных наблюдений на актинометрических станциях СССР [Основные..., 1986] установлены количественные связи месячных величин положительной составляющей радиационного баланса с дефицитом влажности воздуха. Примечательно, что для всех географических зон форма связи в общих чертах одинакова. Причины нелинейности и гистерезиса этой связи состоит в том, что положительная составляющая радиационного баланса является одной из составляющих уравнения (4.30), куда входят еще затраты тепла на испарение, турбулентный теплообмен с атмосферой и изменение теплосодержания почвогрунтов. Дефицит влажности воздуха характеризует лишь интенсивность турбулентного обмена с атмосферой. В соответствии со знаком теплообмена в почвогрунтах и интенсивностью затрат тепла на испарение изменения дефицита влажности воздуха в первом полугодии происходит по выпуклой кривой, а во втором – по прямой. Эффект гистерезиса увеличивается с ростом континентальности, т. е. при движении с севера на юг и с запада на восток. Так как при этом происходит возрастание засушливости климата, то отмеченная особенность связи может быть полезным инструментом косвенного отражения степени засушливости.

Полученные зависимости откорректированы по имеющимся данным наблюдений за реальные месяцы для условий Беларуси [Справочник..., 1971] (рис. 4.1).

Найдена аналитическая аппроксимация восходящей и нисходящей ветвей петли гистерезиса.

Восходящая ветвь (период роста дефицитов влажности воздуха)

$$R_{Mi}^+ = 2,26 + 6,77 \cdot \lg d_{Mi} \quad \text{при } d_{Mi} \leq d_{Mi+1}. \quad (4.31)$$

Нисходящая ветвь (период спада дефицитов влажности воздуха)

$$R_{Mi}^+ = 2,26 + 6,77 \cdot \lg d_{Mi} \quad \text{при } d_{Mi} > d_{Mi+1}, \quad (4.32)$$

где  $d_{Mi}$  – среднемесячные величины дефицита влажности воздуха, *мм*.

Зависимости (4.31) и (4.32) (рис. 4.1) характеризуются коэффициентами корреляции соответственно  $r=0,965 \pm 0,006$ ;  $r=0,945 \pm 0,010$  и значениями критерия Фишера соответственно  $F=14,31$ ;  $F=9,45$ , что значительно больше допустимого при 1% уровне значимости  $F_{(107,106,1\%)}^T = 1,643$ .

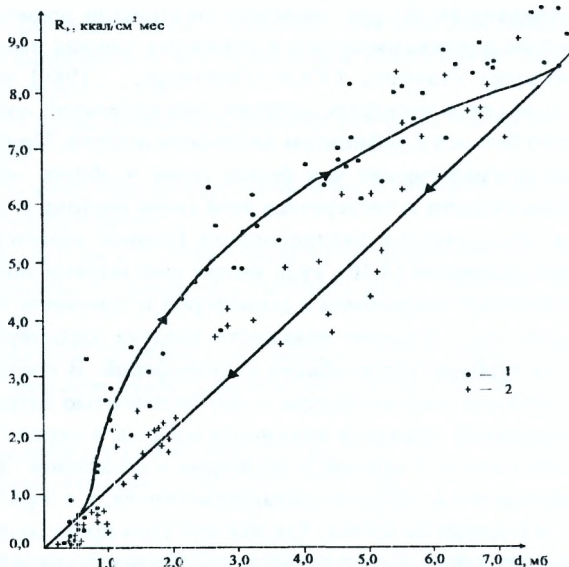


Рис. 4.1. Зависимости дефицитов влажности воздуха от положительной составляющей радиационного баланса. Период роста (1) и период спада (2) дефицитов влажности воздуха

Следовательно, полученные уравнения статистически значимы. Без учета явления гистерезиса осредненная связь имеет вид:

$$R_{Mi}^+ = 1,79 + 6,59 \cdot \lg d_{Mi} , \quad (4.33)$$

$$r=0,937 \pm 0,007; F=8,29 > F_{(215,214,1\%)}^T = 1,533$$

При выполнении расчетов в  $\text{мДж/м}^2$  в зависимости вводится переводной коэффициент  $K=41,9$ .

Для определения годовых величин положительной составляющей, кроме месячных величин дефицита влажности воздуха можно использовать сумму температур воздуха выше  $10^\circ\text{C}$ . Так, для Восточной Европы может быть использована следующая зависимость:

$$E_{mk} = 195,5 \cdot \frac{\sum t_{10^\circ\text{C}}}{1000} + 395,4 . \quad (4.34)$$



С учетом затрат тепла на таяние снега и мерзлоты зависимость (2.44) приведена к виду:

$$E_{mk} = 195,5 \cdot \frac{\sum t_{10^{\circ}C}}{1000} - 3555,3 - 1,67 \cdot \frac{\sum t_{<0^{\circ}C}}{h_{CH} + 20}, \quad (4.35)$$

где  $h_{CH}$  – высота слоя снежного покрова к началу его интенсивного таяния, см.

В справочной литературе данные о суммах среднесуточных температур воздуха выше  $10^{\circ}C$  имеются не по всем метеорологическим пунктам, а ежегодные сведения о  $\sum t > 10^{\circ}C$  не публикуются. Однако имеются сведения о средних многолетних и ежегодных среднемесячных температурах воздуха, по которым можно подсчитать их суммы за теплый период, т. е. за IV – X месяцы. Между термическими характеристиками обнаруживается тесная связь

$$t_{>0^{\circ}C} = 32,80 \cdot t_{IV-X} - 585,82. \quad (4.36)$$

$$(r=0,990 \pm 0,003)$$

Тогда с учетом зависимости (2.46),  $E_m$  определяется как:

$$E_{mk} = 6,4 \cdot t_{IV-X} + 281. \quad (4.37)$$

Для определения  $P_{\Gamma}^{+}$  рекомендуется следующая зависимость:

$$P_{\Gamma}^{+} = 6,8 - 0,082 \cdot R_{\Gamma}, \quad (4.38)$$

где  $R_{\Gamma}$  – сконцентрированный радиационный баланс, ккал/см<sup>2</sup>.

Положительный радиационный баланс  $R^{+}$  для территории СНГ может быть достаточно точно рассчитанным по его связи с  $R_{\Gamma}$ , используя зависимость:

$$R_{\Gamma}^{+} = 1,17 \cdot R_{\Gamma} + 2,10, \quad (r=0,995 \pm 0,002) \quad (4.39)$$

Для реальных месяцев предлагается аналогичная зависимость:

$$R_{mi}^{+} = 1,0 \cdot R_M + 0,77, \quad (4.40)$$

$$(r=0,985 \pm 0,002; F=45,67 > F_{(215,214,1\%)}^T = 1,533).$$

С учетом формулы (4.39), зависимость (4.38) примет вид:

$$P_{mi}^{+} = \left( 6,65 - 0,071 \cdot \sum_1^{12} R_{mi}^{+} \right) \cdot K_i, \quad (4.41)$$

где  $K_i$  – коэффициенты, зависящие от месяцев.

## 5. МЕТОДЫ ОПРЕДЕЛЕНИЯ РАЗЛИЧНЫХ ФАКТОРОВ И УСЛОВИЙ ВЛИЯНИЯ НА ПРОЦЕССЫ И ЯВЛЕНИЯ

В природопользовании, при постановке прикладных комплексных исследований, экспериментальные данные часто требуется разбить на группы, отличающиеся между собой в количественном отношении, и установить сходство (различие) между ними. Например, определить степень влияния географических условий на ход тех или иных природных процессов и явлений. Лучшие всего этим требованиям отвечает дисперсионный анализ, который широко используется при решении именно практических задач. Дисперсионный анализ позволяет установить с определенной долей уверенности влияние на изучаемый объект каждого из исследуемых факторов в отдельности или в определенных их сочетаниях, т. е. реально ли влияние отдельного фактора или этот эффект можно считать результатом случайных ошибок, а также получены ли выборочные средние из одной и той же гипотетической генеральной совокупности со средним значением  $\bar{x}$ . Необходимым условием использования дисперсионного анализа является разбивка каждого учитываемого фактора не менее чем на две группы. Если влияние факторов нельзя выразить количественными показателями, то они могут быть представлены качественными показателями, выраженными в виде баллов. Дисперсионный анализ базируется на законе распределения отношения средних квадратов (дисперсий):

$$S_1^2 / S_2^2 = F , \quad (5.1)$$

где  $S_1^2$  – среднеквадратическая ошибка средних отдельного опыта;  $S_2^2$  – суммарная среднеквадратическая ошибка средних всех опытов. Использование методов дисперсионного анализа позволяет дать ответ на следующие вопросы: 1) значимо ли влияет изучаемый фактор на воспроизводимость и в целом на результат? 2) если установлено значимое влияние какого-либо фактора на результат в целом, то начиная с какого уровня фактора это влияние действует; значимо ли различаются выборочные средние между собой? 3) какой количественной мерой можно оценить степень установленного влияния? Сущностью дисперсионного анализа является расчленение общей суммы квадратов отклонений и общего числа степеней свободы на части – компоненты, соответствующие структуре

эксперимента, и оценка значимости действия и взаимодействия изучаемых факторов по F-критерию. С этой целью дисперсия разделяется на независимые слагаемые, которые затем сравниваются между собой. Допустим, в результате измерения величины  $M$  получено значение  $X$  и пусть на процесс измерения влияют случайные независимые факторы  $A$  и  $B$ . Тогда отклонение  $M-X = \alpha + \beta + \gamma$ , где  $\alpha$  – отклонение под влиянием фактора  $A$ ,  $\beta$  – то же под влиянием фактора  $B$ , а  $\gamma$  – под влиянием остальных, неучтенных факторов, причем,  $\alpha$ ,  $\beta$  и  $\gamma$  – независимы. В этом случае, дисперсия  $D(M-X) = D(\alpha + \beta + \gamma)$ , а  $DX = (D\alpha + D\beta + D\gamma)$ , где  $D\alpha$  – характеризует влияние фактора  $A$ ,  $D\beta$  – влияние фактора  $B$ ,  $D\gamma$  – влияние остальных, неучтенных факторов. Дисперсия  $D\gamma$  называется *остаточной дисперсией*. Для оценки значимости факторов  $A$  и  $B$  сравниваются соответствующие дисперсии  $D\alpha$  и  $D\beta$  с остаточной дисперсией  $D\gamma$ . Если исследуется влияние одного фактора, то говорят об однофакторном дисперсионном анализе, при исследовании влияния двух факторов – о двухфакторном дисперсионном анализе и т.п.

Рассмотрим простейшие способы применения дисперсионного анализа.

### 5.1. Однофакторный дисперсионный анализ

При однофакторном дисперсионном анализе обычно изучается действие одного фактора на  $m$  – уровнях  $k > 2$ , при равном числе определений (измерений) на каждом уровне  $n$ . Пусть фактор имеет  $m$  – уровней. Из каждого уровня делается выборка из  $n$  – элементов. Общее количество выбранных элементов обозначается как  $N = m \cdot n$ . Вся выборка представляет собой матрицу:

$$X = \begin{pmatrix} X_{11} & X_{12} & \dots & X_{1m} \\ X_{21} & X_{22} & \dots & X_{2m} \\ \dots & \dots & \dots & \dots \\ X_{n1} & X_{n2} & \dots & X_{nm} \end{pmatrix}. \quad (5.2)$$

Полагая, что данная выборка сделана из нормально распределенной генеральной совокупности, и задавая уровень значимости  $\alpha$ , необходимо проверить гипотезу  $H_0$  о равенстве средних значений на всех уровнях фактора  $-\mu_1 = \mu_2 = \dots = \mu_m$ . При альтернативной гипотезе  $H_1$  не все сред-

ние значения  $\mu_1$  должны быть равными. В качестве статистики используется величина  $F$ , определяемая по аналогичной с (5.1) зависимости:

$$F = S_1^2 / S_2^2, \quad (5.3)$$

в которой  $S_1^2$  – дисперсия, характеризующая влияние исследуемого фактора (факторная дисперсия);  $S_2^2$  – дисперсия, характеризующая влияние остальных факторов (остаточная дисперсия). Если гипотеза  $H_0$  верна, то случайная величина  $F$  имеет  $F$ -распределение со степенями свободы  $m-1$  и  $N-m=m \cdot (n-1)$ . При проверке гипотезы  $H_0$  используется правосторонняя критическая область, исходя из условия:

$$P(F > f_\alpha) = \alpha. \quad (5.4)$$

Если значение статистики входит в критическую область, то гипотеза  $H_0$  о равенстве средних значений на всех уровнях фактора отвергается, т. е. считается значимым влияние исследуемого фактора. В противном случае принимается гипотеза  $H_0$ , т. е. считается, что значимость влияния фактора не установлена. При определении  $F$  находится сумма квадратов отклонений элементов выборки относительно общего среднего арифметического:

$$Q = \sum_{j=1}^m \cdot \sum_{i=1}^n (X_{ij} - \bar{X})^2, \quad (5.5)$$

где  $\bar{X} = \frac{1}{N} \cdot \sum_{j=1}^m \cdot \sum_{i=1}^n X_{ij}$ , которая, в свою очередь, может быть разделена на

два независимых слагаемых –

$$Q_1 = n \cdot \sum_{j=1}^m (\bar{X}_j - \bar{X})^2, \quad (5.6)$$

$$Q_2 = \sum_{j=1}^m \cdot \sum_{i=1}^n (X_{ij} - \bar{X}_j)^2. \quad (5.7)$$

так, чтобы выполнялось равенство  $Q = Q_1 + Q_2$ ; здесь  $\bar{X}_j = \frac{1}{n} \cdot \sum_{i=1}^n X_{ij}$ , при

$j=1, 2, \dots, m$  – групповое среднее. Суммой квадратов межгрупповых отклонений, характеризующих влияние исследуемого фактора, является сумма квадратов отклонений групповых средних относительно общей средней (сумма  $Q_1$ ). С другой стороны,  $Q_2$  представляет собой сумму

квадратов отклонений значений выборки относительно групповых средних, так называемую сумму квадратов внутри групповых отклонений. Эта сумма характеризует влияние остальных, неучтенных факторов. Имея суммы  $Q, Q_1, Q_2$ , можно вычислить соответствующие дисперсии:

$$S^2 = \frac{Q}{m \cdot n - 1}; \quad S_1^2 = \frac{Q_1}{m - 1}; \quad S_2^2 = \frac{Q_2}{m \cdot (n - 1)}. \quad (5.8)$$

Две последние дисперсии используются при вычислении  $F$ . При практических вычислениях, величины  $Q$  и  $Q_1$  находятся обычно по выборке, а  $Q_2$  — определяется как разность:

$$Q_2 = Q - Q_1. \quad (5.9)$$

В простейших случаях, на каждом уровне фактора, выбирается одинаковое количество объектов исследований, но нахождение сумм по (5.5) — (5.7) достаточно сложно, и более приемлемые формулы можно получить, преобразуя выражения (5.5) и (5.6), когда:

$$Q = \sum_{j=1}^m \left( \sum_{i=1}^n X_{ij}^2 \right) - \frac{1}{m \cdot n} \cdot \left( \sum_{j=1}^m \left( \sum_{i=1}^n X_{ij} \right) \right)^2; \quad (5.10)$$

$$Q_1 = \frac{1}{n} \cdot \sum_{j=1}^m \left( \sum_{i=1}^n X_{ij} \right)^2 - \frac{1}{m \cdot n} \cdot \left( \sum_{j=1}^m \left( \sum_{i=1}^n X_{ij} \right) \right)^2. \quad (5.11)$$

Тогда, обозначив  $R_j = \sum_{i=1}^n X_{ij}^2$  и  $L_j = \sum_{i=1}^n X_{ij}$ , имеем:

$$Q = \sum_{j=1}^m R_j - \frac{1}{m \cdot n} \cdot \left( \sum_{j=1}^m L_j \right)^2; \quad (5.12)$$

$$Q_1 = \frac{1}{n} \cdot \sum_{j=1}^m L_j^2 - \frac{1}{m \cdot n} \cdot \left( \sum_{j=1}^m L_j \right)^2. \quad (5.13)$$

По исходной матрице (выборке) вычисляются суммы элементов и их квадратов по столбцам  $L_j$  и  $R_j$ , при  $j=1, 2, \dots, m$ . Далее производится замена переменных  $Y_{ij} = X_{ij} - C$ . Целесообразно их выбрать близкими к общему среднему. В результате замены (для  $Y_{ij}$ ) получаются следующие формулы:

$$Q = \sum_{j=1}^m P_j - \frac{1}{N} \cdot \left( \sum_{j=1}^m T_j \right)^2; \quad (5.14)$$

$$Q_1 = \frac{1}{n} \cdot \sum_{j=1}^n T_j^2 - \frac{1}{N} \cdot \left( \sum_{j=1}^m T_j \right)^2, \quad (5.15)$$

где  $N=m \cdot n$ ;  $P_j = \sum_{i=1}^n Y_{ij}^2$  и  $T_j = \sum_{i=1}^n Y_{ij}$ , при  $j=1, 2, \dots, m$ .

На практике не всегда удается гарантировать одинаковое количество элементов на каждом уровне фактора. Если количество элементов на  $j$ -м уровне, обозначенное через  $n_j$ , при  $j=1, 2, \dots, m$ , то объем выборки составит:

$$N = \sum_{j=1}^m n_j. \quad (5.16)$$

Формула (5.15) может быть записана в виде:

$$Q_1 = \sum_{j=1}^m \frac{1}{n_j} \cdot T_j^2 - \frac{1}{N} \cdot \left( \sum_{j=1}^m T_j \right)^2, \quad (5.17)$$

а суммы  $P_j$  и  $T_j$  как  $P_j = \sum_{i=1}^{n_j} Y_{ij}^2$  и  $T_j = \sum_{i=1}^{n_j} Y_{ij}$ . Тогда дисперсии определяются по зависимостям:

$$\sigma^2 = \frac{Q}{N-1}; \quad \sigma_1^2 = \frac{Q_1}{m-1}; \quad \sigma_2^2 = \frac{Q_2}{N-m}. \quad (5.18)$$

## 5.2. Двухфакторный дисперсионный анализ

Если в дисперсионный анализ включить несколько факторов, влияющих на результативный признак, то они должны быть независимыми друг от друга. При обработке данных исходной информации алгоритм расчетов аналогичен однофакторному дисперсионному анализу.

Определим влияние метеорологических условий (фактор I) и водохозяйственных мероприятий (фактор II) на урожай биомассы трав в различных агроландшафтах.

Здесь имеет место обработка данных с двумя факторами, каждый из которых делится на две группы. Для этого составляется комбинационный (двухфакторный) дисперсионный комплекс (табл. 5.1). Каждый фактор характеризуется тремя наблюдениями (повторностями). Аналогичную схему можно использовать для двухфакторного анализа с большим числом групп и повторностей в каждом факторе.



Таблица 5.1

## Двухфакторный дисперсионный комплекс

Повторяемость опыта по фак- тору (II)	Биомасса, кг/м <sup>2</sup>		$\frac{\sum Y_i}{\sum Y_i^2}$	$(\sum Y_i)^2$	$\bar{Y}$
	Группы по фактору (I)				
	1992 год (сухой)	1994 год (влажный)	4	5	6
1	2	3	4	5	6
Группа фактора (II) (неосуществленный агроландшафт)					
Первая	5/25	4/16	9/41		
Вторая	6/36	5/25	11/61		
Третья	5/25	6/36	11/61		
$\Sigma\Sigma$	16/86	15/77	31/163	961	5,16
Группа фактора (II) (осуществленный агроландшафт)					
Первая	3/9	5/25	8/34		
Вторая	4/16	6/36	10/52		
Третья	4/16	6/36	10/52		
$\Sigma\Sigma$	11/41	17/97	28/138	784	4,66
$\sum X_i$					
$\sum X_i^2$	27/127	32/174	59/301	$\Sigma(\sum Y_i)^2=1745$	
$(\sum X_i)^2$	729	1024		$\Sigma(\sum X_i)^2=1753$	
$\bar{M}$	4,50	5,33		$\bar{M}_{общ} = 4,90$	

Примечание.  $X_i$  – варианты опыта по фактору (I),  $Y_i$  – то же по фактору (II). В числителе – урожай биомассы, в знаменателе – квадрат чисел

Суть двухфакторного дисперсионного анализа можно представить равенством:

$$Q=Q_1+Q_2+Q_3+Q_4+Q_5, \quad (5.19)$$

где  $Q$  – общая сумма квадратов;  $Q_1$  и  $Q_2$  – соответственно сумма квадратов отклонений для факторов (I) и (II);  $Q_3$  – сумма квадратов отклонений, имеющих место при взаимодействии факторов (I) и (II);  $Q_4$  – сумма квадратов отклонений по повторностям;  $Q_5$  – остаточная сумма квадратов отклонений неучтенных факторов.

Общая сумма квадратов отклонений находится как:

$$Q=(\Sigma\Sigma X_i^2, Y_i^2 - (\Sigma\Sigma X_i, Y_i)^2 / N)=(301-(59^2 / 12))=10,92,$$

где  $N=12$  – общий объем выборки; сумма квадратов отклонений по фактору (I) как:

$$Q_1 = (\sum(\sum X_i)^2 - (\sum \sum X_i Y_i) / n_x) / k_x = (1753 - 59^2 / 2) / 6 = 2,08,$$

где  $n_x=2$  – число групп фактора (I);  $k_x=6$  – число вариант в каждой отдельной сумме; сумма квадратов отклонений по фактору (II) вычисляется аналогично:

$$Q_2 = (\sum(\sum Y_i)^2 - (\sum \sum X_i Y_i) / n_y) / k_y = (1745 - 59^2 / 2) / 6 = 0,75;$$

сумма квадратов отклонений, вызываемых взаимодействием факторов (I) и (II), определяется следующим образом:

$$Q_3 = (\sum(\sum Z_i^2) - (\sum \sum X_i Y_i)^2 / n_z) / k_z - Q_1 - Q_2 = (891 - 59^2 / 4) / 3 - 2,08 - 0,75 = 4,08,$$

где  $\sum(\sum Z_i^2) = (16^2 + 15^2 + 11^2 + 17^2) = 891$  – сумма квадратов сумм значений вариант по группам выборки комбинационной таблицы;  $n_z=4$  – число сумм вариант по группам;  $k_z=3$  – число слагаемых вариант в каждой группе выборки; сумма квадратов отклонений по повторностям  $Q_4$  определяется как -  $Q_4 = (\sum(\sum X_i)^2 - (\sum \sum X_i Y_i) / n_{x,y}) / k_{x,y} = (1171 - 59^2 / 3) / 4 = 2,67,$

где  $n_{x,y}=3$  – число сумм по повторностям;  $k_{x,y}=4$  – число слагаемых в каждой сумме; сумма квадратов сумм  $X_i$ , вычисленная как:

$$\sum(\sum X_i)^2 = ((5+4) + (3+5))^2 + ((6+5) + (4+6))^2 + ((5+6) + (4+6))^2 = 1171;$$

сумма квадратов отклонений по остаточному варьированию составляет:

$$Q_5 = Q - Q_1 - Q_2 - Q_3 - Q_4 = 10,92 - 2,08 - 0,75 - 4,08 - 2,67 = 1,14;$$

число степеней свободы для  $Q$  будет -  $v=N-1=11$ ; для  $Q_1$  и  $Q_2$  – соответственно равно числу градаций фактора минус единица -  $v_1=n_1-1=2-1=1$ ,  $v_2=n_2-1=2-1=1$ ; для  $Q_3$  определится как -  $v_3=v_1 \cdot v_2=1 \cdot 1=1$ ; для  $Q_4$  – равно числу повторностей минус единица -  $v_4=3-1=2$ ; для  $Q_5$  этот показатель определяется следующим образом -  $v_5=v-v_1-v_2-v_3-v_4=11-1-1-1-2=6$ .

Полученные расчетным путем характеристики сведены в табл. 5.2. Показатели дисперсии (табл. 5.2) вычисляются путем деления значений сумм квадратов отклонений на соответствующие значения степеней свободы. Фактический критерий Фишера определяется путем деления каждой из величин дисперсий на значение остаточной дисперсии.

Таблица 5.2

## Результаты двухфакторного дисперсионного анализа

Варьирование данных	Сумма квадратов отклонений (Q)	Степень свободы (v)	Дисперсия ( $\sigma^2$ )	Критерий Фишера	
				$F_\phi$	$F_{0,05}^T$
Общие по опыту	10,92	11	0,99	5,21	4,03
По фактору (I)	2,08	1	2,08	10,94	5,99
По фактору (II)	0,75	1	0,75	3,94	5,99
По взаимодействию факторов (I) и (II)	4,08	1	4,08	21,47	5,99
По повторностям	2,67	2	1,34	7,05	5,14
Остаточное	1,14	6	0,19	1,00	—

Анализируя критерии Фишера, можно заключить, что влияние исследуемых параметров на биомассу существенно (в целом по опыту, по фактору - I, по взаимодействию факторов и по повторностям, т. е. во всех рассмотренных случаях) при вариации опытных данных

$V = \sqrt{\sigma_{общ}^2 / M_{общ}} \cdot 100\% = \sqrt{0,99 / 4,9} \cdot 100\% = 20,0\%$  и  $F_{тм} > F_{0,05}^T$ ; действие группы факторов (II) на исследованных агроландшафтах не доказано  $F_{тм} < F_{0,05}^T$ .

Оценку результатов эксперимента можно сделать и по критериям наименьшей существенной разности и критерию Стьюдента. Для вычисления НСР и  $t$  находится ошибка среднего арифметического  $m_M$  всего опыта и ошибка разности средних по формулам:

$$m_M = \sqrt{\sigma_{осм}^2 / N} = \sqrt{0,19 / 12} = 0,1258;$$

$$m_d = \sqrt{2 \cdot \sigma_{осм}^2 / n} = \sqrt{2 \cdot 0,19 / 6} = 0,25;$$

$$НСР = m_d \cdot t_{0,05}^T = 0,25 \cdot 2,45 = 0,61,$$

в которых  $n$  – численность меньшей из сравниваемых частных групп. По критерию Стьюдента сравниваются средние арифметические данные по осушенному и неосушенному агроландшафтам:

$$t = (M_{y,1} - M_{y,2}) / m_{d_y} = (5,16 - 4,66) / 0,25 = 2,00.$$

Критерий Стьюдента –  $t_{0,05}^T = 2,45$ , при  $P=0,95$  для  $\nu=6$ . Таким образом, на биомассу трав в агроландшафтах не влияет мелиорация (т. е. фактор II), так как  $t_{\phi} = 2,0 < t_{0,05}^T = 2,45$ , при  $P=0,95$ ; метеорологические условия (фактор I) достоверно влияют на биомассу трав, при  $P=0,95$ . Выводы, сделанные с использованием критериев Фишера и Стьюдента, совпадают. В заключение обычно определяется точность опыта, которая составляет:

$$P = (m_M / M_{\text{общ}}) \cdot 100\% = 0,1258/4,9 \cdot 100\% = 2,56\%.$$

Точность опыта признается достаточно высокой, поскольку  $P < 3\%$ . Коэффициент варьирования опытных данных:

$$V = \frac{\sqrt{\sigma_{\text{об}}^2}}{M_{\text{общ}}} \cdot 100\% = \frac{\sqrt{0,99}}{4,9} \cdot 100\% = 20,0\%,$$

также незначителен, что удовлетворяет требованиям опыта.

### 5.3. Критерии статической значимости флюктуаций природных процессов

Вследствие того, что вычислять, анализировать спектр и подбирать фильтр довольно-таки трудно, часто применяются более простые оценки значимости отклонений, т. е. различных видов неслучайностей в геофизических рядах.

В этом случае статистический анализ должен дать последовательные и наиболее надежные ответы на два вопроса, касающиеся временного ряда: 1-ый – случаен ли временной ряд или нет? 2-ой – если нет, то какая характерная форма неслучайности содержится в ряде? На практике обычно нужно выполнять анализ рядов, в которых характерные варианты случайностей неизвестны (или, по крайней мере, сомнительны).

Прежде всего необходимо решить следующую важнейшую задачу. Многие известные критерии значимости обоснованы только в том случае, если частотное распределение переменной исследуемого временного ряда имеет Гауссову, т. е. нормальную или, в особых случаях, какую-то другую форму распределения. Если нас удовлетворяет, что эмпирическое распределение в действительности то же самое, или, по крайней мере, почти то же самое, что и распределение, от которого зависит обоснованное, можем использовать этот критерий. В противном случае необходимо будет обратиться к обобщенному не параметрическому критерию значи-

мости, обладающему тем преимуществом, что для него безразлична действительная форма распределения. Как известно, у Кендалла и Стьюарта критерии, безразличные к формам распределения частоты, названы «здравыми»; более того, ряд критериев, в которых первоначально предполагалось Гауссово распределение, вполне применим к данным, распределение которых значительно отклоняется, по крайней мере, в некоторых отношениях от Гауссова распределения.

Анализ климатологических рядов требует признания, что самая общая форма неслучайности в любых климатологических рядах определяется процессом Маркова первого порядка или это действительный тренд.

#### *Общие критерии случайности*

Тренд, периодичность, эффект связанности членов ряда или другие виды неслучайности могут оказывать большие влияния на величину серийной корреляции при сдвиге на 1, чем при больших сдвигах в ряде. Это обстоятельство придает большое потенциальное значение коэффициенту серийной корреляции при сдвиге на 1 как критерию неслучайности вообще.

Если мы имеем ряд длиной  $N=46$  и хотим сравнить выборочное  $r_1$  с нулевой гипотезой, чтобы  $r_1$  было не больше, чем соответствующее значение случайности, то можем выбрать одностороннюю 95%-ную точку вероятности  $t_g$ , которая по таблице распределения Гаусса будет равна  $t_g=1,645$ . Затем подставим в уравнение

$$(r_1)_t = \frac{-1 \pm t_g \sqrt{N-2}}{N-1} \quad (5.20)$$

это значение  $t_g$  и  $N=46$ , а числитель взяв со знаком «+», при  $N=45$  по-

$$\text{лучим: } (r_1)_t = \frac{-1 + 1,645\sqrt{43}}{44} = +0,223.$$

Тогда и только тогда, когда выборочное значение  $r_1$  больше 0,223, можем сделать вывод, что ряд действительно не случаен (при 95% уровне значимости).

Если получено отрицательные значения  $r_1$ , то это означает, что ряд содержит явное высокочастотное колебание и в таком исключительном случае предпочтительным может оказаться 2-сторонний критерий.

Если частотное распределение значений ряда значительно отклоняется от Гауссовского (нормального) распределения, необходимо проверить значение  $r_1$  каким-то другим способом, а именно, способом, предложенным Вальдом и Волфовидцем.

Учитывая сложность этого способа, Снейерс [Снейерс, 1976] предложил упрощенный вариант критерия, основанного на рангах который хотя и не является хорошим критерием случайности, но все же может быть использован при необходимости.

По методу Снейерса, значение каждого из  $N$  членов ряда  $x_i$  заменяется его рангом  $K_i$ , который изменяется от 1 для наименьшего значения до  $N$  – для наибольшего. Затем производится расчет циклической статистики

$$R'_c = \frac{\sigma \sum_{i=1}^N K_i K_{i+1}}{N(N+1) \cdot (2N+1)}, \quad (5.21)$$

который, однако, в случае нециклической статистики можно успешно модифицировать для обычного применения:

$$R' = \frac{\sigma \sum_{i=1}^{N-1} K_i K_{i+1}}{(N^2 - 1) \cdot (2N+1)}. \quad (5.22)$$

Критерий значимости можно записать так

$$(R')_i = \frac{(3N+2) \pm tg \sqrt{N-2.8}}{4N+2}. \quad (5.23)$$

Необходимо заметить, что  $R'_i$  и  $R'$  нельзя сравнить с  $r_i$ , поскольку в числителях уравнений (5.22) и (5.23) нет поправочного члена, учитывающего ненулевую среднюю  $K_i$ . Как отмечает Снейерс, ожидаемое значение  $R'_i$  (которое приблизительно то же, что и  $R'$ ) для условия нулевой случайности равно  $(3N+2)/(4N+2)$ . Допускается использование критерия Снейерса в тех случаях, когда нецелесообразно вычислять  $F(s)$  Вальда – Волфовитца.



## Отношение Фон-Неймана

Критерий случайности для неопределенных вариантов, который тесно связан с сериальным нециклическим коэффициентом корреляции со сдвигом на 1 член, есть отношение Фон Неймана. Его можно использовать вместо  $r_i$  критерия, каждый раз, когда распределение частотных значений ряда приближается к Гауссовскому распределению. Использование этого критерия представляет собой интерес по нескольким причинам. Во-первых, его легче рассчитать. Во-вторых, он определяется как начало нециклической статистики. И в-третьих, его распределение для нулевого условия Гауссовского случайного ряда точно известно.

Отношение Фон-Неймана определяется как отношение среднего квадратичного последовательного отклонения к дисперсии

$$V = \frac{N}{N-1} \frac{\sum_{i=1}^{N-1} (x_i - x_{i+1})^2}{\sum_{i=1}^N x_i^2 - \frac{1}{N} \left( \sum_{i=1}^N x_i \right)^2}. \quad (5.24)$$

Для большого  $N$ , при условии случайности,  $V$  — есть распределение Гаусса с ожидаемым значением  $2N/N-1$  и дисперсией, приближенно заданным в следующем виде:  $\frac{4(N-2)}{(N-1)^2}$ .

Использование отношения Фон-Неймана в критериях значимости такое же, как и для  $r_i$ , но нужно иметь в виду, что обычные варианты случайности приведут к малым значениям  $V$ .

Это ясно видно из уравнения связи между  $V$  и  $r_i$ , которое приближенно имеет вид:

$$V = 2(1 - r_i). \quad (5.25)$$

Следовательно, за критерий можно принять уравнение:

$$(V)_t = \frac{2N \pm 2t_g \sqrt{N-2}}{N-1}, \quad (5.97)$$

в числителе которого обычно берется знак «-», а значение  $t_g$  — из распределения Гаусса, соответствующее одностороннему критерию при заданном уровне значимости. Затем определяется выполняется ли условие  $V < (V)_t$ ; если это условие выполняется, то предполагается, что ряд неслу-

чаен. Если же  $V$  не меньше  $(V)_0$ , принимаем «нулевую» гипотезу, то есть ряд случаен.

*t-критерий Стьюдента для выявления различия между 2-мя средними*

Часто в анализе климатических флуктуации нас может интересовать не столько общее поведение временного ряда, сколько сравнение средних значений ряда за два различных периода наблюдений, в случае, когда резкое изменение средней сомнительно, такая постановка вопроса очевидна. Однако критерий, который применяется в этом случае оказывается способным обнаружить непостоянство средней, включая тренд. Кроме того, этот особый критерий целесообразен тем, что его можно применять к данным с произвольным распределением частоты, кроме случая, когда длины двух периодов, выбранных для сравнения, неравны [Кендалл, Стьюарт, 1961].

Критерий обладающий этими ценными свойствами – это известный  $t$ -критерий Стьюдента, для обнаружения различий между двумя средними. Если две рассматриваемые средние обозначить через  $\bar{X}_1$  и  $X_2$ , соответственно, числа значений, из которых вычисляется каждая средняя, обозначим через  $N_1$  и  $N_2$ , то статистика критерия вычисляется по уравнению:

$$t_d = \frac{((X_1 - \bar{X}_2) - (M_1 - M_2)) \cdot \sqrt{n_1 + n_2 - 2}}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \cdot \sqrt{(n_1 - 1) \cdot \sigma_1^2 + (n_2 - 1) \cdot \sigma_2^2}}. \quad (5.26)$$

Здесь  $(M_1 - M_2)$  – предлагаемая разность между  $\bar{X}_1$  и  $\bar{X}_2$ , согласно нулевой гипотезе, которая при наличии случайности принимается равной нулю. Величины  $\sigma_1^2$  и  $\sigma_2^2$  – соответственно выборочные рассеяния  $X_i$  в двух периодах наблюдений, а  $N_i \cdot S_i^2$  вычисляется обычным способом по уравнению:

$$N_i \sigma_i^2 = \sum_{i=1}^N x_i^2 - \frac{1}{N_i} \left( \sum_{i=1}^{N_i} x_i \right)^2. \quad (5.27)$$

Для «нулевой» гипотезы случайности, распределение  $t_d$ , заданное уравнением (5.27), следует  $t$ -распределение Стьюдента для  $(N_1 + N_2 - 2)$  степеней свободы.

Обычно критерий должен основываться на сравнениях величины  $t_d$  с 95%-ми точками вероятности  $t$ -распределения, соответствующего двусторонней форме критерия. Если  $t_d$  лежит за пределами этого  $t$ -распределения, необходимо принимать разность между двумя выборочными средними как доказательство неслучайности. В ряде случаев, когда вариант случайности можно предсказать априори, чтобы получить один знак в числителе уравнения (5.27), можно сравнивать  $t_d$  с более точной 95%-ой точкой  $t$ -распределения, соответствующей одностороннему критерию.

*Критерий Крамера для сравнения средних подпериодов со средней всего периода*

Иногда необходимо исследовать устойчивость длинного ряда наблюдений, сравнивая общую среднюю за весь период со средними определенных частей этого периода. В этом случае можно использовать критерий Крамера для определения – не будут ли разности средних больше, чем можно было ожидать по «нулевой» гипотезе случайности.

Пусть  $\bar{X}$  и  $\sigma$  определяются соответственно как среднее и среднее квадратичное отклонение всех данных значений, т. е.

$$x = \frac{\sum_{i=1}^N x_i}{N}; \quad \sigma = \left[ \frac{\sum_{i=1}^N x_i^2}{N} - (\bar{x}_k)^2 \right]^{\frac{1}{2}}, \quad (5.29)$$

а  $\bar{x}_k$  – среднее подпериода  $n$  значений.

Сравнивая с  $\bar{X}$   $\bar{x}_k = \frac{\sum_{i=k+1}^{k+n} x_i}{n}$  и определив  $\tau_k = \frac{(\bar{x}_k - \bar{X})}{\sigma}$ , имеем:

$$t_k = \left[ \frac{n(N-2)}{N-n(1+\tau_k^2)} \right]^{\frac{1}{2}} \tau_k. \quad (5.30)$$

Статистическое  $t_k$  распределяется как  $t$ -распределение Стьюдента, имеющее  $(N-2)$  степеней свободы. Поэтому можно проверить значимость точно таким же способом, как статистику  $t_N$  в уравнении (5.27). Этот критерий можно проверить для любого необходимого числа и выбора подпериодов за весь период наблюдений, исключая случаи, когда подпериоды

могут перекрывать друг друга, интерпретация результатов критерия значимости будет требовать другого особого подхода.

### *Критерий случайности, зависящий от тренда*

Часто получается так, что наиболее вероятный вариант случайности в климатологическом ряду является какой-то формой тренда. Как правило, нет причины предполагать рассматриваемые тренды – линейными, что и определяет целесообразность использовать те критерии случайности, которые в отличие от схем линейной регрессии наименьших квадратов имеют почти постоянную мощность для случая нелинейных трендов.

Можно выделить две формы рангового критерия в зависимости от тренда. Хотя они формулируются несколько неодинаково, но по существу они эквивалентны [Кендалл, Стьюарт, 1961].

### *Ранговая статистика Манна-Кендалла*

Ранговый критерий, зависящий от варианта тренда более прост по сравнению с другими, но в некоторых случаях его применение становится проблематичным.

Прежде чем использовать этот критерий, необходимо рассмотреть только относительные значения всех членов  $x_i$  анализируемого ряда. Прежде чем применить критерий, полезно заменить  $x_i$  на их ранги  $K_i$ , так, чтобы каждый член получал номер от 1 до  $N$ , что отражает его величину по отношению к величинам всех других членов.

Затем вычисляем статистическое  $\tau$  из ряда следующим образом. Сравнивается значение первого члена ряда  $x_1$  (или его ранговое значение  $K_1$ ) со значениями всех последующих членов ряда от 2-го до  $N$ -го и подсчитывается число последующих членов ( $n_1$ ), значения которых превышают  $X_1$  или  $K_1$ . Затем сравнивается значение 2-го члена  $X_2$  (или  $K_2$ ) со значениями последующих членов и подсчитывается число последующих членов ( $n_2$ ), которые превышают  $X_2$  (или  $K_2$ ). Продолжим этот подсчет для каждого члена ряда, заканчивая  $X_N - 1$  (или  $K_N - 1$ ) и его соответствующим номером  $n_N - 1$ . Тогда  $P$  задается суммой  $P = \sum_{i=1}^{N-1} n_i$

Статистическое  $\tau$ , которое нам необходимо, получаем из отношения  $N$  и  $P$ .

$$\tau = \frac{4P}{N(N-1)} - 1. \quad (5.31)$$

Поскольку  $\tau$  распределяется почти как Гауссовское нормальное распределение для  $N > 10$  и имеет предполагаемое нулевое значение и дисперсию равными  $(4N+10) \cdot 9N(N-1)$ , то его можно использовать как основу критерия значимости, сравнивая его со значениями

$$(\tau)_l = 0 \pm t_g \sqrt{\frac{4N+10}{9N(N-1)}}, \quad (5.32)$$

где  $t_g$  – необходимая вероятностная точка гауссовского нормального распределения, соответствующая двухстороннему критерию.

#### *Ранговая статистика Спирмена*

Статистика 2-го ранга, которая определяет случайность в зависимости от варианта тренда (более проста для использования, когда несколько членов временного ряда оказываются равными по значению, а ранги этих членов связаны). Мощность критерия, основанного на  $r_s$ , по существу та же, что и критерия, основанного на  $\tau$ .

При расчете  $r_s$  необходимо, чтобы ряд  $x_i$  при исследовании сначала преобразовывался в его эквивалент ранга  $K_i$  вышеописанным способом, как произвольное для расчета  $\tau$ . Для каждого члена ряда вычисляем величину  $d_i = K_i - i$  или, другими словами, разность между рангом  $i$ -го члена и самим номером этого члена, где  $i$  колеблется от 1 до  $N$ . Тогда  $r_s$  можно проверить по уравнению:

$$r_s = 1 - \frac{6 \sum_{i=1}^N d_i^2}{N(N^2 - 1)} \quad (5.33)$$

Для  $N > 8$ , значение  $r_s$  можно проверить на значимость, определив  $t$  в уравнении

$$t = r_s \sqrt{\frac{N-2}{1-r_s^2}}, \quad (5.34)$$

и сравнив его с вероятностными точками  $t$ -распределения Стьюдента для  $(N-1)$  степеней свободы. Как и в случае  $\tau=95\%$ , точку вероятности для двухстороннего критерия нужно использовать обычным способом.

Особенности способа вычисления  $\tau$  и  $r_s$  и использования каждого в критерии значимости рассмотрим на следующем примере.

Имеются средние значения температур за январь за 20 лет на каждой из двух станций А и В и предполагается, что продолжительные антропогенные факторы приводят к росту температуры (тренду) на станции А. Таким образом, анализ должен основываться на ряде различий температуры в пунктах А и В. Эти различия вместе с различной статистикой, полученной по станциям, даются в табл. 5.3.

Применив сначала критерий, основанный на  $\tau$  вычисляем  $P$ , суммируя различные значения  $n$  ( $P=105,5$ ). Вычислив  $P$  и имея  $N=20$ , находим  $\tau$  по уравнению (5.34):  $\tau = +0,111$ .

Найденное значение  $\tau$ , в свою очередь, нужно сравнить с пределами ( $\tau$ ), задаваемые уравнением (5.32). Так как значение  $t_g$  в Гауссовском нормальном распределении, соответствующее двухсторонней 95% точке распределения, составляет 1,96, можем найти  $\tau$ :  $(\tau)_l = \pm 0.318$ .

Поскольку  $\tau$  находится в пределах от  $-0.318$  до  $+0.318$ , то оно незначительно и, таким образом, последовательность различий в температуре между станциями «А» и «В» должна приниматься случайной. Практически это означает, что тренда в данных на станции А нет.

Далее, применяем критерий, основанный на  $r_s$ . Находим сумму квадратов  $d_i$ , приведенных в последнем столбце табл. 5.3, и получаем:

$$\sum_{i=1}^{20} d_i^2 = 1147,5$$

Подставив эту величину и  $N=20$  в уравнение (5.33), получим  $r_s = +0,137$ .

Для определения значимости этой величины  $r_s$  уравнение (5.105) и находим  $t = +0,587$ .

Затем сравниваем  $t$  с двухсторонними 95 % точками вероятности  $t$ -распределение Стьюдента при  $(N-2)=18$  степенях свободы, т. е.  $t=\pm 2,101$ . Поскольку, вычисленное  $t$  находится в этих пределах, то мы приходим к выводу, что тренд в этих данных отсутствует.



Таблица 5.3

Данные для вычисления тренда используя ранговую статистику

Номер года $i$	Разность температур $X_i$ ( $^{\circ}\text{F}$ )	Ранг $x_i$ $K_i$	$n_i$	$d_i=K_{i,i}$
1	-0.4	6	13	5
2	1.1	15	5	13
3	1.9	17	3	14
4	-1.2	2	15	-2
5	-1.0	3	14	-2
6	1,0	14	4	8
7	2,0	18	2	11
8	0,6	9	7	1
9	0,2	8	7	-1
10	-0,9	4	9	-6
11	-0,8	5	8	-6
12	0,9	12,5	3,5 <sup>1)</sup>	0,5
13	0,1	7	6	-6
14	0,9	12,5	3	-1,5
15	0,8	11	3	-4
16	2,1	19	1	3
17	1,2	16	1	-1
18	0,7	10	1	-8
19	2,8	20	0	1
20= $N$	-1,4	1	-	-19

*Критерий постоянства изменчивости (критерий Бартлетта)*

Вместо или в дополнение к непостоянству средней, климатическая флуктуация может включать и непостоянство дисперсии или, другими словами, изменение в общей климатической изменчивости. Ни один из критериев случайности, описанных выше, не может измерить непостоянство дисперсии, и для этой цели должны исследоваться другие критерии. Как указывали Кендалл и Стьюарт, обычные статистические критерии дисперсии чрезвычайно чувствительны к различным формам распределения частоты, т. е. имеющиеся критерии в большинстве случаев основываются на допущении Гауссовского нормального распределения частоты. Поэтому нельзя рекомендовать использовать эти критерии в таких климатологических данных, которые значительно отклоняются от Гауссовского нормального распределения.

На практике часто используют сокращенный вариант критерия Бартлетта, когда ряд можно разделить на « $K$ » равных периодов, где  $K \geq 2$ . В каждом из этих подпериодов вычисляется выборочное рассеяние, т. е. величина

$$\sigma_k^2 = \frac{1}{n} \left[ \sum x_i^2 - \frac{1}{n} (\sum x_i)^2 \right], \quad (5.35)$$

где суммы изменяются в пределах  $n$  значений ряда в подпериоде  $K$ .

Из значений  $\sigma_k^2$  выбираем наибольшее, обозначаемое через  $\sigma_{\max}^2$  и сравниваем отношение  $\sigma_{\max}^2 / \sigma_{\min}^2$  со стандартным значением  $K$  и  $n$ .

Необходимо также заметить, что полная форма критерия Бартлетта допускает неравные длины полупериодов и учитывает все значения  $\sigma_n^2$ .

## 6. МЕТОДЫ ПРОСТРАНСТВЕННОГО ОБОБЩЕНИЯ ГИДРОМЕТЕОРОЛОГИЧЕСКОЙ ИНФОРМАЦИИ

Пространственное обобщение гидрометеорологической информации является заключительным этапом статистической обработки результатов наблюдений с их представлением в виде информационных полей. Основное содержание пространственного обобщения информации: оценка статистических показателей пространственной структуры полей, картирование и районирование метеорологических, климатических, гидрологических и эколого-ландшафт-ных характеристик. Исследование элементов пространственной структуры приобретает важное значение в связи с необходимостью представления рассматриваемой информации непосредственно в границах природно-экономических (административных) районов, осуществление программы природно-климатического мониторинга и экологического аудита. Поля гидрометеорологических и экологических элементов обычно задаются данными в отдельных точках пространства (по отдельным пунктам или проинтерполированными в узлы регулярной сетки с помощью объективного анализа).

На рис. 6.1 показана принципиальная схема формирования информационных полей региона.

Поскольку каждое значение элемента в координатах поля представлено случайным элементом выборки, то такие информационные поля элементов являются случайными полями. Для описания совокупности рядов в пространстве, используются корреляционные, ковариационные и спектральные функции. Данные функции однозначно связаны между собой и равно пригодны для описания статистической структуры полей элементов, но на практике чаще используются пространственные корреляционные функции (ПКФ) – как более точные, универсальные, менее зависимые от колебаний сезонного и географического характера. Пространственные корреляционные функции случайного поля  $M(\varphi_j, \lambda_j, t_j)$ , в общем случае, характеризуются зависимостью:

$$r_{jk} = \frac{\sum_{i=1}^{n_{jk}} (M_{ij} - \bar{M}_j) \cdot (M_{ik} - \bar{M}_k)}{\sqrt{\sum_{i=1}^{n_{jk}} (M_{ij} - \bar{M}_j)^2 \cdot \sum_{i=1}^{n_{jk}} (M_{ik} - \bar{M}_k)^2}}, \quad (6.1)$$

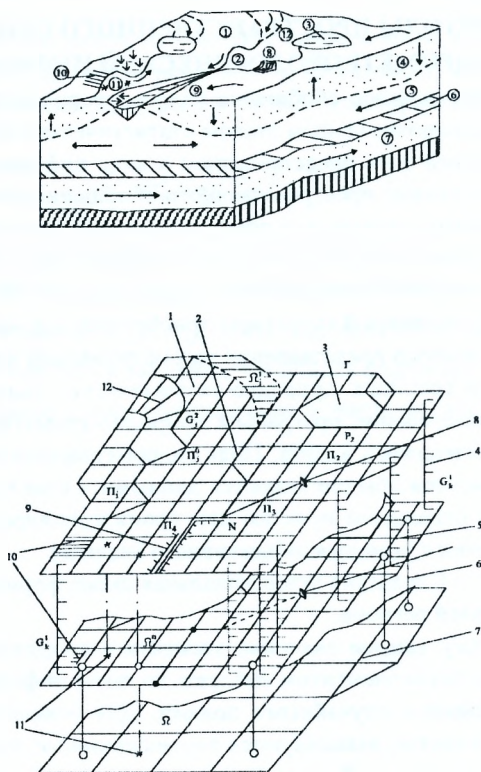


Рис. 6.1. Схема формирования информационных полей региона: а – компонентная, б – информационная структура; 1 – поверхностный сток ( $G_j^2$ ); 2 – речная сеть ( $\Pi$ ); 3 – водоемы ( $\Gamma$ ); 4 – область фильтрационных потоков; 5 – область грунтовых вод ( $\Omega^n$ ); 6 – область напорных подземных вод ( $\Omega$ ); 7 – подземный сток; 8 – инженерные сооружения ( $N$ ); 9 – область пойменных процессов; 10 – область действия поверхностных водохозяйственных систем ( $P$ ); 11 – область действия глубинных хозяйственных систем; 12 – область склонового стока

в которой  $M_{ij}$  – значение элемента гидрометеорологического поля в  $j$ -ой точке с координатами  $\varphi_j$  и  $\lambda_j$  в  $i$ -ый интервал времени;  $\overline{M}_j$  – норма гидрометеорологической величины в  $j$ -ой точке;  $n_{jk}$  – количество интервалов совместных наблюдений за элементом поля.

Практический интерес представляют однородные и изотропные поля. Однородными являются поля с одинаковыми законами распределения вероятностей той или иной величины во всех координатах поля (равенство одноточечных характеристик средних и дисперсий). Если пространственная корреляционная функция поля зависит только от расстояния между рассматриваемыми точками и не зависит от направления между ними, то поле считается изотропным. В случае невыполнения этих условий, говорят об однородности и изотропности поля только относительно пространственных корреляционных функций. Для однородных и изотропных полей ПКФ зависит только от расстояния между наблюдаемыми точками, т. е.  $R=R(\rho)$ . Однако для реальных информационных полей, из-за сложности их структуры, характерна локальная однородность и локальная изотропность, т. е. однородность и изотропность в пределах пространственных и временных моментов, характерных для исследуемых процессов. В зависимости от этого рассматриваются микро-, мезооднородность – изотропность в пределах от нескольких километров до сотен километров и макрооднородность – изотропность от сотен километров и далее. Это связано с тем, что в отличие от микро- и мезоизменчивости, макроизменчивость вызывается, по существу, принципиально иными причинами, связанными с глобальными процессами циркуляции атмосферы, физическая природа которых отличается от природы микро- и мезомасштабных процессов. В случае мезо-, макромасштабных явлений, условия локальной однородности выполняются лишь в горизонтальном направлении, как правило, до расстояний не более 2000 км, хотя для ряда метеорологических полей (осадки, запасы воды в снеге и др.) однородность (изотропность) нарушается на гораздо меньших расстояниях. В любом случае это происходит из-за существования неоднородности полей гидрометеорологических элементов в горизонтальном направлении [Волчек, 2001].

### 6.1. Оценка статистической макроструктуры полей

Кроме пространственных корреляционных функций, к числу наиболее важных статистических характеристик, описывающих пространственную структуру гидрометеорологических и экологических полей, относятся кривая распределения в пространстве (ее параметры: средняя арифметическая –  $\bar{X}_F$  и дисперсия –  $\sigma_F$  по площади –  $F$ ), коэффициент аномальности поля, пространственная структурная функция, а также показатели сходства и различия полей. Осреднение используется как способ получения результатов для сравнения фоновых характеристик рассматриваемых элементов в отдельных районах или их изменений на больших территориях. Осреднение – необходимый этап обобщения исходных данных при анализе полей гидрометеорологических элементов в силу их большой пространственной изменчивости. Существует множество способов площадного осреднения, но наиболее часто используются методы среднего арифметического взвешивания по площадям – квадратам, изолиний и оптимального осреднения. Обобщенное выражение для оценки средних по площади записывается в виде (Колде, 1991):

$$\bar{X}_F = \sum_{i=1}^n \alpha_i \cdot x_i(\varphi_i, \lambda_i), \quad (6.2)$$

где  $\bar{X}_F$  – среднее значение рассматриваемого элемента по площади;  $x_i(\varphi_i, \lambda_i)$  – значение элемента в точке с координатами  $\varphi_i, \lambda_i$ ;  $\alpha_i$  – коэффициент, зависящий от веса и способа осреднения;  $n$  – число метеостанций, используемых при осреднении. При равномерном расположении метео-, мониторинговых станций в условиях равнинного рельефа (когда веса  $\alpha_i \approx 1$ ), чаще всего используется метод арифметического осреднения:

$$\bar{X}_F = \frac{\sum_{i=1}^n x_i(\varphi_i, \lambda_i)}{n}. \quad (6.3)$$

Арифметическая сумма по методу взвешивания по площадям (метод полигонов) вычисляется как:

$$\bar{X}_F = \sum_{i=1}^n X_i(\varphi_i, \lambda_i) \cdot F_i / \sum_{i=1}^n F_i, \quad (6.4)$$



где  $F_i$  – площадь полигона, относящаяся к метеомониторинговой станции  $X_i(\varphi_i, \lambda_i)$ .

При неравномерном распределении пунктов по площади целесообразно применять метод квадратов, который при использовании ЭВМ называется еще методом оптимальной интерполяции. По методу квадратов исследуемая площадь разбивается на регулярную сеть квадратов или точек, данные в которых определяются путем осреднения наблюдаемых значений непосредственно в квадратах или снимаются с карт изолиний. В последующем, для определения  $\bar{X}_F$ , производится осреднение характеристики по формуле (6.2). В методе оптимальной интерполяции весовые множители  $P_j$  находятся, при наличии сведений о статистической структуре поля, путем решения системы уравнений:

$$\sum_{j=1}^n P_j \cdot R_{ij} = R_{0i}, \text{ при } (i=1, 2, \dots, n), \quad (6.5)$$

или в случае, когда учитывается ошибка наблюдений  $\Delta f(\rho_i=0)$  в точке, при  $(\rho_i=0)$ , по соотношению:

$$\sum_{j=1}^n P_j \cdot R_{ij} + P_j \cdot \eta_j^2 = R_{0i}, \text{ при } (i=1, 2, \dots, n), \quad (6.6)$$

в котором  $\eta_j^2 = \sigma_{\Delta x}^2 / \sigma_i^2$  – мера ошибок наблюдений в  $j$ -ой точке;  $R_{0i}$ ,  $R_{ji}$  – значения коэффициентов корреляции истинных значений исследуемой характеристики в точках  $j$  и 0. Оценка пространственной дисперсии  $\sigma_F^2$ , как правило, рассчитывается по формуле, аналогичной обычной формуле оценки дисперсии для случайной выборки:

$$\sigma_F^2 = \frac{\sum_{i=1}^n (x_i(\varphi_i, \lambda_i) - \bar{X}_F)^2}{n}. \quad (6.7)$$

Рассмотрим предельные случаи интерполяции метеорологических полей.

1. Пусть точка, в которой производится интерполяция, совпадает с одной из станций, в которых имеются данные. Тогда вес данных станции, совпадающей с нашей точкой, равен 1, а все остальные – нулю.  $r_0$  совпадает с  $\bar{r}_k (1 \leq k \leq n)$ . Тогда в системе (6.5) столбец коэффициентов при  $R_i$

будет совпадать со столбцом правых частей. Поэтому  $R_k = 1$ ,  $R_i = 0$  ( $i \neq k$ ) будет решением системы (6.5). Ошибка такой интерполяции равна 0, в силу совпадения  $\bar{r}_0$  и  $\bar{r}_k$ .

2. Если точки с данными расположены от интересующей нас точки  $r_0$  настолько далеко, то значения  $f$  в них статистически не связаны с его значением в точке  $\bar{r}_0$ , то все веса равны 0, так что результатом интерполяции является климатическая норма. В этом случае  $P_{oj} = 0$  и система (6.5) принимает вид

$$\sum_{j=1}^n P_j \cdot R_{ij} = 0. \quad (6.8)$$

Единственным решением этой системы является нулевое  $R_i = 0$ . В этом случае мера ошибки интерполяции  $\epsilon = 1$ , а средний квадрат ошибки равен дисперсии рассматриваемого элемента.

3. Рассмотрим случай, когда взаимное расположение точек таково, что значения в точках  $\bar{r}_i$  статически не связаны друг с другом, но хотя бы некоторые из них статистически связаны со значением  $M_{ij}$ .

Весами для интерполирования независимых друг от друга данных являются значения автокорреляционного коэффициента. Если все расстояния  $|\bar{r}_i - \bar{r}_0|$  меньше того, при котором функция  $P(r)$  достигает первого минимума, то данные в более удаленных точках будут входить с меньшим весом.

Для метода оптимальной интерполяции существенно лишь, чтобы статистическая структура поля была задана. Поэтому рассматриваемый здесь метод интерполяции является простейшим вариантом оптимальной интерполяции. Его применение сводится к следующему.

По расположению  $n+1$  точек ( $n$  станций и одного узла сетки) вычисляются расстояния между каждой парой точек.

$$r_{ij} = \sqrt{(x_i - x_{ji})^2 + (y_i - y_j)^2} \quad (6.9)$$

Эти расстояния образуют матрицу  $n(n+1)$  расстояний

$$\begin{array}{cccccc}
 0 & r_{12} & r_{13} & \dots & r_{1n} & r_{10} \\
 r_{21} & 0 & r_{23} & \dots & r_{2n} & r_{20} \\
 \dots & \dots & \dots & \dots & \dots & \dots \\
 \dots & \dots & \dots & \dots & \dots & \dots \\
 r_{n1} & r_{n2} & r_{n3} & \dots & 0 & r_{n0}
 \end{array} \quad (6.10)$$

Первые  $n$  столбцов которой представляют собой симметричную матрицу, поскольку ясно, что  $r_{ij} = r_{ji}$  поэтому нужно вычислять не

$n(n+1)$ , а лишь  $\frac{n^2}{2}$  расстояний.

Далее с помощью ковариационной функции производится переход от матрицы расстояний к матрице ковариаций.

$$\begin{array}{cccccc}
 D & m_{12} & m_{13} & \dots & m_{1n} & m_{10} \\
 m_{21} & D & m_{23} & \dots & m_{2n} & m_{20} \\
 \dots & \dots & \dots & \dots & \dots & \dots \\
 m_{n1} & m_{n2} & m_{n3} & \dots & D & m_{n0}
 \end{array} \quad (6.11)$$

по формуле

$$m_{ij} = m_f \left( \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \right) \quad (6.12)$$

Матрица (6.11) соответствует системе уравнений (6.5)

$$\sum_{j=1}^n m_{ij} \cdot R_{ij} = m_{i0}, \quad (i = 1, 2, \dots, n) \quad (6.13)$$

для определения весов  $R_i$ .

Имея дело с системой линейных алгебраических уравнений, важно представить себе, не является ли она плохообусловленной, т. е. системой, решение которой сильно меняется в результате небольших изменений коэффициентов.

С геометрической точки зрения каждое из уравнений системы (6.5) можно рассматривать как уравнение некоторой плоскости в  $n$ -мерном пространстве. Тогда решением системы являются координаты точки пересечения этих плоскостей.

Система является плохообоснованной в том случае, когда эти плоскости пересекаются друг с другом под малыми углами.

С алгебраической точки зрения, плохая обусловленность системы означает, что главный определитель ее мал по сравнению с теми произведениями коэффициентов, алгебраической суммой которых он является. Поскольку с уменьшением расстояния  $|\vec{r}_i - \vec{r}_j|$  величины  $P_{ij}$  приближаются к единице, то, следовательно, чем гуще сеть станций, тем хуже обусловлена система.

Для плохообусловленных систем ошибки округления могут привести к абсурдным выводам. Этот эффект называют вычислительной неустойчивостью. Ее появление заставляет с осторожностью относиться к выбору метода решения системы (6.5). Наилучшим, по сравнению с методом простой итерации (последовательных приближений) или методом Зейделя, является классический метод Гаусса, как слабо чувствительный к плохой обусловленности систем. Но и в этом случае иногда появляется вычислительная неустойчивость. Однако, если учесть, что данные наблюдений известны с некоторой случайной ошибкой, то вычислительная неустойчивость не проявляется, поскольку учет наличия случайных ошибок наблюдений увеличивает обусловленность системы для определения весов. Рассмотрим его детальнее.

Как было показано выше, случайные ошибки не изменяют значений автокорреляционной (ненормированной) функции для разных точек и увеличивает значение автокорреляционной функции для совпадающих точек, т.е. дисперсии элемента, на средний квадрат ошибки наблюдений в этой точке

$$m_{\vec{r}}(\vec{r}_1, \vec{r}_2) = m_f(\vec{r}_1, \vec{r}_2); \quad (6.14)$$

$$\tilde{f}^{12}(\vec{r}) = m_f(\vec{r}_1, \vec{r}) = \tilde{f}^{12}(r) + \sigma_f^2(\vec{r}), \quad (6.15)$$

где  $f$  – истинное значение элемента,  $\tilde{f}$  – измеренное значение элемента,  $\sigma_f$  – средняя квадратическая ошибка наблюдений. Если разделить (6.14) на (6.15), то получим:

$$\alpha_{ij} = \begin{cases} P_{ij} & \text{при } \vec{r}_i \neq \vec{r}_j \\ 1 + \eta_j & \text{при } \vec{r}_i = \vec{r}_j \end{cases}, \quad (6.16)$$

где  $\eta_j$  – мера ошибок наблюдений за элементом  $f$ - в точке  $\vec{r}_j$ .

Тогда –

$$\varepsilon = 1 - 2 \sum_{i=1}^n R_i \cdot \alpha_{0i} + \sum_{i=2}^n \sum_{j=1}^n R_i \cdot R_j \cdot \alpha_{ij}. \quad (6.17)$$

Подставляя значения  $\alpha_{ij}$  преобразуем (6.17) к виду:

$$\varepsilon = 1 - 2 \sum R_i \cdot P_{ij} + \sum_{i=1}^n \sum_{j=1}^n R_i \cdot R_j \cdot P_{ij} + \sum_{i=1}^n R_i^2. \quad (6.18)$$

Потребуем снова, чтобы выполнилось условие минимума  $\varepsilon$  по  $R_i$ , тогда

$$\sum_{j=1}^n \mu_{ij} \cdot R_j + \eta_i \cdot R_i = P_{0i}. \quad (i=1, 2, \dots, n) \quad (6.19)$$

Последнее выражение служит для отыскания  $\rho_i$ . Если точность наблюдений одинакова, то

$$\sum_{j=1}^n P_{ij} \cdot R_j + \eta \cdot R_i = P_{0i}. \quad (6.20)$$

Помимо своего принципиального значения (учет наличия ошибок наблюдений) введение  $\eta$  играет и другую роль, т. е. с ростом  $\eta$  распределение весов  $R_i$  данных в различных точках сглаживания, а именно, сумма квадратов весов  $\sum_{i=1}^n R_i$  убывает. Это означает, что чем больше  $\eta$ , тем менее сказываются неточности в данных о корреляционной функции. Последнее особенно существенно применительно к густой сети станций. В этом случае недиагональные члены  $P_{ij}$  ( $i \neq j$ ) автокорреляционной матрицы оказываются близкими к диагональным  $P_{ii} = 0$ , так что система (6.5) оказывается плохообусловленной. Это означает, что ошибки округления, неизбежные при решении этой системы могут заметно исказить интерполяционные веса. Введение  $\eta$  т. е. переход к системе (6.19), уменьшает эту опасность, так как диагональные члены увеличиваются. Вышеизложенное неизбежно вытекает из условий влияния случайных ошибок на точность вычисления дисперсии и автокорреляционных функций. Существенно также, что введение  $\eta > 0$  сопровождается уменьшением количества отрицательных весов и уменьшением их абсолютных значений.

Нетрудно обобщить метод оптимальной интерполяции и на случай, когда исходной информацией являются значения в различных точках любых метеорологических элементов, а не только того элемента, значение которого в узле регулярной сетки нас интересует. Будем с этой целью обозначать через  $f'_{ki}$  значение отклонения от нормы  $k$ -го метеорологического элемента в  $i$ -ой точке и предположим, что нам нужно найти значение  $f'_0$  элемента  $f_0$  (его отклонения от нормы) в точке 0 по данным о значениях элемента  $f_1$  в  $n_1$  различных точках, элемента  $f_2$  в  $n_2$  различных точках и т. д. Будем аналогично предыдущему искать величину  $f'_{00}$  в виде линейной комбинации

$$\frac{\tilde{f}'_{00}}{\sqrt{D_{00}}} = \sum_{k=1}^m \sum_{i=1}^{n_k} R_{ki} \frac{\hat{f}'_{ki}}{\sqrt{D_{ki}}} = \sum_{k=1}^m \sum_{i=1}^{n_k} R_{ki} \frac{f'_{k0} + \delta_{ki}}{\sqrt{D_{ki}}}. \quad (6.21)$$

Здесь через  $m$  обозначено количество различных метеорологических элементов, информация о которых используется для получения значений  $\tilde{f}'_{00}$ . Один из этих может совпадать с элементом  $f_0$ . Равным образом точки, в которых известны различные элементы, могут совпадать полностью или частично, а могут и различаться. Количество таких точек  $n$  может быть разным для различных элементов. Через  $\delta_{ki}$  обозначена ошибка наблюдения над значением элемента  $f_k$  в соответствующей точке, так что  $\hat{f}'_{ki} = f_{ki} + \delta_{ki}$  представляет собой наблюдаемое значение, а  $\tilde{f}'_{ki} = f'_{ki} + \delta_{ki}$  — наблюдаемое отклонение от нормы. В формуле (6.21) для сокращения записи использовано краткое обозначение для дисперсий  $D$ , а именно,  $D_{ki}$  представляет собой дисперсию  $k$ -го элемента в  $i$ -ой (для данного элемента) точке, а  $\rho_{ki}$  — искомый вес значения  $\hat{f}'_{ki} / D_{ki}$  для  $k$ -того элемента в  $i$ -ой точке.

Тогда мера ошибки интерполяции  $\varepsilon = \frac{(\overline{f'_{00} - f_{00}})^2}{D_{00}}$  может быть, согласно формуле (6.21) записана в виде

$$\varepsilon = \sum_{k=1}^m \sum_{i=1}^{n_k} \sum_{j=1}^{n_l} R_{ki} \cdot R_{lj} \frac{(f'_{ki} + \delta_{ki}) \cdot (f'_{lj} + \delta_{lj})}{\sqrt{D_{ki} D_{lj}}} - 2 \sum_{k=1}^m \sum_{i=1}^{n_k} R_{ki} \frac{(f'_{ki} + \delta_{ki}) f'_{00}}{\sqrt{D_{ki} D_{00}}} + 1. \quad (6.22)$$



Будем по-прежнему считать ошибки  $\delta_{ki}$  случайными, добавляя к определению таких ошибок требование, чтобы ошибки наблюдений над различными метеорологическими элементами были некоррелированы, т. е. чтобы выполнялись равенства –

$$\delta_{ki} \cdot \delta_{lj} = 0 \text{ при } k \neq l. \quad (6.23)$$

Иначе говоря, предположим, что среднее произведение  $\delta_{ki} \delta_{lj}$  отлично от нуля только тогда, если одновременно  $k=l$  и  $i=j$ . В этом случае оно равно –

$$\delta_{ki}^2 = \eta_{ki} D_{ki}, \quad (6.24)$$

где  $\eta_{ki}$  – мера ошибки наблюдений.

Обозначив через

$$P_{klij} = \frac{f'_{ki} f'_{lj}}{\sqrt{D_{ki} D_{lj}}}, \quad (6.25)$$

с учетом всего сказанного можно переписать формулу (6.22) в виде

$$\varepsilon = \sum_{k=1}^m \sum_{l=1}^m \sum_{i=1}^{n_k} \sum_{j=1}^{n_l} P_{klij} R_{ki} R_{lj} + \sum_{k=1}^m \sum_{i=1}^{n_k} \eta_{ki} R_{ki}^2 - 2 \sum_{k=1}^m \sum_{l=1}^{n_k} P_{k0l0} R_{ki} + 1. \quad (6.26)$$

Условия минимума  $\varepsilon$  по всем  $R_{ki}$

$$\frac{\partial \varepsilon}{\partial p_{ki}} = 0, \quad (6.27)$$

приобретают, согласно (6.26) вид

$$\sum_{l=1}^m \sum_{j=1}^{n_l} P_{klij} \cdot R_{lj} + \eta_{ki} \cdot R_{ki} = P_{k0i0}, \quad (6.28)$$

и образуют систему

$$N = n_1 + n_2 + \dots + n_m = \sum_{k=1}^m n_k, \quad (6.29)$$

линейных алгебраических уравнений с таким же количеством неизвестных  $p_{ki}$ . После решения системы (6.28) можно выполнить интерполяцию по формуле (6.21). Кроме того, используя уравнения (6.28), можно записать выражение (6.26) для  $\varepsilon$  в более простом виде

$$\varepsilon = 1 - \sum_{k=1}^m \sum_{l=1}^m \sum_{j=1}^{n_k} \sum_{i=1}^{n_l} P_{kl ij} \cdot R_{ki} \cdot R_{lj} - \sum_{k=1}^m \sum_{i=1}^{n_k} \eta_{ki} \cdot R_{ki}^2, \quad (6.30)$$

либо, что удобнее,

$$\varepsilon = 1 - \sum_{k=1}^m \sum_{i=1}^{n_k} P_{k0i0} \cdot R_{ki}. \quad (6.31)$$

Эти формулы позволяют производить интерполяцию по данным о любом количестве метеорологических элементов. Однако практически реализация такой интерполяции связана с рядом трудностей.

Некоторые трудности связаны с высоким порядком системы (6.28). Если, например, используются данные о пяти метеорологических элементах в шести точках, то система (6.28) состоит из 30 уравнений. Необходимость составлять и решать такие системы для каждого узла регулярной сетки связана с серьезными требованиями к скорости и объему компьютерной памяти.

Более существенны, однако, трудности, возникающие вследствие необходимости знать все коэффициенты корреляции  $P_{kl ij}$ , входящие в системы вида (6.28).

В соответствии со сказанным выше эти коэффициенты естественно считать значениями однородных и изотропных корреляционных функций – автокорреляционных при  $k = l$  и взаимных корреляционных при  $k \neq l$ . Использование этой гипотезы, разумеется, упрощает дело. Однако, если автокорреляционные функции основных метеорологических элементов известны в настоящее время сравнительно хорошо, то о кросскорреляционных функциях информация весьма невелика.

Дополнительные затруднения связаны с необходимостью обеспечить положительную определенность матриц величин  $P_{kl ij}$ , входящих в системы вида (6.28). Для этого недостаточно, чтобы каждая из функций, автокорреляционных и кросскорреляционных, была положительно определенной; необходимо, чтобы этому требованию удовлетворяла совокупность используемых корреляционных функций. С этой целью можно, например, принять все корреляционные функции  $P_{kl ij}$  подобными, т. е. автокорреляционные функции – совпадающими, а кросскорреляционные – отличающимися от них лишь постоянными множителями.

Положительную определенность матриц обеспечивает и согласование, состоящее в том, что для получения значения отклонения от нормы некоторого элемента  $f_1$  в узле регулярной сетки используются данные наблюдений на станциях как над этим элементом, так и еще над одним элементом  $f_2$ . Формулы для этого случая получатся, очевидно, из (6.21) и (6.29) – (6.31), если положить  $m=2$  и заменить индекс 0 искомого элемента на индекс 1.

Тогда формула примет вид:

$$\frac{\tilde{f}'_{10}}{\sqrt{D^{10}}} = \sum_{k=i=1}^2 \sum_{k=1}^{n_k} R_{ki} \cdot \frac{\hat{f}'_{ki}}{\sqrt{D_{ki}}}; \quad (6.32)$$

$$\frac{\tilde{f}'_{10}}{\sqrt{D^{10}}} = \sum_{i=1}^{n_1} R_{1i} \cdot \frac{\hat{f}'_{ki}}{\sqrt{D_{1i}}} + \sum_{i=1}^{n_2} R_{2i} \cdot \frac{\hat{f}'_{ki}}{\sqrt{D_{2i}}}. \quad (6.33)$$

Соответственно этому уравнение (6.28), примет вид

$$\sum_{i=1}^2 \sum_{j=1}^{n_j} P_{kij} \cdot R_{ij} + \eta_{ki} \cdot R_{ki} = P_{klio}, \quad (6.34)$$

удобно разбить на две подсистемы – при  $K=1$  и при  $K=2$ . Тогда получим:

$$\sum_{j=1}^{n_1} P_{11ij} \cdot R_{1j} + \eta_{1i} \cdot R_{1i} + \sum_{j=1}^{n_2} P_{12ij} \cdot R_{2j} = P_{11i0}; \quad (6.35)$$

$$\sum_{j=1}^{n_1} P_{21ij} \cdot R_{1j} + \sum_{j=1}^{n_2} P_{22ij} \cdot R_{2j} + \eta_{2i} \cdot R_{2i} = P_{21i0}. \quad (6.36)$$

В этой системе коэффициенты корреляции  $P_{1ij}$  представляют собой значения автокорреляционной функции элемента  $f_1$ , коэффициенты  $P_{2ij}$  – значения автокорреляционной функции элемента  $f_2$ ,  $P_{12ij}$  и  $P_{21ij}$  значения взаимной корреляционной функции, описывающей статистическую связь полей элементов  $f_1$  и  $f_2$ .

Тогда формулу (6.31) для  $\varepsilon$  можно в случае оптимального согласования записать в виде:

$$\varepsilon = 1 - \sum_{i=1}^{n_1} P_{11i0} \cdot R_{1i} - \sum_{i=1}^{n_2} P_{21i0} \cdot R_{2i}. \quad (6.37)$$

Рассмотрим более частный случай оптимального согласования, использующий лишь одно значение «дополнительной» информации, а именно, значение элемента  $f_2$  в том самом узле сетки, где нас интересует значение  $f_1$ , в этом случае формула (6.32) принимает вид

$$\frac{f'_{10}}{\sqrt{D_{10}}} = \sum_{i=1}^{n_1} R_{1i} \cdot \frac{\tilde{f}'_{1i}}{\sqrt{D_{1i}}} + R_{20} \cdot \frac{\tilde{f}'_{20}}{\sqrt{D_{20}}}, \quad (6.38)$$

а система (6.33) состоит из  $n_1$  уравнений вида

$$\sum_{j=1} P_{11ij} \cdot R_{1j} + \eta_{1i} \cdot R_{1i} + P_{12i0} \cdot R_{20} = P_{11i0}, \quad (6.39)$$

и одного уравнения

$$\sum_{j=1}^{n_1} P_{210j} \cdot R_{1j} + (1 + \eta_{20}) \cdot P_{2i} = P_{2100}, \quad (6.40)$$

и соответственно

$$\varepsilon = 1 - \sum_{i=1}^{n_1} P_{11i0} \cdot R_{1i} - P_{2100} \cdot R_{20}. \quad (6.41)$$

Уравнение подсистемы (6.37) отличаются от системы уравнений для случая простой оптимальной интерполяции элемента  $f_1$

$$\sum_{j=1}^{n_1} P_{11ij} \cdot R'_{1j} + \eta_{1i} \cdot R'_{1i} = P_{11i0}, \quad (6.42)$$

лишь наличием последнего члена левой части. Предполагая, что взаимная корреляционная функция  $P_{12(r)}$  подобна автокорреляционной функции  $P_{11(r)}$  элемента  $f_1$ , т. е. что имеет место соотношение

$$P_{12i0} = \rho P_{11i0}, \quad (6.43)$$

где множитель  $\rho$  не зависит от  $i$  ( $\rho = \mu_{1200} \leq 1$ ),

подсистема (6.40) примет вид:

$$\sum_{j=1}^{n_1} P_{11ij} \cdot R_{1j} + \eta_{1i} \cdot R_{1i} = (1 - \rho \cdot R_{20}) \cdot P_{11i0}. \quad (6.44)$$

Сопоставление ее с системой (6.42) показывает, что веса  $P_{1i}$  отличаются от весов  $P'_{1i}$  простой оптимальной интерполяции лишь постоянным множителем

$$R_{li} = (1 - \rho \cdot R_{20}) \cdot R'_{li}. \quad (6.45)$$

Аналогичное соотношение имеет место и для меры ошибки согласования и простой интерполяции. Действительно, подставляя (6.43) и (6.45) в формулу (6.41) для меры ошибки согласования  $\varepsilon$ , получаем

$$\varepsilon = 1 - (1 - \rho \cdot R_{20}) \sum_{i=1}^{n_1} P_{1li0} \cdot R'_{li} - \rho \cdot R_{20}. \quad (6.46)$$

Но так как

$$\sum_{i=1}^{n_1} P_{1li0} \cdot R'_{li} = 1 - \varepsilon', \quad (6.47)$$

где  $\varepsilon'$  – мера ошибки простой оптимальной интерполяции, то

$$\varepsilon = (1 - \rho \cdot R_{20}) \cdot \varepsilon', \quad (6.48)$$

что и определяет формулу для веса дополнительной информации.

$$R_{20} = \frac{\rho \cdot \varepsilon'}{1 + \eta_{20} - \rho^2 \cdot (1 - \varepsilon')}. \quad (6.49)$$

Таким образом, веса и мера ошибки для данного случая согласования весьма просто выражаются через соответствующие характеристики оптимальной интерполяции, производимой без привлечения дополнительной информации. В частности, если дополнительная информация представляет собой значение того же элемента  $f_1$ , но определенное другим путем, то  $\rho = 1$ , и формулы (6.48) и (6.49) принимают простой вид

$$\varepsilon = (1 - R_{20}) \cdot \varepsilon', \quad R_{20} = \frac{\varepsilon'}{\varepsilon' + \eta_{20}}. \quad (6.50)$$

## 6.2. Особенности оценки гидрометеорологических полей методами их геофизического разложения

В некоторых случаях целесообразно вместо самих значений метеорологических элементов использование обобщенных параметров индексов атмосферной циркуляции, коэффициентов разложения в ряд по сферическим функциям, полиномов Чебышева и др.

### Полиномы Чебышева

Если некоторую функцию  $F(x)$ , заданную в равноотстоящих точках  $x=1, 2, 3 \dots n$ , требуется представить в виде полинома  $k$ -ой степени, то можно записать:

$$F(x) = A_0 + A_1 \cdot \psi_1 + A_2 \cdot \psi_2 + \dots + A_k \cdot \psi_k, \quad (6.51)$$

где  $A_0, A_1, A_2, \dots, A_n$  коэффициенты разложения функции.

Эти коэффициенты определяются формулами вида:

$$A_k = \frac{\sum_{i=1}^n F(x_i) \cdot \psi(x_i)}{\sum_{i=1}^n \psi_k^2(x_i)}, \quad (6.52)$$

и подбираются так, чтобы сумма ряда  $F(x)$  наилучшим образом отражала заданную функцию  $F(x)$  в смысле быстрой сходимости ряда.

Двухмерное поле (например, геопотенциал) может быть представлено следующим образом:

$$H(x_i, y_j) = \sum_k \sum_s A_{ks} \cdot \psi_k(x_i) \cdot \psi_s(y_j), \quad (6.53)$$

с определением коэффициентов  $A_{ks}$  определяются по формулам:

$$A_{ks} = \frac{\sum_{i=1}^n \sum_{j=1}^k H(x_i, y_j) \cdot \psi_k(x_i) \cdot \psi_s(y_j)}{\sum_{i=1}^n \psi_k^2(x_i) \cdot \sum_{j=1}^n \psi_s^2(y_j)}, \quad (6.54)$$

$$\sum_{i=1}^n \psi_k^2(x_i) = \frac{(k!)^2 \cdot (2k-1) \cdot n \cdot (n^2 - 2^2) \dots (n^2 - k^2)}{2^{2k} \cdot (1 \cdot 3 \cdot 5 \dots (2k+1)^2)}. \quad (6.55)$$

Если имеем дело с полем давления, то каждый коэффициент следует истолковать как индекс некоторой формоциркуляции. Действительно, если направим  $x$  на юг по меридиану, а ось  $y$  на восток вдоль параллели, то полиномы различного порядка будут выражать:

$\varphi_0(x)\varphi_0(y)$  – постоянная или фон поля;

$\varphi_1(x)\varphi_0(y)$  – значения возрастают к югу, оставаясь постоянными вдоль параллели; это чисто зональная циркуляция, а коэффициент  $A_{10}$  будет выражать ее интенсивность;

$\varphi_0(x)\varphi_1(y)$  – чисто меридиональная циркуляция, направленная с юга на север;



$\varphi_1(x)\varphi_1(y)$  – циркуляция типа «деформационного поля» и т. д.

Следует заметить, что последовательность полиномов удовлетворяет свойством ортогональности, т. е.  $\sum_{i=1}^n \varphi_k(x_i)\varphi_e(x_i) = 0$ .

Такая последовательность полиномов удовлетворяет так называемым числам Чебышева, некоторые из которых приведены ниже:

$$\left. \begin{aligned} \psi_0 &= 1; \psi_1 = x - \frac{n+1}{2}; \psi_2 = \psi_1^2 - \frac{n^2-1}{12}; \\ \psi_3 &= \psi_1^3 - \frac{1}{2} \cdot (3 \cdot n^2 - 7) \cdot \psi_1; \\ \psi_4 &= \psi_1^4 - \frac{1}{14} \cdot (3 \cdot n^2 - 13) \cdot \psi_1^2 + \frac{5}{360} \cdot (n^2 - 1) \cdot (n^2 - 9); \\ \psi_{k+1} &= \psi_1 \cdot \psi_k - \frac{k^2 \cdot (n^2 - k^2)}{4 \cdot (4 \cdot k^2 - 1)} \cdot \psi_{k-1}. \end{aligned} \right\} \quad (6.56)$$

#### *Разложение полей в ряд Фурье и по сферическим функциям*

При разложении в ряд Фурье поле представляется суммой гармоник различных волн и амплитуд

$$y(k) = \sum_{m=0}^M C_m \cdot \sin(m \cdot x + \varphi_m), \quad (6.57)$$

где  $m$  – круговые частоты (волновые числа), с которыми связаны периоды

$$T_m = \frac{2 \cdot \pi}{m}, \quad (6.58)$$

где  $U_m$  – начальная фаза.

Разложение может вестись как временных, так и пространственных полей. Пусть  $x$  – географическая широта, так что разложения (6.57) производятся по каждому из избранных кругов широты. Для каждой карты мы получим серию  $C_m$  и  $\varphi_m$ , зависящих от широты  $\varphi$ . Их можно разложить в ряды по каким-либо функциям. Можно в частности, вновь привести разложение по синусам вида  $\sin(n\varphi + \delta n)$ .

В этом случае получится двойной ряд Фурье (разложение будет вестись как по меридиану, так и по параллели). Однако вместо синусов в последнем случае используют полиномы Лежандра. Последние являются решением уравнения Лежандра

$$(1-x^2) \cdot y'' - 2 \cdot x \cdot y' + n \cdot (n+1) \cdot y = 0. \quad (6.59)$$

Полиномы представляют собой линейные комбинации кратных дуг. Полиномы Лежандра принято обозначать символами  $P_n^m$  или  $P_n^m(\cos \theta)$ , где  $\theta$  – полярное расстояние.

Если задана некоторая функция полярного расстояния и долготы  $(\theta, \lambda)$ , то она может быть разложена в ряд по сферическим функциям вида:

$$f(\theta, \lambda) = \sum_{n=0}^{\infty} A_0 \cdot P_n(\cos \theta) + \sum_{m=1}^n (A_m \cdot \cos(m \cdot \lambda) + B_m \cdot \sin(m \cdot \lambda)) \cdot P_n^m(\cos \theta). \quad (6.60)$$

Коэффициенты разложения  $A_{ik}$  и  $B_{ik}$  можно определить из системы заданных значений  $f(\beta, \lambda)$ .

Так как многочлен Лежандра  $P_n(x)$  имеет  $n$  вещественных корней, расположенных симметрично относительно  $\theta = \frac{\pi}{2}$ , т. е. относительно экватора, то этими корнями сфера делится на  $n+1$  широтную зону. Если ограничиться многочленом третьего порядка, то полярные расстояния будут делить сферу на 4 зоны. Эти первые члены каждой строки носят название зональных сферических функций. Вторые и последующие члены каждой строки суть произведения присоединенных функций Лежандра на сумму синусов и косинусов кратных дуг. Такого вида функция называется секторальными сферическими функциями. Сфера делится на  $2m+1$  секторов, имеющих ширину  $\frac{\pi}{m}$ , внутри которых функция имеет один и тот же знак.

Таким образом, сфера делится системой параллелей и меридианов на сферические трапеции. Рассматриваемые члены разложения представляют функцию в этих сферических трапециях и называются тессеральными. Если функция  $f(\theta, \lambda)$  известна в какой-либо точке  $\theta_i, \lambda_i$ , то подставив это значение в левую часть уравнения (6.60), введя  $\theta_i, \lambda_i$  и  $P$  и вычислив их, получим одно уравнение с  $n(n+2)-1$  неизвестными. Если

значения  $f(\theta, \lambda)$  определить в таком же количестве точек, то можно определить и все коэффициенты разложения.

Для прогнозов существенна и зависимость коэффициентов разложения от  $t$  до  $[A_n^m(t), B_n^m(t)]$  и тогда вся последовательность ежедневных карт аналитически может быть записана как:

$$P(\varphi, \lambda, t) = \sum_{n=m}^{10} \sum_{m=n}^4 [A_n^m(+) \cos m\lambda + B_n^m(+) \sin m\lambda] P_n^m(\varphi), \quad (6.61)$$

где  $m$  и  $n$  – целые числа характеризующие размеры волн.

Однако все выше рассмотренные типы разложения полей являются формальными, поскольку они искусственно разбиваются на слагаемые, которые могут иметь некоторый физический смысл, но чаще его иметь не будут.

Поэтому чаще применяются разложения, в которых отдельные составляющие определяются совокупностью изучаемых полей, т. е. естественными составляющими.

#### *Естественные ортогональные составляющие*

Рассмотрим метод аналитического представления последовательности метеорологических полей посредством естественных ортогональных составляющих, пользуясь разложением [Бочаров, 1971]. Практическое применение этого метода наиболее подробно дано в работах М. Л. Юдина и др. [Юдин и др. 1979].

Пусть на некотором участке земной поверхности или данной точке поверхности на различных уровнях дана последовательность некоторых полей. Это может быть либо набор синоптических, аэрологических или океанологических карт, либо вертикальные профили какого-то элемента, либо последовательности любых других геофизических полей.

Данная совокупность полей может быть представлена в виде некоторой функции  $F(t, x)$ , где  $t$  – время, принимающие дискретные значения ( $t=1, 2, \dots, m$ ), а  $x$  – параметр-координата ( $x=1, 2, \dots, n$ ). Таким образом, мы имеем таблицу значений функции  $F(t, x)$

$$\begin{array}{cccccc} F_{11} & F_{12} & F_{13} & \dots & F_{1n} \\ F_{21} & F_{22} & F_{23} & \dots & F_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ F_{m1} & F_{m2} & F_{m3} & \dots & F_{mn} \end{array} \quad (6.62)$$

Здесь первый индекс соответствует порядковому номеру  $t$ , второй – порядковому номеру  $x$ . Другими словами, первая строка характеризует значения функции  $F$  в точках с координатами  $x=1, 2, \dots, n$  в момент времени  $t=1$ , вторая строка – значения функции в тех же точках в момент времени  $t=2$  и т.д.

Представим функцию  $F(t, x)$  в виде следующего элементарного разложения:

$$F(t, x) = \sum_i T(t) \cdot X(x), \quad (6.63)$$

т. е. ищем разложение совокупности полей по некоторым функциям  $X(x)$ , зависящим только от  $x$ . Заранее они нам неизвестны. Весовые коэффициенты  $T(t)$  зависят только от  $t$ .

При определении функций  $X(x)$  и  $T(t)$  наложим условие, чтобы сумма квадратов ошибок разложения (6.64) по всем точкам данной совокупности достигла минимума

$$\Delta = \sum_i \sum_x [F(t, x) - T(t) \cdot X(x)]^2 = \min. \quad (6.64)$$

Приняв следующие обозначения

$F(t, x) = F_{ij}$  ( $i=1, 2, \dots, m$ );  $X(x) = X_j$  ( $j=1, 2, \dots, n$ );  $T(t) = T_i$ , имеем:

$$\Delta = \sum_i \sum_j F_{ij}^2 - 2 \cdot \sum_i \sum_j F_{ij} \cdot T_i \cdot X_j + \sum_i T_i^2 \cdot \sum_j X_j^2. \quad (6.65)$$

Из условий минимума  $\Delta$

$$\left. \begin{array}{l} \frac{\delta \Delta}{\delta T_i} = -2 \sum_j F_{ij} X_j + 2 T_i \sum_j X_j^2 = 0; \\ \frac{\delta \Delta}{\delta X_j} = -2 \sum_i F_{ij} T_i + 2 X_j \sum_i T_i^2 = 0. \end{array} \right\} \quad (6.66)$$

находим  $T_i$  и  $X_j$  –

$$\left. \begin{array}{l} \sum_j F_{ij} X_j = T_i \sum_j X_j^2; \\ \sum_i F_{ij} T_i = X_j \sum_i T_i^2. \end{array} \right\} \quad (6.67)$$

Так как –

$$T_i = \frac{\sum_k F_{ik} X_k}{\sum_k X_k^2}, \quad (6.68)$$

то внося это выражение в левую часть второго уравнения системы (6.67), найдем

$$\sum_k X_k \cdot \sum_i F_{ij} \cdot F_{ik} = X_j \cdot \sum_j X_j^2. \quad (6.69)$$

Обозначим

$$\frac{1}{m} \cdot \sum_i T_i^2 \cdot \sum_j X_j^2 = \lambda, \quad (6.70)$$

$$\frac{1}{m} \cdot \sum_i F_{ij} \cdot F_{ik} = A_{jk}, \quad (6.71)$$

где  $\lambda$  – неопределенный множитель.

Функции  $A_{jk}$  являются корреляционными моментами данной последовательности полей для точек  $x=j$  и  $x=k$ . Другими словами, это есть ненормированные взаимные корреляционные функции либо вида  $m_1(r_j, r_k)$ , если ведется разложение самих значений интересующего нас элемента, либо вида  $\mu(r_j, r_k)$ , если ведется разложение не самих полей, а полей аномалий (отклонений от норм).

Совокупность всех этих комбинаций образует квадратную и симметричную ковариационную матрицу следующего вида:

$$\begin{array}{cccc} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \dots & \dots & \dots & \dots \\ A_{n1} & A_{n2} & \dots & A_{nn} \end{array} \quad (6.72)$$

Диагональные элементы ковариационной матрицы являются средней суммой квадратов чисел каждого столбца матрицы (6.62). Если имеем дело с отклонениями от средней, то это будут дисперсии данной выборки. С учетом введенных обозначений система (6.68) может быть переписана таким образом:

$$\left. \begin{aligned} \sum A_{jn} \cdot X_k &= \lambda \cdot X_j \quad (j=1,2,\dots,n) \\ (A_{11} - \lambda) \cdot X_1 + A_{12} \cdot X_2 + \dots + A_{1n} \cdot X_n &= 0; \\ A_{21} \cdot X_1 + (A_{22} - \lambda) \cdot X_2 + \dots + A_{2n} \cdot X_n &= 0; \\ \dots &\dots \\ A_{n1} \cdot X_1 + A_{n2} \cdot X_2 + \dots + (A_{nn} - \lambda) \cdot X_n &= 0. \end{aligned} \right\} (6.73)$$

Система (6.73) будет иметь решение, отличное от нуля, в том случае, если ее детерминант равен нулю, т. е.

$$\begin{vmatrix} A_{11} - \lambda & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} - \lambda & \dots & A_{2n} \\ \dots & \dots & \dots & \dots \\ A_{n1} & A_{n2} & \dots & A_{nn} - \lambda \end{vmatrix} = 0 \quad (6.74)$$

В матричном виде

$$\{A - \lambda \cdot E\} = 0, \quad (6.75)$$

где  $E$  – единичная матрица.

Уравнение (6.75) есть характеристическое уравнение матрицы  $\{A\}$ , содержащее неизвестное  $\lambda$ . Если разложить определитель (6.74), то это уравнение примет вид алгебраического уравнения 1-го порядка относительно  $\lambda$ , т. е.

$$\lambda^n - P_{n-1} \cdot \lambda^{n-1} - P_{n-2} \cdot \lambda^{n-2} - \dots - P_1 \cdot \lambda - P_0 = 0. \quad (6.76)$$

Оно будет иметь  $n$  решений:  $\lambda_1, \lambda_2, \lambda_3, \dots, \lambda_n$ , которые называются собственными (характеристиками) числами матрицы  $\{A\}$ . Таким образом, первый этап задачи разложения полей по естественным составляющим сводится к вычислению ковариационной матрицы и нахождению ее собственных (характеристических) чисел, которые будем располагать в убывающем порядке. В силу симметричности матрицы  $\{A\}$  все собственные числа действительные и положительно определенные. Таким образом, система (6.72) имеет  $n$  решений, соответствующих собственным числам  $\lambda_1, \lambda_2, \lambda_n$ . Возвращаясь к двойной индексации при  $X$  и  $T$ , получим следующие  $n$  решений:



$$\begin{array}{cccccc}
 x_{11} & x_{12} & \dots & x_{1n} & \text{для } \lambda_1 & \\
 x_{21} & x_{22} & \dots & x_{2n} & \text{для } \lambda_2 & \\
 \dots & \dots & \dots & \dots & \dots & \\
 x_{n1} & x_{n2} & \dots & x_{nn} & \text{для } \lambda_m & 
 \end{array} \quad (6.77)$$

Каждое из этих решений называется собственным вектором ковариационной матрицы  $\{A\}$ . Для каждого собственного вектора  $X_n(x)$  можно определить соответствующую ему функцию времени  $T_n(t)$  по формуле (6.68), которую перепишем в следующем виде:

$$T_{ni} = \frac{\sum_k F_{ik} X_{nk}}{\sum_k X_{nk}^2} \quad (6.78)$$

Окончательное выражение для разложения  $F(x, t)$  будет

$$F(t, x) = \sum_n T_n(t) \cdot X_n(x) \quad (6.79)$$

Следовательно, задача сводится к нахождению ковариационной матрицы, собственных чисел и собственных векторов этой матрицы и определению для каждого собственного вектора значений функции  $T$  для каждого момента времени  $t$ . В результате мы получаем все необходимые параметры, чтобы образовать сумму (6.79), дающую разложение функции  $F(t, x)$

Из теории матриц следует, что собственные векторы симметричной матрицы образуют ортогональную систему, что очень важно при практических приложениях, т. е.

$$\sum_j X_{nj} \cdot X_{gj} = 0 \quad (n \neq g) \quad (6.80)$$

Функции  $T_n(t)$  можно интерпретировать как  $m$ -мерные векторы. Из формул (6.79) и (6.73) следует, что эти функции аналогично функциям  $X_n(x)$  также ортогональны.

При использовании всех членов разложения разложение является совершенно точным. Однако для большинства гидрометеорологических полей основная информация содержится в первых нескольких членах. Из предыдущего ясно, что математический аппарат метода разложения функций в ряды по естественным составляющим сводится к определению ковариационной матрицы и нахождения собственных чисел и собственных векторов матрицы, а коль так, то вновь возникает вопрос о обусловленно-

сти систем уравнений. Эффективным приемом улучшения обусловленности систем является исключение одного из сильно коррелирующих элементов совокупности. Впрочем надо избегать и того крайнего случая – слабой корреляционной связи между элементами случайной функции, т. е. в этом случае введение линейных комбинаций переменных не имеет смысла.

Важнейшим вопросом, возникающим при разложении по ЕОФ, является вопрос о необходимом объеме выборки. М. И. Юдин и др. [Юдин и др., 1979] предлагают путь статистического эксперимента. Оценки могут производиться на основании прямого сопоставления собственных векторов, полученных из различных выборок. Считается, что выборка объемом порядка 100 независимых данных достаточна для вычисления каждого элемента ковариационной матрицы. При этом достаточно точно вычисляется несколько первых собственных векторов.

Оценка количественной устойчивости ЕОФ состоит в степени близости двух единичных векторов  $l$  и  $y$   $\varphi$   $\sum_{k=1}^n y(x_k)\varphi(x_k) = \cos(y, \varphi)$ . Если угол между собственными векторами ковариационной матрицы, полученными из двух различных выборок достаточно мал, то можно говорить о близости этих векторов. Скалярное произведение двух  $n$  – мерных векторов рассматривается как коэффициент корреляции двух случайных величин. Близкими считаются вектора при условии, что величина  $Z$  в три раза и более превосходит значение  $\sigma_z = \frac{1}{\sqrt{n-3}}$

Практическое использование ЕОФ, определяемых статистическим путем, связано с вопросом об их устойчивости при переходе от одной выборки к другой. Если система ЕОФ теряет свою «оптимальность» при переходе к другим выборкам, то она не имеет никаких преимуществ, например, перед обычными ортогональными полиномами. Но, как показали исследования М. И. Юдина и др. [Юдин и др., 1979], устойчивость ЕОФ характерна как в пространстве, так и времени, т. е. первые собственные вектора, на которые приходится большая доля суммарной дисперсии, не испытывают существенных вариаций.

Следует отметить, что быстрая сходимость (первые члены описывают около 90% вариация дисперсии) разложений по ЕОФ создает воз-

возможность малопараметрического описания различных полей метеорологических элементов. Такое описание обладает в среднем минимальной ошибкой по сравнению с любой другой стилизацией, основанной на том же количестве параметров.

### 6.3. Методы оценки степени аномальности полей

В практике часто возникает необходимость оценить степень аномальности поля. Оценку аномальности поля можно вести исходя из отклонений значений элемента от средней в отдельных точках поля (аномалий) по повторяемости аномалий, по весу площади, занимаемой аномалией и т. д. Одной из распространенных, хотя и несколько формальных, характеристик является коэффициент аномальности поля, предложенный А. Н. Багровым:

$$A = \frac{1}{n} \cdot \sum_{i=1}^n \frac{(x_i(\varphi_i, \lambda_i) - \bar{X})^2}{\sigma_i^2}, \quad (6.81)$$

где  $X_i(\varphi_i, \lambda_i) - \bar{X}$  – аномалия;  $\sigma_i^2$  – дисперсия поля в  $i$ -той точке.

Предполагается, что число точек  $n$  достаточно велико и они расположены относительно равномерно по территории. В каждой реализации случайного поля величина  $X_i$  в координатах  $\varphi_i$  и  $\lambda_i$  принимает то или иное значение, в связи с чем возникает вопрос, насколько близки к нему значения  $x_i$  в окружающих точках. Полное статистическое описание одновременного поведения случайного поля требует задания многомерных функций распределения, что осложняет объективное решение задачи описания и интерполяции.

На практике ограничиваются более простыми характеристиками в виде ПКФ –  $R=f(\rho)$ . ПКФ принимает максимальное значение при  $\rho=0$ , с увеличением расстояния  $R=f(\rho)$  – убывает. При очень больших расстояниях между метео-, мониторинговыми станциями, связь между элементами практически отсутствует, приближаясь к нулю.

Алгоритм расчета ПКФ предусматривает проведение следующих операций [Волчек, 1988]:

1) составление матрицы гидрологической характеристики на  $k$  – станциях за  $N$  – лет:

$$\|M_{ij}\|, \text{ при } i = \overline{1, N_j}; j = \overline{1, k}; \quad (6.82)$$

2) расчет одноточечных моментов по всем рядам наблюдений-

а) среднее для каждого пункта

$$\bar{M}_j = \frac{1}{N_j} \cdot \sum_{i=1}^{N_j} M_{ij} ; \quad (6.83)$$

б) среднеквадратическое отклонение

$$\sigma_j = \sqrt{\frac{1}{N_j} \cdot \sum_{i=1}^{N_j} (M_{ij} - \bar{M}_j)^2} ; \quad (6.84)$$

в) среднеквадратическое отклонение выборочной средней арифметической

$$\bar{\sigma}_j = \sigma_j / \sqrt{N_j} ; \quad (6.85)$$

г) коэффициент вариации

$$C_{V_j} = \sigma_j / \bar{M}_j ; \quad (6.86)$$

д) среднеквадратическая ошибка коэффициента вариации

$$\sigma_{C_{V_j}} = \sqrt{\frac{1 + C_{V_j}^2}{2 \cdot N_j}} ; \quad (6.87)$$

3) оценка коэффициентов парной корреляции за совместный период наблюдений

$$r_{jk} = \frac{\sum_{i=1}^{n_{jk}} (x_{ik} - \bar{X}_k) \cdot (x_{ij} - \bar{X}_j)}{\sigma_k \cdot \sigma_j \cdot n_{jk}} ; \quad (6.88)$$

4) осреднение парных коэффициентов корреляции по градациям расстояний  $\Delta r$  (с учетом числа попаданий –  $k_j$  в  $j$ -градацию)

$$\bar{r}_j = \frac{\sum_{i=1}^{k_j} r_{ji} \cdot n_{ji}}{\sum_{i=1}^{k_j} n_{ji}} ; \quad (6.89)$$

5) аппроксимация ПКФ, по всей совокупности коэффициентов корреляции или по средневзвешенным значениям, зависимостями различного типа, выбор которых производится по критерию минимальной остаточной дисперсии и критерию Фишера

$$R(\rho) = \begin{cases} R(0) - \alpha \cdot \rho, \\ \exp(-\alpha \cdot \rho^\beta), \\ (1 + \alpha \cdot \rho) \cdot (\exp - \alpha \cdot \rho), \\ R(0) \cdot \exp(-\rho / \rho_0)^n, \\ \frac{\sin(\beta \cdot \rho)}{\beta \cdot \rho} \cdot \exp(-\alpha \cdot \rho), \\ \exp(-\alpha \cdot \rho) \cdot \cos(\beta \cdot \rho); \end{cases} \quad (6.90)$$

6) оценка ошибки взвешенных средних коэффициентов корреляции по соотношению

$$\sigma_{r_{jk}} = \frac{1 - r_j^2}{\sqrt{n_j - 1}}, \quad (6.91)$$

а для распределения величины:

$$Z_j = \frac{1}{2} \cdot \ln \frac{1 + \bar{r}_j}{1 - \bar{r}_j} + \frac{\bar{r}_j}{2 \cdot (n - 1)}, \quad (6.92)$$

оценка ошибки –

$$\sigma_{Z_{jk}} = 1 / \sqrt{n_j - 3}; \quad (6.93)$$

7) построение соответствующих доверительных интервалов при соответствующих уровнях доверительной вероятности

$$\bar{Z}(\rho_{jk}) - t_{1-P} \cdot \sigma_{Z_{jk}} < Z_{jk} < \bar{Z}(\rho_{jk}) + t_{1-P} \cdot \sigma_{Z_{jk}}, \quad (6.94)$$

где  $t_{1-P} = (Z_{jk} - \bar{Z}(\rho_{jk})) / \sigma_{Z_{jk}}$  – квантиль нормального распределения при заданной доверительной вероятности  $(1-P)$ ;

8) оценка пространственной однородности ПКФ выполняется на основе критериев согласия Колмогорова  $\chi^2$ ,  $\omega^2$ , по распределению Фишера или каким-либо другим способом.

#### 6.4. Оценка точности статистических характеристик структуры поля

Расчет статистических характеристик гидрометеорологических и других полей по ограниченному объему данных не позволяет получить генеральное значение, а дает лишь их оценки. Для практических целей важно знать точность, с которой получены оценки статистической структуры поля. Кроме того, вследствие неоднородности исходных данных и нестационарности материалов наблюдений, связанных с особенностями

измерений, условиями макро- и мезопроцессов (масштабы которых меньше пространственно-временного разрешения системы наблюдений) имеют место определенные неточности (систематические погрешности). Разделить ошибки за счет указанных факторов зачастую невозможно, и приходится их объединять под общим названием случайные ошибки. При измерении или вычислении величины в точке получается не истинное ее значение, а некоторая величина:

$$M'_k = M_k \pm \Delta_{M_k} . \quad (6.95)$$

Ошибка  $\Delta_{M_k}$  включает в себя систематическую погрешность  $\bar{\Delta}_{M_k}$ , одинаковую для всех измерений или вычислений в аналогичных условиях, и случайную величину  $\delta_{M_k}$ , которая может принимать различные значения

$$\Delta_{M_k} = \bar{\Delta}_{M_k} + \delta_{M_k} . \quad (6.96)$$

Погрешность в определении гидрометеорологических элементов составляет:

$$\bar{\sigma}_{M_k}^2 = (M'_k - \bar{M}_k)^2 = \sigma_{M_k}^2 + \Delta_k^2 = \sigma_{M_k}^2 \cdot (1 + \eta_{M_k}^2) , \quad (6.97)$$

где  $\Delta_k^2$  – дисперсия ошибок наблюдений;  $\eta_{M_k} = \Delta_k^2 / \sigma_{M_k}^2$  – мера случайных погрешностей в исходных данных.

Средняя величина, вычисленная по реальным данным, изменяется лишь при наличии систематической погрешности  $\bar{\Delta}_{M_k}$ . На дисперсию наличие систематических погрешностей не оказывает влияния, но она завышается на величину, равную дисперсии ошибок измерения. С коэффициентами корреляции –

$$r_{jk} = r_{jk} \cdot \sqrt{\frac{1 + \eta_{M_j}^2}{1 + \eta_{M_k}^2}} \quad (6.98)$$

дело обстоит иначе. Здесь наличие ошибок в исходных данных, как правило, приводит к их занижению. В случае однородных и изотропных полей естественно считать одинаковыми в разных точках и меры ошибок в исходных данных, следовательно,



$$R(\rho) = \frac{R(0)}{1 + \eta_M^2}, \quad (6.99)$$

где  $R(0)$  – экстраполированное значение эмпирической ПКФ до значения  $\rho=0$ . Отметим также, что при уменьшении расстояния  $\rho$ , корреляционная функция  $R(\rho)$  стремится не к 1, а к некоторой величине:

$$\bar{R}(0) = \frac{1}{1 + \eta_M^2}. \quad (6.100)$$

Исходя из этого, значения эмпирических функций на малых расстояниях могут быть использованы для оценки точности исходных данных. Величина  $R(0)$  получается путем экстраполяции  $R(\rho)$  при малых значениях аргумента. После этого можно получить:

$$\eta_M^2 = \frac{(1 - R(0))}{R(0)}. \quad (6.101)$$

В итоге, приведение эмпирической корреляционной функции к теоретической осуществляется по формуле

$$R(\rho) = \frac{\tilde{R}(\rho)}{R(0)}. \quad (6.102)$$

Наряду с систематическими погрешностями за счет ошибок наблюдений, возможны искажения характеристик статистической структуры полей элементов из-за недостаточных объемов выборок экспериментальных данных, а также из-за различной длины их статистических моментов. Поэтому определенные на эмпирическом материале значения ПКФ всегда имеют разброс, и перед их использованием приходится проводить сглаживание. Алгоритм сглаживания содержит в себе осреднение коэффициентов корреляции для всех пар точек, попавших в некоторую градацию по расстоянию. Полученные значения корреляционной функции сглаживаются для того, чтобы согласовать значения для различных градаций, что позволяет уточнить характеристики для некоторых, недостаточно статистически обеспеченных градаций. В большинстве случаев, довольно надежные результаты получаются при простом арифметическом осреднении коэффициентов корреляции. Но при достаточно высокой связности ( $r > 0,8$ ) необходимо учитывать несимметричность распределения выборочных коэффициентов корреляции, которое приводит к тому, что более высокие их значения являются более точными и должны

учитываться с большим весом. В этом случае предлагается осуществлять осреднение не самих коэффициентов корреляции, а величин  $Z$ , получаемых путем преобразования  $Z$ -Фишера (6.19). Распределение величин  $Z_j$ , как правило, существенно ближе к нормальному закону, и их осреднение является более оправданным. После осреднения величин  $Z_j$  по градациям расстояний соответствующие значения коэффициента корреляции получаются путем обратного преобразования:  $\bar{r} = th(\bar{Z})$ . Расчеты характеристик пространственной структуры связаны с учетом однородности и изотропности, особенно при использовании ПКФ для решения ряда практических задач [Рождественский, Чеботарев, 1975]. Если однородность и / или изотропность поля нарушаются, точность определения коэффициентов корреляции по ПКФ снижается, что, естественно, сказывается в целом на результатах расчетов. Наибольшее практическое распространение получил следующий подход к решению задачи оценки однородности и изотропности. С помощью каких-либо статистических критериев оценивается степень разброса точек на ПКФ. Если разброс точек невелик и удовлетворяет выбранному статистическому критерию, поле считается однородным и изотропным. Необходимым и достаточным условием однородности корреляционной функции в пределах рассматриваемого района по критерию Г. А. Алексеева является выполнение неравенства:

$$|z_{jk} - \bar{Z}(\rho_{jk})| \geq \sigma_{z_{jk}} \approx 31,7\% \text{ или } \geq 2 \cdot \sigma_{z_{jk}} \approx 4,6\%, \quad (6.103)$$

где 31,7 % и 4,6 % соответственно число случаев от общего числа  $C^2(p) = p \cdot (p-1)/2$  эмпирических значений. Другими словами, в пределах квантилей нормального распределения  $t=1, t=2$ , общее эмпирическое число превышений  $K_3$  должно быть теоретически (по нормальному закону распределения) равно числу превышений, т. е.

$$K_3(1) \approx 0,317 \cdot C_\rho^2 = 0,317 \cdot n \cdot (n-1)/2 ; \quad (6.104)$$

$$K_3(2) \approx 0,046 \cdot C_\rho^2 = 0,046 \cdot n \cdot (n-1)/2 , \quad (6.105)$$

где  $n$  – число пунктов наблюдений по территории.

При больших расхождениях между теоретическими и эмпирическими вероятностями «нулевая» гипотеза отвергается и признается аль-

тернативная гипотеза «неоднородности» эмпирической пространственной корреляционной функции. В этом случае поле рассматриваемого элемента должно быть уменьшено. Проверка однородности и изотропности при этом повторяется. Кроме того, в некоторых случаях исключаются пункты, вызывающие неоднородность статистической структуры полей элементов. Однако следует отметить, что достижение однородности и изотропности поля путем исключения части информации, если не проводится специальных исследований по поводу достоверности данных, не всегда правомерно, так как исключение данных приводит к утере ценной, возможно, реальной информации. Альтернативой может служить подход, который основан на выяснении генетических причин разброса точек и решении задач на случай неоднородного и анизотропного поля. ПКФ анизотропного поля зависят не только от расстояния между точками, но и от направления, их связывающего. На использовании этого условия разработана методика оценки изотропности путем анализа схем изокоррелят, построенных относительно различных пунктов – центров корреляции. Этот метод, хотя и наглядный, но трудоемкий, поскольку требует обработки большого объема информации вручную. В связи с этим, следует отметить более наглядный способ представления пространственной связности поля по развернутым пространственным корреляционным функциям (РПКФ). В соответствии с этим методом, осреднение коэффициентов корреляции осуществляется не только по грациям расстояний, но и в зависимости от направления между метеостанциями, которое отсчитывается от параллели или меридиана. Для получаемой таким образом трехмерной поверхности строятся линии равных уровней (изокорреляты). Если линии уровня РПКФ близки к окружностям, принимается гипотеза об изотропности поля; в общем случае по форме изокоррелят можно наглядно проследить направления большей и меньшей связности поля, т.е. проанализировать характер его анизотропности. В отличие от изокоррелят, построенных относительно отдельных станций, линии уровня РПКФ получают путем осреднения всех индивидуальных изокоррелят, в силу чего несущественные индивидуальные детали сглаживаются и четко проявляются основные закономерности анизотропности поля.

### 6.5. Особенности комплексного анализа статистической структуры гидрометеорологических полей

Рассмотрим сущность комплексного анализа статистической структуры поля на примере речного стока и атмосферных осадков Беларуси. Если случайное поле  $\xi(\rho)$  представлено в виде независимых составляющих  $\xi(\rho) = \eta(\rho) + \delta(\rho)$ , где  $\eta(\rho)$  – мелкомасштабная,  $\delta(\rho)$  – крупномасштабная составляющие, то корреляционная функция, обладающая свойством аддитивности по отношению к независимым составляющим поля, может быть представлена как  $R_{\xi}(\rho) = R_{\eta}(\rho) + R_{\delta}(\rho)$ . Для выявления соотношений между мелкомасштабной и крупномасштабной составляющими строятся эмпирические ПКФ по эмпирическим коэффициентам корреляции  $r_{jk}$  и соответствующим им расстояниям  $\rho$  между метео-, мониторинговыми станциями (центрами тяжести водосборов), которые аппроксимируются линейными зависимостями типа:

$$R(\rho) = R(0) - \alpha_r \cdot \rho, \quad (6.106)$$

характеризующими закономерность убывания эмпирических коэффициентов корреляции с увеличением расстояния между пунктами наблюдений. Величина  $R(0)$ , которой определяются значения ПКФ, при  $\rho=0$ , как правило, меньше единицы. Это обусловлено наличием в данных наблюдений случайных ошибок, а также микроклиматических –  $\eta(\rho)$  различий в расположении станций (бассейнов). Хотя эти различия в каждом пункте вызывают систематическое расхождение, при рассмотрении гидрометеорологического поля на большой территории они выступают как случайные. Именно этими различиями в основном и определяется имеющий место значительный разброс коэффициентов корреляции относительно средних величин. При отсутствии ошибок измерений и микроклиматических различий имело бы место –  $R(0)=1$ . В действительности, выполняется соотношение (6.101). Таким образом, при  $R(0)<1$  можно оценить, какая доля изменчивости поля определяется естественной изменчивостью рассматриваемых элементов на территории, а какая возникает за счет случайного размещения станций, погрешностей наблюдений;  $\alpha_r = dR(\rho)/d\rho$  – по физическому смыслу представляет градиент поля, т.е. показывает величину изменения ПКФ на единицу расстояния. Градиент поля  $\alpha_r$  служит характеристикой при совместном анализе и сопос-

тавлении различных гидрометеорологических и экологических полей. Приведение эмпирических ПКФ к теоретическим или, точнее, откорректированных функций осуществляется путем деления каждого члена уравнения (6.106) на  $R(0)$ . В результате чего ПКФ имеет вид:

$$\hat{R}(\rho) = 1 - \hat{\alpha}_r \cdot \rho, \quad (6.107)$$

где  $\hat{\alpha}_r$  – приведенный градиент ПКФ поля. Величины корреляционных функций  $R(0)$ , а также мер ошибок случайных погрешностей в исходных данных  $\eta_M$ , приведенного градиента ПКФ поля  $\hat{\alpha}_r$ , коэффициентов корреляции функций (6.106) стока и осадков одного из регионов  $r$  представлены в табл. 6.1

Таблица 6.1

**Экстраполированные значения основных характеристик ПКФ атмосферных осадков и речного стока в регионе для различных временных интервалов  $R_0$ ;  $\eta_M$ ;  $\hat{\alpha}_r$ ;  $\rho$**

Интервал осредне- ния	Атмосферные осадки				Речной сток			
	$R(0)$	$\eta_M$	$\hat{\alpha}_r \cdot 10^{-3}$	$r$	$R(0)$	$\eta_M$	$\hat{\alpha}_r \cdot 10^{-3}$	$r$
Январь	0,84	0,43	0,76	0,39	0,90	0,33	0,67	0,50
Февраль	0,76	0,56	0,70	0,36	0,83	0,44	0,54	0,42
Март	0,79	0,51	0,65	0,30	0,88	0,37	1,07	0,66
Апрель	0,80	0,50	1,40	0,69	0,82	0,46	1,03	0,56
Май	0,59	0,83	0,99	0,37	0,85	0,42	1,35	0,69
Июнь	0,59	0,83	0,90	0,43	0,73	0,60	1,56	0,62
Июль	0,66	0,71	1,32	0,50	0,80	0,50	1,59	0,70
Август	0,64	0,46	1,33	0,52	0,82	0,46	1,84	0,75
Сентябрь	0,82	0,40	1,01	0,52	0,83	0,44	1,85	0,79
Октябрь	0,86	0,42	0,91	0,42	0,90	0,33	1,82	0,77
Ноябрь	0,85	0,42	0,73	0,40	0,88	0,37	1,40	0,75
Декабрь	0,83	0,44	0,77	0,42	0,87	0,38	1,28	0,75
Год	0,71	0,64	1,08	0,48	0,84	0,43	1,61	0,79

Годовой цикл эмпирических ПКФ месячных значений атмосферных осадков и речного стока представлен на рис. 6.2.

Чтобы заведомо не упрощать картину принятием каких-либо гипотез о виде ПКФ, годовой ход представлен не по аппроксимирующим функциям, а изокоррелятами, полученными путем интерполяции эмпирических коэффициентов корреляции. Шаг интерполяции ПКФ – 0,1. В мезомасштабной области, для различных элементов водного баланса,



вклад крупномасштабной составляющей неодинаков. В метеорологических полях он значительно выше, чем в гидрологических. Поскольку коррелированность поля крупномасштабной составляющей выше, чем мелкомасштабной – значения  $R_{\xi}(\rho)$  завышаются для величин  $\rho$ , не превышающих средний масштаб крупных флуктуаций поля. Поэтому ПКФ элементов водного баланса для большинства интервалов временной дискретизации полей являются выгнутыми, подчиняющимися экспоненциально-степенной зависимости типа:

$$R(\rho) = \exp(-\alpha \cdot \rho^{\beta}) . \quad (6.108)$$

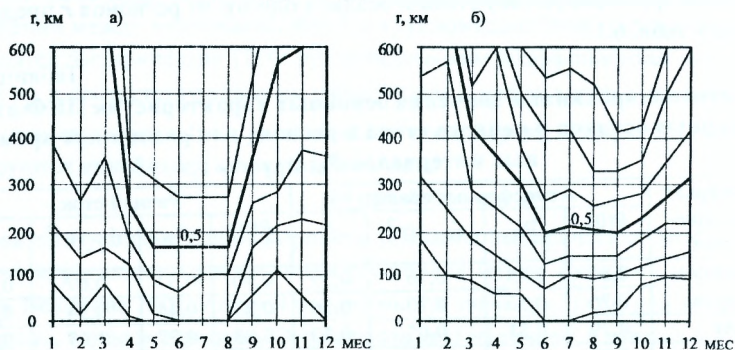


Рис. 6.2. Годовой цикл пространственных корреляционных функций: а) атмосферных осадков; б) речного стока (при шаге интерполяции – 0,1)

Приведенная зависимость одинаково хорошо описывает поля всех элементов водного баланса, в том числе атмосферных осадков и речного стока. Недостатком зависимостей типа (6.35) является то, что они неопределимы при отрицательных коэффициентах корреляции. Значения параметров, характеризующих статистическую структуру полей атмосферных осадков и речного стока региона для различных периодов осреднения, а также коэффициентов корреляции и критериев Фишера представлены в табл. 6.2.

Линии регрессии  $R(\rho)$  атмосферных осадков для всех месяцев года вогнуты, так что с увеличением расстояния между метеостанциями убывание связности полей осадков замедляется. Для всех месяцев имеет место неравенство  $R(0) < 1$ , которое увеличивается в летние месяцы, когда



роль локальной неоднородности растет за счет конвективных осадков. Наименьший разброс точек относительно эмпирической линии регрессии  $R(\rho)$  отличается в зимние месяцы. В пространственной коррелированности месячных сумм атмосферных осадков четко выражен годовой ход связности их полей в холодный период, которая существенно выше, чем в теплый период. Минимальная коррелированность наблюдается в летние месяцы, что вызвано несколько большей масштабностью конвективных процессов. Отметим и такую особенность, как некоторое увеличение разнородности осадков по территории в феврале и октябре за счет перехода от твердых осадков к жидким и наоборот.

Таблица 6.2

### Характеристика статистической структуры атмосферных осадков и речного стока региона

Интервал осреднения	Атмосферные осадки				Речной сток			
	$\alpha$	$\beta$	$r$	$F$	$\alpha$	$\beta$	$r$	$F$
Январь	0,03	0,47	0,94	10,5	0,02	0,53	0,95	14,1
Февраль	0,05	0,41	0,96	13,5	0,04	0,39	0,97	19,5
Март	0,07	0,34	0,87	4,17	0,01	0,67	0,97	12,7
Апрель	0,02	0,71	0,96	20,7	0,02	0,62	0,99	64,3
Май	0,09	0,40	0,97	13,5	0,01	0,76	0,98	24,4
Июнь	0,11	0,36	0,99	14,9	0,02	0,67	0,97	20,1
Июль	0,07	0,42	0,92	7,01	0,01	0,73	0,96	22,0
Август	0,07	0,48	0,95	11,7	0,01	0,91	0,96	20,7
Сентябрь	0,02	0,61	0,97	26,8	0,01	0,93	0,98	39,8
Октябрь	0,02	0,58	0,92	10,9	0,01	1,04	0,97	33,7
Ноябрь	0,02	0,50	0,94	10,7	0,01	0,83	0,98	45,3
Декабрь	0,03	0,49	0,96	17,0	0,01	0,70	0,96	18,4
Год	0,07	0,43	0,93	8,40	0,01	0,91	0,98	36,4

Проверка показала, что при доверительных вероятностях 68,3% и 95,4% пространственные корреляционные функции атмосферных осадков неоднородны. Анизотропность поля атмосферных осадков можно оценить с помощью коэффициентов анизотропности поля ( $\chi$ ), которые определяются путем деления градиента поля на градиент ориентированного поля. Годовой ход коэффициентов анизотропности поля атмосферных осадков представлен на рис. 6.3. Поля изокоррелят для каждого месяца имеют вид эллипсов, большая ось которых ориентирована в направлении преобладающего переноса воздушных масс.

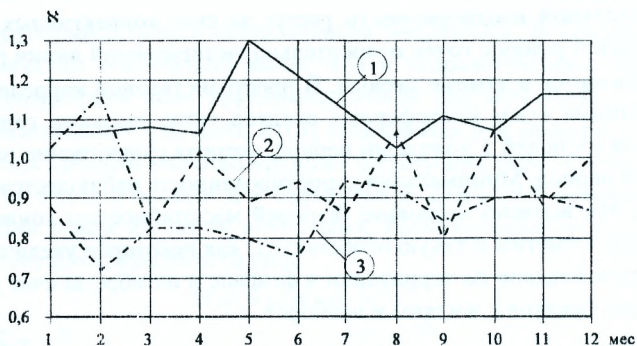


Рис. 6.3. Годовой ход анизотропии пространственной корреляции среднемесячных сумм атмосферных осадков Беларуси по направлениям: 1 – северо-восток – юго-запад; 2 – запад – восток; 3 – северо-запад – юго-восток

Так, в январе-феврале анизотропию поля сумм атмосферных осадков определяет западный перенос воздушных масс; в мае, июне, июле – северо-восточный. Наименьшая анизотропия полей осадков наблюдается в апреле, августе, ноябре. На рис. 6.4 показаны изокоррелянты сумм атмосферных осадков за май.

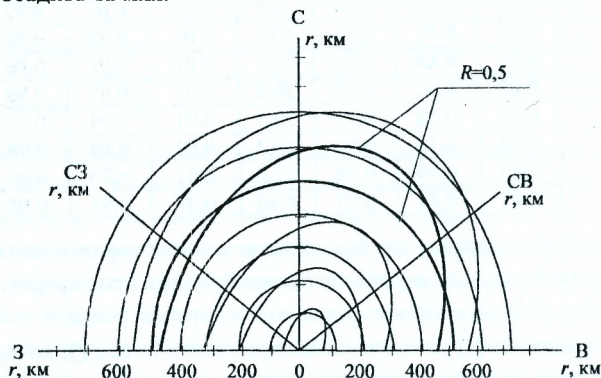


Рис. 6.4. Поле изокоррелянт атмосферных осадков Беларуси в мае (при шаге интерполяции – 0,1)

### 6.6. Особенности восстановления стоковых характеристик методом пространственной интерполяции

Под восстановлением стока имеется в виду пространственная интерполяция гидрологических величин, наблюдаемых в реперных пунктах, с целью их косвенного получения при отсутствии данных наблюдений в искомом пункте, а также корректировка в сторону уточнения измеренных данных при наличии в них антропогенных составляющих. Оптимальная пространственная интерполяция (ОПИ) сводится к следующему. Значение гидрологической величины в любой точке поля вычисляется как:

$$M_{i0} = \sum_{j=1}^k P_j \cdot M_{ij} , \quad (6.109)$$

где  $M_{ij}$  – значение элемента в  $i$ -ый срок на  $j$ -ый влияющий аналог;  $P_j$  – весовые коэффициенты;  $k$  – число влияющих рек-аналогов. Чаще уравнение (6.109) решается для отклонений от среднего, т. е. в виде:

$$M_{i0} = \overline{M}_0 + \sum_{j=1}^k P_j \cdot \Delta M_{ij} , \quad (6.110)$$

где  $\overline{M}_0$  – норма гидрологической характеристики в точке интерполяции;  $\Delta M_{ij}$  – отклонения от нормы на  $j$ -ом аналоге в  $i$ -ый срок.

После определения состава влияющих аналогов составляется система линейных уравнений для определения интерполяционных весов  $P_j$  по формуле (6.5). Количество влияющих рек-аналогов для каждого отдельного случая выбирается в зависимости от наличия данных по стоку за конкретный срок наблюдений, а также от расстояния между гидропостами, поэтому совокупность влияющих рек-аналогов не остается постоянной, а система (6.5) также индивидуальна для различных периодов. Реальное число гидропостов, привлекаемых к процедуре интерполяции, может меняться от двух до шести. Для получения результатов с требуемой для практики точностью обычно достаточно уже трех аналогов. После решения системы (6.5) интерполяцию по точкам поля можно проводить по уравнению (6.109). Для интерполяции по уравнению (6.110) необходимо определить норму гидрологической характеристики в заданной точке интерполяции одним из способов: по соответствующим картам; ОПИ по значениям норм. Выполненная интерполяция имеет среднюю квадратическую ошибку

$$\varepsilon^2 = 1 - \sum_{j=1}^k r_{0j} \cdot P_j . \quad (6.111)$$

По величине  $\varepsilon$  просто оценить возможную точность интерполяции, используя ее для поиска грубых ошибок в данных наблюдений. При анализе результатов интерполяции стока, можно обнаружить ошибки двух типов: ошибки в измерениях; ошибки при пространственной интерполяции, которые, в свою очередь, зависят от различий в условиях формирования стока и подбора аналогов. В большинстве случаев, близкое расположение гидропостов обеспечивает сходность физико-географических условий, высокие значения парных коэффициентов корреляции и хорошее совпадение вычисленных и измеренных величин. Среднеквадратическая ошибка оптимальной пространственной интерполяции (ОПИ) месячных величин стока, в среднем, составляет 10 – 20%, хотя в отдельные месяцы могут наблюдаться и большие отклонения. Годовые (сезонные) значения гидрологических характеристик определяются с большей достоверностью.

В качестве примера на рис. 6.5 и 6.6 приведены результаты интерполяции модулей стока за теплый период (IV...X – месяцы) реки Ница – створ Соколище (реки – аналоги: Дрисса – створ Демехи; Дрисса – створ Дерновичи; Свольня – створ Пользино) и за год р. Котра – с. Котра (реки-аналоги: Скиделька – ст. Скидель; Невища – ст. Половня; Свислочь – ст. Сухая Долина).

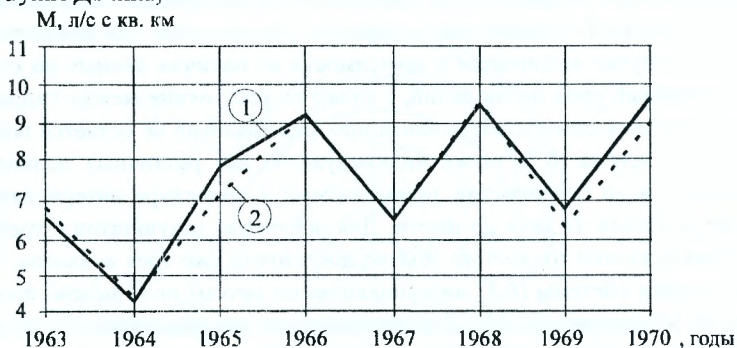


Рис. 6.5 Наблюдаемые (1) и вычисленные (2) значения стока р. Ница – ст. Соколище (теплый период)

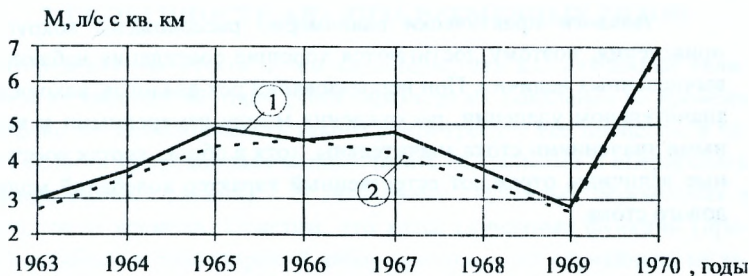


Рис. 6.6. Наблюдаемые (1) и вычисленные (2) значения годового стока р. Котра – ст. Котра

Суммарная относительная ошибка интерполяции и исходных данных составляет 5 – 10 процентов от средних величин модуля стока. При восстановлении месячных величин стока контроль осуществляется путем сопоставления суммы месячных величин (за год) с годовыми значениями. Если невязка получается меньше допустимой, ее распределяют помесячно пропорционально абсолютным величинам стока. В противном случае, требуется дополнительный анализ как исходной информации, так и репрезентативности аналогов. В табл. 6.3 представлены результаты интерполяции модулей норм годового стока реки Цны – створ Дятловичи при благоприятном выборе влияющих рек-аналогов (Ясельда – ст. Сенин; Бобринь – ст. Парохонск; Птичь – ст. Лучицы).

Таблица 6.3

**Наблюдаемые и вычисленные значения модулей годового стока  
р. Цна – ст. Дятловичи**

Характеристики	Месяцы					
	январь	февраль	март	апрель	май	июнь
Наблюдаемый сток	1,77	2,10	6,18	16,1	6,29	2,49
Вычисленный сток	2,09	2,18	6,01	14,0	6,98	2,83
Среднеквадратическая ошибка	0,18	0,02	0,03	0,3	0,11	0,14

→ продолжение таблицы 6.3

Характеристики	Месяцы					
	июль	август	сентябрь	октябрь	ноябрь	декабрь
Наблюдаемый сток	1,19	0,75	0,75	1,05	2,32	2,55
Вычисленный сток	1,73	1,08	1,05	1,00	2,53	3,04
Среднеквадратическая ошибка	0,45	0,24	0,17	0,02	0,09	0,19

Аналоги практически равномерно расположены вокруг пункта приведения, поэтому достигаются хорошие совпадения наблюдаемых и вычисленных величин. При использовании рек-аналогов, находящихся на значительном удалении, расхождения между измеренными и вычисленными значениями стока значительны, хотя в общих чертах восстановленные величины отражают естественный характер колебаний модулей годового стока.



## 7. ОСОБЕННОСТИ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ

Определяя расчетным путем характеристики как случайные величины, мы не касались методик учета характера их появления во времени и пространстве. Необходимость такого учета требует все процессы и явления в области природопользования считать случайными и оперировать только временными рядами, т. е. значениями случайных переменных в зависимости от времени. Отметим, что класс случайных функций (процессов) включает в себя широкий набор процессов (от случайных, не зависящих от предшествующих значений, до строго периодических, легко предсказываемых). По числу реализаций их разделяют на многомерные и одномерные, а по зависимости их характеристик от времени – стационарные, в которых средняя амплитуда и характер колебаний  $X(t)$  существенно не изменяются во времени, и нестационарные – имеющие определенную тенденцию изменяться во времени. С известным приближением, на некоторых временных отрезках, нестационарные процессы могут считаться стационарными. Например, если нестационарность функции вызвана изменяющимся во времени математическим ожиданием –  $m_x(t)$ , то процесс легко может быть приравнен к стационарному преобразованием типа:

$$\left. \begin{aligned} x(t) - m_x(t) &= \Delta_x(t); \\ x(t) / m_x(t) &= k(t), \end{aligned} \right\} \quad (7.1)$$

приводящим к математическому ожиданию, соответственно равному нулю или единице. Случайный процесс называется стационарным, в широком смысле, если его математическое ожидание не зависит от времени  $t$ , а корреляционная функция –  $r_x(t)$  является только функцией интервала времени:

$$m_x(t) = m_x = \text{const}; \quad r_x(t, t + \tau) = r_x. \quad (7.2)$$

По множеству реализаций наличия (отсутствия) связей между характеристиками они бывают эргодические и неэргодические. Эргодическое свойство стационарного процесса состоит в том, что каждая отдельная выборка может характеризовать всю генеральную совокупность. Среднее (по времени) и другие характеристики случайного процесса можно приближенно определять по одной, достаточно длинной выборке. В качестве формального признака используется оценка затухания корре-

ляции до нуля (если автокорреляционная функция  $r(t)$ , при  $t \rightarrow \infty$ , стремится к нулю, – процесс является эргодическим). Если автокорреляция при увеличении интервалов между сечениями  $t_2, t_1$  стремятся к некоторой (отличной от нуля) величине, то такая функция не является эргодичной. Эргодичность и стационарность – это два различных свойства случайных процессов. Каждый эргодический процесс является стационарным, тогда как стационарный процесс не обязательно должен быть эргодическим. Благодаря свойству эргодичности стационарных процессов, отпадает необходимость исследовать большое число выборок, данные о которых, как правило, отсутствуют, но достаточно исследовать одну выборку в течение длительного периода. Это свойство некоторых случайных процессов, как и свойство стационарности, существенно облегчает решение как гидрометеорологических задач, так и любых других задач рационального природопользования. Поскольку практические наблюдения гидрометеорологических и других природных явлений все же ограничены во времени, можно признать справедливым равенство во времени среднего значения множества выборок и среднего значения лишь с некоторой степенью достоверности [Жуковская, Мучник, 1976].

Тип закона распределения случайной функции в каждом сечении может быть любым.

По типу корреляционных функций следует различать процессы без последствий (непереходные, при  $\tau_{\max} > 1$ ) и процессы с последствием – переходные. Случайные процессы, в которых корреляционные связи ( $r(t, t') \neq 0$ ), при ( $t=t'$ ), называются цепями Маркова, при этом, если для каждого момента времени вероятность любого состояния системы в будущем зависит только от состояния системы в настоящий момент ( $t_0$ ) и не зависит от предшествующих значений, математическая модель процесса называется простой цепью Маркова [Алехин, 1963]. В таком процессе автокорреляция имеет место только между смежными членами ряда:

$$r(t, t+1) = r(\tau=1) \neq 0. \quad (7.3)$$

В обратном случае последовательность называется сложной цепью Маркова.

По характеру аналитического описания они могут быть с наличием или отсутствием тренда и в общем виде запись процесса  $Z(t)$  описывается зависимостью:

– при наличии трендовой составляющей

$$Z(t) = X_0(t) + X_T(t) + X_U(t) + \xi(t) , \quad (7.4)$$

– при отсутствии тренда

$$Z(t) = X_0(t) + \xi(t) , \quad (7.5)$$

где  $X_0(t)$  – постоянная (средняя);  $X_T(t)$  – трендовая;  $X_U(t)$  – циклическая;  $\xi(t)$  – случайная (шумовая) независимая составляющие.

В зависимости от класса стационарного эргодического процесса более развернутое аналитическое представление для одномерных стационарных временных процессов (рядов) может записываться следующими соотношениями:

$$X(t) = A \cdot \cos(\omega \cdot t) + \xi(t) ; \quad (7.6)$$

$$X(t) = \xi(t) + \alpha \cdot \xi(t-1) ; \quad (7.7)$$

$$X(t) = \beta \cdot X(t-1) + \xi(t) ; \quad (7.8)$$

$$X(t) = P(t) + q(t) \cdot \xi(t) , \quad (7.9)$$

в которых  $P(t)$  и  $q(t)$  – полиномы по  $t$ . Уравнение (7.6) описывает линейный циклический процесс, у которого  $A$  – амплитуда,  $\omega$  – основная угловая частота. С его помощью, в частности, можно было бы описать годовой ход температуры воздуха, сглаженных сумм атмосферных осадков или суммарного испарения с поверхности почвы. Уравнение (7.7) описывает процессы скользящего осреднения, где  $\alpha$  – весовой коэффициент, учитывающий предысторию. Уравнение (7.8) характеризует простой авторегрессионный процесс, называемый также марковским процессом 1-го порядка или простой инерционностью. Здесь настоящее значение  $X(t)$  определяется предыдущим  $X(t-1)$ . Независимо от ожидаемого типа колебаний, в целях физической интерпретации, ряд обычно раскладывается на сезонную, циклическую компоненты, компоненту тренда и случайный остаток. В связи с этим анализ временных рядов целесообразно проводить поэтапно: 1) выделение периодических, регулярных и сезонных циклов (годовой, сезонный, суточный ход); 2) выделение нерегулярных циклов (тренд, непериодические, квази-периодические составляющие); 3) сглаживание и фильтрация отдельных частот; 4) проверка на случайность

колебаний; 5) анализ однородности колебаний во времени и в пространстве; 6) прогноз колебаний.

Практика анализа временных рядов показывает, что общий ход или колебание любой случайной величины во времени представляет собой сумму нескольких колебаний. Некоторые из них носят периодический характер, другие – непериодический. Поэтому анализ любых временных рядов может проводиться как с целью разделения их на периодические и непериодические компоненты, так и для изучения каждой из компонент в отдельности.

### **7.1. Особенности анализа периодических колебаний**

#### *а) Выделение регулярных циклов*

К периодическим колебаниям относятся годовой и суточный ход гидролого-климатических, тепловоднобалансовых и других характеристик. Годовой ход таких характеристик лучше всего анализировать вычислив средние значения для каждого месяца или сезона и представив их в графическом виде в зависимости от времени. Так же можно представить и средние суточные величины, но подобный ряд, безусловно, даст случайные колебания, обусловленные короткопериодическими нерегулярными изменениями. Суточный ход обычно анализируется либо по результатам наблюдений в течение только одного месяца, либо по средним величинам за весь год, годовой ход анализируется – по результатам наблюдений в один и тот же час суток – по средним суточным данным.

#### *б) Гармонический анализ*

Гармонический анализ получил наибольшее распространение при исследовании периодического хода гидрометеорологических параметров. В задачу гармонического анализа входит разложение функций на простейшие периодические. Такой анализ позволяет понять физическую сущность периодических колебаний. Исходя из основных принципов математического анализа, любую функцию, заданную в каждой точке интервала, можно представить бесконечным рядом синусоидальных и косинусоидальных функций (рядом Фурье). Если в анализируемом ряду имеется конечное число точек, то все результаты наблюдений могут быть выражены и представлены конечным числом (суммой) синусов и косинусов. Первая (основная) гармоника имеет период, равный длине всего ис-

следуемого периода. Вторая гармоника имеет период, равный половине основного периода, третья – период, равный третьей части основного и т. д. Вообще, если число наблюдений ( $n$ ), то число гармоник ( $n/2$ ). Различные гармоники выделяются таким образом, что каждую из них можно рассматривать как независимый объект и объяснять разными физическими причинами. Часто ход характеристики не может быть объяснен полно, в то время как отдельные гармоники этому объяснению поддаются. Однако в каждую гармонику в отдельности, не обязательно вкладывать отчетливый физический смысл. Так будет всякий раз, когда периодическая функция не носит синусоидального характера. Не всегда определяются все ( $n/2$ ) – гармоник. Обычно изменение периодической функции достаточно хорошо описывается первыми двумя или, в крайнем случае, – тремя гармониками. Но в случае непериодических функций дело обстоит совершенно иначе. Полную сумму, задающую случайную переменную  $X(t)$ , можно записать в виде:

$$X(t) = \bar{X} + \sum_{i=1}^{n/2} \left[ A_i \cdot \sin\left(\frac{2 \cdot \pi}{P} \cdot i \cdot t\right) + B_i \cdot \cos\left(\frac{2 \cdot \pi}{P} \cdot i \cdot t\right) \right]. \quad (7.10)$$

Временная сумма есть среднее плюс сумма всех  $n/2$  – гармоник. Здесь  $P$  – основной период или полный период периодической функции;  $i$  – номер гармоники. Следует отметить, что  $P$  не всегда равно  $n$ . Если наблюдения производить каждые два часа одного дня, то  $n=12$ , а  $P=24$ . Первые два члена суммы (7.10) проходят полный цикл за один основной период, а быстрее всего меняется последний член, имеющий период  $2 \cdot P/n$ , и если существуют какие-либо более короткие периоды, то их можно обнаружить только на основании более частых наблюдений.

Гармонический анализ начинается с нахождения коэффициентов приведенного ряда, каждый из которых может быть вычислен независимо от другого по следующим формулам:

$$A_i = \frac{2}{n} \cdot \sum_{t=1}^n X(t) \cdot \sin\left(\frac{2 \cdot \pi}{P} \cdot i \cdot t\right); \quad (7.11)$$

$$B_i = \frac{2}{n} \cdot \sum_{t=1}^n X(t) \cdot \cos\left(\frac{2 \cdot \pi}{P} \cdot i \cdot t\right), \quad (7.12)$$

где  $i$  – может иметь любое целое значение от 1 до  $n/2-1$ . Для последней гармоники  $A_n=0$ ,  $B_n$  – равно величине, получаемой из (7.12), но деленной на 2, т. е.  $B_n/2$ .



Далее рассчитываются значения амплитуды  $i$ -ой гармоники и фазы (времени, при котором  $i$ -ая гармоника имеет максимум —  $\varphi_i$ ):

$$C_i = \sqrt{A_i^2 + B_i^2} \quad ; \quad \varphi_i = \frac{P}{2 \cdot \pi \cdot i} \cdot \operatorname{arctg} \left( \frac{A_i}{B_i} \right). \quad (7.13)$$

В связи с тем, что  $\operatorname{arctg}(A_i / B_i)$  имеет два значения (в интервале от 0 до  $2 \cdot \pi$ ), для выбора правильного решения используется дополнительное условие:

$$\varphi_i = \frac{P}{2 \cdot \pi \cdot i} \cdot \operatorname{arcsin}(A_i / C_i). \quad (7.14)$$

Каждая гармоника учитывает некоторую часть полной дисперсии  $\sigma(t)$ . Если несколько первых гармоник учитывают значительную часть полной дисперсии  $\sigma(t)$ , то дальнейшие расчеты нецелесообразны. Дисперсия за счет  $i$ -ой гармоники равна  $C_i^2 / 2$  для всех гармоник, за исключением последней, для которой она равна  $C_i^2$ . Часть дисперсии, учитываемая  $i$ -ой гармоникой, может быть представлена в виде отношения величины  $C_i^2 / 2$  или  $C_i^2$  к полной дисперсии  $\sigma_x^2$ . Поскольку никакие две гармоники не будут учитывать одну и ту же часть дисперсии  $\sigma(t)$ , то дисперсии, учитываемые различными гармониками, можно складывать.

#### *в) Исключение регулярных циклов*

С целью анализа нерегулярных колебаний временного ряда необходимо после выделения и анализа регулярных циклов исключить последние, т.е. вычесть их из исходных данных так, чтобы можно было проанализировать оставшийся временной ряд. Если период регулярного цикла короче, чем предполагаемые периоды нерегулярных колебаний, то для исключения регулярного цикла можно использовать только результат наблюдения в одной и той же точке цикла или использовать средние из всех результатов наблюдений за полный регулярный цикл. Так, для исключения суточного хода температуры воздуха можно использовать только результат наблюдения в конкретный час каждых суток (например, в полдень) или только средние суточные температуры. Если период регулярного цикла длиннее, чем период нерегулярных колебаний, то результат каждого наблюдения может быть выражен как отклонение от средне-



го (нормы). Например, если временной ряд состоит из средних месячных температур воздуха, то каждое значение можно заменить разностью между средней месячной температурой и климатической нормой температуры для того же месяца.

В качестве примера выполним гармонический анализ временного ряда средних месячных температур воздуха. Исходный ряд представлен в табл. 7.1.

Таблица 7.1

**Среднемесячные температуры воздуха по данным метеостанции**

Месяц	$t$ °С	Месяц	$t$ °С	Месяц	$t$ °С
I	-4,7	V	13,6	IX	13,1
II	-3,8	VI	16,9	X	7,7
III	0,4	VII	18,4	XI	2,6
IV	7,3	VIII	17,4	XII	-2,0

1). В зависимости от  $n$  и  $P$  (для всех значений « $a$ » и « $b$ ») составляется таблица множителей ( $A_i$ ,  $B_i$ ) (табл. 7.2):  $\left[ \frac{2}{n} \cdot \left( \sin \frac{2 \cdot \pi}{P} \cdot i \cdot t \right) \right]$  и

$\left[ \frac{2}{n} \cdot \left( \cos \frac{2 \cdot \pi}{P} \cdot i \cdot t \right) \right]$ . Для последней гармоники  $i=n/2$  множители определяются как:  $\left[ \frac{1}{n} \cdot \left( \cos \frac{2 \cdot \pi}{P} \cdot i \cdot t \right) \right]$ .

2). Значения анализируемого ряда (табл. 7.1) умножаются на соответствующие множители (табл. 7.2) и произведения (в столбцах) суммируются. Суммы (по столбцам) являются коэффициентами Фурье. Далее, по (7.13), определяются амплитуда  $C_i$  и дисперсия гармонического колебания  $i$ , равная  $\sigma_i = C_i^2 / 2$ ; результаты расчетов вносятся в табл. 7.3.

3) Поскольку  $S^2=67,85$ , доля дисперсии, учитываемая первой гармоникой, определяется как:  $\frac{C_1^2}{2 \cdot S^2} = 0,9069 \approx 91\%$ , второй гармоникой

как  $-\frac{C_2^2}{2 \cdot S^2} = 0,00319 \approx 0,3\%$ .



Таблица 7.3

**Результаты гармонического анализа средних месячных температур воздуха по данным метеостанции**

$i$	$A_i$	$B_i$	$C_i$	$\varphi_i$	$C_i^2/2$
1	-6,2419	-9,1708	11,0936	-1,11413	61,5335
2	-0,5628	0,3417	0,6584	-0,9789	0,2167
3	0,1033	0,3167	0,3332	0,2007	0,0555
4	-0,1587	4,1441	1,1550	-0,0658	0,6670
5	-0,0081	0,0374	0,0382	-0,0820	0,0007
6	--	0,0083	0,0083	0,0000	0,0000

Как и следовало ожидать, первая гармоника, описывающая годовой ход температуры воздуха, учитывает большую часть общей дисперсии. Взятые вместе две первые гармоники описывают 95% суммарного изменения температуры. Вычислять дальнейшие гармоники нет необходимости.

### 7.2. Особенности выделения и анализ нерегулярных циклов

Остаточные члены временного ряда после исключения из исходного ряда периодов регулярных колебаний не представляют собой явно выраженных циклов. Для такого временного ряда обычно характерно несколько типов колебаний:

- короткопериодные флуктуации настолько малого временного масштаба, что они проходят за половину (меньше половины) периода между, следующими друг за другом наблюдениями. Такие циклы не могут быть изучены из-за недостаточной частоты наблюдений; их влияние может быть в значительной степени исключено с помощью сглаживающих, низкочастотных фильтров;

- медленное, постепенное изменение случайной переменной в течение всего анализируемого периода, называемое трендом; тренд никогда не длится бесконечно, а скорее является частью колебаний с периодами, длительность которых сравнима с периодом наблюдений (тренды эффективно исключаются с помощью высокочастотной фильтрации);

- нерегулярные колебания, характеризующиеся промежуточным временным масштабом.

Рассмотрим основные методы выделения и анализа выявленных нерегулярных циклов.

## а) метод интегрально-разностных кривых

Метод интегральных или интегрально-разностных кривых базируется на интегральном ряде, члены которого равны  $(x_1), (x_1 + x_2), (x_1 + x_2 + x_3), \dots, (x_1 + x_2 + \dots + x_n)$ , или интегрально-разностным рядом с членами  $(d_1), (d_1 + d_2), (d_1 + d_2 + d_3), \dots, (d_1 + d_2 + \dots + d_n)$ , при этом  $d_i = x_i - \bar{x}$ .

Интегрирование (суммирование) периодических или циклических колебаний увеличивает амплитуды подобных колебаний пропорционально длине соответствующего периода и облегчает выделение низкочастотных колебаний. Метод интегрально-разностных кривых может быть рекомендован, когда важны кумулятивные свойства ряда, например, при интерпретации рядов осадков, стока, суммарного испарения и т. д. Однако, как и в случае скользящего осреднения, при этих методах, вследствие увеличения внутрирядной связности суммируемых последовательностей  $x_i, d_i$ , может возникать ложная цикличность. Кроме того, даже при исходном бессвязном ряде в середине интегрально-разностного ряда дисперсия его членов будет наибольшей, и наложение случайных ошибок может увеличивать амплитуду случайных колебаний в этой части ряда.

На рис. 7.1 представлены хронологические функции скользящих средних годовых расходов воды р. Припять – г. Мозырь.

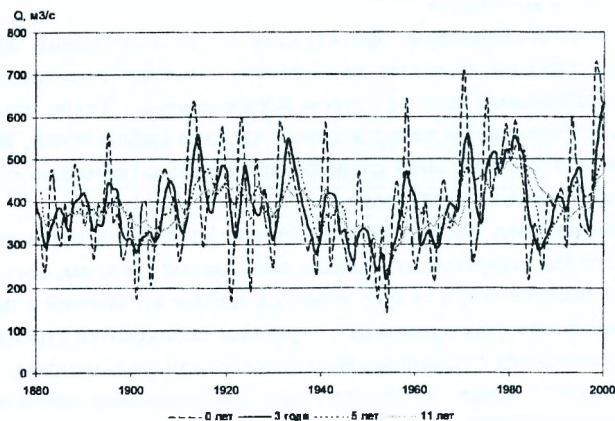


Рис. 7.1. Хронологические функции скользящих средних годовых расходов воды р. Припять – г. Мозырь

## б) Выделение тренда

Поскольку тренды используются для оценки тенденции будущих значений ряда, исключение тренда - одна из основных задач анализа нерегулярных колебаний. Обычно, тренд выделяется по методу наименьших квадратов способом скользящей средней или по определенной, характерной для данного ряда аналитической формуле, т. е. находятся средние  $\bar{x}$  и соответствующие отклонения аномалий  $d$  от среднего  $\bar{x}$  или уровня аналитической кривой  $x(t)$ . Алгебраически можно показать, что при переходе от уровней к их разностям  $d$  исключается влияние колебаний ряда. Если тренд ряда может быть представлен прямой  $\bar{y}_t = \alpha + b \cdot t$ , то, обозначая последовательные моменты времени через  $(t_1, t_2, \dots, t_n)$ , получим для:

$$\left. \begin{aligned} t = 1 - & y_1 = \alpha + b + d_{y_1}; \\ t = 2 - & y_2 = \alpha + 2 \cdot b + d_{y_2}; \\ & \dots \dots \dots \dots \dots \dots \dots \dots \\ t = n - & y_n = \alpha + n \cdot b + d_{y_n}. \end{aligned} \right\} \quad (7.15)$$

Найдем первые разности -

$$\left. \begin{aligned} \Delta'_1 = y_2 - y_1 &= b + (d_{y_2} - d_{y_1}); \\ \Delta'_2 = y_3 - y_2 &= b + (d_{y_3} - d_{y_2}); \\ & \dots \dots \dots \dots \dots \dots \dots \dots \\ \Delta'_n = y_n - y_{n-1} &= b + (d_{y_n} - d_{y_{(n-1)}}). \end{aligned} \right\} \quad (7.16)$$

Так как во всех разностях присутствует одна и та же константа  $b$ , то очевидно, что колебания рассчитанных разностей  $\Delta$  зависят только от  $d_y$ , т. е. влияние общей тенденции (тренда) механически исключается. Для случая, когда исходные данные изменяются по параболе 2-го порядка  $\bar{y}_t = \alpha + b + c \cdot t^2$ , получим, при  $(t_1 = y_1), (t_1 = \alpha + b + c + d_{y_1})$ , при  $(t_2 = y_2), (t_2 = \alpha + 2 \cdot b + 4 \cdot c + d_{y_2})$  и т. д. Находим первые разности:

$$\left. \begin{aligned} \Delta'_1 &= b + 3 \cdot c + (d_{y_2} - d_{y_1}); \\ \Delta'_2 &= b + 5 \cdot c + (d_{y_3} - d_{y_2}); \\ & \dots \dots \dots \dots \dots \dots \dots \dots \\ \Delta'_n &= b + n \cdot c + (d_{y_n} - d_{y_{(n-1)}}) \end{aligned} \right\} \quad (7.17)$$



Как видно, первые разности содержат, кроме постоянного  $b$ , еще и переменные слагаемые  $3c, 5c, 7c, \dots$ . Чтобы добиться устранения влияния общей тенденции, на основе первых разностей  $\Delta'$ , рассчитаем вторые разности  $\Delta''$ :

$$\left. \begin{aligned} \Delta_1'' &= \Delta_2' - \Delta_1' = 2 \cdot c + (d_{y_3} - 2 \cdot d_{y_2} + d_{y_1}); \\ \Delta_2'' &= \Delta_3' - \Delta_2' = 2 \cdot c + (d_{y_4} - 2 \cdot d_{y_3} + d_{y_2}); \\ &\dots\dots\dots \\ \Delta_n'' &= \Delta_n' - \Delta_{n-1}' = 2 \cdot c + (d_{y_n} - 2 \cdot d_{y_{(n-1)}} + d_{y_{(n-2)}}) \end{aligned} \right\} \quad (7.18)$$

Отсюда, колебания  $\Delta'$  определяются только величинами  $d_y$ , так как  $2c$  – величина постоянная во всех вторых разностях. Таким образом, для дальнейших исследований следует оперировать не с исходными рядами, а с их разностями различного порядка: при линейном тренде – с первыми разностями, при параболическом – с вторыми разностями, при аппроксимации зависимости параболой  $n$ -го порядка или полиномами – с разностями  $n$ -го порядка. Простейший прием экстраполяции динамики рядов или их тренда следующий:

1) проводится обобщающая линия, отражающая тенденцию ряда на основе визуального впечатления о расположении фактических точек;

2) определяется постоянство и рассчитывается средний абсолютный прирост за последние годы –  $\Delta = (y_n - y_1)/(n - 1)$ . Далее, он последовательно прибавляется к оценке тренда на последний срок столько раз, на сколько периодов экстраполируется ряд. Можно также уменьшить последний член тренда на коэффициент роста в степени, пропорциональной периоду экстраполяции;

3) устанавливаются соотношения изменений показателей динамики сравниваемых рядов на основе корреляции между ними.

На рис. 7.2 показано выделение тренда изменения скорости ветра по метеостанции Василевичи за май.



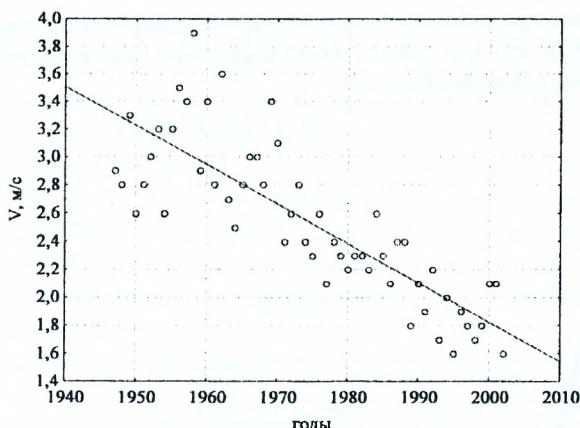


Рис. 7.2. Изменения скорости ветра по метеостанции Василевичи за май

*в) Автокорреляционный анализ*

Автокорреляция означает корреляцию параметра с самим собой. Коэффициенты автокорреляции являются коэффициентами линейной корреляции между значениями временного ряда в данный момент времени и его же значениями в последующий момент времени. Для практических целей коэффициенты автокорреляции определяются по формуле:

$$r_x(\tau) = \frac{\sum_{i=0}^{n-\tau-1} (x_i - \bar{x}) \cdot (x_{i+\tau} - \bar{x})}{(n - \tau - 1) \cdot \sigma_x^2}, \quad (7.19)$$

где  $\tau$  – запаздывание;  $\bar{x}$  – среднее значение ряда;  $\sigma_x^2$  – его дисперсия.

Если запаздывание мало, то в природных процессах коэффициенты автокорреляции обычно положительны, поскольку для природных процессов характерна устойчивость. По мере увеличения запаздывания коэффициент автокорреляции уменьшается и может стать отрицательным. Зависимость между коэффициентом автокорреляции и периодом запаздывания обычно представляется автокоррелограммой (рис. 7.3).

В зависимости от динамики развития исследуемого процесса корреляционные функции имеют различную степень внутрирядной связности и, как следствие, – различный вид и форму аппроксимации. Например, корреляционная функция чисто случайного ряда имеет максималь-

ное значение при сдвиге  $\tau=0$ , при остальных сдвигах  $r(\tau)$  – она равняется нулю. Достаточно простой вид имеет корреляционная функция процессов по типу простой цепи Маркова:

$$r(\tau) = [r(1)]^\tau \approx r^{-\alpha(\tau)}, \quad (7.20)$$

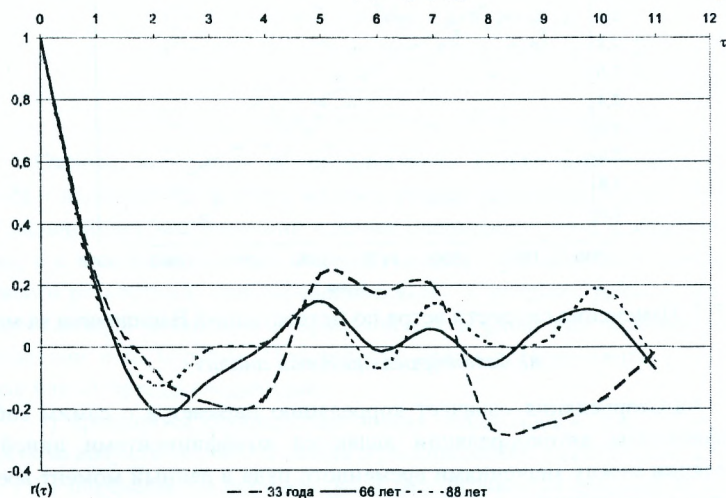


Рис. 7.3. Автокорреляционные функции годового стока р. Припять для выборок 33 года, 66 и 88 лет при смещении до 11 лет

Корреляционная функция  $r(\tau)$  случайного ряда (периодом  $m$ -лет), сглаженного скользящим образом, определяется равенством:

$$r(\tau) = (m - \tau) / m, \quad (7.21)$$

а ряда из суммы синусоид  $x(t) = \sum \alpha_k \cdot \sin(\omega_k \cdot t)$  – соотношением

$$r(\tau) = \sum b_k \cdot \cos(\omega_k \cdot \tau). \quad (7.22)$$

Представление о пространственной структуре коэффициентов автокорреляции можно получить из рис. 7.4. При построении карты коэффициентов автокорреляции использованы данные по 161 речному створу Беларуси. Изолинии коэффициентов автокорреляции годового стока рек образуют локальные группировки и приурочены к возвышенностям.

Однако корреляционные функции, несмотря на их универсальность, часто не поддаются интерпретации. Кроме того, непосредственно по ним трудно установить сопряженность колебаний с тем или иным пе-

риодом. В связи с этим при анализе структуры случайных процессов используются спектральные характеристики, позволяющие оценить распределение амплитуд (энергии) колебаний в частотной области. То обстоятельство, что коэффициенты автокорреляции отличны от нуля, означает, что многие из статистических оценок неприменимы, поскольку при их выводе предполагается, что данные временного ряда независимы. Получив автокорреляционную функцию, необходимо решить вопрос о погрешностях ее определения. Важно отметить, что значения ее ординат содержат не только детерминированную (причинно обусловленную), но и случайную составляющую.

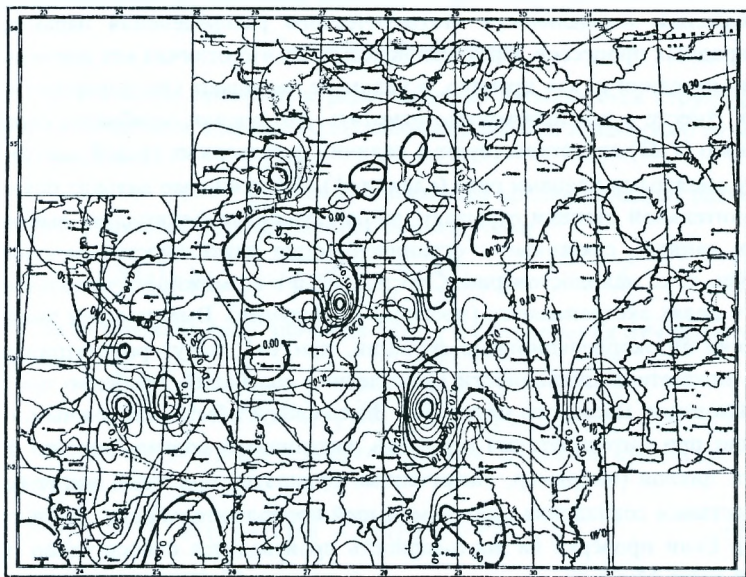


Рис. 7.4. Пространственная структура коэффициентов автокорреляции годового стока рек Беларуси

Среднеквадратическое отклонение последней находится по соотношению:

$$\sigma_{r(x)} = \frac{1 - (r(\tau))^2}{\sqrt{N - \tau - 1}}, \text{ при } \tau = 1, 2, \dots, m. \quad (7.23)$$

В случае, когда величины  $r(\tau)$  малы, отклонения коэффициентов автокорреляции от их истинных значений распределены по нормальному закону. Тогда доверительные границы (Н-нижняя, В-верхняя) для оценки автокорреляционной функции, с доверительной вероятностью 95%, могут быть найдены как:

$$r_H(\tau) = r(\tau) - \frac{2}{\sqrt{N-\tau}} \cdot (1-r^2(\tau)) ; \quad (7.24)$$

$$r_B(\tau) = r(\tau) + \frac{2}{\sqrt{N-\tau}} \cdot (1-r^2(\tau)) . \quad (7.25)$$

Необходимо отметить, что данные зависимости получены в предположении нормальности статистического распределения параметров природных процессов. Практика показывает, что отличия «от нормальности» невелики и, как правило, в целом не нарушают стационарности рядов. Вместе с тем необходимо отметить, что оценки ошибок по соотношению (7.23) свидетельствуют о наличии достоверных связей только между смежными членами ряда (сдвиг  $\tau=1$ ). Оценки более дальних связей в значительной степени отражают лишь случайные флуктуации выборочных данных, связанные с ограниченностью длительности имеющихся рядов, т. е. малодостоверны. Это исключает возможность использовать их в целях экстраполяции (прогноза на будущее). Вместе с тем эмпирические автокорреляционные функции, при  $\tau > 1$ , могут рассматриваться как показатели цикличности многолетних колебаний, особенно при использовании массового материала. Здесь следует быть осторожным, особенно при интерпретации периодов, являющихся совпадением по фазе двух циклов (например, увеличенные ординаты, при  $\tau=25$ , могут быть следствием совпадения циклов меньшей продолжительности:  $\tau=8$  и  $\tau=17$  лет). Если проверка на достоверность производится относительно простой цепи Маркова, т.е. проверяется гипотеза об отсутствии внутрирядных связей, например, при  $\tau > 3$  лет, то построение доверительных интервалов производится при объеме информации:

$$n' = \frac{n-1}{\sqrt{1-2,2 \cdot r_1 \cdot (1+r_1)^2}} , \quad (7.26)$$

а оценка среднего квадратического отклонения как:

$$\sigma_{r(\tau)} = \frac{[1 - r^2(\tau)]}{\sqrt{n'}}. \quad (7.27)$$

В общем случае, независимо от вида последовательности  $x_1, x_2, \dots, x_n$ , построение доверительных интервалов  $r(\tau)$  осуществляется по формуле:

$$r_{\alpha\tau} = t_{\alpha} \cdot \sigma_{r(\tau)} = 0, \quad (7.28)$$

где  $r_{\alpha\tau}$  – доверительные границы  $r(\tau)$  для данного  $\tau$ ;  $t_{\alpha}$  – критерий Стьюдента с  $(n-\tau-1)$  – степенями свободы при данном уровне значимости  $\alpha$ ;  $\sigma_{r(\tau)} = 0$  – среднее квадратическое отклонение выборочных значений  $r(\tau)$ , определяемое по системе зависимостей (7.26) – (7.28).

### 2) Структурная функция

Для анализа временных рядов широко используется структурная функция, введенная впервые А. Н. Колмогоровым, и независимо от него, О. А. Дроздовым.

Под структурной функцией понимается средний квадрат разности значений функции  $X(t)$ , разделенных промежутком  $\tau$ :

$$B_1(\tau) = \overline{[x(t) - x(t + \tau)]^2} = \frac{1}{N \cdot \tau} \sum_{i=1}^{N-\tau} [x_i(t) - x_i(t + \tau)]^2 \quad (7.29)$$

Если брать не сами значения  $X(t)$ , а их отклонения от среднего то получим:

$$B_2(\tau) = \overline{[x'(t) - x'(t + \tau)]^2}. \quad (7.30)$$

Между структурной и корреляционной функцией существует определенная связь:

$$B_2(\tau) = \overline{[x'(t)]^2} - 2\overline{[x'(t)x'(t + \tau)]} + \overline{[x'(t + \tau)]^2}. \quad (7.31)$$

Поскольку  $\overline{[x'(t)]^2} = \overline{[x'(t + \tau)]^2} = \sigma^2$ , то  $B_2(\tau) = 2[\sigma^2 - m_2(\tau)]$ .

При  $\tau \rightarrow \infty$ ,  $m_2(\tau) \rightarrow 0$ , отсюда  $B_2(\infty) = 2\sigma^2$ ,  $m_2(\tau) = \frac{1}{2}[B_1(\infty) - B_2(\tau)]$ .

Пронумеровав, получим:

$$\frac{m_2(\tau)}{\sigma^2} = \mu_1(\tau) = 1 - \frac{B_2(\tau)}{2\sigma^2} \quad (7.32)$$



Последняя формула устанавливает простую связь между структурной и корреляционной функцией.

Использование структурной функции вместо корреляционной во многих случаях удобнее, например, в тех случаях, когда нас интересуют не абсолютные значения величин, а только разности их в различные моменты времени. С другой стороны, определение структурной функции по эмпирическим данным связано с меньшими ошибками, поскольку случайные большие отклонения оказывают большее влияние при вычислении корреляционной функции.

При использовании структурной функции поля исключается из рассмотрения влияние изменений «среднего поля», т. е. влияние медленной эволюции возмущений.

Для подсчета необходимо вычисление произведений, что сложнее, чем вычисление разностей при вычислении  $B(\tau)$ .

Существенно также, что наличие неоднородностей внутри используемой выборки сильно сказывается на результатах вычисления по формуле для  $m(\tau)$  и меньше на  $\sigma(\tau)$ .

Для годовой гармоники  $m(\tau)$  завышается на  $1/2$  квадрата амплитуды, а структурная функция не завышается. В случае нескольких гармоник  $B(\tau)$  завышается меньше, чем  $m(\tau)$ .

Для характеристики пространственной и временной изменчивости метеорологических элементов обычно используется структурная функция  $b_f(\rho_1, \rho_2)$ , описывающая средний квадрат разности величин в различных пунктах или в различные моменты времени:

$$b_f(\rho_1, \rho_2) = \overline{[f(\rho_1) - f(\rho_2)]^2}. \quad (7.33)$$

На практике структурная функция природного поля  $b_f(\rho)$  определяется ошибкой, равной удвоенной дисперсии случайной ошибки метода измерений  $\sigma_0^2$  (рис. 7.5). Эта средняя квадратическая ошибка метода измерений определяется с помощью экстраполяции на нуль измеренной структурной функции. «Истинная» структурная функция  $\beta_{ист}(\rho)$  определяется смещением измеренной на  $2\sigma_0^2$  вниз, к началу оси координат:

$$\beta_{изм}(\rho) = \beta_{ист}(\rho) + 2\sigma_0^2. \quad (7.34)$$



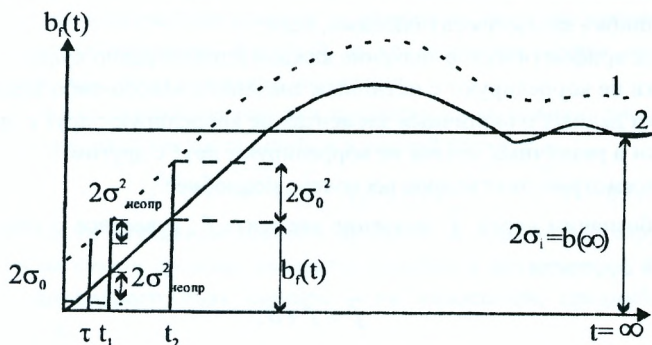


Рис. 7.5. Схематический вид измеренной (1) и истинной (2) структурно-временной функции  $t_1 - t_2/3$

Временное (пространственное) разрешение (дискретность) — временной интервал (расстояние) между наблюдениями ( $\Delta t$ ,  $\Delta r$ ) выбирается исходя из масштабов изучаемых процессов.

*д) Особенности вычисления структурных и корреляционных функций макромасштабных метеорологических полей*

К статистически однородным и изотропным метеорологическим полям применима эргодическая теорема, согласно которой корреляционная и структурная функция могут быть определены по одной реализации случайной функции. Учитывая, что эргодичности для метеорологических полей нет, т. к. число дискретных величин не так и велико и отклонения от норм получаются по тому же материалу, что и отклонения, то для получения точных значений структурной или корреляционной функции необходимо, чтобы:

1) отдельные ситуации отличались теми «случайными» свойствами, статистическая структура которых подлежит исследованию. Во всех иных отношениях ситуации должны быть тождественны друг другу;

2) количество данных должно быть настолько велико, чтобы введение дополнительных данных, в пределах заданной точности, не изменило вычисленного значения  $\mu$ ;

3) каждое из данных должно быть точным. На точность вычисления структурной и корреляционной функции оказывают влияния и ошибки в исходных данных.

Ошибки являются случайными, если:

- а) среднее арифметическое значение каждой ошибки равно нулю;
- б) ошибки не коррелируют с истинным значением какого-либо элемента;
- в) ошибки данных о различных элементах не коррелируют друг с другом;
- г) ошибки в различных точках не коррелируют друг с другом.

Рассмотрим этот вопрос несколько подробнее.

Обозначим через  $\tilde{f}$  значение элемента  $f$ , известное с некоторой ошибкой  $\delta_f$ , тогда

$$\tilde{f} = f + \delta_f, \quad (7.35)$$

где  $f$  - истинное значение элемента.

Если ошибка  $\delta_f$  чисто случайна, то имеет место равенство при всех  $r$  и  $f$   $\overline{\delta_f(r)} = 0$  при всех  $r_1, r_2, f, \varphi$ , т. е.  $\overline{f'(r_2)\delta_\varphi(r_2)} = 0$   $\delta_f(r_1)\delta_\varphi(r_2) = 0$ , за исключением  $f = \varphi$  и  $r_1 = r_2$ .

Средний квадрат ошибки данных об элементе  $f$  в точке  $r$  обозначим через

$$\frac{\sigma_f^2(r)}{\delta_f^2(r)} = \sigma_f^2(r). \quad (7.36)$$

Отношение его к дисперсии элемента назовем мерой случайной ошибки

$$\eta_f(r) = \frac{\sigma_f^2(\bar{r})}{f^2(r)}. \quad (7.37)$$

Для однородного поля

$$\eta_f = \frac{\sigma_f^2}{mf(0)}, \quad (7.38)$$

можно отметить следующее:

- а) чисто случайные ошибки не влияют на значения ненормированных взаимных корреляционных функций

$$\begin{aligned} m_{\tilde{f}\tilde{\varphi}}(r_1, r_2) &= \tilde{f}'(r_1)\tilde{\varphi}'(r_2) = [f'(r_1) + \delta_f(r_1)] \times [\varphi'(r_2)] = \\ &= \overline{f'(r_1)\varphi'(r_2)} + \delta_f(r_1)\delta_\varphi(r_2) = m_{f\varphi}(r_1, r_2) \end{aligned}, \quad (7.39)$$

б) чисто случайные ошибки не влияют на значение ненормированных автокорреляционных функций, за исключением случая, когда автокорреляционная функция равна дисперсии.

в) чисто случайные ошибки приводят к завышению дисперсии на величину среднего квадрата ошибки в данной точке, т. е.

$$\tilde{f}^{r^2}(r) = \overline{f^{r^2}}(r)[1 + \eta_f(r_2)] \text{ или } \overline{\tilde{f}^{r^2}} = \overline{f^{r^2}}(1 + \eta), \quad (7.40)$$

г) чисто случайные ошибки приводят к завышению структурной функции на сумму средних квадратов ошибок в двух точках, в частном случае однородного поля ошибок – на удвоенный, средний квадрат ошибки, т. е.

$$B_{\tilde{f}}(r_1 r_2) = B_f(r_1 r_2) + \sigma_f^2(r_1) + \sigma_f^2(r_2). \quad (7.41)$$

Если поле однородно и изотропно по отношению к структурной функции, то  $B_{\tilde{f}}(\rho) = b_f(\rho) + 2\sigma_f^2$ . Так как  $B_f(0) = 0$ , то

$$B_{\tilde{f}}(0) = 2\sigma_f^2. \quad (7.42)$$

На (7.42) основан косвенный метод оценки ошибок исходных данных.

Допустим, нужно экстраполировать кривую  $B_{\tilde{f}}(\rho)$ , построенную по исходным данным, на нуль. Тогда ордината точки, в которой  $\rho=0$ , дает удвоенный средний квадрат ошибки. После этого для получения истинной структурной функции надо вычесть из всех значений  $B_{\tilde{f}}(\rho)$  величину  $2\sigma_f^2$ , т. е. провести кривую, параллельную исходной и проходящую через начало координат. Практически определить  $\sigma_f^2$  трудно, т. к. при малых  $\rho$  может иметь место существенная кривизна  $B_{\tilde{f}}(\rho)$ ;

д) случайные ошибки приводят к занижению значений нормированной автокорреляционной функции в отношении  $\frac{1}{(1 + \eta_s)}$  и

$$\mu_{\tilde{f}}(\rho) = \frac{\mu_f(\rho)}{1 + \eta_f}; \quad (7.43)$$

е) чисто случайные ошибки изменяют значение нормированной структурной функции

$$B_j(\rho) = B_j(\rho) \frac{1 + \frac{\eta_j}{\beta_j(\rho)}}{1 + \eta_j}, \quad (7.44)$$

приближая его к единице.

Ошибки исходных данных включают в себя и микроклиматические ошибки, обусловленные влиянием мелкомасштабных особенностей подстилающей поверхности. Исключить эти ошибки очень трудно, т. к. не ясно, какие неоднородности надо исключить, а какие оставить, какое применение мыслится для статистических характеристик.

Одним из фактов неоднородности выборок является годовой ход. Часто, чтобы объем выборки был велик, в выборку включается ситуация, относящиеся к разным сезонам года или месяцам. При этом, наряду с теми отклонениями от средних, структуру которых мы хотим изучить, будет иметь место отклонения от средних, обусловленных влиянием годового хода элементов внутри выборки. Поскольку результаты должны применяться к каждому физическому моменту надо исключить влияние этой неоднородности.

Рассмотрим особенности учета случайных ошибок на точность вычисления статистических характеристик.

Средняя квадратическая  $\sigma_{\bar{x}}$  среднего значения  $\bar{X}$  в случае нормального закона определяется соотношением  $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{N}}$ .

Если между данными имеется корреляционная зависимость, то используется более сложная формула:  $\sigma_{\bar{x}} = \frac{\sigma_t}{\sqrt{N}} \sqrt{\frac{1 + \mu_1(\tau)}{1 - \mu_1(\tau)}}$ . Здесь  $\tau = 1$

По этим формулам можно оценить число случаев  $N$ , необходимых для вычисления среднего значения с заданной ошибкой  $\sigma_{\bar{x}}$ :

$$N \geq \frac{\sigma^2}{\sigma_{\bar{x}}^2}; \quad N \geq \frac{\sigma^2}{\sigma_{\bar{x}}^2} \frac{1 + \mu_1(\tau)}{1 - \mu_1(\tau)}. \quad (7.45)$$

Ошибка определения среднего квадратического отклонения  $\sigma_{\sigma}$  определяется формулой

$$\sigma_{\sigma} = \frac{\sigma}{\sqrt{2N}}. \quad (7.46)$$

Если наблюдается отклонение от нормального распределения, то надо учитывать эксцесс распределения

$$\sigma_{\sigma} = \frac{\sigma}{\sqrt{2N}} \sqrt{\frac{2+\Theta}{2}}. \quad (7.47)$$

Средняя квадратическая ошибка коэффициента корреляции оценивается по формуле

$$\sigma_r = \frac{1-r^2}{\sqrt{N-1}}. \quad (7.48)$$

Эта формула справедлива при  $r \ll 1$ .

В других случаях используют такой параметр  $z = \frac{1}{2} \ln \frac{1+r}{1-r}$ , согласно Фишеру последний имеет нормальное распределение

$$\sigma_z = \sqrt{\frac{1}{N-3}}.$$

Количество данных для подсчета коэффициента корреляции с точностью 5% вычисляется по формуле:

$$N \geq \frac{4}{(z')^2} + 3, \text{ а } \frac{z'}{\sigma_z} \geq 2. \quad (7.49)$$

Существует еще один источник ошибок при вычислении автокорреляционных функций – нарушение однородности и изотропности. В связи с этим М. И. Юдин [Юдин, 1979] предлагает изучать структурные функции не самого метеорологического элемента, а его разделенных разностей различного порядка. Таким образом, хотя и невозможно восстановить структурную функцию самого элемента, но зато можно исследовать тонкие свойства последней.

#### *е) Взаимная корреляционная функция*

Взаимная корреляционная функция характеризует зависимость значений одного временного ряда в данный момент от значений другого временного ряда в другой момент времени. Если имеются аномалии двух рядов –

$$\left. \begin{array}{l} \Delta x_1, \Delta x_2, \dots, \Delta x_\tau, \dots, \Delta x_N \\ \Delta y_1, \Delta y_2, \dots, \Delta y_\tau, \dots, \Delta y_N \end{array} \right\} \quad (7.50)$$

то оценку функции взаимной корреляции можно представить в виде:

$$r_{xy} = \begin{cases} \frac{1}{(N - \tau) \cdot \sigma_x \cdot \sigma_y} \cdot \sum_{i=1}^{N-\tau} \Delta x_i \cdot \Delta y_{i+\tau}, & \text{при } \tau \geq 0, \\ \frac{1}{(N - |\tau|) \cdot \sigma_x \cdot \sigma_y} \cdot \sum_{i=1}^{N+|\tau|} \Delta x_i \cdot \Delta y_{i+|\tau|}, & \text{при } \tau \leq 0. \end{cases} \quad (7.51)$$

Первая из формул описывает зависимость  $y_{i+\tau}$  от  $x_\tau$ , вторая, наоборот,  $x_{i+|\tau|}$  от  $y_i$ . В противоположность автокорреляционной функции взаимная корреляционная функция необязательно имеет максимум при  $\tau=0$ . Максимальное значение  $r_{xy}(\tau)$  указывает на сдвиг между анализируемыми рядами, при котором эти ряды наиболее тесно связаны. И автокорреляционная функция, и функция взаимной корреляции могут быть использованы в прогнозировании значения данной величины в зависимости от значения этой или другой величины в предшествующий момент времени -

$$x(t + \tau) = r(\tau) \cdot x(\tau) + \omega(\tau) \cdot \sqrt{1 - r^2(\tau)}, \quad (7.52)$$

где  $r(\tau)$  – коэффициент корреляции между  $x(t+\tau)$  и  $x(t)$ ;  $\omega(\tau)$  – некоторая некоррелированная с  $x(\tau)$  функция, которая обычно определяет ошибку прогноза. В результате, по (7.52) получаем так называемый статистический метод прогноза по одной точке предыстории. Прогнозируемая величина  $x(t+\tau)$  при этом будет всегда близка к последнему значению  $x(t)$ . Отметим, что включение в прогностическое соотношение переменных за более ранние сроки не всегда приводит к положительным результатам, так как чаще всего оказывается, что прогностические уравнения типа (7.52) нарушаются при проверке на других периодах. Многие свойства взаимно корреляционных функций идентичны свойствам автокорреляционных функций, так как свойство четности заменяется симметричностью нулевого сдвига, т. е.  $r_{xy}(\tau) \neq r_{xy}(-\tau)$ , а фазы колебаний взаимодействующих процессов не всегда совпадают, то следует, что гармоники с одинаковыми фазами и периодами колебаний на графике взаимной корреляции будут образовывать всплески, соответствующие максимальным амплитудам циклов при нулевой разности фаз. Различие же частот, наоборот, приводит к взаимной корреляции, близкой к нулевым значениям.



*ж) Автоспектральный анализ*

В тех случаях, когда процесс складывается из составляющих разных временных масштабов (явление, широко распространенное в гидрометеорологии), знание структуры процесса во временной области часто оказывается недостаточным. При решении многих задач необходимо знать распределение интенсивности процесса между составляющими различных временных масштабов, т. е. необходимо выполнить описание случайного процесса в частной области. Для этой цели используется спектральное разложение процесса. Спектр временного ряда аналогичен оптическому спектру, который показывает вклад различных длин волн или частот в энергию заданного источника света. Физический смысл спектра временного ряда состоит в том, что он показывает вклад колебаний с различными частотами в полную дисперсию временного ряда и, следовательно, может быть получен с помощью метода гармонического анализа. Если вычислить все  $n/2$  – гармоник и построить график полуквadrата их амплитуды, как функцию частоты, то разброс точек окажется очень большим. При этом, если построить спектры для двух отдельных частей одного стационарного временного ряда, то отдельные точки спектра будут занимать совершенно различное положение, поскольку отдельные резкие пики в нерегулярных колебаниях временного ряда носят чисто случайный характер. Поэтому с помощью спектрального анализа обычно не пытаются определить амплитуды отдельных гармоник. Его целью является нахождение сглаженного спектра, который остается одинаковым для различных частей одного и того же временного ряда. При этом рассматривается не спектр данного короткого временного ряда, а спектр бесконечного длинного временного ряда, из которого данный временной ряд представляет собой короткую случайную выборку. Для длинного ряда можно определить сглаженный спектр, используя соответствующую методику предельного перехода. Задачей практического спектрального анализа является оценка этого сглаженного спектра на основании данного короткого ряда. Спектральное разложение отображает стационарную случайную функцию в виде разложения на периодические колебания различных частот  $\omega_j$ . Если имеется какой-либо квазигармонический процесс, представляемый набором гармоник:

$$x(t) = \bar{x} + \sum_{i=1}^m [A_i \cdot \cos \omega t_i + B_i \cdot \sin \omega t_i], \quad (7.53)$$

то его спектром является функция

$$C_i = \sqrt{A_i^2 + B_i^2}, \quad (7.54)$$

описывающая распределение амплитуд гармоник по различным частотам. В отличие от простых колебательных движений стационарный случайный процесс описывается каноническими разложениями случайных процессов типа

$$x(t) = \bar{x}(t) + \sum \vartheta_i \cdot \varphi_i(t), \quad (7.55)$$

где  $\bar{x}(t)$  – математическое ожидание случайной функции;  $\vartheta_1, \vartheta_2, \dots, \vartheta_m$  – некоррелированные случайные величины с математическим ожиданием, равным нулю.

Плотность распределения дисперсий по частотам непрерывного спектра характеризует спектральную плотность дисперсии или спектральную плотность стационарной случайной функции  $S(x)$ . Отсюда дисперсия будет представлена в виде:

$$D_x = \int_0^{\infty} S_x(\omega) \cdot d\omega. \quad (7.56)$$

Иногда при решении практических задач вместо спектральной плотности  $S(x)$  пользуются нормированной спектральной плотностью  $\Psi_x(\omega)$ , связанной преобразованиями Фурье:

$$r_x(\tau) = \int \Psi_x(\omega) \cdot \cos(\omega \cdot \tau) \cdot d\omega; \quad (7.57)$$

$$\Psi_x(\omega) = \frac{1}{\pi} \cdot \int r_x(\tau) \cdot \cos(\omega \cdot \tau) \cdot d\tau. \quad (7.58)$$

Если положить  $\tau=0$  и учесть, что  $r_x(\tau=0)=1$ , получим  $\int_0^{\infty} \Psi_x(\omega) \cdot d\omega = 1$ , т. е. полную площадь, ограниченную графиком нормированной плотности, равной единице. Имея аналитическое выражение корреляционной функции, можно легко определить вид спектральной функции. Так, спектральная плотность процесса (по типу простой цепи

Маркова), корреляционная функция которой описывается формулой  $r(\tau) = r^{-\alpha(\tau)}$ , с учетом выражения (7.58), будет определяться как:

$$\Psi(\omega) = \frac{a}{\pi \cdot \omega^2} + d^2. \quad (7.59)$$

Если аналитическое выражение корреляционной функции неизвестно или аналитическое описание носит сложный характер, то аппроксимация подинтегральной функции осуществляется кривой с соответствующим переходом от точного значения ее площади к приближенному значению, как суммы площадей конечного числа трапеций:

$$\Psi(\omega) \approx \frac{2}{\pi} \cdot \sum_{i=3}^{m-1} r_i \cdot t_i \cdot \left( \frac{\sin(\omega \cdot t_i)}{\omega \cdot t_i} \right) \cdot \left( \frac{\sin(\omega \cdot \Delta_i)}{\omega \cdot \Delta_i} \right), \quad (7.60)$$

где  $r_i = r_{\tau_i} - r_{\tau_{i-1}}$  – разности ординат соседних переломных точек ломаной кривой функции;  $\Delta_i = (\tau_i - \tau_{i-1})/2$  – полуразности абсцисс соседних точек;  $t_i = \tau_i + \Delta_i$  – абсциссы середин между этими точками;  $\omega = 2\pi/T$  – круговая частота;  $m$  – максимальный сдвиг  $t_{max}$ . Расчет спектральной функции ведется по выражению:

$$\Psi_i = \frac{r}{m} + \frac{2}{m} \cdot \sum_{\tau=1}^{m-1} \left[ r_{\tau} \cdot \cos\left(\frac{360}{2 \cdot m} \cdot i \cdot \tau\right) \right] + (-1)^m \cdot \frac{r_m}{m}, \quad (7.61)$$

где  $i$  – номер гармоники ( $i=1, 2, \dots, n$ );  $m$  – максимальный сдвиг;  $r_0$  и  $r_m$  – значения  $r(\tau)$ , при  $\tau=0$  и  $\tau=m$ . Значения  $\Psi_0$  и  $\Psi_m$  нужно уменьшить в 2 раза. Период колебания и номер гармоники связаны соотношением:

$$T = 2 \cdot m / i. \quad (7.62)$$

Отсюда период первой гармоники принимается равным  $2 \cdot m$ , второй –  $m$ , третьей  $2 \cdot m/3$ , и т. д. Приближенная оценка доверительной границы  $I_a$  когерентности, при уровне вероятности  $P$ , производится по формуле Гудмана:

$$I_{\beta} [H(\omega)] = \sqrt{1 - P^{1/\ell-1}}, \quad (7.63)$$

в которой  $\ell = (2 \cdot N - m/2)/m$  – число степеней свободы  $N$  данных реализаций;  $m$  – число запаздываний. Значения когерентности при 1% и 5%-ом уровнях значимости для различного числа степеней свободы приведены в табл. 7.4.

Таблица 7.4

Доверительные границы когерентности  $H(\omega)$ 

Уровень значимости, %	Степени свободы			
	4	10	20	40
1	0.89	0.63	0.46	0.33
5	0.80	0.53	0.38	0.27

Из табличных данных следует, что при числе степеней свободы  $\ell = 20$ , когерентность, равная 0,38 и больше, будет недостоверна в одном случае из 20. При использовании спектрального анализа для решения каждой конкретной задачи необходимо исследовать соотношения между временным шагом выборки  $\Delta t$ , длиной выборки  $N$ , максимальным сдвигом корреляционной функции  $m_{max}$  для расчета спектра, числом степеней свободы  $\ell = 2 \cdot N \neq m$  и нормированной стандартной ошибкой  $\sigma_{S(\omega)}$  оценок функции спектральной плотности. С одной стороны, число сдвигов  $m$  должно быть мало по сравнению с длиной выборки, число степеней свободы  $\ell$  – возможно большим. Это сохранит определенную степень статистической надежности. С другой стороны, число сдвигов должно быть достаточно большим, чтобы получить большее разрешение по полосе частот; при этом надежность статистических оценок в пределах частотной полосы уменьшается. Обычно выбор максимального временного сдвига  $\tau_{max} = m$  производится, исходя из возможной точности расчетов спектральной плотности, в частности, по формулам:

$$\tau_{2\%} = 2 \cdot \pi \cdot n / 50; \quad \tau_{5\%} = 2 \cdot \pi \cdot n / 120; \quad \tau_{10\%} = 2 \cdot \pi \cdot n / 10, \quad (7.64)$$

где  $\tau_{2\%}$ ,  $\tau_{5\%}$ ,  $\tau_{10\%}$  – значения максимального сдвига при допустимой погрешности расчета –  $S(\omega)$  соответственно 2; 5; 10%.

Большинство авторов при 50 – 100-летних рядах наблюдений, принимают –  $\tau_{max} = 10$  – 30 лет. Спектральные функции, рассчитанные по выборочным данным, будут отличаться от спектра генеральной совокупности. Как и в случае оценок одномерных распределений, при оценке значимости спектра целесообразно использовать нуль-гипотезу. Нуль-гипотеза заключается в предположении отсутствия гармонических колебаний в спектре исследуемого ряда на фоне спектра реализации «белого» (горизонтальная линия – нулевой континуум) или «красного» шума (убывающая экспоненциальная линия). Предполагается, что исходная выбор-

ка не случайна, а ее значения распределены по нормальному закону. Гипотеза проверяется в результате сравнения  $S_x(\omega)$  со значениями  $S(\omega)$  заданной обеспеченности, принимаемыми за границы доверительного интервала  $I_\beta(S_\omega)$ . В этом случае спектральные оценки полагаются приблизительно распределенными по значению  $\chi^2$  и нормированными на число степеней свободы  $\ell$ :

$$S(\omega) = \chi^2 / \ell, \quad (7.65)$$

где  $\ell$  – число степеней свободы, определяемое как:

$$\ell = (2 \cdot N - 0,5 \cdot \tau_m) / \tau_m; \quad (7.66)$$

$N$  – объем выборки, используемый для оценки спектра;  $\tau_m$  – максимальный сдвиг на коррелограмме. Тогда для отыскания доверительного интервала  $I_\beta(S_\omega)$  справедливо равенство:

$$I_\beta(S_\omega) = \bar{S}_x(\omega) \cdot \chi^2 / \ell, \quad (7.67)$$

в котором  $S_x(\omega)$  – средний уровень спектральной плотности (приравняваемый к «белому шуму»), вычисляемый в интервале значений автокорреляционных функций в пределах сдвигов (запаздываний от 1 до  $m$ ). Выход пиков на спектрограмме за границы заданных  $I_\beta(S_\omega)$  будет свидетельствовать о достоверности выявляемых частот колебаний. Вероятностные точки распределения  $\chi^2 / \ell$  даны в табл. 7.5.

Если имеются только таблицы  $\chi^2$ , а не  $\chi^2 / \ell$ , то при исследуемом уровне значимости  $\alpha$  получаются отысканием значения  $\chi_\alpha^2$ , которые соответствуют значениям  $\ell$ , с последующим делением на  $\ell$ . При значительном вкладе в исследуемый процесс «красного шума» марковского процесса первого порядка, средний уровень спектральной плотности (линейный в силу стационарности случайного процесса) превращается в нелинейный, убывающий к низким частотам. Оценка континуума «красного шума» осуществляется с использованием выражения:

$$S_k(\omega) = \bar{S}_x(\omega) \cdot \frac{1 - r_x^2(\tau_1)}{1 + r_x^2(\tau_1) - 2 \cdot r_x(\tau_1) \cdot \cos \frac{\pi \cdot k}{\tau_m}}, \quad (7.68)$$

которое содержит значения автокорреляционной функции единичного  $\tau_1$  и максимального  $\tau_m$  сдвигов, превышающих, при  $\tau_1$ , значение  $r(\tau)=0,4$ .

Таблица 7.5

Вероятностные (по Халду) точки распределения ( $\chi^2 / \ell$ )

Степени свободы, $\ell$	Вероятность в процентах				
	1	5*	95*	99	99,9
2	0,010	0,052	3,000	4,605	6,908
5	0,111	0,229	2,214	3,017	4,103
10	0,256	0,394	1,831	2,321	2,959
20	0,413	0,543	1,570	1,877	2,266
30	0,498	0,616	1,459	1,696	1,990
50	0,594	0,695	1,350	1,523	1,733
60	0,625	0,720	1,318	1,473	1,660
80	0,669	0,755	1,274	1,404	1,560
100	0,701	0,699	1,243	1,358	1,494
200	0,782	0,841	1,170	1,247	1,338
400	0,843	0,887	1,119	1,172	1,238
1000	0,899	0,928	1,075	1,107	1,144

Примечание: \*) – вероятность, эквивалентная 95%-ой точке значимости для одно-стороннего критерия спектрального минимума

Сглаженный спектр, как указывалось, может быть получен методом гармонического анализа с последующим сглаживанием амплитуд всех отдельных гармоник, посредством какой-либо алгебраической процедуры осреднения. Этот метод громоздок и в настоящее время реализуется на ЭВМ с использованием так называемого алгоритма быстрого преобразования Фурье. Винером и Хинчином отмечалась возможность вычисления спектра через корреляционные функции; Блэкман и Тьюки реализовали эту идею практически. Сегодня этот метод практически вытеснен другими методами, но процедура Блэкмана – Тьюки для небольших запаздываний сохраняет свою актуальность [Дроздов, 1984]. Алгоритм вычисления спектра с использованием процедуры Блэкмана–Тьюки имеет вид:

1) если среднее временного ряда  $x(i)$ , состоящего из  $n$  – членов, не равно нулю, то его вычисляют и вычитают из всех значений ряда;

2) для  $m+1$  – значений индекса  $t$  вычисляются значения автокорреляционной функции --



$$r_x(\tau) = \frac{\sum_{i=0}^{n-\tau-1} x_i \cdot x_{i+\tau}}{(n-\tau-1) \cdot \sigma_x^2}, \text{ пр } \tau = 0, 1, \dots, m; \quad (7.69)$$

3) выбирается корреляционное окно –

а) Хеннинга

$$V_m^{(1)}(\tau) = \frac{1}{2} \cdot (1 + \cos \frac{\pi \cdot \tau}{m}); \quad (7.70)$$

б) Хемминга

$$V_m^{(2)}(\tau) = 0,54 + 0,46 \cdot (1 + \cos \frac{\pi \cdot \tau}{m}); \quad (7.71)$$

в) Парзена

$$V_m^{(3)}(\tau) = \begin{cases} (1 - 6 \cdot \frac{\tau}{m})^2 \cdot (1 - \frac{\tau}{m}), & \text{при } \tau < \frac{m}{2}; \\ 2 \cdot (1 - \frac{\tau}{m})^3, & \text{при } \tau > \frac{m}{2}, \end{cases} \quad (7.72)$$

и с помощью одного из них (окон) сглаживается автокорреляционная функция

$$\bar{r}_x(\tau) = r_x(\tau) \cdot V_m^{(l)}(\tau); \quad (7.73)$$

4) посредством интегрирования (методом трапеции) для различных частот вычисляются спектральные оценки:

$$S_x(k) = 2 \cdot \left( 1 + 2 \cdot \sum_{\tau=1}^{m-1} \bar{r}_x(\tau) \cdot \cos \frac{\pi \cdot \tau \cdot k}{m} \right), \text{ при } 0 \leq k \leq m; \quad (7.74)$$

график выборочных спектральных оценок целесообразно строить в логарифмическом масштабе, поскольку построение доверительного интервала для спектра сводится к откладыванию околорыборочной спектральной оценки одного и того же интервала для всех частот; доверительный интервал для логарифма спектра рассчитывается по зависимостям:

$$\lg S_x(k) + \lg \frac{\vartheta}{x_{\vartheta}(1-\alpha/2)}; \quad \lg S_x(k) + \lg \frac{\vartheta}{x_{\vartheta}(\alpha/2)}, \quad (7.75)$$

в которых  $\vartheta$  – число степеней свободы (для окон: Хеннинга –  $\vartheta = 2,667 \cdot \frac{n}{m}$ ; Парзена –  $\vartheta = 3,71 \cdot \frac{n}{m}$ ).

Хотя энергетический спектр различен при различных видах случайностей, но он содержит и определенные недостатки.

Первое – если имеются точные периодичности во временных рядах, то спектр необязательно представит их так очевидно как это может сделать классический гармонический анализ.

Второе – спектр теряет всю информацию о фазе колебаний (периодичных или нет), содержащуюся в исходном ряду, поэтому должны быть использованы другие методы анализа, если нужно излучить информацию о фазе.

Третье – если ряд характеризуется очень сильной периодичностью (например, годовой ход температуры как отражение последовательных среднемесячных значений температуры), то на спектр могут оказать существенное влияние утечка информации на других длинах волн через щели спектрального окна. Эту проблему можно обойти путем преобразования исходных данных ряда, перед тем как делается спектральный анализ. Как пример преобразования рассмотрим ряд средних месячных значений температуры нескольких лет, которые содержат строго годовой ход. Этот годовой ход может быть исключен полностью путем трансформации исходной температуры  $T_i$  в новый ряд  $T'_i$  при помощи уравнения:  $T'_i = T_i - \bar{T}$ . Здесь  $\bar{T}$  – средняя наблюдаемая температура для всех значений ряда, принадлежащего к тому же календарному месяцу, что и  $T_i$ . В этом примере информация о годовом ходе будет опущена из спектра ряда  $T'_i$ , но она может быть обнаружена другими способами, например, классическим гармоническим анализом  $T$  значений. Подобным образом, в случае осадков, преобразование ряда месячных сумм  $R_i$  может быть осуществлено путем трансформации ряда  $R_i$  во второй ряд

$$R'_i = \frac{R_i}{R}, \quad (7.76)$$

где  $\bar{R}$  – наблюдаемое среднее  $R_i$  в том же календарном месяце, что и  $R_i$ .

Как и в случае классического гармонического анализа, надо быть осторожным относительно появления «фиктивной волны» при спектральном анализе. Если исходные ряды состоят из мгновенных наблюдений, сделанных через постоянные интервалы времени, то вариация короткой длины волны может быть неправильно воспринята как вариация длинной волны.

Вообще нежелательный эффект «фиктивной волны» может быть преодолен в спектральном анализе путем замены мгновенных наблюдений средними за некоторый отрезок времени.

### з) Фильтрация временных рядов

Спектральный состав временного ряда можно описать с помощью его статистических характеристик. Если исходный временной ряд содержит некоторые частоты или периоды, которые в данный момент не представляют интереса для исследователя, амплитуда этих волн может быть уменьшена с помощью статистической фильтрации. При этом изменяется спектр исходного временного ряда. Можно отфильтровать как низкие и высокие частоты, так и каждую частоту в отдельности. Операции «фильтрации» в гидрометеорологии и природопользовании выполняется обычно численными методами при следующем математическом описании

$$\tilde{x}_i = \sum_{m=-M}^M h_m \cdot x_{i+m}, \quad (7.77)$$

где  $x$  – исходный ряд;  $\tilde{x}$  – отфильтрованный ряд;  $h_m$  – весовые коэффициенты фильтра;  $M$  – параметр, определяющий число весов, которое для симметричных фильтров равно  $2 \cdot M + 1$ .

Изменение спектра исходного ряда в результате фильтрации может быть оценено с помощью частотной характеристики фильтра, которая представляет собой отношение амплитуды колебания данной частоты после фильтрации к амплитуде до фильтрации. Это отношение меняется вместе с частотой. Частотная характеристика любой дискретной симметричной сглаживающей или фильтрующей функции выражается следующим равенством:

$$R(f) = h_0 + 2 \cdot \sum_{m=1}^n h_m \cdot \cos(2 \cdot \pi \cdot m \cdot f \cdot \Delta t), \quad (7.78)$$

где  $R(f)$  – частотная характеристика;  $f$  – частота;  $h_m$  –  $m$ -ый вес, причем число  $m$  отличается от главного весового коэффициента  $h_0$ ;  $\Delta t$  – интервал времени между последовательными наблюдениями во временном ряде. Весовые функции и частотные характеристики наиболее употребительных низкочастотных фильтров приведены ниже:

1) скользящее равновесное среднее; весовая функция фильтра:

$$h_m = \begin{cases} \frac{1}{2 \cdot M + 1}, & \text{при } |m| \leq M; \\ 0, & \text{при } |m| > M, \end{cases} \quad (7.79)$$

где  $-2 \cdot M + 1$  – интервал сглаживания.

Частотная характеристика фильтра:

$$R(f) = \frac{\sin(\pi \cdot f \cdot (2 \cdot M + 1))}{\pi \cdot f \cdot (2 \cdot M + 1)}. \quad (7.80)$$

Простое осреднение значений величины в пределах соседних отрезков времени есть частный случай скользящего равновесного осреднения. Скольжение осуществляется здесь «скачками», длина которых равна длине интервала осреднения. Можно сказать, что, несмотря на широкое распространение, простое осреднение является едва ли не худшим видом сглаживания. Каждое значение исходного ряда, в отличие от других фильтров, используется лишь однажды. Кроме того, фазы ряда гармоник меняются на  $180^\circ$ , при сравнительно плохой фильтрации этих гармоник. Поэтому простое осреднение дает искаженное представление о процессе;

2) фильтр Бартлетта (треугольный)

$$h_m = \begin{cases} \frac{2}{2 \cdot M + 1} - \frac{4 \cdot |m|}{(2 \cdot M + 1)^2}, & \text{при } |m| \leq M; \\ 0, & \text{при } |m| > M; \end{cases} \quad (7.81)$$

$$R(f) = \left[ \frac{2 \cdot \sin\left(\frac{\pi \cdot f \cdot (2 \cdot M + 1)}{2}\right)}{\pi \cdot f \cdot (2 \cdot M + 1)} \right]^2; \quad (7.82)$$

характеристика (7.81), в отличие от (7.80), есть неотрицательная функция;

3) фильтр Тьюки

$$h_m = \begin{cases} \frac{1 + \cos\left(\frac{2 \cdot \pi \cdot m}{2 \cdot M + 1}\right)}{2 \cdot M + 1}, & \text{при } |m| \leq M; \\ 0, & \text{при } |m| > M; \end{cases} \quad (7.83)$$

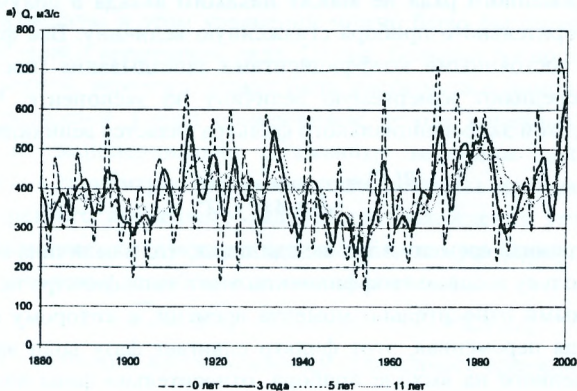
$$R(f) = \frac{\sin\left(\frac{\pi \cdot f \cdot (2 \cdot M + 1)}{2}\right)}{\pi \cdot f \cdot (2 \cdot M + 1)} \cdot \frac{1}{1 - (f \cdot (2 \cdot M + 1))^2}. \quad (7.84)$$

Этот фильтр характерен тем, что боковые лепестки его частотной характеристики весьма малы. К недостаткам фильтра Тьюки, так же, как и фильтра Бартлетта, следует отнести сравнительно медленное затухание его частотной характеристики. Располагая характеристиками низкочастотных фильтров, легко осуществить выделение гармоник с высокими частотами. Частотная характеристика высокочастотного фильтра  $R_H(f)$  просто выражается через частотную характеристику сглаживающего фильтра [Юдик, 1979].

$$R_H(f) = 1 + R_L(f). \quad (7.85)$$

Значения весовой функции высокочастотного фильтра равны по величине и обратные по знаку значениям весовой функции низкочастотного фильтра, за исключением центрального веса  $h_0$ , который дополняет до единицы величину центрального веса низкочастотного фильтра.

Результаты сглаживания колебаний годового стока по р. Припять – г. Мозырь приведены на рис. 7.6, где изображены сглаженные скользящие средние 3-, 5-, 11 летние колебания годового стока с постоянным (рис. 7.6а) и биномиальным весовым коэффициентами, симметрично убывающими от центрального члена осреднения (рис. 7.6б).





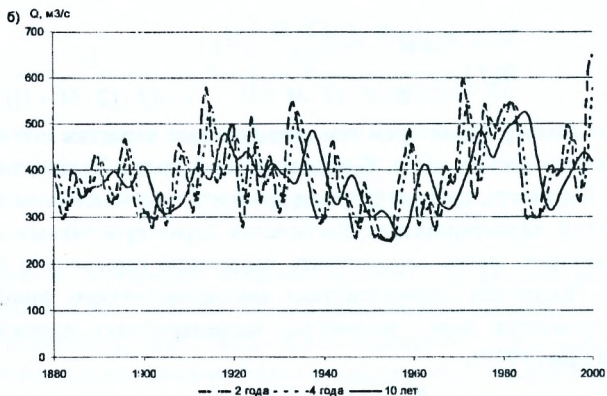


Рис. 7.6. Многолетние колебания речного стока р. Припять – г. Мозырь: а – скользящие средние арифметические; б – сглаженные колебания с использованием биномиального фильтра

#### и) Экспоненциальное сглаживание

Оно называется экспоненциальным, так как вклад различных величин временного ряда в полученную на выходе сглаженную величину экспоненциально убывает в зависимости от интервала времени, прошедшего до момента, к которому отнесена осредненная величина. Будущие значения временного ряда не вносят никакого вклада в полученную на выходе измерительного прибора сглаженную величину. Все физические приборы с постоянными коэффициентами запаздывания (постоянными времени) осредняют измеряемую величину по экспоненте. Частотной характеристикой экспоненциального фильтра является величина

$$R(f) = \frac{1}{\sqrt{1 + 4\pi^2 f^2 \lambda^2}}, \quad (7.86)$$

где  $\lambda$  – постоянная времени в тех же единицах, что и величина обратная  $f$ .

Поскольку весовые коэффициенты этого типа фильтра не являются симметричными относительно момента времени, к которому относится фильтрованная переменная, этот фильтр сдвигает фазу волн временного ряда, полученного на выходе прибора, относительно фазы волн той же частоты в исходном ряде. Этот сдвиг фаз является функцией частоты и выражается следующим образом:



$$\varphi = \operatorname{arctg}(-2\pi f\lambda), \quad (7.87)$$

где  $\varphi$  – сдвиг фаз.

Угол  $\varphi$  всегда отрицателен и лежит в пределах от 0 до  $90^\circ$ . Таким образом, по сравнению с исходными фильтрованные волны всегда запаздывают. В приведенных формулах не фигурирует величина  $\Delta t$ , поскольку временные ряды, экспоненциально сглаживаемые физическими приборами, являются непрерывными, а не дискретными.

Кроме того, с помощью обратного сглаживания можно получить не только истинный спектр временного ряда, но и оценить сам истинный временной ряд. Обратная сглаживающая функция для обращений экспоненциального сглаживания получается из дифференциального уравнения постоянным коэффициентом запаздывания. Это уравнение имеет следующий вид:

$$X - \bar{X} = \lambda \frac{d\bar{X}}{dt}, \quad (7.88)$$

где  $X$  – истинное значение переменной,  $\bar{X}$  – показание прибора, производящего сглаживание.

Решив это уравнение относительно  $X$ , получим:

$$X = \bar{X} + \lambda \frac{d\bar{X}}{dt}. \quad (7.89)$$

Производную в этом уравнении можно было бы определить графически из записи временного ряда. Однако большая точность вычисления производной, вероятно, может быть достигнута с помощью метода конечных разностей. Перед применением метода конечных разностей непрерывные экспоненциально сглаженные величины должны быть сперва сняты в равномерно распределенные моменты времени для получения дискретного временного ряда. После нахождения этих величин можно использовать следующую конечно-разностную форму дифференциального уравнения для обратного сглаживания ряда:

$$X_t = \bar{X}_t = \lambda(\bar{X}_t - \bar{X}_{t-1}) = (1 - \lambda)\bar{X}_t + \lambda\bar{X}_{t-1}, \quad (7.90)$$

где  $X_t$  и  $\bar{X}_t$  величины несглаженного и сглаженного временных рядов в момент  $t$ , а  $\bar{X}_{t-1}$  – сглаженное значение в момент  $t-1$ , т. е. момент, предшествующий  $\Delta t$ .

Постоянная времени  $\lambda$  измеряется в единицах  $\Delta t$ . Для обеспечения приемлемой точности восстановления истинного временного ряда интервал  $\Delta t$  между данными должен быть малой долей  $\lambda$ , скажем,  $1/4$  или меньше. Это уравнение для обратного сглаживания увеличивает амплитуду высокочастотных колебаний, не нарушая заметно низкочастотные колебания, и тем самым позволяет восстановить исходное соотношение спектральных компонент временного ряда. Полное обращение первоначального сглаживания невозможно, так как волна, которая ранее в результате сглаживания полностью была исключена из временного ряда, никогда не может быть восстановлена.

#### *к) Спектрально-временной анализ (СВАН)*

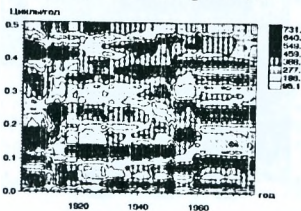
В последнее время для более тонких исследований временных рядов применяется спектрально-временной анализ (СВАН). В этом методе спектры вариации вычисляются на скользящих временных отрезках (временное окно) и изображаются в виде СВАН-диаграмм. Длина окна не должна быть слишком малой, поскольку при этом уменьшается точность спектрального анализа, а также не дается четкого представления о низких частотах. Однако завышенная длина окна также не дает полной информации, так как при этом будут сглаживаться высокочастотные колебания.

Периоды гармоник (или обратные им величины – частоты) на СВАН-диаграммах откладывают на вертикальной оси; время, отвечающее середине окна, – на горизонтальной. Глубина тона отвечает соответствующей амплитуде. Более сильная зачерненность на диаграмме соответствует большей амплитуде спектра. Повторяемость доминирующих циклов выражается в виде более или менее продолжительных зачерненных полос. Этот признак показывает продолжительность существования циклических изменений.

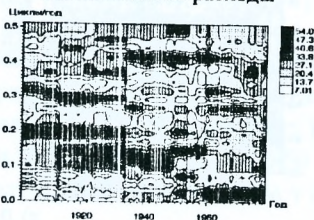
Определение параметра хаотизации также представляет собой вид спектрально-временного анализа. На оси абсцисс откладывается календарное время, а на оси ординат – степень «заполненности» спектра. Монохроматическому процессу соответствует нулевой уровень, а белому шуму – единица.

Полученные СВАН-диаграммы максимальных и минимальных расходов воды рек Беларуси приведены на рис. 7.7.

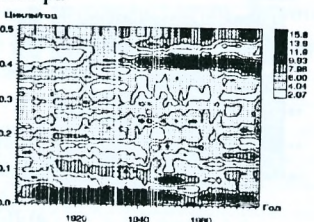
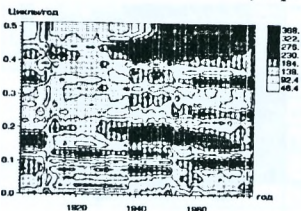
Максимальные расходы



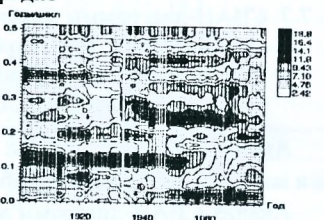
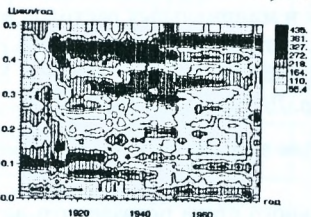
Минимальные расходы



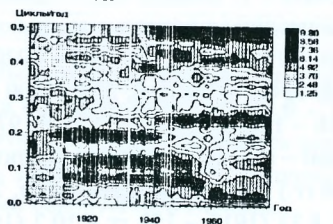
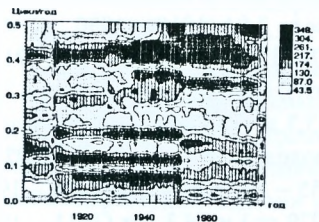
а) Припять – г. Мозырь



б) Неман – г. Гродно



в) Западная Двина – г. Витебск



г) Березина – г. Бобруйск

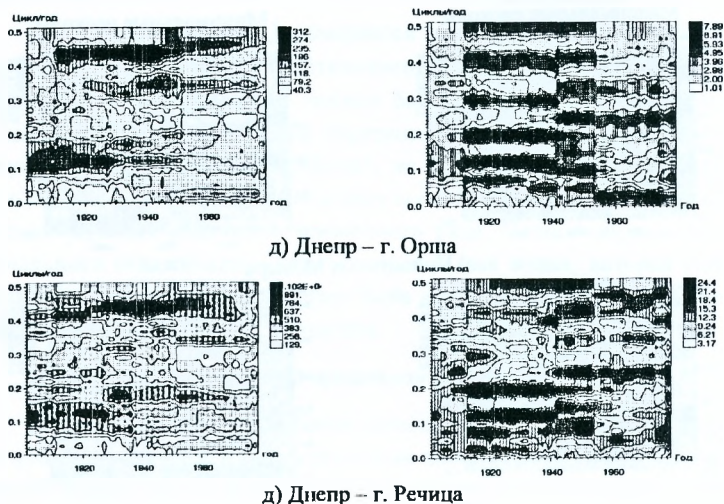


Рис. 7.7. СВАН-диаграммы максимальных и минимальных расходов воды рек Беларуси: а) р. Припять – г. Мозырь, б) р. Неман – г. Гродно, в) р. Западная Двина – г. Витебск, г) р. Березина – г. Бобруйск, д) р. Днепр – г. Орша, е) р. Днепр – г. Речица

Анализ диаграмм показал следующие закономерности максимального и минимального стоков (табл. 7.6).

Анализ СВАН-диаграмм максимального стока показал, что для р. Припять – г. Мозырь характерны достаточно много непродолжительных выделяющихся циклов длительностью около 2,04 (1910 – 1930 гг.), 2,27 (1930 – 1946 гг.), 2,38 (1900 – 1920 гг.), 2,70 (1930 – 1962 гг.), 4,17 (1910 – 1944, 1942 – 1980 гг.), 5,88 (1900 – 1910 гг.), самый мощный цикл – 7,69 (1916 – 1942 гг.), 20,0 (1936 – 1978 гг.), 33,3 лет (1928 – 1976 гг.). Для р. Неман – г. Гродно характерны циклы с продолжительностью 2,33 (1927 – 1980 гг.) – самый мощный цикл, 2,94 (1955 – 1980 гг.), 5,88 (1927 – 1976 гг.), 8,33 (1912 – 1952 гг.), 10,2 (1953 – 1979 гг.) и 14,3 лет (1911 – 1952 гг.). Для р. Западная Двина – г. Витебск возможно выделение следующих циклов 2,13 (1947 – 1980 гг.), 2,32 (1909 – 1976 гг.) – наиболее мощные циклы, 2,78 (1942 – 1967 гг.), 3,17 (1922 – 1954 гг.), 5,00 (1942 – 1961 гг.), 9,09 (1900 – 1933 гг.) и 12,5 лет (1917 – 1946 гг.). Для Березины у г. Бобруйска наблюдаются циклы с продолжительностью 2,04 (1948 – 1976 гг.),



самый мощный цикл – 2,40 (1910 – 1977 гг.), 2,98 (1937 – 1977 гг.) и три цикла 5,05, 8,13, 14,3 лет практически с одинаковой продолжительностью (1910 – 1952 гг.). Для максимального стока реки Днепр как у г. Орша, так и у г. Речица характерны три цикла, при этом первые являются самыми мощными. Их продолжительность составляет 2,27 (1910 – 1971 гг.), 2,86 (1920 – 1978 гг.), 7,69 (1900 – 1932 гг.) и 2,28 (1910 – 1964 гг.), 5,56 (1927 – 1969 гг.), 8,70 лет (1900 – 1938 гг.) соответственно для Днепра у г. Орша и у г. Речица.

Таблица 7.6

**Цикличность временных рядов расходов воды рек Беларуси  
(СВАН-анализ)**

Река-створ	Период, год
<b>Максимальный сток</b>	
Припять – г. Мозырь	2,04; 2,27; 2,38; 2,70; 4,17; 5,88; 7,69; 20,0; 33,3
Неман – г. Гродно	2,33; 2,94; 5,88; 8,33; 10,2; 14,3
Западная Двина – г. Витебск	2,13; 2,32; 2,78; 3,17; 5,00; 9,09; 12,5
Березина – г. Бобруйск	2,04; 2,40; 2,98; 5,05; 8,13; 14,3
Днепр – г. Орша	2,27; 2,86; 7,69
Днепр – г. Речица	2,28; 5,56; 8,70
<b>Минимальный сток</b>	
Припять – г. Мозырь	2,44; 3,45; 5,26; 6,67; 7,69; 33,4
Неман – г. Гродно	2,38; 14,3; 41,7
Западная Двина – г. Витебск	2,43; 2,74; 4,02; 5,13; 8,33; 50,0
Березина – г. Бобруйск	2,02; 2,39; 4,13; 5,10; 8,33; 12,5; 33,3
Днепр – г. Орша	2,03; 2,44; 3,13; 3,45; 3,91; 5,05; 8,33; 14,3; 29,4; 50,0
Днепр – г. Речица	2,04; 2,22; 3,03; 4,03; 5,00; 8,22; 13,3; 33,3

Для СВАН-диаграмм максимального стока наблюдается общая картина преобладания мощных циклов с обертонами около 2,0 - 3,0 в верхней части диаграмм, кроме Припяти, что можно объяснить различными условиями формирования стока, рельефом местности, геологическими и гидрогеологическими условиями Белорусского Полесья.

Анализируя СВАН-диаграммы минимального стока, можно отметить, что для Припяти у г. Мозырь возможно выделение следующих циклов – 2,44 (1948 – 1954 гг.), 3,45 (1934 – 1970 гг.), 5,26 (1900 – 1934 гг.), самые мощные циклы – 6,67 (1900 – 1944 гг.) и 7,69 (1932 – 1960 гг.), 33,4 (1951 – 1980 гг.). Для Немана у г. Гродно характерны циклы с продолжительностью 2,38 (1940 – 1979 гг.), мощные циклы 14,3 (1930 – 1980 гг.) и

41,7 (1900 – 1955 гг.). Западная Двина у Витебска имеет несколько выдающихся циклов с периодами 2,43 (1929 – 1973 гг.), 2,74 (1900 – 1929 гг.), 4,02 (1932 – 1978 гг.), 5,13 (1913 – 1944 гг.), самый мощный цикл – 8,33 (1900 – 1959 гг.) и 50,0 (1942 – 1978 гг.) Для Березины у Бобруйска доминирующие периоды складываются из циклов с продолжительностью 2,02 (1938 – 1980 гг.), 2,39 (1939 – 1971 гг.), 4,13 (1938 – 1980 гг.), сравнительно мощные циклы 5,10 (1908 – 1947 гг.) и 8,33 (1912 – 1956 гг.), 12,5 (1912 – 1960 гг.) и 33,3 года (1960 – 1979 гг.). Для Днепра у Орши выделяются циклы с периодом 2,03, 2,44, 3,13, 3,45, 5,05, 8,33, 14,3 (1912 – 1954 гг.), 3,91 (1933 – 1980 гг.), 29,4 (1932 – 1954 гг.), 50,0 (1954 – 1980 гг.), в то время как Днепра у Речицы они следующие – 2,04 (1912 – 1955 гг.), 2,22 (1938 – 1978 гг.), 3,03 (1942 – 1978 гг.); 4,03 (1938 – 1980 гг.), наиболее выдающиеся циклы 5,00 (1912 – 1955 гг.) и 8,22 (1912 – 1955 гг.), 13,3 (1913 – 1954 гг.), 33,3 года (1913 – 1936 гг.)

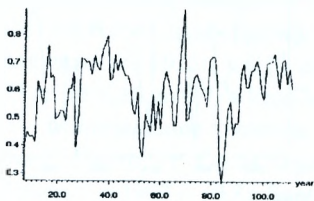
В отличие от СВАН-диаграммы максимального стока от минимального существенно различаются для рек Беларуси. Это может быть связано с генетическими условиями формирования минимальных расходов, различными уровнем антропогенной нагрузки на водосборы рек и особенностями грунтового питания рек различных регионов Беларуси. Существенный разброс в значениях продолжительности циклов минимального стока (часто встречаются циклы с обертонами порядка 30 – 50) возможен в связи с относительной неорганизованностью (отсутствием дружности) формирования минимальных расходов по отношению к максимальным, имеющим часто непродолжительные пикообразные гидрографы.

На рис. 7.8 приведены распределения параметра хаотизации для различных рек Беларуси.

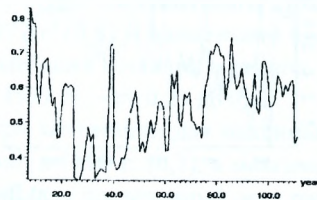
Для р. Припять – г. Мозырь в начале 60-х гг. отмечается спад параметра хаотизации до 0,28, а затем постепенный подъем до 0,72. В 1949 г. произошел скачок до 0,9. В основном размах колебаний от 0,4 до 0,8 для максимального стока. В 1900 – 1920 гг. параметр снизился до 0,33 и для минимального стока, затем происходит рост, размах колебаний от 0,4 до 0,7. Для р. Припять – г. Мозырь как для максимального, так и для минимального стоков в последние годы XX века наблюдается рост этого параметра.



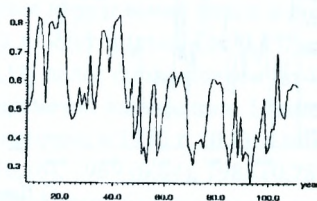
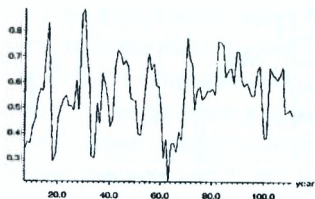
Максимальные расходы



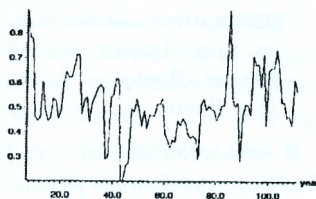
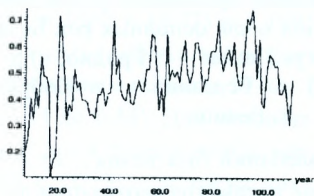
Минимальные расходы



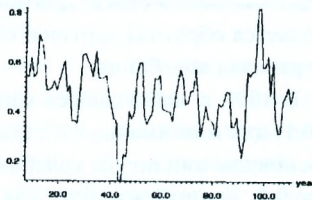
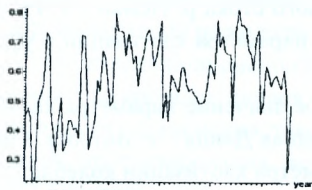
а) Припять – г. Мозырь



б) Неман – г. Гродно



в) Западная Двина – г. Витебск



г) Березина – г. Бобруйск

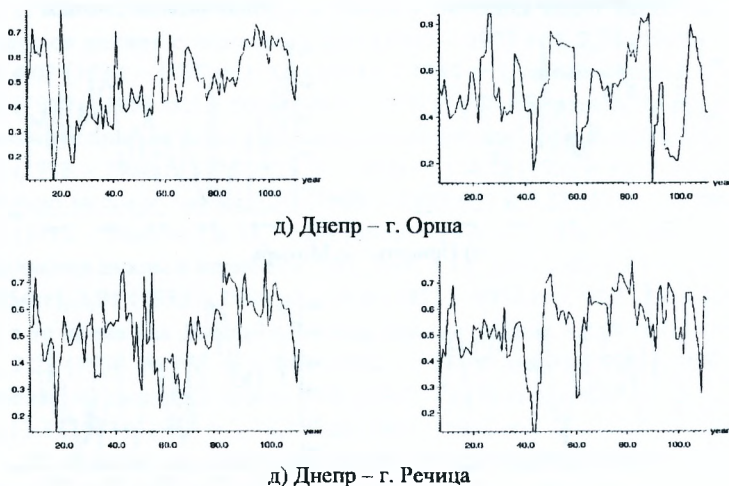


Рис. 7.8. Распределение параметра хаотизации временных рядов максимального и минимального расходов воды основных рек Беларуси: а) р. Припять – г. Мозырь, б) р. Неман – г. Гродно, в) р. Западная Двина – г. Витебск, г) р. Березина – г. Бобруйск, д) р. Днепр – г. Орша, е) р. Днепр – г. Речица

В начале прошлого века размах колебаний был велик – от 0,3 до 0,9. В 1944 г. происходит резкое понижение параметра хаотизации до 0,2, а затем постепенный рост и снижение размаха колебаний (от 0,45 до 0,75) для максимального стока. Для минимального стока р. Неман – г. Гродно наблюдается обратная картина: снижение параметра хаотизации и увеличение размаха колебаний.

В 1898 г. наблюдалось минимальное значение параметра хаотизации (0,1) для максимального стока р. Западная Двина – г. Витебск. Однако в основном амплитуда колебаний параметра хаотизации колеблется от 0,35 до 0,7. Данный параметр для минимального стока имеет наименьшее значение в 1925 г. (0,2) и затем постепенно увеличивается, а наибольший пик наблюдается в 1966 г. (0,9).

Для максимального стока р. Березина – Бобруйск параметр хаотизации растет до 1939 г., размах колебаний в этот период составляет от 0,35 до 0,8. Период 1939 – 1961 гг. параметр снизился и находился в пре-

делах 0,45 – 0,6, затем наблюдается рост параметра (до 0,8) и в 1979 г. опять начинается спад до 0,25 (1991 г.). Для минимального стока этой реки характерен наибольший пик в 1977 г. (0,8) и наименьшее значение наблюдалось в 1925 г. (0,1), размах колебаний составил в период 1881 – 1921 гг. от 0,36 до 0,7, в период 1929 – 1971 гг. – от 0,25 до 0,6.

Минимальное значение параметра хаотизации для максимального стока р. Днепр – г. Орша зафиксировано в 1898 г. (0,1), а максимальное значение – в 1903 г. (0,8). Затем наблюдается постепенный рост параметра с размахом колебаний от 0,3 до 0,7. Иначе обстоит дело с минимальным стоком этой реки; здесь наблюдается чередование пиков и спадов: с начала периода наблюдений и до 1910 г. идет подъем (от 0,37 до 0,84), затем до 1924 г. происходит спад до 0,18, далее снова подъем до 1939 г., и с 1941 г. начинается возрастание параметра хаотизации. Наименьшее значение этого параметра было зафиксировано в 1970 г. (0,1), а пики значений приходятся на 1910 (0,84), 1933 (0,76), 1969 (0,84), 1985 (0,78).

Размах колебаний хаотического параметра для максимального стока р. Днепр – г. Речица составляет от 0,14 до 0,8; наибольшие значения были в 1967 г. (0,77) и в 1981 г. (0,78), а наименьшее – в 1898 г. (0,14). Для всего графика наблюдается постепенное возрастание параметра за исключением периода 1936 – 1947 гг. (размах колебаний был небольшим от 0,24 до 0,44). Для минимального стока р. Днепр – г. Речица параметр хаотизации определяется тремя пиками – 1893г. (0,69), 1931 г. (0,73), 1964 г. (0,78) – и тремя спадами 1924 г. (0,13), 1940 г. (0,26), 1990 г. (0,28). Как видно из графика, максимальные и минимальные значения параметра с каждым циклом увеличиваются, и наблюдается общая тенденция возрастания параметра хаотизации.

Спектрально-временной анализ дает средства для отслеживания динамики различных циклов гидрологических рядов, а также оценки возможности применения такого анализа при создании статистических моделей прогнозирования водного режима рек. Совместный анализ СВАН-диаграмм и временных распределений параметров хаотизации показывает устойчивость циклов, а, следовательно, возможность построения прогностических оценок для минимального стока воды рассматриваемых рек и в меньшей степени – для максимальных расходов воды.

## 8. ОСНОВЫ СТАТИСТИЧЕСКОГО ПРОГНОЗИРОВАНИЯ ПРИРОДНЫХ ПРОЦЕССОВ

В гидрометеорологических исследованиях понятие статистического прогноза обычно связывается с задачей экстраполяции (интерполяции) и сглаживания случайного процесса. По известной реализации:

$$z(t) = x(t) + y(t), \quad (8.1)$$

в которой  $x(t)$  – детерминированная составляющая;  $y(t)$  – случайная составляющая; в случае  $y(t)=0$  (процесс без ошибок) прогноз сводится к чистой экстраполяции. В случае наличия ошибок  $y(t)$ , прежде чем определить истинное значение реализации  $x(t)$  в некоторый момент  $t+\Theta$ , необходимо отделить его от ошибки наблюдения. Это задача о сглаживании (фильтрации) случайного процесса. Задача об экстраполяции тесно связана со сглаживанием, так как реализация случайного процесса включает в себя ошибки измерения. При этом, задача экстраполяции сглаживанием состоит в том, чтобы по имеющейся реализации (8.1) на промежутке  $t_0+\Theta$  дать прогноз реализации  $x(t)$  в момент  $t+\Theta$ , при  $\Theta>0$ . При  $\Theta<0$  – имеет место задача интерполяции со сглаживанием. В математической постановке задачи предполагается, что математические ожидания процессов –  $m_x(t)$ ,  $m_y(t)$ , их автокорреляционные и взаимные корреляционные функции заданы. При этом обычно считают  $m_x(t)$  и  $m_y(t)$  равными 0. В противном случае, вместо  $x(t)$  и  $y(t)$  рассматриваются их центрированные случайные функции (аномалии) –  $\Delta x(t)$ ,  $\Delta y(t)$ . Математическое решение задачи статистического прогноза сводится к получению наилучшего результата по всему множеству реализаций, т. е. к нахождению такого оператора  $L$ , который в применении к множеству реализаций  $z(t)$  давал бы наилучшее, в некотором смысле, значение реализации  $x(t_0+\Theta)$

$$x(t_0 + \Theta) = L(x(t) + y(t)), \quad (8.2)$$

Для оценки качества прогнозирования вводится критерий качества прогнозирования, так или иначе связанный с ошибкой прогноза, – средним квадратом ошибки:

$$e^{-2}(t_0 + \Theta) = M\left(\left(x(t_0 + \Theta) - \hat{x}(t_0 + \Theta)\right)^2\right) \Rightarrow \min. \quad (8.3)$$

Чтобы вычислить предсказанное значение, нужно уметь выбрать правило вычисления ожидаемой оценки  $\hat{x}(t_0 + \Theta)$  – алгоритм прогноза.

Алгоритм предсказания  $\hat{x}$  должен связать с ним предысторию процесса и его вероятностные характеристики. Кроме того, качество алгоритма определится дисперсией ошибки прогноза.

Рассмотрим простейшие алгоритмы прогноза.

### 8.1. Особенности сверхдолгосрочного прогнозирования гидрометеорологических характеристик

Как показано выше, для долгосрочных и сверхдолгосрочных прогнозов, а также для прогнозов изменения климата могут служить такие предикторы, которые имеют большую «память».

В настоящее время считается, что поиск источников информации (предикторов, имеющих большую «память») о многолетних колебаниях гидрометеорологических характеристик является одной из основных задач современной климатологии.

Все существующие основные предпосылки и методы прогнозирования на долгие сроки могут быть классифицированы следующим образом:

- экстраполяция эмпирических или теоретических периодичностей (циклическостей) и случайных процессов;
- преемственность крупномасштабных особенностей циркуляции атмосферы;
- устойчивость (инерционность) климатических характеристик;
- диагноз погоды и климата (методы мировой погоды);
- множественная корреляция;
- разложение метеорологических полей по естественным ортогональным составляющим и полиномам Чебышева;
- численное моделирование.

В общем виде статистическая модель гидрометеорологических процессов может быть представлена следующим образом:

$$X(t) = f(t) + S(t) + \eta(t) + \xi(t), \quad (8.4)$$

где  $X(t)$  – реализация стохастического процесса, которая содержит:  $f(t)$  – трендовую компоненту,  $S(t)$  – периодическую компоненту,  $\eta(t)$  – стационарную стохастическую компоненту,  $\xi(t)$ , – компоненту «белого шума».

Успех прогнозирования, основанного на экстраполяции временных рядов, определяется стационарностью ряда и величиной квазиперио-



дической вариации, которая, в свою очередь, зависит от внутренних факторов.

Привлекательность метода экстраполяции периодичностей (циклическостей) состоит в том, что он позволяет предсказывать климат с большой заблаговременностью. Хотя метод и не выделяет годы экстремальных климатических явлений, он позволяет предвидеть систематические тенденции изменения климата. Доступность информации в данный момент не является критическим ограничением для таких прогнозов.

Другие возможные подходы связаны с подбором годов-аналогов по фазам солнечной активности и экстраполяцией ранее выявленных закономерностей связи *Солнце — нижняя атмосфера* с учетом влияния солнечной активности на устойчивость внутриатмосферных связей.

Особенно много попыток прогноза с учетом солнечной активности было сделано при выделении структурных элементов (циклов) и экстраполяции этих составляющих. При этом успешность прогноза снижается по мере увеличения периода экстраполяции, поскольку со временем связи прошлого с будущим постепенно размываются за счет включения новых факторов или изменения характера их действия. В основе этого метода лежит представление об однозначной причинности в природе (одинаковые причины при одинаковых обстоятельствах должны приводить к одинаковым результатам), т. е. детерминированности природных процессов, чего обычно не наблюдается.

Достаточно проблематичным является и использование длинных рядов, так как удлинение не оправдано по физическим соображениям. Ведь за большой промежуток времени существенно меняется вековой уровень различных факторов, определяющих климат, что приводит к следующему: выбранный большой период осреднения плохо характеризует период, состоящий из одного или нескольких десятилетий, поскольку гидрометеорологические ряды не всегда являются стационарными.

Не менее проблематичен и учет в гидрометеорологических процессах циклических колебаний с большими периодами, так как для ряда таких циклов их длительность может быть детерминированной, а амплитуда и фаза — случайными.

Следует отметить, что в последние годы достигнут некоторый успех в прогнозировании погоды на долгие сроки при использовании пока-



зателей изменчивости свойств подстилающей поверхности, а также крупномасштабных особенностей общей циркуляции атмосферы, таких, как интенсивность Южного и Северо-Атлантического колебаний и др. Использование статистических методов ограничивается следующими обстоятельствами: а) поток информации о крупномасштабных особенностях параметров климатической системы пока не доступен в реальном масштабе времени; б) в ряде случаев отсутствуют однородные ряды климатических параметров. Случается, что климатические параметры удовлетворяют условию (а), но не удовлетворяют условию (б), например, спутниковые данные. Рутинные наземные и морские наблюдения климатических параметров удовлетворяют условию (б), но зачастую не удовлетворяют условию (а).

Все это обусловило тот факт, что методы численного моделирования пока не получили достаточного развития в прогнозировании погоды на долгие сроки и прогнозировании климата. Это и определяет достаточно широкое использование статистических методов и диагностических исследований, которые в будущем могут быть базой численного моделирования.

## **8.2. Особенности применения статистических методов в прогнозах погоды**

Прежде чем перейти к статистическим методам в практике прогнозов погоды, укажем на возможные корни погоды, т. е. попытаемся назвать возможных конкурентов, которые могут «делать погоду»:

а) тепловые воздействия подстилающей поверхности, главным образом, океанов на атмосферу. Убедительно показано, что океан обладает огромной памятью, а, следовательно, корни прогностических зависимостей лежат в океане. Термическое состояние последнего может служить предиктором в долгосрочных прогнозах погоды;

б) фазовое превращение тепла в атмосфере в связи с влагооборотом. Исследования О. А. Дроздова и др. [Дроздов, 1994] показали, что корни прогностических зависимостей многим связаны с этим фактором;

в) поступление добавочной энергии от Солнца и Космоса значительно стимулирует развитие атмосферных процессов.

Рассмотрим некоторые методы прогнозов, где в качестве рабочего аппарата используются методы математической статистики.

1). *Метод мировой погоды*. Является старейшим методом. Его краеугольное положение состоит в следующем: крупномасштабные процессы атмосферной циркуляции развертываются не изолированно в отдельных частях земного шара, а находятся в взаимодействии с процессами того же масштаба, происходящими в других областях. Такая связь определяется тем, что атмосфера представляет непрерывную среду и (если не считать фронтальных разделов) возмущения, возникшие на данном участке атмосферы должны передаваться в соседние части, т. е. отрицательные аномалии давления в одних местах должны сопровождаться положительными в других, т. к. общая масса атмосферы неизменна.

Рабочим аппаратом метода мировой погоды является линейная множественная корреляция. С помощью этого метода выбираются предикторы. Для ряда предикторов можно найти физическую интерпретацию, например, выявлены колебания крупного масштаба в атмосферной циркуляции (южное колебание, Северно-Атлантическое, Северо-Тихоокеанское) связанные, по-видимому, с влиянием подстилающей поверхности и, возможно, другими факторами (солнечная активность).

При использовании метода мировой погоды допускается, что: а) временные ряды метеорологических элементов являются линейными; б) временные зависимости устойчивы во времени, если разница между коэффициентами, вычисленными для двух сроков, не превышает суммы стандартных погрешностей этих коэффициентов.

К сожалению, оба эти допущения не выполняются. Уже само наличие циклических процессов в атмосфере не дает уверенности в устойчивости связей. Кроме того, наличие тренда оказывает отрицательное влияние на устойчивость прогностических связей. Однако последующий опыт доказал приемлемость метода линейной корреляции для практики прогнозов, поскольку, очевидно, этот метод улавливает некоторые соотношения, вызванные циклическостью, с одной стороны, а с другой стороны, не все прогностические связи определяются одной циклическостью. Не случайно в последние годы вновь обратились к методу мировой погоды. ЭВМ дали возможность для вычисления коэффициентов корреляции и построения уравнений регрессии с использованием большего материала.

Коэффициенты уравнений регрессии при этом получили название эмпирических функций влияния (ЭФВ). Основная идея ЭФВ заключается

в сочетании гидродинамической теории прогноза со статистическими методами экстраполяции. Форма связи между прогнозируемым значением метеорологического элемента и различными физическими величинами или предсказателями, а также сами эти факторы, принимаются исходя из результатов гидродинамической теории. Подбор же коэффициентов и параметров, входящих в уравнение связи, производится методами математической статистики.

Метод ЭФВ получил свое название по аналогии с теоретическими функциями влияния, входящими в решения уравнений динамики. Физическую основу этого метода представляет система уравнений гидродинамики, решение которой относительно тенденции геопотенциала в общем виде может быть записано так:

$$\frac{\partial H}{\partial t} = \iiint_V g F dv, \quad (8.5)$$

где  $g$  – функция Грина (теоретические функции влияния);  $F$  – функция, определяющая изменение геопотенциала.

Несмотря на то, что функция  $F$  – не описывает всех факторов, вызывающих изменение геопотенциала, статистическим путем можно учесть влияние тех из них, физическая природа которых недостаточно ясна или невозможна их количественная оценка. При практической реализации интегралы типа (8.5) заменяются суммами парных произведений

$$\delta H = \sum_{i=1}^n C_i F_i, \quad (8.6)$$

где  $F_i$  – значение функции  $F$  в некоторых точках или разные моменты времени;  $C_i$  – коэффициенты, определяемые приближенно.

Значение функции (8.6) в некоторой точке  $M$ , находящейся на расстоянии  $\rho$  от начала координат является тем весом, с которым значения величины  $F$  в этой же точке входят в интеграл

$$\left( \frac{\partial H}{\partial t} \right)_{\rho=0} = \frac{1}{2\pi} \int_0^{2\pi} F(\rho, y) \ln \frac{1}{\rho} \rho d\rho y, \text{ который является решением уравнения}$$

Пуассона для плоскости  $\gamma^2 \frac{dH}{dt} = F$ . Здесь  $\ln \frac{1}{\rho} = q$ .

Последнее выражение представляет собой стандартное уравнение множественной регрессии. Методы математической статистики позволя-

ют определить коэффициенты  $C_i$ , являющиеся конечно-разностными аналогами теоретических функций влияния, используя значения  $F_i$  и  $\delta H$  за прошлые годы.

Физически эти коэффициенты характеризуют «вес» функций в окружающих точках и изменении геопотенциала в данной точке и зависят от вида уравнения, его конечно-разностного аналога и выбранной системы точек. Исходным моментом ЭФВ является выбор функций  $F$ . Если  $F$  есть функция только линейных комбинаций значений метеорологического элемента, то прогностическая схема считается линейной. Если в качестве  $F$  приняты нелинейные функции начальных полей (например, уравнение тенденции геопотенциала), то такая схема является нелинейной.

В последние годы проведены обширные работы по выбору предикторов исходя из физических представлений о взаимодействии энергетических процессов на подстилающей поверхности и в атмосфере. Характеристики состояния подстилающей поверхности рассматриваются как косвенные показатели этих процессов.

Не описывая другие методы прогнозов погоды отметим, что они широко используются циркуляционные индексы, поскольку они отражают не только адвекцию воздушных масс, но и другие факторы – радиационные и конденсационные. Понятно, что эффективное излучение в случае арктического вторжения будет иметь другое значение, чем при адвекции атлантического воздуха.

И в заключение отметим, что в настоящее время любые, без исключения, методы прогнозов используют в большей или меньшей мере статистические методы прогнозов.

### 8.3. Вероятностное прогнозирование значений случайного процесса

Оценка значения в реализации случайного процесса (в силу случайности физических явлений) в будущем (в момент времени –  $t$ ) не может быть вычислена по точной формуле, но может быть описана в вероятностном виде. В случае стационарного и эргодического процессов  $P(x, t)$  не зависит от времени и может быть определена по единственной реализации  $x(t)$  как:

$$P(x) = P(x(t) \leq \xi) = \lim_{T \rightarrow \infty} \left( \frac{T(x(t) \leq \xi)}{T} \right), \quad (8.7)$$

где  $T[x(t) \leq \xi]$  – общее время, в течение которого реализация  $x(t)$  находится не выше уровня  $\xi$ . В этом случае значение  $x(t)$  в произвольный момент времени не превышает данного значения  $\xi$ . При  $x \rightarrow -\infty$  интегральная функция распределения  $P(x) = P(x, t)$  стремится к 0, при  $x \rightarrow \infty$  – стремится к 1. По характеру изменения функции распределения от 0 до 1 различаются случайные процессы с разной вероятностной структурой. Можно сказать, что плотность вероятности определяет скорость изменения функции распределения, поскольку вероятность различных событий можно находить интегрированием плотности вероятности на кривой:

$$P(x_1 \leq x(t) \leq x_2) = \int_{x_1}^{x_2} P(x) \cdot dx = P(x_2) - P(x_1). \quad (8.8)$$

Функция распределения существует как для непрерывных, так и для прерывных случайных величин и является универсальной характеристикой случайных величин, так как плотность характеризует их с вероятностной точки зрения. Зная функцию распределения случайной величины, можно найти вероятность ее попадания на заданный участок, которая равна приращению функции распределения на этом участке. При  $x_1 \rightarrow 0$ , получим  $x_2$ ,

$$P(0 \leq x(t) \leq x_2) = \int_{-\infty}^{x_2} P(x) \cdot dx = P(x_2), \quad (8.9)$$

т. е. площадь под графиком плотности вероятности левее точки  $x_2$  равна значению дифференциальной функции распределения в точке  $x_2$ . Таким образом, прогнозирование вероятности того или иного элемента может быть осуществлено при знании или прогнозировании функции распределения. Задача прогнозирования при использовании вероятностных моделей заключается в определении по кривой распределения вероятностей величины параметра  $x$ , такого, когда вероятность  $P(x)$  равна заданному значению  $P$ . Следует помнить, что точность прогноза, с вероятностной точки зрения, в этом случае, будет зависеть от точности прогноза функции распределения.



#### 8.4. Прогноз по последнему значению

Прогнозирование по последнему значению реализации (инерционный прогноз), заключается в том, что в качестве предсказанного значения  $\hat{x}(t_0 + \Theta)$  принимается значение  $x(t_0)$ :

$$\hat{x}(t_0 + \Theta) = x(t_0). \quad (8.10)$$

Предсказанное значение здесь не зависит от предыстории прогноза (предыстория представлена лишь одной точкой – последним значением  $x(t_0)$ ), а вероятностные характеристики не учитываются совсем. Алгоритм прогноза заключается в умножении значения последнего наблюдения  $x(t_0)$  на 1, т. е. не требует выполнения никаких вычислительных операций. Прогноз, таким образом, можно выполнить, ничего не зная о процессе, не производя никаких вычислений. Однако точность прогноза очень низкая. Возможная ошибка прогноза по алгоритму (8.10) здесь определяется как:

$$e(t_0 + \Theta) = x(t_0 + \Theta) - x(t_0), \quad (8.11)$$

а ее средний квадрат  $e^2$ , если  $m_x=0$ ,

$$\bar{e}^2(\Theta) = M((x(t_0 + \Theta) - x(t_0))^2) = \sigma_x^2 - 2 \cdot r_x(\Theta) + \sigma_x^2 = 2 \cdot (\sigma_x^2 - r_x(\Theta)). \quad (8.12)$$

Средний квадрат ошибки прогноза растет от 0, при  $\Theta=0$ , когда  $r_x(0) = \sigma_x^2$ , до  $2 \cdot \sigma_x^2$ , при  $\Theta=\infty$ , когда  $r_x(\infty) = 0$ . Об истинном качестве этого способа прогноза можно говорить после сравнения полученной ошибки с ошибками других алгоритмов и способов прогноза. Простота этого способа обеспечила ему широкое распространение.

#### 8.5. Прогноз по математическому ожиданию

Прогнозирование по математическому ожиданию заключается в том, что в качестве предсказанного значения  $\hat{x}(t_0 + \Theta)$  принимается математическое ожидание  $m_x$ . Как и в предыдущем способе, предсказанное значение здесь не зависит от времени прогноза  $\Theta$ . Различие заключается в том, что, хотя не требуется никакой информации о предыстории, необходимы сведения о свойствах процесса – о его математическом ожидании. Алгоритм прогноза не требует никаких вычислительных операций. Ошибка прогноза вычисляется по зависимости:



$$e(\Theta) = x(t + \Theta) - m_x, \quad (8.13)$$

и представляет собой отклонение процесса от среднего в момент  $t_0 + \Theta$ .

Средний квадрат ошибки не зависит от времени прогноза и равен дисперсии прогноза:

$$\bar{e}^2(\Theta) = M((x(t_0 + \Theta) - m_x)^2) = \sigma_x^2. \quad (8.14)$$

При малой заблаговременности  $\Theta$  прогноз по последнему значению явно предпочтителен, однако, после получения некоторой критической величины заблаговременности прогноза  $\Theta^*$ , когда  $\bar{e}^2(\Theta^*) = \sigma_x^2$ , метод прогноза по математическому ожиданию дает большую точность. Наконец, при  $\Theta \rightarrow \infty$ , квадрат ошибки прогноза по математическому ожиданию вдвое меньше, чем по последнему отсчету. Так, предсказывая расход воды в реке на несколько дней, руководствуясь инерцией, ориентируемся на ее текущее состояние, совершенно игнорируя средние многолетние величины. Наоборот, пытаясь предвидеть летом весеннее половодье, напротив, прежде всего ориентируемся на «норму» половодья.

### 8.6. Статистический прогноз по одной точке

Стационарный эргодический процесс может быть как ансамблем реализаций, так и одной реализацией неограниченной длительности. Сечения ансамбля представляют собой случайные величины, функция распределения которых отождествляется с одномерной функцией распределения процесса. Обозначим случайную величину  $x(t_0)$  как сечение процесса в момент  $t_0$  – через  $x$ , а сечение  $x(t_0 + \Theta)$  – через  $y$  и будем рассматривать систему двух случайных величин  $x, y$  – последнего значения предыстории и предсказанного значения. Компоненты системы  $x$  и  $y$  подчинены одномерным нормальным законам –  $N_1(m_x, \sigma_x)$ ,  $N_2(m_y, \sigma_y)$ . Алгоритм прогноза в рассматриваемом способе формулируется так, что в качестве предсказанного значения  $\hat{x}(t_0 + \Theta)$  выступает условное математическое ожидание  $m_{y/x}$  величины  $y$ , при условии, что  $x = x(t_0)$  –

$$\hat{x}(t_0 + \Theta) = m_{y/x}, \quad (8.15)$$

где аналогично уравнению регрессии:

$$m_{y/x} = m_y + r \frac{\sigma_y}{\sigma_x} \cdot (x - m_x); \quad (8.16)$$

$$\sigma_{y/x} = \sigma_y \cdot \sqrt{1 - r^2}. \quad (8.17)$$

Известно, что закон, постулированный при условии, что первая компонента  $x$  приняла определенное значение, называется условным законом распределения и имеет вид:

$$f_{(y/x)} = \frac{f(x, y)}{f(x)} = \frac{1}{\sigma_y \cdot \sqrt{2 \cdot \pi} \cdot \sqrt{1 - r^2}} \cdot \exp\left(-\frac{1}{2(1 - r^2)}\right) \cdot \left(\frac{y - m_y}{\sigma_y} - r \cdot \frac{x - m_x}{\sigma_x}\right)^2 \quad (8.18)$$

или

$$f_{(y/x)} = \frac{1}{\sigma_y \cdot \sqrt{2 \cdot \pi} \cdot \sqrt{1 - r^2}} \cdot \exp\left(-\frac{1}{2(1 - r^2)}\right) \cdot \left(y - m_y - r \cdot \frac{\sigma_y}{\sigma_x} \cdot (x - m_x)\right)^2. \quad (8.19)$$

С учетом (8.18) и (8.19), получим плотность нормально распределенной условной случайной величины:

$$f_{(y/x)} = \frac{1}{\sqrt{2 \cdot \pi} \cdot \sigma_{y/x}} \cdot \exp\left(-\frac{y - m_{y/x}}{2 \cdot \sigma_{y/x}^2}\right). \quad (8.20)$$

Из (8.20) следует, что, при изменении одной из компонент, вид закона распределения второй компоненты не меняется, а меняется лишь его параметр  $m_{y/x}$  – условное математическое ожидание (8.16). Условная дисперсия (8.17) от значения  $x$  также не зависит. Зависимость  $m_{y/x}$  от  $x$  линейна и называется регрессией  $y$  на  $x$ . Ошибка прогноза по соотношению (8.16) определяется как -

$$e(t_0 + \Theta) = x(t_0 + \Theta) - \hat{x}(t_0 + \Theta) = y - m_{y/x} \quad (8.21)$$

и представляет отклонение случайной величины  $y$  от своего условного математического ожидания, а средний квадрат ошибки  $-e^2(\Theta) = M(y - m_{y/x})^2$  равен условной дисперсии  $\sigma_{y/x}^2$ . Учитывая (8.17) и (8.18), установим, что:

$$\hat{x}(t_0 + \Theta) = m_{x/y} = m_y + r_{x/y} \cdot \frac{\sigma_y}{\sigma_x} \cdot (x - m_x) \quad (8.22)$$

и

$$e^2(\Theta) = \sigma_{y/x} = \sigma_y^2 \cdot (1 - r^2). \quad (8.23)$$

Поскольку процесс  $x(t)$  стационарен, математические ожидания и дисперсии сечений одинаковы:  $\sigma_y = \sigma_x = \sigma$ ;  $m_y = m_x = m$ . Коэффициент корреляции  $r_{xy}$  равен значению нормированной корреляционной (автокорреляционной) функции

$$r_{xy} = r(\Theta). \quad (8.24)$$

Теперь алгоритм прогноза прост:

$$\hat{x}(t_0 + \Theta) = m + r(\Theta) \cdot (x(t_0) - m), \quad (8.25)$$

а средний квадрат ошибки оказывается зависящим от  $\Theta$ :

$$e^{-2} = \sigma^2 \cdot (1 - r(\Theta)). \quad (8.26)$$

Алгоритм предполагает знание отклонения процесса от среднего в момент  $t_0$ , т. е. одной точки предыстории, знание нормированной корреляционной функции и математического ожидания.

### 8.7. Динамико-стохастический метод сверхдолгосрочного прогноза

Математический аппарат прогнозирования динамико-стохастическим методом содержит в своей основе предположение о стационарности прогнозируемых процессов во времени и наличии в процессе внутрирядных связей (даже при сдвиге степени  $\tau > 1$ ). Задача линейного экстраполирования (прогноза) стационарной случайной последовательности, удовлетворяющей условиям:  $m_x = \text{const}$ ,  $r_x = r(\tau) \neq 0$ , заключается в необходимости подбора таких действительных коэффициентов  $k_\tau$ , при заданных  $m > 0$  и  $\Theta > 0$ , при которых линейная комбинация:

$$\hat{x}_0(t_0 + \Theta) = \bar{x} + k_1 \cdot \Delta x_{t-1} + k_2 \cdot \Delta x_{t-2} + \dots + k_m \cdot \Delta x_{t-m} = \bar{x} + \sum_{\tau=1}^m k_\tau \cdot \Delta x_{t-\tau} \quad (8.27)$$

является наиболее точным приближением к случайной величине  $\hat{x}(t)$ . В соотношении (8.27):  $\bar{x}_0 \cdot (t_0 + \Theta)$  – прогнозируемое на момент  $t = t_0 + \Theta$  значение исследуемой величины  $x(t)$ ;  $\bar{x}$  – среднее значение  $x(t)$ ;  $\Delta x_{t-1}$ ,  $\Delta x_{t-2}$

...,  $\Delta x_{t-m}$  – последовательность предшествующих моменту  $t$  значений величины  $x(t)$  в отклонениях от  $\bar{x}$ ;  $k_1, k_2, \dots, k_m$  – коэффициенты обратной связи, определяемые путем решения системы уравнений:

$$\left. \begin{aligned} k_1 + k_2 \cdot r_2 + \dots + k_m \cdot r_{m-1} &= r_1, \\ k_1 \cdot r_1 + k_2 + \dots + k_m \cdot r_{m-2} &= r_2, \\ \dots & \\ k_1 \cdot r_{m-1} + k_2 \cdot r_{m-2} + \dots + k_m &= r_m \end{aligned} \right\}, \quad (8.28)$$

где  $r_1, r_2, \dots, r_m$  – последовательность значений корреляционной функции  $r(\tau)$ ;  $m$  – оптимальный период обратной связи.

Для определения периода обратной связи  $m_{opt}$  рекомендуется выполнять проверочное прогнозирование, при значениях  $m$ , последовательно увеличивающихся от 1 до 30 лет. В качестве оптимального выбирается то значение, при котором ошибка прогноза  $e^{-2}$  становится минимальной. Оценка точности прогнозов, как правило, производится по последовательности эмпирических коэффициентов связи между фактическими  $\Delta x_{if}$  и прогностическими  $\Delta x_{ip}$  аномалиями ряда. При этом прогностическое значение  $\Delta x(t+\theta)$  отыскивается по уравнению авторегрессии вида:

$$\Delta x(t + \Theta) = \sum_{k=0}^m \alpha_k \cdot \Delta x(t - k). \quad (8.29)$$

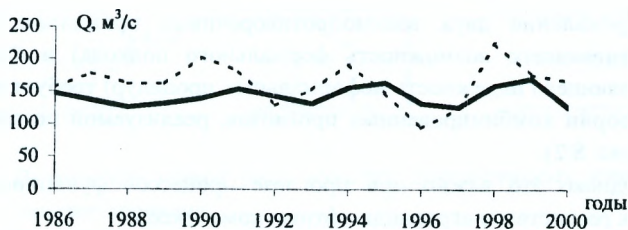
Коэффициенты  $\alpha_k$  для каждого заданного значения  $\theta$  определяются исходя из условия минимума ошибки экстраполяции при решении системы уравнений:

$$r_{\Delta x}(\Theta + j) = \sum_{k=1}^m \alpha_k \cdot r_{\Delta x}(k - j); \quad \text{при } j=1, 2, \dots, m, \quad (8.30)$$

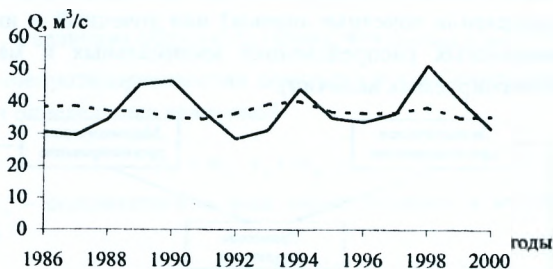
где  $r_{\Delta x}(\tau)$  – корреляционная функция отклонений.

Число слагаемых  $m$  в сумме  $\sum_{k=1}^m$  следует выбирать таким, чтобы корреляционные моменты  $r_{\Delta x}(k-j)$  определялись по данным наблюдений в  $m$  – точках с требуемой надежностью. На рис. 8.1 приведены результаты прогноза годового стока некоторых рек Беларуси с заблаговременностью 1 год.

Днепр – Речица



Березина – Бобруйск



Припять – Мозырь

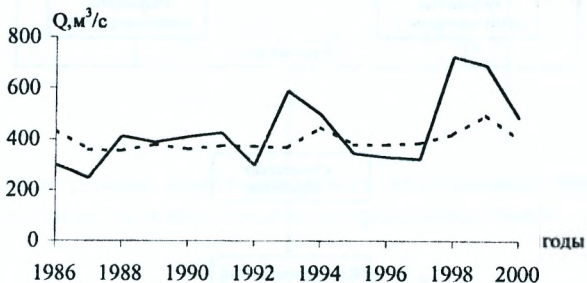


Рис. 8.1. Результаты прогноза годового стока рек Беларуси с заблаговременностью один год: — — — фактический сток, - - - прогнозные величины стока

### 8.8. К вопросу о точности прогнозирования

Проявление двух взаимопротиворечивых принципов Беллмана (ограничивающего возможность формального подхода) и Форрестера (определяющего надежность неформальных процедур) требует использования теории комбинированных прогнозов, реализуемой по следующей схеме (рис. 8.2).

Однако это важно при прогнозе процессов функционирования больших геосистем и агроландшафтных комплексов.

Одним из ключевых моментов такого комбинирования является определимость противоречивости, так как математическое прогнозирование дает точечный и интервальный прогнозы при данном уровне вероятности, а эвристический – точечный и его дисперсию (при нормальном законе распределения точечных оценок) или точечный и интервальный (при равновероятных распределениях минимальных и максимальных значений прогнозируемых величин).



Рис. 8.2. Схема комбинированного прогноза



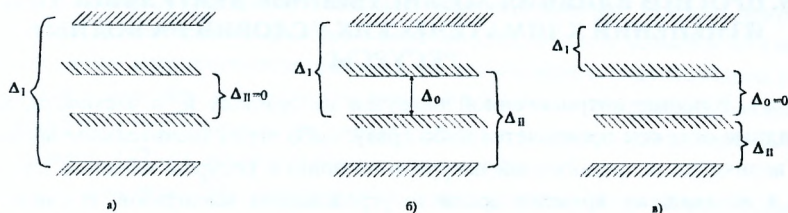


Рис. 8.3. Возможные схемы расположения доверительных интервалов прогнозирования

Для случая б) характерно частичное перекрытие интервалов, а для случая в) – отсутствие общей области ( $\Delta_0=0$ ).

Для любого случая результаты прогнозирования будут «непротиворечивы», если точечные прогнозы принадлежат области  $\Delta_0$ , т. е.  $\hat{y}_3, \hat{y}_m \in \Delta_0$ , а величина области  $\Delta_0$  такова, что  $\Delta_0/\Delta_{\min}(I, II) \geq k = 0,5$ .

При «непротиворечивости» результатов комбинированный точечный прогноз определяется величиной

$$\hat{y}_k = \omega_3 \cdot \hat{y}_3 + \omega_m \cdot \hat{y}_m, \quad (8.31)$$

где  $\omega_3$  и  $\omega_m$  – соответственно веса эвристического и математического прогнозов и

$$\left. \begin{aligned} \omega_3 &= D(\hat{y}_m) / D(\hat{y}_3) + D(\hat{y}_m); \\ \omega_m &= D(\hat{y}_3) / D(\hat{y}_3) + D(\hat{y}_m); \\ \omega_m + \omega_3 &= 1. \end{aligned} \right\} \quad (8.32)$$

Для определения комбинированного интервального прогноза, решив задачу композиции двух законов распределения, имеем

$$\tilde{y}_k = \omega_3 \cdot \tilde{y}_3 + \omega_m \cdot \tilde{y}_m. \quad (8.33)$$

## 9. ПРОГНОЗ ВЛИЯНИЯ ХОЗЯЙСТВЕННОЙ ДЕЯТЕЛЬНОСТИ И ИЗМЕНЕНИЙ КЛИМАТИЧЕСКИХ УСЛОВИЙ НА ВОДНЫЕ РЕСУРСЫ

Влияние антропогенной нагрузки на природу и, в частности, на водные объекты проявляется либо сразу, либо через значительное время. Увеличение антропогенной нагрузки на водные ресурсы началось давно и к настоящему времени достигло угрожающих масштабов. В связи с этим наряду с разработкой систем мер по охране вод, устранению наступившего загрязнения и истощения водных ресурсов необходимо прогнозирование их состояния. Многообразие видов хозяйственной деятельности, накладываемых на циклические климатические изменения, а также несовершенство применяемых прогностических методов делают задачу долгосрочных прогнозов изменений водных ресурсов весьма трудной, так как при этом необходимо оценивать не только количественные, но и качественные характеристики возможных изменений.

Существуют самые различные методы количественной оценки влияния антропогенных факторов на гидрологический режим поверхностных и подземных вод, которые могут быть объединены в две группы (Шикломанов, 1976, 1979):

- методы, основанные на исследовании многолетних колебаний стока в опорных гидростворах в совокупности с анализом изменения естественных метеорологических факторов и развития хозяйственной деятельности;

- воднобалансовые методы, отдельно учитывающие водозабор, изменения элементов водного и теплового балансов в бассейне реки в результате воздействия каждого вида хозяйственной деятельности в отдельности.

Методы, входящие в первую группу, позволяют произвести оценку влияния на гидрологический режим водосбора всего комплекса антропогенных факторов или отдельных видов хозяйственной деятельности: орошительных и осушительных мелиораций руслового регулирования, урбанизации, вырубки лесов. Общим преимуществом методов этой группы является возможность получить некоторые результаты на основании многолетних гидрометеорологических материалов без постановки сложных и дорогостоящих экспериментов. Недостатком методов является необходимость иметь длительные периоды наблюдений, охватывающих различные этапы развития хозяйственной деятельности на водосборе. Методы первой группы не вскрывают физической сущности процессов

на водосборе, не позволяют надежно выделить влияние каждого природного и хозяйственного фактора в отдельности, не дают возможности оценить роль антропогенного фактора, не получившего значительного развития в бассейне или если его влияние находится в пределах точности измерения гидрологических характеристик в замыкающем створе (Шикломанов, 1976). Сложности возникают в случае:

- на водосборе одновременно действуют несколько антропогенных факторов разной направленности;
- отсутствуют многолетние гидрометрические наблюдения;
- необходим прогноз количественных изменений водных ресурсов на перспективу.

В связи с этим широко используются воднобалансовые методы, позволяющие вскрыть физическую сущность процессов, происходящих на водосборе под влиянием хозяйственных мероприятий, оценить роль каждого антропогенного фактора не только в прошлом, но и в будущем.

К этой группе относятся методы математического моделирования, описывающие процессы, происходящие на водосборе, системой уравнений (при разных начальных и граничных условиях, а также при различных значениях параметров). Методы позволяют в короткие сроки количественно оценить влияние самого разнообразного сочетания естественных и антропогенных факторов на гидрологические характеристики, то есть рассчитать прогнозный гидрологический режим и водный баланс водосбора с учетом различных вариантов хозяйственного освоения территорий. Наиболее широко методы моделирования применяются при изучении движения подземных вод и трансформации паводочных волн в руслах рек и водохранилищах, для оценки влияния горных выработок, крупных водозаборных сооружений, осушительной дренажной сети, обводнении и т. п. При использовании математического моделирования необходимы надежные исходные данные, но также полнота и обоснованность принятых расчетных уравнений, так как даже для одних и тех же водосборов многочленные параметры уравнений оказываются неустойчивыми или мало обоснованными.

К недостаткам балансовых методов следует отнести низкую точность измерений и расчетов отдельных элементов водного и теплового баланса, а также трудности при переносе данных, полученных на стоковых площадках, орошаемых и распаханых массивах, малых бассейнах, на крупные речные водосборы. В то же время наличие на крупных водосборах не одного, а не скольких видов хозяйственной деятельности также осложняет задачу прогнозирования.

Следует отметить, что для разработки перспективных долгосрочных планов устойчивого развития различных территорий первостепенное значение имеют вопросы возможных изменений водных ресурсов не только под влиянием хозяйственной деятельности, но и климата.

### **9.1. Оценка современных трансформаций водного режима под воздействием технических мелиораций и хозяйственной деятельности на водосборах\***

Мелиоративные мероприятия в пределах разных речных водосборов представляют сложный многосторонний комплекс, отдельные элементы которого по-разному влияют на водный режим.

Как показывает проведенный анализ послемелиоративных изменений годового стока с крупных водосборов площадью в несколько тысяч квадратных километров, они невелики и лежат в пределах погрешностей расчетов ( $\pm 15\%$ ). Поэтому при оценке возможных изменений водных ресурсов их можно не учитывать. Построенные совместные разностные интегральные кривые речного стока р. Припять – г. Мозырь (площадь водосбора  $101000 \text{ км}^2$ ) и атмосферных осадков метеостанции Василевичи подтверждают вышесказанное.

Наиболее заметно осушение сказывается на водных ресурсах малых водосборов, площадью до  $2000-3000 \text{ км}^2$ . Наблюдения на таких водосборах показывали, что в первые годы после осушения за счет уменьшения суммарного испарения и сработки запасов грунтовых вод годовой сток может увеличиваться на  $20 - 30\%$ , а меженный – на  $50 - 70\%$ , иногда и более. Максимальные расходы уменьшаются в зависимости от характера осушительной сети на объекте и состояния аккумулирующей емкости зоны аэрации. Минимальные месячные и суточные расходы летней межени устойчиво возрастают в  $1,5 - 2$  раза, вследствие улучшения условий дренирования грунтовых вод и облегчения стока летних осадков. В речном стоке повышается доля подземного питания, что улучшает его структуру. После сельскохозяйственного освоения осушенных земель возможно некоторое увеличение суммарного испарения и соответствующее снижение годового стока до первоначальной величины, а иногда и ниже ее. Сохраняется устойчиво повышенная летняя межень.

Для выявления пространственной структуры изменения годового стока использованы данные гидрометрических наблюдений за расходами воды рек Беларуси. Имеющиеся ряды наблюдений были разбиты на два

\* Раздел написан совместно с к.т.н. В. В. Лукшей

периода: с начала наблюдений по 1965 год (начало крупномасштабных мелиораций) и с 1966 года по 2000 год. При этом выбраковывались ряды с периодом наблюдений менее 15 лет хотя бы за один из периодов. После выбраковки осталось 62 реки-створа, для которых определены величины изменения годового стока как

$$k_i = \frac{Q_{cp2} - Q_{cp1}}{\bar{Q}}, \quad (9.1)$$

где  $Q_{cp1}$  и  $Q_{cp2}$  – средние значения годового стока за период до 1965 года и с 1966 до настоящего времени соответственно;  $\bar{Q}$  – норма годового стока.

Анализ изменения стока позволил выявить ряд рек (Вить – с. Борисовщина, Копаювка – с. Черск и др.), где антропогенные воздействия были наиболее существенными (изменение площади водосбора, перераспределение стока, зарегулированность и др.). Остальные значения  $k_i$  находятся в пределах от -0,281 до 0,333 и были картированы (рис. 9.1) с использованием координат центров водосборов исследуемых рек-створов.

Анализ пространственной структуры изменения годового стока (рис. 9.1) позволяет утверждать, что в северной и центральной частях Беларуси, менее подверженных мелиоративным воздействиям, изменений годового стока практически не произошло. В северо-западной части расходы воды незначительно уменьшились за период 1966 – 2000 гг. В то время как для южной и юго-западной частей Беларуси произошло увеличение годового стока за период 1966 – 2000 по сравнению с периодом до 1965 года.

Таким образом, крупномасштабные мелиорации внесли значительный вклад в увеличение годового стока малых рек Белорусского Полесья путем сработки вековых запасов воды верхних горизонтов земной поверхности.

Аналогичные исследования проведены для максимального и минимального стока рек Беларуси. Наблюдается уменьшение максимальных расходов после 1965 года по всей территории Беларуси, только для 2 рек-створов изменения имеют положительные значения (р. Бобрик – ст. Парохонск, р. Вить – с. Борисовщина), для остальных рек-створов колеблются в пределах от -1,028 (р. Вилия – с. Вилейка) до -0,016 (р. Припять – г. Туров).



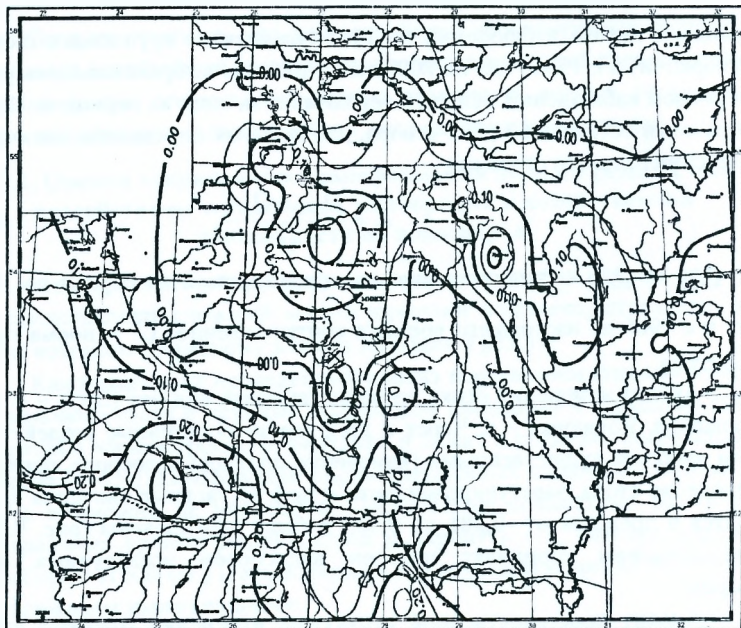


Рис. 9.1. Пространственная структура изменения годового стока рек Беларуси

Также четко прослеживается картина увеличения минимальных расходов после 1965 года, только для 4 рек-створов изменение стока имеет отрицательные значения, для остальных колеблется в пределах от 1,202 (р. Вить – с. Борисовщина) до 0,007 (р. Виляя – ст. Залесье).

Полученные результаты практически совпадают с аналогичными результатами других авторов, в частности А. Г. Булавко [Булавко, 1971], о влиянии мелиорации и хозяйственной деятельности на водосборах на сток рек.

Анализируя полученные результаты, можно отметить, что изменения различных видов стока после 1965 года характерны не только для рек Белорусского Полесья, но и для остальных рек Беларуси, что можно связать не столько с массовыми мелиорациями, сколько с вековыми колебаниями гидрометеорологических элементов и речного стока. Влияние же антропогенной составляющей (например, мелиорации) на максимальный и минимальный сток должно рассматриваться в каждом конкретном случае индивидуально.



Детальный анализ гидрографов частного модулей стока рек Лесная – с. Замосты ( $F=1840 \text{ км}^2$ ,  $\bar{q}=4,4 \text{ л/(с км}^2\text{)}$ ) и Ясельда – г. Береза ( $F=1040 \text{ км}^2$ ,  $\bar{q}=4,6 \text{ л/(с км}^2\text{)}$ ) за 1960 ( $q=4,35; 4,68$ ), 1970 ( $q=8,61; 9,63$ ) и 2000 ( $q=3,68; 5,18$ ) годы позволил выявить конкретный вклад антропогенных факторов в годовой ход речного стока (рис. 9.2).

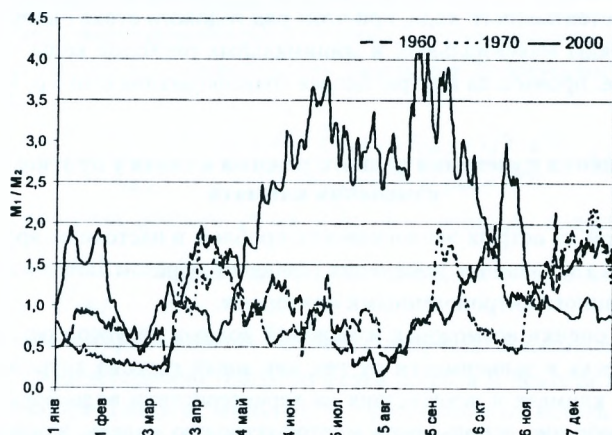


Рис. 9.2. Гидрографы частного модулей стока р. Лесная – с. Замосты ( $M_1$ ) и р. Ясельда – г. Береза ( $M_2$ )

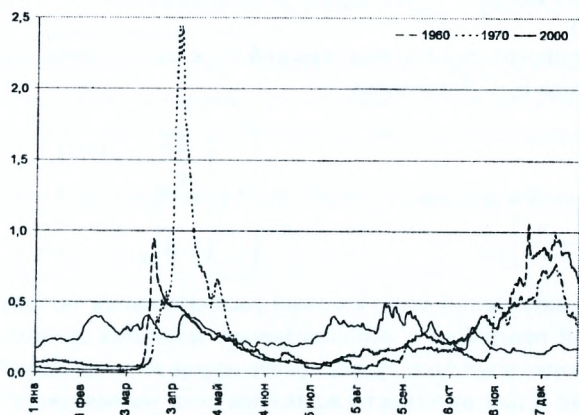


Рис. 9.3. Гидрографы реки Ясельда – г. Береза

Построенное в 1985 году для целей разборазведения водохранилище Селец трансформировало гидрограф р. Ясельда – г. Береза (рис. 9.3) в основном из-за полезных попусков в нижний бьеф.

Анализируя изменения гидрографов р. Ясельда у г. Береза можно констатировать факт антропогенного воздействия. Так, разностная интегральная кривая с 1985 г. для р. Ясельда – г. Береза начинает отдаляться от своего естественного хода, при этом ряд годового стока однородный, в то время ряды максимальных и минимальных расходов воды – неоднородные, т. е. произошла внутригодовая трансформация стока р. Ясельда – г. Береза.

## 9.2. Оценка изменения водного режима в связи с прогнозами изменения климата

Одной из острых экологических проблем в настоящее время является угроза глобального изменения (главным образом потепления) климата, вызванного антропогенными факторами.

Для оценки возможных изменений водных ресурсов рек Белорусского Полесья в зависимости от тех или иных гипотез антропогенного изменения климата и воздействия на характеристики водосборов наиболее целесообразно использовать многофакторную модель, в основе которой лежит стандартное уравнение водного баланса участка суши с независимой оценкой основных элементов баланса (атмосферные осадки, суммарное испарение и климатический сток) в годовом аспекте [Лукша, Волчек, 2002].

Среднегодовое климатическое сток рассчитывался с использованием метода В. С. Мезенцева

$$U_{кл} = X - E_c, \text{ или } U_{кл} = X - E_0 \left( 1 + \frac{\left( \frac{E_0}{W_{нв}} + V^{1-r} \right)^{-n}}{\frac{X}{W_{нв}} + V} \right)^{-1/n}, \quad (9.2)$$

где  $U_{кл}$  – климатический сток, который рассчитывается по специальной компьютерной программе с использованием исходных данных (осадки, дефицит влажности воздуха, сумма температур за вегетационный период) по ближайшей к центру тяжести водосбора реки метеостанции, мм;  $X$  – атмосферные осадки, мм;  $E_c$  – суммарное испарение, мм;  $E_0$  – максимально возможное испарение, мм;  $W_{нв}$  – значение наименьшей влагоемкости

деятельного слоя почвы, мм;  $V = \frac{W}{W_{нв}}$  – относительная влажность почвы

на начало расчетного интервала;  $r$  – параметр, зависящий от водно-физических свойств и гранулометрического состава почвогрунтов, значение которого в нашем численном эксперименте принималось 1,5;  $n$  – параметр, учитывающий гидравлические условия стока ( $n=3,0$ ).

Модель формирования речного стока с использованием взаимосвязи климатического и руслового стока малых рек Белорусского Полесья имеет вид:

$$Y_p = Y_{кл} \cdot \frac{\varphi^{n_2} \cdot \lambda^{n_3} \cdot (L+1)^{n_4} \cdot I_{ср.р.}^{n_5} \cdot H_{ср.}^{n_6} \cdot (F_{заб.л.} + F_{сух.л.} + 1)^{n_9} \cdot \rho^{n_{10}}}{(A + A_{кр})^{n_1} \cdot (F_{оз.} + 1)^{n_7} \cdot (F_{бол.} + 1)^{n_8} \cdot (F_{расп.} + 1)^{n_{11}}} \cdot A \cdot 10^3 / t; \quad (9.3)$$

где  $Y_p$  – величина речного стока м<sup>3</sup>/с;  $A$  – площадь водосбора, км<sup>2</sup>;

$A_{кр} = \frac{0,98}{\rho^2} + \frac{1,5 \cdot L}{\sqrt{I_{ср.р.}}}$  – первая критическая площадь, т. е. та площадь во-

досбора, при которой водоток начинает получать подземное питание, км<sup>2</sup>;  $\varphi$ ,  $\lambda$  – соответственно географическая широта и долгота центра тяжести водосбора (градусы, минуты);  $L$  – расстояние от истока реки, км;  $I_{ср.р.}$  – средний уклон основного тальвега реки, промилле;  $H_{ср.}$  – средняя высота водосбора, м;  $F_{оз.}$ ,  $F_{бол.}$ ,  $F_{заб.л.}$ ,  $F_{сух.л.}$  – площадь, занятая озерами, болотами, заболоченным и сухим лесом, % к площади водосбора ( $A$ );  $\rho$  – коэффициент густоты речной сети, км/км<sup>2</sup>;  $F_{расп.}$  – распаханность водосбора, %;  $t$  – период осреднения, с.

Коэффициенты  $n_1 \dots n_{11}$  (показатели степени), полученные с использованием компьютерной программы статистических расчетов «Statistica 5.5» методом множественной регрессии, приведены в табл. 9.1.

Таблица 9.1

 Значения показателей степени  $n_1 \dots n_{11}$  в уравнении (9.3)

Месяц	$n_1$	$n_2$	$n_3$	$n_4$	$n_5$	$n_6$	$n_7$	$n_8$	$n_9$	$n_{10}$	$n_{11}$
Апрель	-0,148	-2,268	0,860	0,034	-0,132	1,446	-0,222	0,048	0,174	-0,058	-0,239
Май	-0,120	2,172	-1,912	0,094	0,042	-0,471	0,093	0,051	0,311	0,351	-0,080
Июнь	-0,024	2,723	-0,942	0,031	0,215	-1,638	0,084	0,025	0,235	0,122	-0,077
Июль	0,026	1,247	-0,588	-0,074	0,248	-0,930	0,132	0,063	0,320	0,195	0,113
Август	-0,121	3,731	-1,567	0,168	0,030	-1,755	-0,012	-0,015	0,017	0,597	-0,229
Сентябрь	-0,269	2,837	-1,336	0,368	-0,200	-0,766	-0,070	-0,019	-0,179	0,807	-0,544
Октябрь	-0,103	1,434	-1,727	0,148	-0,136	0,183	-0,106	0,084	-0,018	0,543	-0,191
Теплый период	-0,118	0,504	-0,374	0,107	-0,022	-0,056	-0,063	0,042	0,122	0,248	-0,147
Год	-0,103	0,097	-0,711	0,046	-0,024	0,356	-0,045	0,055	0,122	0,064	-0,025

Проверка модели на 14 независимых створах рек Белорусского Полесья (не входящих в разработку модели) показала близкую сходимость рассчитанных и измеренных месячных и годовых расходов (рис. 9.3), что позволяет использовать разработанную модель для практических расчетов.

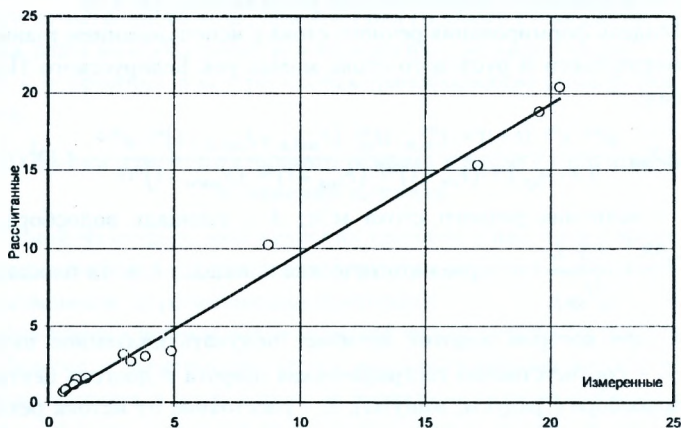


Рис. 9.3. Зависимость измеренных и рассчитанных по разработанной модели значений годового стока рек Белорусского Полесья ( $r=0,994$ )

Для проведения численного эксперимента возможного антропогенного влияния на водные ресурсы рек Белорусского Полесья нами отобраны 12 малых рек Брестской области. Основываясь на анализе существующих в настоящее время оценок возможного антропогенного изменения климата и водосборов рек, численный эксперимент проведен по следующим вариантам:

**вариант 1** – средняя годовая температура воздуха увеличивается на  $2^{\circ}\text{C}$  по сравнению с современным уровнем при неизменном количестве атмосферных осадков;

**вариант 2** – уменьшение годовых атмосферных осадков на 10 % с неизменной температурой воздуха;

**вариант 3** – годовые атмосферные осадки уменьшаются на 10 %, а средняя годовая температура воздуха увеличивается на  $2^{\circ}\text{C}$ ;

**вариант 4** – заболоченность (осушение) и залесенность (вырубка леса) водосбора уменьшаются, а густота речной сети (создание несовершенных мелиоративных систем) и распаханность (интенсивное выращи-

вание сельскохозяйственных культур) увеличиваются на 5, 10, 20 и 30% от существующих в настоящее время при неизменных климатических условиях;

**вариант 5** – заболоченность и залесенность водосбора увеличиваются на 10% каждая, а густота речной сети и распаханность уменьшается на 20% от существующих в настоящее время при неизменных климатических условиях.

Использование вышеперечисленных вариантов позволит комплексно оценить трансформацию речного стока с точки зрения изменения климатических условий (1, 2, 3 варианты) и увеличения (4 вариант) или уменьшения (5 вариант) антропогенного воздействия на водосборы рек.

Исходя из расчетов, на основе указанных выше уравнений с принятыми гипотезами, сделаны следующие выводы:

– по первому варианту (увеличение температуры на 2 °С) речной сток уменьшится в среднем на 10%, а суммарное испарение может увеличиться в среднем до 4,7% (табл. 9.2 и 9.3).

Таблица 9.2

**Величины изменения речного стока по варианту 1, в % к существующему**

Река-створ	IV	V	VI	VII	VIII	IX	X	IV-X	Год
Копаявка - с.Черск	-12,1	-23,9	-24,8	-22,6	-20,0	-17,1	-14,3	-16,9	-10,7
Мухавец - г.Пружаны	-8,10	-18,4	-22	-22,0	-19,3	-17,0	-14,3	-15,3	-10,7
Жабинка - с.М.Жабинка	-11,8	-23,6	-24,5	-22,0	-19,8	-17,6	-15,6	-16,7	-10,7
Рыта - с.М.Радваничи	-12,1	-23,9	-24,8	-22,6	-20,0	-17,1	-14,3	-16,9	-10,7
Лесная - с.Замосты	-11,8	-23,8	-25,2	-23,5	-19,8	-19,2	-14,8	-16,9	-10,4
Пульва - г.Высокое	-12,7	-23,9	-24,5	-21,5	-18,8	-19,2	-14,8	-17,3	-11,3
Гривда - г.Ивацевичи	-7,40	-18,2	-24,0	-21,7	-19,5	-18,1	-13,5	-14,7	-9,60
Яссьлда - г.Береза	-7,80	-18,4	-22,1	-22,3	-20,2	-18,4	-14,9	-15,3	-10,2
Бобрик - ст.Парохонск	-7,50	-17,5	-23,4	-20,4	-17,7	-16,1	-11,4	-14,2	-9,00
Цна - с.Дятловичи	-8,10	-18,7	-24,8	-22,0	-19,0	-17,2	-13,5	-15,1	-9,40
Лань - с.Локтыши	-5,40	-16,1	-24,1	-21,1	-19,1	-17,2	-11,1	-13,1	-8,40
Мышанка - с.Березки	-5,90	-16,4	-22,0	-21,4	-18,9	-15,3	-11,1	-13,5	-8,80
<b>Среднее</b>	<b>-9,20</b>	<b>-20,2</b>	<b>-23,9</b>	<b>-21,9</b>	<b>-19,3</b>	<b>-17,5</b>	<b>-13,6</b>	<b>-15,5</b>	<b>-10,0</b>

Анализ табл. 9.2 и 9.3 дает ясное представление об уменьшении речного стока при увеличении температуры, так как при этом увеличивается суммарное испарение, особенно в летние месяцы (июнь, июль, август). Можно отметить также асинхронность в изменениях величин среднемесячного речного стока и суммарного испарения, например, в апреле при среднем уменьшении стока на 9,2% испарение увеличилось на 8,3%, что можно объяснить пиком весеннего половодья (интенсивное таяние



снегов при прохождении максимальных расходов) и повышенной влажностью воздуха в этот период.

Таблица 9.3

**Величины изменения суммарного испарения по варианту 1, в % к существующему**

Река-створ	IV	V	VI	VII	VIII	IX	X	IV-X	Год
Копаявка - с.Черск	7,5	2,7	1,7	2,1	3,0	4,0	5,3	3,6	4,2
Мухавец - г.Пружаны	8,4	5,2	3,1	2,4	3,2	3,9	5,7	4,2	4,8
Жабинка - с.М.Жабинка	7,5	2,7	1,7	2,0	2,9	4,0	5,3	3,5	4,3
Рыга - с.М.Радваничи	7,5	2,7	1,7	2,1	3,0	4,0	5,3	3,6	4,2
Лесная - с.Замосты	7,6	2,7	1,6	1,9	3,1	3,8	4,9	3,5	4,2
Пульва - г.Высокое	7,1	2,3	1,6	1,9	3,1	3,8	4,9	3,4	4,1
Гривда - г.Ивацевичи	8,9	5,7	2,2	2,7	3,5	4,2	6,0	4,4	5,0
Ясельда - г.Береза	8,6	5,3	3,1	2,4	3,3	3,9	5,2	4,2	4,8
Бобрск - ст.Парохонск	9,0	6,0	2,7	3,4	4,2	5,1	6,6	4,9	5,6
Цна - с.Дятловичи	8,7	5,4	2,2	2,7	3,7	4,3	6,0	4,4	5,1
Лань - с.Локтыши	9,3	6,5	2,6	2,9	3,7	4,2	6,1	4,7	5,2
Мышанка - с.Березки	9,3	6,4	3,4	3,0	3,8	5,0	6,8	4,9	5,4
<b>Среднее</b>	<b>8,3</b>	<b>4,5</b>	<b>2,3</b>	<b>2,5</b>	<b>3,4</b>	<b>4,2</b>	<b>5,7</b>	<b>4,1</b>	<b>4,7</b>

– по второму варианту (уменьшение осадков на 10%) речной сток может уменьшиться на 24,5%, а суммарное испарение – в среднем, на 5,4% (табл. 9.4 и 9.5). При этом максимальное уменьшение стока наблюдается в июле (29,7%), а минимальное – в апреле (23,8%), а суммарное испарение – соответственно в июле (7,0 %) и апреле (4,2%).

Таблица 9.4

**Величины изменения речного стока по варианту 2, в % к существующему**

Река-створ	IV	V	VI	VII	VIII	IX	X	IV-X	Год
Копаявка - с.Черск	-24,5	-27,3	-30,1	-30,1	-30,2	-29,5	-27,8	-27,9	-24,9
Мухавец - г.Пружаны	-24,5	-28,1	-29,2	-29,2	-28,7	-28,4	-28,2	-27,7	-25
Жабинка - с.М.Жабинка	-23,8	-26,9	-29,8	-30,8	-31,1	-30,2	-27,4	-27,9	-24,8
Рыга - с.М.Радваничи	-24,2	-27,3	-30,1	-30,1	-30,2	-29,5	-27,8	-27,8	-24,9
Лесная - с.Замосты	-25,7	-27,3	-29,6	-30,3	-30,4	-29,8	-28,6	-28,3	-25,7
Пульва - г.Высокое	-24,7	-27,6	-28,6	-30,7	-31,4	-29,8	-27,6	-28,2	-25,4
Гривда - г.Ивацевичи	-23,9	-26,1	-27,6	-27,9	-29,0	-29,8	-28,1	-26,8	-24,5
Ясельда - г.Береза	-24,9	-27,9	-29,1	-30,0	-29,6	-27,9	-27,9	-27,8	-25,2
Бобрск - ст.Парохонск	-22,5	-26,4	-29,1	-29,2	-28,4	-27,7	-27,4	-26,6	-23,7
Цна - с.Дятловичи	-23,4	-26,7	-27,7	-28,8	-28,2	-26,9	-25,8	-26,4	-23,6
Лань - с.Локтыши	-21,3	-25,6	-29,0	-30,2	-29,3	-29,7	-29,2	-26,2	-23,2
Мышанка - с.Березки	-21,7	-25,6	-27,7	-28,5	-28,9	-27,6	-26,2	-25,8	-23,1
<b>Среднее</b>	<b>-23,8</b>	<b>-26,9</b>	<b>-29,0</b>	<b>-29,7</b>	<b>-29,6</b>	<b>-28,9</b>	<b>-27,7</b>	<b>-27,3</b>	<b>-24,5</b>



Таблица 9.5

**Величины изменения суммарного испарения по варианту 2, в % к существующему**

Река-створ	IV	V	VI	VII	VIII	IX	X	IV-X	Год
Копаяовка - с.Черск	-4,57	-6,72	-7,12	-7,34	-6,88	-6,51	-5,69	-6,43	-5,64
Мухавец - г.Пружаны	-4,41	-6,15	-6,69	-6,98	-6,66	-6,18	-5,84	-6,22	-5,55
Жабинка - с.М.Жабинка	-4,56	-6,7	-7,11	-7,22	-6,93	-6,54	-5,69	-6,41	-5,65
Рыга - с.М.Радваничи	-4,57	-6,72	-7,12	-7,34	-6,88	-6,51	-5,69	-6,43	-5,64
Лесная - с.Замосты	-4,83	-6,56	-7,01	-7,4	-7,11	-6,57	-6,06	-6,51	-5,72
Пульва - г.Высокое	-4,85	-6,39	-7,08	-7,42	-6,94	-6,57	-6,06	-6,49	-5,77
Гривда - г.Ивацевичи	-4,19	-5,96	-6,68	-6,59	-6,79	-6,77	-5,95	-6,15	-5,36
Ясельда - г.Береза	-4,36	-6,26	-6,86	-7,02	-6,67	-6,42	-5,9	-6,29	-5,55
Бобрик - ст.Парохонск	-3,73	-5,59	-6,76	-6,43	-5,99	-5,74	-4,91	-5,68	-4,80
Цна - с.Дятловичи	-4,19	-6,01	-6,86	-6,75	-6,26	-5,99	-5,14	-5,97	-5,06
Лань - с.Локтыши	-2,93	-5,29	-6,73	-6,86	-6,61	-6,57	-5,94	-5,87	-5,08
Мышанка - с.Березки	-3,31	-5,47	-6,28	-6,52	-6,34	-5,88	-5,45	-5,68	-4,85
<b>Среднее</b>	<b>-4,20</b>	<b>-6,20</b>	<b>-6,90</b>	<b>-7,00</b>	<b>-6,70</b>	<b>-6,40</b>	<b>-5,70</b>	<b>-6,20</b>	<b>-5,40</b>

Здесь отмечается синхронное уменьшение речного стока и суммарного испарения при уменьшении величин атмосферных осадков, так как количество поступающей влаги уменьшилось, соответственно уменьшилась и возможность ее испарения, что связано с меньшей влажностью почвы и увеличением ее водопоглощающей способности.

– по третьему варианту (уменьшение осадков на 10% и увеличение температуры на 2 °С) сток уменьшился в среднем на 29,3% (июль – 45,2%, апрель – 26,6%), а суммарное испарение увеличивается в апреле на 6,2% и уменьшается в июле на 5,1% при среднем уменьшении около 0,7% (табл. 9.6 и 9.7).

Таблица 9.6

**Величины изменения речного стока по варианту 3, в % к существующему**

Река-створ	IV	V	VI	VII	VIII	IX	X	IV-X	Год
Копаяовка - с.Черск	-30,2	-44,0	-46,7	-46,4	-45,6	-43,4	-38,1	-37,6	-30,6
Мухавец - г.Пружаны	-25,5	-37,5	-43,4	-44,7	-43,1	-40,9	-38,6	-35,4	-30,1
Жабинка - с.М.Жабинка	-30,0	-43,5	-46,2	-46,3	-45,3	-43,2	-39,1	-37,3	-30,4
Рыга - с.М.Радваничи	-30,2	-44,0	-46,7	-46,4	-45,6	-43,4	-38,1	-37,6	-30,6
Лесная - с.Замосты	-31,1	-43,8	-46,6	-46,9	-45,3	-43,8	-39,3	-37,9	-31,1
Пульва - г.Высокое	-31,3	-43,0	-45,9	-46,8	-44,7	-43,8	-39,3	-38,1	-31,5
Гривда - г.Ивацевичи	-24,3	-36,4	-44,2	-43,4	-43,2	-42,6	-37,1	-33,9	-28,5
Ясельда - г.Береза	-25,4	-37,8	-43,6	-45,5	-44,0	-42,5	-38,8	-35,5	-29,7
Бобрик - ст.Парохонск	-23,4	-35,8	-44,0	-43,8	-41,5	-39,8	-35,1	-33,4	-27,6
Цна - с.Дятловичи	-24,6	-37,0	-44,8	-44,1	-42,2	-41,9	-36,0	-34,1	-27,8
Лань - с.Локтыши	-21,0	-34,2	-44,4	-44,2	-43,5	-42,4	-36,7	-32,0	-26,8
Мышанка - с.Березки	-21,7	-34,3	-42,3	-43,5	-42,7	-39,5	-34,3	-32,4	-27,1
<b>Среднее</b>	<b>-26,6</b>	<b>-39,3</b>	<b>-44,9</b>	<b>-45,2</b>	<b>-43,9</b>	<b>-42,2</b>	<b>-37,5</b>	<b>-35,4</b>	<b>-29,3</b>

Таблица 9.7

**Величины изменения суммарного испарения по варианту 3, в  
% к существующему**

Река-створ	IV	V	VI	VII	VIII	IX	X	IV-X	Год
Копаяовка - с.Черск	4,6	-3,3	-5,7	-6,0	-5,1	-3,7	-1,3	-3,0	-1,7
Мухавец - г.Пружаны	6,5	0,5	-3,2	-4,8	-4,1	-3,0	-0,9	-1,7	-0,6
Жабинка - с.М.Жабинка	4,6	-3,3	-5,7	-6,2	-5,3	-3,7	-0,9	-3,0	-1,7
Рыга - с.М.Радваничи	4,6	-3,3	-5,7	-6,0	-5,1	-3,7	-1,3	-3,0	-1,7
Лесная - с.Замосты	4,6	-3,3	-5,7	-6,2	-5,3	-4,0	-1,3	-3,1	-1,8
Пульва - г.Высокое	3,8	-3,5	-5,7	-6,2	-5,1	-4,0	-1,8	-3,2	-2,0
Гривда - г.Ивацевичи	7,3	1,4	-4,1	-4,2	-3,8	-3,1	0,4	-1,3	-0,1
Ясельда - г.Береза	6,7	0,8	-3,4	-5,0	-4,1	-3,0	-1,3	-1,7	-0,6
Бобрик - ст.Парохонск	7,6	2,2	-3,5	-3,6	-2,6	-1,2	1,6	-0,4	0,8
Цна - с.Дятловичи	7,1	1,1	-4,2	-4,3	-3,2	-2,3	0,8	-1,2	0,1
Лань - с.Локтыши	8,4	3,1	-3,7	-4,3	-3,5	-2,8	0,4	-0,8	0,2
Мышанка - с.Березки	8,2	2,9	-2,4	-3,9	-3,2	-1,4	2,1	-0,3	0,6
<b>Среднее</b>	<b>6,2</b>	<b>-0,4</b>	<b>-4,4</b>	<b>-5,1</b>	<b>-4,2</b>	<b>-3,0</b>	<b>-0,3</b>	<b>-1,9</b>	<b>-0,7</b>

Речной сток оказался очень чувствительным к одновременному уменьшению количества атмосферных осадков и увеличению температуры воздуха: значения стока существенно уменьшились для июня на 44,9%, для июля на 45,2%, для августа на 43,9%, что можно объяснить небольшими расходами во время летней межени и большим (по отношению к остальным месяцам года) абсолютным уменьшением количества осадков (в летние месяцы выпадает наибольшее количество осадков). Также в летние месяцы речной сток формируется в основном за счет атмосферных осадков, а не талых вод, как в весенние, поэтому и наблюдается максимальное уменьшение стока летом.

– по четвертому варианту (заболоченность и залесенность водосбора уменьшаются, а густота речной сети и распаханность увеличиваются на 5, 10, 20 и 30% соответственно) средние значения изменений речного стока для исследуемых рек-водосборов приведены в табл. 9.8.

Анализ табл. 9.8 позволяет выявить тенденцию постепенного перехода от уменьшения стока в апреле-июле к его увеличению в августе-октябре, при этом переход через «нулевые» значения изменений приходится на вторую половину июля. Можно отметить, что одновременное осушение болот, вырубка леса, создание новых мелиоративных систем и увеличение процента пахотных земель уменьшает речной сток весеннего половодья и увеличивает его в осенние месяцы. Хотя влияние этих антропогенных воздействий на речной сток неоднозначно, возможно компонентное исследование влияния каждого из них на сток рек и про-

гнозирование количественного изменения среднемесячных расходов рек Белорусского Полесья. Явно прослеживается тенденция увеличения средних величин изменений речного стока в зависимости от степени антропогенных воздействий, но 20-30% изменение параметров водосбора при существующем уровне развития мелиоративного строительства и эксплуатации уже созданных мелиоративных систем нереально, поэтому хозяйственная деятельность в бассейнах рек не повлияет существенным образом в ближайшем будущем на речной сток.

Таблица 9.8

**Средние величины изменения величины речного стока по варианту 4, в % к существующему**

Степень антропогенного воздействия	IV	V	VI	VII	VIII	IX	X	IV-X	Год
5%	-2,48	-0,41	-1,08	-0,40	1,90	2,24	1,29	-0,25	-0,69
10%	-4,88	-1,03	-2,16	-0,80	3,81	4,93	2,59	-0,50	-1,38
20%	-9,48	-2,47	-4,32	-2,41	7,14	9,87	5,18	-1,49	-2,98
30%	-13,89	-4,74	-7,19	-4,82	10,48	15,25	7,44	-2,72	-4,82

– по пятому варианту (заболоченность и залесенность водосбора увеличиваются на 10%, а густота речной сети и распаханность уменьшаются на 20%) средние значения изменений речного стока для исследуемых рек-водосборов приведены в табл. 9.9.

Таблица 9.9

**Величины изменения речного стока по варианту 5, в % к существующему**

Река-створ	IV	V	VI	VII	VIII	IX	X	IV-X	Год
Копаяювка - с.Черск	7,89	-4,14	0,00	-4,86	-8,37	-7,25	-7,00	-1,58	0,24
Мухавец - г.Пружаны	7,39	-4,74	0,00	-4,60	-9,03	-7,69	-7,41	-1,90	-0,32
Жабинка - с.М.Жабинка	7,7	-4,45	0,35	-4,47	-8,07	-7,35	-6,98	-1,46	0,00
Рыга - с.М.Радваничи	7,82	-3,97	0,36	-4,48	-8,13	-7,08	-7,14	-1,63	0,00
Лесная - с.Замосты	8,02	-4,04	0,00	-4,42	-8,21	-7,14	-6,93	-1,38	0,25
Пульва - г.Высокое	7,87	-4,32	0,00	-4,50	-8,13	-7,41	-6,91	-1,42	0,00
Гривда - г.Ивацевичи	7,99	-4,23	0,00	-4,80	-8,22	-7,31	-7,12	-1,30	0,25
Ясельда - г.Береза	7,81	-4,04	0,32	-4,52	-8,68	-7,37	-7,35	-1,56	0,22
Бобриск - ст.Парохонск	7,84	-4,04	0,25	-4,27	-8,13	-7,59	-7,09	-1,62	0,00
Цна - с.Дятловичи	7,78	-4,19	0,31	-4,81	-8,33	-7,47	-7,26	-1,60	0,20
Лань - с.Локтыши	8,05	-4,05	0,00	-4,45	-8,24	-7,10	-6,80	-1,31	0,00
Мышанка - с.Березки	7,86	-4,16	0,28	-4,51	-8,14	-7,03	-6,93	-1,39	0,18
Среднее	7,82	-4,12	0,36	-4,42	-8,10	-7,17	-7,12	-1,49	0,00

Данный вариант рассматривался как наиболее вероятный в ближайшем будущем в связи с уменьшением хозяйственной деятельности на водосборе. Данные табл. 9.9 подтверждают предположение о незначи-

тельном влиянии антропогенной нагрузки по сравнению с климатическими факторами на средний годовой сток (изменения около 0) и о постепенном восстановлении естественного режима рек, т. е. возможно увеличение максимального стока (до 8%) и уменьшение минимального (также до 8%). Это означает, что если влияние антропогенных факторов на сток малых рек будет постепенно уменьшаться при отсутствии климатических изменений, наиболее вероятен вариант возврата к естественному внутригодовому распределению стока (как в период до 1966 года).

Таким образом, наиболее неблагоприятным прогнозом развития антропогенного изменения речного стока для рек Белорусского Полесья является третий вариант – может произойти уменьшение стока до 45%. При наложении на этот вариант 10%-ного антропогенного воздействия на водосбор реки уменьшение среднего годового стока может достигнуть 50%. Наиболее вероятным прогнозом на ближайшие 5 лет можно считать пятый вариант, т. е. существенного изменения годового стока не произойдет, будет наблюдаться небольшое увеличение максимального и снижение минимального стока.

Более детально проведен анализ влияния климатических и антропогенных факторов на средний годовой сток р. Ясельда – г. Береза, результаты которого представлены в табл. 9.10. Эти результаты подтвердили предположение с том, что самым неблагоприятным прогнозом изменения речного стока является третий вариант, имеющий максимальные изменения как обеспеченности (с 50 до 85%), так и коэффициента вариации (с 0,47 до 0,54).

Таблица 9.10

**Изменение обеспеченности и коэффициента вариации годового стока р. Ясельда – г. Береза в зависимости от варианта воздействия**

№ варианта	Исходный	1	2	3	4
Обеспеченность, %	50	62,4	72,9	85,0	52,0; 54,1; 56,2; 58,0
Изменение обеспеченности, %	0,0	24,8	45,8	70,0	4,0; 8,2; 12,4; 16,0
Коэффициент вариации $C_v$	0,47	0,49	0,52	0,54	0,47; 0,47; 0,48; 0,48
Изменение коэффициента вариации $C_v$ , %	0,0	4,1	9,6	13,0	0,0; 0,0; 2,0; 2,0

Установлено, что увеличение средней годовой температуры воздуха с одновременным уменьшением количества атмосферных осадков неоднозначно влияет на изменение суммарного испарения, т. е. при некото-



ром сочетании изменения осадков и температуры наблюдается перелом в изменении суммарного испарения. Анализ матрицы средней годовой температуры и суммарного годового испарения при постепенном увеличении температуры и уменьшении осадков показал, что для р. Ясельда – г. Береза такой переломной точкой является увеличение средней годовой температуры на 0,6 °С при одновременном 2 %-ном уменьшении количества атмосферных осадков. При дальнейшем увеличении температуры и уменьшении количества осадков происходит постепенное уменьшение суммарного испарения.

В последнее десятилетие (1990 – 2000 гг.) произошло изменение количества атмосферных осадков на юге страны, поэтому нами были проведены расчеты по определению изменения норм речного стока с использованием реальных измеренных среднемесячных и годовых расходов. Результаты расчетов приведены в табл. 9.11.

Таблица 9.11

**Отношения величин среднемесячных и годовых расходов за 1990-2000 гг. к нормам за весь период инструментальных наблюдений**

Река-створ	I	II	III	IV	V	VI	VII	VIII	IX	X	XI	XII	Год
Копаювка – с. Черск	1,00	1,04	1,22	1,09	1,26	1,97	1,55	1,21	1,20	1,35	1,52	1,15	1,22
Рыга – с.М.Радваничи	1,12	1,38	1,51	1,58	1,62	1,81	1,61	1,40	1,37	1,56	1,62	1,37	1,46
Лесная – с. Замосты	0,96	1,20	1,20	1,32	1,00	1,10	1,01	1,29	1,10	1,28	1,33	1,17	1,17
Пульва – г. Высокое	1,17	0,98	0,96	1,31	1,12	1,19	1,36	2,30	1,51	2,18	1,48	1,29	1,08
Ясельда – г. Береза	0,78	0,77	1,42	2,15	1,19	0,81	0,58	0,49	0,45	0,66	1,12	1,05	0,94
Цна – с. Дятловичи	0,62	0,64	0,73	1,17	0,96	0,83	0,62	0,47	0,57	0,59	0,77	0,77	0,80

Анализ табл. 9.11 показывает, что почти для всех рассмотренных створов и месяцев произошло изменение величин стока при среднем уменьшении суммарного годового количества осадков на 5%. Эти изменения неоднозначны – для рек бассейна Западного Буга (Копаювка, Рыга, Лесная, Пульва) наблюдается увеличение значений расходов, для Ясельды и Цны, которые относятся к бассейну Припяти – уменьшение. Здесь также значительно сказалась неравномерность распределения атмосферных осадков внутри года за последние 10 лет по сравнению со средним распределением за весь период. Количество осадков уменьшилось в апреле-мае и особенно сильно в августе. Эти изменения количества выпавших осадков явно прослеживаются на реках бассейна Припяти – Ясельде и Цне, когда коэффициенты уменьшения стока в августе составили 0,49 и 0,47 соответственно, т. е. норма за 1990-2000 годы в этот месяц уменьшилась практически на 50%.

В связи с этим нами предпринята попытка перераспределить осадки с периода май-июнь на июль-август – уменьшить на 20% в мае-июне и увеличить на 20% в июле-августе. Для детального анализа выбран створ р. Ясельда – г. Береза, как наиболее близко расположенный к метеостанции г. Пинск, которая, в свою очередь, имеет достаточно длинные ряды наблюдения за осадками и дефицитами влажности (55 лет). Полученные изменения значений речного стока приведены в табл. 9.12.

Таблица 9.12

**Величины измеренных и рассчитанных значений модулей речного стока р. Ясельда – г. Береза при различных вариантах использования атмосферных осадков в полученной модели (л/с с км<sup>2</sup>)**

Модули речного стока	Апрель	Май	Июнь	Июль	Август	Сентябрь	Октябрь	Теплый период	Год
Измеренные	12,73	4,98	2,64	2,41	2,46	3,75	4,25	4,75	4,79
Рассчитанные	12,42	5,45	3,11	2,21	2,19	2,85	3,40	4,48	4,49
Ошибка, %	-2,40	9,40	17,6	-8,40	-11,0	-24,0	-20,0	-5,60	-6,20
Рассчитанные при использовании осадков за 1990 – 2000 гг.	12,2	5,23	2,85	1,98	1,86	1,88	2,9	4,13	4,08
Рассчитанные при перераспределении значений осадков на 20% с V-VI на VII-VIII	12,32	4,68	2,15	1,90	2,25	2,04	3,05	4,06	4,15
Процент отклонения по двум последним вариантам	1,00	-10,5	-24,6	-4,00	21,0	8,50	5,20	-1,70	1,70

Как видно из табл. 9.12, значения модулей речного стока, рассчитанные при перераспределении атмосферных осадков на 20% с мая-июня на июль-август отличаются от рассчитанных при использовании средних значений атмосферных осадков за 1990 – 2000 гг. в среднем на 1,7% за год. Внутри же года равномерность распределения значительно нарушена по сравнению с исходной.

Можно констатировать тот факт, что в мае-июне наблюдается уменьшение значений речного стока (на 10,5 и 24,6% соответственно), а в августе-сентябре увеличение (21,0 и 8,5%). В июле же наблюдается небольшое уменьшение (4%), хотя значение суммарных осадков увеличилось с 83,6 до 97,0 мм (на 16%). Этот факт можно объяснить некоторой



инертностью речного стока, т. е. выпавшие на водосбор атмосферные осадки не сразу, а с некоторым запаздыванием попадают в русло реки, что и подтверждается полученными расчетами, т. е. предлагаемая модель (9.2) применима для расчета значений речного стока с учетом атмосферных осадков, выпавших в предыдущие месяцы.

Полученные результаты моделирования изменения стока рек Белорусского Полесья в зависимости от варианта воздействия хорошо коррелируют с выводами, которые были сделаны ранее по результатам исследования изменения стока для рек бассейна Днепра. Их прогноз изменения стока р. Днепра составил от 25 до 40% и более в зависимости от вариантов воздействия, которые основывались на моделях общей циркуляции атмосферы, глобального потепления климата и росте содержания  $\text{CO}_2$  в атмосфере.

Прогнозируемое потепление климата вызовет существенные изменения водного режима рек. Дальнейшее развитие работы связано с совершенствованием воднобалансовых моделей и детализацией зависимостей «климат-сток» для получения более надежных оценок и выводов по проведению мероприятий своевременной адаптации водного хозяйства к изменениям условий формирования местных водных ресурсов.

### **9.3. Региональная оценка антропогенных влияний на сток рек Белорусского Полесья**

В настоящее время разработано большое количество методов и способов оценки антропогенных изменений в зависимости от конкретных физико-географических условий исследуемой территории, наличия и качества исходной информации, вида хозяйственной деятельности, а также временного расчетного интервала, применяются известные приемы восстановления естественного стока, основанные на уравнении водного баланса или же по аналогии. При использовании методов аналогии приходится сталкиваться с рядом трудностей. Так, в условиях интенсивного освоения земель почти не осталось водосборов, не затронутых мелиорациями, поэтому не во всех случаях аналоги полностью отвечают предъявляемым к ним требованиям, кроме того, имеют место факты прекращения гидрометеослужбой гидрологических наблюдений, что существенно осложняет их использование для оценки антропогенного воздействия. Состав исходных данных, особенно по малым рекам, не позволяет применять для анализа изменений стока метод водного баланса.

Оценка антропогенной составляющей на основании анализа полей стока лишена этих недостатков. Суть методики заключается в оценке

различий ПКФ, построенных по данным всего поля за периоды до и после активного воздействия хозяйственной деятельности.

Как известно, любая исходная гидрологическая величина определяется с той или иной степенью точности, что можно представить в следующем виде

$$M = \dot{M} \pm \Delta M_u \pm \Delta M_M \pm \Delta M_A, \quad (9.4)$$

где  $\dot{M}$  – истинная величина гидрологической характеристики, которая формируется в естественных условиях;  $\Delta M_u$  – ошибка за счет погрешностей измерительных приборов;  $\Delta M_M$  – ошибка методики определения;  $\Delta M_A$  – влияние антропогенных факторов.

Средняя квадратическая суммарная погрешность исходных величин объективно выявляется с помощью отраженной эмпирической корреляционной функцией пространственной  $R(\rho)$  или временной  $R(\tau)$  по экстраполированному значению  $R(0)$  отвечающему нулевому расстоянию  $\rho = 0$  или нулевому сдвигу во времени  $\tau = 0$ .

При измерении гидрологической величины с весьма малым шагом времени или расстояния  $\Delta t$ ,  $\Delta r$ , которые существенно меньше периода мелких колебаний типа «белого шума»  $\Delta t^*$ ,  $\Delta r^*$ ,  $\Delta t \ll \Delta t^*$ ,  $\Delta r \ll \Delta r^*$ , то в этом случае погрешности из-за дискретности измерения будут близки к нулю.

Для оценки антропогенной составляющей стока строятся ПКФ за периоды до и после активных антропогенных воздействий. Исходя из независимости различных видов погрешностей, т. е. дисперсия суммарной погрешности равна сумме дисперсий этих погрешностей. Среднюю квадратическую погрешность, вызванную антропогенными факторами определяют как:

$$\sigma_{\Delta} M_A = (\sigma_M^2)^{0,5} - (\sigma_{\Delta}^2 M_u + \sigma_{\Delta}^2 M_M)^{0,5} = \sigma_{\Delta} \dot{M} (R(0) - R(0)^*)^{0,5}, \quad (9.5)$$

где  $\sigma_M$  – среднее квадратическое отклонение исходных значений гидрологической величины;  $R(0)$ ,  $R(0)^*$  – экстраполированное значение эмпирической ПКФ, построенным по данным о стоке соответственно до и после антропогенных воздействий.

Значимость полученной величины  $\sigma_{\Delta} M_A$  оценивается на основе известных статистических критериев. Если установлено, что  $\sigma_{\Delta} M_A$  вызвана не случайными ошибками, зависящими от длины ряда, расчленения

его на периоды и т. д., то сама величина  $\Delta M_A$  характеризует антропогенную составляющую стока исследуемой территории.

Выявление конкретных водосборов, подверженных влиянию хозяйственной деятельности, проводится посредством анализа изменений парных коэффициентов корреляции и выявляются водосборы со статистически различными изменениями коэффициентов.

Для реализации изложенного подхода разработан комплекс программ, позволяющих из стандартных наблюдений на малых водосборах выделить антропогенную составляющую и тем самым восстановить естественный сток. Для оценки возможностей методики проведен численный эксперимент на материалах малых рек Белорусского Полесья. Данный район широко охвачен мелиорациями и достаточно полно представлен гидрологической информацией, кроме того, выполненные ранее другими авторами исследования, позволяют сопоставить полученные результаты. Рассматриваемая территория представлена 26 гидрометрическими створами на реках с площадями водосборов от 67 до 2560 км<sup>2</sup>. Расчеты выполнены для трех характерных расходов: наибольшего весеннего, наименьшего летнего и годового. Ряд наблюдений условно делится на два периода: с момента наблюдений до 1966 года, как начала крупномасштабного мелиоративного строительства, и с 1966 года – по настоящее время.

В исследованиях использованы следующие расходы: наибольшие весенние (В), наименьшие летние (Л) и средние годовые (Г) (рис. 9.3). Случаи обнаружения статистически различимых изменений коэффициентов корреляции свидетельствуют о наличии антропогенных воздействий. Так, сравнение ПКФ для периода, не подверженного мелиоративным воздействиям ( $R(0)_в = 0,774$ ;  $R(0)_л = 0,687$ ;  $R(0)_г = 0,853$ ), с ПКФ для периода крупномасштабного мелиоративного строительства ( $R(0)_в^* = 0,734$ ;  $R(0)_л^* = 0,351$ ;  $R(0)_г^* = 0,742$ ) указывает на статистически значимые различия между ними ( $R(0)_в = 0,040$ ;  $R(0)_л = 0,326$ ;  $R(0)_г = 0,111$ ).

Различия наиболее существенны для минимального стока (47,5%), что может быть связано с изменением общей водности реки, величины водосборной площади, размера площади осушительных болот и заболоченных земель, густоты осушительной сети и степени канализации. В Полесье преобладают мелкозалежные торфяники на хорошо проницаемых песках, глубокая дренажная осушительная сеть и канализованные русла рек врезаны в подстилающий грунт, поэтому здесь отмечается существенное увеличение минимального стока (приблизительно на 50%) за

период 1966 – 2000 гг. по сравнению с периодом с начала наблюдений до 1966 года.

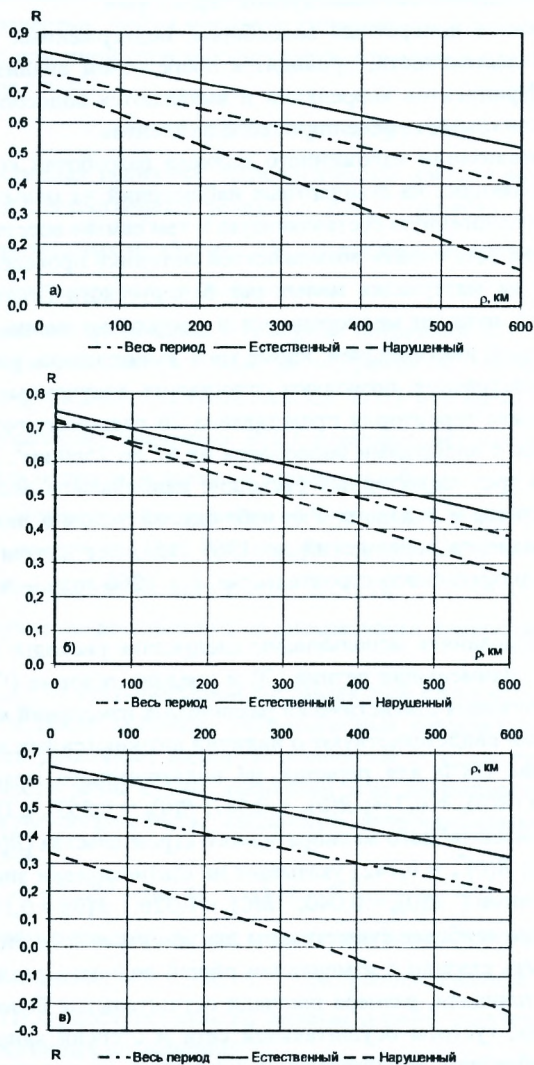


Рис. 9.3. Пространственно-корреляционные функции годового (а), максимального (б) и минимального (в) стока рек Белорусского Полесья за весь период наблюдений, естественный (до 1966 года) и нарушенный (1966–2000 гг.) периоды

На начальном этапе мелиоративных воздействий изменение максимального стока происходило неоднозначно. На тех водосборах, где произошло нарушение однородности рядов максимальных расходов воды и слоев стока, наблюдалось как увеличение значений, так и снижение. Это объясняется тем, что в результате проведения мелиоративных работ на водосборе создается сложное сочетание различных условий, оказывающих разностороннее влияние на формирование максимального стока рек в период весеннего половодья. Значительное увеличение аккумулирующей емкости осушенных площадей на водосборе вызывает потерю талых вод и уменьшение максимальных расходов, а искусственное увеличение густоты гидрографической сети одновременно с регулированием рек способствует формированию повышенных максимальных расходов. В связи с тем, что оба фактора действуют одновременно, характер и величина изменения максимального стока зависит от того фактора, влияние которого после преобразования естественных условий водосбора оказывается более сильным и играет в формировании основную роль.

После установления равновесного состояния отмечается стабильное уменьшение максимального стока (приблизительно на 25%), при этом оно происходит более «организованно» как во времени, так и в пространстве в отличие от процессов увеличения минимального стока. Поэтому отмечается незначительное уменьшение значения коэффициента корреляции  $R(0)_в - 5,2\%$ .

При этом максимальное изменение годового стока происходит на водосборах, расположенных в пределах плоских низменностей, а также с широкими заболоченными поймами. Чем больше была заболоченность водосбора, тем больше величина изменения годового стока после мелиорации. Она увеличивается и с увеличением доли осушенных болот и заболоченных земель. Согласно результатам наших исследований для территории Белорусского Полесья произошло увеличение годового стока за период 1966–2000 гг. по сравнению с периодом до 1965 г. приблизительно на 20%. При этом изменение коэффициента корреляции  $R(0)_г$  составило (13%).

Также исследовались изменения статистических параметров ПКФ в зависимости от продолжительности расчетного периода. Были отображены 26 рядов (рек-створов) речного стока Белорусского Полесья с наиболее продолжительными рядами наблюдений (не менее 25 лет). Исследовались ряды наблюдений среднего годового, максимального и минимального речного стока с использованием методики построения ПКФ с нарастающим пятилетним периодом наблюдений по следующим временным



интервалам: до 1966 года (до периода начала интенсивных мелиораций на водосборах рек), до 1970, до 1975, до 1980, до 1985, до 1990, до 1995 и до 2000 года. В результате получены ПКФ и их параметры для каждого из периодов, последние представлены на рис. 9.4.

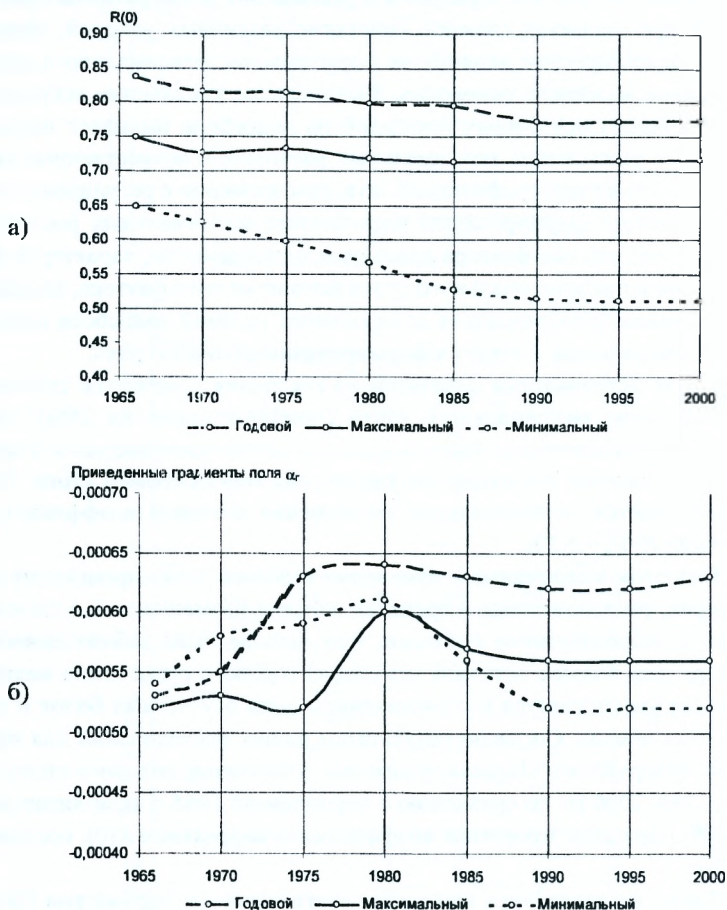


Рис. 9.4. Зависимости коэффициента  $R(0)$  (а) и градиента поля ( $\alpha_r$ ) (б), входящего в уравнение пространственно-корреляционных функций стока рек Белорусского Полесья, от нарастающего пятилетнего периода наблюдений (до 1966, до 1970, до 1975, до 1980 и т. д.)



Анализ полученных результатов позволяет с достоверностью сказать, что, начиная с 1990 года, значения как  $R(0)$ , так и  $\alpha_r$  практически не меняются. Это характеризует стабильность (изменение величин в пределах ошибки их расчета – 5%) изменения величины коэффициента корреляции на единицу расстояния между центрами водосборов, что позволяет говорить о прекращении интенсивных антропогенных воздействий на речной сток после массовых мелиораций 1965 – 1985 гг. в бассейнах водосборов рек Белорусского Полесья. При этом, начиная с 1990 г., происходит постепенное и равномерное восстановление естественных процессов формирования речного стока.

#### 9.4. Прогнозирование катастрофических изменений

Прогнозирование катастрофических изменений (скачков), т. е. значимых качественных изменений в структуре природных процессов, относится к группе достаточно сложных математических задач [Ивченко, 1998].

Высокая цена ошибочных решений при прогнозировании гидрометеорологических процессов обуславливает необходимость обращения к методологии системно-информационного анализа сложных процессов и базирования исследований не на классических моделях, а на моделях, сформированных на рандомизации параметров закона Пуассона или использующих аппарат производящих функций, принцип максимума неопределенности и лагранжевые вероятностные распределения.

Принципиальным отличием скачков в развитии гидрометеорологических процессов, от скачков в большинстве эргономических и экономических процессов и систем, является неопределенность и искажаемость «массовой» информации о прогнозируемых условиях функционирования атмосферы и гидросферы.

Построение математических моделей для таких процессов требует использования вариационных принципов с разработкой методов построения экстремальных законов распределения их параметров в условиях ограниченной информации по тенденциям развития факторных систем, как высшего, так и низшего порядков.

В качестве количественной оценки качественных изменений параметров процессов при скачкообразном их переходе из  $i$ -го состояния в  $(i+1)$ -ое состояние целесообразно принять момент времени, начиная с которого последующий процесс эволюционного развития будет осуществляться по другой траектории с новыми начальными данными.

Моделирование ступенчатых процессов возможно на детерминистической или стохастической основе [Шведовский, 2002].

Детерминистическая модель ступенчатых процессов может базироваться на следующей функции –

$$z = \sum_{i=0}^n a_i \cdot n \cdot (t - t_i), \text{ при } n \cdot (t - t_i) = \begin{cases} 1, \text{ если } (t - t_i) > 0 \\ 0, \text{ если } (t - t_i) \leq 0 \end{cases}, \quad (9.6)$$

где  $z$  – мера, определяющая уровень процесса;  $a_i$  – величина скачка меры  $z$ , связанная с появлением  $i$ -ой «разладки» эволюционного развития процесса;  $t$  – момент реализации  $i$ -го скачка, определяемый через функционал распределения  $F(z)$  и  $G(t)$

$$J_o[F(z), G(t)] = \frac{\sigma_z \cdot \sigma_t}{(\bar{z} + J_2) \cdot (\bar{t} + J_2)}. \quad (9.7)$$

Такой подход позволяет по маргинальным распределениям величин и коэффициентам корреляции определить совместную плотность, ее параметры и построить моделирующий алгоритм

$$f(z_1, \dots, z_n) = \frac{\partial^n F(z_1, \dots, z_n)}{\partial z_1 \dots \partial z_n}. \quad (9.8)$$

Если координаты вектора  $Z$  независимы, то

$$F(z_1, \dots, z_n) = \prod_{i=1}^n F_i(z_i), \quad (9.9)$$

и тогда многомерная случайная величина  $Z$  моделируется путем последовательного моделирования каждой из ее независимых координат  $(z_i, i = \overline{1, n})$ , подчиненных закону распределения  $F_i(z_i)$ .

Если координаты зависимы, то

$$f(z_1, \dots, z_n) = f_1(z_1) \cdot f_2(z_2 / z_1) \dots f_n(z_n / z_1, \dots, z_{n-1}), \quad (9.10)$$

и тогда моделирование случайной величины можно свести к последовательному моделированию координат с условными плотностями распределений, при инверсировании функции с помощью операторных рядов.

Фактически реализация модели требует отыскания двух функций  $a_i = f_i(i)$  и  $t = f_2(j)$ , одна из которых возрастающая  $f_1(i)$ , а вторая  $f_2(j)$  – убывающая, т. е. по мере развития индустриального общества величина техногенных скачков возрастает. Однако эти модели не в полной мере учитывают случайные факторы, объективно присущие гидрометеорологическим процессам.

Исходя из принципа максимума неопределенности, наиболее приближенной к истинному состоянию будет модель двумерного показательного распределения Гумбеля

$$f(z, t) = \frac{1}{x_o \cdot t_o} \cdot \exp \left[ \frac{z}{z_o} - \frac{t}{t_o} \right] \cdot \left[ 1 + P \cdot \left( 2 \cdot \exp \frac{z}{z_o} - 1 \right) \cdot \left( 2 \cdot \exp \frac{t}{t_o} - 1 \right) \right], \quad (9.11)$$

где  $z_o$  и  $t_o$  – средние величины скачков и периодичность их появления;  $p$  – параметр закона распределения и  $p = 4r_{zt}$ ;  $r_{zt}$  – коэффициент корреляции.

С учетом проявления случайностей для маргинальных (частных) распределений плотность может быть описана в виде –

$$n(z, t) = f(z) \cdot q(t) \cdot \{ 1 + \gamma \cdot [1 - 2 \cdot F(z)] \cdot [1 - 2 \cdot G(t)] \}, \quad (9.12)$$

где  $q$  – параметр закона распределения, определяемый линейной функцией коэффициента корреляции и  $q = r_{zt} \cdot J_o[F(z), G(t)]$ ;  $J_o[F(z), G(t)]$  – функционал, определяемый маргинальным распределением.

Наиболее простой моделирующий алгоритм, полученный нами для локальных, относительно однородных процессов, имеет вид:

$$\left. \begin{aligned} z &= z_o \ln \alpha_1; \quad t = t_o \cdot \ln \frac{2k}{\sqrt{(1+k)^2 - 4k\alpha_2 - 1 + a}}; \\ k &= p \cdot \left( 2 \cdot \exp \frac{z}{z_o} - 1 \right); \quad \alpha_1 = F(z); \quad \alpha_2 = F_{2/z}(t/z) \end{aligned} \right\}. \quad (9.13)$$

Так как зачастую ограниченность объема информации затрудняет определимость маргинальных распределений  $F(z)$  и  $G(t)$ , то используя метод квантилей экстремального распределения, построенного по распределению малой выборки, нами определена и сглаженная функция квантилей, которая может быть прообразом моделирующего алгоритма и

$$Z = \sum_{k=1}^{n-1} C_k \cdot p^k; \quad t = \sum a_k \cdot q^k. \quad (9.14)$$

Используя эти маргинальные квантильные функции, можно моделировать процессы, используя бутстреп – процедуры полей случайных коррелированных величин.

Не менее проблематичной является и учет наличия в выборке аномальных значений. Наиболее оптимальным путем решения этой проблемы будет отыскание специального критерия, учитывающего степень взаимной зависимости измерений многомерной случайной величины.

Основой критерия принимаем ковариационную матрицу вида

$$Q = \begin{vmatrix} D_z & K_{zt} \\ K_{tz} & D_t \end{vmatrix}, \quad (9.15)$$

где  $D$  – дисперсия;  $K$  – корреляционный момент случайной величины  $Y = (z, t)$ .

Сопоставление матриц для выборки с аномальным значением и без него можно найти меру их отличия  $v = \det Q^- / \det \bar{Q}$ , где  $\det Q^- (\bar{Q})$  – определитель матрицы.

Взяв за основу распределение Вейсбула, имеем –

$$F(v) = 1 - \exp\left(- (v/v_0)^m\right), \quad (9.16)$$

где  $m$  и  $v_0$  – параметры формы и масштаба, зависящие от объема выборки  $n$ , не зависящие от вида закона распределения.

Зная вид параметров закона распределения, легко определяются критическое значение величины  $v_{кр}$  для заданного объема выборки и уровень значимости  $\alpha$ , т. е. –

$$v_{кр} = v(-\ln(1-\alpha))^{1/m}. \quad (9.17)$$

Следует также отметить, что не менее перспективным в анализе влияния вариации потоков событий на оценку продолжительности сформированной тенденции процессов является и метод подбора, базирующийся на рандомизации интенсивности перехода  $\lambda$  и последующем осреднении вероятностей состояния факторных систем, с учетом маргинального распределения параметра  $\lambda$ .

Определяя суммарный оператор факторных систем в виде –

$$A_z = \int_a^b K(z, t) z(t) dt = U(z), \quad (9.18)$$

где  $z(t)$  – искомая функция из пространства  $F$ ;  $U(z)$  – заданная функция из пространства  $Q$ ;  $F$  и  $Q$  – метрические пространства, и задаваясь подклассом возможных решений  $M (M \subset F)$ , определяем приближенное решение, для которого элемент  $z_0$  из множества  $M$  принимает значение с нижней границей, т. е.

$$p_u(A_{z_0}, u) = \inf_{z \in M} p_u(A_z, u). \quad (9.19)$$

Имея, например, информацию о времени пребывания процесса в определенном состоянии, т. е. выборку  $T_{pk} (k = 1, \dots, n)$ , модель можно представить в следующем виде –

$$\int_0^{\infty} (1 - itm)^{-1} \frac{\lambda \cdot (\lambda_m)^{s-1} \cdot \exp(-\lambda m)}{T(s)} dm = \frac{1}{n} \cdot \sum_{k=1}^n \exp(it \cdot T_{pk}). \quad (9.20)$$

Его решение позволяет определить величины –

$$S = \frac{2m_{Tp}^2}{\sigma_{Tp}^2 - m_{Tp}^2} \quad \text{и} \quad \lambda = \frac{2m_{Tp}}{\sigma_{Tp}^2 - m_{Tp}^2}, \quad (9.21)$$

где  $T_p$  – время пребывания процесса в определенном состоянии.

Следует указать и на возможность использования метода «псевдо-состояний», сущность которого в том, что состояние факторных систем с немарковскими потоками переходов заменяют эквивалентной группой фиктивных состояний с марковскими потоками переходов. А это позволяет использовать аппарат теории марковских процессов.

### 9.5. Особенности прогнозирования в условиях множественной неопределенности

Концептуальная схема прогнозных исследований процессов в любых условиях имеет вид (рис. 9.5).

Совокупность же прогнозных задач может быть описана следующим образом [Шведовский, 1998]:

– известны начальные ( $A_1$ ) и конечные ( $A_T$ ) состояния системы. Необходимо определить состояние  $A_k$ , где  $m > k > T$ ;

– известно множество последовательных состояний ( $A_1, A_2, \dots, A_{m-1}$ ). Необходимо определить конечное состояние  $A_k$ ;

– известны промежуточные и конечное состояния системы ( $A_2, A_3, \dots, A_m$ ) или конечное состояние  $A_m$ . Необходимо определить начальное  $A_1$  или некоторые промежуточные состояния  $A_2, A_3$ ;

– известны все состояния системы ( $A_1, A_2, \dots, A_m$ ). Необходимо отыскать

набор элементов системы  $a_1, a_2, \dots, a_m$  и ее структуру  $\overrightarrow{a_1}, \overrightarrow{a_2}, \dots, \overrightarrow{a_m}$ ;

– известен набор элементов системы ( $a_1, a_2, \dots, a_m$ ). Необходимо описать поведение системы с переходами последовательных состояний  $\overrightarrow{A_1}, \overrightarrow{A_2}, \dots, \overrightarrow{A_m}$ ;

– известен последовательный переход системы по состояниям ( $\overrightarrow{A_1}, \overrightarrow{A_2}, \dots, \overrightarrow{A_m}$ ). Необходимо отыскать структуру перехода состояний и

основных элементов ( $\overrightarrow{a_1}, \overrightarrow{a_2}, \dots, \overrightarrow{a_m}$ ).

Для решения этих задач в целом применимы следующие группы решений и подходов (рис. 9.6).







Конфликтная природа процессов требует формализации неопределенной (неоднозначной, недостоверной, неизвестной) информации методами математических теорий. Наиболее эффективными являются теории вероятности, ошибок, интервальных средних, субъективных вероятностей, многозначной логики нечетких множеств и нечетких мер и интегралов [Ивченко, 1998].

Применимость всех этих теорий для решения конкретных прогнозных задач зависит от особенностей учета ими факторов неопределенности (табл. 9.13).



Рис. 9.6. Возможные группы решений и подходов прогнозных задач

Анализ табл. 9.13 позволяет отметить, что одной из наиболее эффективных математических теорий при решении проблем множественностью неопределенностей является теория нечетких интегралов, множеств и, особенно, нечетких мер, базирующиеся на нечетких процессах.

Таблица 9.13

**Применимость математических теорий для решения прогнозных задач**

Учитываемая характеристика неопределенности	Возможности теорий по учету факторов неопределенности						
	Вероятности	Ошибок (интервальных моделей)	Интервальных средних	Субъективных вероятностей	Многозначной логики	Нечетких множеств	Нечетких мер и интервалов
Физическая числовая неопределенность	+	+	+	+	-	+	+
Физическая нечисловая неопределенность	+	-	+	+	+	+	+
Противоречия между точностью и неопределенностью	-	-	+	+	+	+	+
Возможность количественной оценки неопределенности	+	-	-	+	-	+	+
Эффективность формализации полного незнания	-	+	+	+	+	+	+
Требования жесткого определения всех событий, факторов и характеристик	-	+	+	-	+	+	+
Возможность эффективного учета взаимовлияния неопределенности	-	-	-	-	+	-	+
Возможность получения оптимистических и пессимистических оценок и уровня доверия к ним	+	-	+	+	-	+	+
Единство подхода к представлению точных, неполных, неопределенных и нечетких знаний	-	-	-	-	-	-	+
Возможность работы с неопределенной информацией на базе малых статистических выборок	-	+	+	-	+	+	+

*Примечание:* (+) – возможен, (-) – невозможен учет факторов неопределенности.

Отметим, что нечеткая мера является обобщенным понятием вероятностной меры, свободной от ряда ограничений. При этом сама мера –

это функция множества  $m:P(x) \rightarrow R$ , где  $P(x)$  – множество всех подмножеств  $x$ , и  $R$  – множество действительных чисел [Бочарников, 1998].

С математической точки зрения нечеткая мера  $q(\bullet)$  является однопараметрическим расширением вероятностной меры, которая удовлетворяет условиям ограниченности, монотонности и непрерывности.

Что касается понятия нечетких процессов, то под ними необходимо понимать процесс, состояние которого в каждый момент времени  $t=T$  может быть описано некоторым распределением нечеткости  $\mu_t(\omega) \in F(\Omega)$  на пространстве состояний процесса  $\Omega$ .

Состояние любой гео- и экосистемы может быть определено совокупностью динамических процессов, которые формируются действием совокупности внешних и внутренних факторов  $x = \{x\}$ , при этом для фактора  $x_k \in x$  возможное состояние процесса ограничивается некоторым нечетким процессом, описываемым нечетко-интегральным уравнением вида:

$$\sigma_{x_k} \left( \frac{\omega}{t_{np}} \right) = \int_T h_{x_k}(\omega, t_{np}) \cdot \int_{\Psi T(\omega)} f_{t_{np}}(\omega) \cdot g_\omega(\bullet), \quad (9.22)$$

где  $t_{np} \in T$ ;  $T$  – нечеткий временной интервал прогноза;  $\sigma_{x_k} \left( \frac{\omega}{t_{np}} \right)$  – функция распределения нечеткой меры (нечеткости) на  $T$ , связывающей пространства  $T$  и  $\Omega$ , т. е.  $\sigma_{x_k} \left( \frac{\omega}{t_{np}} \right): T \rightarrow [01]$ ;  $\int(\bullet)$  – нечеткий интеграл;  $g_\omega(\bullet)$  – расширенная нечеткая мера;  $h_{x_k}(\omega, t_{np})$  – функция преобразования нечетной динамической системы, определяющей её динамику и  $h_{x_k}(\omega, t_{np}): (\Omega \times T) \times \Omega \rightarrow [01]$ ;  $f_{t_{np}}(\omega)$  – функция распределения плотности произвольного нечеткого множества.

Зная или задаваясь важностью влияния различных факторов  $p(x): X \rightarrow [01]$ , прогнозирование состояния системы (объекта) сводится к нахождению такого нечеткого процесса, который бы агрегировал исходный процесс с учетом функции  $p(x)$ , т. е.

$$\bar{S}_{t_{np}}(\omega) = F(\sigma_{x_k}(\omega, t_{np}), p(x)), \quad (9.23)$$

где  $F$  – оператор оперирования функции.

Таким операторам может выступать нечеткий интеграл на множестве  $X$  по некоторой нечеткой мере  $\bar{\omega}_x(\bullet)$ . А так как агрегирующие

свойства нечеткого интеграла определяются распределением меры  $\bar{\omega}_x(\bullet)$ , то в зависимости от нее он может определяться как операцией со свойствами объединения, так и операцией со свойствами нечетких множеств со следующими мерами:

– для объединения множеств

$$q_F = \int_x \eta(x) \cdot \bar{\omega}(\bullet), \quad (9.24)$$

где  $\eta(x_j) = \frac{n-j}{n-1}$ ,  $x_j \in X$ ;

– для пересечения множеств

$$l_F = 1 - q_F, \quad (9.25)$$

С учетом функции важности  $p(x)$  нечетно-интегральное уравнение принимает вид:

$$\sigma'_{x_k}(\omega, t_{np}) = (p(x_k) l_F) \cdot \sigma_{x_k}(\omega, t_{np})^{(p(x_k) q_F)}, \quad (9.26)$$

и соответственно –

$$\bar{S}_{i, np}(\omega) = \int_x \sum (x_k | \omega, t_{np}) \cdot \bar{\omega}_x(\bullet | \omega), \quad (9.27)$$

где  $\bar{\omega}_x(\bullet | \omega)$  – условная нечеткая мера оператора  $F$ .

Для описания дискретного нечеткого процесса можно использовать уравнение вида:

$$\mu_{i+1}(\omega) = \int_{\Omega_1} h(\omega, \omega') \cdot \int_{\Omega} \mu_i(\omega) \cdot q(\bullet), \quad (9.28)$$

где  $\mu_i(\omega)$  – функция, описывающая состояние нечеткого процесса в  $i$ -ый момент времени;  $q(\bullet)$  – нечеткая мера на пространстве состояний;  $h(\omega, \omega')$  – нечеткое отношение, реализующее оператор вход-выход на расчетном временном интервале и  $h(\omega, \omega') = \mu_{i+1}(\omega') \cdot \chi_{H_{\mu_{i+1}}}(\omega)$ ,  $\chi_{H_{\mu_{i+1}}}(\omega)$  – характеристическая функция множества для фиксированного  $\omega' \in \Omega$ .

Мера соответствия  $\mu^m(\omega)$  истинному состоянию нечеткого процесса может определяться мерой возможности (оптимистический) или мерой необходимости (пессимистический вариант), т. е.

$$\begin{cases} v(\mu^u, \mu^m) = \min_{\omega \in \Omega} \{ \mu^u(\omega); \mu^m(\omega) \} \\ v(\mu^u, \mu^m) = \max_{\omega \in \Omega} \{ \mu^m(\omega); 1 - \mu^u(\omega) \} \end{cases} \quad (9.29)$$

Для модели нечеткого процесса на пространстве  $\Omega$  нечетко-дифференциальное уравнение имеет вид –

$$fd\mu(\omega) = [h(\omega u) \cdot \bar{c}(\omega u)] fdf_r^*(\omega), \quad (9.30)$$

где  $\bar{c}(\omega u)$  – нечеткая функция управления, заданная на пространстве  $U$  – значений управляющего воздействия ( $u \in U$ );  $h(\omega u)$  – оператор нечеткой динамической системы (объекта);  $f_r^*(\omega)$  – нечеткий процесс на  $\Omega$  пространстве ( $\omega \in \Omega$ ).

Полагая, что эффективность управления нечеткой динамической системой определяется множеством критериев  $\theta(v)$  с нечеткой мерой их важности  $q_\theta(\bullet): 2^\theta \rightarrow [01]$ , в общем случае потери (ухудшение, изменение)  $l(v, u)$  по каждому из показателей (факторов, условий)  $v \in \Omega$  зависят от выбора управления  $u(t, \omega) \in U$  в конкретный момент времени и для конкретного состояния системы, т. е.  $l(v, u): \theta \times U \rightarrow [01]$ .

При этом нечеткое отношение  $l(\omega, u)$  характеризует распределение меры возможности потерь по  $v \in \Omega$  при  $u \in U$ , а  $l'(v, u)$  – меру выгоды и  $l'(v, u) = 1 - l(v, u)$ .

Отсюда соответственно максимально возможная выгода по критерию  $v \in \theta$  при выборе управления из подмножества  $E \leq U$  определится соотношением

$$j = \max_{u \in E} [1 - l(v, u)], \quad (9.31)$$

а минимально возможные потери  $V$  соотношением

$$v = 1 - \max_{u \in E} [\chi_E(v, u) \cdot (1 - l(v, u))], \quad (9.32)$$

где  $\chi_E(\bullet)$  – характеристическая функция множества  $E \leq U$ .

И тогда соответственно выигрыш по всем критериям в текущий момент времени определится зависимостью:

$$l'_\theta(u) = \int_\theta l'(v, u) \cdot g_\theta(\bullet), \quad (9.33)$$

а интегральный выигрыш определится функционалом:

$$I = \int_\Gamma l'_\theta(u) \cdot \int_{\omega, (\bullet|\omega)} f_r(\omega) \cdot q(\bullet), \quad (9.34)$$

где  $q(\bullet): 2^\Omega \rightarrow [01]$ ,  $f_r: \Omega \rightarrow [01]$  – нечеткий процесс на  $\Omega$  задающей временную нечеткость динамики нечеткого динамического процесса.

Отсюда в соответствии с принципом оптимальности Беллмана функциональное неравенство для определения оптимального управления примет вид:

$$I(u^*) \geq \max_{u \in U} \int_{\theta} I'(u, v) \cdot q_{\theta}, \quad (9.35)$$

где функция выигрыша  $I'(u, v)$  является нечетким аналогом функции Беллмана.

Если рассматривать случай, когда управление определяется не только временем  $t \in T$ , но и состоянием системы (объекта)  $\omega \in \Omega$ , т. е. выбором нечеткой стратегии управления  $S_t(\omega): U \times \Omega \rightarrow [01]$  в момент  $t \in T$ , то нечеткий выигрыш определяется нечетким отношением  $I(v, S_t(\omega): \theta \times (U \times \Omega)) \rightarrow [01]$  с расчетными зависимостями:

$$\varphi_{\tau}(u) = \int_{\mu_t(\omega)} b(u, \omega) \cdot q(\bullet); b(u, \omega) = \int_{\theta} [1 - I(u, v) \cdot \sigma_{\theta}(\bullet | \omega)], \quad (9.36)$$

где  $\mu_t(\omega)$  – текущее состояние системы (нечеткое подмножество);  $\sigma_{\theta}(\bullet | \omega)$  – условная нечеткая мера важности критериев  $v \in \theta$  в состоянии  $\omega \in \Omega$ .

Отсюда основу решения любых экологических проблем методами теории нечетких интегралов, множеств и мер составляет формализация нечетких данных.

Рассмотрим формализацию нечетких данных для оценки уровня риска  $d \in D$  формирования критического критического состояния водного режима Белорусского Полесья (табл. 9.7).

При формализации использованы  $q$  – нечеткие меры, следующим образом зависящие от параметра нормировки:

$$\left. \begin{aligned} M_n \in -1 < \lambda < 0; & M_{sz} \in \lambda = -1; & M_H \in \lambda > 0; \\ M_D \in \lambda > 0; & M_{sp} \in \lambda = 0. \end{aligned} \right\} \quad (9.37)$$

Отметим также, что любые четкие данные представимы примитивным классом мер, т. е. мерами Дирака и

$$\mu(d) = \begin{cases} 1, & x_0 \in A; \\ 0, & x_0 \notin A, \end{cases} \quad -\infty < \lambda < \infty \quad (9.38)$$

где  $x_0$  – заданный элемент в пространстве (носителе меры)  $X$ .

Бесспорно, предложенная формализация нечетких данных не ограничивают всего спектра возможностей формализации. При необходимости, в каждой конкретной решаемой задаче, могут использоваться и другие варианты формализации, позволяющие более широко описать спектр разнородных и малодостоверных данных.



Таблица 9.7

## Формализация нечетких данных для оценки уровня риска

Описание данных	Формализованное представление данных
Полная уверенность, что риска нет	$\mu(d) = \begin{cases} 0, & d \in D \setminus \{6\} \\ 1, & d = 6 \end{cases}$
Полная уверенность, что риск есть, но тяжело оценить его значение	$\mu(d) = \begin{cases} 1, & \mu(\cdot) = M_{\text{вз}}(\cdot), d \neq 6 \\ \lambda \in [0,1] \mu(\cdot) = M_n(\cdot), M_D(\cdot), d \neq 6 \\ 0, & \mu(\cdot) = M_H(\cdot), d = 6 \end{cases}$
Полная уверенность, что риск есть, однако известно, что его значение от минимального до допустимого	$\mu(d) = \begin{cases} 1, & d \in [2,3] \\ 0, & d \notin [2,3] \\ 0, & d = 6 \end{cases}$
Полная уверенность, что риск есть, но значение его четко неизвестно	$\mu(d) = \begin{cases} \varphi(d), & d \in D \setminus \{6\} \\ 0, & d = 6 \end{cases}$ $\varphi(d)$ — распределение нечёткости для риска низкого уровня
Полная уверенность, что риск допустимый	$\mu(d) = \begin{cases} 0, & d \in D \setminus \{3\} \\ 1, & d = 3 \end{cases}$
Вполне правдоподобно, что есть риск достаточно высокого уровня, но имеется и ненулевая возможность $\lambda$ что риска нет	$\mu(d) = \begin{cases} M_n, & d \in D \setminus \{5\} \\ \lambda, & d = 6 \end{cases}$ $M_n$ — распределение меры правдоподобия для риска высокого уровня
Неизвестно есть риск или нет, но если есть, то его величина вообще неизвестна	$\mu(d) = \begin{cases} 1, & \mu(\cdot) = M_{\text{вз}}(\cdot) \\ \lambda \in [0,1] \mu(\cdot) = M_n(\cdot), M_D(\cdot) \\ 0, & \mu(\cdot) = M_H(\cdot) \end{cases}$
Вполне возможно, что риска нет, но имеется и не нулевая возможность, что он есть и не выше критического	$\mu(d) = \begin{cases} \lambda, & d < 3 \\ 0, & d \geq 3 \\ 1, & d = 6 \end{cases}$

*Примечание:* 1 – событие возможно; 2 – событие невозможно;  $M_n$  – мера правдоподобия;  $M_{\text{вз}}$  – мера возможности;  $M_n$  – мера необходимости;  $M_D$  – мера доверия;  $M_{\text{вп}}$  – мера вероятности. Оценка риска:  $D = \{1 - \text{отсутствует}, 2 - \text{минимальный}, 3 - \text{допустимый}, 4 - \text{критический}, 5 - \text{недопустимый}, 6 - \text{неизвестно, является ли это риском}\}$ ;  $\mu$  – функция принадлежности;  $\lambda$  – параметр нормировки.

На сегодня теория нечетких множеств, мер и интегралов в решении проблем прогнозирования процессов и соответственно состояния и управления гео- и экосистемами практически не используется. Имеющиеся комплексы программных продуктов Fuzzy-технология и Expro-2000 позволяют лишь дать анализ данных, рисков и оценку событий при наличии информации, не вызывающей должного доверия (словесной, разнородной, низкокачественной) и недостаточной известности факторов будущего.

Отсюда разработка методики решения прогнозных задач в условиях неопределенности на базе нечетких процессов в настоящее время весьма актуальна.

## ЛИТЕРАТУРА

- Алексеев Г. А.* Объективные методы выравнивания и нормализации корреляционных связей. – Л.: Гидрометеиздат, 1971. – 363 с.
- Алексеев Г. А.* Пространственная интерполяция наблюдаемых гидрометеорологических элементов, их статистических параметров квантилей с оценкой случайных погрешностей // Труды IV Всесоюзного гидрологического съезда – Т. 3. – Л.: Гидрометеиздат, 1975. – С. 430 – 444.
- Алехин Ю. М.* Статистические прогнозы в геофизике. – Л.: ЛГУ, 1963. – 59 с.
- Анохин Ю. А., Острогомилский А. Х.* Некоторые вопросы математического моделирования процессов циркуляции веществ в природных геофизических средах / Проблемы экологического мониторинга и моделирования экосистем. – Л.: Гидрометеиздат, 1979. – С. 147 – 160.
- Аношко В. С., Трофимов А. М., Широков В. М.* Основы географического прогнозирования. – Минск: Высшая школа, 1985. – 239 с.
- Антонцев С. Н., Епихов Г. П., Кошеваров А. А.* Системное математическое моделирование процессов водообмена. – Новосибирск: Наука (СО), 1986. – 361 с.
- Белов П. Г.* Способ системного прогнозирования технического риска // Проблемы безопасности при чрезвычайных ситуациях. ВИНТИ. – М., 1994. – Вып. 4. – С. 26 – 33.
- Беляев В. И.* Теория сложных систем. – Киев: Наукова думка, 1978. – 216 с.
- Борисов А. Н.* Обработка нечетной информации в системах принятия решений. – М.: Радио и связь, 1989, – 304 с.
- Боровиков В. П., Боровиков И. П.* Статистический анализ и обработка данных в среде Windows. – М.: Информационно-издательский дом «Филинь», 1997. – 608 с.
- Бочариков В. П.* Fuz zy – технология. Математические основы. Практика моделирования в экономике. – С. – Пб.: Наука, РАН, 2001. – 328с.
- Бочариков В. П.* Модель управляемого непрерывного нечетного процесса на основе нечетно-интегрального уравнения. Проблемы управления и информатики – Киев: КМУГА, 1998. – С. 72 – 77.
- Бочаров М. К.* Методы математической статистики в геофизике. – М.: Мысль, 1971. – 371 с.

*Брукс К., Карузере Н.* Применение статистических методов в метеорологии/ Пер. с англ. Е.Ф. Ивановой, Л.Л. Френкеля / Под ред. Н.А. Багрова. – Л.: Гидрометеиздат, 1963. – 416 с.

*Булавко А. Г.* Водный баланс речных водосборов – Л.: Гидрометеиздат, 1971. – 304 с.

*Бурлибаев М. Ж., Шведовский П. В., Волчек А. А.* Концептуальные основы оптимизации решений экологических проблем/ Матер. МНТК «Проблемы гидрометеорологии и экологии». – Алматы: КазНИИМОСК, 2001. – С. 353 – 357.

*Верещагин М. А., Наумов Э. П., Шанталинский К. М.* Статистические методы в метеорологии. – Казань: Казанский университет, 1990. – 110 с.

*Волчек А. А.* Автоматизация гидрологических расчетов// Тр. междунац. конф. «Водохозяйственное строительство и охрана окружающей среды»/ Биберах – Брест – Ноттингем, 1998. – С. 55 – 60.

*Волчек А. А.* Характер синхронных колебаний стока рек Беларуси. Брест, Вестник БГУ – 2001. – №2(18) – С.31 – 36.

*Волчек А. А., Лукша В. В.* Антропогенное воздействие на водные ресурсы рек// Материалы международной конференции «Природнае асяроддзе Полесся: сучасны стан і яго змены» / ОПП НАН Беларусі. – Т. 1 – Брест, 2002. – С. 177 – 182

*Волчек А. А., Лукша В. В.* Временная изменчивость пространственно-корреляционных функций речного стока рек Белорусского Полесья// Материалы IV международной научно-практической конференции «Экологические проблемы Полесья и сопредельных территорий»/ Гомельский гос. универ. им. Ф. Скорины. – Гомель, 2002. – С. 47 - 49

*Волчек А. А., Лукша В. В.* Пространственная структура изменения годового стока рек Беларуси// Материалы V международной научно-практической конференции «Экологические проблемы Полесья и сопредельных территорий»/ Гомельский гос. универ. им. Ф. Скорины. – Гомель, 2003. – С. 32 – 34

*Волчек А. А., Шведовский П. В., Образцов Л. В.* Математические модели в природопользовании. – Минск: БГУ, 2002. – 281 с.

*Волчек А. А., Шведовский П. В., Пойта П. С.* Математические методы в природопользовании и природообустройстве. – Минск: БГУ, 2003. – 294 с.

*Гандин Л. С.* Объективный анализ метеорологических полей. – Л.: Гидрометеиздат, 1963. – 267 с.

- Гандин Л. С., Казан Р. Л.* Статистические методы интерполяции метеорологических данных. – Л.: Гидрометеиздат, 1976. – 383 с.
- Герасимович А. И.* Математическая статистика. – Минск: Высшая школа, 1983. – 279 с.
- Гергиевский В. Ю., Ежов А. В.* и др. Возможные изменения водных ресурсов и водного режима в бассейне Днепра при различных сценариях потепления климата // Материалы научно- практической конференции «Водные ресурсы и устойчивое развитие экономики Беларуси». – Минск: ЦНИИКИВР, 1996. – С. 21 – 23.
- Гуляев А. И.* Временные ряды в динамических базах данных. – М.: Наука, 1989. – 209 с.
- Гумбель Э.* Статистика экстремальных значений. – М.: Мир, 1965. – 392 с.
- Гурарий В. И., Науменко Г. Н.* Приложение теории полезности к принятию водоохранных мероприятий// Проблемы охраны вод. Харьков, 1978. – Вып. 9. – С. 94 – 102.
- Девдориани А. С., Грейар В. Л.* Роль кибернетических методов в изучении и преобразовании природных комплексов / Изв. АН СССР. Серия геогр. Наук, 1978. – Вып. 6. – С. 94 – 102.
- Доспехов Б. А.* Методика полевого опыта (с основами статистической обработки результатов исследований). – 5-е изд., доп. и перераб. – М.: Агропромиздат, 1985. – 351 с.
- Дрейпер Н., Смит Г.* Прикладной регрессионный анализ: В 2-х кн. Кн. 1/ Пер. с англ. – 2-е изд., перераб. и доп. – М.: Финансы и статистика, 1986. – 366 с.
- Дроздов О. А.* Колебания естественного увлажнения в связи с анализом антропогенных изменений климата и увлажнения// Водные ресурсы. – 1990. – №2. – С. 5 - 15.
- Дроздов О. А.* О надежности использования аналогов прошлого для прогнозов водного режима рек на будущее // Водные ресурсы. — 1994. – № 4 – С. 7 – 12.
- Дэвис Дж. С.* Статистический анализ данных в геологии: Пер. с англ. в 2 кн./ Пер. В. А. Голубевой/ Под ред. Д. А. Родионова. Кн. 1 – М.: Недра, 1990. – С.319 – 427.
- Ивченко Б. П., Мартыщенко Л. А.* Информационная экология. – С.-Пб.: Нормедиздат, 1998. – 201 с.
- Изменение климата Беларуси и их последствия//* Под общ. ред. В. Ф. Логинова. – Минск: ОДО «Тонпик», 2003. – 330 с.

- Изменения климата и использование климатических ресурсов//* Под общ. ред. П. А. Ковриго. – Минск: БГУ, 1979. – 219 с.
- Исаев А. А.* Статистика в метеорологии и климатологии. – М.: МГУ, 1988. – 248 с.
- Колде Я. К.* Практикум по теории вероятностей и математической статистике. – М.: Высшая школа, 1991. – 157 с.
- Константинов А. Р., Химин Н. М.* Применение сплайнов и метода остаточных отклонений в гидрометеорологии. – Л.: Гидрометеоздат, 1983. – 267 с.
- Крамер Г.* Математические методы статистики. – М.: Мир, 1975. – 648 с.
- Крапивин В. Ф., Свирезжев Ю. М., Тарко А. М.* Математическое моделирование глобальных биосферных процессов. – М.: Наука, 1982. – 273 с.
- Куржиновский А. В.* Управление и наблюдение в условиях неопределённости. – М.: Наука, 1997. – 238 с.
- Кучмент Л. С.* Математическое моделирование речного стока. – Л.: Гидрометеоздат, 1972. – 191 с.
- Лисичкин В. А.* Теория и практика прогностики. – М.: Наука, 1972. – 224 с.
- Литтл Р. Дж. А., Рубин Д. Б.* Статистический анализ данных с пропусками/ Пер. с англ. – М.: Финансы и статистика, 1990. – 336 с.
- Логинов В. Ф.* Влияние мелиорации на региональный климат Беларуси // Природные ресурсы. – 1997. – № 1. – С. 24 – 27.
- Логинов В. Ф.* Изменение числа экстремальных климатических явлений в Беларуси за период инструментальных наблюдений // Природопользование. – Минск. – 1997. – Вып. № 3. – С. 24 – 25.
- Логинов В. Ф.* Климат Национального парка «Припятский»// Тр. Национального парка «Припятский»/ Туров – Мозырь, 1999. – С. 30 – 34.
- Логинов В. Ф.* Причины и следствия климатических изменений. – Минск, 1997. – 320 с.
- Логинов В. Ф.* Статистические оценки солнечно-атмосферных связей. Предпосылки использования солнечно-земных связей в прогнозах погоды и теории климата// Тр. Всесоюз. конф. по взаимосвязи метеорол. явлений и процессов в околоземн. косм. пространстве/ Москва, 9 – 11 января 1985 г. – М., 1987. – С. 59 – 68.



- Логинов В. Ф.* Управление гидрометеорологическими данными. – Минск: БГУ, 2002. – 174 с.
- Логинов В. Ф., Витченко А. Н.* Климатические ресурсы Беларуси и их изменение// Природные ресурсы. – 1999. – № 1. – С. 24 – 35.
- Логинов В. Ф., Гирская Э. И., Молодых В. А.* Закономерности изменчивости температуры при различных уровнях климатообразующих факторов// География и природ. ресурсы. – 1981. – № 2. – С. 29 – 34.
- Логинов В. Ф., Иконников В. Ф.* Межгодовая изменчивость увлажненности территории Беларуси и колебания уровня озер// Прикладная лимнология. Лимнологическое и геоморфологическое обоснование рационального природопользования. – Минск. – 2002. – Вып. 3. – С. 220 – 228
- Логинов В. Ф., Калинин М. Ю., Иконников В. Ф.* Современное антропогенное воздействие на водные ресурсы Беларуси. – Минск, 2000. – 283 с.
- Логинов В. Ф., Константинова Т. С., Микуцкий В. С.* Оценка влияния последствий антропогенной деятельности на изменение климата //Природныересурсы. – 2000. – №4. – С. 19 – 24.
- Логинов В. Ф., Кузнецов Г. П., Микуцкий В. С.* Математические методы в исследовании климата Беларуси// Природопользование. – Минск – 2002. – Вып. 8. – С. 75 – 87.
- Логинов В. Ф., Мельник В. И.* Современные тенденции изменения основных агроклиматических показателей теплового периода научные и прикладные аспекты оценки изменения климата и использования климатических ресурсов// Тр. межд. конф. – Минск, 2000. – С. 112 – 113.
- Логинов В. Ф., Микуцкий В. С.* Изменение амплитуды годового хода температуры в Беларуси// Природные ресурсы. – 1999. – № 3. – С. 121.
- Логинов В. Ф., Микуцкий В. С.* Спектрально-временной анализ Северо-Атлантического колебания и температуры Беларуси// Природопользование. – Минск – 2001. – Вып. 7. – С. 74 – 77.
- Логинов В. Ф., Микуцкий В. С., Каждан Е. Н.* Использование моделей общей циркуляции атмосферы для оценки климата в Беларуси// Природопользование.– Минск.– 2000. – Вып. 6.– С. 30 – 31.
- Лукша В. В., Шведовский П. В.* К проблеме определения срока функционирования агроландшафтных систем до формирования критических уровней экологической надёжности// Вестник БГТУ/ Водохозяйственное строительство, теплоэнергетика и экология. – Брест. – 2002. – №2. – с. 51 – 53.

*Львовский Е. Н.* Статистические методы построения эмпирических формул – М.: Высшая школа, 1988. – 239 с.

*Мартыщенко Л. А.* Математические задачи теории малых выборок и их приложение к испытаниям сложных технических систем. – Л.: МО СССР, 1975. – 216 с.

*Мартыщенко Л. А., Панов В. В.* Моделирование распределений, заданных характеристическими функциями. – М.: Кибернетика – 1985. – №3. – С. 19 – 26.

*Математические методы оценки агроклиматических ресурсов/ В. А. Жуков, А. П. Полевой, А. Н. Витченко.* – Л.: Гидрометеоздат, 1989. – 207 с.

*Международное руководство по методам расчета основных гидрологических характеристик.* – Л.: Гидрометеоздат, 1984. – 247 с.

*Мишустин Н. А.* Генетический метод расчета нормы годового стока неизученных рек и временных водотоков Белорусского Полесья// НТИ «Мелиорация и водное хозяйство» – 1979. – № 12. – С. 12 – 17.

*Мостеллер Ф., Тьюки Дж.* Анализ данных и регрессия: В 2-х вып./ Пер. с англ. Ю. Н. Благовещенского/ Под ред. и с предисл. Ю.П. Адлера. – М.: Финансы и статистика – 1982. – Вып.1; Вып.2 – 317 с.; 239 с.

*Никифоров И. В.* Последовательное обнаружение изменений свойств временных рядов. – М.: Наука, 1983. – 198 с.

*Никульченко Н. Н.* Использование регрессионного анализа для прогнозирования стока и восстановления среднемесячных расходов// Гидрохимические материалы. Мониторинг, самоочищение и математическое моделирование качества воды водных объектов. – Л.: Гидрометеоздат – 1989. – Т. VI. – С. 166 – 177.

*Павличенко Л. М.* Система многомерных статистических моделей анализа неполных эколого-гидрогеологических данных// Вопросы изучения водных ресурсов Центральной Азии. – Алматы: Галым, 1993. – С. 89 – 103.

*Пановский Г. А., Брайер Г. В.* Статистические методы в метеорологии. – Л., Гидрометеоздат, 1967. – 211 с.

*Перегудов Ф. Н., Тарасенко Ф. П.* Введение в системный анализ. – М., Наука, 1989. – 477 с.

*Поллард Дж.* Справочник по вычислительным методам статистики / Пер. с англ. В.С. Занадворова / Под ред. и с предисл. Е.М. Четыркина. – М.: Финансы и статистика, 1982. – 344 с.

**Поляк И. И.** Методы анализа случайных процессов и полей в климатологии. – Л., Гидрометеиздат, 1976. – 256 с.

**Пособие П1** – 98 к СНиП 2.01.14-83 «Определение расчетных гидрологических характеристик». – Минск: Министерство архитектуры и строительства Республики Беларусь, 2000. – 174 с.

**Привольский В. Е.** Климатическая изменчивость (стохастические модели, предсказуемость, спектры). – М., Наука, 1984. – 183 с.

**Пясковский Р. В., Померанец К. С.** Наводнения. Математическая теория и предсказания. – Л.: Гидрометеиздат, 1982. – 176 с.

**Репин С. В., Шейн С. А.** Математические методы обработки статистической информации с помощью ЭВМ. – Минск: Университетское, 1990. – 128 с.

**Рождественский А. В., Чеботарев А. И.** Статистические методы в гидрологии. – Л.: Гидрометеиздат, 1974. – 424 с.

**Сачок Г. И., Цуркова Т. Ф.** Математико-картографическое моделирование условий Белоруссии. – Минск: Наука и техника, 1984. – 253 с.

**Статистические методы в гидрологии.** – Л.: Гидрометеиздат, 1970. – 270 с.

**Статистические методы в природопользовании/ В. Е. Валуев, А. А. Волчек, П. В. Шведовский и др.** – Брест: БПИ, 1999. – 251 с.

**Тейлор Дж.** Введение в теорию ошибок/ Пер. с англ. – М.: Мир, 1985. – 272 с.

**Уилкс С.** Математическая статистика. – М.: Наука, 1967, – 508 с.

**Уланова Е. С., Сиротенко О. Д.** Методы статистического анализа в агрометеорологии. – Л.: Гидрометеиздат, 1968. – 198 с.

**Феллер В.** Введение в теорию вероятностей и ее приложение. – М.: Мир, 1987. – 572 с.

**Цветков В. Я.** Информатизация: Создание современных информационных технологий. – М.: Наука, 1990. – 412 с.

**Чернышев М. К.** Математическое моделирование иерархических систем. – М.: Наука, 1998. – С. 44 – 49.

**Чернышев М. К., Гаджиев М. Ю.** Математическое моделирование иерархических систем – М.: Наука, 1983. – 191 с.

**Чертко Н. К.** Математические методы в физической географии. – Минск: Университетское, 1987. – 151 с.

**Шведовский П. В.** Мелиорация и природная среда. – Минск: Ураджай, 1984. – 160 с.

**Шведовский П. В., Богомолов Ю. Г.** Системное математическое моделирование процессов природного водообмена: Экспрес-информ. – М. – 1988. – Сер.4 – Вып. 2. – С.8 – 16.

**Шведовский П. В., Бурлибаев М. Ж., Волчек А. А.** Концептуальные основы оптимизации решений экологических проблем// Проблемы гидрометеорологии и экологии: Материалы междунар. Конференции/ КазНИИМОСК.– Алматы, 2001.– С. 353 – 358.

**Шведовский П. В., Лукиш В. В.** Особенности математического моделирования скачков в развитии экологических систем и процессов// Вестник БГУ, Брест. – 2001. – № 2(18).– С. 28 – 31.

**Шебеко В. Ф.** Изменения климата под влиянием мелиорации. – Минск: Ураджай, 1977. – 209 с.

**Шелутко В. А.** Статистические модели и методы исследования многолетних колебаний стока. – Л.: Гидрометеиздат, 1984. –159 с.

**Шикломанов И. А.** Антропогенные изменения водности рек. – Л.: Гидрометеиздат, 1979. – 258 с.

**Эколого-социальные** аспекты освоения водно-земельных ресурсов и технологий управления режимами гидромелиораций / П. В. Шведовский, В. Е. Валуев, А. А. Волчек. – Минск: Ураджай, 1998. – 363 с.

**Юдин Д. Б.** Задачи и методы стохастического программирования. – М.: Мир, 1979. – 392 с.

## СОДЕРЖАНИЕ

	стр.
ПРЕДИСЛОВИЕ .....	3
1. ОСНОВНЫЕ ХАРАКТЕРИСТИКИ ГИДРОМЕТЕОРОЛОГИЧЕСКОЙ ИНФОРМАЦИИ И ЕЕ УПОРЯДОЧЕНИЕ ПО ПРОСТРАНСТВЕННЫМ И ВРЕМЕННЫМ ПРИЗНАКАМ .	4
2. СИСТЕМЫ СБОРА И МЕТОДЫ (ТЕХНОЛОГИИ) ОБРАБОТКИ ГИДРОМЕТЕОРОЛОГИЧЕСКИХ ДАННЫХ .....	10
2.1. Общие принципы .....	10
2.2. Теоретические основы и методы статистической обработки и анализа данных .....	22
2.3. Особенности проверки статистических гипотез о видах и свойствах распределения .....	36
2.4. Алгоритмы и методы контроля информации .....	49
2.5. Особенности автоматизированной обработки гидрометеорологической информации .....	56
2.6. Обработка режимной гидрометеорологической информации .....	65
3. ПРИНЦИПЫ ВЫЯВЛЕНИЯ И АНАЛИЗ ФУНКЦИОНАЛЬНЫХ СВЯЗЕЙ СЛУЧАЙНЫХ ВЕЛИЧИН. ....	73
3.1. Общие принципы выявления функциональных связей .	73
3.2. Особенности анализа нелинейных корреляционных зависимостей .....	82
3.3. Особенности выявления корреляционных связей многофакторных процессов .....	88
3.4. Особенности выбора оптимального числа предсказателей .....	102
3.5. Оценка значимости и тесноты связи между качественными признаками .....	104
3.6. Корреляция числовой случайной переменной с дихотомической .....	110
3.7. Особенности выявления тесноты нелинейных связей (информационно-энтропийный анализ) .....	110
4. МЕТОДЫ МАТЕМАТИЧЕСКОГО ОПИСАНИЯ И ВЗАИМООЦЕНКИ ПАРАМЕТРОВ ВЫЯВЛЕННЫХ ЗАКОНОМЕРНОСТЕЙ .....	115
4.1. Уравнение линейной регрессии с одним переменным фактором .....	115
4.2. Нелинейная парная регрессия .....	118
4.3. Линейная множественная регрессия .....	122
4.4. Нелинейная множественная регрессия .....	126

	стр.
4.5. Основные аспекты выбора оптимальной регрессионной модели . . . . .	128
5. МЕТОДЫ ОПРЕДЕЛЕНИЯ РАЗЛИЧНЫХ ФАКТОРОВ И УСЛОВИЙ ВЛИЯНИЯ НА ПРОЦЕССЫ И ЯВЛЕНИЯ . . . . .	134
5.1. Однофакторный дисперсионный анализ . . . . .	135
5.2. Двухфакторный дисперсионный анализ . . . . .	138
5.3. Критерии статической значимости флюктуаций природных процессов . . . . .	142
6. МЕТОДЫ ПРОСТРАНСТВЕННОГО ОБОБЩЕНИЯ ГИДРОМЕТЕОРОЛОГИЧЕСКОЙ ИНФОРМАЦИИ . . . . .	153
6.1. Оценка статистической макроструктуры полей . . . . .	156
6.2. Особенности оценки гидрометеорологических полей методами их геофизического разложения . . . . .	167
6.3. Методы оценки степени аномальности полей . . . . .	177
6.4. Оценка точности статистических характеристик структуры поля . . . . .	179
6.5. Особенности комплексного анализа статистической структуры гидрометеорологических полей . . . . .	184
6.6. Особенности восстановления стоковых характеристик методом пространственной интерполяции . . . . .	189
7. ОСОБЕННОСТИ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ . . . . .	193
7.1. Особенности анализа периодических колебаний . . . . .	196
7.2. Особенности выделения и анализ нерегулярных циклов . . . . .	201
8. ОСНОВЫ СТАТИСТИЧЕСКОГО ПРОГНОЗИРОВАНИЯ ПРИРОДНЫХ ПРОЦЕССОВ . . . . .	238
8.1. Особенности сверхдолгосрочного прогнозирования гидрометеорологических характеристик . . . . .	239
8.2. Особенности применения статистических методов в прогнозах погоды . . . . .	241
8.3. Вероятностное прогнозирование значений случайного процесса . . . . .	244
8.4. Прогноз по последнему значению . . . . .	246
8.5. Прогноз по математическому ожиданию . . . . .	246
8.6. Статистический прогноз по одной точке . . . . .	247
8.7. Динамико-стохастический метод сверхдолгосрочного прогноза . . . . .	249
8.8. К вопросу о точности прогнозирования . . . . .	252
9. ПРОГНОЗ ВЛИЯНИЯ ХОЗЯЙСТВЕННОЙ ДЕЯТЕЛЬНОСТИ И ИЗМЕНЕНИЙ КЛИМАТИЧЕСКИХ УСЛОВИЙ НА ВОДНЫЕ РЕСУРСЫ . . . . .	254



	стр.
9.1. Оценка современных трансформаций водного режима под воздействием технических мелиораций и хозяйственной деятельности на водосборах . . . . .	256
9.2. Оценка изменения водного режима в связи с прогнозами изменения климата . . . . .	260
9.3. Региональная оценка антропогенных влияний на сток рек Белорусского Полесья . . . . .	271
9.4. Прогнозирование катастрофических изменений . . . . .	277
9.5. Особенности прогнозирования в условиях множественной неопределенности . . . . .	281
ЛИТЕРАТУРА . . . . .	291

***Логинов Владимир Федорович***

1940 года рождения, окончил арктический факультет Ленинградского Высшего инженерного морского училища им. адмирала С. О. Макарова в 1963 году.

Доктор географических наук, профессор, академик Национальной академии наук Беларуси. Директор ГНУ «Институт проблем использования природных ресурсов и экологии Национальной академии наук Беларуси».

Автор более 340 научных работ, в том числе 15 монографий и брошюр, 2 справочников, 3 учебных пособий.

Основное направление научных исследований – изменение климата и их последствия в различных отраслях экономики.

***Волчек Александр Александрович***

1955 года рождения, окончил Брестский инженерно-строительный институт в 1978 г. по специальности «гидромелиорация».

Кандидат географических наук, доцент. Заместитель директора по научной работе ГНУ «Отдела проблем Полесья Национальной академии наук Беларуси».

Автор более 300 научных работ, в том числе 5 монографий, 2 пособий, 5 учебных пособий для ВУЗов, 3 изобретений.

Основное направление работы – водные балансы речных водосборов, колебания и изменения их элементов.

***Шведовский Петр Владимирович***

1946 года рождения, окончил Белорусскую сельскохозяйственную академию в 1967 г. по специальности «гидротехническое строительство».

Кандидат технических наук, профессор. Профессор Брестского государственного технического университета.

Автор более 500 научных работ, в т. ч. 7 монографий, 7 учебных пособий, 108 изобретений и патентов.

Основное направление работы – экологические проблемы, агротрансформация ландшафтов и их моделирование.

Научное издание

**Логинов Владимир Федорович**  
**Волчек Александр Александрович**  
**Шведовский Петр Владимирович**

**ПРАКТИКА ПРИМЕНЕНИЯ СТАТИСТИЧЕСКИХ МЕТОДОВ  
ПРИ АНАЛИЗЕ И ПРОГНОЗЕ ПРИРОДНЫХ ПРОЦЕССОВ**

ISBN 985-6584-93-0



Ответственный за выпуск: П. В. Шведовский  
Компьютерный набор: Н.Н. Шпендик, Н. Г. Кулик  
Корректор: Никигчик Е. В.  
Редактор: Строкач Т. В.

Лицензия №02330/0133017 от 30.04.2004 г.

Подписано в печать 1.11.2004 г. Формат 60x84 <sup>1</sup>/<sub>16</sub>. Бумага писчая.  
Гарнитура Таймс. Печать офсетная. Усл. печ. л. 17,7. Уч. изд. л. 19,0.  
Тираж 200 экз. Заказ № 1090.

Отпечатано с оригинал-макета заказчика на ризографе учреждения  
образования «Брестский государственный технический университет».

Лицензия №02330/0148711 от 30.04.2004 г.