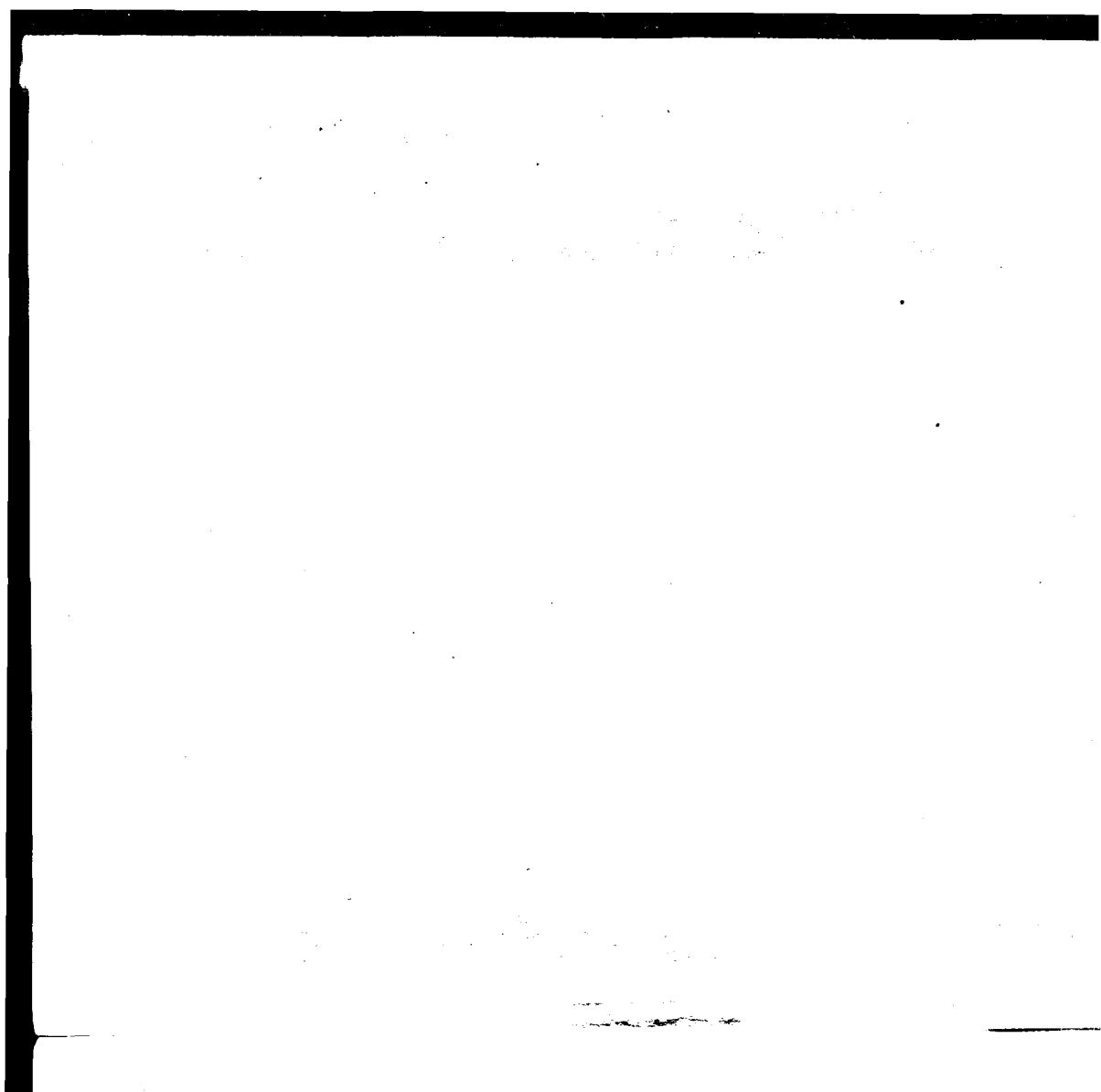
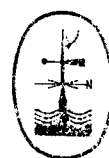




Ч.П.СМИРНОВ, П.А.ВАЙНОВСКИЙ, Ю.Э.ТИТОВ

**СТАТИСТИЧЕСКИЙ
ДИАГНОЗ
И ПРОГНОЗ
ОКЕАНОЛОГИЧЕСКИХ
ПРОЦЕССОВ**



САНКТ-ПЕТЕРБУРГ ГИДРОМЕТЕОИЗДАТ 1992

УДК 551.46

Рецензент: д-р геогр. наук В. Р. Фукс (Санкт-Петербургский государственный университет)

Научный редактор: д-р физ.-мат. наук В. А. Рожков (Санкт-Петербургское отделение Государственного океанографического института)

Крупномасштабная изменчивость океанологических характеристик в пространстве и во времени рассматривается как многомерный случайный процесс, количественное описание (диагноз) и прогнозирование которого осуществляется на основе применения общей методологии многомерного статистического анализа. Освещается широкий круг вопросов, связанных с развитием данного подхода в прикладных океанологических исследованиях: обобщаются современное состояние теории методов и накопленный опыт их практического применения, оцениваются возможные пути оптимизации решения типовых океанологических задач методами многомерной статистики.

Книга предназначена для научных работников и инженеров, специализирующихся в области прикладного статистического анализа океанологической информации.

Large-scale variations of oceanological characteristics in space and time are considered like upon a multi-dimentional stochastie process in the monograph.

Quantaty desorption and prediction of this process is made on the base of using a general methodology of multi-variable statistical analysis.

Up to date theory of methods and experience of using these ones in practic are generalized and summarized. Possible ways of optimization for solving typical oceanological tasks by multi-dimensional statistical methods are assessed and illustrated by examples of different oceanological process.

The book is intended for scientists, engineers which are specializing in the field of applied statistical analysis of oceanological information.

С 1805040600-072 30-92
069(02)-92

ISBN 5-286-00315-5

© Н. П. Смирнов, П. А. Вайновский,
Ю. Э. Титов, 1992

ПРЕДИСЛОВИЕ

Настоящая книга задумывалась как продолжение серии монографий по проблемам статистической океанологии, выпускаемых Гидрометеиздатом. Наличие ряда серьезных работ, посвященных первичному статистическому анализу океанологических наблюдений [27, 87], корреляционно-спектральному анализу временных рядов [69, 151], в которых глубоко проработаны вопросы применения этих методов в океанологических исследованиях, позволило авторам сразу сконцентрировать свое внимание на обсуждении методологии статистического исследования многомерных океанологических процессов.

Широко распространяющийся в последнее десятилетие взгляд на океан как на сложную динамическую систему заставляет по-новому решать задачи наблюдения за ним и его описания. Понимание многопараметрического многообразия изучаемых явлений предопределило выработку стратегии комплексных натурных экспериментов, в результате проведения которых аналитики получают огромное количество потенциально связанных между собой физико-химических, биологических параметров состояния океана. Многомерность и стохастичность исследуемых процессов, неоднозначность их отображения через изменчивость регистрируемых параметров, в свою очередь, потребовало применения соответствующих методов многомерного статистического анализа (МСА), которые позволяют количественно описать динамику процессов в терминах случайных функций и полей.

Общая теория методов МСА достаточно глубоко разработана, на сегодняшний день накоплен значительный положительный опыт их практического применения (в медицине, социологии, экономике). В океанологии процесс накопления опыта применения МСА отличается известным своеобразием. Основополагающие принципы применения многомерного анализа в отечественной гидрометеорологии были заложены в 60-е—70-е годы работами А. М. Обухова, Н. А. Багрова, Г. В. Грузы, Л. С. Гандина и др. В океанологических исследованиях широкое распространение получили идеи ленинградской школы многомерной статистики, успехи которой в изучении крупномасштабных океанологических процессов связывают с именем Ю. В. Николаева.

Вспышка популярности применения методов МСА в 60-е годы сменилась затем весьма осторожным к ним отношением. Надежда на всеобъемлемость и общедоступность методологии МСА оправ-

дались далеко не полностью. В последнее десятилетие наметился отрыв теории многомерной статистики от практики ее применения в океанологии. Теория интенсивно развивалась в направлении оптимального решения как классических, так и неклассических статистических задач (анализ нестационарных, неоднородных данных). Исследователи-океанологи, однако, не торопились внедрять новые методы МСА в свою практическую деятельность. И если раньше это можно было оправдать отсутствием необходимой вычислительной техники, то сейчас такой тезис уже все больше выходит из моды. Число ЭВМ повсеместно растет, а психология исследователя в применении мощного аппарата многомерной статистики меняется пока значительно медленнее.

Отмеченная тенденция объективно отражает реальный уровень статистического диагноза и прогноза многомерных океанологических процессов, который определяется тем, что широкий круг океанологов явно недостаточно знаком с методологией МСА. Поэтому авторы, давая более или менее общий обзор прикладного многомерного анализа океанологических процессов, намеренно уделяли особое внимание разъяснению узловых моментов, наиболее актуальных сегодня для повышения эффективности применения методов МСА в практической океанологии. С этой точки зрения в работе раскрываются три основных вопроса:

- что собой представляет в настоящее время методология многомерного статистического анализа;
- как, когда и в каких целях следует применять МСА для диагноза и прогноза океанологических процессов;
- каких результатов следует ожидать от применения методов МСА.

Общая классификация решаемых прикладных задач и статистических методов с учетом многомерности океанологической информации приводится во введении. Здесь же изложена рабочая концепция статистического диагноза и прогноза океанологических процессов и на основе общих принципов моделирования случайных процессов сформулирована логическая последовательность применения методов.

В главах 1—3 последовательно и систематизированно излагаются основы теории и вычислительной практики методов многомерной статистики. Объединение методов по главам выполнено в соответствии с прикладной классификацией методов и определяется функциональным назначением последних при решении океанологических задач. Изложение каждого метода завершается обзором литературных источников по вопросам его применения в океанологии. В главе 4 рассматриваются статистико-вероятностные модели прогнозирования океанологических процессов. Обзор широкого спектра разнообразных моделей завершается анализом возможностей статистической комплексации нескольких альтернативных прогнозов и выработки оптимальной прогностической модели. Диагностированная структура исходных данных определяет границы применения конкретных статистических про-

гностических моделей. С этих позиций обсуждается проблема статистической предсказуемости исследуемых океанологических процессов.

В главе 5 обобщены результаты статистического диагноза некоторых крупномасштабных океанологических процессов, на примере которых демонстрируются практические возможности применения обсуждаемой методологии. Сочетание нескольких методов диагноза, взаимное сопоставление их результатов, при необходимости повтор расчетов с измененными начальными условиями — все это является основой построения удачной обобщенной модели процесса в целом.

В главе 6 собраны результаты прогнозирования океанологических процессов на основе результатов диагноза. Сравнительный анализ эффективности различных прогностических моделей здесь сочетается с оценением применимости двух различных подходов — прогноза исходных наблюдаемых переменных и прогноза диагностированных обобщенных параметров различных океанологических процессов.

Хотя вопросы прогнозирования многомерных океанологических процессов выделены в отдельную главу, читатель может заметить, что общий центр тяжести книги несколько смещен в сторону диагноза. Такое соотношение, по нашему мнению, отражает реальную картину в прикладной океанологии. Судя по имеющимся публикациям, вряд ли можно сделать обоснованный вывод о близости окончательного решения проблемы статистического прогнозирования многомерных океанологических процессов. Скорее это тема специального исследования, для которого сегодня сформулированы только некоторые отправные гипотезы и намечены потенциально эффективные пути проработки решения.

В заключении кратко обобщаются важнейшие выводы, полученные в предшествующих главах, на основании сопоставления которых делается попытка оценить ближайшую перспективу в проблеме статистического диагноза и прогноза многомерных океанологических процессов.

В процессе работы над книгой удалось собрать воедино значительное число публикаций по рассматриваемой тематике. Стремясь сохранить целостность изложения материала, авторам нередко приходилось отказываться от детального обсуждения приводимых в них результатов и ограничиваться ссылками. Тем не менее список литературы можно рекомендовать к использованию как прикладной библиографический указатель по применению методов МСА в океанологии, который будет полезен при проработке частных методологических вопросов исследования заданных процессов в конкретных географических регионах.

Возможности применения МСА в океанологических прикладных исследованиях никоим образом не исчерпываются рамками материала, изложенного в настоящей работе. Некоторые методы, такие, например, как многомерный спектральный анализ, многомерный дисперсионный анализ, авторы сознательно не рассматри-

вали в силу их специфичности. Обсуждение теоретических вопросов и анализ результатов расчетов ограничивались рамками линейных статистических моделей, тогда как вопросы нелинейного взаимодействия не рассматривались. Не прорабатывалась в полном объеме и тематика исследования векторных океанологических процессов. Перечисленные ограничения в принципе не должны сильно повлиять на логику излагаемых вопросов в целом. Дискуссионность отдельных выдвигаемых положений и гипотез должна, на наш взгляд, послужить скорейшей выработке общей точки зрения по этой перспективной проблеме анализа данных. Поэтому все замечания, дополнения и предложения будут с благодарностью приняты.

Авторы выражают глубокую признательность научному редактору В. А. Рожкову и рецензенту В. Р. Фуксу, усилиями которых логика рукописи во многом обрела окончательную форму.

ВВЕДЕНИЕ

Многомерность и стохастичность проявлений океанической изменчивости не позволяют сразу однозначно проследить, количественно описать и физически интерпретировать механизмы колебаний. Методы статистического анализа служат инструментом выявления и описания подобных взаимосвязей для решения задач диагноза и прогноза океанологических процессов. Точность и надежность получаемых при этом выводов во многом определяется количеством и разнообразием исходной обрабатываемой информации.

Вся информационная база накапливается в трех плоскостях: пространство — время $X(l, t)$, пространство — признаки $X(l, z)$, признаки — время $X(z, t)$, которые в совокупности образуют куб данных с осями: пространство (l), время (t), переменные (z) (рис. В.1). Как же располагаются данные наблюдений в таком кубе? Единичное наблюдение x за единственной переменной в единственной точке пространства в заданный момент времени представляет собой единичную ячейку куба данных. Последовательность из N наблюдений $\{x_1, x_2, \dots, x_N\}$ является вектором, располагающимся вдоль соответствующей оси куба.

Например, регистрация температуры воды в одной точке в моменты времени $t_1, t_2, \dots, t_N$ приводит к получению вектора наблюдений $\mathbf{X}(t)$, который в кубе данных будет параллелен оси времени. Единичное наблюдение вертикального распределения температуры воды приводит к получению вектора данных, параллельного оси пространства: $\mathbf{X}(l)$. Последовательное во времени многократное зондирование вертикального распределения температуры воды в одной точке приведет к получению плоскости данных, располагающейся в кубе параллельно плоскости признаков: $\mathbf{X}(t, l)$. Если же совместно с температурой многократно наблюдать другие параметры состояния океана (соленость, скорость течений, химические компоненты среды), то в результате получается куб данных, заполненный во всех трех выбранных координатных осях: $\mathbf{X}(t, l, z)$.

В методах одномерного анализа данных рассматривается последовательность наблюдений во времени или в пространстве и изучается характер изменчивости (периодический, стохастический, ...), однородность колебаний, связность вектора наблюдений [104, 123, 211]. При этом из рассмотрения практически выпадает плоскость наблюдения за различными параметрами. И при

всей своей эффективности методы одномерной статистики не позволяют с необходимой полнотой проанализировать сложный многопараметрический процесс.

Методы парного анализа рядов (корреляционный, регрессионный, спектральный) преследуют цель статистического описания

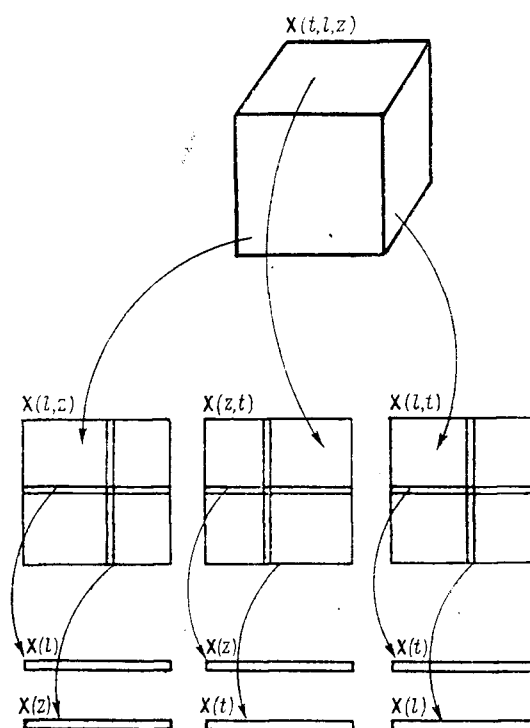


Рис. В.1. Куб океанологических данных: составляющие его плоскости и векторы наблюдений.

взаимосвязи двух векторов наблюдения во времени или в пространстве: $X_1(t)$, $X_2(t)$; $X_1(l)$, $X_2(l)$. Здесь анализируется как каждый вектор в отдельности, так и степень их парной синхронной и асинхронной взаимосвязи, что позволяет уже более полно по сравнению с одномерным анализом изучить статистическую структуру процесса.

И лишь в методах многомерного статистического анализа (МСА) появляется реальная возможность исследовать совместно множество векторов наблюдений. Методы МСА направлены на решение задач статистического анализа матриц из N наблюдений (объектов) за M переменными (признаками). В качестве матриц здесь выступают числовые таблицы, полученные в результате среза куба данных вдоль любой из граней. Это могут быть мат-

рицы наблюдений в плоскостях переменные — время $X(z, t)$, переменные — пространство $X(z, l)$ или время — пространство $X(t, l)$. Равноправное включение в анализ оси признаков является несомненным достоинством МСА, которое позволяет исследовать куб данных в целом. Совместные распределения наблюдаемых характеристик анализируются в терминах векторов и матриц, с помощью которых методами МСА дается количественное описание скрытых взаимосвязей исходных параметров и раскрывается содержание моделей диагноза и прогноза исследуемых многомерных океанологических процессов.

Понятие диагноза

Как указывает Словарь иностранных слов, диагноз — это процесс определения существа болезни и состояния больного на основании его всестороннего обследования. В применении к практической океанологии существо термина меняется незначительно. Не будучи сформулировано соответствующим ГОСТом, оно может быть определено следующим образом.

Диагноз — это последовательность процедур обобщения, систематизации и интерпретации наблюдаемой изменчивости океанологических характеристик с целью формулировки достоверной модели происходящих процессов.

Диагноз предполагает два основных этапа: анализ и синтез. Анализ, смысл которого состоит в разложении объекта исследований на составные части, является начальным этапом исследований. В него могут быть включены многочисленные методы преобразования данных наблюдений и расчетов для описания составных частей и отдельных явлений процессов. Основное содержание любого аналитического метода — физического, эмпирического, статистического или численного — состоит в том, чтобы выявить и описать составные элементы происходящих процессов.

Синтез представляет собой заключительный этап диагноза и направлен на объединение выявленных элементарных составных частей или простейших свойств в некое единое целое. Основной принцип подобного объединения состоит в логическом построении законченной картины происходящих событий на основе выявленных на этапе анализа закономерностей и взаимосвязей элементов. По сути синтез включает многие из методов анализа, но применяемые в иной содержательной и целевой функции — для объединения и систематизации уже известных составных частей. Кроме того, синтез включает собственно интерпретацию выстроенной системы закономерностей. Интерпретация понимается непосредственно как физическое объяснение выявленных закономерностей, как формулировка общего физического механизма возникновения и развития происходящих в океане явлений.

В понятие диагноза авторы настоящей работы вкладывают всю совокупность процедур преобразования разнородной океанологической информации в процессе изучения функционирования си-

стемы «океан» в прошлом или настоящем. Поэтому в диагноз в качестве составных частей входят два основных типа диагностических процедур: физический и статистический диагноз. Эти понятия включают в себя соответствующие методы исследования, которые различаются по технологическому или вычислительному признаку.

К физическому диагнозу отнесены процедуры и методы определения состояния океана, построенные на основе описания океанологических явлений и процессов путем расчета и анализа разнообразных гидрофизических характеристик. По своему содержанию физический диагноз традиционен. Он включает в себя прежде всего типично океанологические методики вычислений непосредственно ненаблюдаемых, а потому вторичных характеристик состояния океана. Это огромный класс параметров, характеризующих физические, динамические, энергетические поля. Перечисление подобных характеристик заняло бы много места. К счастью, большинство из них представлены в океанографических таблицах, что не только облегчает их оценку, но и демонстрирует основной принцип отнесения океанологических характеристик к данному классу: все эти параметры функционально через посредство известных физических законов связаны с другими первичными характеристиками и образуют набор переменных, однозначно характеризующих конкретные виды океанологических процессов.

Смысл физического диагноза и состоит в том, чтобы охарактеризовать происходящие в океане изменения через систему известных и принятых в океанологии терминов и категорий. К последним следует прежде всего отнести элементы теплового баланса (энтальпию, потоки тепла), элементы горизонтальной и вертикальной структуры вод (водные массы, фронты и переходные зоны, круговороты и системы циркуляции) и т. п. Применительно к конкретным видам терминов сложилась специфическая методология анализа.

Как правило, методология физического анализа представляет собой своеобразный конгломерат, состоящий из эмпирических приемов количественного анализа в сочетании с практическими методами качественной обработки и описания. Цементирует подобное объединение многолетний опыт практических исследований, который зачастую является не результатом направленных систематических исследований океана, а следствием обобщения частных интересов практической деятельности.

Так, например, методология T, S -анализа водных масс [23, 125] представляется фундаментальным достижением науки, которое позволяет понять и описать важнейшие процессы формирования структуры вод. В то же время методика описания ледовых условий или состояния поверхности моря по балльной шкале представляется результатом реализации чисто прикладных требований мореплавания.

Большинство методов физического анализа океанологических процессов сочетает основы количественного анализа наблюдений

с разнообразными эвристическими приемами сбора, подготовки и обработки океанологической информации. Эвристические правила являются одновременно базой для последующей интерпретации количественных оценок в океанологических терминах.

В целом практика физического анализа океанологических процессов представляется сложившейся, законченной и весьма эффективной методологией обработки океанологической информации.

Принципы статистического диагноза

Статистический диагноз предполагает применение статистических методов для обработки, анализа и интерпретации океанологической информации во всем ее многообразии. Методология статистического диагноза базируется на разнообразных методах статистического анализа и представления данных.

Статистический диагноз потенциально имеет самые широкие возможности практического применения в океанологических исследованиях. Эта перспектива обусловлена большим разнообразием статистических методов преобразования информации. Однако применение статистического диагноза на практике пока весьма ограничено. Причин тому несколько.

Во-первых, очевидно, нет глубокого понимания в среде исследователей-прикладников самой системы статистических вычислений, а также внутренней взаимообусловленности различных статистических методов. Это по сути и сдерживает использование всего многообразия методик для формулирования и проверки статистических гипотез, принятия статистически обоснованных решений всестороннего количественного исследования взаимосвязей океанологических характеристик. Думается, что именно недостаток «прикладного понимания» сущности статистической методологии порождает определенное недоверие к ней. В результате диагноз сводится к практическому применению лишь отдельных методологических приемов, которые, хотя и не являются оптимальными для решения конкретных задач, зато представляют собой отлаженную, устоявшуюся, исторически сложившуюся схему, не грозящую исследователям какими-либо неожиданными проблемами.

Во-вторых, применение современных методов статистического анализа должно быть подкреплено соответствующей технической базой. Характеристики имеющихся ЭВМ порой однозначно определяют границы статистического анализа информации на практике. В отличие от гидродинамических расчетов статистика требует не столько быстродействия ЭВМ, сколько большого объема памяти. Оперативная и внешняя память однозначно определяют возможности применения многих эффективных подходов, в том числе методов многомерного статистического анализа.

Наконец, третья причина заключается в крайне медленном накоплении океанологической информации на технических носителях ЭВМ. Как известно, эффективность обработки больших мас-

сивов данных во многом зависит от уровня организации информации, чем определяется доступность натуральных материалов для обработки. Достоверность самих статистических выводов однозначно связана с длиной исследуемых выборок. Поэтому накопление информации и ее организованное хранение можно рассматривать как прямой путь к развитию статистических методов диагноза.

Сущность статистического диагноза сводится к количественному анализу множества океанологических характеристик, по-

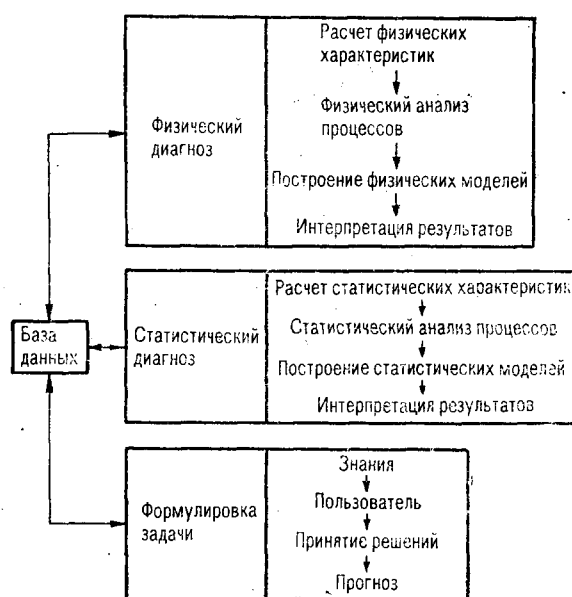


Рис. В.2. Функциональная схема диагноза океанологических процессов.

строению оптимальных статистических моделей наблюдаемых процессов, дальнейшему описанию и интерпретации закономерностей и физических механизмов формирования изменчивости океана. Наибольшую популярность методология статистического анализа получила при первичном анализе рядов наблюдений (построение гистограмм, оценка первичных статистических моментов), описании взаимосвязей рядов (парная корреляция и регрессия), вскрытии закономерностей временной изменчивости океанологических характеристик (гармонический, спектральный, взаимоспектральный анализ). В то же время методы многомерного анализа в целом используются пока значительно слабее, что никак не соответствует реальной мощности этих методов.

Попытаемся обобщить приводимые рассуждения о физическом и статистическом диагнозе в виде простой наглядной схемы, которая суммировала бы методические идеи различных океанологических направлений (рис. В.2).

Блок статистического диагноза оказывается непосредственно связанным с тремя узлами системы обработки: базами данных, физическим диагнозом и формированием знаний пользователя. В соответствии с этими взаимосвязями можно сформулировать основные задачи статистического диагноза, которые условно разделяются на следующие три группы.

1. Задачи подготовительной обработки данных, которые включают процедуры формирования требуемых массивов исходной информации из баз исходных данных и на основе результатов вычислений вторичных океанологических показателей в блоке физического диагноза.

2. Задачи статистического анализа информации, представленные в виде набора последовательно решаемых конкретных частных вопросов оптимального статистического анализа материала в соответствии с выбранной последовательностью разложения информации. В частности, представлены задачи выявления периодических составляющих временной и пространственной изменчивости, количественной оценки связности колебаний океанологических характеристик в пространстве или во времени.

3. Задачи этапа синтеза, когда полученные количественные оценки статистических взаимосвязей надо объединить в виде статистически достоверных моделей происходящих явлений, а затем физически обоснованно интерпретировать механизмы функционирования полученных моделей в океанологических терминах и в соответствии с принятыми при анализе физическими постулатами.

Успешность функционирования представленной блок-схемы определяется балансом между уровнем знания пользователя по результатам диагноза и эффективностью принимаемых на основе знания практических решений. Отрицательная разность этих двух составляющих формулирует требуемые задачи дальнейшего исследования. Положительная разность предопределяет возможность надежного дальнейшего использования знания, например, в виде прогнозов развития океанологических процессов во времени или в пространстве.

Успешное решение задач статистического диагноза океанологических процессов во многом предопределяется эффективностью применения конкретных методов статистической обработки информации. Формально блок статистического диагноза представляет собой последовательность нескольких методов статистики, выбор и конкретизация взаимодействия которых осуществляются в соответствии с общими целями исследователя. Тем не менее следует выделить несколько основополагающих принципов подбора и расположения статистических методов в общей цепочке диагноза океанологических процессов.

Принцип дополнения. Согласно этому принципу, подбираемые методики должны приводить к таким результатам, которые позволили бы построить наиболее полную картину происходящих явлений. Тем самым подчеркивается необходимость исключения заранее предопределенного многократного дублирования получаемых

результатов. Вряд ли необходимо классифицировать выборку, для которой предварительно подтверждены все критерии ее однородности. Однако можно объяснить реальную необходимость многократной классификации выборки по различным метрикам сходства и правилам классификации для того, чтобы выбрать затем наиболее устойчивую и физически обоснованную структуру классов. Смысл данного правила заключается в том, чтобы последующий результат расширял наши представления об объекте исследования, которые получены в итоге предшествующего диагноза.

Принцип сочетаемости методов вводится для того, чтобы подчеркнуть необходимость соблюдения последовательного усложнения применяемых методов исследования. Более простые методы статистической обработки позволяют получить весьма грубый, приблизительный результат, но зато с минимальными затратами. Более точный метод чаще всего более сложен и менее экономичен. Как правило, усложнение вычислений и увеличение точности результата повсеместно сопровождается резким сужением рамок применения методики. Высокая надежность статистической оценки — во многом следствие строгого соблюдения всего набора жестких ограничений при расчетах. Иллюстрацией сказанного может быть известная в гидрометеорологии проблема средних многолетних норм температуры воды. Приблизительная оценка среднего арифметического по длительному ряду наблюдений оказывается ничем не хуже тонких оценок математического ожидания, выполняемых по различным более сложным методам в приближении нестационарности конкретного вида.

Принцип преемственности при своей идеальной реализации должен позволить исследователю добиться максимальной достоверности результатов вычислений при минимизации риска. Он состоит в возможности циклически последовательной расстановки методов статистического диагноза океанологических процессов. Под циклом понимается группа методов статистической обработки, формирующих замкнутый круг вычислений, эффективность которых контролируется пользователем. Например, при расчете главных компонент может быть сформирован цикл вычислений каждой компоненты с восстановлением ее значений. Исследователь, контролируя сходимость разложения и интерпретируемость компонент, определяет необходимость дальнейшего разложения либо завершение этих процедур. Организация вычисления по принципу цикличности должна обеспечить в процессе диагноза достижение необходимого баланса между точностью исходной информации, эффективностью вычислительных алгоритмов диагноза и уровнем физической интерпретации получаемых результатов вычислений.

Последовательность статистического диагноза

Представляется возможным реализовать на практике перечисленные наиболее общие принципы статистического диагноза в виде

конкретных схем статистических вычислений. Остановимся на одном из возможных вариантов, который оказался весьма эффективным при исследовании крупномасштабных океанологических процессов.

Предлагаемая схема статистического диагноза (рис. В.3) формировалась на основе рассматриваемых выше принципов как со-

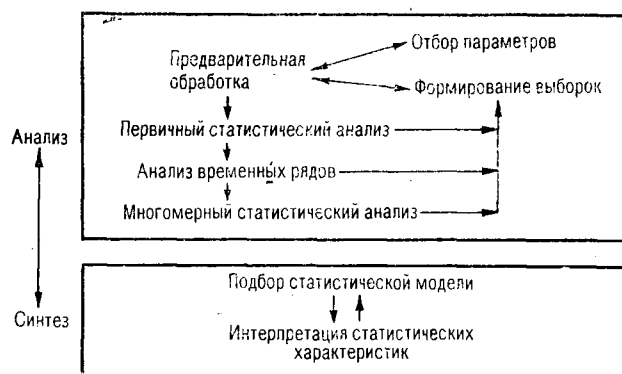


Рис. В.3. Функциональная схема статистического диагноза океанологических процессов.

ставная часть общего процесса обработки океанологической информации. Этап анализа представлен четырьмя основными группами методов обработки. Предварительная обработка обособлена как самостоятельный блок, чтобы подчеркнуть важность и необходимость тщательного начального выбора параметров статистического анализа.

Очевидно, что конкретный набор переменных следует формировать исходя из условий постановки задачи диагноза. В результате физического диагноза уже появляется возможность на качественном уровне оценить перспективность включения дополнительных параметров в список исследуемых взаимосвязей. В данный список уже обоснованно включены те переменные, которые непосредственно не наблюдаются, но могут быть рассчитаны в процессе физического диагноза.

Первичный статистический анализ в статистическом диагнозе может рассматриваться и как самоцель исследования, и как промежуточная ступень вычислений. В любом случае его результат представляется фундаментом достоверной интерпретации устойчивости и изменчивости исследуемых параметров, однородности и стационарности отдельных составляющих исследуемых процессов.

Следует выделить отдельно методы оценки значений первичных статистических моментов и алгоритмы вычисления одно- и двумерных эмпирических гистограмм как процедуры расчета важных эмпирических характеристик исследуемых океанологических процессов. Получаемые статистические оценки — среднее, дисперсия, коэффициенты вариации, асимметрии, эксцесса, параметры

распределения вероятностей представляют собой важную информацию о внутренней структуре временной или пространственной изменчивости процессов. На их основе формируется начальное представление о фоновой изменчивости исследуемых явлений в целом.

Анализ временных рядов выделен в отдельный блок, поскольку в нем объединены в логически взаимосвязанную цепочку вычислений алгоритмы количественной оценки степени закономерности изменения океанологических характеристик во времени. Идеология временного анализа наблюдений подробно и всесторонне рассмотрена в целом ряде монографий, посвященных проблемам вероятностного анализа океанологических процессов [69, 151, 193]. Здесь важно подчеркнуть смысловую нагрузку, вкладываемую в эту методику при комплексном статистическом анализе многомерных океанологических процессов.

Выделить закономерность изменения временного ряда — значит получить достоверные оценки характеристик его случайности или закономерности изменения во времени. Очевидно, что выявление закономерностей представляется наиболее интенсивным результатом временных рядов. Под закономерностью понимается прежде всего скрытая ритмика колебаний океанологических характеристик, которая интерпретируется как проявление закономерных механизмов формирования исследуемых процессов внешними или внутренними факторами. Поэтому наиболее интенсивное развитие получили методы гармонического анализа Фурье, периодограммного анализа, спектрального анализа.

Океан — сложная система с иерархической структурой взаимосвязей, которая обладает целой гаммой характерных собственных и вынужденных частот колебаний. Именно структурированность системы и позволяет, приблизительно оценив частотную структуру исследуемых рядов, практически сразу отнести изучаемые явления к определенной ступени в иерархии изменчивости океана. Опыт, однако, показывает, что набор частот характерных колебаний ограничен и практически полностью описан в диапазоне изменчивости от минут до тысячелетий с достаточной для современных практических и теоретических запросов точностью и надежностью.

Что же дальше? Оказалось, куда большим набором степеней свободы обладает частотная структура взаимодействий отдельных составляющих климатической системы. Описание взаимосвязей колебаний океанологических характеристик в частотной области выполняется методами взаимоспектрального анализа, который в оценках спектров синхронных и асинхронных колебаний, когерентности и фазы колебаний позволяет получить представление о наличии ритмических взаимодействий двух временных рядов.

Еще одна группа проблем анализа временных рядов связана с описанием долгопериодных непериодических колебаний в исходных процессах, которые интерпретируются термином «тренд». Проблему описания тренда вряд ли следует считать тривиальной,

и, по-видимому, она еще далека от своего решения.

Традиционно тренд вызывает своеобразное недовольство исследователей как назойливая помеха в работе. Причина тому — нарушение условий стационарности временных рядов, что резко снижает эффективность классических методов корреляционного и спектрального анализа данных. Действительно, нестационарность по математическому ожиданию, по амплитудно-фазовым характеристикам колебаний снижает достоверность оценок детерминированных составляющих, которые в большинстве случаев на практике получаются по алгоритмам расчетов для стационарных выборок. Именно поэтому превалирует стремление всеми доступными средствами избавиться от трендов в исследуемых последовательностях.

Осуществляется такое желание различными методами, в основе которых лежит по сути простая идея раздельного описания различных колебаний во временных рядах. И трендовая изменчивость ($A(t)$) при таком подходе трактуется как самостоятельная изменчивость, непосредственно не связанная с более короткопериодной ($B(t)$). Результат состоит в описании общей изменчивости $X(t)$ океанологических процессов в виде аддитивной модели:

$$X(t) = A(t) + B(t).$$

Одной из важнейших черт всех предшествующих методов анализа океанологической информации следует признать почти экспоненциальный рост числа расчетных переменных, характеризующих ту или иную сторону исследуемых процессов.

Так, всего к двум исходным временным рядам — температуры и солености — уже на первом этапе предварительной обработки может быть добавлен еще один — плотность. В результате первичного анализа исследователь получит еще десяток количественных характеристик. Наконец, после анализа скрытой структуры колебаний рабочая информация увеличивается еще в несколько раз за счет оценок авто- и взаимокорреляционных функций, авто- и взаимных спектральных функций, параметров разложения Фурье, которые при аддитивной записи модели исходных процессов формируются раздельно для трендовой и короткопериодной компонент. Фактически наблюдается уже избыток информации, систематизация которой требует специфических методов, а именно методов многомерного статистического анализа.

Прикладная классификация методов МСА

Основная цель подавляющего большинства методов МСА состоит в таком преобразовании исходных данных, которое позволило бы исследователю наглядно и достаточно полно представить себе сущность ведущих взаимосвязей в анализируемом материале. Разнообразие возможных теоретических подходов обусло-

вило возникновение и развитие целой совокупности методов МСА, каждый из которых по-своему интерпретирует изменчивость данных. К числу наиболее распространенных методов МСА следует отнести [12, 94, 95, 177]:

- 1) регрессионный анализ,
- 2) компонентный и факторный анализ,
- 3) канонический корреляционный анализ,
- 4) классификацию или кластер-анализ,
- 5) дискриминантный анализ.

На фундаментальной основе этих методов реализовано около сотни различных модификаций. Задача исследователя состоит в том, чтобы, хорошо ориентируясь в методах, выбрать и использовать наиболее эффективные. Применимость МСА на практике определяется и проверяется точностью соответствия построенной статистической модели натурным данным, а кроме того, физической интерпретируемостью полученных в модели взаимозависимостей. Низкий уровень результатов приложения методов МСА говорит не столько о слабости самого метода, сколько о несоответствии метода исходным данным в поставленной задаче или о неграмотном его использовании. Успех достигается только при осознанном, обоснованном и умелом применении этих методов на практике. Соответствие методов МСА решаемым задачам представляет поэтому особый интерес. В самом общем виде этот вопрос рассмотрен в табл. В.1, которая демонстрирует лишь наиболее отработанные в практической океанологии способы решения типовых задач методами МСА. Все множество целей прикладных океанологических исследований здесь разбито на три основные

Таблица В.1

Прикладная классификация методов МСА

Цели	Принципы	Методы
Описание взаимосвязей (моделирование, прогноз)	Количественное описание взаимосвязей признаков на основе корреляции и регрессии	Множественная регрессия (линейная, нелинейная). Канонический корреляционный анализ
Выявление и описание скрытой структуры данных	Преобразование признаков к ортогональным составляющим	Разложения на ортогональные алгебраические полиномы (Чебышева, Фурье). Метод главных компонент. Факторный анализ
Выделение подмножеств объектов наблюдений; типизация, классификация, районирование	Поиск естественного разделения облака наблюдений на однородные классы, оценка принадлежности конкретных наблюдений к выделенным классам	Кластер-анализ (автоматическая классификация). Дискриминантный анализ

группы, для каждой из которых определены принципы и конкретные методы решения подобных общих задач.

Проблема статистического описания взаимосвязей с целью построения моделей явлений и процессов для диагноза и прогноза поставлена на первом месте. Это не случайно, поскольку она остается главным итогом большинства решаемых прикладных задач. Кроме того, данную проблему приходится решать и в процессе других исследований, связанных с классификацией процессов или анализом их структуры. Традиционный способ статистического описания в рассматриваемом аспекте основан на применении методов множественного корреляционного и регрессионного анализа, посредством которых исследователь описывает взаимозависимости между исследуемыми признаками.

Априорно накоплена большая информация о потенциальном наличии взаимосвязей, а здесь требуется уже облечь качественные выводы в математическую форму — в уравнения множественной регрессии. При этом множественный регрессионный анализ представляется ведущим подходом в решении задач описания. В океанологических приложениях этой методики получен целый ряд важных научных и практических результатов. Так, применение множественной регрессии позволяет описать зависимость плотности от температуры, солености, давления в эмпирических уравнениях состояния морской воды. Эта же методика используется при составлении расчетных и прогностических формул для вычисления ледовых условий в зависимости от метеорологических характеристик. Аналогичный подход развит и при исследовании параметров волнения в зависимости от скорости ветра, длины разгона и т. п.

Нахождение скрытой структуры в гидрометеорологических данных подразумевает выявление основных закономерностей пространственно-временных переменных. В океанологии эта проблема решается в трех основных аспектах: как анализ структуры взаимосвязей между параметрами состояния океана, как анализ структуры пространственных масштабов изменчивости океанологических характеристик и как анализ структуры временной изменчивости процессов.

Главный принцип реализации решаемых задач состоит в разложении исходной выборки на ряд ортогональных (взаимонезависимых) составляющих. Способ разложения выбирается так, чтобы каждая ортогональная составляющая описывала какую-то характерную взаимосвязь переменных или обобщала типовой масштаб пространственно-временной изменчивости.

Наиболее простыми способами разложения следует признать процедуры разложения по алгебраическим полиномам, среди которых популярность заслужили полиномы Чебышева и Фурье. Метод главных компонент также получил широкое распространение, хотя он представляется несколько более сложным. Метод факторного анализа пока еще мало применялся в практической океанологии, однако в сопредельных областях знания (метеоро-

логия, биология, социология, экономика) он зарекомендовал себя весьма хорошо. За эффективность факторного решения нередко приходится платить сложностью вычислительных процедур.

Процесс научной систематизации накопленных знаний неразрывно связывается с классификацией, типизацией объектов исследования. Особой прикладной и теоретической значимостью обладают проблемы классификации водных масс, классификации форм льда, классификации вертикальной структуры вод и т. п.

Класс (тип) явления или процесса представляется важным результатом обобщения накопленной информации об океане, поскольку в нем отражены наиболее общие закономерные черты изменчивости океана и характерные особенности взаимосвязей простейших элементов. В океанологических исследованиях распространены типизации временного хода процессов, типизации пространственной структуры полей, типизации взаимосвязей признаков.

Не менее важной следует считать задачу отнесения новой информации к одному из ранее выделенных классов состояния океана. Подобный тип задач решается, как правило, применением метода дискриминантного анализа, в котором на основе знания статистических свойств ранее выделенных классов удается с минимальной ошибкой определить принадлежность новой информации к одному из них. На практике дискриминантный анализ эффективно применялся при распознавании типа вертикальной структуры профиля температуры воды, при диагнозе и прогнозе типа ледовых условий.

Последовательность статистического диагноза многомерных процессов

Вся логика рассматриваемых методов МСА подчинена описанию многопараметрических систем. Центральное место в МСА, несомненно, занимают две группы методов — классификационные методы и методы структурного анализа. Определение их соотношения в рамках отдельного исследования является, по-видимому, основным вопросом определения необходимости применения МСА вообще.

Приемлемы два варианта последовательности МСА в практике океанологических исследований. На рис. В.4 приведены детальные схемы применения МСА. Очевидно, что в обоих случаях конечная цель статистической обработки сохраняется неизменной — добиться наиболее достоверного и надежного описания существующей изменчивости океанологических процессов. Однако достижение цели происходит различными путями, которые определяются структурой исходных данных, выявляемой на предыдущих этапах статистической обработки океанологических параметров.

Перед каждым исследователем стоит проблема выбора последовательности применения методов МСА и вопрос принципиальной необходимости использования конкретных методов. Решить эти задачи потенциально возможно не на интуитивном, а на

вполне осмысленном уровне. Допустимость применения методов МСА к исследуемой выборке предопределена результатами первичного статистического анализа и выводами анализа временных рядов. Принципы выбора возможных путей применения МСА следует искать прежде всего в выполнении гипотез о нормальности эмпирических распределений исходных переменных и о стационарности, однородности исследуемых рядов.



Рис. В.4. Функциональная схема последовательности многомерного статистического анализа океанологической информации.

Необходимость первоочередной классификации исследуемого набора переменных может быть в достаточной степени обоснована при невыполнении статистических гипотез нормальности одномерных и двумерных совместных законов распределения, особенно в случае проявления многомодальности. Именно последняя особенность представляется во многом решающей для принятия решения о классификации исследуемых процессов. По опыту авторов, первоначальное проведение структурного анализа подобных данных методами разложения в систему ортонормированных базисов нередко приводит к малоэффективным статистическим моделям. Причина получения таких результатов лежит в низкой информативности оценок параметров корреляции и регрессии, на которых базируется большинство эмпирических методов разложения.

Следует напомнить исследователям-практикам, что подобная закономерность вполне обоснована известным требованием соблюдения нормальности двумерного распределения вероятностей переменных, для которых строится линейная корреляционная оценка взаимосвязи [8, 75, 96, 211]. Если гипотеза нормальности эмпи-

рического закона распределения не выполняется с вероятностью P (%), то имеет смысл рассчитывать на получение содержательного результата разложения в том случае, если точность разложения будет не выше P (%).

Важно подчеркнуть, что сочетание, взаимоувязывание нескольких методов МСА на сегодняшний день представляется назревшей необходимостью, которая подкрепляется практическим опытом применения методов, теоретической обоснованностью такого объединения и техническими возможностями современных ЭВМ. При таком подходе принцип сочетаемости методов исследования сложных океанологических процессов может реализоваться на новом уровне понимания полученных статистических выводов. Сочетание методов применительно к разным координатным осям (например, пространство-переменные и переменные — время) позволяет описать структуру взаимосвязей всего куба многомерной океанологической информации, а значит, воссоздать более полную картину формирования объекта исследования.

Приводимые ниже примеры диагноза океанологических процессов послужат конкретными аргументами в пользу высказанных соображений.

Прогноз океанологических процессов

Проведение диагностических исследований не является самоцелью. Чаще всего они предваряют заключительный и наиболее сложный этап — прогнозирование океанологических процессов. Эта задача стоит особенно остро сегодня, когда цена принятия обоснованных и экономически эффективных решений непосредственно зависит от качества исходной гидрометеорологической информации и прогностических оценок. При этом наиболее существенные успехи в предсказании всегда были обусловлены глубоким пониманием происходящих в океане процессов и самих используемых методов экстраполяции, на базе которых верифицируются физические механизмы прогнозируемых явлений.

Интуитивно ясно, что методы диагноза должны быть неразрывно связаны между собой при исследовании конкретных процессов. Однако на практике этого зачастую не происходит, и вызвано это отсутствием соответствующих теоретических и методических проработок. Настоящая работа имеет целью, в частности, продемонстрировать эффективность и перспективность применения методов МСА также для решения задач прогноза многомерных океанологических процессов.

Понятие «прогноз» в современной прикладной океанологии формулируется как научно обоснованное предвидение изменений в состоянии элементов режима океана. Проблема предсказания океанологических характеристик нетривиальна в силу сложности процессов, происходящих в океане, с одной стороны, и недостаточности научно обоснованного знания, позволяющего адекватно описывать процессы, с другой. Общая теория океанологических

процессов создает систему взглядов, подчас не лишенную отдельных противоречий, на базе которой разрабатываются конкретные методы анализа и прогноза океанологических характеристик. Однако из нее отнюдь не следует универсальный алгоритм, гарантирующий эффективный прогноз. Наблюдения в океане редко допускают однозначную интерпретацию фиксируемой изменчивости и чаще всего служат лишь основой формулировки гипотез, нуждающихся в тщательной проверке. Многообразие факторов, прямо и косвенно влияющих на временную изменчивость океанологических характеристик, затрудняет однозначную формализацию причинно-следственных зависимостей. Поэтому статистические методы диагноза и прогноза океанологических характеристик находят все более широкое применение в практике. По аналогии с [153] статистический прогноз можно определить как вероятностное суждение о перспективах развития и возможных состояниях океанологического процесса в будущем, сформулированное на основе статистической модели эволюции процесса.

Принципы статистического прогнозирования

В процессе прогностического исследования одной из важнейших задач является разработка методологических вопросов с целью повышения обоснованности прогнозов. Следование хорошо известным методологическим принципам построения прогностических моделей в определенной степени позволяет исследователю избежать очевидных ошибок. Напомним эти принципы.

1. *Согласованность модели со структурой исходных данных наблюдений.* Из этого принципа следует, что прежде чем переходить к разработке конкретного метода прогноза, необходимо дать описание структуры данных, т. е. провести их корректный статистический анализ.

2. *Полнота, адаптивность и эволюционность прогностической модели.* Модель должна описать широкий диапазон изменений изучаемого процесса, самонастраиваться при описании конкретного класса процесса, а ее параметры должны изменяться в зависимости от точности воспроизведения. Следование этим принципам позволяет глубже понять особенности разрабатываемой модели и классы параметров, которые она описывает, т. е. дополнительно диагностировать изучаемые процессы.

3. *Простота модели.* Согласно этому принципу, модель обеспечивает наглядность для исследователя и потребителя (заказчика), чему способствует блочное представление модели (в виде программных модулей и универсальных программ). Заменяемость блоков, их совершенствование в конечном итоге позволяют исследователю определить возможности модели, т. е. провести верификацию получаемых прогностических моделей и вынести окончательную объективную оценку используемой модели.

Классификация методов прогноза

Прогнозирование многомерного процесса сейчас представляется несколько обособленной задачей в практике современных прогнозистических исследований. Для океанологии более типично прогнозирование одномерного процесса, что именуется прогнозом элемента режима или прогнозом океанологической характеристики. Многосвязность, многомерность собственно самого процесса формирования изменчивости рассматриваемой характеристики при этом отходит на второй план (опосредованная параметризация через уравнение множественной регрессии) либо вообще исчезает при использовании уравнений автопрогноза. В настоящей работе делается попытка сохранить многомерность описания процесса и в диагнозе, и в прогнозе. При этом полагается, что важнейшее содержание прогноза многомерного процесса состоит как раз в предсказании взаимообусловленной эволюции параметров m -мерного вектора наблюдаемых переменных. Поэтому на первое место выходит не отвлеченное прогнозирование отдельных параметров, а разработка прогностической модели описания многомерных взаимосвязей параметров процесса.

Количество возможных методических подходов к статистическому прогнозированию многомерных процессов предопределяется содержательными результатами диагноза, с одной стороны, и целевой функцией прогноза, с другой. В табл. В.2 представлены три наиболее важные цели разработки прогностической модели

Таблица В.2

Прикладная классификация методов прогноза многомерных процессов

Цели	Принципы	Методы
Прогноз структуры процесса	Предсказание эволюции обобщенных параметров в модели ортогонального разложения	Экстраполяция членов разложения методами автопрогноза (авторегрессия, скользящее среднее, Фурье-гармоники)
Прогноз класса состояния процесса	Предсказание эволюции процесса в терминах смены выделенных и интерпретированных классов состояния	Экстраполяция функции плотности вероятности сохранения или изменения класса методами байесовского оценивания и теории марковских процессов
Прогноз отдельной переменной многомерного процесса	Предсказание характеристики с учетом влияния всех параметров процесса	Экстраполяция на основе учета асинхронных взаимосвязей исходных параметров методами множественного регрессионного анализа (линейная, полиномиальная, спектральная, гребневая, пошаговая регрессия)

в прикладных океанологических исследованиях. Кратко охарактеризуем соответствующие им принципы и методические приемы, которые будут предметом дальнейшего обсуждения.

Прогноз структуры многомерного процесса реализуется на основе экстраполяции во времени результатов ортогонального разложения матриц исходных наблюдений методами компонентного и факторного анализа. Изменчивость во времени общих факторов (главных компонент) обобщает статистически значимую взаимосвязанную динамику исходных переменных. При этом происходит существенное сжатие информации — число общих факторов практически в два раза меньше числа исходных переменных. Взаимонезависимость членов разложения обеспечивает эффективность применения методов автопрогноза с самостоятельным подбором оптимального экстраполяционного уровня для каждого члена разложения.

Выраженная многомодальность исследуемого многомерного процесса является существенной предпосылкой для диагноза и прогноза его динамики в терминах классов состояния. Оценка будущего состояния процесса при этом производится на основе экстраполяции функции плотности вероятности методами теории марковских процессов и байесовского оценивания. Прогностические значения исходных параметров задаются через их оценки математического ожидания и дисперсии, соответствующие данному классу.

В ряде практических ситуаций прогнозирование многомерных процессов сводится к требованию экстраполяции единственной наблюдаемой переменной процесса. В принципе не составляет труда осуществить решение этой задачи с помощью двух рассмотренных подходов. Опираясь на диагноз структуры процесса, можно осуществить прогнозирование путем экстраполяции значительно ограниченного числа членов ортогонального разложения, которые статистически достоверно связаны с экстраполируемой переменной. Другой вариант — построение прогностической модели на основе учета асинхронных взаимосвязей собственно исходных параметров процесса с экстраполируемой переменной. Традиционно эта задача решается методами множественного регрессионного анализа, а в зависимости от устойчивости взаимосвязей и меры их линейности выбирается оптимальная регрессионная модель.

Общая последовательность обработки многомерных океанологических процессов

Современные физико-статистические гидрометеорологические прогнозы в своем большинстве реализуют две самые общие идеи:

1) многопараметрический прогноз, при котором эволюция процесса описывается путем привлечения нескольких наиболее устойчивых внешних асинхронных взаимосвязей, трактуемых в терминах причинно-следственных механизмов;

2) автопрогноз, когда будущее развитие процесса ставится в зависимость только от его же изменчивости в прошлом.

Характерно, что применение в океанологии многопараметрических схем прогнозов прямо не связано с полным исчерпанием потенциальных возможностей методов автопрогноза. Во всяком случае авторам не известны случаи конкретных практических доказательств на фактическом материале такого соотношения двух методических приемов. Скорее, наоборот, скачок к многопараметрическим схемам прогнозирования зачастую происходит после неудачной попытки автопрогноза с применением одного из простейших методов тренд-анализа или фурье-аппроксимации. В то же время реальные возможности широкого набора методов статистико-вероятностного автопрогноза остаются вообще вне поля зрения.

Развитие и совершенствование статистических методов позволяет взглянуть на соотношение многопараметрических прогнозов и автопрогнозов с новых позиций, непосредственно базируясь на результатах диагноза многомерных океанологических процессов. Применение методов МСА в диагнозе создает заманчивую возможность достоверно описать исследуемый океанологический процесс в терминах замкнутых статистико-вероятностных моделей с минимальным числом обобщенных взаимонезависимых параметров — главных компонент, общих факторов, классов. Прогностическая ценность таких параметров очевидна. Новое содержание получает классическая методология автопрогноза в применении к обобщенным показателям диагноза.

Новизна предлагаемого подхода как раз и состоит в переосмыслении понятия многомерности применительно к диагнозу и прогнозу. Многомерность не рассматривается исключительно как множественность исходных предикторов, наблюдаемых или рассчитываемых непосредственно для прогностического уравнения. Она трактуется как многомерность самого предиктанта, который достоверно описывает состояние океанологического процесса и содержит достаточное количество информации для автопрогноза. Поэтому при диагнозе каждый наблюдаемый параметр состояния океанологического процесса изучается не обособленно, а в совокупности его статистических взаимосвязей с другими переменными, что и позволяет в конечном счете сформировать обобщенные показатели структуры и динамики процесса. Содержание прогноза тогда трансформируется в решение задачи эффективного предсказания обобщенных параметров диагноза без необходимости явного учета внешних связей, что существенно упрощает вычислительную сторону построения прогностических схем.

Самое общее формализованное представление диагностической модели МСА при таком подходе может быть зафиксировано матричным уравнением вида

$$X(t, z, l) = F(t, z) \cdot A(z, l) + E(t, z, l).$$

Здесь X — матрица наблюдаемых параметров состояния океана;

F — матрица обобщенных параметров, получаемых в результате диагноза, A — матрица коэффициентов взаимосвязи; E — матрица ошибок, не описываемых обобщающей диагностической моделью.

При решении прогностической задачи осуществляется предсказание обобщенных параметров, которые представлены в матрице $\hat{F}$. Их поведение во времени описывается широким классом экстраполяционных моделей вида

$$\hat{F}(t + \tau, z) = F(t, z) \cdot B(t, z) + \varepsilon(t, z),$$

в которых матрица прогностических статистико-вероятностных коэффициентов B связывает прошлое F и будущее $\hat{F}$ состояния океанологического процесса. Здесь также появляется и остаточная часть ε , в которую входят ошибки прогнозирования.

Взаимосвязь диагноза и прогноза осуществляется не только через единую для двух рассматриваемых этапов матрицу обобщенных параметров F . В случае если матрица остатков диагностических моделей отличается от процесса типа белого шума, она также может быть привлечена для самостоятельного прогнозирования. Тогда в прогностическую модель будет включен еще один член, ответственный за прогноз самой матрицы ε .

На рис. В.5 приведена схема общей последовательности обработки многомерных океанологических процессов. Важнейшие аспекты выделяемых этапов диагноза и прогноза уже обсуждались выше и будут далее детально анализироваться. Здесь целесообразно дать характеристику взаимосвязей отдельных крупных блоков, чтобы предварительно оценить потенциальные возможности применения такого подхода на практике. Основной рабочий принцип организации взаимосвязей блоков состоял в реализации идеи обратной связи отдельных методов обработки и этапов анализа. Включение обратных связей в известной мере усложняет процесс обработки информации. Их применение диктуется существующей неопределенностью знаний о природе многомерных океанологических процессов. Поэтому практически нереально выработать сегодня единственно возможную последовательность обработки и представления данных, которая удовлетворила бы широким запросам различных направлений прикладного анализа наблюдений. Применение обратных связей в принципе должно обеспечить подбор оптимальной диагностической и прогностической модели в рамках конкретной задачи и имеющейся в наличии информации.

Расшифруем смысл обратных связей следуя общей последовательности диагноза и прогноза. Имеющаяся информация об изменчивости наблюдаемых процессов не может исчерпывающе охарактеризовать процесс, если не будет дополнена непосредственно ненаблюдаемыми физическими характеристиками, которые вычисляются на основании применения известных физических законов гидротермодинамики. Получаемые характеристики дополняют спи-

сок исходных параметров и включаются в процедуру статистического анализа.

Статистический анализ взаимодействует непосредственно с физической интерпретацией полученных частных количественных ре-

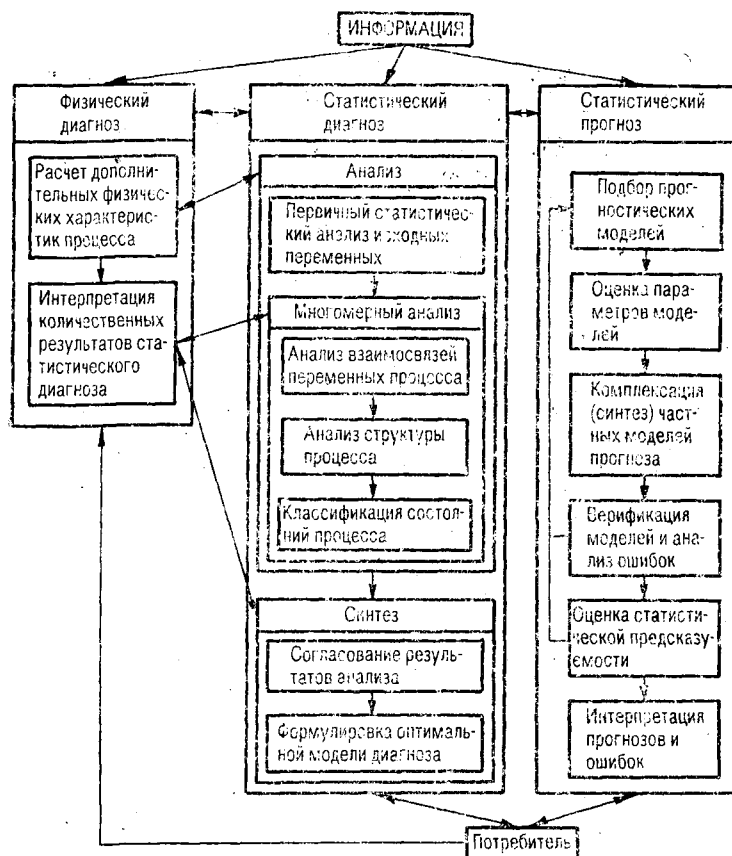


Рис. В.5. Функциональная схема диагноза и прогноза океанологических процессов.

зультатов применения отдельных методов, низкая успешность которой может быть причиной выполнения повторных вычислений, направленных на получение более достоверных физических выводов о природе исследуемых процессов.

На этапе синтеза результатов расчетов может возникнуть необходимость повторных аналитических расчетов, чтобы лучше увязать результаты различных методов и построить законченную общую статистическую модель процесса. Согласованность результатов расчетов в рамках общей модели процесса должна быть физически обоснованна, чем объясняется необходимость возврата

к физическому диагнозу. При необходимости расширения общей модели диагноза набор исходных параметров может быть дополнен вторично, с учетом выявленных неточностей интерпретации.

Двухсторонняя связь диагноза и прогноза выражает идею построения прогностической зависимости на основе результатов диагноза и, наоборот, уточнение диагностических расчетов с учетом выполнения требований к построению прогностических моделей. Возврат к диагнозу может быть обусловлен неудовлетворительным качеством прогнозов, слабой предсказуемостью модели, если перечисленные недостатки не удастся устранить в процессе повторного пересчета прогностических моделей. При подборе моделей подразумевается, что исследователь выберет сразу несколько типов моделей, которые в большей или меньшей степени соответствуют результатам диагноза.

Любая прогностическая схема — это статистическая модель развития процесса. Она дает возможность заглянуть в будущее, на основе «своей интерпретации» истории развития процесса, которая зачастую не во всем адекватна фактическим данным. Это и приводит к противоречивым выводам, с помощью которых трудно дифференцировать полезность различных схем прогнозов океанологических характеристик. Причиной тому является «настроенность» модели на прогноз определенного типа процессов. В результате перестройки процессов на другой тип модель перестает работать, а при повторении исходного типа процессов — опять дает удовлетворительные результаты. Другие модели лучше прогнозируют «свои» типы процессов, но и им свойствен тот же недостаток: достоверно прогнозировать процессы другого типа не удается.

В связи с этим становится ясно, что эффективность прогнозирования зависит не только от вида модели, но и, что очень важно, от типа прогнозируемых процессов. Многие прогностические модели являются адаптационными (второй принцип), и изменение их структуры, подстройка параметров непосредственно определяются процессом, т. е. возникает эффект влияния процесса на модель. Это может привести к тому, что модель в результате коррекции выходит не на «свои» процессы и полностью расстраивается. Поэтому необходимо выявить, в каких типовых ситуациях экономически выгодно использовать те или иные прогностические модели.

Из этого следует, что прежде чем переходить к прогнозу, необходимо проанализировать и классифицировать процессы, при которых прогнозирование ведется успешно и при которых происходят срывы, выяснить их физическую сущность. Фактически диагностический блок должен включать как диагноз физической природы процессов, так и диагноз их связи с конкретными методами статистического диагноза и прогноза. В дальнейшем, по желанию исследователя, прогностический блок может включать такие модели, которые наилучшим образом в смысле минимизации ошибок описывают определенный класс процессов.

Таким образом, изложенный методологический аспект сводится к тезису о работе прогностических моделей лишь на «своих» типах процессов, которые можно идентифицировать по степени детерминированности и характеру развития во времени. Отсюда следует принципиальный вывод о том, что не существует единственной универсальной методики прогноза океанологических характеристик. Впрочем этот вывод относится ко многим другим природным процессам [135, 142, 143, 146, 151, 171, 225], что приводит к пониманию необходимости одновременного использования нескольких прогностических моделей. Подобная процедура получила название комплексирования прогнозов. В этом видится резерв повышения качества гидрометеорологических прогнозов [35, 36, 43, 44, 76, 216, 245].

Методология построения общей прогностической модели (модели-симбиоза) сводится к четырем основным этапам вычислений:

- 1) коллективизация (отбор) исходных моделей,
- 2) гибридизация моделей (выбор метода комплексации),
- 3) реализация модели (оценка параметров),
- 4) верификация модели (сравнительный анализ ошибок).

Среди проблемных задач статистического прогноза особое место занимает задача увеличения пределов статистической предсказуемости. Она во многом обусловлена выбором оптимальной модели, описывающей устойчивые во времени закономерности изменения прогнозируемой океанологической характеристики. Комплексация нескольких альтернативных (дополняющих друг друга) моделей является одним из путей увеличения как пределов статистической предсказуемости, так и улучшения оправдываемости прогнозов. Этот результат достаточно ясен: комплексная модель отражает основные особенности в изменчивости прогнозируемой характеристики и может перенастраиваться в случае смены типа прогнозируемого процесса.

Полученные оценки статистической предсказуемости помогают правильно распорядиться используемыми прогностическими моделями при оперативном прогнозировании подобрать оптимальные частные модели и оптимизировать комплексацию, определить и использовать эффективную заблаговременность прогноза для заданной точности.

Интерпретация полученных прогнозов как завершающий этап прогнозирования представляет собой описание прогнозируемых значений в терминах задаваемых потребителем с указанием доверительных интервалов оценок и физической трактовки развития многомерного процесса на период прогноза.

Глава 1

ОПИСАНИЕ ВЗАИМОСВЯЗЕЙ ПРОЦЕССОВ

1.1. Множественный регрессионный анализ. Множественная линейная регрессия

Постановка задачи

Аппарат множественной линейной регрессии (МЛР) представляется на сегодняшний день одним из наиболее хорошо разработанных средств математической статистики. Как и многие другие идеи, концепция регрессионного анализа была сформулирована Гауссом более 100 лет назад. Методика нашла широкое применение практически во всех прикладных науках, что позволило существенно развить ее и адаптировать к решению многих задач [70, 80, 139, 140, 163, 200].

В практике гидрометеорологических исследований метод МЛР наиболее широко применяется для описания и предсказания отдельных переменных на основе статистического учета влияния нескольких влияющих факторов. Кроме того, метод МЛР часто применяется для решения вопросов, связанных с интерполяцией океанологических полей в пространстве как по вертикали, так и по горизонтали. Важно отметить, что МЛР выступает в практике не только как самостоятельная методика, но и как составная часть ряда других методов исследования зависимостей переменных [12, 72, 144].

Рассматриваемая ниже достаточно общая постановка задачи МЛР может быть распространена и на ряд других моделей регрессионного анализа. В частности, на случай полиномиальной регрессии, основное уравнение которой может быть сформировано на основе линейной регрессии при соответствующей замене переменных.

Нас интересует возможность описания изменчивости зависимой переменной как функции влияния группы независимых переменных $x_1, x_2, x_3, \dots, x_M$. Предположим, что влияние каждой из M независимых переменных x_j на зависимую переменную y можно описать как линейное. Тогда основное уравнение множественной линейной регрессии будет иметь вид

$$y_i = a_0 + \sum_{j=1}^M a_j (x_{ij} - \bar{x}_j) + e_i = \hat{y}_i + e_i, \quad (1.1)$$

где i — номер наблюдения из выборки значений длиной N , ($i = 1, \dots, N$); j — номер независимой переменной x_j ($j = 1, \dots, M$); $a_0, a_1, a_2, \dots, a_M$ — коэффициенты линейной регрессии; $\bar{x}_j$ — среднее значение j -й переменной ($\bar{x}_j = (1/N) \sum_{i=1}^N x_{ij}$); $\hat{y}_i$ — значение

i -го наблюдения зависимой переменной, восстановленное с помощью регрессионного уравнения; e_i — отклонение восстановленного значения $\hat{y}_i$ от наблюдаемого y_i .

Решение задачи построения модели множественной линейной регрессии сводится к поиску таких значений коэффициентов регрессии $a_0, a_1, a_2, a_3, \dots, a_M$, которые позволили бы восстановить изменчивость зависимой переменной y с минимальными ошибками e .

Поскольку истинные значения коэффициентов регрессии исследователю не известны, то выполняется статистическая оценка этих коэффициентов, основанная на минимизации дисперсии ошибок s_R :

$$s_R = \frac{1}{N-1} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \rightarrow \min.$$

Традиционным средством поиска значений $a_0, a_1, \dots, a_M$, соответствующих данному условию, является известный метод наименьших квадратов, эффективность применения которого, а следовательно, точность уравнения МЛР, зависят от выполнения следующих простых условий:

— необходимо, чтобы ошибки регрессии имели нулевое среднее:

$$\bar{e} = \frac{1}{N} \sum_{i=1}^N e_i = 0$$

(здесь под знаком $\bar{e}$ понимается оператор математического ожидания ряда e),

— влияющие признаки должны быть статистически независимы между собой:

$$(\overline{x_j, x_l}) = 0 \text{ при } j \neq l.$$

Другими словами, они не должны между собой коррелировать

$$r(x_j, x_l) = 0.$$

При выполнении перечисленных свойств вычисление значений коэффициентов регрессии сводится к решению системы нормальных уравнений вида

$$\begin{bmatrix} M & \sum x_{i1} & \sum x_{i2} & \dots & \sum x_{iM} \\ \sum x_{i1} & \sum x_{i1}^2 & \sum x_{i1}x_{i2} & \dots & \sum x_{i1}x_{iM} \\ \dots & \dots & \dots & \dots & \dots \\ \sum x_{iM} & \sum x_{iM}x_{i1} & \sum x_{iM}x_{i2} & \dots & \sum x_{iM}^2 \end{bmatrix} \cdot \begin{bmatrix} a_0 \\ a_1 \\ \dots \\ a_M \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_{i1}y_i \\ \dots \\ \sum x_{iM}y_i \end{bmatrix}.$$

где суммирование (Σ) проводится по всей длине выборки из N наблюдений. Очевидно, что при большом числе переменных ($M > 3$) приводимая математическая поэлементная запись становится излишне громоздкой. Более удобна матричная запись. Рекомендуем читателю предварительно освежить в памяти основные понятия матричной алгебры, краткая сводка которых приводится в [12, 86].

Для перехода к матричным обозначениям введем следующие понятия:

y — вектор-столбец наблюдений длиной N значений, $y^T = \{y_1, \dots, y_N\}$, центрированный на среднее арифметическое значение

$$\bar{y} = a_0 = \frac{1}{N} \sum_{i=1}^N y_i;$$

X — матрица значений независимых переменных, состоящая из N строк и M столбцов (i -я строка есть вектор наблюдений за M признаками, j -й столбец есть вектор N значений j -й переменной);

A — вектор-столбец коэффициентов регрессии $A^T = \{a_1, a_2, \dots, a_M\}$,

E — вектор-столбец N ошибок регрессионной модели.

В новых обозначениях основное уравнение МЛР примет вид

$$Y = X \cdot A^T + E.$$

Система нормальных уравнений для оценки коэффициентов регрессии запишется следующим образом:

$$(X^T \cdot X) \cdot A = X^T \cdot Y,$$

где индекс «т» соответствует транспонированной матрице или вектору (см. [3, 12, 86]). Очевидно, что решение системы нормальных уравнений путем переноса матрицы $(X^T \cdot X)$ в правую часть системы запишется в виде

$$A = (X^T \cdot X)^{-1} \cdot X^T \cdot Y. \quad (1.2)$$

Полученное в виде (1.2) решение задачи МЛР обладает рядом ценных для исследователя свойств:

— по определению, вектор значений коэффициентов регрессии $A^T = \{a_1, \dots, a_M\}$ минимизирует суммарную квадратическую ошибку описания изменчивости зависимой переменной, что позволяет утверждать оптимальность данной модели среди всех других моделей подобного класса для выбранного набора переменных;

— сами коэффициенты регрессии a_j являются линейными функциями наблюдений Y и представляют собой несмещенные оценки истинных (но неизвестных) значений коэффициентов регрессии;

— оценки коэффициентов регрессии обладают минимальной дисперсией среди любых других линейных оценок подобного рода.

Введем в рассмотрение еще одну важную характеристику модели МЛР, которая называется коэффициентом множественной корреляции R_0 . Данный показатель является необходимой количественной характеристикой подобия наблюдаемых y и вычисляемых $\hat{y}$ значений зависимой переменной. Вычисляется коэффициент множественной корреляции по формуле

$$R_0 = \frac{\sum_{i=1}^N (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^N (y_i - \bar{y})^2 \sum_{i=1}^N (\hat{y}_i - \bar{\hat{y}})^2}}. \quad (1.3)$$

Величину R_0^2 называют коэффициентом детерминации, поскольку она является простым обобщением коэффициента корреляции $r^2(x, y)$ на случай многих независимых переменных и пропорциональна доле дисперсии переменной y , описываемой данной моделью регрессии. В случае одной независимой переменной выполняется тождество $R_0^2 = r^2(x, y)$. Нетрудно видеть, что при $R_0^2 = 1$ имеет место полное соответствие наблюдаемых и вычисляемых значений зависимой переменной.

Интерпретация коэффициента множественной корреляции (КМК) несколько отличается от понимания классического коэффициента парной корреляции. Дело в том, что КМК имеет пределы изменения от 0 до 1, и, следовательно, он не показывает знака связей между y и совокупностью $x_1, \dots, x_m$. Кроме того, величина КМК оказывается больше или равна максимальному коэффициенту парной корреляции $r_j(x_j, y)$ в рассматриваемой матрице данных.

Интерпретировать КМК следует как величину общей линейной связи, зависимой переменной и выбранных независимых переменных, а также как показатель эффективности модели МЛР. Последнее утверждение следует из того факта, что квадрат КМК численно равен доле дисперсии переменной y , которая может быть описана данным уравнением МЛР.

Геометрическая интерпретация МЛР

Представим себе координатную систему, образованную тремя осями: температурой, соленостью и условной плотностью воды (рис. 1.1). Плотность является зависимой переменной от температуры T и солености S . Искомое уравнение множественной регрессии будет записано в виде

$$\rho_i = a_0 + a_1 T_i + a_2 S_i.$$

Здесь величина коэффициента $a_0 = \frac{1}{N} \sum_{i=1}^N \rho_i$ будет задавать уровень среднего значения плотности из имеющейся выборки. Вычисляемые значения ρ образуют плоскость, наклоненную к осям T и S . Величина коэффициента регрессии $a_1 = \operatorname{tg}(\alpha_1)$ будет соответствовать углу наклона плоскости регрессии к плоскости (ρ, T) . Величина коэффициента $a_2 = \operatorname{tg}(\alpha_2)$ будет соответствовать углу наклона плоскости регрессии к плоскости (ρ, S) . Величина

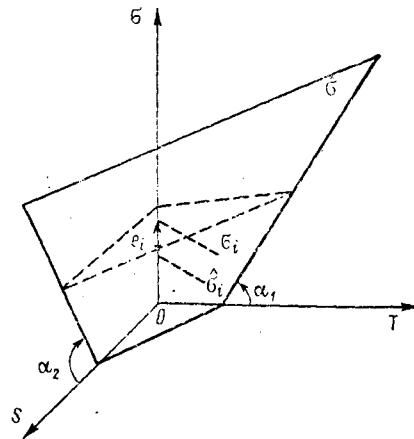


Рис. 1.1. Геометрическая интерпретация уравнения линейной регрессии.

ошибки e_i будет эквивалентна расстоянию от наблюдаемого значения ρ_i до его проекции на плоскость регрессии $\hat{\rho}_i$.

Вычисление по модели МЛР

Рассмотрим общий алгоритм вычислений по схеме МЛР.

1. Вычисляются оценки средних арифметических и стандартных отклонений всех исходных параметров:

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i, \quad \bar{x}_j = \frac{1}{N} \sum_{i=1}^N x_{ij},$$

$$s_y = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2, \quad s_{x_j} = \frac{1}{N-1} \sum_{i=1}^N (x_{ij} - \bar{x}_j)^2.$$

Составляются векторы средних и стандартных отклонений: $\bar{\mathbf{x}}, \mathbf{s}_x$.

2. Вычисляется полная корреляционная матрица $\mathbf{R}$, составленная из парных коэффициентов корреляции между всеми переменными:

$$\mathbf{R} = \frac{1}{N} \mathbf{X}^T \mathbf{X} = \begin{bmatrix} r_{yy} & r_{yx_1} & r_{yx_2} & \dots & r_{yx_M} \\ r_{x_1y} & r_{x_1x_1} & r_{x_1x_2} & \dots & r_{x_1x_M} \\ \dots & \dots & \dots & \dots & \dots \\ r_{x_My} & r_{x_Mx_1} & r_{x_Mx_2} & \dots & r_{x_Mx_M} \end{bmatrix}.$$

Коэффициенты корреляции вычисляются, как правило, по традиционным формулам и проверяются на значимость по критерию Стьюдента.

3. Вычисляются определители для матрицы корреляций и ее миноров. В методике МЛР рассматриваются определители матриц вида

$$D = \begin{bmatrix} 1 & r_{yx_1} & r_{yx_2} & \dots & r_{yx_M} \\ r_{x_1y} & 1 & r_{x_1x_2} & \dots & r_{x_1x_M} \\ \dots & \dots & \dots & \dots & \dots \\ r_{x_My} & r_{x_Mx_1} & r_{x_Mx_2} & \dots & r_{x_Mx_M} \end{bmatrix}.$$

Миноры для расчета определителей вида D_{yx_j} строятся путем исключений y -й строки и x_j -го столбца из матрицы R .

4. Вычисляются коэффициенты регрессии по формулам (1.2) или

$$a_j = \frac{s_y}{s_{x_j}} \frac{|D_{xyj}|}{|D_{yy}|}.$$

5. На основании полученных значений коэффициентов регрессии строится уравнение регрессии и вычисляются значения $\hat{y}$:

$$\hat{y}_i = a_0 + \sum_{j=1}^M (x_{ij} - \bar{x}_j) a_j.$$

6. Оценивается полный или множественный коэффициент корреляции по формуле (1.3) или

$$R_0 = \sqrt{\frac{1 - |D|}{|D_{yy}|}}.$$

7. Вычисляются оценки стандартного отклонения для коэффициентов регрессии и проверяется значимость регрессионных параметров по критерию Стьюдента (t_v):

$$s_{a_j} = \sqrt{\frac{N s_y}{(N - M) (|D| / |D_{yy}|)}}, \quad t_v \geq \frac{a_j}{s_{a_j}}.$$

8. Проверяется общая адекватность модели МЛР для исходных данных путем расчета эмпирического критерия Фишера:

$$F^* = s_D / s_R,$$

где

$$s_D = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2, \quad s_R = \frac{1}{N-M-1} \sum_{i=1}^N (y_i - \hat{y}_i)^2.$$

Если рассчитанная величина соотношения F^* превышает табличное значение F_τ при заданном уровне значимости и степенях свободы $v_1 = M$ и $v_2 = N - M - 1$, то гипотеза об адекватности построенной модели принимается.

Приведенная выше схема оценки параметров регрессионной модели является наиболее общей и часто встречается в програм-

мах стандартного математического обеспечения статистического анализа данных на ЭВМ различных классов. Для ЭВМ класса ЕС схема линейного регрессионного анализа реализована как в советских пакетах прикладных программ [4, 77, 80, 185], так и в зарубежных [12, 70, 139, 140].

Сложности реализации полной схемы на мини-ЭВМ связаны с проблемами представления и обработки матриц данных в условиях ограниченной мощности подобных компьютеров. Расширенные возможности крупных ЭВМ позволяют не только выполнить собственно расчет по данной схеме МЛР, но провести более глубокое исследование регрессионных зависимостей выбранных переменных.

Так, в большинстве пакетов программ, перечисленных здесь, может быть проведено пошаговое исключение или перебор независимых переменных с целью выбора оптимального числа признаков.

Кроме того, появляются дополнительные возможности по предварительному преобразованию исходных признаков с целью их наиболее выгодного представления в модели МЛР (с помощью известных методов линеаризации). При проверке адекватности модели МЛР значительную помощь исследователю могут оказать графики остатков, которые могут быть получены в пакетах BMDP, STATGRAF [3, 73, 122, 140].

Применение МЛР в океанологических исследованиях

Исторически сложилось так, что уравнения регрессии использовались в океанологических исследованиях практически с момента зарождения науки об океане. От расчета параметров единственного уравнения регрессии со временем океанологи переходят к применению в целом всей методологии регрессионного анализа. Эта тенденция оформилась наиболее четко в 70-е годы, когда начали решаться задачи построения оптимальной регрессионной модели для конкретной выборки наблюдений. Сейчас методология регрессионного анализа прочно внедрилась в прикладные океанологические исследования. Однако неизменными остались трудности практического применения регрессионного анализа в океанологии, которые чаще всего связаны со следующими проблемами:

- подбором оптимального вида регрессионного уравнения,
- надежной оценкой регрессионных параметров.

Рассмотрим опыт регрессионного анализа в океанологии именно в аспекте перечисленных проблем.

Регрессионный подход лежит в основе описания фундаментальных физических законов развития океанических процессов. На принципах линейной регрессии базируются формулы расчета характеристик нарастания и таяния льда [62], параметров теплового взаимодействия океана и атмосферы [61] и переменных в уравнениях состояния морской воды [27].

Ретроспектива последних десятилетий четко показывает быструю эволюцию большинства регрессионных уравнений. Изна-

начально это были малопараметрические уравнения низких порядков. По мере углубления знаний о физических законах и накопления больших архивов наблюдений исследователи пытаются максимально точно описать регрессионными формулами выявленные взаимосвязи океанологических характеристик. Поэтому уравнения регрессии изменялись, как правило, в сторону усложнения их вида.

Во-первых, вводилось в рассмотрение все большее число независимых переменных, которые описывали вновь выявленные корреляционные взаимосвязи.

Во-вторых, совершенствовалась форма описания зависимостей: в виде независимых переменных использовались агрегированные характеристики, которые представляют собой различные комбинации исходных параметров физических процессов.

Весьма ярко такая тенденция прослеживается в регрессионных уравнениях состояния морской воды. Современная океанология использует большой набор эмпирических уравнений, начиная от простейшего

$$\rho - \rho_0 = a_1 T + a_2 S$$

и кончая формулами Чанга—Миллера [87], где взаимосвязь плотности, температуры, солености и давления описывается смешанным полиномом 5-го порядка. Вид конкретного уравнения диктуется необходимым уровнем точности аппроксимации плотности воды и определяется во многом исходной выборкой наблюдаемых $\rho - T - S$ соотношений. Однако в основе регрессионных расчетов сохраняет неизменным принцип линеаризации исследуемых зависимостей с последующей оценкой регрессионных параметров по методу наименьших квадратов [3, 78].

При исследовании пространственного распределения океанологических полей регрессионный подход нашел широкое применение для решения задач интерполяции [25, 27]. Общепринятые в океанологии принципы построения карт на основе восстановления значений в узлах сеточной области используют идею аппроксимации полей как функций от географических координат. Подобная зависимость описывается регрессионным уравнением, порядок которого, как и вид переменных регрессии, подбирают исходя из условия достоверного описания пространственных аномалий заданного масштаба.

Регрессионный подход лежит и в основе методики объективной интерполяции и синхронизации океанологических полей [53, 57, 209]. Здесь регрессионные уравнения оцениваются для каждого узла сеточной области как функции от ближайших точек наблюдений. Согласование и синхронизация полей достигается применением того же подхода уже к различным океанологическим характеристикам и к рядам наблюдений во времени [27, 101, 222, 228, 261].

Регрессионное описание изменчивости океанологических характеристик во времени носит принципиально важный характер при исследовании океана и признано важнейшим методом выявления

физических закономерностей. Потенциальная взаимозависимость различных океанологических процессов в пространстве и времени объясняет стремление описать ее в терминах регрессионных моделей и предсказать возможные тенденции развития процессов.

Успешное построение регрессионных моделей временной изменчивости нескольких океанологических характеристик играет важную роль, поскольку позволяет достоверно трактовать выявленные сопряженные колебания в рамках известных физических законов функционирования гидрофизических, гидрохимических и других процессов.

Пример тому — промыслово-океанологические исследования, цель которых состоит в достоверном описании взаимных колебаний океанических и биологических характеристик [212], на основе которого в дальнейшем формулируется закономерность воздействия среды на биологические объекты [108].

Особое значение подобные регрессионные зависимости приобретают при попытке восстановления пропусков в исторических архивах океанологических наблюдений. Однако здесь исследователь сталкивается с проблемой неоднородности выявленных зависимостей во времени. Не секрет, что долгопериодные колебания в климатической системе порождают в выборочных рядах эффекты нестационарности по математическому ожиданию и дисперсии.

Нестационарность, в свою очередь, прослеживается в неустойчивости корреляционных связей и, следовательно, в оценке регрессионных параметров. Прежде всего подобные влияния приводят к снижению общей адекватности построенных регрессионных моделей. Наиболее часто применяемый способ преодоления перечисленных трудностей связан с заменой обобщающего подхода на локальный.

Другими словами, вместо единственной общей для всего массива данных регрессионной модели производится построение нескольких частных уравнений регрессии. Последние наилучшим образом аппроксимируют взаимосвязи океанологических переменных на конкретном участке их изменчивости и не работают на других наблюдениях.

Таким образом, применение регрессионного анализа позволяет построить кусочно-линейные модели и добиться локально-оптимальных описаний океанологических характеристик. Примером такого описания может быть опыт регрессионного анализа термохалинных полей океана [25], в результате которого получен набор регрессионных зависимостей температуры и солености для различных глубин и регионов Мирового океана.

Другой пример связан с классической задачей регрессионного описания вертикального профиля того или иного океанологического параметра. Сложная стратификация океана, наличие нескольких структурных зон, внутри которых существенно различаются физические механизмы формирования вертикальных распределений океанологических характеристик приводят к существенному усложнению обобщенных регрессионных моделей профилей

температуры, солености, плотности. Поэтому широко распространены кусочно-линейные модели, в которых параметры уравнений регрессии подбираются раздельно для поверхностной и глубинной структурных зон [54, 55, 89, 106, 166].

При статистическом описании временной изменчивости океанологических характеристик исследователи вынуждены детально описывать неоднородности, вызванные сезонным ходом большинства процессов верхнего слоя океана. Соответственно подбираются локальные по времени регрессионные уравнения для описания особенностей вертикального распределения параметров верхнего слоя вод и взаимосвязей основных термодинамических характеристик атмосферы и океана [47, 54].

Особая сфера применения множественного регрессионного анализа принадлежит океанологическим прогнозам. Наиболее часто множественная регрессия рассматривается как форма описания асинхронных взаимосвязей зависимой и независимых переменных. Составление уравнения происходит так, чтобы текущее колебание независимых переменных предопределяло будущую изменчивость зависимой. Тем самым координата «время» трансформируется в координатную ось сопряженных по фазе колебаний — в фазовое пространство. Эмпирический подбор степени асинхронности переменных обеспечивает совпадение фаз колебаний и, следовательно, достоверную оценку корреляционно-регрессионных параметров [1, 61, 104].

Подобная методология хорошо зарекомендовала себя в прогнозе тепловых, динамических и ледовых характеристик состояния океана [62, 65, 102, 130, 219].

В процессе подбора лучших независимых переменных для оптимального регрессионного уравнения исследователь нередко наталкивается на странный парадокс: даже большой набор физически вполне обоснованных независимых переменных не гарантирует оптимальности регрессионного уравнения. Причины такого парадокса лежат в невыполнении основополагающих требований при составлении регрессионных моделей.

В целом ряде случаев природная взаимообусловленность океанологических процессов порождает линейную статистическую взаимосвязь независимых переменных между собой. Этот факт противоречит изначальному ограничению модели множественной регрессии. Его неучет приводит к вычислительной неустойчивости, а в конечном счете — к малоэффективной оценке регрессионных параметров и низкой адекватности регрессионного уравнения в целом.

Второй аспект той же проблемы связан с чисто океанологическими трудностями, вызванными малым числом исходных данных. В ряде случаев на практике число наблюдений оказывается равным числу включаемых в уравнение независимых переменных или даже не превосходит последнего. В результате переопределенность решения не позволяет добиться состоятельных оценок регрессионных параметров, и модель теряет свою ценность. Как правило,

в подобных регрессионных моделях не удастся интерпретировать смысл описываемых регрессионных взаимосвязей.

Очевидно, необходимо максимально тщательно отобрать исходные переменные еще до начала оценивания регрессионных параметров. Однако что делать, если «плохие» переменные все же просочились в модель? В этом случае выбор наилучшего регрессионного описания осуществляется последовательным перебором всех возможных вариантов исходных независимых переменных. Метод перебора — великое множество. Это пошаговая регрессия, методики выметания и включения, направленного выбора и т. п. [3, 15, 122, 140, 144, 157, 158, 163].

Общими для всех перечисленных методик остаются критерии оценки оптимальности уравнения регрессии: высокая адекватность модели по дисперсионному критерию Фишера, максимальное описание изменчивости переменных по коэффициенту детерминации и достоверность оценок коэффициентов регрессии по критерию Стьюдента.

1.2. Метод канонического корреляционного анализа

Постановка задачи корреляционного анализа

Практика современных океанологических исследований нередко ставит задачу количественной оценки взаимосвязи гидрометеорологических процессов в различных акваториях Мирового океана.

Так, например, изучение взаимосвязей в режиме гидрометеорологических полей Черного и Каспийского морей, в первую очередь, требует оценки степени сопряженности колебаний в различных точках исследуемых акваторий. Набор временных рядов температуры (солености, уровня и т. п.) можно проанализировать по результатам вычисления простых парных коэффициентов корреляции между параметрами пар точек, принадлежащих различным акваториям. Однако такой путь позволит выявить лишь локальные связи, но не даст возможности обобщить их для исследуемых акваторий в целом.

Метод канонического корреляционного анализа (МККА) направлен как раз на то, чтобы описать уровень обобщенной взаимосвязи двух (или нескольких) целостных объектов исследования по результатам совместного анализа совокупности наблюдаемых параметров. Математическая логика МККА сводится к поиску таких линейных ортогональных преобразований исходных признаков, которые одновременно описывали бы наиболее общие черты для каждого объекта и в то же время по возможности высоко коррелировались между собой. Подобные преобразования называются каноническими переменными. А корреляция между ними называется канонической.

Очевидно, что идея МККА во многом перекликается с ранее рассмотренным подходом в МГК. Принципиальное отличие состоит

в том, что в МККА исследователь сталкивается с необходимостью ортогонального преобразования так называемой блочной матрицы исходных данных, каждый блок которой — это выборка наблюдений для одного из объектов исследования.

Пусть имеются исходные данные в виде двух наборов X и Y из наблюдений за M_1 и M_2 переменными, соответственно. Наборы

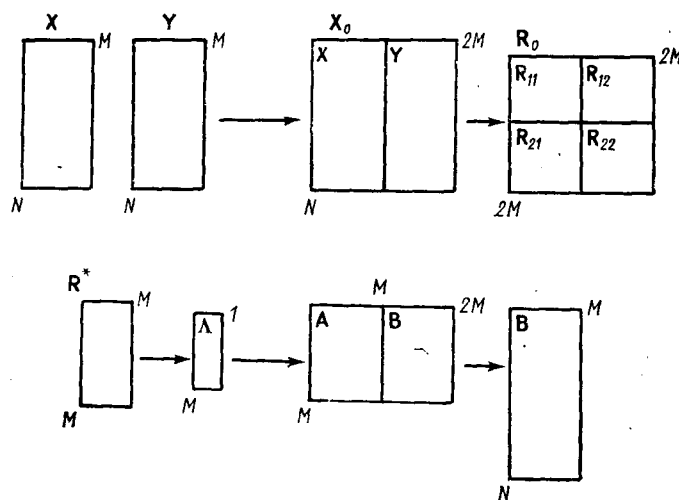


Рис. 1.2. Схема общей последовательности решения задачи канонического корреляционного анализа двух выборок наблюдений X и Y .

представим в виде прямоугольных матриц, которые состоят из N строк и M_1 и M_2 столбцов (рис. 1.2). Объединим матрицы в одну обобщенную матрицу данных, состоящую из N строк и $M = M_1 + M_2$ столбцов. Обобщенная матрица данных при этом состоит как бы из двух блоков: 1-й блок — матрица X , 2-й блок — матрица Y . Для оценки взаимосвязи переменных вычислим коэффициенты парной корреляции между всеми переменными, входящими в обобщенную матрицу данных.

В результате вычислений получим обобщенную корреляционную матрицу R , состоящую из четырех блоков: 1-й блок — R_{11} — характеризует коррелированность переменных внутри матрицы X , 2-й блок — R_{22} — характеризует связность переменных внутри матрицы Y , 3-й и 4-й блоки характеризуют степень коррелированности между переменными матриц X и Y — R_{12} и R_{21} .

Суть МККА состоит в поиске максимально коррелированных между собой линейных комбинаций исследуемых групп переменных вида:

$$U_i = A_i X, \quad V_i = B_i Y, \quad (1.4)$$

где U_i и V_i — канонические переменные, A_i и B_i — коэффициенты перехода к каноническим переменным от исходных выборок.

В процессе вычислений соблюдаются следующие условия:
 — математическое ожидание канонических переменных равно нулю:

$$\bar{U}_i = \bar{V}_i = 0,$$

— дисперсия канонических переменных нормирована и равна 1:

$$(\bar{U}_i^2) = (\bar{V}_i^2) = 1.$$

Взаимосвязь канонических переменных двух выборок определяется значением коэффициента канонической корреляции ρ_i :

$$\rho_i = (\bar{U}_i, \bar{V}_i),$$

который удовлетворяет условию максимизации этой связи:

$$\rho_i = \mathbf{A}_i^T \mathbf{R}_{12} \cdot \mathbf{B}_i = \mathbf{B}_i^T \cdot \mathbf{R}_{21} \cdot \mathbf{A}_i \rightarrow \max.$$

После того как найдена первая пара канонических переменных, описывающих наиболее высоко связанную долю дисперсии исследуемых выборок, переходят к вычислению последующих пар канонических переменных. При этом соблюдается условие максимального объяснения оставшейся межгрупповой корреляции при одновременном соблюдении условия некоррелированности новых канонических переменных U_{i+1} , V_{i+1} с предшествующими. Число вычисляемых таким образом пар канонических переменных не превышает наименьшего числа переменных в наблюдаемых выборках.

На каждом этапе вычисления подлежат оценке и канонические корреляции ρ_i .

Алгебраическая проблема МККА сводится к решению системы матричных уравнений вида

$$\begin{bmatrix} -\mathbf{L} \cdot \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{R}_{21} & -\mathbf{L} \cdot \mathbf{R}_{22} \end{bmatrix} \cdot \begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix} = 0, \quad (1.5)$$

где $\mathbf{R}_{ij}$ — блоки обобщенной корреляционной матрицы, $\mathbf{L}$ — множитель Лагранжа, $\mathbf{A}$ и $\mathbf{B}$ — искомые коэффициенты перехода.

Приведенная матричная запись может быть переписана в виде

$$\begin{cases} -\mathbf{L} \cdot \mathbf{R}_{11} \cdot \mathbf{A} + \mathbf{R}_{12} \cdot \mathbf{B} = 0, \\ \mathbf{R}_{21} \cdot \mathbf{A} - \mathbf{L} \cdot \mathbf{R}_{22} \cdot \mathbf{B} = 0. \end{cases}$$

Из этой системы уравнений получается основное уравнение МККА:

$$|\mathbf{R}^* - \Lambda \mathbf{I}| = 0. \quad (1.6)$$

Множитель $\mathbf{R}^* = \mathbf{R}_{11}^{-1} \cdot \mathbf{R}_{12} \cdot \mathbf{R}_{22}^{-1} \cdot \mathbf{R}_{21}$ представляет собой квадратную несимметричную корреляционную матрицу размерностью M строк и M столбцов. В результате процесс решения сводится к уже рассматриваемой выше проблеме собственных чисел и собственных векторов матрицы $\mathbf{R}^*$.

Раскрывая определитель, стоящий в скобках выражения (1.6), можно вычислить собственные числа Λ , подстановка которых

в данное уравнение позволяет решить его относительно собственных векторов A . Вычисление канонической переменной проводится по формуле (1.4), а соответствующий этой переменной канонический коэффициент корреляции оценивается по формуле

$$\rho_i = \sqrt{\lambda_i}. \quad (1.7)$$

Тем самым однозначно определяется вид канонической переменной для второго набора:

$$B = \frac{1}{\lambda} \cdot R_{22}^{-1} \cdot R_{21} \cdot A. \quad (1.8)$$

Вычислительная схема МККА

Процедура проведения канонического корреляционного анализа может быть разбита на несколько основных этапов, которые могут легко программироваться [34, 73, 77].

1. Вычисление парных коэффициентов корреляции для всех наблюдаемых переменных во всех выборках и составление обобщенной корреляционной матрицы R .

2. Разбиение матрицы R на блоки и расчет вспомогательной матрицы:

$$R^* = R_{11}^{-1} \cdot R_{12} \cdot R_{22}^{-1} \cdot R_{21}.$$

3. Расчет собственных чисел λ вспомогательной матрицы R^* на основании характеристического уравнения, получаемого из условия равенства нулю определителя матрицы:

$$|R^* - \lambda I| = 0.$$

4. Расчет собственных векторов (A) вспомогательной матрицы R^* путем решения системы уравнений вида

$$A \cdot (R^* - \lambda I) = 0.$$

5. Вычисление коэффициентов собственного вектора (B) для второго набора Y по формуле

$$B = \frac{1}{\lambda} \cdot R_{22}^{-1} \cdot R_{21} \cdot A.$$

6. Вычисление коэффициентов канонической корреляции (ρ) по формуле

$$\rho_i = \sqrt{\lambda_i}.$$

7. Оценка M значений канонических переменных U и V :

$$U = A \cdot X,$$

$$V = B \cdot Y.$$

8. Оценка относительной информативности канонических переменных в описании общей изменчивости исследуемых переменных:

$$d_i = \lambda_i / \sum_{j=1}^M \lambda_j.$$

Применение ККА в океанологии

Канонический корреляционный анализ для океанологических процессов и полей применялся пока весьма осторожно. На основе анализа имеющихся в океанологической литературе публикаций по МККА трудно сделать однозначный вывод о причинах такой осторожности, тем более, что двадцать лет назад в работах Н. А. Багрова [15—19] и Г. А. Карпеева [90, 91] был проведен тщательный сравнительный анализ преимуществ применения этого метода для решения гидрометеорологических задач. В частности, было показано, что каноническое преобразование позволяет сохранить «прогностическую информацию» без искажений, а роль МККА в прогностике во многом эквивалентна роли МГК при решении диагностических задач.

Позднее сделанные выводы неоднократно подтверждались результатами эмпирических расчетов по анализу взаимосвязей метеорологических и океанологических полей [58, 230], а также при выявлении прогностических зависимостей этих полей [22, 134, 148, 203, 253].

Рассматривая в целом все имеющиеся результаты канонического анализа в океанологии, можно наметить обобщенную схему применения МККА:

- к исследованию привлекаются выборки в координатных осях пространство—время;
- авторы стараются подобрать векторы наблюдений одной (или близкой) размерности;
- число исследуемых выборок, как правило, равно двум;
- при решении диагностических задач исследуемые выборки синхронизированы по времени; при оценке прогностических зависимостей, как правило, рассматриваются асинхронные корреляционные моменты;
- интерпретация характеристик канонических разложений осуществляется в основном аналогично интерпретации главных компонент.

Этот метод недостаточно широко применяется в практике океанологических исследований, чтобы можно было раскрыть все его достоинства и недостатки уже сейчас. Остановимся лишь на основных из них, известных на сегодняшний день.

1. Наличие большого числа высоких коэффициентов парной корреляции в обобщенной матрице данных R_0 приводит к затруднениям обращения этой матрицы при решении проблемы собственных чисел. В связи с чем ухудшается точность оценивания канонических корреляций и самих канонических переменных.

2. Величина канонической корреляции в классическом МККА является, по определению, всегда положительной, поскольку оценивается из выражения (1.7). Очевидно, что реальный знак взаимосвязей наборов исходных переменных отнюдь не всегда положителен. Поэтому на практике исследователь вынужден, оценивая знак канонической корреляции, привлекать дополнительные оценки коэффициентов парной корреляции и общие физические соображения.

3. Основным недостатком постановки задачи канонического анализа следует считать его узкую ориентированность на определение показателей, в наибольшей степени объясняющих межгрупповую корреляцию наборов переменных, но не саму дисперсию исходных наборов. Вследствие этого при плохой сходимости разложения обобщенной корреляционной матрицы R_0 по собственным числам возникает ситуация неинформативности найденной зависимости ввиду малости дисперсии внутри исследуемых наборов, приходящейся на полученные канонические переменные.

Если вклад первых двух-трех канонических переменных относительно низок (менее 50—60 %), то целесообразно либо изменить исследуемую выборку, либо отказаться от данной методики.

4. Оценки канонических переменных оказываются весьма чувствительными к выполнению условий стационарности исходных исследуемых рядов. Поэтому перед проведением МККА целесообразно убедиться в выполнении гипотезы стационарности и при необходимости провести соответствующие преобразования исходных данных.

2.1. Метод главных компонент

Постановка задачи, основное уравнение

Идея метода главных компонент (МГК) состоит в представлении сложного процесса как суммы более простых взаимонезависимых составляющих (членов разложения). Однако в отличие от разложений по полиномам Чебышева или Фурье в МГК структура параметров разложения наперед не задается. Она однозначно и объективно определяется только внутренним содержанием исследуемого процесса и уровнем коррелированности анализируемых рядов наблюдений между собой. Поэтому метод главных компонент часто называют еще разложением по естественным (или эмпирическим) ортогональным функциям [155].

Правила МГК таковы, что каждый член разложения вбирает в себя сходную изменчивость нескольких исходных переменных и в обобщенном виде представляет ее как собственную изменчивость. В этом смысле МГК выступает как инструмент эффективного сжатия информации и такого ее обобщения, которое позволяет в малом числе рассчитанных параметров разложения отобразить и проанализировать целостную изменчивость разнообразных характеристик состояния природных систем. При этом из разрозненных в общем-то наблюдений удается выделить то общее, которое позволяет восстановить сущность происходящих изменений, нередко скрытых непосредственно от глаз наблюдателя.

Исходные данные наблюдений для МГК формируются в виде таблиц, расположенных в координатах: пространство—время, пространство—признаки или время—признаки. В каждом конкретном случае возникает свой уровень обобщения материала.

В координатах пространство—время с помощью МГК удается выявить основные масштабы колебаний в пространстве и во времени данной исследуемой переменной. В координатах пространство—признаки основное внимание уделяется описанию сферы влияния каждого из признаков и возможности обобщения их сопряженной изменчивости. В координатах время—признаки преследуется цель обобщенного описания объекта исследования как системы, зависящей от множества разнородных параметров.

Возникает вопрос: зачем переходить от уже имеющихся данных наблюдений к каким-то новым переменным? Дело в том, что для подавляющего большинства реальных ситуаций наблюдаемая изменчивость параметров не позволяет исследователю досконально разобраться в сущности происходящих процессов. В данных

всегда содержатся шумовые компоненты, точно не известны ни пространственные, ни временные масштабы изменчивости, часть параметров наблюдения может быть связана между собой и описывать один и тот же процесс. Поэтому и ставится задача разложения сложного процесса на простые колебания таким образом, чтобы: 1) сохранив общую точность описания перейти к новым взаимонезависимым переменным, 2) каждая из новых переменных объединяла в себе одну из наиболее важных тенденций в исходных данных, 3) все новые переменные были упорядочены по степени уменьшения их вкладов (влияний) в общую изменчивость.

Очень важно, что вся последовательность столь важных проблем решается технически просто — линейным преобразованием исходных данных или, другими словами, путем поэлементного умножения каждого признака на коэффициенты линейной связи данного признака с новой переменной, отвечающей заданным свойствам. Основной вопрос МГК состоит как раз в том, чтобы наилучшим образом найти эти коэффициенты связи исходных переменных и главных компонент.

Математическая формулировка модели МГК состоит в следующем. Пусть имеется набор данных из N наблюдений за M переменными. Это равнозначно наличию M векторов наблюдений вида $\mathbf{x} = \{x_1, x_2, x_3, \dots, x_M\}$, которые образуют матрицу данных $\mathbf{X}$ из N строк и M столбцов. Матрице данных в соответствие ставится матрица значений главных компонент $\mathbf{F}$, состоящая также из N строк и M столбцов. Каждый столбец в матрице $\mathbf{F}$ соответственно описывает изменение одной главной компоненты f_j , которая представляется в виде вектора $\mathbf{f}_j = \{f_{1j}, f_{2j}, f_{3j}, \dots, f_{Nj}\}$. Число переменных M и число главных компонент совпадают. Число наблюдений N исходных данных и число значений каждой компоненты совпадают. Тогда основное уравнение МГК приобретает вид

$$\mathbf{X} = \mathbf{F} \cdot \mathbf{A}^T. \quad (2.1)$$

Здесь матрица $\mathbf{A}$ представляет собой матрицу коэффициентов связи переменных и главных компонент. Матрица $\mathbf{A}$ содержит M строк и M столбцов. Каждый столбец матрицы $\mathbf{A}$ содержит коэффициенты связи данной компоненты со всеми исходными признаками. Согласно правилам матричной алгебры, матрица $\mathbf{A}$ в уравнении (2.1) транспонирована $\mathbf{A}^T$ (т. е. повернута на 90°).

Раскрывая содержание уравнения (2.1), можно записать формулировку МГК для i -го наблюдения j -й переменной в матрице данных:

$$x_{ij} = \sum_{k=1}^M a_{ik} f_{kj}. \quad (2.2)$$

Геометрическая интерпретация МГК

Наиболее просто рассмотреть этот вопрос на примере анализа всего двух исходных переменных. Пусть это будут два ряда наблюдений за температурой воды на двух различных горизонтах.

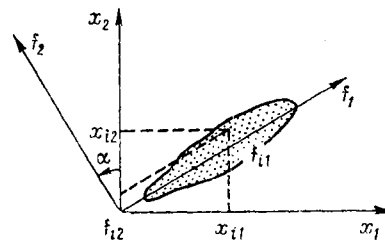
Облако наблюдений представим в координатах: ось абсцисс—изменение температуры на первом горизонте (x_1), ось ординат—изменение температуры на втором горизонте (x_2). Если облако изменчивости имеет вид эллипса (рис. 2.1), то это указывает на наличие связанных колебаний температуры воды на исследуемых горизонтах.

Суммарная изменчивость колебаний, соответствующая главной диагонали эллипса, может быть оценена по размаху колебаний на каждом горизонте:

$$\sigma_{\Sigma}^2 = \sigma^2(x_1) + \sigma^2(x_2).$$

Рис. 2.1. Геометрическая интерпретация метода главных компонент.

x_1, x_2 — оси исходных признаков; f_1, f_2 — оси главных компонент; α — угол поворота осей.



Главные компоненты можно рассматривать как оси новой ортогональной системы координат. Тогда процесс поиска главных компонент следует интерпретировать как процедуру поворота координатных осей. Переориентация осей производится по следующим правилам: первая ось (главная компонента) направляется вдоль оси максимального рассеяния наблюдений, вторая ось указывает направление максимального рассеяния остатков наблюдений за вычетом влияния предыдущей компоненты и при этом обязательно будет перпендикулярна первой оси.

При таком подходе переход от координат признаков x_1, x_2 к координатам компонент f_1, f_2 осуществится простым поворотом системы координат на угол α против часовой стрелки, в результате чего изменится проекция наблюдений на оси компонент. Очевидно, что дисперсия проекций наблюдений на ось первой компоненты будет существенно больше, чем на вторую компоненту, хотя суммарная дисперсия не изменится (потому что конфигурация облака в силу линейности преобразования координат не изменилась):

$$\sigma_{\Sigma}^2 = \sigma^2(f_1) + \sigma^2(f_2).$$

Данная геометрическая интерпретация позволяет по-новому понять содержание собственных векторов. Матрица собственных векторов A содержит в себе коэффициенты перехода от исходных признаков к главным компонентам. С точки зрения тригонометрии эти коэффициенты перехода по сути косинусы и синусы углов поворота главных компонент относительно исходных признаков.

Отсюда легко записать матрицу A в тригонометрическом виде:

$$A = \begin{bmatrix} \cos \alpha & -\sin \alpha \\ \sin \alpha & \cos \alpha \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

Приводимая геометрическая интерпретация МГК показывает, как в данной методике происходит процесс сжатия информации. Это выражается в ориентации первых главных компонент вдоль осей максимального рассеяния облака наблюдений. В результате большая часть изменчивости исследуемого процесса оказывается сосредоточенной в первых компонентах, которые и носят название главных. А каждая последующая компонента описывает, по определению, все меньшую дисперсию колебаний, и чем выше номер компоненты, тем менее она информативна.

Свойства главных компонент

Они могут быть сформулированы на основе анализа основного уравнения метода и его геометрической интерпретации следующим образом:

1) ГК между собой линейно независимы (ортогональны); следовательно, их попарная корреляция равна 0:

$$r(f_i, f_j) = 0, \quad \cos(f_i, f_j) = 0;$$

2) компоненты описывают изменчивость данных таким образом, что первая ГК описывает максимальную дисперсию облака наблюдений, вторая ГК ортогональна первой и описывает максимальную остаточную дисперсию, третья ГК ортогональна предыдущим ГК и описывает максимальную остаточную дисперсию и т. д.:

$$\sigma^2(f_1) \geq \sigma^2(f_2) \geq \dots \geq \sigma^2(f_n);$$

3) применение МГК позволяет оптимальным образом среди всех других линейных преобразований описать дисперсию исследуемых переменных;

4) использование ГК позволяет описать исходную информацию с минимальным искажением геометрической структуры облака наблюдений в пространстве ГК;

5) вклад ГК в описание общей дисперсии переменных пропорционален квадрату соответствующего ей собственного числа корреляционной матрицы исходных признаков;

6) вычисляемые главные компоненты оптимальным образом описывают изменчивость исходных переменных, что непосредственно следует из максимизации функционала:

$$V = \sum_{k=1}^M \sum_{j=1}^M r^2(f_k, x_j) \rightarrow \max,$$

где $r(f_k, x_j)$ есть коэффициент корреляции j -го исходного признака и k -й компоненты.

Математический аппарат МГК

Как было показано, основное уравнение МГК записывается в матричной форме:

$$\mathbf{X} = \mathbf{F} \cdot \mathbf{A}^T.$$

Задача МГК сформулирована выше как задача поиска линейных преобразований f_i со свойствами:

— максимального исчерпания дисперсии каждой ГК $\overline{(f_i^2)} \rightarrow \max$,

— взаимонезависимости между ГК $\overline{(f_i, f_k)} = 0$.

Для поиска ГК требуется прежде всего определить коэффициенты связи каждого j -го признака с каждой k -й компонентой, которые образуют матрицу $\mathbf{A}$ —весов главных компонент (или коэффициентов разложения). Это становится возможным, если принять условие полного восстановления корреляции исходных признаков по матрице $\mathbf{A}$:

$$\mathbf{R} = \mathbf{A} \cdot \mathbf{A}^T, \quad (2.3)$$

где $\mathbf{R}$ — матрица корреляции исходных переменных размерностью M строк на M столбцов.

В этом случае вся процедура МГК сводится практически к поиску собственных векторов корреляционной матрицы признаков, ибо условие (2.3) выполняется только в случае, если матрица $\mathbf{A}$ составлена из M собственных векторов корреляционной матрицы $\mathbf{R}$.

Процесс поиска собственных векторов начинается с оценки собственных чисел корреляционной матрицы путем решения характеристического уравнения:

$$|\mathbf{R} - \lambda \mathbf{I}| = 0. \quad (2.4)$$

Смысл решения матричного уравнения (2.4) состоит в поиске M корней характеристического уравнения для определителя вида

$$\begin{vmatrix} r_{11} - \lambda & r_{12} & r_{13} & \dots & r_{1M} \\ r_{21} & r_{22} - \lambda & r_{23} & \dots & r_{2M} \\ r_{31} & r_{32} & r_{33} - \lambda & \dots & r_{3M} \\ \dots & \dots & \dots & \dots & \dots \\ r_{M1} & r_{M2} & r_{M3} & \dots & r_{MM} - \lambda \end{vmatrix} = 0,$$

который приравняется к нулю.

Второй этап — решение систем линейных уравнений для определения матрицы собственных векторов:

$$\mathbf{A} \cdot (\mathbf{R} - \lambda \mathbf{I}) = 0. \quad (2.5)$$

Матричное выражение (2.5) последовательно раскрывается M раз в виде системы M уравнений вида

$$\begin{cases} a_{i1}(r_{11} - \lambda_i) + a_{i2}r_{12} + \dots + a_{iM}r_{1M} = 0, \\ a_{i1}r_{21} + a_{i2}(r_{22} - \lambda_i) + \dots + a_{iM}r_{2M} = 0, \\ \dots \\ a_{i1}r_{M1} + a_{i2}r_{M2} + \dots + a_{iM}(r_{MM} - \lambda_i) = 0. \end{cases}$$

Каждая система содержит соответствующее значение собственного числа (λ_i) и при решении дает M корней — значений i -го собственного вектора, которые составляют столбцы матрицы A .

Дополнительным условием поиска собственных векторов ставится условие нормированности дисперсии элементов каждого собственного вектора:

$$\sum_{j=1}^M a_{ij}^2 = 1$$

для каждого из $i = 1, \dots, M$ векторов.

Следует иметь в виду, что задачи вычисления собственных (характеристических) чисел и собственных (характеристических) векторов являются традиционными задачами линейной алгебры, не представляют собой ничего сложного и проделываются с известной легкостью уже многие десятилетия [3, 30, 34, 81, 86, 94].

После поиска собственных чисел и собственных векторов A осталось решить две задачи. Первая из них — оценить важность каждой компоненты. Эта оценка непосредственно следует из факта равенства дисперсии каждой ГК соответствующему ей собственному числу. Потому относительный вклад i -й компоненты в описание общей дисперсии переменных описывается формулой:

$$d_i = \frac{\lambda_i}{\sum_{j=1}^M \lambda_j}. \quad (2.6)$$

Последняя необходимая подробность в МГК — оценка самих значений ГК, т. е. векторов f_j . По традиции это делается путем использования классической линейной регрессии. Причем матрица коэффициентов регрессии (B) вычисляется на основе вектора собственных чисел (Λ) и матрицы собственных векторов (A):

$$B = \Lambda^{1/2} C,$$

где матрица C является обратной по отношению к матрице собственных векторов A^T : $C = (A^T)^{-1}$. В поэлементной записи эта формула примет вид

$$b_{ij} = \sqrt{\lambda_j} c_{ij}.$$

В общем виде выражение для восстановления матрицы главных компонент выглядит следующим образом:

$$F = X \cdot B, \quad (2.7)$$

или в покомпонентной записи для i -го значения j -й компоненты:

$$f_{ij} = \sum_{k=1}^M x_{ik} b_{kj},$$

где $k = 1, \dots, M$ — номер исходной переменной.

Применение МГК в океанологии

По сложившейся в прикладной океанологии традиции МГК широко применяется при исследовании прежде всего пространственно-временной структуры процессов, причем наиболее часто — для анализа изменчивости тепловых условий на поверхности океана [20, 60, 67, 105, 111, 265]. Метод успешно применяется для разложения колебаний как по пространству, так и по времени. Закономерности временной изменчивости главных компонент представляют собой обобщенные тенденции колебаний исследуемых полей, что успешно используется для прогностических целей [6, 61, 64, 71, 91]. (Подробнее этот аспект рассматривается в главе 6.)

Особая область применения классического МГК связана в океанологии с проблемой выделения структуры вод. Решение проблемы проводится традиционно в координатных осях пространство—признаки. Признаками выступают наблюдаемые значения гидрофизических и гидрохимических характеристик. Цель применения МГК состоит в том, чтобы через главные компоненты описать наиболее существенные взаимосвязи океанологических характеристик и в географических координатах определить структуру исходных полей. Под элементами структуры при таком подходе понимаются основные водные массы [32, 37, 99, 207], вертикальная расслоенность вод океана [149, 201], типы вод и фронты [96], вихревые динамические образования [48, 49, 112] и т. п. Конечный результат применения МГК однозначно определяется выбором набора исследуемых признаков, их расположением в пространстве и масштабом осреднения [114, 138, 183, 184, 202, 223, 236, 241, 249, 251, 264, 271, 278].

Одна из ведущих проблем, дискутируемая в процессе применения МГК, состоит в оценке числа членов разложения, оптимального для физической интерпретации и количественного описания свойств исследуемых полей. Наиболее часто используется методика, основанная на оценке потенциальной ошибки при расчетах параметров разложения [59, 66, 68, 73, 114, 118, 214, 236, 240, 258]:

$$\delta\lambda_j = \lambda_j \sqrt{2/N}.$$

Интерпретируются лишь те компоненты, для которых $\lambda_j > \delta\lambda_j$. Подобный критерий в целом неплохо зарекомендовал себя в прикладных исследованиях. Однако в ряде сложных случаев оказался малоэффективным в силу того, что дисперсия шумовых колебаний превышает дисперсию циклической компоненты. Для таких случаев используется так называемая Q -статистика [81, 255, 256, 257]:

$$Q = \sum_{\tau=1}^M \frac{r^2(\tau)}{\sigma^2(\tau)},$$

где $r(\tau)$ — оценка автокорреляционной функции главной компоненты времени на сдвиге τ , $\sigma^2(\tau)$ — выборочная дисперсия этой оценки.

Сравнение рассчитанного значения Q -статистики с χ^2 — критерием Пирсона позволяет включить в анализ те компоненты, которые в своей структуре содержат статистически значимые колебания. Идея о содержательном анализе тех компонент, которые имеют значимую структуру автокорреляционных функций, успешно используется при изучении колебаний термических и ледовых характеристик [102, 128, 154, 175].

Однако исчерпать все проблемы анализа сложных пространственно-временных колебаний гидрометеорологических процессов в рамках классического МГК не удалось, что и послужило поводом к развитию фундаментальных теоретических основ метода.

Техника ГК в частотной области представляется уже более совершенной по сравнению с традиционным МГК [126, 270]. Она базируется на принципах получения ортогонального представления матрицы кросспектральных функций нескольких временных рядов. Получаемое решение содержит уже фазовую структуру элементарных волновых колебаний, которая описывается с помощью представления собственных векторов как наборов комплексных чисел.

Техника ГК в частотной области основывается на понятии кроссковариационной матрицы $C(\tau)$, которая составлена из коэффициентов кросскорреляции $c_{ij} = \overline{(x_i(t + \tau), x_j(t))}$. Вычисляемая на основе $C(\tau)$ матрица спектральной плотности порядка $M \times M$

$$S = \sum_{\tau=-\infty}^{\infty} c(\tau) \exp(-i\tau\lambda)$$

имеет M собственных чисел λ , связанных с комплексными ортогональными собственными векторами A .

Прекрасный пример успешного применения такой техники ГК содержится в работе [275], где исследовалась волновая структура крупномасштабных долгопериодных аномалий температуры воды в северной части Тихого океана. В результате статистического анализа удалось выявить и описать процесс распространения аномалий температуры из юго-западной части исследуемого района в направлении на северо-восток.

Одно из наиболее перспективных направлений развития методологии главных компонент связано с техникой комплексных ГК. Причина этого заключается в стремлении исправить один из основных недостатков классического МГК, который связан с невозможностью надежного описания прогрессивных, движущихся в пространстве волн.

Классический МГК позволяет достоверно охарактеризовать стоячие волны. Если же в исследуемой пространственно-временной структуре наблюдений имеются прогрессивные волны, то МГК искаженно представит их в виде набора стоячих колебаний. Путь к описанию пространственно-временных волн в МГК лежит через изменение методики составления корреляционной (ковариационной) матрицы R . Первый шаг был сделан, когда матрица R была

составлена из асинхронных коэффициентов кросскорреляционных функций [231, 232, 241, 273, 274].

Однако эффективность такой методики не всегда высока в силу сложности подбора оптимальных коэффициентов кросскорреляции.

Разработка методологии комплексного анализа ГК (КГК), по-видимому, является сейчас высшей ступенью развития метода главных компонент [109, 235]. Для применения КГК необходимо осуществить гильбертово преобразование исходных временных рядов, после чего на основе комплексного представления временных рядов выполняется расчет их комплексных кросскорреляционных функций, которые объединяются в корреляционную матрицу. Получаемые затем комплексные собственные векторы и главные компоненты характеризуют амплитудно-фазовую структуру исследуемых колебаний в пространстве и во времени.

Опыт применения этого метода в океанологических исследованиях весьма ограничен, что не позволяет оценить возможные скрытые недостатки комплексных главных компонент.

2.2. Метод факторного анализа

Постановка задачи и основное уравнение

Эффективность исследования природных процессов с помощью метода главных компонент была одной из главных причин развития фундаментальных и теоретических основ принципа ортогонального разложения. Развитие МГК в естествознании шло по пути добавления к традиционной схеме метода дополнительных вычислительных этапов, направленных на решение одной из перечисленных выше проблем компонентного анализа.

В результате на сегодняшний день в многомерной статистике оформилась специализированная область ортогонального преобразования данных, которая носит название метода факторного анализа (МФА). Современный МФА, развиваясь на базе классического МГК, представляет собой сложный комплекс вычислительных процедур, схема которого напоминает ветвящееся дерево.

Исходя из целей и задач исследования, особенностей исходных данных ученые с помощью ЭВМ выбирает из множества вариантов решения проблемы ортогонального разложения по МФА тот, который представляется оптимальным в данных условиях и позволяет достичь наиболее простого и физически понятного результата.

Идеология факторного анализа базируется на предположении о том, что регистрируемые в природе характеристики не являются сами по себе причинами происходящих изменений. Они скорее представляют собой следствия или индикаторы влияния внешних и внутренних сил, скрытых от наблюдателя за сложной динамикой взаимосвязей наблюдаемых переменных. Каждая из влияющих сил воздействует сразу на несколько наблюдаемых переменных. И, хотя переменные по-своему реагируют на вынуждающие воздей-

ствия, их междоусобная коррелированность содержит то общее зерно, которое и описывает изменение влияющей силы. Кроме этих общих сил или факторов (как их принято называть в МФА), в природе присутствуют и другие вынуждающие силы, которые являются частными для каждого из наблюдаемых признаков и не влияют на другие. Подобные факторы называются специфическими, поскольку отражают специфику изменения данной конкретной переменной. Дополнительно в наблюдения и измерения всегда вкрадываются случайные ошибки и неточности, которые именуются случайными факторами.

Обобщая сказанное, можно представить общую изменчивость наблюдаемых признаков в виде трех слагаемых:

$$X = F \cdot A^T + E,$$

где X — изменчивость наблюдаемых признаков, $F \cdot A^T$ — изменчивость скрытых общих факторов, E — изменчивость специфических факторов плюс изменчивость случайных факторов.

На практике весьма трудно разделить сферы влияния специфических и случайных факторов и описать их отдельно. Для этого требуется весьма тонкий анализ данных. Поэтому в классическом МФА изменчивость специфических и случайных факторов рассматривается совместно как единое целое. Это ограничение несколько загроубляет результат, но позволяет сосредоточиться на описании основных общих факторов, формирующих генеральную изменчивость признаков.

Обращаясь к такой сложной многосвязной системе, какой является Мировой океан, можно найти массу натуральных примеров, соответствующих приведенной схеме взаимодействия факторов. Так, глобальное поле температуры поверхности океана будет находиться под влиянием прежде всего гелио-геофизических сил, формирующих суточный, сезонный и вековой ход температуры. Поэтому эти силы могут рассматриваться как факторы, общие для всего океана. Одновременно каждая точка океана имеет локальные особенности взаимодействия с атмосферой и нижележащими слоями океана, которые могут описываться как специфические факторы.

Наконец, точность наблюдений температуры в океане весьма неодинакова по пространству и существенно зависит от вида измерительных приборов и условий измерений. Эта группа причин изменения температуры может описываться влиянием случайных факторов.

Все содержание методики МФА направлено как раз на то, чтобы с максимальной точностью и достоверностью проверить гипотезу о наличии малого числа влияющих факторов и получить их статистическое описание. Принимаемая в факторном анализе схема формирования изменчивости наблюдаемых переменных (x) под влиянием скрытых общих для нескольких переменных факторов (f) и индивидуальных, специфических ошибок наблюдения (e) представлена на рис. 2.2 [4, 73, 86].

Влияние факторов на наблюдаемые переменные осуществляется и оценивается на основе модельных коэффициентов линейной связи (a_{ij}), которые показывают степень связанности i -го общего фактора и j -й переменной. Запишем на основании приведенных рассуждений алгебраическое выражение модели МФА:

$$X = F \cdot A^T + E. \quad (2.8)$$

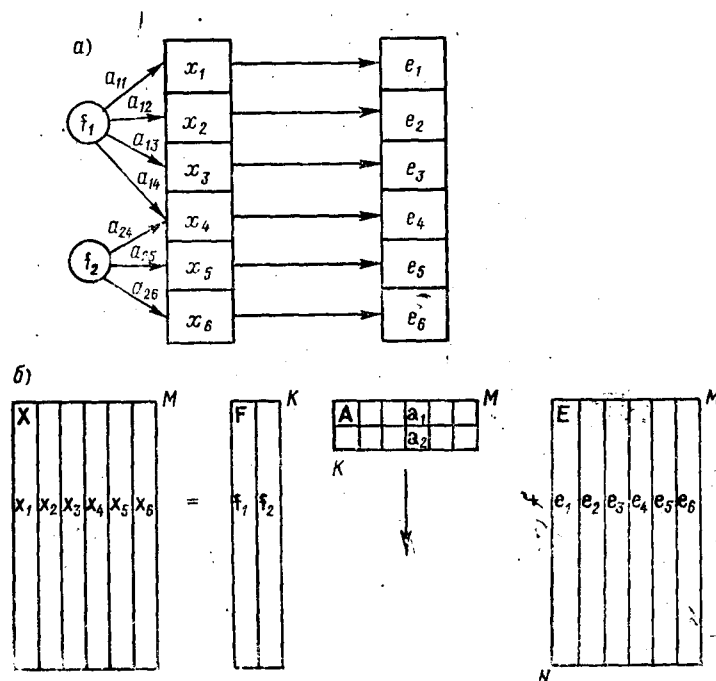


Рис. 2.2. Схема взаимосвязей переменных в модели факторного анализа для покомпонентного (а) и матричного (б) представления.

Здесь X — матрица исходных данных, состоящая из N наблюдений (строки) за M переменными (столбцы); F — матрица значений общих факторов, состоящая из N значений K факторов; A — матрица коэффициентов связи общих факторов и исходных переменных, составленная из M строк и K столбцов; E — матрица остатков или специфических факторов, содержащая N наблюдений за M переменными.

Для того чтобы получить единственное решение представленного уравнения, необходимо ввести некоторые дополнительные условия на вид матриц A , F и E . При формулировке этих условий традиционно предполагается, что из бесконечного числа математических моделей процессов, протекающих в природе, нам необходимо получить максимально простую модель, которая одновре-

менно вписывалась бы в корреляционную структуру матрицы исходных данных [94, 95].

Накладываемые условия можно сформулировать следующим образом:

— общие факторы $f_1, f_2, \dots, f_k$ должны быть взаимонезависимы (некоррелированы):

$$r(f_i, f_j) = 0 \text{ при } i \neq j;$$

— общие факторы $f_1, f_2, \dots, f_k$ должны быть нормированы (безразмерны) и приведены к единичной длине $\sigma^2(f_i) = 1$;

— общие факторы не должны быть связаны с ошибками и специфическими факторами

$$r(f_i, e_j) = 0 \text{ при любых } i, j;$$

— ошибки не должны коррелировать между собой

$$r(e_i, e_j) = 0 \text{ при } i \neq j;$$

— число общих факторов не должно превышать половину числа наблюдаемых переменных $K \leq M/2$;

— общий фактор должен описывать большую часть изменчивости как минимум двух переменных.

Переписывая основное уравнение МФА в покомпонентном виде, получаем:

$$x_{ij} = \sum_{p=1}^k f_{ip} a_{pj} + e_{ij}, \quad (2.9)$$

где x_{ij} — значение i -го наблюдения j -й переменной, f_{ip} — значения K общих факторов для i -го наблюдения, a_{pj} — веса (нагрузки) K общих факторов на j -ю переменную, e_{ij} — ошибка наблюдения или специфический фактор для i -го наблюдения j -й переменной.

Геометрическая интерпретация МФА

Геометрическая интерпретация факторного анализа несколько усложнена по сравнению с МГК вследствие более сложной вычислительной процедуры. Дело в том, что в МФА происходит проектирование облака исходных наблюдений из M -мерного пространства признаков в K -мерное пространство общих факторов (рис. 2.3).

Учитывая, что $K < M$, мы неизбежно приходим к выводу о появлении искажений облака наблюдений при его проектировании в пространство меньшей размерности. На первом этапе МФА используемая схема расчетов практически эквивалентна процедуре вращения исходных координат с целью ориентации осей общих факторов вдоль направлений максимальной коррелированности наблюдений (переход от x_1, x_2, x_3 к f_1, f_2, f_3 на рис. 2.3 а), что в целом соответствует процедуре МГК.

Далее осуществляется оптимизация факторного решения, для чего из последующих расчетов удаляются оси специфических факторов (f_3), изменчивость вдоль которых не выходит за рамки допустимых ошибок. Одновременно производится проектирование облака наблюдений на оставшиеся оси общих факторов (рис. 2.3 б). Другими словами, сокращаются те оси рассеяния, в проекции на которые исходное облако наблюдений представляется сферой с минимальной дисперсией наблюдений. Такое сокращение факторных

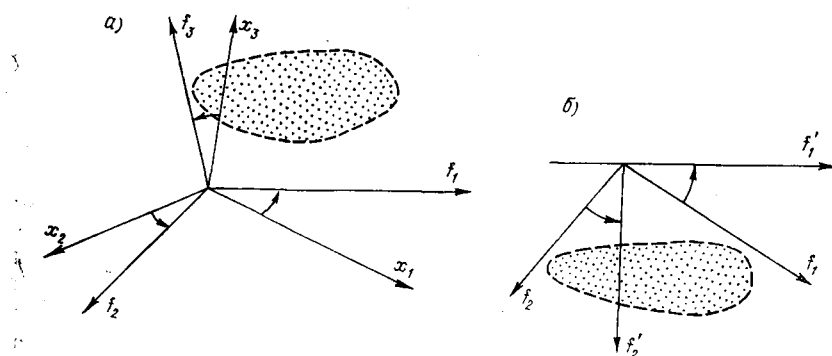


Рис. 2.3. Геометрическая интерпретация метода факторного анализа.

а) переход от осей исходных признаков x_1, x_2, x_3 к осям начальных факторов f_1, f_2, f_3 ; б) переход к общим факторам f_1, f_2, f_3 в результате ортогонального вращения осей начальных факторов.

осей позволяет, оставив предполагаемые ошибки наблюдений на осях специфических факторов, сохранить важнейшие исходные наблюдения в проекциях на оси общих факторов (f_1 и f_2 на рис. 2.3 б).

Теперь уже в новом K -мерном пространстве ищется оптимальное размещение факторных осей. Оси вращаются для максимально точного расположения факторов вдоль направлений наибольшего рассеяния, что позволяет достичь условия простого описания структуры факторного решения (см. рис. 2.3 б).

В ряде случаев простая структура факторов может быть недостижима в терминах ортогонального вращения осей. Однако задача легко решается в проекциях на неортогональные факторные оси, угол между которыми $\beta \neq 90^\circ$. Такой тип описания переменных называется косоугольным вращением и представляет собой эффективное средство описания сложных факторных структур. Поскольку вычислительная процедура получения косоугольного решения весьма сложна, она здесь рассматриваться не будет.

Поиск факторного решения

Процесс поиска факторного решения представляет собой совокупность частных процедур матричного анализа, направленных на решение конкретных задач. При этом практически каждый из рас-

смаатриваемых ниже этапов может иметь иной способ решения, выбор которого определяется умением исследователя и общей направленностью вычислений. Здесь же будет рассмотрен только один из возможных вариантов факторного решения, который прост и иллюстративен.

Итак, процесс решения начинается уже с составления исходной матрицы данных. В матрицу данных включаются те переменные, которые, по мнению исследователя, априорно позволяют решить проблему поиска скрытых факторов. Это могут быть как реально наблюдаемые характеристики состояния океана, так и их различные производные (градиенты по направлению, потоки веществ и т. п.). При составлении матрицы данных важно помнить, что число наблюдений должно превышать число переменных хотя бы в 3—5 раз, что необходимо для получения устойчивых оценок факторного решения.

Далее производится вычисление средних и стандартных отклонений переменных. Затем матрица данных стандартизуется [86, 95]. По стандартизованным значениям вычисляются коэффициенты парной корреляции между переменными, из которых составляется матрица корреляций R . Матрица R — симметричная, квадратная и состоит из M строк и M столбцов. Вся вычислительная процедура МФА строится на основе разложения полученной корреляционной матрицы. Поэтому естественно желать, чтобы вычисленные оценки коэффициентов корреляции отражали реальные тенденции в исходных наблюдениях.

Для удобства вычислений основное уравнение МФА записывается в корреляционной форме:

$$R = A \cdot A^T + U^2, \quad (2.10)$$

где общая корреляция переменных разделена на две составляющие: систематическую коррелированность, представленную в виде произведения $A \cdot A^T$, и коррелированность случайную, свойственную влиянию ошибок и специфических факторов и представленную в виде матрицы U^2 . Предположение о наличии малого числа общих факторов приводит к необходимости решения двойной задачи — оценки числа скрытых факторов и определения вида взаимосвязи факторов и исходных признаков. Практически решение этих задач состоит в построении итеративной процедуры вычисления векторов факторных нагрузок $a = \{a_1, a_2, a_3, \dots, a_m\}$, на основе которых составляется матрица A факторных нагрузок, выполняется матричное умножение $A \cdot A^T$ и проверяется ряд условий, накладываемых априорно на матрицу корреляций остатков U^2 , которая оценивается как остаток:

$$U^2 = R - A \cdot A^T.$$

Вид матрицы U^2 определяется условиями независимости ошибок различных признаков. Поэтому $u_{ij} = 0$ при $i \neq j$. Матрица U^2 —

диагональная, квадратная, порядка M . На ее диагонали находятся нормированные дисперсии остатков:

$$U^2 = \begin{bmatrix} u_{11}^2 & 0 & 0 & \dots & 0 \\ 0 & u_{22}^2 & 0 & \dots & 0 \\ 0 & 0 & u_{33}^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & u_{MM}^2 \end{bmatrix}.$$

В покомпонентной записи выражение (2.10) примет вид

$$r_{ij} = \sum_{p=1}^k a_{ip} a_{jp} + u_{ij}^2 = h_{ij}^2 + u_{ij}^2,$$

из которого следует, что коррелированность двух переменных r_{ij} содержит систематическую долю изменчивости h^2 , которая интерпретируется общими факторами, и случайную долю u^2 , которая понимается как остаток и не содержит физического смысла.

Выделенная систематическая изменчивость возникает под влиянием общих факторов на объект исследования, поэтому h_{ij}^2 называется общностью j -й переменной. Величина u_{ij}^2 , формируемая влиянием специфических факторов, характерных только для данной j -й переменной и наличием случайных ошибок, называется характерностью j -й переменной. Соотношение общности и характерности является важным показателем эффективности введения данной переменной в факторный анализ. Чем ближе общность к единице, тем выше вероятность эффективного факторного разложения.

Наоборот, при увеличении характерности u^2 затруднительно выделить общий для нескольких переменных фактор и достаточно точно описать его. Поэтому целесообразно предварительно оценивать изменчивость исследуемых переменных, обусловленную случайными ошибками наблюдений, измерений или вычислений. Полученная оценка может быть принята в качестве начальной в модели факторного анализа и исключена из процесса вычисления общих факторов путем замены диагональных элементов матрицы корреляций на значения общностей $r_{ii} \rightarrow h_{ii}^2$. Эта процедура называется редуцированием корреляционной матрицы:

$$R^h = R - U^2 = \begin{bmatrix} r_{11} - u_{11}^2 & r_{12} - u_{12}^2 & \dots & r_{1M} - u_{1M}^2 \\ r_{21} - u_{21}^2 & r_{22} - u_{22}^2 & \dots & r_{2M} - u_{2M}^2 \\ \vdots & \vdots & \ddots & \vdots \\ r_{M1} - u_{M1}^2 & r_{M2} - u_{M2}^2 & \dots & r_{MM} - u_{MM}^2 \end{bmatrix}.$$

Далее производится решение проблемы собственных чисел и собственных векторов редуцированной корреляционной матрицы R^h :

— ищется решение характеристического уравнения вида

$$|\mathbf{R}^h - \Lambda \mathbf{I}| = 0, \quad (2.11)$$

что дает оценку вектора собственных чисел Λ ;

— решаются системы линейных однородных уравнений вида

$$\mathbf{A}_0(\mathbf{R}^h - \Lambda \mathbf{I}) = 0, \quad (2.12)$$

что позволяет оценить собственные вектора $\mathbf{A}_0$.

После вычисления собственных векторов, которые принимаются в качестве начальных приближений соответствующих факторных нагрузок, переходят к оценке числа общих факторов, достаточных для воспроизведения матрицы корреляций $\mathbf{R}^+$. Для этого оценивают матричное произведение:

$$\mathbf{R}^+ = \mathbf{A}_0 \cdot \mathbf{A}_0^T,$$

которое дает оценку матрицы восстановленных корреляций, и вычисляют уровень различий наблюдаемых и восстановленных корреляций

$$\mathbf{R}^h - \mathbf{R}^+ = \mathbf{D}.$$

Две эти операции повторяют многократно, начиная с одного собственного вектора в матрице $\mathbf{A}_0^{(1)}$ и с каждым разом увеличивая число векторов на единицу: $\mathbf{A}_0^{(2)}$, $\mathbf{A}_0^{(3)}$, ..., $\mathbf{A}_0^{(M)}$. Как только величина остатков $\mathbf{D}$ удовлетворяет принятому уровню значимости модели, исследователь прекращает добавление собственных векторов и принимает число общих факторов модели равным данному. Полученные собственные вектора являются коэффициентами перехода из M -мерного пространства исходных переменных в K -мерное пространство общих факторов с дисперсией ошибки порядка

$$d_u = \sum_{i=K+1}^M \lambda_i / \sum_{j=1}^K \lambda_j. \quad (2.13)$$

Далее необходимо оптимизировать решение МФА уже в сокращенном факторном пространстве. Этот этап называется вращением или поиском простой структуры решения. Суть этапа состоит в том, чтобы провести операцию поворота осей общих факторов и перейти от матрицы $\mathbf{A}_0$ к матрице $\mathbf{A}$ так, чтобы по возможности все имеющиеся факторные нагрузки были максимально приближены либо к ± 1 , либо к 0. Такое условие зачастую позволяет избежать половинчатых решений и более четко очертить сферу влияния факторов на переменные.

На практике процедура вращения сводится к итеративному многократному умножению вида

$$\mathbf{A} = \mathbf{A}_0 \cdot \mathbf{T},$$

где $\mathbf{T}$ — матрица поворота размерностью M строк и M столбцов с взаимонезависимыми и нормированными столбцами. Квадратная

матрица $\mathbf{T}$ по сути содержит косинусы и синусы угла поворота (α) принятой системы координат, обозначенной матрицей начальных факторных нагрузок $\mathbf{A}_0$, к новой системе координат, обозначенной матрицей $\mathbf{A}$. Пример формирования подобной матрицы поворота $\mathbf{T}$ и геометрическая интерпретация матричного умножения приводится в [86]. Обычно вращение осей проводится в несколько этапов многократным поворотом системы координат $\mathbf{A}_0$ на малый угол α , величина которого выбирается равной 10 или 20°. Вращение осей координат приостанавливается, когда достигнут максимум дисперсии факторных нагрузок, приходящихся на каждый признак:

$$S^2 = \frac{1}{K} \sum_{j=1}^K (a_{ij}^2) - \frac{1}{K} \left(\sum_{j=1}^K a_{ij} \right)^2 \rightarrow \max.$$

Полученные значения коэффициентов a_{ij} принимаются в качестве искоемых факторных нагрузок.

Последним этапом факторного анализа традиционно является процедура вычисления значений общих факторов и оценки матрицы остатков. Оценка значений общих факторов $\mathbf{F}$, как правило, осуществляется на основе классической схемы множественной линейной регрессии. Матричное уравнение для этой операции имеет вид

$$\mathbf{F} = \mathbf{X} \cdot (\mathbf{A} \cdot \mathbf{A}^T)^{-1} \cdot \mathbf{A}^T. \quad (2.14)$$

Здесь необходимо обратить внимание на оператор $(\mathbf{A} \cdot \mathbf{A}^T)^{-1}$, который является обратной матрицей восстановленных корреляций. Дело в том, что точность проведения процесса обращения в данном случае определяет надежность оценки значений общих факторов.

Вычисление остатков проводится в результате оценки матричной разности:

$$\mathbf{E} = \mathbf{X} - \mathbf{F} \cdot \mathbf{A}^T.$$

Рассмотренная выше схема МФА получила название метода главных факторов и довольно широко применяется на практике. На сегодняшний день разработан еще целый ряд процедур оценки факторных нагрузок и вращения общих факторов, с которыми можно познакомиться в специальной литературе [75, 204].

Практические применения МФА в океанологии

В океанологической практике данный метод стал применяться только последние полтора десятилетия, т. е., очевидно, позже всех других методов многомерной статистики. Первые попытки приложения МФА к решению океанологических задач следует отнести ко второй половине 70-х годов [96, 150, 169]. Накопленный опыт позволяет лишь наметить основные проявившиеся тенденции использования МФА в океанологии.

На пути внедрения МФА сразу встало несколько серьезных препятствий, основное из которых состояло в непонимании раз-

личий между МФА и методом главных компонент в некоторых прикладных работах. Разночтения были вызваны явным запаздыванием издания доступной широкому кругу прикладников литературы по МФА, которая появилась лишь в последние годы [73, 81, 86]. МФА общепризнан как самостоятельная техника многомерной статистики.

Тем не менее вопрос о потенциальном сходстве или различии методов главных компонент и факторного анализа оставался до недавнего времени открытым для гидрометеорологов. Попыткой ответа на этот вопрос можно считать те работы, в которых проводился сравнительный анализ результатов компонентного и факторного разложений для конкретных выборок. Не избежали такого подхода и авторы [39, 169, 170]. Противоречивость получаемых выводов заставляет вновь обратиться к основным теоретическим постулатам рассматриваемых методов.

Наиболее обоснованной выглядит сегодня следующая трактовка решения данного вопроса: сама структура исследуемых данных предопределяет меру соответствия факторного и компонентного разложений. Наиболее вероятна высокая связность получаемых решений, если при выполнении гипотез многомерного нормального распределения исходных данных достигается высокая сходимость разложения корреляционной матрицы по собственным числам. В этом случае наличие высоких внедиагональных корреляций в матрице R обеспечивает устойчивость оценок собственных чисел и векторов. В этом случае компонентное разложение чаще всего близко соответствует решению по методу главных факторов.

Наиболее существенные различия МГК и МФА прослеживаются для ситуаций сложной внутренней структуры исследуемых переменных, когда компонентное разложение имеет невысокую скорость сходимости, а дисперсия неоднозначно распределяется по компонентам. Именно здесь процедуры устойчивого оценивания факторного анализа позволяют получить более достоверные начальные оценки факторных нагрузок. А последующие процедуры ортогонального или неортогонального вращения общих факторов окончательно «выжимают» информацию из имеющегося решения до получения простой объясняемой структуры взаимосвязей переменных.

Следовательно, в последнем случае исследователь сознательно удаляется в область факторного анализа от легко достижимого, но неудовлетворительного результата компонентного разложения.

Методологию факторного анализа нельзя отнести к категории «лобовых методов». Скорее это «мастерское оружие», которое гарантирует качественный результат при умелом вычислении параметров факторного разложения. Разнообразие методических приемов при этом служит необходимым условием достижения оптимального решения.

Опыт показывает [238, 247], что традиционная методика главных факторов в последнее время все чаще заменяется в практике различных прикладных исследований более современными мето-

дами максимального правдоподобия, минимальных остатков и т. п. Этап оценки факторных нагрузок сегодня предоставляет исследователю самые широкие возможности получения устойчивого решения. Подробное описание практически всех популярных методов вычисления факторных нагрузок приводятся в работе К. Иберла [81].

Проблема вращения факторных осей чаще всего сейчас решается в ортогональном варианте, сохраняющем статистическую взаимонезависимость результирующих факторов [96, 220, 226, 237, 263]. Однако в случае анализа сложных многоуровневых взаимосвязей требуемой простой структуры достичь в ортогональном пространстве общих факторов не удастся. Как правило, это случается для облаков наблюдений чрезвычайно сложной формы [205, 262]. Для того чтобы добиться адекватного решения в классе линейных статистических моделей, осуществляется неортогональный поворот факторных осей, приводящий к косоугольному пространству общих факторов. Тогда получаемые факторы оказываются линейно взаимосвязанными и образуют новый уровень описания исследуемых процессов [122, 176].

Матрица остатков E в факторном анализе занимает особое место уже потому, что она не задается нулевой априорно. Более того, не отвергается возможность проведения факторного анализа самой матрицы E остатков и специфических факторов [30, 81]. Цель такого анализа состоит в исследовании структуры малодисперсных взаимосвязей переменных, которые могут играть важную роль в системах с высокой избирательной чувствительностью.

В океанологических исследованиях факторный анализ успешно применяется в трех основных направлениях. Построение малопараметрических физико-статистических моделей функционирования природных систем является наиболее распространенной задачей факторного анализа. В этом случае достигается полное решение факторной задачи, включая достоверную оценку значений факторов и формирования регрессионных уравнений связи исходных наблюдаемых параметров и вычисленных статистических общих факторов. Простая структура решения позволяет, кроме того, дать вполне однозначную физическую интерпретацию полученных оценок общих факторов.

Примером такого подхода могут быть работы Ю. Л. Семенова [164, 165] по исследованию физико-химических процессов формирования гидрологического режима Каспийского моря. Используя в общем традиционные для подобных задач исходные ряды гидрохимических и гидрологических характеристик, Семенову удается на основе факторного решения построить логически завершенную и количественно обоснованную модель изучаемых процессов.

Аналогичный подход, используемый в [229, 236], позволил сформулировать локальную физико-статистическую схему процессов взаимодействия океана и атмосферы в прибрежной зоне Тихого океана.

Второе направление использования факторного анализа связано с проблемами гидрометеорологической классификации. В данном аспекте используется эффект «уплотнения» облаков наблюдений в факторном пространстве.

Дело в том, что размытые облака наблюдений, проектируясь на оси общих факторов, приобретают простую структуру взаимного расположения, которая сводится к уплотнению наблюдений вдоль осей наиболее сильного взаимодействия переменных. Получаемая простая структура упаковки исследуется визуально или специальными методами [82, 168]. В океанологии прикладное содержание такой классификации состоит в оптимизации пространственного районирования исследуемых акваторий [2, 46, 150, 173].

Наконец, третье направление использования МФА связано с прогностической ценностью факторного разложения. Повышенная по сравнению с МГК устойчивость общих факторов к стохастической изменчивости в исследуемых временных рядах объясняет интерес к ним как к потенциальным предикторам.

В принципе достижение простой структуры факторного решения при таком подходе должно частично снимать наболевший вопрос об отборе информативных для прогноза членов разложения. Однако этот вопрос пока еще далек от окончательного решения в океанологической практике, хотя прогноз общих факторов уже успешно апробирован в промыслово-океанологических исследованиях [96, 103, 108].

3.1. Метод автоматической классификации

Процессы, протекающие в океане, различные явления и объекты, с которыми сталкивается океанолог, отличаются большим разнообразием. Трудно найти два процесса или два объекта, абсолютно похожих друг на друга. В то же время у отдельных процессов и объектов всегда можно выявить общие свойства и особенности, позволяющие объединить их в группы, типы, классы.

В более широком понимании класс можно отождествить с образом, ибо в теории распознавания под образом понимают множество явлений, объединенных общими свойствами. Необходимость введения понятия образа (класса) обусловлена практической потребностью оперировать более общими категориями, чем понятие индивидуального процесса или объекта, а также невозможностью обозреть все процессы или объекты, входящие в тот или иной класс [12, 107, 182].

Рассматривая классификацию как основу той или иной модели, прежде всего необходимо иметь в виду, что она должна не только обобщать информацию, но и отражать отдельные стороны объективной действительности. В научных классификациях мы имеем дело с абстрактными образами, классами, свойства которых очень многочисленны. Отсюда возникает необходимость оперировать образами, заданными в многомерном пространстве признаков.

Вот почему задача распознавания имеет прямое отношение к задачам многомерного анализа. Однако в этом случае она может решаться только на основе машинных методов классификации, т. е. методов автоматической классификации, в которых исследователь задает только алгоритм решения задачи.

Пусть задано множество Ω в пространстве M признаков, которое образуется выборкой X из N наблюдений за M переменными. Требуется статистически обоснованно и достоверно выделить k подмножеств (групп, классов наблюдений) $\omega_1, \omega_2, \omega_3, \dots, \omega_k$:

$$\Omega = \omega_1 \cup \omega_2 \cup \dots \cup \omega_k.$$

Для выделения подмножеств заранее накладываются следующие условия:

- 1) получаемые подмножества не будут пересекаться:

$$\omega_i \cap \omega_j \text{ при любых } i \neq j;$$

другими словами, каждый вектор x_i , содержащий M значений наблюдаемых переменных $x_i = \{x_{i1}, x_{i2}, \dots, x_{iM}\}$, будет принадлежать только одному из выделенных подмножеств:

2) каждое подмножество не должно быть пустым

$$\omega_i \neq 0 \text{ при } i = 1, 2, \dots, k,$$

т. е. в каждом выделенном классе должно находиться хотя бы одно наблюдение x_i . Отсюда с очевидностью следует, что число выделенных классов в итоге не может превышать числа наблюдений $k \leq N$. В то же время число классов всегда больше нуля — $k \geq 1$.

Постановка задачи

Задача классификации представляется одной из важнейших в любой области научного знания. Уровень классификации, ее достоверность определяют действительный уровень современного знания о предмете исследования. Проблема классификации в океанологии прежде всего связывается с вопросами классификации водных масс, с проблемой районирования Мирового океана по тем или иным признакам, с исследованиями по типизации ледовых условий или проблемой выявления классов характерной изменчивости климата океана и т. д.

Многообразие прикладных проблем породило многообразие методов автоматической классификации, которые приспособлены по своей чувствительности к решению конкретных достаточно узких задач. Единого общего подхода, одинаково хорошо работающего во всех ситуациях, на сегодняшний день, по-видимому, нет. Поэтому, рассматривая вопросы классификации применительно к исследованиям океана, обратимся прежде всего к тем методам, которые хорошо себя зарекомендовали при решении чисто океанологических задач.

Что такое классификация? Это процесс разбиения множества объектов на некоторое число подмножеств (классов). Каждое подмножество, включающее в себя объекты, похожие друг на друга по известному набору отличительных признаков, и называется классом. Таким образом, итог классификации можно представить в виде наборов объектов, расположенных в пространстве признаков и разделенных границами классов. Однако данная интерпретация — не единственная.

Не менее важно в процессе классификации упорядочить объекты по степени их взаимной связанности и выявить структуру объединения объектов в классы. Этот аспект классификации прослеживается путем построения «дерева классификации». Такое дерево приведено на рис. 3.1. Подобное дерево позволяет проанализировать соподчинение классов и объектов и построить иерархию классов.

При данном графическом описании классификации вместо единственной застывшей картины разбиения получается набор

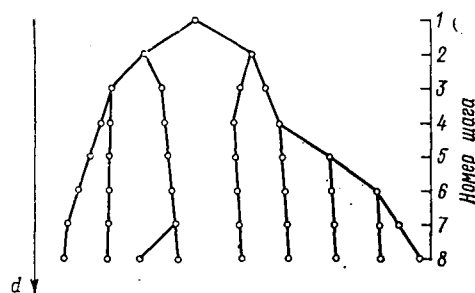
возможных разбиений, каждое из которых достойно физической интерпретации.

Конкретные задачи классификации можно сформулировать в такой последовательности:

- выбор исходных переменных,
- выбор количественного выражения сходства наблюдений,
- выбор алгоритма классификации,
- интерпретация результатов.

Рис. 3.1. Представление классификации наблюдений в виде дерева решения (дендрограммы).

d — сходство классов.



Выбор системы исходных переменных предваряет собственно вычислительную процедуру, поэтому может быть рассмотрен отдельно, вместе с вопросами геометрической интерпретации метода автоматической классификации (МАК).

Геометрическая интерпретация МАК

Рассмотрим этот вопрос на примере анализа водных масс с помощью всего трех переменных: температуры, солености и содержания кремния. Исходные признаки образуют признаковое пространство, которое может быть представлено в виде трех взаимортогональных осей координат (рис. 3.2).

Реальные наблюдения представлены в этом признаковом пространстве как точки, которые образуют неоднородное по плотности облако наблюдений. Интуитивно понятно, что каждая водная масса характеризуется близкими значениями наблюдаемых параметров. Следовательно, водная масса может быть представлена в этом пространстве признаков как некий замкнутый объем, внутри которого точки наблюдений образуют достаточно плотную компактную группу. Подобный объем можно отделить от оставшихся точек искусственной поверхностью раздела. Объединив все данные наблюдения в подобные замкнутые объемы, мы тем самым решим задачу классификации.

Вернемся к проблеме выбора переменных. Пусть данные наблюдений за содержанием кремния отсутствуют. Тогда все облако наблюдений будет спроектировано в плоскость T , S -координат. При этом становится очевидным, что ранее выделяемые классы, типы водных масс, здесь пересекаются и не могут быть выделены,

поскольку плотность упаковки наблюдений примерно равномерна во всем облаке.

Данный пример наглядно демонстрирует, что успешная классификация наблюдений достижима лишь тогда, когда выбранные признаки потенциально позволяют это сделать на основании имеющейся неоднородности облака данных. Поэтому выбор при-

знаков, как считает большинство исследователей, фактически предрешает исход классификации.

Каким же образом лучше выбрать признаки? Для решения некой частной специфической задачи (например, типизации ледовых условий) набор признаков ограничивается физикой класси-

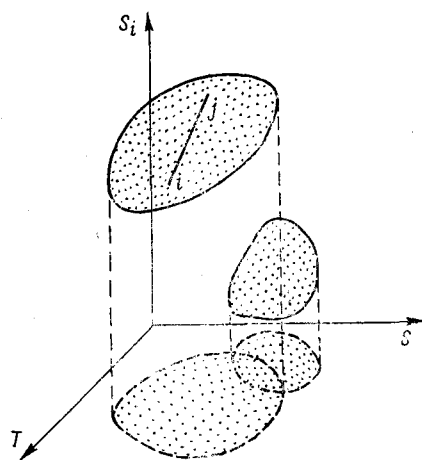


Рис. 3.2. Геометрическая интерпретация классификации в пространстве трех (температура, соленость, кремний) и двух (температура, соленость) признаков.

фицируемого явления. В случае же создания универсальной классификации крупного природного объекта, например при районировании Мирового океана, требуется привлечение всех имеющихся в наличии наблюдаемых параметров.

Всегда ли привлечение большого числа признаков позволит успешно решить задачу? Ведь известная «теорема о гадком утенке» гласит, что любую пару схожих объектов разделяет точно то же число предикатов, что и любую пару несхожих объектов. Опыт показывает, что уже при относительно небольшом числе признаков достигается устойчивость классификации.

Это объясняется тем, что большинство измеряемых признаков, описывающих процесс или явление, статистически взаимосвязаны (в противном случае действительно наблюдался бы полный хаос). Поэтому присоединение новых признаков к некоторому достаточно большому набору не принесет никакой существенной информации о взаимном расположении объектов в признаковом пространстве, так как среди этого набора, весьма вероятно, найдется один или несколько признаков, имеющих довольно высокую корреляцию с каждым из вновь присоединенных.

Другими словами, можно представить себе некоторый конечный набор признаков, позволяющий отразить структуру изучаемых природных объектов. Однако необходимо всегда помнить, что этот большой набор признаков в любом случае отражает целевое назначение классификации, как бы мы ни пытались абстрагиро-

ваться от нее. И в этом смысле, в смысле формулирования цели классификации, любая классификация остается субъективной.

В практике океанологических исследований мы очень редко можем иметь столь большой набор признаков, который можно было бы рассматривать как насыщенный. Поэтому в большинстве случаев приходится опираться на традиционный опыт подбора переменных.

Количественная оценка сходства наблюдений

Количественным выражением сходства наблюдений по данному набору признаков может быть простейшее расстояние между двумя точками в пространстве переменных (см. рис. 3.2). Подобная мера сходства называется расстоянием Евклида и определяется по формуле

$$d_{ij} = \left[\sum_{k=1}^M w_k (x_{ik} - x_{jk})^2 \right]^{1/2}, \quad (3.1)$$

где x_{ik} — значение k -й переменной для i -го наблюдения, w_k — относительный вес данного k -го признака.

Расстояние Евклида удобно использовать в случае совпадения размерностей всех признаков, если переменные взаимонезависимы и известны их относительные вклады.

В случае когда эти условия не выполняются, может быть использовано расстояние Махаланобиса, при вычислении которого допускается различие размерности отдельных переменных. В векторной форме расчетная формула расстояния Махаланобиса может быть записана следующим образом:

$$d_{ij} = (\mathbf{x}_i - \mathbf{x}_j)^T \cdot \mathbf{S}^{-1} \cdot (\mathbf{x}_i - \mathbf{x}_j), \quad (3.2)$$

где $(\mathbf{x}_i - \mathbf{x}_j)$ — вектор-столбец разностей значений M признаков на i -м и j -м наблюдениях, $\mathbf{S}$ — матрица ковариаций переменных размерностью M строк и M столбцов.

Расстояние Махаланобиса эффективно характеризует сходство наблюдений в случае коррелированности переменных.

Нередки ситуации, когда наблюдаемые переменные выражены не в количественной, а в качественной шкале отсчета, например при исследовании различных шкал балльности льда, интенсивности ветра, волнения либо при анализе последовательности ситуаций наличия или отсутствия какого-то явления. Здесь использование приведенных выше мер сходства нецелесообразно. Более эффективно в подобных случаях применять расстояние Хемминга:

$$d_{ij} = \sum_{k=1}^M |x_{ik} - x_{jk}|.$$

Здесь x_{ik} — значение величины k -го признака в номинальной шкале.

Расстояние Хемминга описывается только целочисленными значениями, которые ограничиваются снизу нулем, что соответствует ситуации полного совпадения значений всех переменных для двух наблюдений. Здесь необходимо отметить, что использование расстояния Хемминга весьма эффективно в тех случаях, когда исходные переменные содержат случайные ошибки или шумы. Тогда разбиение области изменения признаков на несколько интервалов и взятие номеров интервалов в качестве значений переменных позволяет во многом устранить влияние шумовой компоненты и получить более достоверную классификацию.

Перечисленные выше меры сходства характеризуют близость наблюдений в линейном пространстве и оказываются эффективной числовой характеристикой при поиске классов, имеющих форму сферы или эллипса в многомерном пространстве переменных. Однако такая простая форма облаков рассеяния наблюдается отнюдь не всегда, что заставляет искать более сложные нелинейные выражения для мер сходства. Весьма часто для этих целей используют различные потенциальные функции, которые получаются в результате нелинейных преобразований простых линейных мер сходства:

$$P_{ij} = 1/(1 + d_{ij}),$$

$$P_{ij} = 1 + \exp(d_{ij}).$$

Эти потенциальные функции более эффективны при классификации скоплений наблюдений в форме подковы, бублика и т. п.

Выбор конкретной меры сходства целиком зависит от исследователя, здесь открыт широкий простор для научного поиска.

Вычислительная процедура МАК

С общих позиций все алгоритмы МАК можно подразделить на вероятностные и эвристические. Первые основаны на принципе минимизации среднего риска неправильной классификации наблюдений и здесь рассматриваться не будут.

Эвристические алгоритмы используются в океанологических задачах пока более часто. Это закономерное следствие их достаточно высокой эффективности, которая достигается применением весьма несложных вычислительных процедур. Сущность эвристических алгоритмов классификации в том и состоит, что в них реализуется на математической основе общепринятые логические, интуитивные правила поиска однородных классов. Поэтому вычислительные процедуры уже априорно настроены на конкретную форму выделяемых классов. Выявляются два типа процедур эвристической классификации: 1-й тип предпочтителен для выделения подмножеств, изотропных в пространстве признаков, 2-й тип реализует принципы выделения изоморфных (подобных по форме) классов.

Изотропные подмножества традиционно понимаются как шарообразные по форме взаимного положения наблюдений. В многомерном пространстве признаков эта форма трактуется как гиперсферическая. Поэтому и вычислительные процедуры носят название «алгоритмов гиперсфер» [4, 74, 79]. Общий смысл их сводится к следующему.

Для каждого наблюдения $x_i = \{x_{i1}, x_{i2}, \dots, x_{iM}\}$ строится в признаковом M -мерном пространстве гиперсфера радиусом ρ . Затем определяется число наблюдений, попавших внутрь гиперсферы, и формируется подмножество ω_i , которое сформировано наблюдениями, удаленными от исходного на расстояние d , меньшее заданного радиуса ρ : $d \leq \rho$.

В качестве первого класса выбирается тот, внутрь которого попало максимальное число соседних наблюдений. Для оставшихся наблюдений процедура построения повторяется до исчерпания всей выборки.

Алгоритмы классификации изоморфных классов направлены на выявление структур, подобных по форме взаимного расположения наблюдений в многомерном пространстве признаков. Интуитивно ясно, что многообразие взаимосвязей наблюдений порождает самые различные формы сгущений наблюдений в пространстве признаков.

В практике анализа гидрометеорологических процессов сгущения наблюдений имеют вид вытянутых «цепочечных» образований. Поэтому широкую популярность в океанологических исследованиях заслуженно завоевал метод построения кратчайшей связывающей сети (КСС) [125, 172].

Суть метода состоит в последовательном попарном объединении близлежащих наблюдений в общую сеть. Так, лягушка, прыгая по болоту с одного листа на другой, ближайший к ней листок, объединит своими прыжками все листки в единую сеть, которая будет пройдена с минимальными затратами. Вычислительная логика метода КСС состоит в том, чтобы, взяв произвольную начальную точку наблюдений, найти наиболее похожее на нее наблюдение по минимальному расстоянию между точками среди всех других точек, еще не вошедших в сеть. Присоединив эту точку к сети, необходимо искать минимум расстояния до оставшихся точек относительно всех тех точек, которые уже включены.

После объединения в сеть всех точек наблюдений производится собственно разбиение на компактные классы. Для этого задается критическое расстояние $d_{кр}$. По условию $d_{ij} \leq d_{кр}$ производится разбиение сети на отдельные отрезки, которые и являются искомыми классами.

Особое внимание следует обратить на проблему оптимальной классификации. В самой постановке задачи классификации можно заметить неоднозначность конечного результата. Поэтому оптимизация разбиения на классы переносится на заключительный этап вычислений, когда оценивается относительная достоверность

нескольких результатов классификации с помощью дополнительных эмпирико-статистических критериев.

Чаще всего выбор оптимального уровня классификации осуществляется на основе дисперсионного соотношения вида:

$$Q = J_2/J_1 \text{ или } Q = J_2 - J_1, \quad (3.3)$$

где J_1 — среднее по всей выборке расстояние между наблюдениями, расположенными внутри классов (среднее внутриклассовое расстояние); J_2 — среднее расстояние между наблюдениями, расположенными в разных классах (среднее межклассовое расстояние).

Смысл критерия Q состоит в том, что он показывает отношение среднего уровня различий классов к средней плотности упаковки наблюдений внутри класса. Очевидно, максимум величины Q соответствует классификации, в которой наблюдения упакованы в плотные классы, максимально удаленные друг от друга. Такой вариант разбиения выборки на классы и представляется оптимальным.

Расчетные формулы для J_1 и J_2 имеют вид

$$J_1 = \frac{1}{k} \sum_{j=1}^k \left(\frac{1}{N_j} \sum d_{lp} (x_l, x_p \in \omega_j) \right),$$

$$J_2 = \frac{1}{k-1} \sum_{j=1}^{k-1} d_{lp} (x_l \in \omega_j, x_p \notin \omega_j),$$

где k — число выделенных классов; N_j — число наблюдений в j -м классе; d_{lp} — расстояние между парами наблюдений, принадлежащих к одному (для J_1) или к разным (для J_2) классам.

В любом алгоритме классификации есть параметр, определяющий чувствительность метода к структуре расположения наблюдений и в конечном счете формирующий классы. В алгоритмах гиперсфер такой параметр есть задаваемый радиус гиперсферы ρ . Чем меньше ρ , тем более плотные классы будут найдены, и, наоборот, с ростом ρ плотность классов уменьшается.

Как правило, первоначальный выбор радиуса гиперсферы основывается на предварительных результатах анализа статистической структуры неоднородности наблюдений. Вариация ρ позволяет получить несколько различных классификаций, из которых следует выбрать оптимальную.

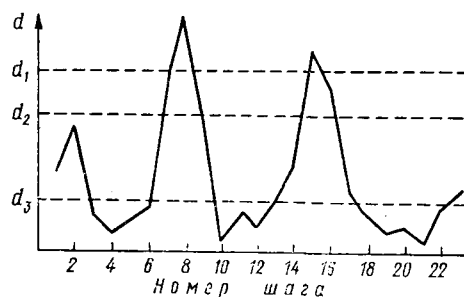
Для нерархических методов, к которым относится рассмотренный выше метод КСС, все возможные варианты классификации уже представлены в полученной сети. Здесь параметром, определяющим классификацию, является величина критического расстояния $d_{кр}$, которая может варьироваться исследователем в пределах от минимального для данной выборки наблюдений (получаем N классов) до максимальной для выборки (получаем 1 класс).

Смысл подобной оптимизации нетрудно понять из рис. 3.3, где представлен график расстояний между объектами, последова-

тельно объединяемыми в сеть. Очевидно, что максимумы значений d на этом графике соответствуют ситуации, когда присоединяются к сети весьма удаленные объекты, возможно принадлежащие к другой группе. Внутри же классов, где сходство объектов велико, расстояния между ними малы, что прослеживается на графике как впадины. Тогда, задавая некий критический уровень $d_{кр}$, мы тем самым определяем границу между классами. При максимальной величине $d_{кр}$ мы получим минимальное число классов, и разбиение будет обладать малой подробностью. При малой величине $d_{кр}$ получается излишне детальная классификация. Опти-

Рис. 3.3. График объединения наблюдений в кратчайшую связывающую сеть.

Штриховые прямые — возможные уровни критического расстояния между классами.



мизируя по критерию Q величину критического расстояния между классами, исследователь выбирает тот вариант разбиения, который в среднем для данной совокупности объектов дает наибольшее взаимное удаление классов при одновременной высокой внутренней их компактности.

Применение МАК в океанологических исследованиях

Одна из первых проблем, с которой неизбежно сталкивается океанолог в практике применения МАК, состоит в поиске эффективной методики классификации данной конкретной выборки наблюдений. Применяя классификационные методы, исследователь априорно не имеет достоверных представлений о структуре классов. Скорее, он располагает лишь предварительными соображениями, которые строятся на количественных оценках первичного статистического анализа, общеизвестных физических механизмах формирования исследуемых процессов и собственной интуиции.

Число классов, их форма и принадлежность к ним конкретных наблюдений в лучшем случае только угадываются. Выбирая методику классификации, следует иметь в виду, что конечный результат во многом зависит от правильности и обоснованности этого выбора.

На сегодняшний день в океанологических исследованиях получили широкое распространение все перечисленные алгоритмы, и пока трудно предпочесть какой-нибудь из них. В целом оформилась лишь общая тенденция использовать по возможности простые вычислительные схемы. Поэтому методы классификации,

разработанные на основе построения кратчайшей связывающей сети и кратчайшего незамкнутого пути, получили известность в 60-е — 70-е годы [125, 172]. Простота вычислений, требующая совсем небольшого ресурса ЭВМ в сочетании с наглядностью и интерпретируемостью получаемых результатов, во многом объясняют такую популярность.

Во второй половине 70-х годов все чаще стали использоваться различные алгоритмы таксономии, основанные на построении гиперсфер в пространстве признаков [79, 209]. В целом положительный опыт применения этих методов в океанологии, тем не менее, выявил сложность построения оптимальных классификаций, обусловленную низкой достоверностью настройки алгоритмов к структуре выделяемых классов. Поэтому в конце 70-х — начале 80-х годов, когда была начата обработка крупных архивов наблюдений, применение МАК вышло на новый уровень. На этом этапе авторы, добиваясь оптимального результата, осуществляли многократное классифицирование выборок различными методами, с разными подходами к оценке сходства наблюдений [56, 92, 161, 185]. Особое внимание уделялось проблеме надежного выделения классов сложной формы, в связи с чем применялись усовершенствованные формулы расчета нелинейных оценок межклассовых и внутриклассовых расстояний [178, 213, 215, 217].

В конце 80-х годов впервые в океанологии была применена методика классификации, основанная на теории «нечетких множеств» [110, 242]. В этом подходе отвергнут основной исходный постулат классической классификации, согласно которому, выделяемые классы не должны пересекаться.

Наоборот, само понятие нечеткого множества основано на явном допущении пересечения классов, при котором отдельные наблюдения могут принадлежать одновременно двум или нескольким соседним классам [97]. В этом подходе принципиально по-новому трактуется граница класса. Если в классической постановке — это гиперплоскость, разделяющая наблюдения в различных классах, то в терминах нечетких множеств граница понимается как область пересечения классов с принадлежащими ей наблюдениями. Новый подход позволяет в процессе классификации океанологических процессов особое внимание уделить переходным зонам, зонам трансформации свойств и т. п.

Накопленный опыт убеждает, что методология автоматической классификации применялась для решения большого числа прикладных океанологических задач, по-видимому, наиболее часто — применительно к задаче классификации водных масс различных регионов Мирового океана [24, 28, 38, 46, 56, 88, 98, 260].

Причин такого применения МАК было множество, но основными были, очевидно, две: во-первых, желание автоматизировать традиционно «ручной» способ выделения водных масс; во-вторых, стремление наиболее достоверно охарактеризовать водные массы с учетом большого числа гидрофизических, химических и биологических параметров.

Особо выделяется классификационный подход при выявлении вертикальной структуры вод океана в соответствии с распределениями различных гидрофизических и гидроакустических параметров [56, 100, 167]. Данная работа позволила установить типовые параметры стратификации вод и затем районировать акваторию океана в соответствии с полученной классификацией.

В перечисленных задачах классификация выполнялась в координатных осях «пространство—переменные», а временная изменчивость определяла результат лишь косвенно. Классификация гидрометеорологических процессов с учетом пространственной и временной изменчивости наблюдений играет решающую роль в районировании океана. В данном случае классы, как правило, отождествляются с типовыми пространственными распределениями океанологических и метеорологических полей и отражают общую направленность физических процессов [123, 130, 149, 174, 185]. Поэтому предметом физического анализа служат здесь как основные особенности типовых полей, так и последовательность смены классов во времени.

Проблема интерпретации, т. е. физического осмысления полученных результатов формального разбиения наблюдений, представляется заключительным этапом МАК, завершающим процесс вычислений. Одновременно это может быть и первая ступень к дальнейшему исследованию возможностей применения классификации к данной выборке.

Дело в том, что физическая интерпретация классов, полученных в результате разбиения, удается отнюдь не всегда. Хорошо, если полученные группы однородных объектов позволяют исследователю соотносить их с уже ранее известными физическими свойствами предмета исследований. Для подобной процедуры целесообразно провести вычисление средних значений признаков внутри каждого из классов, дисперсий и вариаций признаков внутри классов, оценить корреляционные связи между признаками, другими словами, выполнить первичный анализ наблюдений по каждому из полученных классов. Полученные оценки с учетом доверительных интервалов их изменчивости позволяют провести количественное сравнение результатов классификации с уже накопленным ранее материалом исследований.

Другой важный шаг интерпретации состоит в том, чтобы изобразить полученные классы наблюдений в привычных пространственно-временных координатах. Перенос результатов классификации в пространственные координаты называется районированием. В этом процессе возможны сложности из-за несоответствия числа формально выделенных классов и числа однородных в географических координатах районов. Это явление обусловлено наличием нескольких удаленных природных районов с близкими значениями наблюдаемых признаков.

Так, возвращаясь к рассмотренному выше примеру анализа водных масс, следует напомнить, что субтропические водные массы с одинаковыми характеристиками наблюдаются и в Атлан-

тическом, и в Тихом, и в Индийском океанах. Поэтому в случае различия числа однородных классов и однородных районов исследователю следует продолжить классификацию, предприняв действия, необходимые для получения более полной, подробной картины разбиения.

В целом ряде практических ситуаций проблема адекватной классификации не столько зависит от количества исходных наблюдений, сколько определяется набором включенных признаков. Применительно к МАК проблема подбора признаков имеет иное содержание по сравнению с другими методами многомерной статистики.

Преобразование исходной матрицы признаков

При классификации часть исходной информации всегда теряется. Это объясняется тем, что объединение в той или иной степени различных объектов в один класс связано с исключением информации, характеризующей индивидуальные черты каждого объекта. Очевидно, подобного рода индивидуальные черты являются более мелкими по сравнению с более общими свойствами, присущими всем объектам какого-либо класса. Как же выделить эти наиболее общие черты, свойства, присущие большинству объектов? Этого можно достигнуть за счет преобразования исходной информации.

Преобразование исходной информации отражает одну из главных идей проблемы распознавания образов и классификации. Решение многих трудноформализуемых задач в значительной степени облегчается умением преобразовать исходную информацию к немногочисленным признакам, достаточно полно характеризующим изучаемые объекты. Такого рода обобщенные признаки позволяют не обращать внимания на несущественные для классификации детали.

Характер преобразования исходных данных зависит и от их вида. В океанологии, как и в гидрометеорологии в целом, основными методами преобразования являются метод главных компонент и факторный анализ.

Важно то, что указанные методы позволяют получить систему взаимно независимых признаков, что значительно упрощает алгоритмы классификации. МГК получил широкое применение для преобразования информации при выделении водных масс.

3.2. Метод дискриминантного анализа

Постановка задачи

Процесс накопления знаний об океане неизбежно связан с сопоставлением новых данных и уже ранее накопленной и обработанной информации. Обработанная информация об океане, как пра-

вило, представляет собой структуру знаний в форме обобщений. Обобщением знаний выступают классы или типы явлений и процессов, выявленные в результате исследования всего накопленного архива данных.

Полученные типы представляют собой наиболее вероятные, устойчивые состояния исследуемых объектов, для которых не только определены статистические характеристики, но также сформулированы физические законы их возникновения, существования и взаимоперехода. Собственно соотнесение новой информации с уже классифицированной информацией и является основной задачей применения метода дискриминантного анализа (МДА). Сама методика при этом выступает как формальное математическое правило, согласно которому, путем статистического сравнения характеристик исследуемых наблюдений и выделенных ранее классов производится оценка принадлежности данного вектора наблюдения.

Применение МДА позволяет свести к минимуму возможные ошибки неправильного определения принадлежности наблюдения к выделенным классам, что определяет эффективность методики на практике.

МДА в полной мере относится к области знаний, называемой «распознаванием образов» [17, 130]. Статистические характеристики выделенных ранее классов представляют собой образы; на соответствие которым анализируется образ нового вектора наблюдения, составленный из аналогичных статистик.

Классическим примером использования теории распознавания образов в океанологических исследованиях может служить проблема типизации вертикального распределения плотности воды в Мировом океане.

Акватория океана уже достаточно подробно районирована по типам вертикального распределения плотности воды. Для каждого региона определены статистические характеристики распределения плотности (среднее, дисперсия, асимметрия), увязан водоедин с распределением плотности и комплекс других параметров состояния океана — термохалинный режим, динамика вод, оптические и акустические характеристики, распределение биологических и химических показателей. Если удастся объективно определить принадлежность вновь наблюдаемого вертикального профиля плотности к одному из выделенных классов, то, следовательно, удастся получить весь комплекс сопутствующей данному классу информации и разобраться в физике протекающих процессов.

Перейдем непосредственно к формулировке задач метода дискриминантного анализа. В процессе применения МДА требуется решить две основные задачи:

- сформулировать математическое правило отнесения объектов к одному из выделенных классов,
- определить принадлежность нового вектора наблюдений к классу.

В качестве исходной используется архивная выборка данных, уже разделенная на однородные классы — так называемая обучающая выборка X , состоящая из N наблюдений за M переменными. Наблюдения разнесены по k классам. Для каждого i -го класса вычислены первичные статистические характеристики (среднее для M параметров $x_i = \{\bar{x}_{1i}, \bar{x}_{2i}, \dots, \bar{x}_{Mi}\}$, дисперсии для M параметров $s_i = \{s_{1i}, \dots, s_{Mi}\}$) и построены эмпирические гистограммы $P_i(x_1, x_2, \dots, x_M)$.

По обучающей выборке необходимо сформулировать решающее правило. Под правилом понимается математическое выражение — дискриминантная функция, по величине которой можно сделать однозначный вывод о принадлежности вектора наблюдений к конкретному классу.

Геометрическая интерпретация МДА

Решение поставленной в МДА задачи может быть пояснено на простом геометрическом примере. Для этого представим вы-

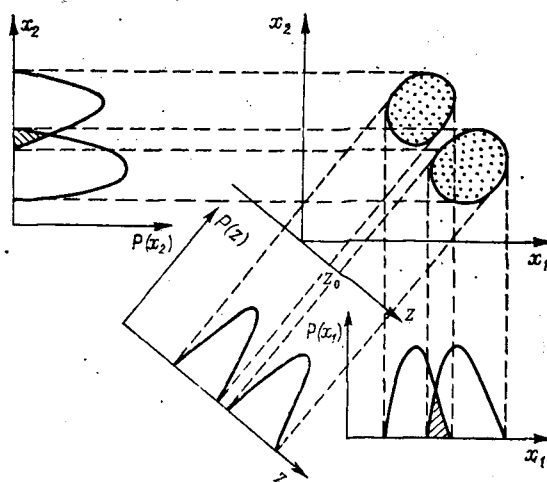


Рис. 3.4. Геометрическая интерпретация метода дискриминантного анализа.

x_1, x_2 — исходные признаки; $P(x_1), P(x_2)$ — плотности вероятности наблюдений по признакам; z — дискриминантная функция; $P(z)$ — плотность вероятности распределения наблюдений в проекции на функцию.

борку исходных данных, состоящую из N наблюдений за $M = 2$ переменными в виде облаков рассеяния в плоскости координат (рис. 3.4). Пусть имеются два отчетливо выделяемых облака данных, которые априорно объединены в два класса: 1-й и 2-й.

Для каждого класса вычислены значения величин дисперсии и математического ожидания по каждому из двух признаков.

Кроме того, на каждый из признаков построены обобщенные гистограммы облаков рассеяния: $P_1(x)$ и $P_2(x)$.

Как следует из приводимых гистограмм, часть значений наблюдений попадает в область пересечения кривых, внутри которой не представляется возможным точно определить принадлежность наблюдений классам. Смысл дискриминантного анализа состоит в поиске решающего правила, минимизирующего ошибки классификации. Геометрически это эквивалентно поиску такой прямой z , в проекции на которую облака классов будут максимально удалены друг от друга, а площадь области пересечения гистограмм различных классов станет минимальной. Тогда на прямой z , которая называется дискриминантной функцией, окажутся спроектированными и границы классов z_0 . Точка z_0 , которая характеризует границу классов, называется дискриминантным индексом.

Переноса рассуждения из двумерного пространства признаков в многомерное, следует иметь в виду, что вместо прямой линии дискриминантная функция будет геометрически отображаться плоскостью раздела классов (гиперплоскостью для числа признаков более трех).

Вычисление дискриминантных функций

Предположим, что эмпирические гистограммы распределения наблюдений по классам соответствуют нормальному закону. Рассмотрим построение и расчет дискриминантных функций на основе правила Байеса, которое обеспечивает минимизацию ошибки неправильной классификации. Обозначим через c_1 ошибку неверного отнесения объекта из 1-го класса во 2-й, а через c_2 — ошибку неверного отнесения объекта из 2-го класса в 1-й. Тогда решающее правило заменяется следующим образом:

$$\begin{cases} x \in A_1, & \text{если } \frac{P_1(x)}{P_2(x)} \geq \frac{c_2 p_2}{c_1 p_1}, \\ x \in A_2, & \text{если } \frac{P_1(x)}{P_2(x)} < \frac{c_2 p_2}{c_1 p_1}. \end{cases} \quad (3.4)$$

Здесь $P_i(x)$ — плотность распределения вероятностей появления данного наблюдения в классе A_i , p_i — априорно задаваемая величина вероятности того, что это наблюдение проведено над объектом из i -го класса.

Значения p_i задаются исследователем до проведения дискриминантного анализа на основе предположения о вероятности появления в будущем наблюдения, принадлежащего i -му классу. Нередко в случае отсутствия подобной информации вероятность p_i задается одинаковой для всех классов $p_1 = p_2 = \dots = p_k$, что соответствует ситуации равновероятного отнесения объекта к любому из классов.

Полученное общее выражение можно упростить в предположении:

— равенства векторов дисперсий признаков по классам: $s_1 = s_2$;

— равенства ошибок неправильного распознавания: $c_1 = c_2$.

Тогда, прологарифмировав выражение (3.4), получим явное выражение для дискриминантной функции:

$$\ln \left\{ \frac{P_1(x)}{P_2(x)} \right\} = z(x) - z_0 = \begin{cases} \geq 0, & x \in A_1, \\ < 0, & x \in A_2. \end{cases}$$

Полагая, что дискриминантная функция $z(x)$ линейно зависит от параметров, будем искать ее значение для i -го наблюдения на основе классического регрессионного подхода в виде

$$z(x) = z_0 + \sum_{j=1}^M b_j x_{ij} = z_0 + \mathbf{B} \cdot \mathbf{x}_i^T. \quad (3.5)$$

Здесь z_0 — свободный член функции, вычисляемый по формуле

$$z_0 = \frac{1}{2} \sum_{j=1}^M b_j (\bar{x}_{1j} + \bar{x}_{2j}) = \frac{1}{2} \mathbf{B} \cdot (\bar{\mathbf{x}}_1^T + \bar{\mathbf{x}}_2^T),$$

$\mathbf{B}$ — вектор-столбец коэффициентов дискриминантной функции, который рассчитывается по формуле

$$\mathbf{B} = \mathbf{S}^{-1} \cdot (\bar{\mathbf{x}}_1^T - \bar{\mathbf{x}}_2^T), \quad (3.6)$$

где $\mathbf{S}$ — матрица ковариаций признаков, полученная по всей выборке.

Полученная в результате дискриминантная функция имеет конкретные значения для любого вектора наблюдений из рассматриваемой выборки. Граница раздела классов задается дискриминантным индексом, который был вычислен как свободный член z_0 .

На практике процедура распознавания образов с помощью методики МДА сводится к расчету значения дискриминантной функции для нового вектора наблюдений $\mathbf{x}_i = \{x_{i1}, x_{i2}, \dots, x_{iM}\}$. Если величина дискриминантной функции $z(\mathbf{x}_i)$ больше нуля, то новый вектор наблюдений следует отнести к классу A_1 , и, наоборот.

Изложенный байесовский подход в дискриминантном анализе при своей простоте и эффективности имеет один существенный недостаток. Он требует выполнения условия нормальности распределения выборок исходных классов. В практике экспериментальных исследований удовлетворить такому требованию удается отнюдь не всегда. Альтернативой байесовскому методу может служить непараметрический подход, предложенный Р. Фишером.

Достоинством непараметрического МДА является то, что здесь не накладывается условий на вид вероятностного распределения по классам, и для решения задачи распознавания необходимо знать только принадлежность наблюдений классам, оценки

векторов средних значений признаков по классам $\bar{x}_1, \bar{x}_2$, оценку обобщенной для всей выборки матрицы ковариаций S между признаками.

Стратегия непараметрического МДА базируется на неизменном предположении линейности дискриминантной функции $z(x)$. Поиск коэффициентов дискриминантной функции производится на основе удовлетворения условию максимизации разности средних значений между классами:

$$(\bar{x}_1 - \bar{x}_2) \rightarrow \max.$$

Представим это условие в матричном виде, выделив вектор средних значений первого класса $\bar{x}_1$, вектор средних значений второго класса $\bar{x}_2$ и вектор средних значений для всей выборки данных $\bar{x}_0 = \{\bar{x}_1, \bar{x}_2, \dots, \bar{x}_M\}$. Составим матрицы различий средних для классов в результате векторного умножения:

$$U_1 = (\bar{x}_1 - \bar{x}_0) \cdot (\bar{x}_1 - \bar{x}_0)^T,$$

$$U_2 = (\bar{x}_2 - \bar{x}_0) \cdot (\bar{x}_2 - \bar{x}_0)^T.$$

Составим из матриц U_1 и U_2 обобщенную матрицу различий средних:

$$U = N_1 U_1 + N_2 U_2,$$

где N_1 и N_2 — число наблюдений в 1-м и 2-м классах.

Составленная матрица различий средних U определяет в пространстве из M исходных переменных ориентацию эллипса рассеяния различий классов. Большая ось эллипса различий совпадает по направлению своей ориентации с расположением дискриминантной функции. Поэтому значения первого собственного вектора матрицы U будут одновременно коэффициентами дискриминантной функции.

Целесообразно в данном случае строить дискриминантную функцию следующим образом. Сначала необходимо решить характеристическое уравнение вида

$$|U - \lambda S| = 0 \quad (3.7)$$

и определить первое максимальное собственное число λ_1 , затем подставить λ_1 в систему уравнений вида

$$B \cdot (U - \lambda S) = 0 \quad (3.8)$$

и решить ее относительно первого собственного вектора B_1 . Коэффициенты собственного вектора $b_{11}, b_{12}, \dots, b_{1M}$ будут коэффициентами дискриминантной функции.

Свободный член z_0 дискриминантной функции в этом случае вычисляется по формуле

$$z_0 = \frac{\bar{x}_1 \cdot s_1 + \bar{x}_2 \cdot s_2}{s_1 + s_2},$$

где $s_1 = \frac{1}{N_1} \sum_j (z_j - \bar{z}_1)$ — значения дисперсии величин дискриминантной функции z для тех векторов, которые принадлежат к первому классу; s_2 — то же для второго класса.

Применение МДА в океанологии

Основное содержание МДА состоит в оптимальном разбиении объектов на классы по выработанному эффективному математическому правилу (дискриминантной функции). Поэтому прикладное значение МДА можно рассматривать в двух аспектах: в результатах собственно классификации исходных наблюдений и в эффективности полученной дискриминантной функции для других аналогичных вычислений.

Физически интерпретируемые, достоверные результаты группировки наблюдений с помощью дискриминантных функций обусловили серьезный практический интерес гидрометеорологов к этому методу еще в конце 60-х годов [132, 198].

Анализ накопленного опыта применения МДА [83] уже в 70-е годы показал, что наиболее эффективен метод в случае априорно известных характеристиках исследуемых классов наблюдений. Другими словами, необходимость применения обучающей выборки стала очевидной. Поэтому в дальнейшем применение МДА строилось на базе обработки многомерных выборок наблюдений, классифицируемость которых на заданные классы предопределялась самой структурой данных. Такими классами стали грозы, гололед, ливневые осадки в метеорологическом приложении [84, 85], ледовые фазы, формы льда, типы волнения в океанологии [145].

Современный этап применения МДА для классификационных задач характеризуется внедрением этой методики в область обработки результатов дистанционного (космического) зондирования океана. Выявляемые классы уже заранее определены — это прежде всего суша, океан, облачность. Многомерность поступающей информации предопределяется наличием многозональных снимков земной поверхности. Смысл применения МДА состоит в эффективном дешифрировании снимков в заданных исследователем терминах (классах) [10, 11, 116, 254].

Аналогичный подход в последнее время развивается и для обработки сигналов авиационных и наземных радиолокационных станций в применении к анализу локальных неоднородностей морской поверхности и льда [224, 252]. В подобных приложениях значительную роль играет скорость обработки информации. Относительная простота и высокая скорость вычислительных алгоритмов МДА обусловили его популярность для успешного решения задач оперативной обработки различных видов изображений.

Важное прикладное содержание дискриминируемых классов предопределили еще одно направление применения МДА. Это направление — прогноз развития гидрометеорологических процессов.

В данной области процедуры дискриминантного анализа обеспечивают представление исторических архивов гидрометеорологических наблюдений в виде чередования отдельных классов [9, 63, 108, 110, 135, 156, 227].

Достоверность каждого класса подкрепляется конкретными значениями целого набора гидрометеорологических характеристик (предикторов), формирующих «вектор климатического состояния» [62, 117]. Предсказание будущего класса с соответствующим ему вектором состояния формируется либо на основе найденной вероятностной закономерности взаимной смены классов, либо на основе регрессионной экстраполяции отдельных предикторов. Подобный подход особенно успешно зарекомендовал себя в прогнозах опасных явлений [130, 152], а также ледовых условий [127, 129].

Однако в целом ряде случаев достоинства МДА, обеспечившие ему заслуженную популярность, одновременно становятся недостатками, снижающими качество дискриминации. Так, простота классического линейного МДА обусловлена априорным требованием равенства ковариационных матриц для различных классов. На практике это требование отнюдь не всегда статистически проверяется. Неравенство же дисперсий внутри классов приводит к смещению оцениваемых дискриминантных индексов самих классов и их границ. В любых процедурах классификации основная проблема сводится к надежному выделению границ классов. Линейный МДА может приводить к серьезным погрешностям оценки границ классов в случае статистически значимых различий ковариационных матриц классов.

Избавиться от подобного недостатка в известной мере возможно, если сохранить в расчетных формулах дискриминантной функции внутриклассовые ковариации [198, 206, 208]. Однако в этом случае дискриминантная функция становится нелинейной. Гидрометеорологи давно обратили внимание на эту проблему. Нелинейный дискриминантный анализ уже неоднократно применялся для решения прикладных задач [180, 227, 279].

Вторая принципиальная проблема МДА связана с видом функции распределения признаков внутри классов. Традиционно предполагается наличие многомерного нормального распределения признаков. На практике искажение эмпирического закона распределения приводит непосредственно к ошибкам в расчетах вероятностей принадлежности наблюдений к классам.

На практике эта проблема решается либо дополнительным подбором информативных признаков, либо применением непараметрических МДА. Идея подбора признаков, оптимальных для дискриминирования наблюдений, чаще всего решается с помощью методики пошагового дискриминантного анализа [206, 259, 279].

Как и в пошаговом регрессионном анализе, здесь реализованы принципы последовательного исключения или последовательного включения признаков. Критериями эффективности отбора признаков принимаются вычисляемые оценки D^2 -расстояния Махалано-

биса и уровень выполнения гипотез соответствия нормальному закону распределения.

Техника непараметрического МДА сегодня продолжает развиваться. Методика МДА, основанная на решении проблемы собственных чисел и векторов блочной ковариационной матрицы классов, нашла свое применение для решения многих океанологических задач в работах Ю. В. Николаева [130]. Этот подход, известный под названием канонического дискриминантного анализа, остается и предметом теоретических исследований [94, 95].

Выводы статистического диагноза дают представление о механизме многомерного процесса, который определен с заданной точностью через конечное число известных статистических переменных, что можно считать достаточной основой для разработки прогностических схем процессов. Рассмотрим далее эту проблему, оставаясь в рамках двух основных диагностических подходов — структурного и классификационного представления многомерных процессов. Основная задача настоящей главы — определить, какие методы статистико-вероятностного прогнозирования обобщенных показателей следует предпочесть в конкретных ситуациях. Множественность возможных методов прогноза в свою очередь выдвигает задачу эффективного объединения (комплексирования) нескольких методов с целью выработки оптимального прогноза. Обсуждая существенные преимущества такого методического приема по сравнению с традиционными нельзя тем не менее не учесть и его объективные ограничения, которые диктуются самой статистической предсказуемостью исследуемого процесса. Это понятие в применении к многомерным океанологическим процессам еще не было проанализировано в прикладной литературе, хотя представляет практический интерес, поскольку позволяет от абстрактной модельной вседозволенности прогноза вернуться к жизненным реалиям океанологического прогнозирования.

4.1. Предсказание структуры процессов

Структура океанологических процессов в результатах ортогонального разложения матрицы наблюдаемых переменных обобщена в двух наборах статистических параметров — это набор собственных векторов или факторных нагрузок и набор значений главных компонент или общих факторов. Нагрузки компонент и факторов описывают достоверные взаимосвязи наблюдаемых переменных и, как правило, устойчивы во времени. Каждый из общих факторов характеризует эволюцию процесса во времени. Поэтому суть прогнозирования структуры океанологического процесса состоит в задаче экстраполяции общих факторов (или главных компонент).

В соответствии с обозначениями, принятыми в главе 2, запишем основное уравнение модели факторного анализа с учетом прогностических задач:

$$\mathbf{X} = \mathbf{F} \cdot \mathbf{A}^T + \mathbf{E}.$$

Здесь $\mathbf{X}$ — матрица исходных наблюдений, число строк в которой $i = 1, \dots, N$ соответствует числу наблюдений; число столбцов $j = 1, \dots, M$ задает мерность процесса. Смысл применения модели МФА состоит в предсказании M -мерного вектора $\hat{\mathbf{x}}_{(i)} = \{\hat{x}_{(i1)}, \hat{x}_{(i2)}, \dots, \hat{x}_{(iM)}\}$ исходных переменных с учетом их взаимосвязей на интервал заблаговременности (τ) при условии сохранения определенного уровня ошибок ($\mathbf{E}$). Логика прогноза состоит в переходе от исходного M -мерного пространства наблюдений ($\mathbf{X}$) к сокращенному K -мерному пространству обобщенных функций ($\mathbf{F}$), последующей экстраполяции функций $\hat{\mathbf{F}}$ на интервал времени τ и заключительном восстановлении прогностических значений M -мерного вектора $\hat{\mathbf{x}}_{(i+\tau)}$. Основные достоинства данного подхода состоят в следующем.

Отпадает необходимость прогнозирования шумовой компоненты в исходных наблюдениях. Основное внимание сосредоточивается на прогнозе систематической изменчивости в наблюдаемых параметрах, которая описывается матрицей

$$\hat{\mathbf{X}} = \hat{\mathbf{F}} \cdot \mathbf{A}^T.$$

Ведущие статистические закономерности эволюции процесса аккумулированы в функциях $\mathbf{F}$, что повышает статистическую предсказуемость последних по сравнению с предсказуемостью отдельных наблюдаемых параметров. При этом сокращается число формально экстраполируемых переменных: M -мерный процесс прогнозируется в сокращенном K -мерном пространстве, где $K \ll M$. Для МФА всегда выполняется условие $K \leq M/2$, причем отбор числа прогнозируемых переменных f_j определяется не только общими для модели статистическими критериями, но и способностью конкретной обобщенной функции описывать физически информативную долю изменчивости процесса.

Упрощается собственно процедура экстраполяции обобщенных функций. Полученное K -мерное пространство описания процесса задано взаимноортогональными осями, что определяет статистическую взаимонезависимость обобщенных функций между собой на период прогноза. Следовательно, каждая из них может быть экстраполирована во времени с помощью простейших малопараметрических уравнений автопрогноза:

$$\hat{f}(i + \tau, j) = \sum_{l=1}^p b^{(\tau)}(j, l) f(i - l + 1, j). \quad (4.1)$$

Здесь $\hat{f}$ — прогностическое значение j -го общего фактора (компоненты), полученного на момент времени $i + \tau$ с упреждением τ ;

$\hat{f}$ — значение фактора, вычисленное по диагностической модели;
 b — l -й коэффициент прогностической модели p -го порядка для j -го общего фактора с упреждением τ . Взаимонезависимость общих факторов позволяет отдельно для каждого из них подобрать оптимальный тип и порядок прогностической модели, чтобы добиться минимизации дисперсии суммарной ошибки экстраполяции.

В результате вычислений формируется вектор прогностических оценок K общих факторов:

$$\hat{f}(i + \tau) = \{\hat{f}(i + \tau, 1), \hat{f}(i + \tau, 2), \dots, \hat{f}(i + \tau, K)\},$$

который проектируется в M -мерное пространство наблюдений с помощью диагностированной матрицы перехода (A):

$$\hat{x}(i + \tau) = \{\hat{x}(i + \tau, 1), \hat{x}(i + \tau, 2), \dots, \hat{x}(i + \tau, M)\} = \hat{f}(i + \tau) \cdot A^T.$$

Прогностическая оценка значений общих факторов неизбежно будет содержать ошибки, которые в свою очередь приведут к возникновению ошибок в прогностическом векторе исходных переменных:

$$x(i + \tau) - \hat{x}(i + \tau) = \hat{e}(i + \tau).$$

Таким образом в матрице остатков (E) появляются добавочные ошибки ($\hat{e}$), которые вместе с ошибками диагноза (e) определяют общее качество модели диагноз + прогноз. Каждая ошибка прогнозирования j -й исходной переменной $\hat{e}(i + \tau, j)$ будет включать в себя суперпозицию ошибок прогноза каждого из K общих факторов $\hat{e}^{(f)}$

$$x(i + \tau) = \hat{f}(i + \tau) \cdot A^T + e^f(i + \tau) \cdot A^T, \quad (4.2)$$

причем вклад ошибок будет пропорционален вкладу соответствующих членов разложения, который задается соответствующими коэффициентами в матрице A . Минимизации ошибок прогнозирования исходных переменных можно ожидать только при условии возможно более точной экстраполяции наиболее важных (с максимальными вкладами) членов разложения — факторов или компонент.

Методы экстраполяции общих факторов

Одно из важнейших достижений диагноза на основе применения ортогональных разложений состоит в аккумуляции связанной изменчивости исходных переменных на первых членах разложения. Поэтому первые общие факторы (главные компоненты) времени оказываются рядами, последовательные значения которых взаимозависимы, что отражается в соответствующей структуре автокорреляционных функций и автоспектров. Этим объясняется возможность построения эффективных прогностических оценок для первых общих факторов на основе моделей авторегрессии порядка

$p(AP - p)$. Статистическая структура первых общих факторов выражает наиболее крупномасштабные закономерности формирования исследуемых океанологических процессов, которые, как правило, обусловлены либо автоколебаниями внутри системы, либо общими для исследуемых параметров внешними ритмическими воздействиями (суточные, сезонные и межгодовые вариации).

Накопленный опыт применения AP -моделей к прогнозу крупномасштабных гидрометеорологических процессов показывает, что оптимальный порядок модели колеблется в пределах $p = 1 \dots 10$, а чаще всего сохраняется на уровне $p = 1 \dots 4$ [142, 146, 151, 212]. Такие модели позволяют описать ряд типов процессов — инерционные, стохастические, квазиритмические; все зависит от порядка модели и подобранных коэффициентов.

Эти вопросы здесь подробно не обсуждаются, поскольку методика расчета вектора коэффициентов авторегрессии (**В**) для общих факторов ничем не отличается от классической, которая подробно изложена в известных монографиях [33, 93, 146, 216].

Экстраполяция членов разложения высоких порядков

Благодаря известным фильтрующим свойствам методов факторного и компонентного анализа мелкомасштабная, стохастическая изменчивость исходных переменных группируется в значениях компонент и факторов больших номеров. Спектральная структура таких колебаний приближается к белому шуму, что существенно ограничивает возможности экстраполяции данной части суммарной изменчивости. В ряде случаев прогноз таких факторов выполняется на основе модели скользящего среднего порядка q ($СС - q$) либо с применением смешанной модели ($АРСС - p, q$) [33, 146, 216]. Учитывая малую значимость этих факторов в общей динамике процесса, их детальным прогнозированием можно пренебречь и ограничиться их оценкой как остаточных разностей значений фактических и прогнозируемых общих факторов.

Экстраполяции факторов с неоднородной структурой

Простейший случай неоднородности связывается с модуляцией во времени первичных статистических характеристик общих факторов — оценок математического ожидания и дисперсии. Для таких факторов не представляется возможным построить единственную прогностическую модель с постоянными коэффициентами во времени. Наоборот, возникает необходимость на каждом шаге усваивать новую информацию и адаптировать прогностические коэффициенты. Для применяемых прогностических моделей авторегрессии это было сделано следующим образом. Сначала на шаге i пересчитывались:

оценка математического ожидания

$$\bar{f}(i) = \bar{f}(i-1) + \frac{1}{N_i} (\bar{f}(i) - \bar{f}(i-1)),$$

оценка среднего квадратического отклонения

$$\sigma(i) = \sigma(i-1) + \frac{1}{N_i} (|\hat{f}(j) - \bar{f}(i)| - \sigma(i-1)),$$

оценка коэффициентов автокорреляции для сдвигов $l = 1, \dots, p$

$$r_i(l) = r_{i-1}(l) + \frac{1}{N_i} [(f(i) - \bar{f}(i))(f(i-l) - \bar{f}(i)) - r_{i-1}(l)].$$

Затем производился пересчет коэффициентов регрессии для автопрогноза j -го общего фактора (индекс j для простоты опущен).

Другой весьма эффективный для решения прогностических задач алгоритм прогноза неоднородных временных рядов основан на использовании процедуры оптимальной фильтрации Калмана [40, 220, 250, 269]. Применение этого алгоритма для моделей авторегрессии позволяет быстро перестраивать коэффициенты регрессии с учетом самых последних изменений в общих факторах и исходных переменных. Так, для модели авторегрессии первого порядка $AR-1$ соответствующее выражение имеет вид

$$\hat{f}(i) = b(1, i) \hat{f}(i-1), \quad (4.3)$$

$$b(1, i) = b(1, i-1) + s(i) \hat{f}(i-1) [f(i-1) - \hat{f}(i-1) b(1, i-1)],$$

$$s(i) = s(i-1) - \frac{(s(i-1) \hat{f}(i-1))^2}{1 + s(i-1) \hat{f}^2(i-1)},$$

где $s(i)$ — коэффициент усиления Калмана, значения которого определяют уровень усвоения изменчивости данных моделью на i -м шаге; $b(1, i)$ — коэффициент авторегрессии на i -м шаге, $\hat{f}$ — прогнозируемое значение общего фактора. Параметры s и b оцениваются по рекуррентным соотношениям и являются своеобразными показателями динамики исследуемого процесса. С их помощью эффективно используется в модели память процесса в сочетании с быстрой кратковременной перестройкой процесса. Поэтому такая модель названа в дальнейшем изложении динамико-стохастической.

4.2. Предсказание классов состояния процессов

Попытка прогноза класса на практике чаще всего обусловлена двумя причинами — это либо единственный эффективный путь прогноза эволюции триггер-процесса со множеством квазиустойчивых состояний, либо просто последний шанс вообще хоть как-то оценить тенденцию развития стохастического процесса. Содержание проблемы прогноза состоит в том, чтобы не только выявить и описать устойчивые состояния многомерного процесса, что может быть сделано уже в рамках статистического диагноза, но и качественно описать закономерность сменяемости во времени найденных состояний.

Как и в случае прогноза структуры процесса, здесь экстраполируются некие обобщенные показатели процесса. В качестве таких показателей выступают классы состояния, которые в результате диагноза характеризуются векторами центров классов и векторами внутриклассовой вариации наблюдаемых параметров. Центром i -го класса чаще всего выступает M -мерный вектор средних значений наблюдений в данном классе $\bar{x}_i = \{\bar{x}_{i1}, \bar{x}_{i2}, \dots, \bar{x}_{iM}\}$. Однородность класса часто характеризуется через вектор средних квадратических характеристик $\sigma_i = \{\sigma_{i1}, \sigma_{i2}, \dots, \sigma_{iM}\}$. Поэтому прогноз класса в конечном итоге сводится к прогнозу этих параметров — x_i и σ_i .

Схема прогноза

Пусть матрица X отображает N наблюдений M -мерного случайного процесса, класс состояния которого необходимо спрогнозировать. Результатом диагностических исследований являются « K » однородных классов $\omega_i, i=1, 2, \dots, k$, каждый из которых представляет типовое взаимообусловленное состояние параметров

процесса через векторы средних $\bar{x}_i$ и вариаций σ_i . Строго говоря, каждый класс характеризуется M -мерной эмпирической функцией распределения. Если она статистически значимо отличается от нормального закона распределения, то характеристиками классов следует взять другие статистические параметры, такие как мода, размах, квантили, которые для каждого класса будут также представлены в виде M -мерных векторов. Оценка выборочных статистик обсуждается применительно к решению задач гидрометеорологии в работах [65, 69, 76, 87, 96, 155, 212]. Сохраняя далее обозначения векторов $\bar{x}_i, \sigma_i$, будем иметь в виду потенциальную возможность их замены, если этого потребует достижение более точных прогностических оценок.

Прогноз класса состояния процесса строится на основе вероятностных процедур оценки сменяемости классов. Сменяемость оценивается по каталогу классов, который составляется в результате диагноза. Каталог представляет собой аппроксимацию развития процесса X во времени в терминах классов в виде последовательности W длиной N значений, где каждому вектору исходных наблюдений $x_i = \{x_{i1}, x_{i2}, \dots, x_{iM}\}$ поставлен в соответствие один из k классов. Таким образом, M -мерный процесс представляется в виде ранжированной последовательности классов, содержащей целые значения номеров классов от 1 до k , которая и будет экстраполирована статистико-вероятностными методами. Схематично перечисленные операции можно записать в виде

$${}_N X^M \Rightarrow {}_K \Omega^M \Rightarrow {}_N W^1 \Rightarrow {}_K P^K \Rightarrow \omega_i \Rightarrow x_i, \sigma_i.$$

Здесь представляет интерес тот факт, что собственно в прогностической матрице P само по себе время отсутствует. Оно заменено вероятностями p_{ij} смены одного класса другим: ω_j на ω_i .

Матрица вероятностей $\mathbf{P}$ фактически описывает динамику M -мерного процесса не в пространстве наблюдений, а в фазовом пространстве состояний. Поэтому содержание прогноза сводится к поиску наиболее вероятного состояния (фазы), которая последует за текущей фазой развития процесса (стратегия 1) или к поиску вероятности состояния априорно заданной фазы процесса (стратегия 2).

Стратегия одношагового прогноза

Общим для большинства стратегий вероятностного прогнозирования остается постулат о том, что матрица переходных вероятностей классов $\mathbf{P}$ является исчерпывающей характеристикой для разработки прогноза. Формулировка прогноза состоит в экстраполяции вероятностей появления классов на период заблаговременности с помощью правила, выработанного на основе принятой вероятностной модели, каталога классов.

Основой выработки правила прогноза служит чаще всего предположение о марковости и однородности последовательности классов $\mathbf{W}$, которая проверяется с помощью χ^2 -критерия Пирсона [44, 65, 269]. Матрица $\mathbf{P}$ при выполнении данной гипотезы аппроксимируется через обычные условные вероятности появления классов:

$$p_{ij} = n_{ij}/N,$$

где n_{ij} — число случаев наблюдения класса ω_i после класса ω_j . Тогда реализация стратегии состоит в том, чтобы определить класс ω_i с максимально возможной переходной вероятностью в текущей (j) фазе развития процесса. Схематично эту логическую цепочку можно записать в следующем виде:

$$\omega_j \Rightarrow \max_{i=1, K} (p_{ij}) \Rightarrow p^{\max} \Rightarrow \omega_i.$$

Существенно более сложная ситуация возникает для реализации стратегии в случае невысокой вероятности появления p_i прогнозируемых классов ω_i . При этом существенно возрастает неопределенность оценки значений условных переходных вероятностей p_{ii} , p_{ij} и снижается качество прогнозирования на основе классического марковского подхода к анализу вероятностей. Реальной альтернативой может быть байесовская статистика, в которой логика оценки вероятностей строится с учетом количественной оценки неопределенности вырабатываемых прогностических параметров. Для этого вводится понятие апостериорной переходной вероятности, которая вычисляется по формуле

$$p_{ij}^{\phi} = \frac{p_i p_{ij}}{\sum_{l=1}^K p_l p_{il}} a_p \hat{x}_{ip}, \quad (4.4)$$

и содержит, в частности, оценки условных переходных вероятностей p_{ij} и априорных климатических вероятностей p_i появления

класса ω_i . Тем самым для каждой фазы процесса с помощью байесовского подхода рассчитываются условные функции распределения вероятностей, на основе анализа которых реализуются прогностические оценки. На рис. 4.1 схематизировано изображена эволюция многомерного процесса в терминах вероятностной динамики априорно известных классов состояния. Очевидно, что каждый раз фаза развития процесса характеризуется специфической функцией вероятности. Заметна изменчивость номера класса

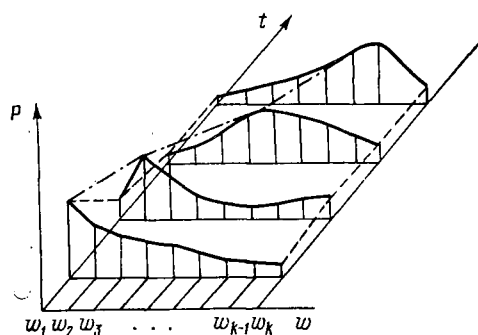


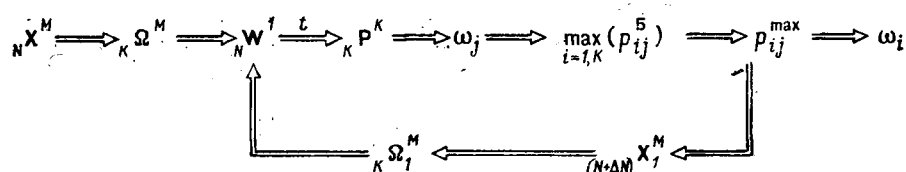
Рис. 4.1. Эволюция во времени многомерного процесса в терминах вероятностей динамики (p) априорно заданных классов состояния (w).

с максимумом переходной вероятности $p_{ij}^{\max}$ и колебания от фазы к фазе самой величины $p^{\max}$. Сохраняя обсуждаемую стратегию, ее можно реализовать простой подстановкой байесовских переходных вероятностей p_{ij}^{δ} вместо p_{ij} . Тогда прогностические оценки вероятностей можно выделить на рисунке штрихпунктирной кривой. Проекция этой кривой на плоскость (p, t) позволяет проследить по времени степень стохастизации процесса в каждой фазе его развития. Чем выше значение апостериорной вероятности ($p_{ij}^{\max} \rightarrow 1$), тем больше проявляется детерминированность в развитии процесса, когда сменяемость пары классов ω_j и ω_i во времени однозначно предопределена. Наоборот, при уменьшении апостериорной вероятности $p_{ij}^{\max} \rightarrow 0$ можно утверждать стохастизацию процесса в рамках заданного набора классов состояния, что приведет к снижению эффективности рассматриваемой прогностической методологии.

Усвоение информации при прогнозе класса

Стохастизация процесса, которая прослеживается в существенном снижении $p_{ij}^{\max}$, может рассматриваться как фактор обратной связи, если в прогностической стратегии предусмотреть возвратную цепочку. Дело в том, что нередко снижение $p_{ij}^{\max}$ обусловлено изменением внутренней структуры взаимосвязей наблюдаемых параметров, которые не укладываются в принятую на

этапе диагноза схему классификации. Поэтому необходимо вернуться к задаче оптимальной классификации, включив в диагностическую процедуру разбиения МАК всю вновь накопленную информацию. Логическая схема расчетов на момент времени t примет вид



Здесь X_1 — матрица исходных данных, дополненная наблюдениями, Ω_1 — скорректированное разбиение на классы, а W — обновленный каталог классов.

В заключение параграфа следует отметить, что в океанологической практике байесовские прогностические оценки используются пока крайне редко. Можно говорить о наличии лишь единичных публикаций по этой тематике [155, 224, 267]. Перспектива применения данного подхода связывается с возможностью принятия на его базе эффективных экономических решений.

4.3. Комплексирование прогностических моделей

Как это видно из предыдущих параграфов, трудно рекомендовать единственный оптимальный метод прогноза многомерных процессов. Чаще всего исследователь располагает возможностью реализовать прогноз по нескольким прогностическим моделям, выбирая затем наилучшую оценку вектора состояния процесса. Этот выбор определяется не всегда оптимально и во многом обусловлен субъективными факторами. Между тем существует реальная возможность формализовать процедуру оптимизации прогноза многомерного процесса, если использовать идеи статистической комплексации одномерных прогнозов и соответствующие формальные вычислительные алгоритмы.

Собственно, сама идея комплексации нескольких ненадежных методов прогнозов для выработки единственной оптимальной прогностической оценки обсуждалась в 60-е годы в работах Н. А. Багрова и Е. П. Борисенкова применительно к проблемам метеорологии. Соответствующая методология формировалась позднее также с учетом специфики метеорологических прогнозов [14, 35, 36, 65]. Попытаемся обобщить известные методы комплексации прогнозов временных рядов, чтобы в дальнейшем перейти к обсуждению возможности комплексации многомерных прогнозов.

Основное содержание методологии комплексации сейчас ограничено рамками линейных комплексирующих моделей вида

$$\hat{x}_{ik} = \sum_{p=1}^l a_p \hat{x}_{ip}, \quad (4.5)$$

где $\hat{x}_k$ — ряд комплексированных прогнозов, $i = 1, 2, \dots, N$; $\hat{x}_{ip}$ — прогностическая оценка величины x по p -му методу прогноза на i -й момент времени; a_p — коэффициенты комплексации. Методы комплексации различаются в зависимости от вида используемых прогностических моделей (количественные или качественные), а также по методу расчета коэффициентов комплексации. При оценке коэффициентов a_p используют несколько подходов. Самый простой — комплексация среднеарифметическим, когда коэффициенты равны между собой:

$$a_1 = \dots = a_l = 1/l.$$

Другой путь — регрессионная комплексация, когда значения a_p соответствуют коэффициентам регрессии $\hat{x}_p$ на x . Расчет проводится по вычислительной схеме множественной линейной регрессии (см. главу 1), что позволяет на рассматриваемой исходной выборке минимизировать среднюю квадратическую ошибку комплексного прогноза ε которая оценивается по рабочей формуле

$$\varepsilon = \sigma_x \sqrt{|\mathbf{D}|/|\mathbf{D}_{11}|}.$$

Здесь $|\mathbf{D}|$ — определитель, а $|\mathbf{D}_{11}|$ — соответствующий минор полной корреляционной матрицы прогнозов, σ_x — среднеквадратическое отклонение наблюдаемого ряда x .

Точность комплексированных прогнозов будет при таком подходе описываться через коэффициенты корреляции между наблюдаемыми x и прогнозируемыми $\hat{x}_p$ значениями. В случае взаимонезависимости рядов прогностических оценок соответствующая формула имеет вид

$$\varepsilon = \sigma_x \sqrt{1 - \sum_{j=1}^l r^2(x, \hat{x}_j)}.$$

Проанализируем возможности комплексации прогнозов для многомерного процесса $\mathbf{X}$, составленного из N наблюдений за M переменными (признаками). Пусть для каждого признака существует возможность дать N последовательных во времени прогнозов с использованием l различных методов. Тогда трехмерную матрицу прогнозов размером $N \times M \times l$ необходимо оптимизировать относительно матрицы наблюдений $N \times M$. Попытаемся для этой цели использовать методику регрессионной комплексации, для чего в предположении однородности исходных наблюдений представим матрицу $\mathbf{X}$ в виде вектора-столбца $(N \times M) \times 1$. Аналогичным образом преобразуем матрицу комплексируемых прогнозов $\hat{\mathbf{X}}$ в двумерную матрицу из $(N \times M)$ строк и l столбцов. Для

расчета комплексированных прогнозов при таком представлении исходных данных достаточно оценить вектор из l коэффициентов регрессии A и составить уравнение:

$$_{(N \times M)} X^l = _{(N \times M)} \hat{X}^l \cdot {}_l A^l + \varepsilon.$$

Соответствующие вычисления можно выполнить по алгоритму множественной линейной регрессии (см. главу 1), а среднюю квадратическую ошибку комплексации — оценить по формулам, приведенным выше.

Успешность регрессионной комплексации и минимизация ошибок будут определяться двумя условиями:

- статистической однородностью вектора наблюдений X ,
- устойчивостью взаимосвязей $r(x, \hat{x}_j)$.

В случае если матрица наблюдений представляется как неоднородная выборка N наблюдений M переменных, целесообразно обратиться к вероятностной комплексации [44, 65, 76, 227]. При этом последовательность наблюдений представляется как чередование k -классов (состояний) процесса, в терминах которых даются и прогнозы по l методам. Тогда комплексный прогноз в вероятностной форме может быть записан следующим образом:

$$P = \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_j \\ \vdots \\ P_k \end{bmatrix} = \left[\sum_{i=1}^l a_i p_{ij} \right], \quad (4.6)$$

где a_i — априорно заданный вес i -го метода прогнозирования, p_{ij} — вероятность прогнозирования i -м методом j -й градации состояния процесса. Решающее правило для оценки весов отдельных методов может задаваться на основе формулы Байеса. В результате расчетов комплексный прогноз многомерного процесса будет представлен таблицей вероятностей появления всех возможных состояний с учетом всех исходных методов прогнозирования.

4.4. Статистическая предсказуемость океанологических процессов

Оценка изменчивости океанологических процессов непосредственно связана с анализом их предсказуемости. Для успешного решения этой задачи, во-первых, требуется построение адекватных математических моделей исследуемых процессов; во-вторых, необходимо описание прогностических свойств этих моделей, и в частности определение пределов предсказуемости процессов на различных пространственно-временных масштабах. Основная задача классической теории предсказуемости состоит в том, чтобы

оценить скорость накопления существующих малых ошибок в прогностической модели. В любой модели предсказуемость будет зависеть от природы ошибок принятой модели процесса, а также от ошибок наблюдения процесса, положенного в основу прогноза. Океан понимается как сложная многокомпонентная система, взаимодействие между составляющими которой носит стохастический характер. В соответствии с этим определением вполне уместно обратиться к статистическим моделям, которые дают представление об эволюции гидрометеорологических явлений во времени как о случайном процессе $X(t)$ и позволяют поставить задачу об изучении возможностей статистической экстраполяции этого процесса. Статистические модели находят все более эффективное применение для целей диагноза и прогноза гидрометеорологических процессов. С их помощью описывается статистическая структура и, как следствие этого, устанавливаются пределы статистической предсказуемости исследуемых процессов [151]. Корректное решение данной задачи во многом зависит от вида исходной информации (мерности вектора признаков, дискретности наблюдений) и вида выбранной статистической модели.

Предсказуемость одномерной модели процесса

В настоящее время при исследовании океанологических процессов большинство авторов отдают предпочтение моделям авторегрессии 1-го порядка (АР-1). Именно их параметры ложатся в основу выводов о статистической структуре временных рядов и пределах статистической предсказуемости. Это положение во многом объясняется тем, что атмосферные процессы описываются спектральной моделью типа «белого шума», а океанологические процессы — моделью «красного шума». На основании этого следует вывод о незначительной инерционности гидрометеорологических процессов, а пределом их статистической предсказуемости является упреждение в один шаг по времени.

Данные результаты однозначно определяют предсказуемость 1-го и 2-го рода [244]. Напомним, что предсказуемость 1-го рода — это временной интервал упреждения (τ), при котором средняя квадратическая ошибка прогнозирования (σ_e^2) становится близкой к величине дисперсии процесса (σ_{cp}^2).

Таким образом, статистическая предсказуемость 1-го рода количественно характеризуется ошибкой прогноза: чем меньше ошибка, тем выше предсказуемость. В практике морских прогнозов обычно используют относительную дисперсию ошибок прогноза d или критерий качества прогнозирования Q :

$$d = \sigma_e^2 / \sigma_{\Phi}^2, \quad Q = \sqrt{d}, \quad (4.7)$$

где σ_e^2 — дисперсия ошибки прогноза, σ_{Φ}^2 — дисперсия прогнозируемого ряда.

Очевидно, чем меньше величины d и Q , тем лучше подогнана прогностическая модель. В случае $d = Q = 0$ прогнозируемый процесс является детерминированным и, следовательно, однозначно описывается в рамках детерминистических моделей. При $d = Q = 1$ процесс представляет собой белый шум и наилучшим прогнозом является климатический прогноз, т. е. прогноз среднего значения.

Другой оценкой эффективности прогнозирования является косвенная оценка d по коэффициенту корреляции r между фактическими и спрогнозированными значениями в зависимости от упреждения прогноза [151]:

$$d = \sqrt{1 - r^{2\tau}}, \quad (4.8)$$

где τ — предел статистической предсказуемости. Определение предела статистической предсказуемости τ , как следует из [244], непосредственно зависит от параметров применяемых статистико-вероятностных моделей, и в частности от относительных дисперсий ошибок прогнозирования

$$\tau = \frac{\ln(1-d)}{2 \ln r}, \quad (4.9)$$

где d принимается меньшим 1. Заметим, что все приведенные критерии являются статистическими оценками, точность которых определена доверительными интервалами входящих в них параметров. В этой связи приведем еще один критерий для определения предела предсказуемости:

$$\tau \leq z - 2\sigma_z, \quad (4.10)$$

где z — параметр Фишера ($z = \frac{1}{2} \ln \frac{1+r}{1-r}$); σ_z — теоретическая среднеквадратическая ошибка расчета ($\sigma_z = 1/(N-3)$).

Анализ гидрометеорологических рядов с помощью линейных статистических моделей проводился многими авторами [35, 65, 69, 135, 146]. В результате получены важные соотношения между параметрами моделей и возможными ошибками прогноза. В частности, установлено, что прогнозирование имеет смысл для $d \leq 0,9$. Однако в качестве практических критериев прогноза выступают более конкретные и довольно жесткие условия: $0,674\sigma_\Phi$ и $0,8\sigma_\Phi$ — значения, соответствующие $0,45d$ и $0,74d$. Поэтому представляется важным установить соотношения, связывающие оправдываемость прогноза при заданных критериях точности с параметрами статистико-вероятностных моделей.

Так, уровень $d = 0,45$ обеспечивает оправдываемость прогноза при условии невыхода ошибки прогноза за $0,674\sigma_\Phi$ в 100 % случаев. Уровень $d = 0,74$ обеспечивает 100 %-ную оправдываемость прогноза при $\varepsilon \leq 0,8\sigma_\Phi$. Из приводимых соотношений несложно установить линейную связь между d и оправдываемостью метода прогноза при заданном уровне ошибки, из которой следует, что уровень $d \leq 0,74$ обеспечивает в среднем оправдавшийся прогноз по критерию $\varepsilon \leq 0,674\sigma_\Phi$ (рис. 4.2).

Эти результаты, по нашему мнению, являются важными для практика-прогноста, так как дают возможность установить не только пределы статистической предсказуемости используемых методов, но и выделить благоприятные и неблагоприятные для прогнозирования статистико-вероятностные методы, географические районы, временные интервалы и сами прогнозируемые характеристики. Фактически стохастическое моделирование позволяет обобщить, классифицировать и заложить основы для физического диагноза и прогноза гидрометеорологических процессов. Предсказуемость 2-го рода — это возможные изменения прогнозируемой характеристики, зависящей от изменений предикто-

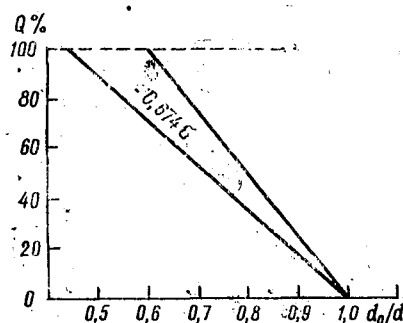


Рис. 4.2. Теоретическая зависимость оправдываемости прогнозов Q от относительной дисперсии ошибок прогнозирования d_0/d .

ров. Для определения предсказуемости 2-го рода необходимо построить адекватную физическую модель исследуемого процесса. Затем, варьируя параметры модели, определить границы изменения предиктанта. Понятно, что и для определения предела статистической предсказуемости 1-го рода нужно хотя бы в первом приближении построить модель прогнозируемого процесса. Таким образом, очевидна связь между предсказуемостью 1-го и 2-го рода, которая определяется типом модели.

В [44] показано, что описать различные океанологические процессы какой-либо одной моделью не удастся. То же относится и к описанию поведения каждого отдельного процесса во времени: на различных участках временного интервала работают различные статистико-вероятностные модели. Поэтому вывод об адекватности моделей АР-1 гидрометеорологическим процессам и их относительно небольшом пределе предсказуемости следует считать лишь первым приближением к реальности.

Идея комплексации элементарных прогностических моделей и выработка оптимального прогноза имеют свои преимущества и с точки зрения анализа статистической предсказуемости. Предсказуемость процесса в рамках комплексной модели, как правило, выше средней предсказуемости исходных элементарных прогностических моделей. Это становится очевидным, если учесть теоретически доказанное условие качества комплексной регрессионной модели [13, 14, 36, 44, 65, 153]: комплексный прогноз всегда не хуже лучшего прогноза, включенного в комплексацию. В тер-

минах предсказуемости первого рода это утверждение соответствует условию минимизации дисперсии ошибок прогноза и увеличению корреляции фактических и прогнозируемых значений комплексной модели, что приводит к росту максимального предела предсказуемости τ .

Предсказуемость многомерного процесса

Сохраняя основные понятия статистической предсказуемости при переходе к анализу многомерных процессов, необходимо сделать ряд уточнений. Предсказуемость многомерных процессов будет рассматриваться с позиций оценки предсказуемости обобщенных статистических параметров, характеризующих конкретный тип диагностической модели.

Предсказуемость структуры многомерного процесса можно оценить на основе анализа прогностических свойств модели ортогонального разложения вида

$$\hat{\mathbf{X}} = \hat{\mathbf{F}} \cdot \mathbf{A}^T + \hat{\mathbf{E}}.$$

Здесь вектор значений взаимонезависимых компонент (факторов) является прогнозируемой характеристикой, которая в обобщенном виде передает важнейшие черты структуры исследуемого процесса $\mathbf{X}$. Как было показано еще в [71, 91, 96, 115, 154], первые члены разложения $(\mathbf{f})$, дисперсия которых как минимум вдвое превышает дисперсию возможных ошибок наблюдения, описывают наиболее долговременные изменения структуры исследуемых процессов. В терминах автокорреляции это свойство эквивалентно увеличению радиуса корреляции и повышению устойчивости автокорреляционных функций для первых членов разложения по сравнению со средними характеристиками автокорреляционных функций исходных переменных.

Степень повышения предсказуемости пока не выражалась теоретически, хотя очевидна ее зависимость от размерности исходного процесса и уровня сжатия информации на первых членах разложения.

В силу взаимонезависимости членов разложения предсказуемость каждого из них может оцениваться самостоятельно по приведенным выше формулам для d , Q , τ . Каждая главная компонента характеризует специфическую особенность структуры процесса. Поэтому, оценивая предсказуемость каждой компоненты, исследователь получает возможность детально проанализировать природу предсказуемости многомерного процесса в рамках данной модели. Суммарная предсказуемость многомерного процесса оценивается как сумма элементарных оценок для членов разложения:

$$D = \sum_{j=1}^k \alpha_j d_j,$$

где α_j — вес j -й компоненты, k — число главных компонент.

Применение методов МАК позволяет в результате диагноза построить модель эволюции исследуемого процесса в терминах смены классов состояния. Таким образом многомерная последовательность векторов наблюдений X преобразуется в одномерную последовательность классов состояния (ω), оценка предсказуемости которой может проводиться на базе классической теории предсказуемости. Для этого необходимо получить количественную оценку автокоррелированности процесса, которой может быть модифицированный ранговый коэффициент корреляции Спирмена:

$$r(1) = 1 - 6 \sum_{i=1}^{N-1} (\omega(i) - \omega(i+1))^2 / (N^2 - 1) N,$$

где $r(1)$ — коэффициент ранговой автокорреляции при сдвиге $\tau = 1$; $\omega(i)$ — номер класса состояния в момент времени i ; N — длина выборки. Наличие такой оценки в принципе позволяет дать характеристику предсказуемости 1-го и 2-го рода на базе вычисления критериев τ и d .

Однако при наличии выраженной многомодальности процесса классический подход оказывается малоэффективным и возникает необходимость обратиться к понятию квантовой предсказуемости [65, 153], содержание которого сводится к вероятностному анализу устойчивости сохранения отдельных состояний и их взаимных переходов. Выборочные оценки устойчивости для различных фаз состояния могут сильно различаться между собой, что делает неинформативной среднюю оценку предсказуемости процесса и является закономерным отображением потери смысла в том случае, если число возможных состояний процесса больше двух.

Оценивание предсказуемости при усвоении информации

Оценивание предсказуемости становится весьма актуальным вопросом при построении адаптивных моделей процесса, учитывающих обратную связь модель — информация. Рассмотрим два варианта моделей, как и в п. 4.1, 4.2.

Изменение предсказуемости модели ортогональных разложений может быть однозначно связано с динамикой коэффициентов автокорреляции в процессе усвоения новой информации. Автокоррелированность главных компонент (факторов) может подвергаться выборочной изменчивости, что будет приводить к колебаниям предсказуемости. Предсказуемость первого рода (τ_i) перестает быть константой и на момент времени i может оцениваться по формуле

$$\tau_i = \frac{\ln(1 - d_i)}{2 \ln r_i},$$

где дисперсия ошибок (d_i) и корреляция фактических и прогнозных оценок (r_i) вычисляются по всей выборке наблюдений или по ее репрезентативной части, включающей «независимые» прогнозы, непосредственно предшествующие данному моменту.

Предсказуемость 2-го рода тогда будет оцениваться по вариативности главных компонент, значимость выборочных колебаний автокоррелированности которых будет мерой чувствительности прогностической модели к изменениям предикторов.

Усвоение новой информации в моделях «классов состояния» в простейшем случае будет приводить к варьированию устойчивости появления и взаимных переходов выделенных классов. Соответствующие колебания переходных вероятностей будут прослежены при условии использования байесовых процедур оценки (см. п. 4.2), которые дадут представление об изменении локальной предсказуемости отдельных структур процесса. Что касается изменения общей предсказуемости, то колебания предсказуемости 1-го рода можно оценить с помощью коэффициента ранговой корреляции Спирмена.

Глава 5

РЕЗУЛЬТАТЫ СТАТИСТИЧЕСКОГО ДИАГНОЗА ОКЕАНОЛОГИЧЕСКИХ ПРОЦЕССОВ

В предшествующих главах были рассмотрены некоторые частные вопросы прикладных океанологических исследований, решаемые с привлечением какого-либо единственного метода многомерной статистики. Обращение к более общим задачам диагноза океанологических процессов предопределяет, как было показано в главе 4, необходимый переход к последовательному применению нескольких методов МСА. Ниже приводятся примеры, которые наглядно демонстрируют реальные и потенциальные возможности применения рассматриваемой методологии диагноза в трех типовых задачах практической океанологии: 1) описание пространственно-временной изменчивости температуры поверхности океана; 2) описание трехмерной структуры вод по совокупности гидрофизических характеристик; 3) описание водных масс.

Изложение результатов расчетов проводится поэтапно с обязательной оценкой достоверности полученных статистик и их физической интерпретацией, что позволяет показать методологию статистического диагноза в действии. Географическими объектами исследований стали регионы Северной Атлантики, Норвежского моря и моря Беллинсгаузена, для которых решение поставленных задач представляет самостоятельный прикладной интерес.

Специфика океанологических условий в выбранных географических регионах позволила в рамках решаемых задач дополнительно акцентировать некоторые методические вопросы, среди которых — выбор и оценка информативных переменных, взаимосвязи общих и частных (локальных по пространству и времени) диагностических моделей, статистическая идентификация энергоактивных зон и фронтов.

5.1. Диагноз тепловых полей Северной Атлантики

Задача диагноза тепловых полей верхнего слоя океана традиционна для исследований крупномасштабных океанологических процессов. Характеристиками теплового режима при таких исследованиях выступают, как правило, температура поверхности океана или энтальпия (теплосодержание) вод верхнего слоя, которые рассматриваются как важнейшие параметры энергетики самого океана и интенсивности его термодинамического взаимодействия с атмосферой [100, 197, 266]. Решение задачи осуществ-

ляется чаще всего на базе физико-статистического анализа крупных массивов натурных наблюдений, когда выбор методов и последовательности их применения предопределяет физическое содержание результата.

Главная цель проводимого диагноза тепловых полей Северной Атлантики состоит в том, чтобы в результате всесторонней обработки имеющихся массивов наблюдений получить достоверные количественные характеристики пространственных и временных

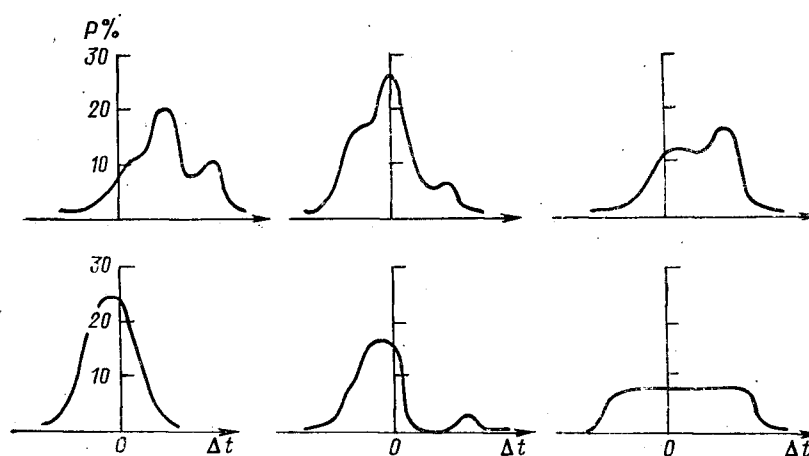


Рис. 5.1. Характерные эмпирические гистограммы рядов аномалий среднемесячной ТПО в Северной Атлантике.

особенностей формирования теплового режима акватории. При этом выделяются две взаимосвязанные задачи статистического описания особенностей климатического сезонного хода и межгодовых вариаций тепловых полей.

В качестве исходных данных использован составленный в ЛГМИ на базе материалов Гидрометцентра СССР архив полей среднемесячной температуры поверхности океана (ТПО) за период 1978—1987 гг. Исследуемая акватория простирается от экватора до 70° с. ш. Распределение температуры описывается временными рядами в 162 узлах пятиградусной сеточной области.

Первичный статистический анализ продемонстрировал наличие временных рядов с принципиально различными характеристиками изменчивости. Неоднородность прослеживается: по средним значениям, по дисперсии рядов, по характеристикам эмпирических функций распределения.

На рис. 5.1 продемонстрированы наиболее типичные эмпирические гистограммы, выявляющие скрытые в рядах ТПО и аномалий среднемесячной ТПО структурные различия. Очевидно, что, наряду с весьма близкими к нормальному закону, встречаются функции распределения существенно асимметричные, с «тяжелыми хвостами», а нередко и многомодальные.

Так, для районов климатических фронтальных зон в океане, особенно для полярной и субполярной, отчетливо выявляется многомодальность гистограмм. Это может быть интерпретировано как результат особой изменчивости ТПО вследствие интенсивных пространственных смещений фронтальных зон и появления в данном пятиградусном квадрате водной массы с пониженной или повышенной относительно вычисленного среднего значения температурой.

В экваториальной зоне Атлантического океана прослеживается обратная тенденция, когда гистограммы имеют слабый эксцесс и выделить характерные моды надежно практически не представляется возможным. В прибрежных зонах океана, где отмечаются кратковременные нарушения фонового речного стока, выявлены гистограммы с резко асимметричным распределением, которое в некоторых районах имеет тяжелые хвосты.

Все перечисленные особенности не только указывают на сложность исследуемых процессов формирования ТПО, но и демонстрируют определенную закономерность пространственного распределения параметров вероятностной структуры.

Гипотеза нормальности эмпирических функций распределения выполняется для 39 % исходных рядов с вероятностью 90 %. Еще для 19 % рядов выполняется гипотеза соответствия эмпирического закона распределению Вейбулла. Вероятностная структура остальных рядов может быть описана либо перечисленными теоретическими законами, но с низкой степенью достоверности, либо с помощью склейки различных распределений.

Оценка статистической структуры временных рядов методами корреляционно-спектрального анализа позволила охарактеризовать принципиальные статистические различия временной изменчивости ТПО по акватории Северной Атлантики. Устойчиво выделяются три диапазона изменчивости ТПО с характерными периодами: 6—9 мес, 13—16 мес, 2—3 года — и, естественно, 12-месячная компонента годового хода ТПО. Полученные диапазоны прослеживаются в изменчивости тепловых условий фактически для всей без исключения акватории, и, следовательно, их неоднородность является наиболее важным в данном случае элементом описания структуры теплового режима.

Более долгопериодная изменчивость, к сожалению, не может надежно фиксироваться используемыми статистическими методами на столь коротких рядах. На спектральных функциях она проявляется в виде трендовой компоненты, которая придает большинству полученных графиков характерный наклон спектров «красного шума».

Полученные результаты спектрального анализа в целом совпадают с количественными оценками более ранних работ по Северной Атлантике, где использовались материалы наблюдений за 1957—1974 гг. [21, 31, 62, 119, 141]. Определенные отличия наших оценок вполне объяснимы выборочной изменчивостью. Однако выявленные различия не выходят за рамки 10 %-ного уровня зна-

чимости, что позволяет говорить о сохранении в 1978—1987 гг. статистической структуры основных колебаний ТПО практически неизменной для Северной Атлантики в целом.

Применение метода главных компонент (МГК) нацелено на описание наиболее общих для всей акватории Северной Атлантики сопряженных колебаний ТПО. Каждая главная компонента по идее должна вобрать характерные особенности пространственных колебаний ТПО. Тогда собственные вектора будут описывать закономерности временной изменчивости наиболее ярких сопряженных колебаний. Разложение выполнено для исходной матрицы данных, составленной из 162 рядов аномалий среднемесячной ТПО длиной по 120 значений.

Сходимость ряда собственных чисел корреляционной матрицы достаточно высока — 90 %-ный уровень описания изменчивости полей ТПО достигается при учете примерно 40 членов разложения. Наиболее существенный вклад вносят первые четыре члена разложения, которые в сумме описывают около 50 % дисперсии наблюдений. Следом по значимости идут еще три члена разложения, описывающих 4—7 % дисперсии. Каждый из оставшихся членов разложения добавляет менее 1 % в описание ТПО. В скорости сходимости собственных чисел наблюдается заметная неоднородность, которая на графике выявляется как своеобразная ступенчатая структура. Следуя теории МГК, нетрудно понять, что каждая ступенька отражает некий характерный масштаб изменчивости ТПО, пространственные и временные характеристики которого необходимо искать в особенностях изменения собственных векторов и компонент. Графики собственных векторов в данном случае анализировать визуально малоэффективно из-за их значительной длины. Поэтому потребовалось обратиться к методикам анализа временных рядов и оценить изменчивость собственных векторов на основе вычисленных автокорреляционных и спектральных функций.

Полученные статистические характеристики убеждают, что первые четыре собственных вектора имеют выраженную периодическую изменчивость на фоне долгопериодной трендовой компоненты. Характерные периодические колебания имеют периоды 2—3 и 5—7 лет (точнее оценить невозможно из-за ограниченности длины выборки). Следующие три вектора в разложении имеют периодические колебания с периодами 6—9 и 14—18 мес. В оставшихся членах разложения столь четкой систематической составляющей выявить не удается.

Как правило, векторы с более высокими номерами не имеют значимых пиков периодических составляющих в межгодовом интервале изменчивости. Статистическая структура этих колебаний близка к спектру красного шума. Одновременно следует отметить существенный тренд, который фиксируется в спектре первого собственного вектора и может быть связан с общим для всей исследуемой акватории Северной Атлантики долгопериодным изменением фоновых температурных условий. Последняя особенность

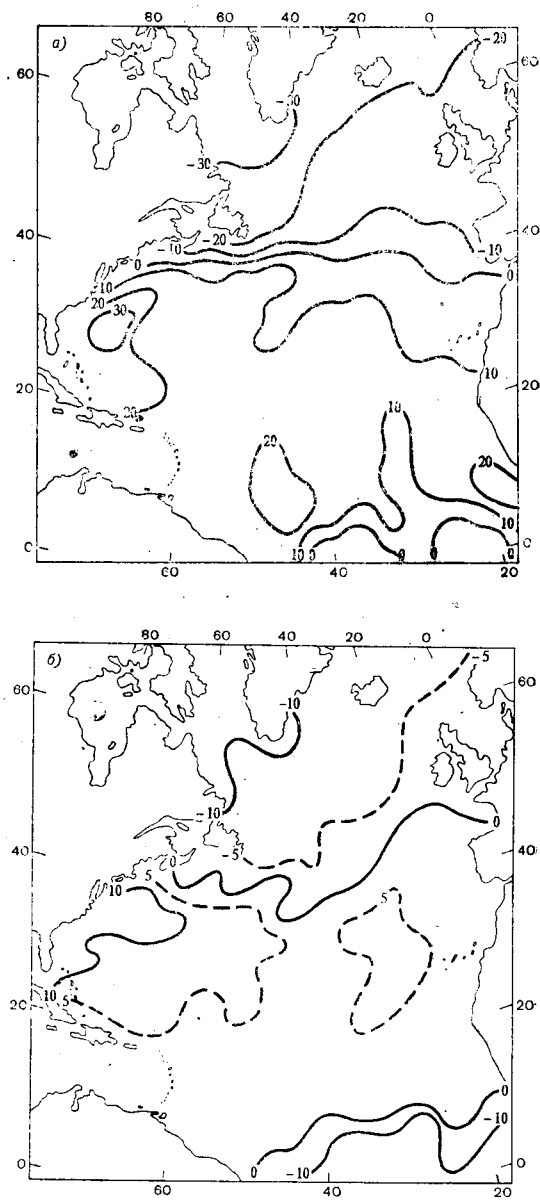


Рис. 5.2. Поля первой (а) и второй (б) главных компонент разложения годового хода ТПО в Северной Атлантике.

принципиально важна для анализа в дальнейшем возможностей эффективного описания изменения ТПО в Атлантике как целостного механизма. Разложение по главным компонентам показывает, что сопряженные и общие для всего океана колебания температуры поверхности океана (рис. 5.2) составляют 20 % от суммарной изменчивости.

Насколько реально отражает полученное разложение действительные процессы формирования теплового режима верхнего слоя Атлантического океана? Общеизвестные недостатки наблюдения температуры воды поверхности в натурных условиях и построения среднемесячных полей ТПО определенным образом снижают доверие к исходному материалу анализа и, следовательно, к достоверности полученных результатов разложения.

Ответить на поставленный вопрос можно, если попытаться сравнить статистическую структуру разложений полей ТПО с разложением более обобщенной и надежной характеристики верхнего слоя океана — энтальпии верхнего слоя вод. Энтальпия как один из основных термодинамических параметров состояния океана связывает и обобщает изменчивость термохалинных условий в заданном объеме морской воды [31, 87, 197]. В данном случае удобно использовать для расчетов статистической структуры поля средних многолетних среднемесячных значений энтальпии верхнего 100-метрового слоя вод в Северной Атлантике в принятой 5-градусной сеточной области.

Результаты компонентного разложения представлены на рис. 5.3 в виде трех полей первых главных компонент, описывающих в сумме 74 % исходной дисперсии [51, 52]. На долю первой компоненты приходится около 40 % общей дисперсии годового хода, и, как следует из приводимых карт, она описывает фоновое изменение тепловых условий в исследуемой акватории. Компонента сохраняет знак на всей акватории Северной Атлантики за исключением локальных областей в Лабрадорском море и у побережья Западной Африки. В поле второй компоненты прослеживается уже ряд противоположных по знаку очагов, доля изменчивости в которых составляет 23 % и связывается прежде всего с фазовыми различиями пространственного распределения годовой гармоники энтальпии.

Пространственная структура второй компоненты поля энтальпии характеризуется своеобразием попарного расположения замкнутых областей перемены знака. Наиболее отчетливо прослеживаются зоны перемены знака, разделяющие юго-западную и северо-восточную области Северной Атлантики. Пространственный масштаб областей сохранения знака второй компоненты составляет порядка 2000 км.

Третья компонента выявляет в пространстве сложную ячеистую структуру полей энтальпии, которая описывает около 11 % дисперсии суммарных колебаний. Одновременно в распределении этой компоненты прослеживается характерная зона контраста, которая пересекает акваторию с юго-запада на северо-восток и окон-

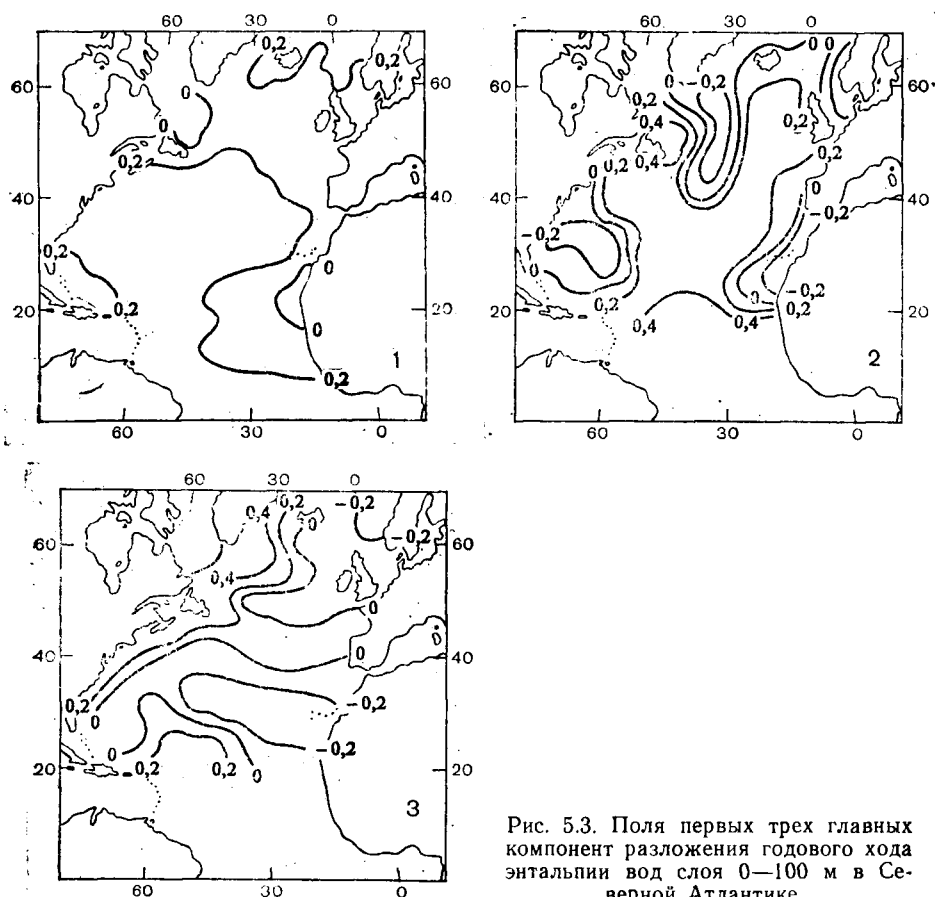


Рис. 5.3. Поля первых трех главных компонент разложения годового хода энтальпии вод слоя 0—100 м в Северной Атлантике.

турирует практически всю зону распространения вод Северо-Атлантического течения от 30 до 60° с. ш.

Первый вектор разложения полей энтальпии не меняет знака во времени и описывает средний годовой фон; второй вектор меняет знак один раз с минимальными значениями в марте и максимальными в августе; третий вектор описывает полугодовую волну в поле энтальпии и меняет знак дважды.

Сравнение полученных разложений полей энтальпии верхнего слоя и ТПО показывает, что структуры пространственного и временного распределения этих параметров теплового режима океана имеют ряд общих черт — в скорости сходимости разложений, закономерностях сезонного хода, особенностях пространственного положения основных очагов синхронных колебаний ТПО и энтальпии.

Взаимное соответствие полученных оценок позволяют утверждать, что в результате компонентного анализа полей ТПО выде-

лены наиболее существенные особенности изменчивости термического режима верхнего слоя воды в Северной Атлантике.

Цель дальнейших диагностических расчетов состоит непосредственно в том, чтобы статистически достоверно выделить в Северной Атлантике районы, однородные по изменчивости тепловых условий верхнего слоя океана. Применяется одна методика обработки данных — автоматическая классификация. Но при этом поочередно используются три массива исходных данных. Каждый массив по-своему характеризует тепловой режим. Сопоставление результатов трех классификаций дает реальную возможность районировать акваторию с высокой надежностью и наглядно интерпретировать смысл полученного разбиения.

Первоначально автоматической классификации подвергались материалы по дисперсии температуры воды каждого из исследуемых пятиградусных квадратов. Классификационная матрица составлена из 12-месячных значений дисперсии ТПО для каждого временного ряда. В результате получено устойчивое разбиение, содержащее три класса изменчивости ТПО: класс с максимальной дисперсией ТПО ($\sigma > 1^\circ\text{C}$), класс с минимальной дисперсией ТПО ($\sigma < 0,7^\circ\text{C}$) и промежуточный класс ($0,7-1,0^\circ\text{C}$).

Выбор границ классов обусловлен естественной группировкой изменчивости тепловых условий на гистограммах. Районирование в пространстве акватории Северной Атлантики позволило получить 12 однородных районов, дисперсия ТПО внутри которых статистически однородна и соответствует принятым критериям границ классов (рис. 5.4).

Полученная классификация носит предварительный характер, но уже выявляет в исследуемой акватории ряд чередующихся в пространстве областей, где изменчивость ТПО весьма значительна. Это относится к району Норвежского моря и северо-западной части океана, где действуют наиболее сильные течения и наиболее выражены фронтальные разделы между различными водными массами.

Продолжая диагностические расчеты, необходимо провести классификацию других характеристик тепловых условий, и прежде всего результатов разложения полей ТПО и энтальпии по главным компонентам. Подобная классификация позволяет получить несколько большее число классов, которые в данном случае выделяются уже по принципу степени сопряженности колебаний ТПО, что обусловлено сложной группировкой дисперсий колебаний ТПО и энтальпии по компонентам. Районирование полученных классов в исследуемой акватории показано на рис. 5.4.

Сравнивая три классификационные картины, нетрудно увидеть качественное сходство полученных разбиений в пространственных координатах, хотя средние параметры классов принципиально различны. Так, первое полученное районирование предопределено сходством дисперсии колебаний ТПО, второе — сопряженностью колебаний ТПО по фазе, третье — сопряженностью колебаний термохалинных процессов в верхнем слое.

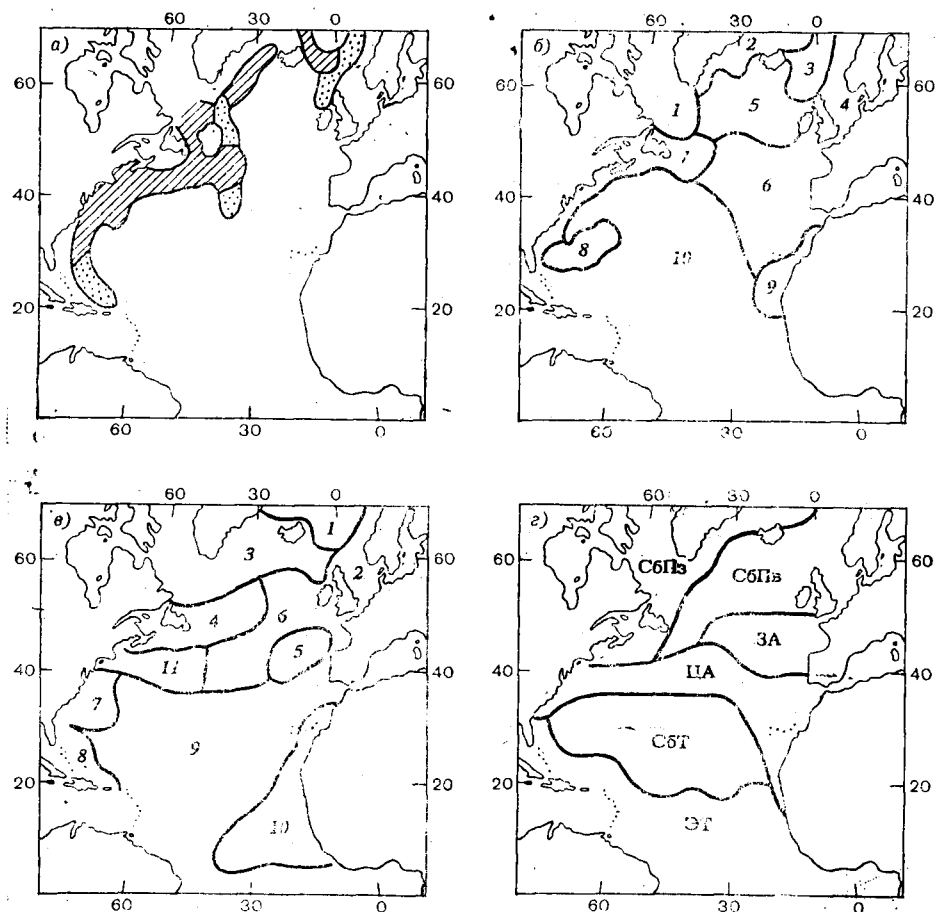


Рис. 5.4. Районирование Северной Атлантики.

а) по дисперсии годового хода ТПО, б) по особенностям годового хода энтальпии, в) по сопряженности изменчивости ТПО, г) по распределению поверхностных водных масс.

Заметны и некоторые различия полученных классификаций.

Они сосредоточены в центральной и южной частях исследуемой акватории Северной Атлантики, где в умеренной и субтропической климатических зонах относительно ослаблен пространственный контраст термических полей. Не удастся по аномальности дисперсий ТПО выделить характерный район, прилегающий к побережью Западной Африки. Аномальность теплового режима здесь прослеживается прежде всего по фазовому сдвигу полугодовой волны энтальпии и характерному 2—3-летнему биению в рядах аномалий ТПО.

Поле энтальпии излишне сглаживает общую картину теплового режима в умеренных широтах (40—55° с. ш.), где практи-

чески исчезает характерный очаг 9—14-месячной ритмики колебаний ТПО.

Общность полученных классификаций проявляется в проведении границ классов на севере и северо-западе Северной Атлантики — прежде всего в области потока Северо-Атлантического течения, в частности на участке от п-ва Флорида до мыса Гаттерас. Здесь обособляется собственно Гольфстрим и его южная ветвь, которые классифицируются и по дисперсии, и по энтальпии, и по коррелированности временного хода аномалий ТПО.

Следующая общая область расположена к востоку от о. Ньюфаундленд, где наблюдаются интенсивные флуктуации параметров теплового режима в зоне взаимодействия Северо-Атлантического и Лабрадорского течений. Здесь наблюдается значительное увеличение дисперсии ТПО, которое связывается с наличием многочисленных больших аномалий среднемесячных значений ТПО от многолетних норм. Эти аномалии столь велики, что специфичность теплового режима в данном районе отчетливо прослеживается в годовом ходе энтальпии верхнего слоя вод.

Наконец, определенная регулярность появления значительных аномалий ТПО с интервалом 2—5 лет предопределяет выявление особого класса в данном районе в результате статистической обработки временных рядов ТПО.

Район Норвежского моря и непосредственно прилегающей к нему части акватории Северной Атлантики также представляет собой зону, в которой полученные классификации имеют ряд общих черт. В частности, в структуре полей ТПО отчетливо прослеживается деление района Норвежского моря на области влияния атлантических и полярных климатических зон, которые определяют наличие однородных районов на юго-востоке и северо-западе моря. По результатам классификации дисперсии ТПО и годового хода энтальпии можно проследить положение границ климатических структурных зон, для которых характерен повышенный уровень изменчивости теплового режима верхнего слоя вод.

Кроме того, пространственное расположение границ выделяемых классов таково, что Норвежская ЭАЗО как бы обособлена от всей акватории Северной Атлантики.

На чем же основано выделение Норвежского моря в отдельный класс изменчивости тепловых условий практически по всем исходным характеристикам? Для того чтобы ответить на этот вопрос, необходимо вернуться на предшествующий этап диагноза и попытаться дать общую количественную оценку сходства и различия формирования фоновых тепловых условий в Норвежском море относительно других районов Атлантики. В этом смысле особый интерес представляет возможность выявить различия теплового режима в основных энергоактивных зонах Северной Атлантики. Для решения подобной задачи весьма эффективен метод канонического корреляционного анализа.

Блочная матрица исходных данных для оценки канонических переменных составлена из трех блоков, куда вошли временные

ряды ТПО, принадлежащие трем районам энергоактивных зон. В каждом блоке находятся ряды, описывающие соответственно изменчивость тепловых условий в Норвежской, Ньюфаундлендской ЭАЗО и ЭАЗО Гольфстрим.

Результаты расчетов показывают (табл. 5.1), что для характеристики фоновых тепловых условий в ЭАЗО необходимо использовать не менее двух канонических переменных, вклад которых существенно различен.

Таблица 5.1
Эффективность канонического разложения
полей ТПО — вклад (%) в общую
дисперсию канонической переменной

ЭАЗО	Переменная	
	1	2
Норвежская	59	19
Ньюфаундлендская	94	2
Гольфстрим	94	3

Характерно, что высокая сопряженность колебаний ТПО в ЭАЗО Гольфстрим и Ньюфаундлендской приводит к сосредоточению изменчивости уже в первом члене разложения, который характеризует фоновый годовой ход тепловых условий. Наоборот, в Норвежской ЭАЗО подобной точности описания ТПО не удастся достигнуть даже при учете двух канонических переменных, которым в сумме удастся описать лишь 78 % дисперсии ТПО. О возможном физическом механизме, порождающем столь значимую разнородность полей ТПО, можно судить по результатам анализа пространственного распределения соответствующих первых двух векторов блочной корреляционной матрицы данных (рис. 5.5).

Как видно из приводимого рисунка, первый вектор характеризует фоновое изменение ТПО в целом во всех исследуемых ЭАЗО Северной Атлантики. Сохранение знака первого собственного вектора внутри ЭАЗО Гольфстрим и Ньюфаундлендской показывает высокую синхронность сезонного хода тепловых условий в этих ЭАЗО, тогда как в Норвежской ЭАЗО отмечается повышенная примерно в 10 раз по сравнению с другими ЭАЗО пространственная изменчивость этого вектора, что указывает на определенную асинхронность колебаний ТПО в этом районе.

Данная особенность становится вполне очевидной из структуры второго собственного вектора. Во всех исследуемых ЭАЗО второй вектор меняет знак, причем нулевая линия располагается вблизи положения климатических фронтальных зон фронта Гольфстрима, субполярного и полярного фронта. Учитывая, что вторая каноническая переменная во времени имеет хорошо выраженную периодичность колебаний в 6 мес и 2—3 года, можно прийти

к выводу о высокой значимости именно этих колебаний в формировании фонового распределения ТПО в Норвежской ЭАЗО. Подобная значимость определяется не только высокой относительной амплитудой подобных колебаний в Норвежском ЭАЗО, но и явным сдвигом их фаз при переходе через полярный фронт в Норвежском море. Последняя особенность следует из очевидной про-

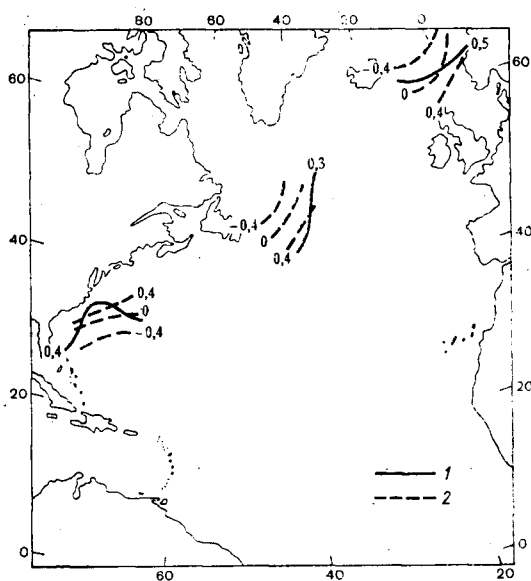


Рис. 5.5. Поля векторов канонического разложения полей ТПО в основных энергоактивных зонах Северной Атлантики.

1 — изолинии первого вектора, 2 — изолинии второго вектора.

странственной оппозиции значений второго вектора в Норвежской ЭАЗО.

Выявленные в ходе многомерного анализа особенности изменения полей температуры заставляют обратить особое внимание на обособленность динамики теплового режима, которая наблюдается в наиболее мощных замкнутых очагах распределения главных компонент — в Норвежской ЭАЗО.

Для того чтобы точнее охарактеризовать особенности формирования теплового режима этого района, необходимо изменить границы области, для которой выполняется статистический анализ. Уменьшение площади исследуемого района дает возможность сосредоточиться главным образом на региональных особенностях формирования ТПО.

Район Норвежского моря, входящий в границы Норвежской ЭАЗО, привлекает особое внимание тем, что здесь на относительно небольшой площади происходит интенсивное взаимодействие вод

арктического и атлантического происхождения. Температура воды остается важнейшим показателем взаимодействия генетически различных водных масс в верхнем слое Норвежской ЭАЗО, поскольку контраст тепловых характеристик определяет термодинамику процесса в целом по акватории.

Цель дальнейшего статистического анализа состоит в том, чтобы получить более надежные оценки пространственно-времен-

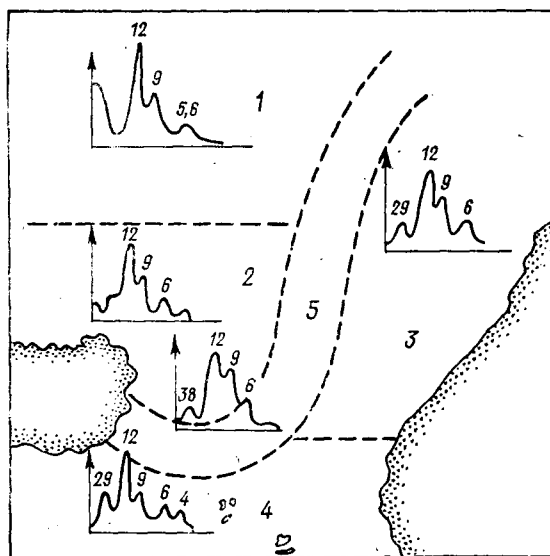


Рис. 5.6. Районирование Норвежской энергоактивной зоны по результатам классификации полей аномалий среднемесячной ТПО.

Границы районов показаны штриховыми линиями, для каждого района приведена характерная спектральная функция ряда ТПО.

ной изменчивости температуры воды в Норвежской ЭАЗО и тем самым уточнить уже имеющиеся результаты. Для этого исходные данные по среднемесячной ТПО были представлены с меньшей дискретностью: каждое поле характеризовалось наблюдениями в узлах сетки с шагом $2,5^\circ$ по широте и долготе. Процесс решения диагностической задачи практически не отличался от рассмотренного выше алгоритма диагноза в Атлантике. Поэтому, учитывая выводы первичного анализа, перейдем непосредственно к обсуждению результатов классификации исходных данных в призначном пространстве общих факторов.

Устойчиво классифицируются пять классов, каждый из которых связан с максимальным влиянием одного из выделенных статистически независимых общих факторов. Районирование по результатам классификации практически мало меняет число однородных районов — их всего семь. Расположение районов показано на рис. 5.6.

В качестве важной характеристики пространственной структуры теплового режима Норвежской ЭАЗО выступает класс 5, который как бы разделяет исследуемую акваторию моря на две части. Такое деление объясняется присутствием полярного фронтального раздела, который разграничивает области влияния субарктической и умеренной климатических зон океана. Следует подчеркнуть естественность подобной классификации ввиду наличия в этом районе очень своеобразных многомодальных эмпирических гистограмм рядов аномалий температуры поверхности океана (см. рис. 5.1).

Как показывает анализ структуры периодических колебаний ТПО, здесь наблюдается увеличение амплитуд колебаний с периодами 2—3 года, 9—11 и 13—14 мес. Кроме того, именно в этой узкой полосе происходит скачкообразное замедление фазы распространения полугодовой гармоник в ходе поверхностной температуры воды. Таким образом, подтверждаются предварительные выводы о термическом режиме Норвежского моря, полученные в результате анализа фоновых тепловых условий на всей акватории Северной Атлантики в целом.

Следующая задача, решаемая на основе применения методов МСА, состоит в описании межгодовых модуляций сезонного хода температуры воды. Статистическая оценка межгодовой изменчивости параметров сезонного хода проводилась путем сравнения разложения коротких выборок среднемесячных аномалий ТПО для каждого года длиной по 12 значений. Исходные матрицы имели размерности 162×12 .

Заметным своеобразием отличается распределение дисперсий исследуемого процесса по отдельным членам разложения. Первые три компонента, характеризующие наиболее крупномасштабные колебания ТПО, описывают от 40 до 60 % суммарной дисперсии. На последующих компонентах прирост описываемой дисперсии резко замедляется, вклад каждого из новых членов разложения практически никогда не превышает 8 %. Уровень 90 %-ной точности описания исследуемой изменчивости достигается только при учете 9—10 членов разложения, что, скорее всего, объясняется возросшей ролью локальных процессов формирования тепловых условий.

Особенности разложений определяются степенью межмесячной корреляции полей аномалий ТПО для конкретного года. В исследуемой выборке отчетливо прослеживаются два типа разложений, которые представлены на рис. 5.7 картами главных компонент полей ТПО 1980 и 1985 гг.

Структура полей главных компонент, аналогичная полученной по данным за 1985 г., встречается весьма часто в исследуемой выборке. Для нее типична общая зональность изолиний пространственного распределения первой компоненты с отчетливо выраженными очагами экстремальных значений в Норвежском море, ЭАЗО Гольфстрим, севернее о. Ньюфаундленд, у побережья Западной Сахары. Временной ход соответствующего ей собственного

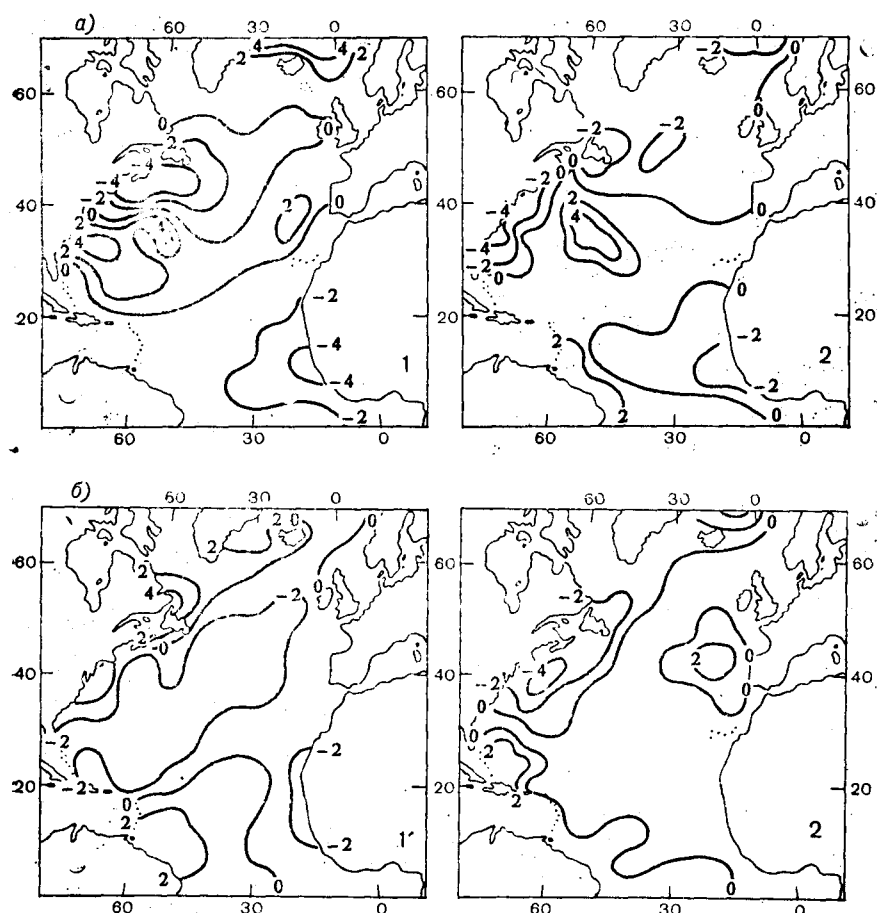


Рис. 5.7. Поля собственных векторов разложения аномалий среднемесячной ТПО в Северной Атлантике в отдельные годы.

а — 1985 г., б — 1980 г.

вектора характеризуется практически полным отсутствием какой-либо закономерной изменчивости в течение всего года.

Вторая компонента описывает изменчивость ТПО главным образом в районе Северо-Атлантического течения на участке от Флоридского пролива до мыса Гаттерас, где выявляются интенсивные очаги ее значений разных знаков.

Кроме того, в поле компоненты выявляется характерная аномалия у побережья Западной Африки. Временной ход соответствующего собственного вектора имеет характерную полугодовую изменчивость.

Компоненты более высоких порядков демонстрируют пространственное распределение с выраженной очаговой структурой и

многократным изменением знака компоненты. В целом основные очаги распределения третьей и четвертой компонент совпадают с уже описанными выше особенностями. Однако особо следует подчеркнуть то, что поле всех четырех главных компонент отчетливо выделяет Норвежское море несколькими замкнутыми изолиниями, а знак этого очага, как правило, не совпадает со знаком поля в центральной Атлантике.

Иная достаточно типичная для исследуемой выборки структура ортогональных разложений наблюдается в 1980 г. (см. рис. 5.7). Здесь наиболее важной чертой следует считать общую меридиональную ориентацию полей первых главных компонент, причем наблюдается сохранение знака первой и второй компоненты в Норвежской зоне и центральной части Северной Атлантики.

Отрицательные значения первой компоненты сосредоточены в Центральной Атлантике, в ЭАЗО Гольфстрим и центральной части Норвежского моря, а также у побережья Западной Африки. Размер горизонтальных замкнутых очагов сохранения знака главных компонент заметно выше по сравнению со структурой полей в 1985 г. (это касается полей и первой и второй компонент) и указывает на общую генерализацию фоновых колебаний тепловых условий на акватории Северной Атлантики в целом.

Поскольку процесс диагноза был длительным, многошаговым и периодически требовал построения локальных моделей, авторы уже внутри процедур анализа пытались синтезировать частные результаты в региональных моделях. Теперь появилась возможность для обобщенной интерпретации всех полученных результатов и построения единой физико-статистической модели формирования тепловых условий в Северной Атлантике.

Физико-статистическая модель структуры теплового режима верхнего слоя вод в Северной Атлантике может здесь формулироваться в терминах сопряженных пространственно-временных колебаний температуры поверхности воды, поскольку проведенный анализ выявляет достаточное соответствие результатов диагностики ТПО и энтальпии верхнего 100-метрового слоя воды. Сопряженные колебания тепловых условий концентрируются в нескольких четко оконтуренных районах исследуемой акватории и фоновое распределение ТПО может быть восстановлено с помощью элементарной линейной суперпозиции выделенных типовых колебаний внутри однородных районов. Содержательная интерпретация полученных элементов модели может быть осуществлена двояко — в терминах водных масс и фронтов либо в терминах энергоактивных зон океана.

В терминах энергоактивных зон динамика изменчивости основных пространственных очагов возникновения аномалий ТПО может быть описана поведением ТПО внутри трех основных зон — Норвежской, Ньюфаундлендской и Гольфстрим — при одновременном учете теплового фона у побережья Западной Африки. Особенности временной изменчивости внутри перечисленных очагов до-

статочны сопряжены между собой и описываются уже первыми членами ортогонального разложения полей ТПО. Точность подобного описания будет находиться в пределах 60—70 %. Повышение эффективности модели может быть связано с более достоверным описанием теплового режима Норвежской ЭАЗО, который, как показывал анализ, отличается аномально высокой пространственной неоднородностью, обусловленной значительным возрастанием роли полугодовых и 2—3-летних квазициклических колебаний в зоне полярного гидрологического фронта в Норвежском море.

Учет перечисленных особенностей позволяет улучшить описание тепловой изменчивости Норвежской ЭАЗО в годовом ходе не менее чем на 20 % и повысить общий уровень качества модели всей акватории на 5—7 %. Дальнейшее повышение качества статистического описания фонового распределения температуры воды в Северной Атлантике возможно при условии восстановления колебаний теплового режима в локальных очагах, расположенных к западу от о. Ирландия, севернее о. Мадейра, вблизи Антильских островов, вдоль восточного побережья Гренландии.

Из-за малой пространственной протяженности перечисленных очагов воспроизводить их модельную изменчивость в заданном масштабе пространственно-временного осреднения не представляется целесообразным. Поэтому дальнейшее увеличение точности описания фоновых тепловых условий в рамках подобной модели может быть выполнено при условии повышения точности самих исходных данных о температуре поверхности океана, погрешности отдельных наблюдений которой могут достигать 1,5 °C [197].

В климатическом аспекте полученные результаты анализа могут удовлетворительно описываться в классической терминологии водных масс поверхностного слоя. Как показывает количественное сравнение совпадения границ классов, полученных в результате классификации полей и по данным T, S -анализа водных масс поверхностного слоя [100], эти результаты районирования практически аналогичны: коэффициент дихотомической корреляции составляет 0,68 и является значимым на принятом уровне доверительной вероятности. Различие методик классификации и районирования и различие исходных данных в данном случае не стали причиной разнородности результатов.

5.2. Диагноз структуры вод Норвежского моря

Исследование термохалинных и динамических процессов формирования режима верхнего слоя вод остается одной из основных проблем в тематике экспериментальных и теоретических работ по Норвежской ЭАЗО [5, 7, 32, 41, 88, 159, 170, 277]. Основными элементами крупномасштабной структуры вод общепризнаны водные массы и фронты (фронтальные зоны), сопряженные колебания которых представляются как обобщенные индикаторы изменений океанологических условий в целом на рассмат-

ваемой акватории. Видимо, еще преждевременно утверждать наличие полностью сложившейся и количественно подтвержденной единой картины формирования структуры вод в Норвежской ЭАЗО даже в климатическом масштабе осреднения. Причины существующих различий в описании и интерпретации механизмов формирования структуры следует искать в реальной сложности происходящих процессов, которые с недостаточной точностью идентифицируются существующими методиками анализа структуры вод. Еще не разработана универсальная методика выделения и описания фронтальных зон и водных масс как взаимосвязанных элементов структуры вод. Пока такое описание проводится раздельно на основании применения принципиально различающихся методов. Поэтому цель применения статистического диагноза состоит в том, чтобы в рамках единого методического подхода: а) получить количественное описание климатических фронтальных зон и водных масс как взаимосвязанных элементов крупномасштабной структуры вод, б) увязать их пространственную изменчивость с процессами перераспределения вещества и энергии в Норвежской ЭАЗО.

В качестве исходных данных для статистического диагноза использованы климатические среднегодовые поля температуры, солености, плотности, содержания кислорода на стандартных горизонтах в узлах сеточной области с дискретностью 1° по параллели и меридиану. Карты пространственного распределения исследуемых параметров выявили значительную горизонтальную изменчивость гидрофизических полей и наличие нескольких зон сгущения изолиний, положение которых не всегда совпадало в пространстве акватории.

Первым этапом МСА стал расчет и анализ двумерных функций вероятности полей температуры и солености. На рельефе функции $P(T, S)$ для слоя (0—200) м достаточно отчетливо (рис. 5.8) выделяются максимумы (вершины), которые соответствуют основным водным массам деятельного слоя Норвежской ЭАЗО. Однако разделение водных масс на рельефе $P(T, S)$ с целью оценки их объемов и расчета T, S -индексов затруднено вследствие замаскированности, размытости ложбин $P(T, S)$, которые принято считать границами водных масс [55]. Размытость ложбин сохраняется и при варьировании дискретности разбиения выборочной изменчивости переменных в процессе расчета. Поэтому можно полагать, что отмеченная особенность является специфической чертой исследуемых океанологических характеристик в Норвежском ЭАЗО. Анализ функций вероятности, построенных с привлечением полей плотности $P(S, \rho)$, $P(T, \rho)$ и кислорода $P(T, O_2)$, привел к аналогичным результатам.

Таким образом, проведенный объемный T, S -анализ позволяет заключить, что ввиду значимого развития фронтальных зон в Норвежской ЭАЗО возникает необходимость идентифицировать их в виде отдельных объемных элементов структуры вод региона, для чего исходный набор признаков должен быть дополнен ин-

формативными параметрами, характеризующими степень трансформации вод.

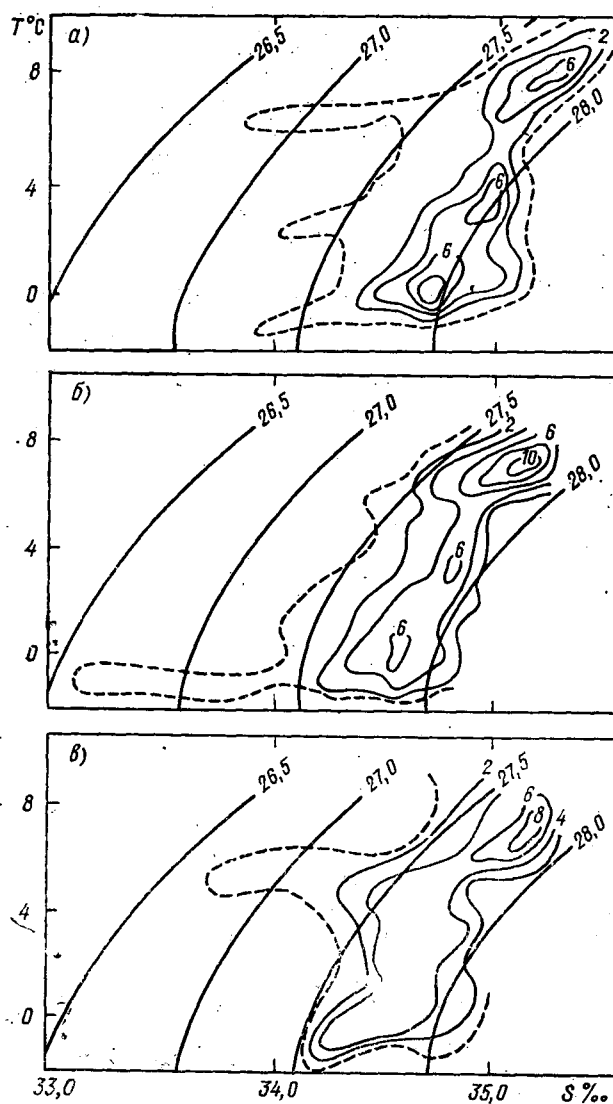


Рис. 5.8. Двумерные гистограммы температура—соленость в слое 0—200 м Норвежской ЭАЭО для среднегодовых (а) и среднесезонных (зима (б) и осень (в)) данных.

По оценкам К. Н. Федорова для фронтальных зон таковыми следует прежде всего считать модули горизонтальных градиентов основных океанологических характеристик [199]. Поэтому

матрица исходных данных для МСА формировалась следующим образом. Конкретный одноградусный квадрат рассматривался как объект исследования. Каждому объекту соответствовал вектор наблюдений, который составлялся из значений восьми признаков: температуры, солености, плотности, содержания кислорода и модулей горизонтальных градиентов этих параметров.

В результате факторного анализа среднегодовых полей пространственная изменчивость восьми исходных параметров была объективно сгруппирована в трех общих факторах при уровне значимости $\alpha = 10\%$. Структура взаимосвязей общих факторов и исследуемых переменных имеет ряд отличительных особенностей (табл. 5.2). Первый общий фактор наилучшим образом связан

Таблица 5.2

**Корреляция общих факторов и исходных признаков
(факторные нагрузки) в деятельном слое
Норвежской ЭАЗО по среднегодовым данным**

Признак	Фактор		
	1	2	3
T	0,9	0,0	-0,2
S	0,3	0,7	0,0
σ	-0,7	0,6	-0,1
O_2	-0,8	0,0	0,2
Γ_T	0,0	0,3	0,9
Γ_S	-0,2	0,1	0,6
Γ_σ	-0,1	0,2	0,9
Γ_{O_2}	-0,1	0,2	0,7
Вес факторов, %	43	29	17

с крупномасштабной изменчивостью поля температуры (коэффициент корреляции $r = 0,9$) и в меньшей степени с изменчивостью поля содержания растворенного кислорода ($r = -0,8$) и плотности ($r = -0,7$). Второй общий фактор характеризуется высокой корреляцией с изменчивостью солености ($r = 0,7$) и плотности ($r = 0,6$). Третий общий фактор формируется изменчивостью градиентных полей и прежде всего градиентами температуры и плотности ($r = 0,9$), а в несколько меньшей степени градиентами солености и содержания кислорода.

Факторная структура демонстрирует значительную несогласованность пространственного распределения основных гидрофизических полей в деятельном слое Норвежской ЭАЗО, что следует из разделения по различным статистически независимым факторам

параметров температуры и солености. Вместе с тем вскрывается взаимосвязанность распределений температуры и содержания кислорода, а также всех градиентных полей. Если взаимообусловленность температуры и содержания кислорода — достаточно очевидный факт, подтверждаемый наблюдениями и теорией [88, 100], то совпадение изменчивости градиентов полей представляется менее очевидным, поскольку рассогласованность исходных полей температуры и солености, казалось бы, соответствует несовпадению в пространстве зон интенсивной трансформации этих полей.

Данную особенность поясняет относительный вклад исходных параметров в суммарную дисперсию: изменчивость температуры и связанная с ней изменчивость кислорода и плотности оказываются наиболее крупным источником дисперсии в исследуемом районе — их доля составляет 43 %, что значительно больше, чем доля солености — 29 %. Вполне обоснованным поэтому будет предположение о наличии подобного соотношения и в третьем факторе, который наилучшим образом связан с трансформацией поля температуры и плотности. Отсюда следует вывод о преимуществе исследования состояния верхнего слоя Норвежской ЭАЗО в целом по полю температуры, чем по изменчивости других параметров.

Учитывая, что каждый фактор обобщает коррелированную изменчивость нескольких переменных, представляет интерес рассмотрения пространственного распределения выделенных факторов (рис. 5.9). Нетрудно видеть, что температурная изменчивость, описываемая первым фактором, действительно преобладает на акватории Норвежской ЭАЗО и особенно в центральной ее части. Преобладающая роль соленостной изменчивости отмечается в прибрежной зоне Скандинавии и на северо-западе исследуемого региона — в полярных водах, выносимых Восточно-Гренландским течением из Арктического бассейна — второй фактор. Пространственное распределение фактора № 3 характеризуется наибольшей неоднородностью, экстремумы локальны и приурочены в основном к району, протянувшемуся вдоль нулевого меридиана и к северу от Фарерско-Исландского порога.

Для объективного районирования акватории Норвежской ЭАЗО проведена автоматическая классификация с использованием в качестве признаков полученных общих факторов. Оптимальность разбиения достигнута при разделении акватории на шесть однородных районов, статистические параметры которых приведены в табл. 5.3, а пространственное расположение — на рис. 5.10. Анализ полученных количественных оценок позволяет идентифицировать районы следующим образом: 1 — район, занятый полярными водами; 2 — атлантическими водами; 3 — атлантическими трансформированными водами (водами Норвежского течения); 5 — смешанными атлантическими и полярными водами. Районы 4 и 6 характеризуются высокими значениями градиентов температуры, солености, плотности, что с учетом географического положения этих районов как цепочек, локализованных на границах

крупных классов позволяет интерпретировать районы 4, 6 как систему фронтальных зон Норвежского ЭАЗО.

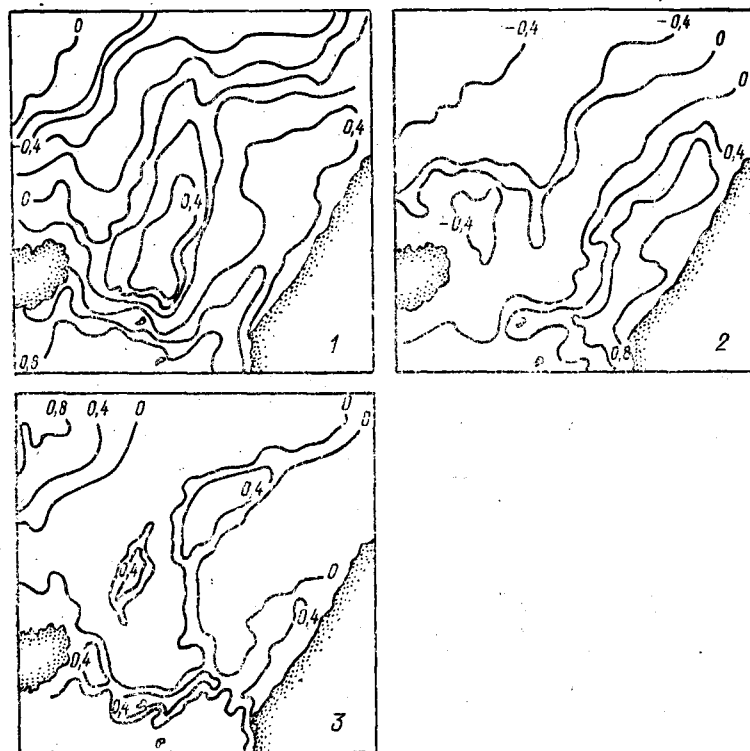


Рис. 5.9. Пространственное распределение трех общих факторов набора гидрофизических полей верхнего слоя Норвежской ЭАЗО.

Таблица 5.3

Статистические характеристики однородных районов, выделенных по результатам анализа среднегодовых полей в деятельном слое Норвежской ЭАЗО

Класс	Параметр			
	$T^{\circ}\text{C}$	$S^{\circ}/_{\text{‰}}$	$\Gamma_T^{\circ}\text{C}/100\text{ км}$	$\Gamma_S^{\circ}/_{\text{‰}}/100\text{ км}$
1	-0,7	34,8	$1,8 \cdot 10^{-1}$	$1,1 \cdot 10^{-4}$
2	2,1	34,9	$4,5 \cdot 10^{-1}$	$1,6 \cdot 10^{-4}$
3	8,6	35,2	$4,1 \cdot 10^{-1}$	$2,0 \cdot 10^{-4}$
4	5,2	35,1	$3,6 \cdot 10^0$	$2,1 \cdot 10^{-3}$
5	3,9	35,1	$5,2 \cdot 10^{-1}$	$6,4 \cdot 10^{-5}$
6	5,3	34,2	$2,4 \cdot 10^0$	$6,7 \cdot 10^{-3}$

Приведенная на рис. 5.10 б структура выделенных районов сохраняется в основных чертах в слое 0—200 м, но глубже происходит значительная перестройка гидрофизических полей.

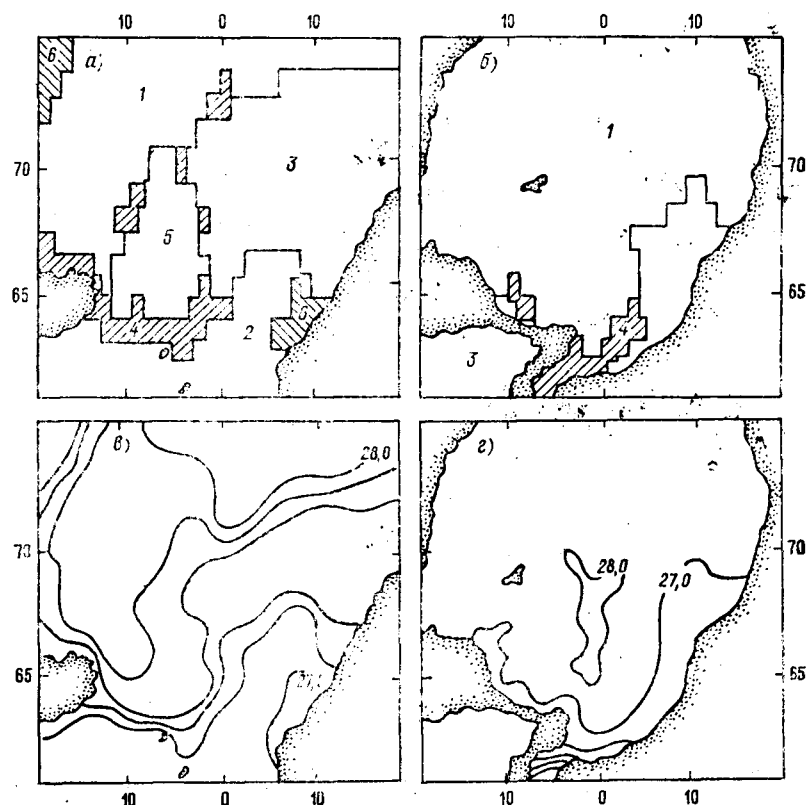


Рис. 5.10. Районирование Норвежской ЭАЭО по результатам классификации среднегодовых гидрофизических в слое 0—200 м (а), 800—1000 м (б) и соответствующие поля условной удельной плотности (в) и (г).

В результате статистической обработки исходных данных вся акватория Норвежской ЭАЭО оказалась разделенной на несколько районов с горизонтальным масштабом $L = 1000$ км, что отражает объективно существующее разнообразие локальных механизмов формирования структуры вод. Очевидно, данный результат является следствием наличия стабильных в пространстве и во времени элементов структуры деятельного слоя исследуемого региона, подобных крупным циклоническим и антициклоническим мезомасштабным круговоротам, которые были зарегистрированы в период проведения натурных наблюдений по программе «ПОЛЭКС—СЕВЕР» [160]. Это предположение подтверждается

результатами расчетов энтальпии вод деятельного слоя по климатическим данным (см. рис. 5.11 б), в которых отчетливо прослеживается наличие замкнутых очагов тепла и холода, близких по масштабу. Высокая интенсивность движения вод внутри выделенных круговоротов отражена в поле удельной потенциальной энергии слоя 0—200 м (см. рис. 5.11 а).

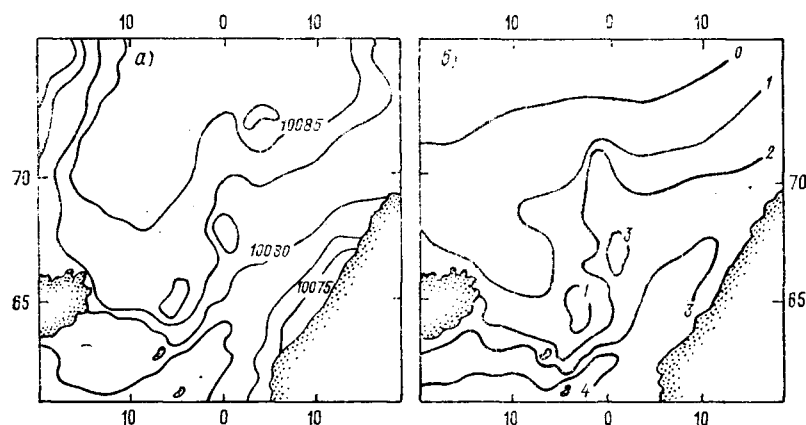


Рис. 5.11. Среднегодовые поля удельной потенциальной энергии (а) и энтальпии (б) слоя 0—200 м.

В слое 300—1000 м характер пространственных неоднородностей существенно меняется: число однородных районов уменьшается, а их горизонтальный масштаб возрастает почти в два раза. Изменяется уровень взаимосвязи между исследуемыми гидрофизическими параметрами: пространственные распределения температуры, солёности, плотности, кислорода оказываются высоко согласованными между собой и объединены в первый общий фактор, а градиентные параметры — во второй.

В результате районирования здесь выделены глубинные воды атлантического и полярного происхождения и полярная фронтальная зона (см. рис. 5.10 г). Полярные и атлантические воды занимают в акватории Норвежской ЭАЗО два однородных района, разделенных полярной фронтальной зоной, расположенной над материковым склоном Скандинавии. Упрощенность взаимосвязей гидрофизических полей может трактоваться прежде всего как следствие единства процессов формирования термохалинной структуры в глубинном слое Норвежской ЭАЗО, обусловленных динамикой крупномасштабных течений.

Представляет интерес различие структуры гидрофизических полей внутри различных водных масс. В табл. 5.4 приведены результаты факторного анализа для атлантических и полярных вод. В атлантической водной массе преобладает влияние термических неоднородностей, которые и формируют основную диспер-

Таблица 5.4

**Корреляция общих факторов и исходных признаков в слое 0—200 м
внутри однородных районов**

Признак	Общие факторы					
	1	2	3	1	2	3
	Атлантические воды			Полярные воды		
T	0,8	0,1	0,3	0,4	0,8	0,2
S	0,3	0,9	—0,2	0,8	0,1	0,5
σ	—0,7	0,6	0,2	0,6	0,6	0,3
O_2	0,9	0,2	0,1	0,3	—0,7	—0,1
Γ_T	0,3	0,5	0,8	0,2	0,6	0,4
Γ_S	0,2	0,1	0,4	0,5	0,2	0,8
Γ_σ	0,3	0,4	0,7	0,4	0,4	0,5
Γ_0	—0,2	0	—0,9	0,2	—0,7	0,3
Вклад, %	49	32	12	37	34	26

сию — 49 %, тогда как в полярной водной массе изменчивость характерна прежде всего для солености и связанной с ней неоднородности поля плотности и горизонтальных градиентов.

Полученный результат отражает основную тенденцию трансформации водных масс: атлантические воды в Норвежском море подвергаются прежде всего термической трансформации и в меньшей степени соленостной, полярные воды — наоборот.

Наличие удовлетворительных результатов статистического диагноза среднегодовых структур вод обусловило попытку реализации аналогичного подхода применительно к среднесезонным гидрофизическим полям в Норвежской ЭАЭО в верхнем двухсотметровом слое. Статистическая обработка того же архива в среднесезонном осреднении полностью адекватна последовательности анализа среднегодовых полей. Поэтому, не вдаваясь в детали статистического анализа, обратимся сразу к заключительной фазе диагноза и рассмотрим основные итоги построения физико-статистической модели формирования сезонного хода гидрологических полей в Норвежской ЭАЭО.

Основой интерпретации механизмов формирования гидрофизических полей служат прежде всего результаты классификации и районирования верхнего слоя акватории Норвежской ЭАЭО в каждом из четырех сезонов. Пространственное расположение выделенных однородных районов представлено на рис. 5.12, а количественные характеристики классов сведены в табл. 5.5.

Наименьшие средние значения горизонтальных градиентов отмечаются в водных массах зимнего сезона. Этот сезон характе-

ризуется повышением интенсивности притока атлантических и полярных вод, а также усилением вертикального конвективного перемешивания, которое в северной части района захватывает значительную толщу вод [41, 124].

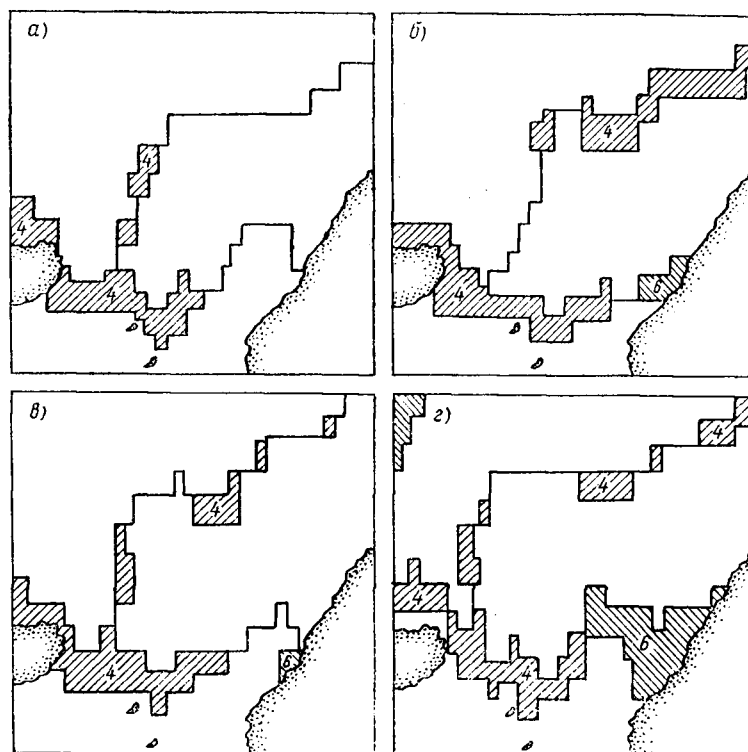


Рис. 5.12. Районирование Норвежской ЭАЭО по результатам классификации гидрофизических характеристик верхнего слоя.

а — зима, б — весна, в — лето, г — осень.

Т, S-характеристики выделенных однородных районов

Таблица 5.5

Класс	Зима		Весна		Лето		Осень	
	T	S	T	S	T	S	T	S
Атлантические воды	6,9	35,2	7,2	35,2	8,2	35,2	8,6	35,2
Полярные воды	-0,3	34,8	-0,1	34,8	0,3	34,8	0,4	34,8
Смешанные воды	2,8	34,9	2,9	34,9	3,2	34,9	3,1	34,9
Полярная фронтальная зона в целом	3,2	35,0	6,0	34,9	4,3	34,9	3,7	34,9
Фарерско-исландский фронт	—	—	5,9	34,6	5,7	34,2	4,2	33,8
Прибрежный фронт	4,7	35,1	5,0	35,0	5,7	35,0	5,6	35,0

В данном случае практически вся значимая трансформация вод сосредоточивается во фронтальных зонах, тогда как водные массы весьма однородны. Наиболее сильно выражена трансформация вод в фарерско-исландской ветви полярной фронтальной зоны (ПВЗ), которая смещена к западу на 200 км по сравнению со среднегодовым ее положением. А ее восточная ветвь, разграничивающая воды Северо-Атлантического и Норвежского течений, оказывается смещенной на север до 67° с. ш.

Для летнего сезона характерно уменьшение адвекции в сочетании с радиационным прогревом вод. Результаты статистического анализа выявляют уменьшение объемов полярных и атлантических вод в верхнем 200-метровом слое и соответствующее изменение T , S -индексов.

На рис. 5.12 отчетливо видно, что адвекция атлантических вод сохраняется интенсивной главным образом в восточной ветви Норвежского течения. При этом граница раздела вод Северо-Атлантического и Норвежского течений смещается на юг до 65° с. ш. На севере Норвежской ЭАЗО на широте $72-74^{\circ}$ выявляется характерный прогиб полярной фронтальной зоны к югу на 300 км, обусловленный, очевидно, усилением циклонической циркуляции в Гренландском море и затоком части холодных полярных вод через подводный хребет Мона в Норвежском море [5, 277].

Особый интерес представляют полученные результаты районирования для переходных сезонов.

Для весеннего сезона результаты отличаются заметным изменением конфигурации фронтальных зон на юге Норвежской ЭАЗО. Как видно из рис. 5.12, фарерско-исландская ветвь ПФЗ, практически не изменив своего положения в западной части моря, в восточной части сместилась на 300 км к побережью Скандинавии. Аналогичный эффект наблюдается в осенний сезон, когда фронтальная зона распресненных стоковых вод граничит с восточной ветвью полярной фронтальной зоны на широте 66° .

Обобщая полученные результаты анализа сезонных полей, можно утверждать, что дважды в год на исследуемой акватории происходит существенная перестройка всей системы фронтальных зон и водных масс. Однако наиболее ярко эти эффекты прослеживаются в районе южнее 65° с. ш., что связано с изменчивостью границ распространения вод атлантического происхождения. Отмеченная особенность может быть интерпретирована как результат периодической модуляции притока атлантических вод в Норвежское море. Выявленная динамика фронтальных зон Норвежской ЭАЗО вполне удовлетворительно увязывается с неоднократно рассматриваемой [197, 212] полугодовой составляющей в потоке Северо-Атлантического течения, минимумы которой отмечены именно в весенний и осенний сезоны.

Статистический анализ показывает, что сезонный ход пространственной структуры полей в Норвежской ЭАЗО определяется изменением взаимосвязей между всеми гидрофизическими параметрами. Это отчетливо прослеживается в результатах фак-

торного анализа восьми исследуемых переменных — температуры, солености, плотности, содержания кислорода, градиентов этих характеристик (табл. 5.6).

Очевидно, что от сезона к сезону меняется сложность обобщенной статистической модели взаимосвязи гидрофизических характеристик: простейшая модель, состоящая из двух общих факторов, наблюдается зимой, наиболее сложная модель (из четырех общих факторов) — осенью; соответственно меняется точность статистической модели.

Данный результат можно трактовать как проявление различной степени согласованности реакции комплекса гидрофизических показателей на суммарное внешнее воздействие. Действительно, в зимний период, когда деятельный слой Норвежской ЭАЭО подвергается интенсивному выхолаживанию, развивающееся конвективное перемешивание является причиной высокой коррелированности температуры, солености, плотности и кислорода.

В полярной фронтальной зоне поддерживаются высокие градиенты в сочетании с интенсивными нисходящими движениями вод [5, 51, 124, 139]. Эти причины обусловили объединение большинства градиентных показателей в один общий фактор, а термохалинных полей — в другой, причем на поверхности относительная дисперсия градиентного фактора оказалась выше, чем термохалинного.

Подобная структура факторного решения сохраняется на горизонте 200 м в течение года без значимых изменений от сезона к сезону. Здесь меняется лишь относительный вес факторов — в силу ослабления перепадов термохалинных характеристик на фронтах, градиентный фактор играет второстепенную роль по сравнению с термохалинным.

Наиболее сложная статистическая модель взаимосвязей гидрофизических параметров выявляется в осенний сезон, когда при четырех общих факторах описывается 87 % дисперсии набора исходных полей. В данный сезон пространственное распределение полей температуры и солености оказывается существенно рассогласованным в силу различия причин, формирующих их аномальность. Распределение солености оказывается чрезвычайно неоднородным и, по сути, определяет аномалии пространственного распределения поля плотности. Изменчивость соленостных компонент дисперсии исходных переменных объединяется в первом общем факторе, который играет наибольшую роль в этот период. Значения соленостных факторов 1, 4 максимальны в восточной части Норвежской ЭАЭО в районе, примыкающем к побережью Скандинавии и, очевидно, связаны с растеканием распресненных вод материкового стока в поверхностном слое толщиной менее 150 м. На горизонте 200 м факторная структура осенних гидрофизических полей практически не отличима от других сезонов.

В весенний и летний сезоны сложность взаимосвязей гидрофизических полей верхнего 200-метрового слоя Норвежского моря носит промежуточный характер от зимы к осени и описывается

Корреляция общих факторов и исходных параметров

Глубина, м	Параметр	Зима		Весна			Лето			Осень			
		1	2	1	2	3	1	2	3	1	2	3	4
0	T	-0,5	0,8	0,2	-0,6	0,7	-0,1	-0,3	0,8	-0,1	-0,2	-0,8	-0,1
	S	0,1	-0,9	0,9	0,1	-0,1	0,9	0	0,1	0,6	-0,1	0,1	0,8
	σ	0,1	-0,5	0,8	0,1	-0,6	0,7	0,2	-0,5	0,5	0,2	0,4	0,7
	O_2	0	-0,9	0,2	0	-0,9	0,1	0,2	-0,9	0,1	0,2	0,9	0,1
	Γ_T	0,9	0,2	0	0,9	0	0,1	0,9	-0,2	0,1	0,9	0,2	0,1
	Γ_S	-0,1	0,2	-0,9	-0,1	0,2	0,9	-0,1	0,1	0,9	0	-0,1	-0,3
	Γ_σ	-0,9	0,2	-0,9	0,1	0,3	-0,9	0	0,3	0,9	-0,1	-0,2	-0,2
	Γ_{O_2}	0,9	0,1	0,1	0,9	0,1	0	0,9	-0,3	0,1	0,9	0,4	0,1
200	T	0,9	-0,1	0,9	0,1		0,9	-0,1		0,9	0,1		
	S	0,7	-0,3	0,7	0,2		0,8	-0,1		0,6	0,2		
	σ	-0,9	0,1	-0,9	0,1		-0,9	0		-0,9	0,2		
	O_2	-1,0	0	-1,0	0		-0,9	0		-1,0	0		
	Γ_T	-0,5	0,8	-1,0	0		-0,2	0,9		0,1	0,9		
	Γ_S	0,3	0,6	0,1	-0,9		0,3	0,5		0,1	0,8		
	Γ_σ	-0,1	0,9	0,1	0,9		0,1	0,9		0,1	0,9		
	Γ_{O_2}	-0,6	0,8	-0,2	0,9		-0,3	0,9		0,1	0,9		

тремя общими факторами. Согласно статистической модели, пространственная изменчивость поля температуры и связанных с ней компонент других параметров играет основную роль в формировании неоднородности — доля температурного фактора составляет около 40 % в весенне-летний период. В отличие от зимнего сезона здесь уже проявляется значимая рассогласованность пространственного распределения температуры и солености, что следует из обособления термической и соленостной компонент по различным общим факторам.

Таким образом, факторная модель выявляет и количественно описывает неоднородность пространственной структуры комплекса исследуемых гидрофизических параметров деятельного слоя Норвежской ЭАЗО как единой системы. Высокая общность параметров, наблюдаемая в зимний сезон, постепенно разрушается под воздействием различных сил; в весенний сезон изменчивость температуры и солености уже вынужденно параметризуется отдельно. Максимум рассогласования параметров выявлен в осенний сезон, когда отдельно параметризуются поля горизонтальных градиентов температуры и солености.

Оценивая информативность отдельных исследуемых переменных на основе построенных факторных моделей, следует выделить преобладающую значимость температуры как признака, наиболее достоверно описывающего пространственные неоднородности исследуемых параметров в зимний, весенний и летний сезоны. Исключением является осень, когда информативность полей температуры и солености в целом сравнима (с некоторым преобладанием значимости последней).

Полученные результаты многомерного статистического анализа могут иметь обоснованную интерпретацию в энергетических терминах. Рассматривая в данном аспекте сезонный ход структуры гидрофизических полей верхнего слоя Норвежской ЭАЗО, следует обратить внимание на неравномерность пространственного накопления и расхода энергии в годовом цикле. Пространственное распределение внутренней и потенциальной энергии в зимний сезон наиболее сглажено и взаимно согласовано. С началом теплого полугодия постепенно возрастает внутренняя энергия и медленно уменьшается потенциальная при нарастающем взаимном рассогласовании.

В осенний сезон отмечается экстремум расхождений пространственного распределения энергий, который сочетается с максимальным числом локальных энергетических аномалий от среднего по акватории фона. Именно летом и осенью наиболее отчетливо прорисовывается пространственная очаговость термодинамической структуры полей Норвежской ЭАЗО, выявленная по среднегодовым данным.

На рис. 5.13 показан меридиональный перенос тепла по сезонам в слое 0—200 м — наличие мезомасштабных очагов накопления тепла характерно именно для лета и осени. Кроме того, на этом рисунке отчетливо видна роль фронтальных разделов

в переносе тепла. В переходные сезоны прослеживается заметное уменьшение потока тепла, приносимого водами Северо-Атлантического течения в Норвежском море. Зональное перемещение фарерско-исландской ветви ПФЗ от сезона к сезону формирует

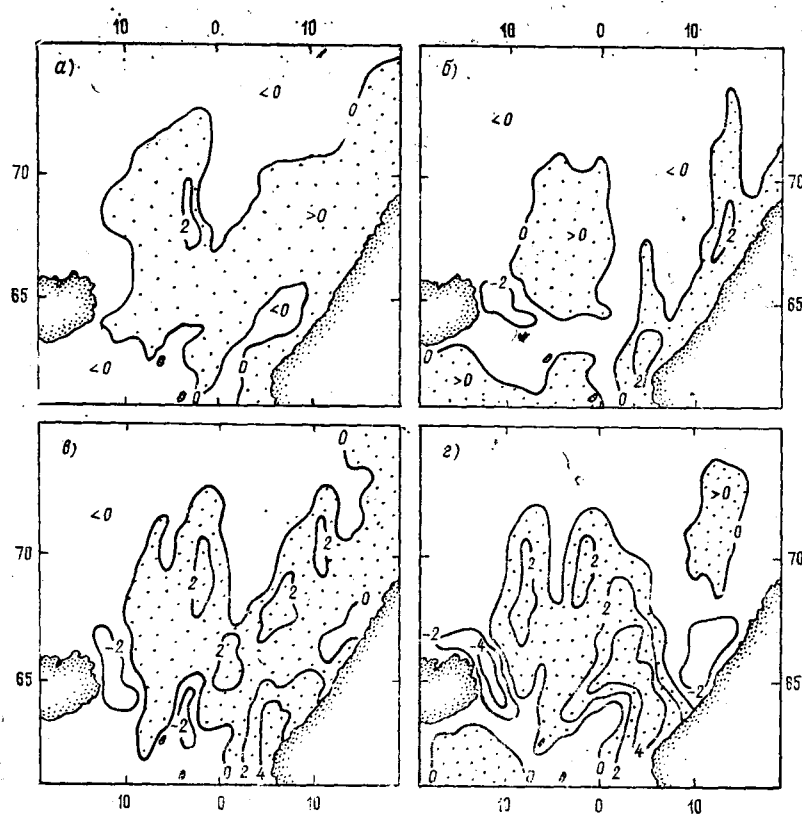


Рис. 5.13. Меридиональный перенос тепла геострофическими течениями в слое 0—200 м в Норвежской ЭАЭО.

а — зима, б — весна, в — лето, з — осень.

ширину и интенсивность меридионального потока тепла в южной и центральной частях Норвежского моря.

Для весеннего сезона это выражается в локализации переноса тепла в узкой струе восточной ветви Норвежского течения располагающейся вдоль побережья Скандинавии. Осенью этот поток тепла отклоняется на широте 66° к западу под влиянием фронтального раздела стоковых вод Скандинавского полуострова. Другими словами, в осенний сезон локальные особенности распределения потенциальной энергии приводят к изменению структуры гидрофизических полей в целом.

Наконец, с переходом от осени к зиме в очень сжатые сроки происходит резкая перестройка взаимосвязей всех гидрофизических полей, уменьшение внутренней энергии и рост согласованности пространственного распределения внутренней и потенциальной энергий.

Полученная пространственная конфигурация выделенных классов демонстрирует наличие в каждом из сезонов года разветвленной системы фронтальных зон, которые располагаются на границах генетически различных водных масс. Величина общей протяженности фронтальных зон в деятельном слое Норвежской ЭАЭО составляет примерно $7 \cdot 10^3$ км, а объем заключенных в них вод — около $5 \cdot 10^3$ км³.

Пространственная конфигурация фронтальных зон может быть условно разделена на два типа: полосы и пятна. Полосы фронтов наиболее распространены в рассматриваемом районе в виде вытянутых зон интенсивной трансформации. Идентифицируются полярная зона и ее основные ответвления — фарерско-шетландская и фарерско-янмайенская. Ширина данных фронтальных зон, как правило, не превышает 200 км, причем на отдельных участках они вырождаются в линию. Последняя особенность наиболее характерна для фарерско-янмайенской фронтальной зоны, где протяженность вырожденных участков составляет до 400 км.

Фронтальные разделы в виде высокоградиентных пятен регистрируются, как правило, в двух районах Норвежской ЭАЭО: на шельфе Скандинавии и на северо-западе региона в зоне Восточно-Гренландского течения. Продольные и поперечные размеры данных фронтальных зон практически равны друг другу и составляют примерно 500 км.

Характерно, что пятна фронтов сосредоточены в верхнем 100-метровом слое и ниже не прослеживаются, тогда как фронтальные разделы в виде полос регистрируются отчетливо в верхнем 200-метровом слое и даже на глубинах до 600 м. Очевидно, данная особенность является следствием различных механизмов образования фронтальных разделов в Норвежской ЭАЭО.

Как известно, во фронтальных зонах происходит трансформация большинства гидрофизических параметров: в работе К. Н. Федорова [199] предложено рассматривать прежде всего изменчивость температуры, солености и плотности как наиболее информативных показателей физической природы фронтов. Для исследования статистической взаимосвязи параметров внутри фронтальных зон воспользуемся методикой факторного анализа.

Построение локальных факторных моделей для фронтальных разделов позволило выявить и количественно описать взаимосвязь полей температуры, солености, плотности. Анализ продемонстрировал неоднородность взаимосвязей и их зависимость от пространственной конфигурации фронтальных зон.

Так, для участка полярной фронтальной зоны вдоль нулевого меридиана и вдоль фарерско-исландского порога отмечается удовлетворительная согласованность термической и соленостной измен-

чивости и ее совпадение по знаку: коэффициент парной корреляции температуры и солености внутри ПФЗ составил $r = 0,6$. При этом взаимосвязь температуры и плотности оказалась выше, чем солености и плотности, что обусловило выдвижение в факторной модели на первое место по значимости термической изменчивости как наиболее важной в ПФЗ. Из табл. 5.7, в которой приведены

Таблица 5.7

Взаимосвязь гидрофизических параметров и общих факторов
внутри фронтальных зон для осеннего сезона в слое 0—200 м

Параметр	Фактор							
	1	2	3	4	1	2	3	4
	Полярная ФЗ				Скандинавская ФЗ			
T	0,8	0,1	0,3	0,2	0,2	0,7	0,1	0,1
S	0,2	0,9	0	0,4	0,9	0,1	0,4	0
σ	0,7	0,4	0,1	0,1	0,8	0,4	0,3	0,1
O_2	-0,7	0,2	-0,2	0	0,1	-0,9	0	0,1
Γ_T	0,4	0	0,8	0,3	0	0,4	0,1	0,9
Γ_S	0,1	0,3	0,4	0,7	0,5	0	0,8	0,1
Γ_σ	0	0,2	0,7	0,6	0,7	0,1	0,7	0,2
Γ_{O_2}	-0,3	-0,1	0,9	0,2	-0,2	0,2	0	0,8
α_T	$2,8 \cdot 10^{-4}$				$0,7 \cdot 10^{-4}$			
β_T	$1,2 \cdot 10^{-4}$				$11,4 \cdot 10^{-4}$			

результаты факторного анализа в осенний сезон, очевидна преобладающая роль термического фактора (фактор 1) в ПФЗ. Здесь же для сравнения приведены результаты для фронтальной зоны у побережья Скандинавии, где структура связи гидрофизических параметров оказывается совершенно иной: корреляция температуры и солености практически не наблюдается ($r = 0,1 \dots 0,2$), поле плотности определяется изменчивостью солености ($r = 0,68$). В факторной модели на первом месте по значимости стоит соленостный фактор.

Наиболее интенсивная изменчивость характеристик фронтальных зон от сезона к сезону отмечается в поверхностном слое толщиной 100 м. Сказанное следует прежде всего связывать с появлением, перемещением, интенсификацией и ослаблением фронтальных пятен на северо-западе и юго-востоке рассматриваемого региона, обусловленных взаимодействием водных масс различной солености. Изменчивость фронтальной зоны на шельфе Скандинавии, очевидно, вызвана тремя основными группами факторов: интенсивностью стока распресненных материковых вод; интен-

сивностью адвекции высокосоленных вод Норвежским течением и распресненных вод Балтийским течением; локальными динамическими процессами, характером приливов, которые определяют скорость перемешивания вод внутри фронтального пятна.

Основными фронтообразующими факторами в данном районе являются потоки Норвежского течения и вод материкового стока — характером их взаимодействия определяется сам факт наличия или отсутствия фронтальной зоны на шельфе Скандинавии. Так, зимой и весной, когда пресный сток минимален, фронтальная зона здесь практически исчезает при одноградусном пространственном осреднении и не оказывает существенного влияния на динамику вод Норвежской ЭАЗО. Максимум стока на реках Скандинавии наблюдается в начале лета [46, 272]. Однако наибольший объем пятна распресненных вод на шельфе имеет место в осенний сезон, что обусловлено наличием глубоких фиордов на побережье Норвегии, в которых происходит накопление пресных вод летом.

В осенний сезон фронтальное пятно распресненных вод на шельфе Скандинавии уже играет значительную роль в гидрофизике Норвежской ЭАЗО в целом; это обусловлено следующими причинами:

- 1) уменьшается адвекция тепла на северо-восток в Баренцевом море вдоль побережья Норвегии,
- 2) ослабляется восточная ветвь Норвежского течения и усиливаются центральная и западная ветви (см. рис. 5.13),
- 3) отрицательная аномалия солёности морской воды обуславливает уменьшение потенциальной энергии слоя 0—200 м.

Фронтальная зона на северо-западе Норвежской ЭАЗО имеет аналогичный сезонный ход, который связан прежде всего с изменчивостью потока Восточно-Гренландского течения и таянием льда на поверхности под действием радиационных факторов. Данная фронтальная зона прослеживается в виде высокоградиентного пятна в слое 0—50 м, где трансформация солёности определяет изменчивость поля плотности — структура факторной модели практически совпадает с моделью скандинавского фронта.

Сезонная изменчивость ПФЗ и ее ветвей наиболее четко прослеживается в поверхностном 100-метровом слое вод. В теплый период года наблюдается тенденция к расширению зоны фронта до 300 км на отдельных участках в южной части Норвежского моря и увеличение перепада солёности и температуры. В осенний сезон отмечается максимальный градиент температуры $\Gamma_T = 3^\circ\text{C}/100\text{ км}$ и солёности $\Gamma_S = 0,1\text{‰}/100\text{ км}$. В зимний и весенний сезоны имеет место обратная тенденция, когда площадь фронтальных зон резко сокращается, фарерско-янмайенская ветвь ПФЗ почти полностью вырождается в линию раздела различных водных масс при данном пространственном осреднении.

Изменение структуры статистических моделей гидрофизических параметров внутри фронтальных зон практически незначительно и находится на грани значимости при уровне вероятности 90 %.

Отмеченная особенность позволяет утверждать наличие стабильных механизмов формирования фронтальных зон центральной части Норвежской ЭАЗО в течение всего года. Превалирующее влияние термической изменчивости на формирование плотностных неоднородностей во фронтальных зонах (см. табл. 5.7) также сохраняется в течение всего года. Таким образом, на основании анализа климатических полей следует заключить, что эффективный мониторинг динамики системы полярной фронтальной зоны в Норвежском море следует прежде всего связывать с исследованием термического режима верхнего слоя океана в данном регионе.

Проведенный корреляционный анализ показал, что наилучшим образом положение ПФЗ описывает изотерма 6°C на поверхности и 5°C на горизонте 100 м в летний сезон; для зимнего сезона — изотермы 5 и 4°C для тех же горизонтов соответственно.

Начиная с горизонта 200 м и глубже проявляется значимая корреляционная связь положения южной части ПФЗ с нижележащим рельефом дна и с нормальной к фронту составляющей градиента глубины. Наиболее высокие значения коэффициентов корреляции характерны для фарерско-исландской ветви ПФЗ; они отмечаются для горизонтов 400 и 600 м и составляют 0,74 и 0,81 соответственно.

Следовательно, на пространственно-временной изменчивости ПФЗ должны сказываться две группы факторов: изменчивость внешних по отношению к Норвежской ЭАЗО крупномасштабных течений Атлантической и Арктической циркуляционных систем; короткопериодная изменчивость термодинамических параметров под влиянием локального взаимодействия океана с окружающей средой.

Первая группа факторов определяет общую конфигурацию системы ПФЗ и ее интенсивность. Вторая группа ответственна за мезомасштабную динамику фронтальных зон. Работа крупномасштабных сил уже рассматривалась в предыдущем параграфе, где была показана изменчивость положения фарерско-исландской ветви ПФЗ под влиянием полугодовой составляющей в Северо-Атлантическом течении. Отмечаемое при этом горизонтальное смещение фронта на несколько сот километров в широтном направлении демонстрирует чувствительность ПФЗ к внешнему воздействию, и лишь ограничения используемого архива данных не позволяют оценить степень подобной изменчивости во всем слое от поверхности до нижней границы потока атлантических вод (600 м).

5.3. Диагноз структуры водных масс в море Беллинсгаузена

Объекты исследований для примеров, рассматриваемых в п. 5.1, 5.2, представляют собой акватории, которые достаточно хорошо изучены в течение последних десятилетий. Причем подобное изучение проводилось как с помощью натурных наблюдений,

так и на основе теории. Поэтому статистический анализ служил инструментом проверки и уточнения в основном уже априорно известных фактов или гипотез.

На примере моря Беллинсгаузена представляется возможность проанализировать эффективность рассматриваемого статистического подхода к изучению водных масс в относительно малоизученном регионе Южного океана. Важным аспектом, требующим самостоятельного изучения, является и ставшее весьма распространенным мнение о крайней сложности вообще исследования и достоверного анализа водных масс в этом районе. Поэтому, приступая к диагнозу, авторы ставили перед собой следующие задачи:

- добиться по возможности объективной и достоверной классификации водных масс в море Беллинсгаузена,
- изучить пространственную структуру распределения климатических водных масс по горизонтали и вертикали во всей акватории,
- оценить методические возможности применения статистического анализа для малоизученных районов Южного океана.

При постановке задачи авторы опирались на те результаты, которые были получены по этому региону в предшествующей работе [162].

В процессе вычислений, следуя общей методологии статистического анализа, можно выделить три основные задачи:

- первичный статистический анализ,
- разложение набора переменных по главным компонентам,
- классификация водных масс в координатах первых компонент.

На этапе синтеза результатов следует выделить процедуры районирования классов, восстановления исходных данных по главным компонентам и физическую интерпретацию полученных результатов в рамках обобщенной климатической модели термохалинного режима моря Беллинсгаузена.

В качестве исходных данных использовались материалы порейсового архива экспедиционных исследований, накопленные в ААНИИ. Архив содержит наблюдения основных гидролого-гидрохимических показателей на стандартных горизонтах для 841 океанографической станции.

Период наблюдений охватывает ряд лет с 1938 по 1978 г. Архив содержит наблюдения шести параметров: температуры, солености, концентраций растворенного кислорода, нитратов, кремния и фосфатов. Однако в шельфовой зоне моря практически отсутствуют данные о концентрациях биогенных элементов и вектор признаков содержит только три характеристики: температуру, соленость и концентрацию кислорода. В связи с подобной структурой исходных данных дополнительно решалась задача поиска наиболее информативных переменных (признаков), минимальное число которых обеспечивало бы достаточно надежную классификацию водных масс в море Беллинсгаузена.

Обратимся к основным результатам статистического анализа. Главный итог первичной статистической обработки состоит в том, что дисперсии исходных признаков различаются между собой на несколько порядков. Так, дисперсия концентраций фосфатов, кремния на 2—3 порядка превышает дисперсию температуры и солености. Подобное соотношение еще раз подтвердило известный факт [162, 199], а именно, слабую значимость классических океанологических характеристик температуры и солености — для ре-

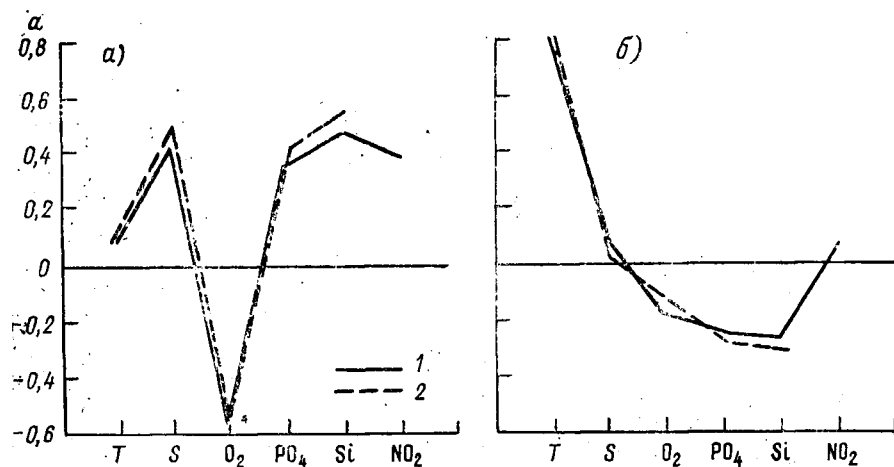


Рис. 5.14. Первый (а) и второй (б) собственные векторы ортогонального разложения гидрофизических полей в акватории моря Беллинсгаузена по шести (1) и пяти (2) признакам.

шения задач анализа водных масс на фоне резко возросшей значимости гидрохимических признаков. Выделяется характерная черта в неоднородности пространственной изменчивости исследуемых признаков. Целая группа признаков (кремний, соленость, кислород) имеет максимальную изменчивость в вертикальном направлении, в 3—10 раз превышающую горизонтальную изменчивость. Наоборот, для температуры более характерна изменчивость по горизонтали, причем, главным образом, в меридиональном направлении.

Расчет эмпирических гистограмм выявляет наличие многомодальных распределений практически для всех исследуемых признаков. Аналогичная особенность прослеживается при рассмотрении двумерных гистограмм признаков, причем особенно ярко — для гистограмм $P(S, Si)$ и $P(S, O_2)$. Гипотеза нормальности для рассматриваемых эмпирических гистограмм не выполняется при априорно заданном уровне значимости 10 %. Гипотезы нормальности оправдываются только для гистограмм трех признаков — солености, кислорода и кремния, но при доверительной вероятности 70—80 %. Температура воды на вычисленных гистограммах характеризуется своеобразной изменчивостью, которая на

качественном уровне не может быть интерпретирована в терминах устойчивой взаимосвязи с другими переменными.

В целом полученные результаты первичного анализа еще раз продемонстрировали чрезвычайную сложность выбора информативных признаков для анализа структуры вод. Проблема информативности признаков может рассматриваться с различных позиций.

Так, при анализе вертикальной расслоенности вод океана оперируют, как правило, информативными с физической точки зрения переменными — температурой, соленостью, плотностью и их вертикальными градиентами. Выполняя классический *T, S*-анализ водных масс, исследователь использует генетическую информативность температуры, солености, кислорода и т. п., которая позволяет охарактеризовать водные массы, восстановив присущие им свойства в момент зарождения. Статистическая информативность признаков может трактоваться как информативность конкретных переменных для устойчивого и состоятельного вычисления требуемых статистических оценок.

В данном случае, когда основой диагноза структуры вод служат методы многомерного корреляционного анализа, термин статистической информативности во многом совпадает с понятием значимости коэффициентов парной корреляции.

Рассматривая полную корреляционную матрицу исходных переменных, можно сразу выделить группу значимых корреляционных оценок, охватывающих пары признаков: концентрация кислорода и кремния, соленость, для которых оценки коэффициентов корреляции колеблются в пределах 0,70—0,95. Менее сильные, но значимые статистические связи прослеживаются для пар: концентрация фосфатов — кислорода ($-0,42$), концентрация фосфатов — кремния ($0,4$), температура — концентрация фосфатов ($-0,58$) и концентрация нитратов — кремния ($0,28$).

Полученные характеристики взаимосвязей исходных признаков существенно отличаются от подобных корреляционных оценок, выполняемых в других районах Мирового океана. Поэтому для окончательного вывода об информативности исходных переменных был выполнен анализ главных компонент по материалам наблюдений на меридиональном разрезе в центральной части моря Беллинсгаузена.

Данный разрез характеризуется общей полнотой и достоверностью наблюдений всех исходных переменных, все гидрологические станции разреза выполнены до дна. Перечисленные особенности и обусловили в конечном счете выбор материала для решения задачи информативности признаков.

В результате вычислений получено разложение со следующими основными характеристиками (рис. 5.14).

1. Наиболее информативен в компонентном разложении набор переменных: соленость и концентрация растворенного кислорода и кремния, суммарная изменчивость которых обобщается в первом члене разложения, описывающем до 77 % суммарной диспер-

сии. Исключение какого-либо параметра из списка перечисленных переменных резко снижает сходимость разложения и ухудшает интерпретацию компоненты вследствие уменьшения эксцесса ее гистограммы.

2. Вторая компонента описывается изменчивостью температуры и концентрации фосфатов. Относительная информативность компоненты в разложении не превосходит 25 %. Следует отметить, что влияние распределения концентрации фосфатов на формирование второго собственного вектора заметно слабее, чем влияние изменчивости температуры. Коррелированность температуры со второй компонентой весьма высока и составляет 0,89.

3. Концентрация нитратов, как показывает анализ, не является информативным признаком для компонентного разложения. Эта переменная частично участвует в формировании первого и второго членов разложения, однако ее влияние не является определяющим в расчетах. Более того, можно утверждать, что присутствие концентрации нитратов в списке исходных переменных способствует размыванию формы первых членов разложения, и облако рассеяния наблюдений в пространстве главных компонент теряет строгую направленную форму.

Основным итогом проведенного анализа статистической информативности признаков является обоснованное решение об исключении концентрации нитратов из списка исследуемых переменных. Для рассматриваемой модели водных масс данная переменная не может считаться статистически информативной, поскольку не удается с помощью используемых методов выявить значимые взаимосвязи с другими характеристиками водных масс. Изменчивость концентрации нитратов весьма специфична и может представлять интерес, если в качестве дополнительных параметров в статистическую модель включить концентрации других форм содержания азота в морской воде. Однако в таком случае статистическая модель описывала бы главным образом цикл превращений азота, что выходит за рамки поставленной задачи.

Второй, не менее важный итог состоит в том, что потенциально возможный уровень описания изменчивости исходных данных в многомерной статистической модели будет находиться в пределах 80—90 % суммарной дисперсии переменных. Полученный интервал вероятности определяется уровнем суммарного описания дисперсии переменных двумя главными компонентами пробного разложения полей. Достижение более высокой точности описания изменчивости гидрофизических полей в море Беллинсгаузена сдерживается заданной структурой исходных данных.

Разложение по методу главных компонент всего имеющегося материала по акватории моря Беллинсгаузена позволило получить более общую количественную оценку пространственной взаимосвязи исследуемых полей. Структура описания дисперсии признаков представлена в табл. 5.8.

Очевидно, что достигнута высокая сходимость разложения, которая обеспечивает целостность статистической модели гидро-

Таблица 5.8

**Характеристики разложения по главным компонентам
гидрофизических полей моря Беллингаузена**

Компо- нента	Число исходных признаков					
	6			5		
	λ	d %	P %	λ	d %	P %
1	2,9	48,9	48,9	3,6	71,0	71,0
2	1,4	23,3	12,2	0,7	15,8	86,8
3	0,7	12,0	84,2	0,5	12,1	98,9
4	0,7	11,1	95,3	0,1	0,7	99,6
5	0,3	2,7	98,0	0,1	0,4	100,0
6	0,1	2,0	100,0			

физических процессов во всем море Беллингаузена. Характерно, что два первых члена разложений уже описывают в сумме около 85 % дисперсии исходных полей, что соответствует необходимой точности восстановления данных. Третий член разложения описывает лишь 7 % дисперсии, что позволяет ограничиться для качественного анализа первыми двумя членами разложения.

Первая компонента формируется суммарным влиянием солёности и концентрации кремния, кислорода и фосфатов, которые сохраняют значимые коэффициенты корреляции с данной компонентой. Вторая компонента практически полностью связана с описанием изменчивости температуры и в меньшей степени — изменчивости концентраций фосфатов и кремния. В третьей компоненте сосредоточено описание взаимосвязей остаточной изменчивости концентрации фосфатов и солёности. Все перечисленные особенности отчетливо прослеживаются на графиках собственных векторов.

Для выявления структуры вод принципиальное значение имеет вид двумерных гистограмм первых главных компонент. Наиболее заметны два характерных сгущения точек наблюдений. Линия разделения «облаков» совпадает с нулевыми значениями компонент, что наводит на мысль о различных условиях формирования соответствующих наблюдений. К этому же заключению можно прийти, если обратить внимание на различную форму облаков и разную их ориентацию в пространстве главных компонент. Кроме того, существуют и более мелкие группы наблюдений, которые прослеживаются на гистограммах с учетом третьей компоненты. Здесь число визуально выделяемых сгущений достигает 5—8, а их взаимное разделение зачастую весьма приблизительно.

Очевидно, что проделанные вычисления уже позволяют составить более точное представление о степени неоднородности исследуемых полей и о согласованности изменения отдельных параметров. Однако для получения возможно более объективной кар-

тины структуры гидрофизических полей требуется осуществить классификацию исходных наблюдений в пространстве главных компонент.

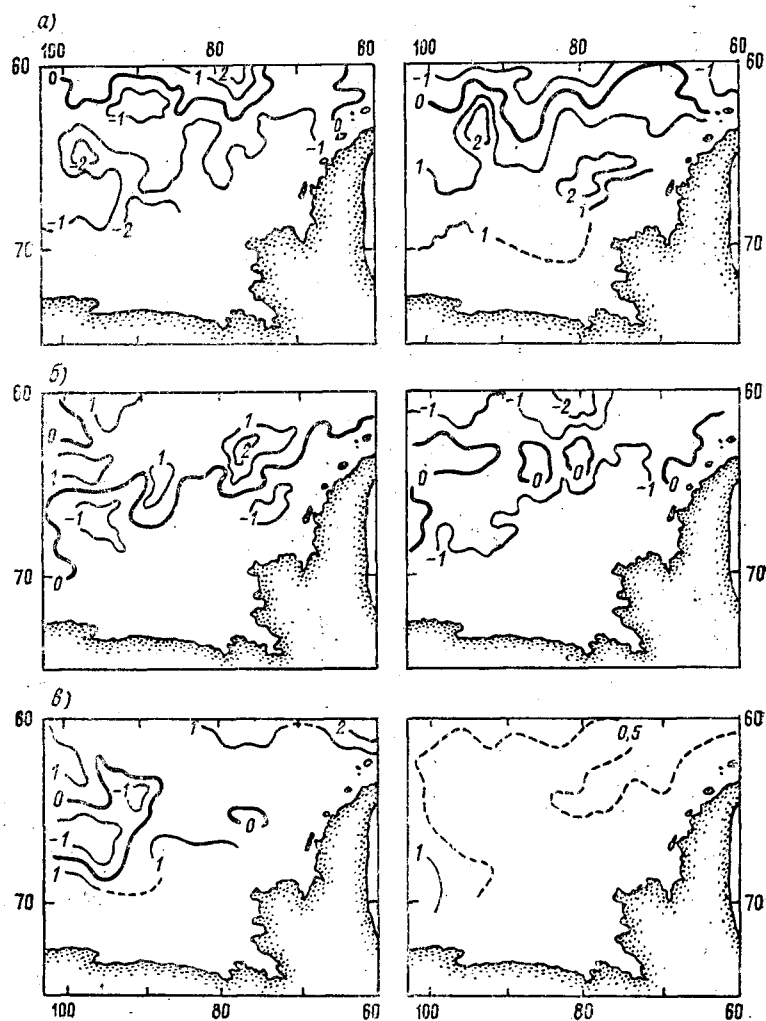


Рис. 5.15. Пространственное распределение первых двух главных компонент разложения на горизонтах.
а — 50 м, б — 200 м, в — 3000 м.

Автоматическая классификация была выполнена в пространстве трех главных компонент по алгоритмам, изложенным выше. В результате классификации получены два наиболее устойчивых разбиения всей выборки наблюдений на классы: 1-й вариант содержит два класса, 2-й — пять классов. Для каждого из классов

рассчитаны оценки средних значений исходных переменных, которые послужили основой для последующей интерпретации ре-

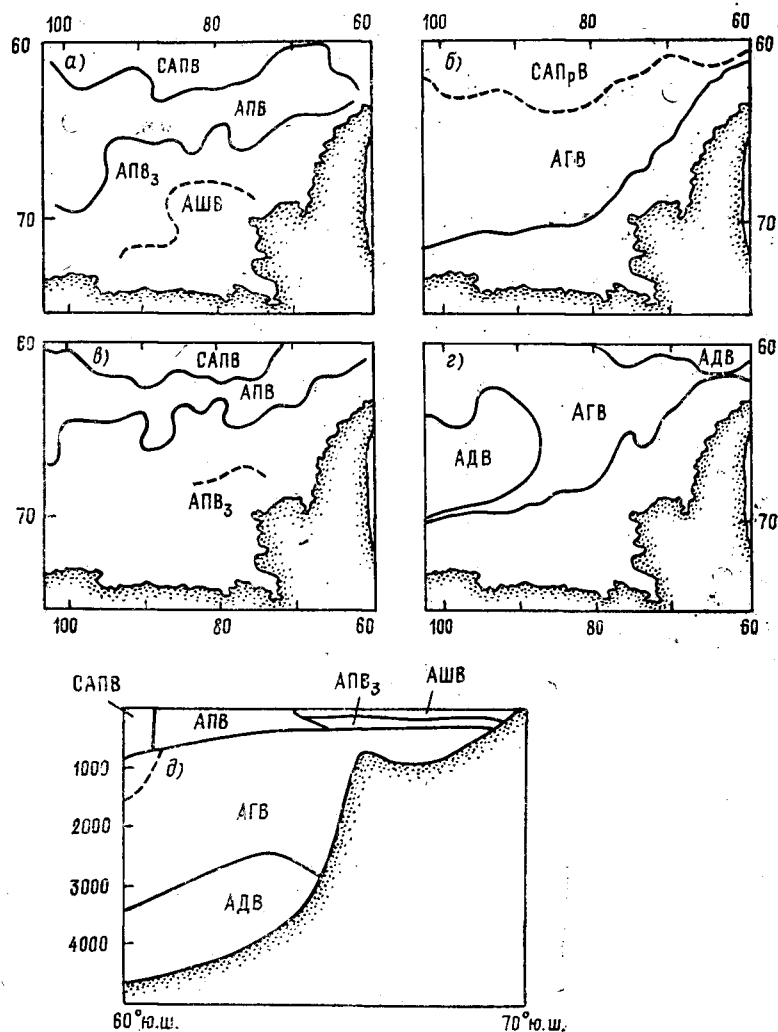


Рис. 5.16. Пространственное распределение основных водных масс моря Беллингаузена по результатам классификации на горизонтах 50 м (а), 200 м (б), 1000 м (в), 3000 м (г) и на меридиональном разрезе по 90° з. д. (д).

зультатов и построения общей физико-статистической модели структуры вод в море Беллингаузена (рис. 5.16).

Интерпретация результатов разложения и классификации гидрофизических полей производилась путем сопоставления особенностей пространственного распределения выделенных классов и

их количественных средних характеристик. В исследуемом районе объективно с вероятностью не менее 93 % исходные данные по комплексу признаков можно подразделить на две крупные структурные зоны — поверхностную и глубинную.

Поверхностная структурная зона состоит из трех однородных водных масс, которые в соответствии с T , S -индексами могут интерпретироваться (табл. 5.9) как антарктическая поверхностная

Таблица 5.9

Средние значения гидрофизических характеристик внутри выделенных классов

Класс	Значения признаков				Водная масса
	T °C	S ‰	O_2 , мг/мл	Si, мг/мл	
1	0,3—2,4	33,8—38,1	7,6—8,3	12—20	АПВ
2	—1,7—1,5	33,9—34,3	8,0—8,4	20—60	АПЗВ
3	3,3—6,8	34,1—38,2	6,9—7,2	4—15	САПВ
4	1,0—2,3	34,1—34,7	4,2—5,1	60—110	ГВВ
5	0,1—0,7	34,7—34,8	1,8—5,3	90—160	ДВВ

(АПВ) — класс 1, антарктическая поверхностная зимняя (АПЗВ) — класс 2, субантарктическая поверхностная — класс 3 (САПВ). Пространственное расположение выделяемых водных масс приведено на рис. 5.16 для горизонтов 50 и 200 м. Там же для горизонтов 1000 и 3000 м показано распределение водных масс глубинной структурной зоны. В этой зоне выделены две основные водные массы: антарктическая глубинная (АГВ) — класс 4, антарктическая донная (АДВ) — класс 5. Вероятность объективного выделения перечисленных пяти водных масс по имеющимся данным составляет 76 %.

Снижение надежности выделения отдельных водных масс может быть, в частности, объяснено слабой точностью классификации ряда наблюдений для антарктической и субантарктической поверхностных водных масс. Эмпирические гистограммы этих классов частично перекрываются и, кроме того, имеют характерную форму с утяжеленными «хвостами». Эта особенность указывает на необходимость более точной интерпретации взаимодействия водных масс в поверхностной структурной зоне. Для этого был произведен повторный расчет по данной схеме диагноза для отдельных локальных разрезов в прибрежной и открытой части моря Беллинсгаузена.

В результате интерпретации полученных результатов можно дополнительно выделить две водные массы — Субантарктическую промежуточную (САПрВ) и антарктическую шельфовую (АШВ). Однако выделение этих водных масс возможно с вероятностью лишь 56 %. Поэтому их локализация указана на рисунках штри-

ховой кривой. Малая надежность идентификации последних двух водных масс объясняется прежде всего отсутствием исходных данных наблюдений. Вместе с тем применяемая методика классификации структуры вод демонстрирует эффективность подхода, когда обособленность водной массы объективно выделяется при наличии 10—14 векторов наблюдений.

На основе полученных результатов можно построить обобщенную схему пространственной структуры вод в море Беллинсгаузена (см. рис. 5.16). Согласно с результатами анализа структуры вод Южного океана в целом [162], данная схема существенно уточняет пространственную картину применительно к морю Беллинсгаузена. Прежде всего отмечено существенное смещение к югу полярной фронтальной зоны, которая вместе с границей проникновения субантарктических вод поднимается в море Беллинсгаузена до 62° — 64° ю. ш. Подобное смещение границ различных климатических зон значительно усиливает меридиональный контраст гидрофизических характеристик в восточной части тихоокеанского сектора Южного океана.

Другая важная особенность полученных результатов состоит в уточнении глубины залегания верхней границы донных вод. Интерпретация результатов статистического диагноза показывает, что в центральной части моря Беллинсгаузена верхняя граница распространения антарктических донных вод составляет 3100—3300 м, тогда как в северной и восточной части моря — 2700—2900 м. Выявленная неоднородность представляется интересной в аспекте дальнейшего исследования проблемы образования донных вод в Южном океане.

Таким образом, проведенный статистический диагноз позволил решить основные поставленные задачи. Интерпретация основных результатов обработки информации демонстрирует достижение необходимой подробности и достоверности в описании структуры вод открытой части моря Беллинсгаузена.

Следует, однако, подчеркнуть, что полученные выводы не распространяются на шельфовую зону моря. Этот район крайне мало описан в исходных данных. Большинство массивов наблюдений для этого района содержит только данные о температуре и солености. Сложность применения классического T, S -анализа и отсутствие дополнительных материалов гидрохимических наблюдений не позволяют пока включить прибрежный район в общую статистическую модель структуры вод моря Беллинсгаузена.

Содержательность и физическая верифицируемость результатов статистического диагноза обосновывает эффективность не только полученной модели, но и перспективность применения подобного подхода для исследования структуры вод в других регионах Южного океана.

**РЕЗУЛЬТАТЫ СТАТИСТИЧЕСКОГО ПРОГНОЗИРОВАНИЯ
ОКЕАНОЛОГИЧЕСКИХ ПРОЦЕССОВ**

Многочисленные эксперименты по методическому прогнозированию обобщенных параметров океанологических процессов позволяют подойти к обоснованию методологии диагноза и прогноза как к единой задаче. Приводимые ниже примеры реализуют простую идею — использование одномерных статистических моделей для прогноза обобщенных параметров процесса. Она положена в основу прогноза температуры воды поверхностного слоя Северной Атлантики, ледовитости Баренцева моря и биомассы зоопланктона Баренцева моря.

Фактически при корректно проведенном диагнозе процесса и выделении устойчивых во времени обобщенных параметров важно подобрать из класса линейных статистических моделей оптимальную, которая позволяет с наименьшей ошибкой экстраполировать основные параметры не более чем на два шага вперед по времени. Этого оказывается достаточно как с методической (прогнозы гидрометеорологических процессов обладают не высокой предсказуемостью [146, 151]), так и с практической точки зрения: увеличение заблаговременности происходит не за счет свойств модели, а в связи с выявляемой устойчивостью базисных или естественных ортогональных функций. Например, главные компоненты характеризуют среднегодовое значение процесса, базисные функции — характер сезонной изменчивости. Тогда окончательным прогнозом будут восстановленные значения произведения главной компоненты на соответствующие базисные функции процесса.

Несмотря на относительную простоту идеи соединения диагностических результатов со статистическими моделями, ее реализация была в определенной степени затруднена сложностью математического аппарата, недостаточностью исходной океанологической информации и отсутствием разработанных автоматизированных систем.

6.1. Прогнозирование тепловых полей Северной Атлантики

Важными причинами, по которым ТПО (или аномалия ТПО—АТПО) оказывается в центре внимания исследователей и прогнозистов-практиков, являются ее значимость как энергетической ха-

рактеристики, определяющей эволюцию океанологических и биологических процессов, доступность наблюдений, относительно высокая точность измерений ТПО и в связи с этим возможность применения статистического анализа, в том числе и многомерного, к накопленным за продолжительное время данным.

В этом параграфе представлены результаты прогнозирования ТПО по различным регионам Северной Атлантики в следующей последовательности:

- а) анализ статистической структуры исходных данных;
- б) прогнозирование временных рядов и верификация;
- в) анализ крупномасштабной структуры ТПО Северной Атлантики, выделение однородных районов методами МСА;
- г) прогнозирование значимых главных компонент и общих факторов, восстановление прогностических значений ТПО по пространству и их верификация.

При этом мы обращаем особое внимание на анализ структуры исходных данных, которая во многом определяет потенциальную предсказуемость ТПО, а также на методический аспект — верификацию различных статистико-вероятностных методов прогноза ТПО, позволяющий выявить принципиальные источники ошибок прогнозов — ошибки анализа данных и ошибки формулировки модели.

Полное представление о случайном процессе можно получить, зная его функцию распределения вероятностей, которая задается двумя первыми статистическими моментами — оценками математического ожидания и дисперсией процесса. В большинстве практических случаев при отсутствии трендов и представлении величин в виде аномалий (отклонений от средних) можно считать, что оценка математического ожидания равна нулю. Тогда оценка дисперсии процесса является достаточной характеристикой для описания функции распределения вероятностей.

На рис. 5.1 представлены основные статистические параметры рядов АТПО. Неоднородная пространственно-временная структура дисперсий и спектральной плотности указывает на разномасштабные процессы формирования теплового режима океана. В целом выделяются районы океана с примерно одинаковой изменчивостью и видом функции спектральной плотности, которая близка к «красному шуму». Выделяются районы и с экстремальными изменениями характеристик, и с вторичными значимыми пиками спектральных значений.

Результаты анализа рядов (см. рис. 5.1) показывают, что выборочные совокупности данных АТПО в своем большинстве подчиняются закону нормального распределения в районах Норвежского моря. При этом асимметрия по абсолютному значению достигает $0,3^\circ\text{C}$. Это значение является, на наш взгляд, характерной ошибкой наблюдений и допустимо для использования в прогностических схемах, поскольку не выходит за пределы погрешности прогнозов по критериям $0,674\sigma$ и $0,8\sigma$. Следовательно, для анализа и прогноза можно применять основные корреляционные

соотношения параметрического анализа и, в частности, коэффициенты корреляции Пирсона.

На устойчивость первых статистических моментов в пределах допустимой погрешности указывает существование линейных трендов со значениями, не превышающими $0,2^\circ\text{C}$ (рис. 6.1). Проверка

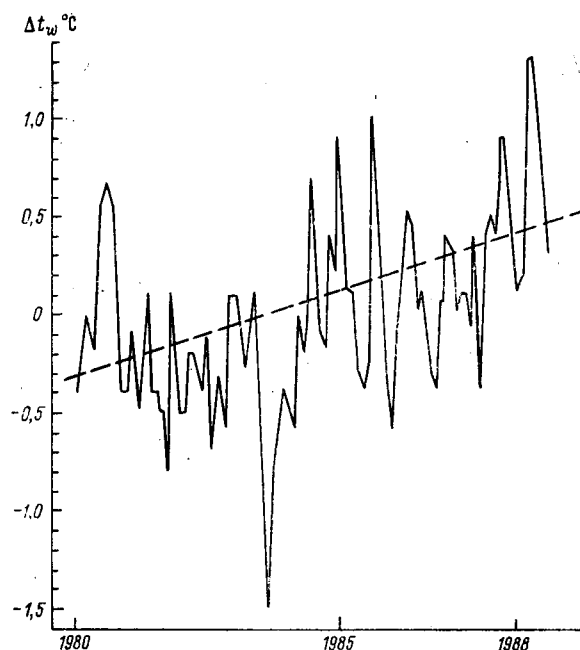


Рис. 6.1. Временной ход среднемесячных аномалий ТПО в Норвежском море в районе судна погоды М.
Линейный тренд показан штриховой линией.

адекватности модели исходным данным ведется, как правило, с помощью первых статистических моментов, таких как дисперсия и коэффициент корреляции, при условии равенства нулю оценки математического ожидания [269]. Являясь важными характеристиками статистического сравнения прогностических и фактических рядов, эти оценки не исчерпывают всех вопросов при проверке модели на адекватность данным. Одним из принципиальных подходов, по-видимому, следует считать верификацию по тем критериям, которые не входят как параметры в модель. Поэтому блок верификации (см. рис. В.5) дополнен процедурами одно- и двумерных функций распределения ошибок прогнозирования [44].

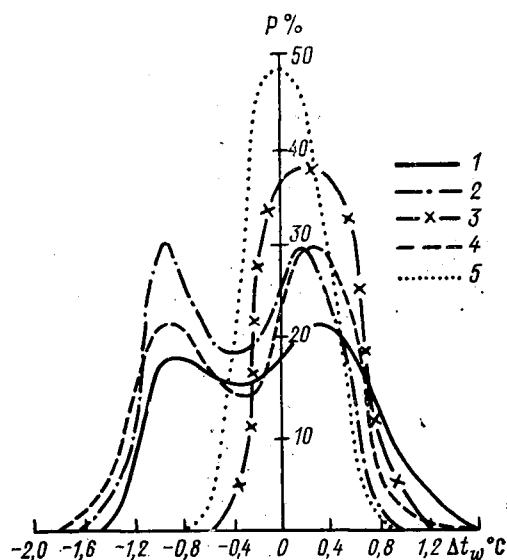
Важным аргументом в пользу применения подобной процедуры служит тот факт, что эмпирические гистограммы позволяют исследовать на устойчивость внутреннюю структуру рядов ошибок прогнозирования в их соответствии с наблюдаемой изменчи-

востью исходных данных. Это принципиально важно при предсказании характерных структурных особенностей данных, фиксируемых в виде резких выбросов, экстремумов, значительных градиентов и т. п.

В этой связи пример, приведенный на рис. 6.2, является весьма показательным. Изображенная функция АТПО по Кольскому меридиану является бимодальной функцией. Она, как видно, не полностью воспроизводится отдельными методами, например методом Фурье-экстраполяции [44]. Лучше она воспроизводится динамико-стохастической моделью (см. п. 4.1), однако и здесь существуют различия, особенно в отрицательной области. Оп-

Рис. 6.2. Эмпирические гistogramмы рядов аномалий среднемесячной температуры воды в слое 0—200 м по Кольскому меридиану.

1 — фактическая, 2 — спрогнозированная методом Фурье, 3 — динамико-статистическим методом, 4 — динамико-стохастическим методом, 5 — методом регрессионной комплексации.



тимальное воспроизведение эмпирической функции вероятностей дает результирующий прогноз, полученный методом статистической комплексации 10 методов (см. п. 4.3).

Верификация различных методов прогноза позволяет более обстоятельно проводить диагноз прогнозируемого процесса. Из анализа рис. 6.2 можно сделать предварительный вывод о природе двух мод функции распределения вероятностей АТПО. Мода с максимумом в положительной области (градация 0—0,4 °C) указывает на существование квазистационарного колебания в АТПО, которое моделируется только методом Фурье суперпозицией десяти гармоник. Это колебание приводит к смещению оценки «норма» приблизительно на 0,2 °C.

Второй пик функции распределения вероятностей АТПО, обусловленный наличием второй моды в области отрицательных значений с максимумом в градации —1,2... —0,8 °C, по своей при-

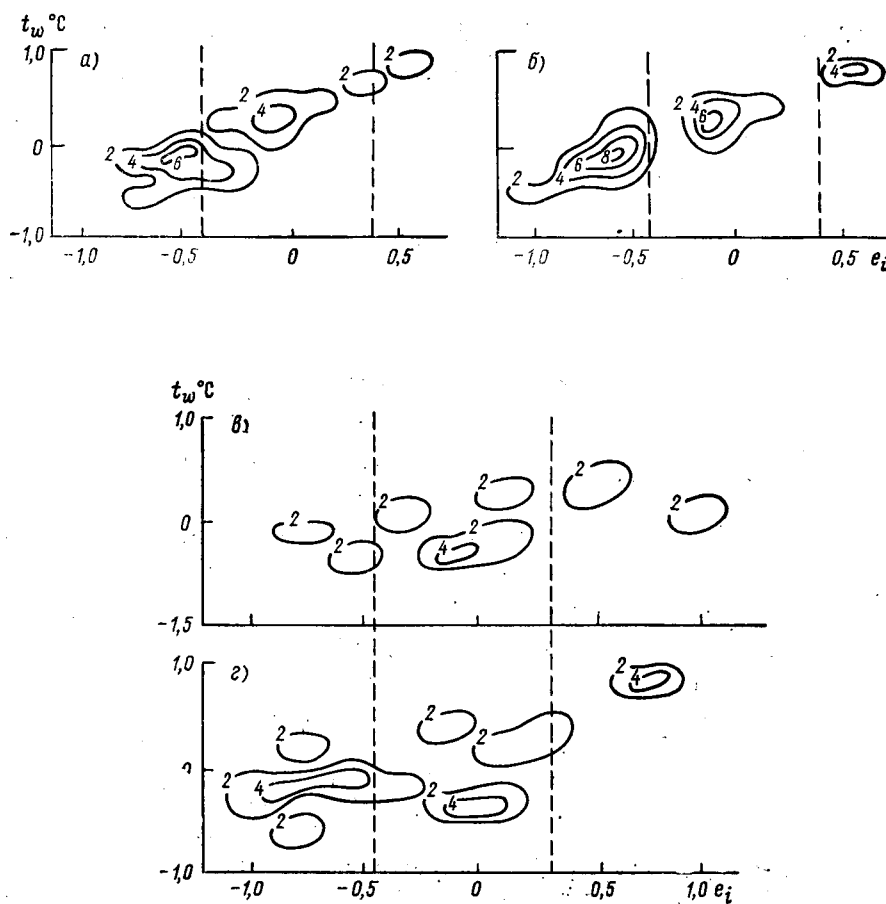
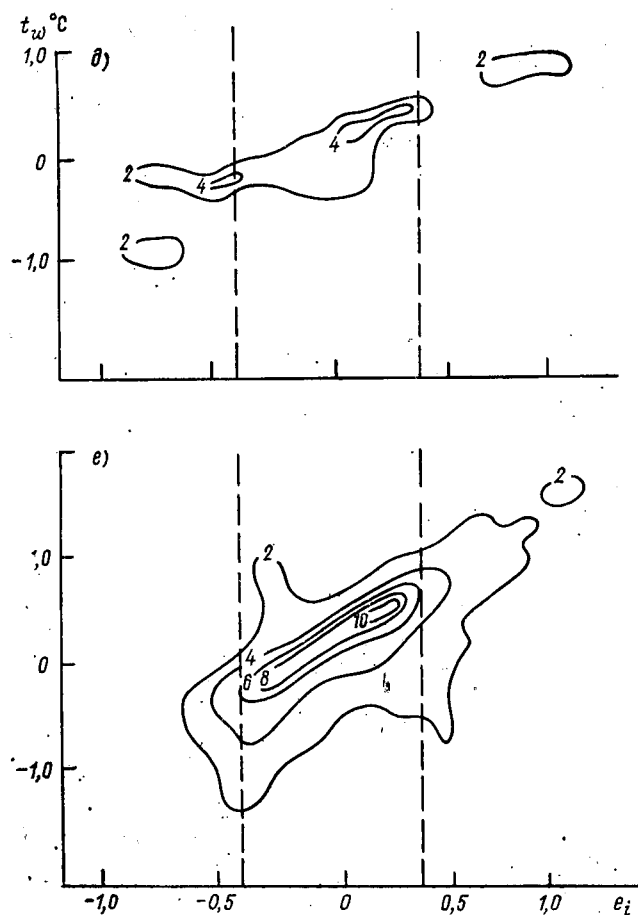


Рис. 6.3. Двумерные гистограммы ошибок e_i прогнозирования аномалий
 а — метод авторегрессии 2-го порядка, б — байесовский, в — динамико-статистический.

роде является случайной (стохастической) компонентой, что моделируется лишь стохастической моделью.

Таким образом, первая мода отражает детерминированную компоненту процесса, вторая — стохастическую. Обе они, а также функция распределения вероятностей процесса в целом (см. рис. 6.2) наилучшим образом воспроизводятся моделью статистической оптимизации всех методов (модель регрессии). Она же является оптимальной в смысле минимизации среднеквадратической ошибки прогнозирования в рамках теории линейных стационарных моделей. Поэтому в среднем двумерная функция распределения ошибок прогнозирования моделью регрессии должна иметь ярко выраженный максимум в интервале допустимой по-



среднемесячной ТПО в районе судна погоды М различными методами.
2 — динамико-стохастический, 6 — Фурье, e — метод регрессионной комплексации.

грешности прогнозирования ($-0,4; 0,4^{\circ}\text{C}$) в отличие от остальных методов. Этот результат демонстрирует рис. 6.3.

Видно, что многие методы имеют вторичные максимумы гистограмм, которые генерируются самими моделями, в то время как закон распределения АТПО в районе судна погоды М является нормальным. Очевидно, что некоторые модели (без адаптации) настроены на «свои процессы», близкие к реальному, о чем говорилось в п. 4.1.

Отметим, что сравнительный анализ всей совокупности полученных критериев статистической предсказуемости позволяет прогнозику получать наиболее полное представление об эффективности применяемых им статистико-вероятностных методов прогноза.

Кроме того, на основе вероятностных методов верификации создается возможность оценки эффективности предсказания крупных аномалий.

Для корректного выделения однородных или однотипных пространственных и временных структур с целью прогноза нами был проведен анализ процессов с помощью методов МСА: методов главных компонент (МГК) и факторного анализа (МФА) (см. п. 5.1).

Прогностическое содержание результатов анализа состоит в следующем.

1. Выделяемые районы указывают на границы изменчивости ТПО, внутри которых основные колебания в виде собственных векторов являются квазистационарными в статистическом смысле и их можно использовать в прогностических моделях. При этом модель прогноза становится многомерной.

2. Каждая из выделенных областей обладает своей степенью взаимосвязанности поля ТПО (матрицы корреляций существенно различны), что позволяет настроить параметры прогностической модели более точно на конкретный тип процесса. Использование одной и той же модели для всего региона Северной Атлантики, как мы уже показали выше, приведет к усреднению оправдываемости и уменьшению эффективности прогнозирования.

3. Включение главных компонент, рассчитанных по естественным ортогональным функциям (е.о.ф.) пространственных или временных координат, позволяет построить объективную многомерную модель, учитывающую влияние адвективных и косвенно метеорологических факторов, как это предлагается в работе [225].

Обратимся к диагнозу тепловых полей Северной Атлантики методами МСА. Объединение (см. рис. 5.2—5.5) трех районов в один позволило оценить влияние пространственных масштабов на основные черты рельефа теплового режима вод Северной Атлантики. В целом структура сохранила свои отличительные особенности: в поле первого собственного вектора четко прослеживается зональность в распределении АТПО, причем зона системы Гольфстрим — Северо-Атлантическое течение выделяется наиболее ярко и имеет максимальные отрицательные значения. Экваториальная и тропическая части Северной Атлантики заняты положительными значениями. Такая картина пространственного распределения первой моды АТПО, очевидно, обусловлена инсоляцией (см. рис. 5.2а).

Поле второго собственного вектора (см. рис. 5.2б) также отражает зональное перераспределение тепла. При этом, однако, разделяются по знаку зоны экваториальной и тропической частей Северной Атлантики. С большим весом, чем в поле первого собственного вектора, выступают районы, подверженные влиянию Гольфстрима, Лабрадорского, Ирмингера и Восточно-Гренландского течений. Таким образом, выявляется роль крупномасштабного перераспределения АТПО посредством адвекции.

Поле третьего собственного вектора отражает меридиональные процессы, которые связаны, скорее всего, с динамикой вод Северной Атлантики и Норвежского моря — это ньюфаундлендский регион, восточная часть тропической Атлантики и центральная часть Норвежского моря.

Наконец, четвертая компонента позволяет достаточно четко выделить зону Гольфстрим — Северо-Атлантическое течение, распространяющуюся до 35° з.д., и зону апвеллинга — в восточной части тропической Атлантики. Следует также отметить очаг в центральной части Северной Атлантики, связанный, возможно, с динамической структурой крупномасштабного антициклонического круговорота в этом районе.

Несмотря на невысокие значения описываемой дисперсии в полях первых компонент, обусловленные большим порядком корреляционной матрицы (120×120), следует сделать вывод о значительной устойчивости и однородности структуры главных черт термического рельефа в поле АТПО, особенно по району Гольфстрим — Северо-Атлантическое течение при переходе от меньших пространственных масштабов отдельных районов к более крупным, охватывающим практически всю акваторию Северной Атлантики и Норвежское море.

Полученные результаты статистического анализа полей АТПО позволяют сформулировать следующие выводы:

1) структура поля АТПО существенно зональна и обусловлена поступлением солнечной радиации,

2) перераспределение тепла в основном осуществляется крупномасштабными течениями,

3) меридиональность в распределении АТПО в основном проявляется в зонах взаимодействия крупномасштабных течений.

По значительным объемам исходных данных удалось фактически классифицировать Северную Атлантику по характеру пространственной неоднородности АТПО. Этот результат, несомненно, является определяющим при выборе модели прогноза АТПО и влияет на ее параметры предсказуемости.

Таким образом, статистическая структура ТПО существенно зависит от пространственно-временных масштабов наблюдений и, несомненно, влияет на предсказуемость как 1-го, так и 2-го рода. Отсюда становится понятно, почему модель Франкиньяла — Хасельмана [250], которая в качестве переменных параметров содержит составляющие уравнения бюджета тепла деятельного слоя океана, явилась достаточно успешной как для диагноза, так и для прогноза тепловых условий в океане. И в первую очередь, это связано с устойчивостью в статистическом смысле влияющих факторов на АТПО и с однородностью самих регионов, где роль того или иного фактора определена.

Другим важным результатом применения стохастической модели [234] явилось увеличение успешности прогнозирования в районе Гольфстрим — Северо-Атлантическое течение [136], где ошибка прогноза уменьшается до 50 %, в то время как по региону

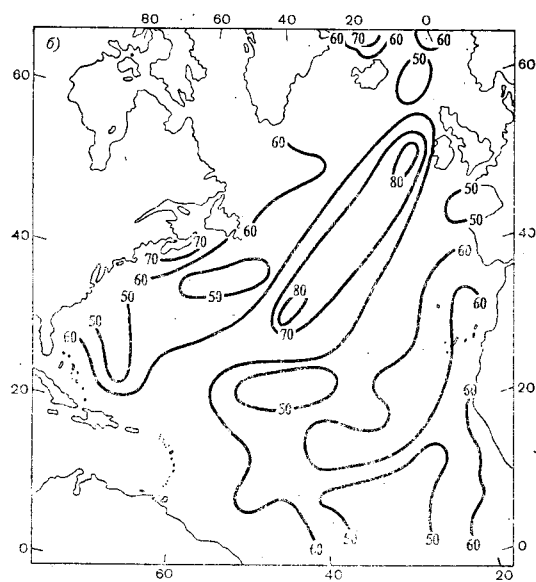
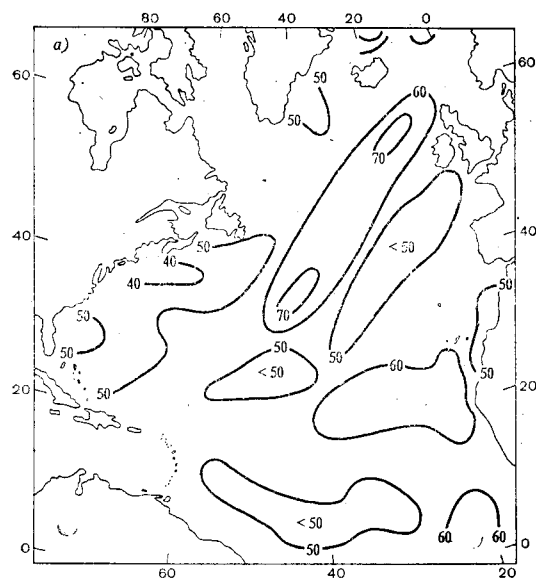
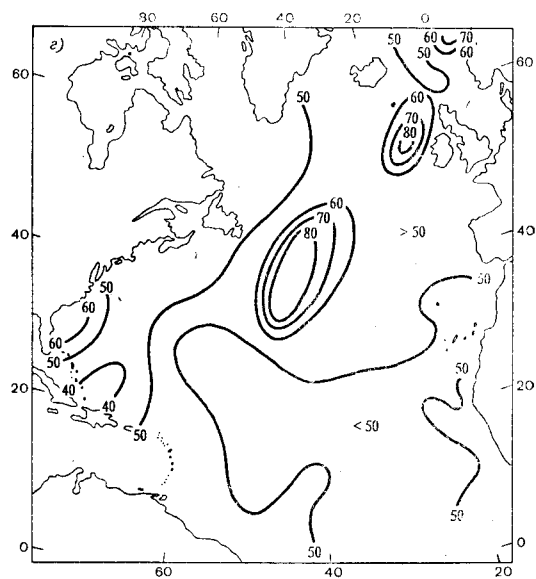
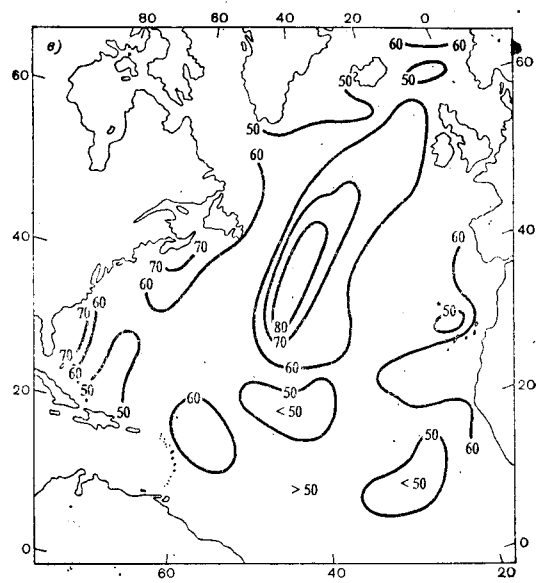


Рис. 6.4. Оправдываемость (%) прогнозов аномалий среднемесячной ТПО в различных
а — байесовский метод, *б* — метод авторегрессии 2-го порядка,



Северной Атлантике по критерию $0,674 \sigma$ для методов.

a — динамико-стохастический, *г* — инерционный,

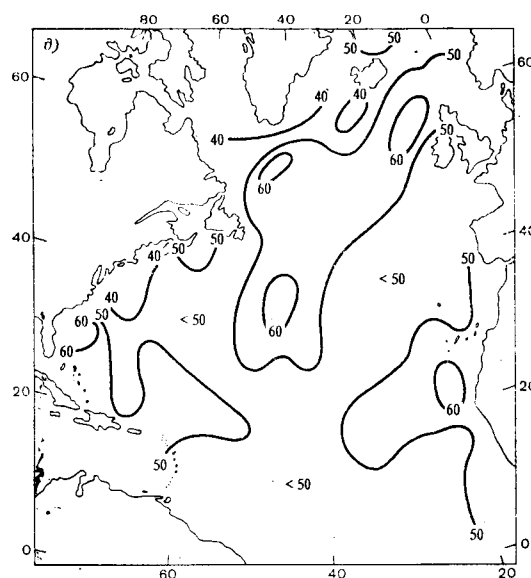


Рис. 6.4. Оправдываемость (%) прогнозов аномалий среднемесячной ТПО в Северной Атлантике по критерию $0,674\sigma$ для различных методов.

d — динамико-статистический метод.

Северной Атлантики в целом дисперсия шума в среднем составляет 70 %, что соответствует 30 %-ному уровню предсказуемости.

Из проведенного статистического анализа и результатов работ [44, 151] следует, что увеличение предсказуемости связано с качеством проводимого анализа, т. е. модели наиболее успешно прогнозировали те процессы и в тех регионах, на которые они были наилучшим образом настроены. Ясно (см. рис. 5.2), что для экваториальной и тропической Атлантики роль перечисленных факторов иная, чем в зоне действия Гольфстрима, т. е. параметры модели для этого региона должны существенно отличаться и поэтому оправдываемость по модели [234] здесь хуже.

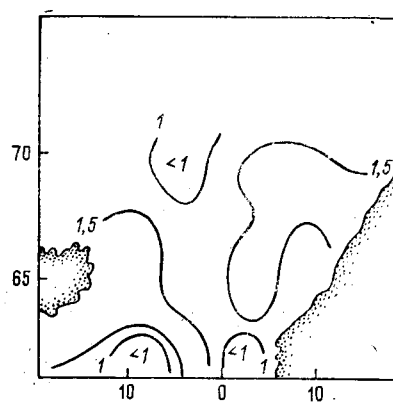
Исходя именно из этих соображений мы применяли различные статистико-вероятностные модели с тем, чтобы определить их возможности в различных регионах Северной Атлантики и Норвежского моря (рис. 6.4). Результаты прогнозирования на независимом ряде наблюдений среднемесячных АТПО с 1981 по 1985 гг. получены впервые по такой значительной акватории, как Северная Атлантика (162 пятиградусные трапеции). Оправдываемость прогнозов колеблется от 50 до 80 %, что составляет 50—70 % (см. рис. 6.4) дисперсии шума.

Интересно отметить уменьшение ошибки прогноза в зоне действия системы Гольфстрим — Северо-Атлантическое течение при

прогнозировании статистико-вероятностными методами первого и второго порядков. Наилучшие результаты получены по данной зоне моделью АР-2.

Сравнение трех методов с результатами прогнозирования с помощью инерционного прогноза показывает их преимущество в среднем на 10—20 %. Повышение оправдываемости инерционного прогноза отмечается в центральной части Северной Атлантики, где она достигает 80 %. Можно предположить, что это связано с сохранением аномалии в антициклоническом круговороте.

Рис. 6.5. Пределы статистической предсказуемости в месяцах рядов аномалий среднемесячной ТПО в Норвежской ЭАЗО по динамико-стохастической модели ($d = 0,9$).



То же можно отметить и в северной части Норвежского моря, где минимальные значения относительной дисперсии ошибки прогноза локализованы в зоне действия квазистационарного антициклонического круговорота, центр которого, как известно [160], расположен в окрестности точки с координатами $\lambda = 2-3^\circ$ в.д. и $\varphi = 68-69^\circ$ с.ш.

Другой важной характеристикой предсказуемости является предел заблаговременности прогноза при заданном критерии точности ϵ . Из соотношений (4.8), (4.9) нетрудно рассчитать τ и построить пространственную картину максимальной заблаговременности прогноза при $d \leq 0,9$ (рис. 6.5). Такой расчет был выполнен для Норвежского моря (26 пятиградусных трапеций).

Он показал, что максимальный предел предсказуемости для среднемесячных АТПО составляет 1,8 мес для зоны действия антициклонического круговорота, Норвежского течения и Фарерской зоны. В среднем же заблаговременность прогноза на один шаг вперед по времени можно считать вполне надежной для прогнозирования статистико-вероятностными методами для данного региона океана.

В северной Атлантике этот предел возрастает до двух-трех шагов, особенно в зоне действия Гольфстрима и Северо-Атлантического течения. В экваториальной и тропической частях максимальный предел статистической предсказуемости не превышает одного

шага вперед по времени. Одной из причин этого является существенная нестационарность рядов АТПО в данном регионе океана.

Повышение эффективности предсказания связывается многими авторами [65, 146, 151] с уменьшением дисперсии непредсказуемой части процесса, т. е. с уменьшением роли шумовой составляющей. Это достигается различными путями: с помощью фильтрации или увеличением пространственно-временных масштабов осреднения.

Естественно, что как использование фильтрации, так и увеличение пространственно-временных масштабов осреднения предсказуемой характеристики дает возможность подобрать наиболее подходящую модель. Однако с практической точки зрения такой подход оказывается малоэффективным, так как не позволяет детально описать, а следовательно, и прогнозировать пространственные или временные особенности процесса.

Изучение и прогнозирование крупномасштабных колебаний океанологических процессов предполагает существование небольшого числа влияющих факторов, которые можно проследить, объяснить и спрогнозировать. Поэтому представление полей или рядов океанологических характеристик в виде суммы первых главных компонент, с одной стороны, отфильтровывает шумовую составляющую процесса, с другой — представляет процесс в виде суперпозиции крупномасштабных колебаний. Моделирование таких колебаний статистико-вероятностными или физико-статистическими методами более эффективно по сравнению с предсказанием непосредственно самих характеристик данными методами, так как в последнем случае ошибки в данных сохраняются. К тому же использование только одномерных статистико-вероятностных моделей не может быть достаточным из-за многочисленных прямых и обратных пространственно-временных связей [120, 142, 146]. В этих же работах делается вывод об увеличении эффективности прогнозирования на 10—20 % при использовании многомерных методов прогноза.

Полученная оценка улучшения предсказуемости обусловлена ролью адвективных факторов в системе пространственных связей ТПО. Проявление роли этого фактора особенно очевидно следует из результатов опытного прогнозирования (рис. 6.6). Учет пространственных связей, которые приводят к повышению эффективности статистического прогноза, можно проводить различными способами: включать в модель АР связи с соседними точками сеточной области, как это предлагается в [143], или прогнозировать первые значимые и описывающие более 80 % дисперсии главные компоненты, рассчитываемые по е.о.ф. горизонтальных координат [251]. Второй способ представляется авторам более привлекательным, поскольку позволяет избежать субъективного выбора числа влияющих точек.

Примером применения многомерной модели АР-1 и байесовской экстраполяции первого порядка могут служить результаты независимых прогнозов по акватории Норвежского моря с исполь-

зованием факторного разложения АТПО (рис. 6.7). В силу того, что разложение выполнялось по площади (матрица исходных данных имела размерность 26×120), факторные нагрузки учитывали взаимосвязь полей во времени.

Результаты прогнозов, выполненные по одномерной и многомерной модели АР-1, относительно невысоки: оправдываемость по критерию $0,8 \sigma$ по площади составила 58 и 65 % соответственно.

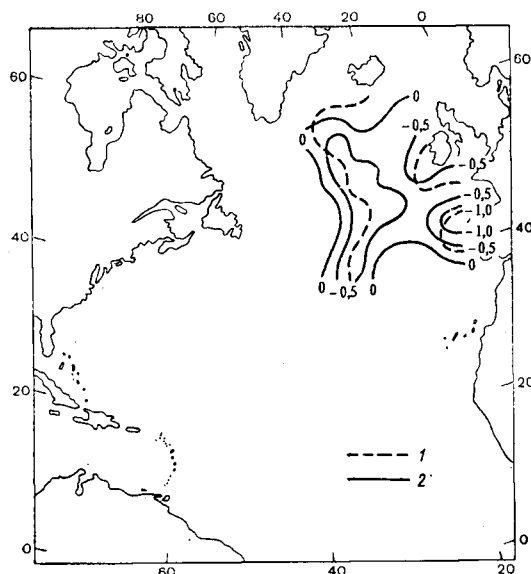


Рис. 6.6. Независимый прогноз поля аномалий среднемесячной ТПО в северо-восточной Атлантике на июль 1981 г. с заблаговременностью 6 мес.

1 — прогностические значения, 2 — фактические.

Несколько выше оправдываемость по моделям байесовской экстраполяции — 65 и 73 % соответственно. Невысокая оправдываемость во многом определялась аномальными условиями, сложившимися в феврале 1984 г. в западной части Норвежского моря: ослабление интенсивности Восточно-Гренландского течения (ТПО повысилась на 2°C). В то же время усилилось Северо-Атлантическое течение: проникновение его вод вызвало увеличение ТПО в Норвежском море на $0,8^\circ\text{C}$.

Важно отметить, что данный конкретный результат отражает известную связь между тепловым состоянием вод основных течений Норвежского моря и его биопродуктивностью: потепление Арктических вод способствует продвижению на север и восток вод Северо-Атлантического течения и вместе с ним теплолюбивых гидробионтов.

Более подробно эта идея развита ниже при анализе и прогнозе биомасс зоопланктона Норвежского моря. В целом же наиболее благоприятными для прогнозирования оказались те районы Норвежского моря, где $d \leq 74\%$ (см. рис. 6.5).

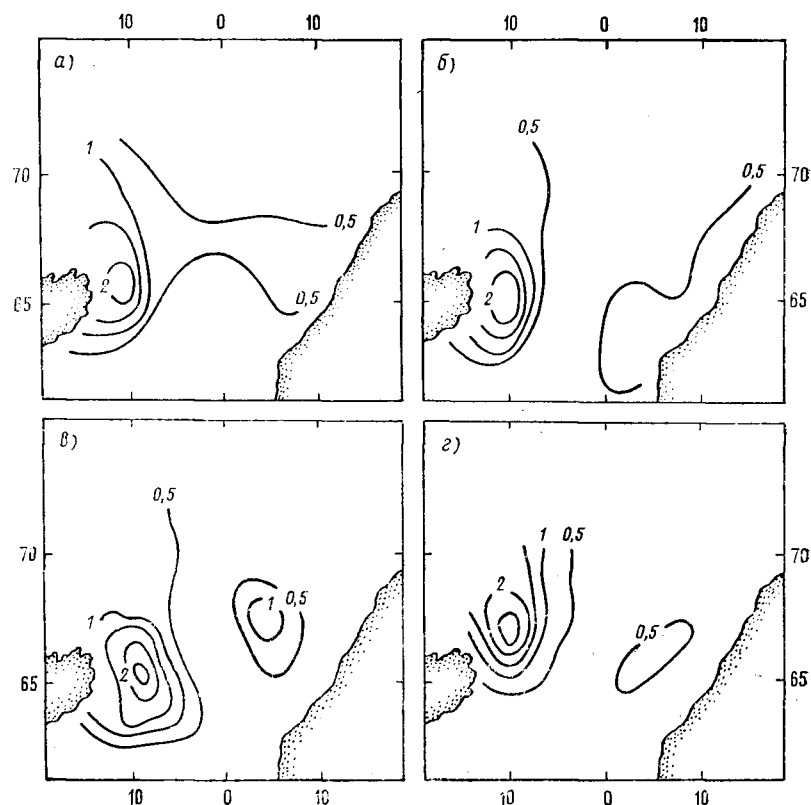


Рис. 6.7. Распределение ошибок прогнозов аномалий среднесесячных ТПО в Норвежском море на февраль 1984 г. различными методами. а — одномерный байесовский метод, б — многомерный байесовский, в — одномерный авторегрессионный, г — многомерный авторегрессионный.

Относительно невысокая предсказуемость элементов климатической системы в рамках линейных одномерных моделей теоретически обоснована в работах [35, 65, 142, 146, 151]. Для атмосферы она составляет не более одного-двух шагов [146], для океана четыре—шесть шагов [151] в зависимости от характеристики климатической системы, дискретности наблюдений и пространственных масштабов осреднения.

Таким образом, увеличение заблаговременности путем экстраполяции исходных рядов более чем на два-три шага по времени, по-видимому, бесперспективно. С практической точки зрения это более очевидно: качество прогноза обычно связывают с невыходом ошибки прогнозирования за $0,674\sigma$ или $0,8\sigma$, что соответствует

приблизительно 0,55 и 0,74 дисперсии процесса. При этом коэффициент корреляции должен быть выше 0,8.

Такой высокий уровень внутрирядной связанности обычно обеспечивается лишь со сдвигом на один, максимум — на два шага вперед [44]. Именно это является основной причиной малой популярности методов экстраполяции временных рядов. Практическая полезность прогностической информации (с заблаговременностью один-два шага вперед) невелика.

Применение аппарата многомерного статистического анализа, по мнению авторов, дает возможность не только повысить эффективность предсказания, но и увеличить заблаговременность прогноза океанологических процессов статистико-вероятностными методами.

Результаты анализа временной структуры рядов первых главных компонент океанологических характеристик a_i [181, 250] указывает на высокую внутрирядную связанность $R_{a_i}(\tau) \geq 0,8$, устойчивость автокорреляционных функций и состоятельность первых статистических моментов: оценок математического ожидания и дисперсии процессов. Для определения статистической устойчивости оценок дисперсий ($\hat{\sigma}^2$) главных компонент, рассчитанных как по среднемесячным АТПО, так и по среднесезонным АТПО, был проведен сравнительный анализ по рядам, сдвинутым относительно друг друга на 12 месяцев и на 4 сезона. Средние значения разности $\Delta\sigma$ по районам 22 пятиградусных трапеций и 10 пятиградусных трапеций оказались равными 0,04 и 0,03° для среднемесячных и среднесезонных значений АТПО соответственно, что указывает на

состоятельность оценок (при средних $\hat{\sigma}_{22} = 0,54^\circ\text{C}$ и $\hat{\sigma}_{10} = 0,47^\circ\text{C}$). Анализ графиков автокорреляционных функций главных компонент (рис. 6.8) позволяет оценить максимальный интервал корреляции, который для всей области северо-восточной Атлантики и Норвежского моря (32 пятиградусные трапеции) составил 7 мес для $\tau(a_1)$, уменьшаясь до 3 мес для $\tau(a_5)$. Фактически он указывает на максимальный предел предсказуемости прогноза — от 7 до 3 мес.

Аналогичный анализ был проведен для среднесезонных значений АТПО. Установлено, что максимальный предел предсказуемости составил 1—2 сезона. Результаты оправдываемости прогнозов приведены в табл. 6.1. Обращает на себя внимание высокая оправдываемость прогнозов с заблаговременностью 1—2 месяца и 1 сезон (более 80 %).

Данный подход реализует задачу прогноза АТПО по значительным акваториям с помощью многомерного статического подхода, в котором учитываются пространственные связи.

Здесь наиболее важным остается вопрос об устойчивости е. о. ф. пространственных координат, так как при достаточно высоких $R_{a_i}(\tau)$ прогностические значения ТПО могут существенно изменяться в случае их неустойчивости. Известно [115], что е. о. ф.

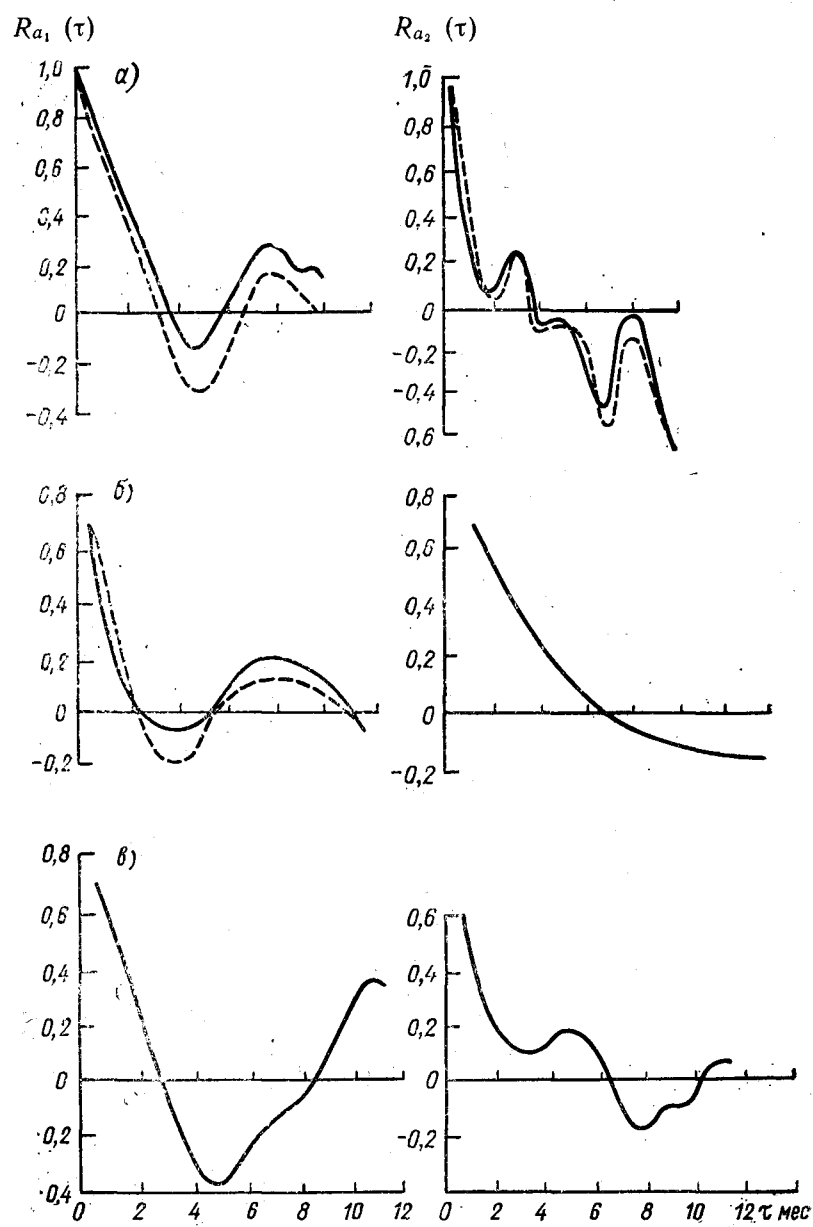


Рис. 6.8. Характерные автокорреляционные функции первых двух главных компонент разложения рядов аномалий среднемесячной ТПО в Норвежском море (а), северо-восточной Атлантике (б), районе Гольфстрима (в).

Таблица 6.1

**Эффективность прогнозирования среднемесячных
и среднесезонных АТПО по 32 пятиградусным
трапециям в Северной Атлантике
(независимый ряд с января по август 1981 г.)**

Заблаговременность	Оправдываемость, %	Эффективность, %
1 мес	88	53
2 мес	69	22
3 мес	66	9
4 мес	67	12
5 мес	84	15
6 мес	69	13
7 мес	88	2
1 сезон (I—II)	81	46
2 сезона (III—V)	78	7
3 сезона (VI—VIII)	72	3

рассчитываются с точностью до знака. Поэтому смена знака е. о. ф. при сдвиге ряда на один и более шагов по времени может привести к неправильному учету вклада прогнозируемой компоненты.

Это особенно важно иметь в виду, если в прогностической модели вида (4.1) каждый следующий прогноз рассматривается как фактическое значение и процедура прогноза повторяется на 1-м, 2-м и последующих шагах. В этой связи интересны результаты, приведенные в табл. 6.2, из которой следует, что устойчивость е. о. ф. достаточно высока: при сдвиге рядов до 60 значений (месяцев) относительно друг друга основные черты рельефа АТПО в поле первых трех е. о. ф. по весьма изменчивому региону Гольфстрима сохраняются.

Таблица 6.2

**Оценка вклада (I) и устойчивости (II) е. о. ф. среднемесячных АТПО,
рассчитанных по 18 пятиградусным трапециям для района Гольфстрима
(Северная Атлантика)**

I			II					
i	$\sum_{i=1}^k \lambda_i / \sum_{i=1}^{18} \lambda_i \%$		1976—1981	1957—1971				
	1957—1971	1976—1981		X ₁	X ₂	X ₃	X ₄	X ₅
1	33	24	X ₁	0,66	—0,53	0,08	—0,12	0,03
2	54	42	X ₂	0,56	0,43	—0,02	0,07	0,46
3	68	55	X ₃	0,24	0,48	0,20	0,17	—0,48
4	74	66	X ₄	0,03	0,06	—0,71	—0,19	0,02
5	80	75	X ₅	0,18	—0,10	0,23	—0,32	—0,25

Таким образом, эффективность прогнозирования с заблаговременностью 1—2 шага по времени по значительным акваториям определяется в основном свойствами $R_{a_i}(\tau)$ при сохраняющихся пространственных крупномасштабных особенностях ТПО. Результаты прогнозирования по данной схеме приведены в табл. 6.3.

Таблица 6.3

**Эффективность прогнозирования среднемесячных АТПО
по 22 пятиградусным трапециям
и среднесезонных АТПО по 10 трапециям
в Северной Атлантике (независимый ряд
с января по июнь 1982 г.)**

Заблаговременность	Оправдыва- емость, %	Эффектив- ность, %
1 мес	73	28
2 мес	73	33
3 мес	64	14
4 мес	77	9
5 мес	77	—3
6 мес	68	0
1 сезон (I—III)	80	20
2 сезона (IV—VI)	80	20

В качестве модели прогноза главных компонент была использована модель АР порядка N , который в каждом конкретном случае выбирался оптимальным по минимуму дисперсии ошибки прогноза [181].

Анализ эффективности прогнозирования ТПО по одномерным и многомерным моделям позволил выявить статистически (подразумевается и физически) значимые связи как по пространству, так и по времени. Однако во всех случаях мы ограничены невысокими пределами предсказуемости — один-два шага вперед по времени. Увеличение этого предела может быть связано, как мы уже говорили, с увеличением пространственного и особенно временного осреднения. Однако при этом происходит разрыв по дискретности.

Так, при прогнозировании среднегодовых АТПО на 1—2 года вперед мы не можем дать прогноз среднемесячных значений по этим же годам, в силу того, что используются при этом лишь ряды среднегодовых АТПО. Это обстоятельство указывает на то, что в многомерных схемах статистико-вероятностных прогнозов учитывается лишь влияние различных пространственных масштабов и не учитывается влияние временных. Совершенно очевидно, что их влияние на регион, по которому дается прогноз, будет различным в зависимости от мезомасштабных, крупномасштабных или глобальных гидрометеорологических процессов.

Учет взаимного влияния различных временных масштабов в стохастических моделях на сегодняшний день практически не

реализуется. Поэтому с целью увеличения предела предсказуемости океанологических характеристик нами предпринята попытка модифицировать данную схему прогноза без привлечения дополнительных параметров (предикторов). Суть ее состоит в получении коэффициентов разложения среднемесячных значений АТПО по е. о. ф. времени для каждого из выбранных квадратов однородной области Северной Атлантики. В этом случае коэффициенты разложения рассчитываются для каждого конкретного года временного ряда, а е. о. ф. соответствуют конкретному месяцу года.

Далее, рассчитывая и анализируя вид автокорреляционных функций коэффициентов разложения, важно выявить значимую корреляционную связь в межгодовом ходе коэффициентов разложения при временном сдвиге в один год (с упреждением в один шаг). Тогда при условии устойчивости е. о. ф. достаточно спрогнозировать коэффициенты разложения одним из статистико-вероятностных методов на один шаг вперед.

Очевидно, что данная схема прогноза, так же как и в случае е. о. ф. пространственных координат, существенно зависит от устойчивости е. о. ф. во времени и не будет приводить к асимптотически ухудшающейся оправдываемости от января к декабрю прогнозируемого года, как в предыдущем методе или одномерных моделях авторегрессии. Существующая ошибка прогноза в этом случае будет распределяться внутри прогнозируемого года и тем самым нивелироваться, повышая оправдываемость прогноза и увеличивая его заблаговременность от 1 до 12 мес.

Прогнозы по вышеприведенной схеме были выполнены для репрезентативных данных АТПО по судам погоды *I*, *J* и *K* с заблаговременностью один год. В табл. 6.4, 6.5 приведены основные

Таблица 6.4

Оценка разложения рядов температуры воды по данным судов погоды *L(I)*, *J*, *K* по е. о. ф. времени

<i>i</i>	λ_i	$\frac{\lambda_i}{\sum_{i=1}^{18} \lambda_i},$ %	$\frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^{18} \lambda_i},$ %	<i>i</i>	λ_i	$\frac{\lambda_i}{\sum_{i=1}^{18} \lambda_i},$ %	$\frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^{18} \lambda_i},$ %
Судно погоды <i>L(I)</i>				Судно погоды <i>J</i>			
1	6,3	52,9	52,9	4	1,2	9,9	79,3
2	2,5	21,0	73,2	5	0,9	7,4	86,7
3	1,1	9,2	83,1	6	0,6	5,4	92,1
4	0,5	4,2	87,3	Судно погоды <i>K</i>			
5	0,5	4,2	91,5	1	5,9	39,0	39,0
6	0,4	3,5	95,0	2	2,0	20,0	59,0
Судно погоды <i>J</i>				3	1,6	17,5	76,5
1	4,6	38,7	38,7	4	0,6	6,8	83,3
2	2,1	17,4	56,1	5	0,6	5,0	88,3
3	1,6	13,3	59,4	6	0,4	4,2	92,5

статистические характеристики главных компонент среднемесячных значений АТПО. Из них следует, что достаточно использовать первые три главные компоненты для прогноза. Их автокорреляционные функции указывают на высокую внутрирядную связь при сдвиге в один год (рис. 6.9). Временной радиус корреляции во многом зависит от структуры поля, куда входит прогнозируемый район [115]: если поле е.о.ф. района вписывается в поле е.о.ф.,

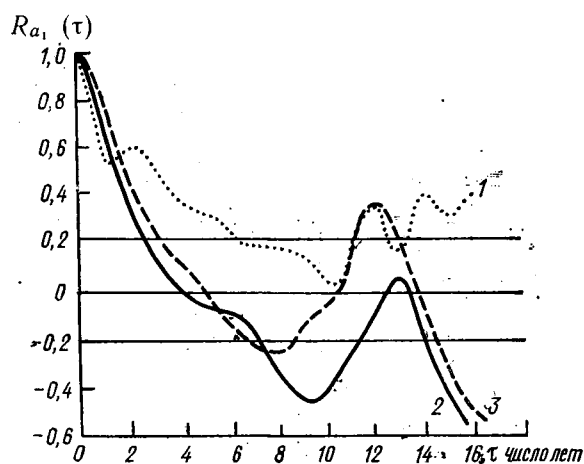


Рис. 6.9. Нормированные автокорреляционные функции первой главной компоненты разложения рядов аномалий среднемесячной ТПО в районах судов погоды L (1), J (2), K (3).

Таблица 6.5

Средние квадратические ошибки расчета автокорреляционной функции среднемесячной температуры воды по данным судов погоды L (1), J , K

a_i	Судно погоды	τ	2σ	r	z	$z-2\sigma$	τ	σ	$\sigma_{a_1}^{L(I)}$	$\sigma_{a_1}^J$	$\sigma_{a_1}^K$
a_1	$L(I)$	1	0,36	0,69	0,85	0,49	1	0,18	0,18	0,11	0,14
		2	0,32	0,37	0,39	0,07	2	0,18	0,18	0,15	0,11
		3	0,36	0,11	0,11	-0,25	3	0,18	0,18	0,17	0,15
	J	1	0,22	0,62	0,72	0,50	4	0,19	0,19	0,19	0,17
		2	0,30	0,40	0,42	0,12	5	0,19	0,19	0,19	0,18
		3	0,31	0,24	0,24	-0,10	6	0,20	0,19	0,19	0,19
	K	1	0,28	0,45	0,43	0,20	7	0,20	0,19	0,19	0,19
		2	0,22	0,64	0,76	0,54	8	0,20	0,17	0,19	0,20
		3	0,30	0,45	0,48	0,18	9	0,21	0,15	0,21	0,20
a_2	$L(I)$	1	0,34	0,26	0,27	-0,07	10	0,21	0,17	0,20	0,21
		2	0,34	0,30	0,31	-0,02	11	0,22	0,20	0,21	0,22
	J	1	0,36	-0,15	0,15	-0,21	12	0,23	0,22	0,22	0,19
		2	0,26	0,16	0,16	-0,10	13	0,24	0,23	0,23	0,23
	K	1	0,31	0,19	0,19	-0,15	14	0,24	0,23	0,23	0,20
		2	0,36	-0,09	0,09	-0,27	15	0,24	0,20	0,19	0,23

рассчитанных по всему региону, то временной радиус корреляции оказывается бóльшим и его следует отнести к районам повышенной информативности. В частности, это относится к районам судов погоды *I* и *K* (см. рис. 6.9), что хорошо согласуется с выводами работы [197]. Важно показать, что выделяемые долговременные колебания не случайны. Для этого были рассчитаны доверительные границы корреляции с 5 %-ным уровнем значимости:

$$z = \frac{1}{2} \ln \frac{1+r}{1-r},$$

где z — параметр Фишера; σ — среднеквадратичная ошибка расчета автокорреляционной функции r ; τ — временной радиус корреляции, который равен нижней границе фишеровского параметра $z - 2\sigma \leq 0$.

Заметим, что при $r = 0,5$ оценка τ становится завышенной. Из табл. 6.5 следует, что максимальным пределом статистической предсказуемости является период от 1 до 3 лет для межгодовых колебаний главных компонент, рассчитанных по среднемесячным АТПО, причем для данных судна погоды *K* он наибольший: для первой главной компоненты — 3 года, а для данных судов погоды *I* и *J* — составляет немногим более 2 лет.

Эти результаты позволили авторам испытать схему долгосрочного прогноза АТПО по районам судов погоды. Основные результаты независимого прогнозирования приведены в табл. 6.6.

Таким образом, проведенное исследование внутригодовой и межгодовой изменчивости АТПО позволяет сделать следующие прогностические выводы.

1. Колебания среднемесячных, среднесезонных и среднегодовых АТПО в Северной Атлантике и Норвежском море удовлетво-

Таблица 6.6

Фактические и спрогнозированные АТПО по данным судов погоды
I, *J*, *K* на 1980 г.

Месяц	<i>I</i>			<i>J</i>			<i>K</i>		
	факт.	прогн.	ошибка	факт.	прогн.	ошибка	факт.	прогн.	ошибка
I	0,1	0,4	0,3	-0,1	0,0	0,1	-0,1	-0,2	0,1
II	0,0	0,4	0,4	0,0	0,0	0,0	-0,4	-0,3	0,1
III	0,0	0,3	0,3	-0,4	0,1	0,5	-0,5	-0,3	0,2
IV	-0,1	0,3	0,4	-0,1	0,2	0,3	-0,6	-0,5	0,1
V	0,2	0,2	0,0	0,7	0,3	0,4	-0,1	-0,5	0,4
VI	0,0	-0,1	0,1	0,0	0,2	0,2	-0,4	-0,7	0,3
VII	-0,2	-0,1	0,1	-0,4	0,0	0,4	-0,4	-0,5	0,1
VIII	-0,3	-0,3	0,0	-0,2	-0,2	0,0	-1,3	-0,6	0,7
IX	0,1	-0,1	0,2	-0,6	0,0	0,6	-0,7	-0,6	0,1
X	-0,4	0,0	0,4	-0,3	0,2	0,5	-0,6	-0,5	0,1
XI	0,0	0,1	0,1	-0,2	0,3	0,5	-0,5	-0,4	-0,1
XII	0,4	0,2	0,2	0,5	0,3	0,2	0,0	-0,1	0,1

рительно описываются статистико-вероятностными моделями 1-го и 2-го порядков.

2. Статистическая структура АТПО существенно различается по акватории Северной Атлантики и Норвежского моря, что приводит к изменениям ошибок прогнозирования с одношаговым упреждением в широких пределах: от 80 % в экваториальной и тропической частях Северной Атлантики и западной части Норвежского моря до 20 % в системе Гольфстрим — Северо-Атлантическое течение.

3. Пределы статистической предсказуемости во многом зависят от масштабов пространственно-временного осреднения и выбранной статистико-вероятностной модели диагноза:

- с увеличением масштабов временного осреднения (сезон, год) предел предсказуемости составляет 2—3 шага,

- с увеличением масштабов пространственного осреднения (экстраполяция главных компонент по акватории, объединяющей несколько пятиградусных трапеций) заблаговременность увеличивается до 4—6 мес,

- прогноз главных компонент по е. о. ф. времени увеличивает предел предсказуемости до 6—8 шагов.

4. Максимальная заблаговременность прогноза АТПО достигается путем статистической оптимизации различных статистико-вероятностных моделей.

6.2. Прогнозирование ледовых условий в Баренцевом море

Важной океанологической характеристикой является также ледовитость. Она играет значительную роль в формировании термического режима морей в силу теплоизолирующих свойств льда: в холодную половину года лед уменьшает выхолаживание, в теплую — препятствует прогреву. Кроме того, ледовые поля существенным образом влияют на биологическую продуктивность моря, так как ледообразование и таяние льда меняют соленость и световой режим верхнего слоя океана. По этим причинам, а также в связи с тем, что ледовитость — индикатор изменений климата, разработка долгосрочных прогнозов ледовых условий приобретает большое прикладное значение.

Для построения и верификации прогностических одномерных и многомерных моделей были использованы ряды наблюдений за ледовитостью Баренцева моря: среднегодовая площадь моря, доля (%) общей площади моря, покрытая льдом, с 1899 по 1981 гг. и среднемесячные значения ледовитости с 1934 по 1989 гг.

Обратимся к результатам независимого прогнозирования среднемесячных значений ледовитости (табл. 6.7) с помощью одномерных статистических моделей. Для выделения детерминированных составляющих процесса на первом этапе исходный ряд был центрирован на среднее значение, затем из него была удалена гармоника с периодом 12 единичных интервалов. Это процедура, как

Таблица 6.7

Эффективность прогнозирования аномалий среднемесячных значений ледовитости Баренцева моря с заблаговременностью 1 мес по ряду с января 1988 г. по декабрь 1989 г.

Метод прогноза	Средняя ошибка	Дисперсия ошибки	Оправдываемость (%) прогнозов			Относительный уровень ошибки прогноза
			Знак	0,67 σ	0,8 σ	
Климатический	-0,729	0,614	8,33	25,00	33,33	1,429
Инерционный	0,008	0,309	91,67	66,67	70,83	0,720
Байесовский	-0,072	0,269	83,33	45,83	58,33	0,626
Динамико-статистический	-0,088	0,264	95,83	70,83	70,83	0,616
Стохастический	-0,132	0,259	91,67	70,83	70,83	0,602
АР-1	-0,100	0,269	91,67	66,67	75,00	0,626
АР-2	-0,115	0,265	95,83	66,67	70,83	0,627
АР-3	-0,123	0,270	95,83	66,67	70,83	0,628
Фурье	-0,388	0,356	54,17	45,83	54,17	0,829
Осреднение	-0,188	0,256	91,67	70,83	70,83	0,596
Регрессионная комплексация	0,000	0,231	91,67	75,00	79,17	0,539

известно, позволяет исключить из временного ряда тренд — годовой ход, а сам ряд представить в виде аномалий:

$$\Delta L_{ij} = L_{ij} - \frac{1}{N \times M} \sum_{i=1}^N \sum_{j=1}^M L_{ij} - \frac{1}{M} \sum_{j=1}^M L_{ij}, \quad (6.1)$$

где L_{ij} — среднемесячные значения ледовитости Баренцева моря, i, j — номер года и месяца соответственно, N — число лет ряда, M — число месяцев.

В соответствии с методологией прогнозирования океанологических процессов (см. главу 4) преобразованный ряд (ΔL_{ij}) экстраполировался различными стохастическими моделями (см. табл. 6.7) с целью построения более эффективной обобщенной модели прогноза. Прогнозы выполнялись с заблаговременностью 1 мес на независимом материале с января 1988 г. по декабрь 1989 г., причем на каждом шаге прогнозирования все параметры моделей пересчитывались с учетом поступившей новой информации.

Из таблицы видно, что оправдываемость прогнозов по отдельным моделям колеблется от 25 до 71 % по критерию 0,67 σ и от 33 до 75 % по критерию 0,8 σ . Статистическая оптимизация этих моделей с помощью процедуры комплексирования позволила повысить качество прогнозирования: для модели регрессии оправдываемость составила 75 % и 79 % по критериям 0,67 σ и 0,8 σ соответственно. Эффективность прогнозирования — превышение оправды-

ваемости прогноза по модели регрессионной комплексации относительно инерционного прогноза — составила 8—9 %. Окончательный прогноз значений ледовитости представляет собой процедуру, обратную (6.1).

В целом, исходя из уровня относительной ошибки прогнозирования, который изменялся от 1,34 до 0,54, можно сделать вывод о высоком пределе статистической предсказуемости лишь для мо-

дели регрессионной комплексации прогнозов (2—3 мес). Для моделей авторегрессии 1-го — 3-го порядков, стохастической и байесовской, а также комплексной модели среднего, он составил приблизительно одно и то же значение — 1—2 мес.

Реализация блока (рис. В.5) стохастического прогнозирования ледовитости дала еще одно подтверждение, как и в случае прогнозирования АТПО, методического результата: повышение эффективности прогнозирования при использовании процедуры ста-

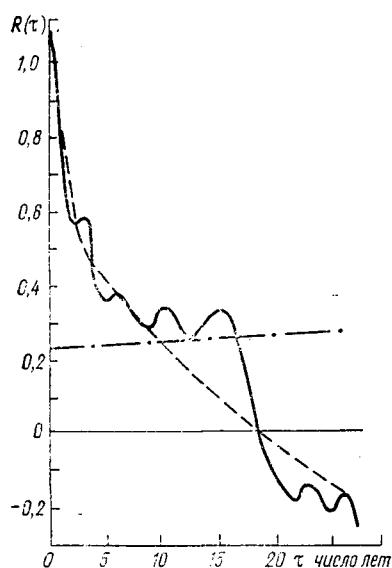


Рис. 6.10. Нормированная автокорреляционная функция ледовитости Баренцева моря.

тистической оптимизации и нескольких прогностических моделей. При этом, однако, пределы статистической предсказуемости определяются исключительно относительным уровнем ошибок прогнозирования (формула (4.7)), которые быстро возрастают с увеличением интервала упреждения прогноза.

Более консервативной, а значит, и более инерционной характеристикой климатической системы является среднегодовая ледовитость. Моделирование ее статистической структуры представляется важным для выделения долгопериодной составляющей процесса. В силу того, что выборочная совокупность подчиняется нормальному закону распределения, использование параметрического статистического анализа вполне уместно.

С целью получения характеристик предсказуемости среднегодовых значений ледовитости нами был выполнен расчет автокорреляционной функции и ее аппроксимация аналитической формулой вида

$$r(\tau) = e^{-\alpha \tau^\gamma} \cos \beta \tau \quad (6.2)$$

с параметрами, равными $\alpha = 0,4$, $\beta = \pi/38$, $\gamma = 0,4$ (рис. 6.10). Это позволило установить адекватность аналитической модели

вида исходному временному ряду, определить относительные дисперсии ошибок прогнозов (d) для различных упреждений прогноза и рассчитать максимальные пределы статистической предсказуемости (заблаговременность) для различных критериев точности прогнозов (табл. 6.8). Высокая инерционность процесса (см. рис. 6.10) (пересечение кривой автокорреляционной функции уровня значимости происходит на сдвиге более 10 лет) не требует усложнения процедуры прогнозирования — синтеза статистических моделей. В этом случае оказывается вполне достаточным использование модели АР-1 с коэффициентами авторегрессии, рассчитанными по формуле (6.2). Оправдываемость независимых прогнозов при этом на сдвигах от 1 до 3 лет остается высокой (см. табл. 6.8).

Таблица 6.8

Эффективность прогнозирования среднегодовых значений ледовитости Баренцева моря

	Заблаговременность, число лет									
	1	2	3	4	5	6	7	8	9	10
d по модели (6.2)	0,48	0,69	0,75	0,80	0,83	0,86	0,88	0,90	0,93	0,94
d по АР-1	0,48	0,68	0,66	0,84	0,88	0,86	0,88	0,90	0,92	0,88
Оправдываемость, %										
по $0,674 \sigma$	95	70	70	55	50		Менее 50 %			
по $0,8 \sigma$	100	85	90	60	55	55	55	Менее 50 %		

Для критерия $d = 0,9$ по формулам (4.8, 4.9) максимальный предел статистической предсказуемости составляет 8—10 лет, для критериев $0,674 \sigma$ и $0,8 \sigma$ он уменьшается до 4—6 лет.

Таким образом, выделение и прогнозирование долгопериодной составляющей процесса достаточно надежно можно осуществлять с помощью простых линейных стохастических моделей первого порядка. Сравнение результатов прогнозирования среднемесячных и среднегодовых значений ледовитости (см. табл. 6.7, 6.8) показывает, что улучшение прогностических свойств стохастических моделей обусловлено увеличением масштабов временного осреднения процесса.

Несмотря на очевидные преимущества как статистической оптимизации прогностических моделей, так и увеличения масштабов осреднения, сохраняется неопределенность — не удастся прогнозировать ряды среднемесячных и среднегодовых значений с упреждением более чем на 1—3 шага. Это характерно для многих гидрометеорологических процессов и связано с тем, что в рамках линейных стохастических моделей ошибка экстраполяции монотонно возрастает с увеличением заблаговременности прогноза. Поэтому, несмотря на значительную инерцию долгопериодных составляющих ледовитости — до 2—3 лет, прогнозы внутригодовой изменчи-

ности с заблаговременностью более 2—3 мес становятся неэффективными.

Устранить этот разрыв помогает методический прием, изложенный во введении, — прогнозирование значимых обобщенных параметров процесса (главных компонент, обобщенных факторов и др.), полученных в результате статистического диагноза процесса, т. е. построение многомерных моделей прогноза.

Обратимся к результатам прогнозирования среднемесячных значений ледовитости Баренцева моря с помощью многомерных моделей типа (4.1). На первом этапе (диагноз) исходная совокупность данных была преобразована МГК (табл. 6.9). Первые пять собственных векторов были выбраны как параметры, достаточные для описания многомерного процесса — в них содержится более 90 % общей дисперсии.

Таблица 6.9

Результаты разложения среднемесячных значений ледовитости Баренцева моря с 1946 по 1989 гг. по естественным функциям времени

Номер компоненты	Дисперсия, %	Интегральная дисперсия, %	Номер компоненты	Дисперсия, %	Интегральная дисперсия, %
1	60,6	60,6	7	1,7	97,5
2	14,4	75,0	8	0,9	98,4
3	8,2	83,2	9	0,6	99,0
4	5,8	89,1	10	0,4	99,5
5	4,4	93,5	11	0,3	99,8
6	2,3	95,8	12	0,2	100,0

Как было показано в п. 6.1 на примерах прогнозирования АТПО, важным условием эффективного прогнозирования по многомерной статистической модели вида (4.2) является устойчивость е.о.ф., рассчитанных либо по пространству, либо по времени. Проверка этого условия проводилась аналогичным образом и для ледовитости: добавление к исходным данным новой информации не изменило структуру е.о.ф. времени, что указывает на их устойчивость на периоде прогнозирования до одного года.

В соответствии с формулой (4.3) первые пять главных компонент среднемесячных значений ледовитости экстраполировались на один шаг (год) вперед по времени с использованием процедуры статистической оптимизации стохастических моделей (см. формулу (4.5)). Восстановление исходных значений позволило составить независимые прогнозы ледовитости Баренцева моря с заблаговременностью 12 мес. В общем случае заблаговременность прогноза определяется пределом статистической предсказуемости главных компонент ледовитости, который составляет 2—3 года, или 24—36 мес. Поэтому для целей фонового прогнозирования расчеты с такой заблаговременностью и месячным разрешением вполне

реальны. Процедура предвычисления ледовитости при этом сохранится прежней (см. формулу (4.2)); меняются лишь значения спрогнозированных главных компонент при постоянстве собственных векторов. Фактически с применением многомерных статистических моделей удастся спрогнозировать разномасштабные процессы.

На рис. 6.11 и 6.12 представлены примеры независимого прогнозирования и распределение ошибок 204 независимых прогнозов ледовитости с 1973 по 1989 годы. Обращает на себя внимание хорошее качественное и количественное описание сезонной структуры аномалий ледовитости. Это, по нашему мнению, составляет главное преимущество перед одномерными моделями, где прогнозируются непосредственно исходные ряды. В этом случае вид прогнозистической кривой относительно фактических значений как бы отстает на один шаг, повторяя все особенности процесса (рис. 6.13). Поэтому вторым преимуществом многомерной модели является синхронное описание (прогнозирование) экстремальных климатических, но не случайных значений в сезонной изменчивости ледовитости. И, наконец, третье преимущество — увеличение заблаговременности прогноза. При этом ошибка прогнозирования не увеличивается пропорционально заблаговременности (рис. 6.12), так как она распределяется в зависимости от устойчивости собственных векторов (е. о. ф.) от месяца к месяцу. Анализ ошибок прогнозирования, рассчитанных как разности спрогнозированных ($\Delta L_{пр}$) и фактических ($\Delta L_{ф}$) значений аномалий ледовитости Баренцева моря, позволяет выявить характерные особенности: а) ошибки прогнозирования относительно равномерно распределяются по всему интервалу независимого прогнозирования; б) максимальные значения ошибок приходятся на весенние и зимние месяцы, минимальные — на летние и осенние.

Эти особенности во многом определяются структурой собственных векторов (е. о. ф.), ответственных за описание внутригодовой изменчивости ледовитости, и межгодовой изменчивостью первых главных компонент, влияющих на распределение ошибок от года к году (рис. 6.14). Средняя оправдываемость по всему интервалу прогнозирования составила 82 %. Более стабильна среднемесячная оправдываемость (рис. 6.14 б), что является косвенным подтверждением факта устойчивости собственных векторов.

Большие изменения оправдываемости прогнозов прослеживаются от года к году (рис. 6.14 а). Минимальные значения приходятся на 1979 г. (59 %), 1981, 1983 и 1984 гг. (67 %), которым соответствуют экстремальные значения первого коэффициента разложения ледовитости. Максимальная оправдываемость прогнозирования составила 100 % в 1980 и 1986 гг. по критерию $0,8\sigma$. В эти годы значения главных компонент были близки к норме, как и в 1974, 1975, 1982, 1985 гг., в которые оправдываемость прогнозов достигала 91 %.

В целом, средняя оправдываемость прогнозов от года к году составила 81 %, а от месяца к месяцу — 82 %, что удовлетворяет

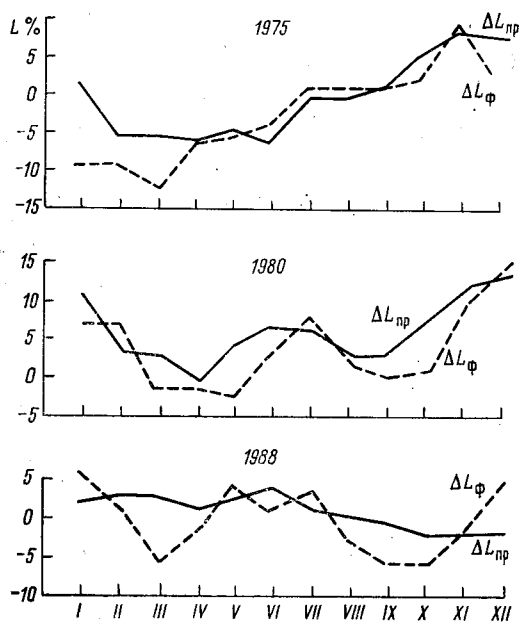


Рис. 6.11. Фактические ($\Delta L_{\text{ф}}$) и спрогнозированные ($\Delta L_{\text{пр}}$) значения аномалий ледовитости Баренцева моря с заблаговременностью 1 год.

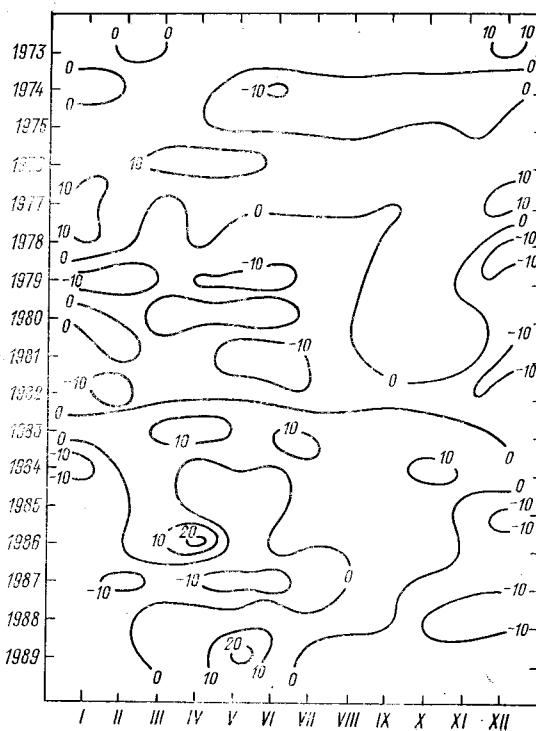


Рис. 6.12. Распределение ошибок независимого прогнозирования аномалий ледовитости Баренцева моря.

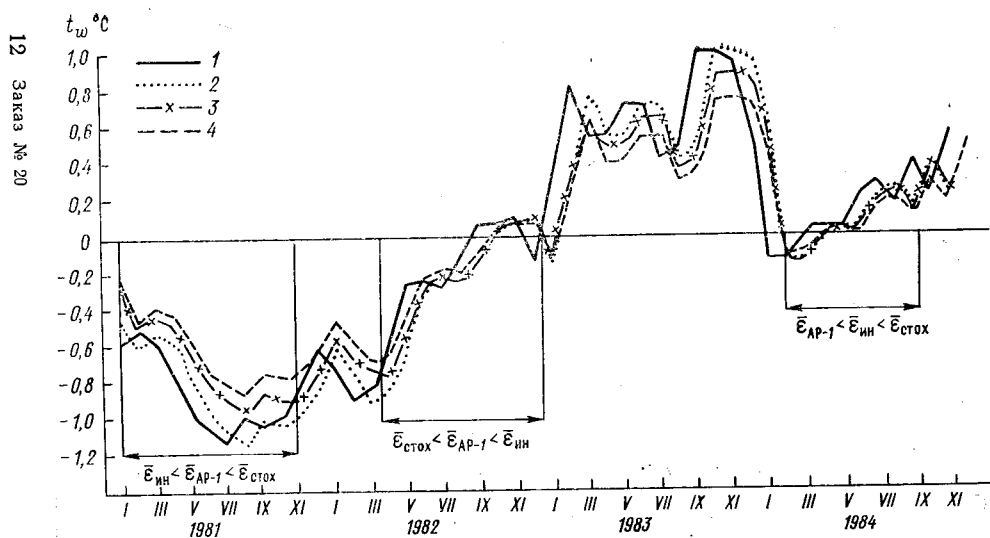
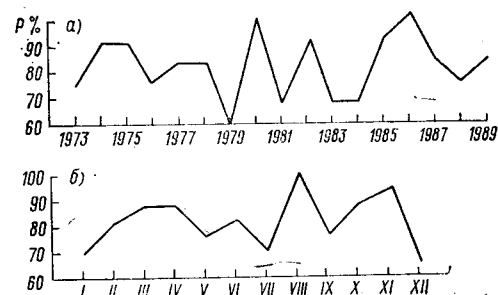


Рис. 6.13. Изменчивость рядов аномалий среднемесячной температуры воды слоя 0—200 м на Кольском меридиане [44].

1 — фактические; 2—4 — спрогнозированные: по инерции (2), динамико-статистическим методом (3), методом авторегрессии 1-го порядка (4).

Рис. 6.14. Оправдываемость P (%) прогнозов межгодовых (а) и внутригодовых (б) аномалий ледовитости Баренцева моря.



требованиям, предъявляемым к долгосрочным гидрометеорологическим прогнозам [61, 65]. Одним из основных выводов, который следует из анализа результатов независимого прогнозирования ледовитости Баренцева моря, состоит в том, что применение многомерных статистических моделей дает существенный выигрыш как в эффективности, так и в заблаговременности прогнозов по сравнению с теми же показателями в случае прогнозирования посредством одномерных статистических моделей. При этом удастся решить две наиболее трудные проблемы: а) разработать схемы прогнозов разномасштабных (по времени или пространству) процессов в рамках одной многомерной прогностической модели; б) добиться относительно равномерного распределения ошибок прогнозирования, а значит, и характеристик качества прогнозов по всему интервалу прогнозирования.

6.3. Прогнозирование пространственного распределения сестона в Норвежском море

В связи с повышенным интересом к океану как основному источнику биоресурсов актуально построение математических моделей, описывающих эволюцию гидробиологических характеристик.

С этих позиций в настоящей работе исследуются взаимосвязи океанологических процессов и биологической продуктивности вод Норвежского моря. Материалом для анализа послужили данные наблюдений по биомассам сестона, которые получены на стандартных океанографических разрезах Норвежского моря с 1968 по 1984 гг. в результате ежегодных июньских комплексных съемок [174]. В качестве основного инструмента реализации системного подхода здесь выступают методы многомерного статистического анализа.

На первом этапе вся исходная информация была подвергнута первичной обработке в целях восстановления пропусков наблюдений и фильтрации ошибок. Эти задачи были решены путем построения полиномиальной регрессионной модели для каждого исследуемого поля биомасс сестона. Значения биомасс представлялись функцией порядка m от широты и долготы. Оптимальный порядок модели выбирался по методу наименьших квадратов сравнительным тестированием функций первого — седьмого порядков. В результате вычислений все поля были представлены восстановленными значениями биомасс сестона в узлах сеточной области с шагом 2° по широте и долготе от $\varphi = 60^\circ$ с. ш. до $\varphi = 75^\circ$ с. ш. и от $\lambda = 15^\circ$ з. д. до $\lambda = 15^\circ$ в. д.

Выбор шага пространственного осреднения полей биомасс обусловлен принятой во ВНИИГМИ — МЦД дискретностью накапливаемых исторических гидрометеорологических данных по региону Северо-Европейского бассейна и согласован с [46]. Такое соответствие дало возможность провести сравнительный анализ различных океанологических и биологических полей.

В этой связи можно понять роль динамики вод Северо-Европейского бассейна в изменчивости теплового режима верхнего слоя Норвежского моря и пространственного распределения биомасс сестона. Реализация методов многомерного статистического

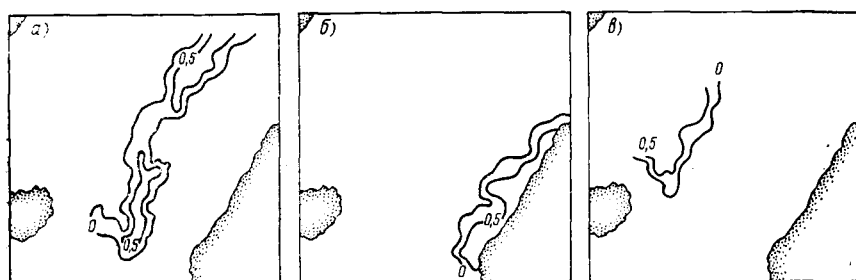


Рис. 6.15. Пространственное распределение первых трех общих факторов разложения полей биомасс сестона в Норвежском море.

а — 1-й фактор, б — 2-й фактор, в — 3-й фактор.

анализа — факторного и кластерного анализа — позволила получить несколько обобщенных параметров (факторов), в которых нашли отражение наиболее важные крупномасштабные черты изменчивости биомасс планктона и океанологических характеристик (рис. 6.15).

Первый общий фактор описывает 30 % суммарной дисперсии восстановленных полей биомасс. Он локализован в центральной части Норвежского моря и вытянут с юго-запада на северо-восток, где четко прослеживается его очаговая структура. Сравнивая пространственную структуру первого общего фактора (рис. 6.15 а) и первого общего фактора океанологических характеристик (см. рис. 5.9), нетрудно сделать вывод о качественном их совпадении: очаги изменчивости планктона и океанологических характеристик практически полностью располагаются во фронтальной зоне (см. рис. 5.12). Центральный очаг располагается на широте 66—68° с. ш. в зоне разделения струй Норвежского течения [46] и конвергенции вод атлантического и полярного происхождения. Южный очаг первого общего фактора биомасс планктона лежит в районе Фарерско-Шетландского пролива, где воды Северо-Атлантического и Норвежского течений встречаются с водами Восточно-Исландского течения, формируя мощную фарерско-исландскую ветвь полярной фронтальной зоны.

Таким образом, наличие развитых фронтальных зон является необходимым условием формирования очагов интенсивного развития биомасс сестона в верхнем слое Норвежского моря. Термический контраст арктических и атлантических вод однозначно характеризует местоположение и интенсивность полярного фронта и, следовательно, предопределяет условия возникновения сопряженных в пространстве очагов развития биомасс сестона.

Зимой фронтальная зона располагается к западу от нулевого меридиана, а в переходные сезоны она распространяется на восток вдоль 62-й параллели. При этом приток атлантических вод в весенний сезон локализуется в узкой восточной ветви Норвежского течения, располагающейся в прибрежной зоне Скандинавии.

Областью влияния второго общего фактора разложения биомасс является только восточная часть Норвежского моря (см. рис. 6.15 б). Влияние фактора распространяется на прибрежную зону Скандинавии (200—400 м. миль). Этот фактор также обладает ярко выраженной очаговой структурой. Размеры очагов — порядка 300 км. Их преимущественная локализация вдоль побережья Скандинавии косвенно указывает на интенсивное развитие процессов взаимодействия Норвежского течения с прибрежными водами, распресненными стоком рек. Хорошо известен тот факт, что вдоль континентального шельфа Скандинавии формируются многочисленные зоны так называемых прибрежных фронтов [199]. Основной вклад в их формирование вносит градиент солености.

Второй общий фактор океанологических характеристик выявляет роль градиента солености в шельфовой зоне Скандинавии (см. рис. 5.9). Его пространственное распределение во многом совпадает со структурой второго общего фактора биомасс сестона (см. рис. 6.15 б). Следовательно, в данном случае можно указать на причину формирования очагов биомасс, связанную с локальной динамикой вод. Однако здесь взаимодействие происходит уже между «солеными» водами Норвежского течения и распресненными водами континентального шельфа Скандинавии. Изменчивость поля солености является косвенным индикатором формирования биомасс в прибрежной зоне Скандинавии.

Третий, четвертый и пятый факторы разложения биомасс планктона описывают приблизительно равное количество дисперсии и в сумме составляют около 10 %. Их пространственная структура во многом схожа и характеризуется наличием многочисленных мезомасштабных очагов диаметром не более 300 км (см. рис. 6.15 в). В распределении третьего фактора можно выделить северо-западный очаг (см. рис. 6.15 в), который также прослеживается в поле третьего общего фактора разложения океанологических характеристик (см. рис. 5.9). Данное сопоставление обосновывает вывод о сопряженном изменении интенсивности Восточно-Исландского течения и усилении мезомасштабных очагов развития биомасс.

В целом приводимые результаты статистического анализа наглядно демонстрируют общность основных выявленных закономерностей пространственного распределения океанологических и биологических характеристик. Это позволяет сформулировать рабочую гипотезу о возможном физическом механизме формирования крупномасштабных структурных изменений полей биомасс сестона Норвежского моря.

Анализ внутригодовой изменчивости положения фронтальной зоны выявил весьма примечательный факт — наличие двух устой-

чивых климатических состояний всего комплекса гидрофизических характеристик в Норвежском море. Первое состояние — усиленный водообмен с Атлантическим океаном, при котором система фронтальных зон южной части Норвежского моря приобретает меридиональную направленность (наиболее часто это происходит в зимний и летний сезоны), и второе состояние — ослабление водообмена с Атлантикой, что проявляется в зональном положении фарерско-шетландской ветви полярной фронтальной зоны и скандинавского прибрежного фронта весной и осенью. Таким образом, дважды в год в пространственной структуре динамических, термохалинных и, очевидно, биологических полей усиливается меридиональная ориентация градиентов промыслово-океанологических характеристик с одновременным увеличением масштабов горизонтальной изменчивости.

Наоборот, при ослабленном водообмене отмечается усиление зональных контрастов на фоне увеличения числа локальных (диаметром 100—300 км) очагов с аномальными свойствами. Переход из одного состояния в другое происходит очень быстро, практически скачком от сезона к сезону (см. рис. 5.12). Это позволяет предложить «триггерный эффект» в качестве основного крупномасштабного механизма, ответственного за структурные изменения сезонных полей биомасс сестона.

Подобный механизм позволяет интерпретировать наблюдаемые резкие межгодовые колебания фонового распределения термохалинных и биологических характеристик, что практически невозможно сделать в терминах моделей, непрерывных по времени.

Следует подчеркнуть, что июнь традиционно понимается как месяц, переходный от весны к лету для основных типов атмосферных процессов и вертикальной структуры вод в Норвежском море [7]. Поэтому даже небольшие колебания параметров климатической системы северного полушария могут приводить к резкой изменчивости полей зоопланктона в период проведения июньских съемок, когда происходит интенсивная региональная перестройка всех процессов от весны к лету и вся экосистема Норвежского моря наиболее чувствительна к внешнему воздействию.

С этих позиций полученные пространственные поля первых общих факторов по сути отражают основные типовые структуры распределения планктона в Норвежском море. Чтобы убедиться в этом, была проведена классификация восстановленных полей биомасс сестона в пространстве общих факторов. В результате классификации получено три класса, которые обобщают изменчивость распределения биомасс сестона главным образом в основных очагах: полярной фронтальной зоне, прибрежной скандинавской зоне и водах Атлантического и Восточно-Исландского течений (рис. 6.16).

Устойчивость полученных классов проверялась методом линейного дискриминантного анализа [198], для чего была построена разделяющая классы линейная дискриминантная функция и оценены дискриминантные индексы классов.

На основании выделенных типов полей был составлен каталог классов пространственного распределения биомасс сестона. Принцип составления каталога достаточно прост: каждое поле планктона количественно сравнивалось с каждым из трех классов и наименьшее различие между наблюдаемым полем и классом было основанием для присвоения данному полю соответствующего номера класса.

Анализ выделенных классов в целом подтверждает выдвинутую гипотезу «триггерного эффекта». Класс 1 (см. рис. 6.16) со-

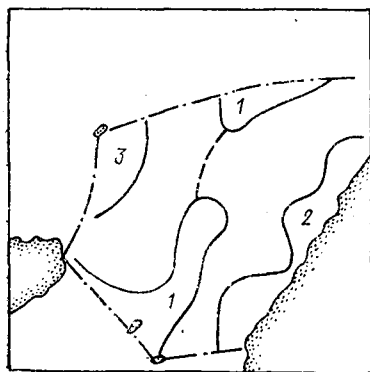


Рис. 6.16. Районирование Норвежского моря по результатам классификации многолетней изменчивости биомасс сестона.

Обозначены зоны 1—3 с максимальными значениями биомасс в соответствующих классах.

ответствует ситуациям интенсивного развития биомасс сестона, а класс 2 характеризует периоды слабого развития планктона.

Важно, что полученная классификация отражает не только общие закономерности фоновое распределения биомасс сестона, но и особенности пространственной текстуры полей.

Для класса 1 типичны крупномасштабные 300—500 км очаги максимальных значений биомасс, расположение которых качественно согласуется с распределением климатических фронтальных зон.

Класс 2, наоборот, демонстрирует резко выраженную локальную очаговую структуру полей биомасс при общем понижении фона.

Выделен класс 3, который характеризуется в целом пониженным фоном биомасс сестона, но в отличие от класса 2 имеет очаг повышенной концентрации зоопланктона к северо-востоку от Исландии в зоне влияния Восточно-Исландского и Ян-Майенского течений.

В терминах «триггерного эффекта» выделенные классы распределяются следующим образом: открытый меридиональный перенос тепла из Атлантического океана в Норвежское море соответствует классу 1; ослабленный перенос тепла — классам 2 и 3 одновременно. Последняя особенность заставляет обратить внимание на особую роль гидрологических процессов северо-западной части Норвежского моря в формировании полей биомасс. Слабая осве-

щенность этого района данными наблюдений биомасс пока не позволяет охарактеризовать взаимосвязь гидрологических и биологических характеристик в обсуждаемых понятиях.

На основе предположения о том, что классифицированные типы полей биомасс устойчивы во времени и укладываются в схему однородных цепей Маркова, был проведен расчет эмпирических вероятностей перехода одного класса в другой (табл. 6.10).

Таблица 6.10

**Матрица переходных вероятностей классов
распределения биомасс сестона**

Класс поля	Класс поля			Σ
	1	2	3	
1	(0,59)	0,47	0,04	1,0
2	0,22	(0,44)	0,34	1,0
3	0,36	0,35	(0,29)	1,0

Очевидно, что наиболее устойчиво сохраняется от года к году класс 1, соответствующий повышенному фону концентраций биомасс во фронтальных зонах. Данный класс с наибольшей вероятностью сменяется классом 2, эквивалентным низкому среднему уровню биомасс сестона с развитой мезомасштабной текстурой.

В свою очередь, это состояние либо сохраняется на следующий год, либо сменяется классом 3. Свойственное классу 3 распределение биомасс имеет в целом низкий фон с очагом относительно высоких значений на северо-западе Норвежского моря. Подобное распределение чаще всего предшествует общему повышению концентраций сестона на всей акватории моря (класс 1). Такова самая общая характеристика межгодового изменения полей биомасс сестона Норвежского моря.

Выполненные расчеты могут быть основой построения вероятностного прогноза типа пространственного распределения зоопланктона в Норвежском море. Прогностическая ценность полученной типизации состоит в том, что в условиях относительно невысокой точности гидробиологической информации предсказание будущего типа поля биомасс позволяет избежать ошибок, неизбежно возникающих при составлении точечных количественных прогнозов.

Кроме того, открывается потенциальная возможность поиска прогностических связей между выявленными типами полей биомасс сестона и типами пространственного распределения гидрофизических характеристик. В дальнейшем предполагается автоматизировать данный подход для целей оперативного диагноза

и прогноза промыслово-океанологических характеристик, формирующих состояние морских экосистем.

В заключение главы подведем итог основным результатам статистического прогнозирования океанологических процессов. В методологическом аспекте следует обратить внимание на два реализованных принципа: принцип единства физического и статистического описания процесса и принцип дополнительности, который состоит в том, что применяя различные статистические модели мы как бы рассматриваем различные «границы» объекта исследования. Эти принципы позволили наилучшим (в статистическом смысле) образом идентифицировать прогностические модели, т. е. определить их параметры — порядок, размерность и степень связанности элементов моделей.

В основе методологического аспекта лежит процедура экстраполяции обобщенных параметров различными статистическими методами, в том числе и моделями, включающими процедуры комплексации. Прогнозы температуры воды поверхностного слоя Северной Атлантики и ледовитости Баренцева моря показали преимущество использования многомерных моделей перед прогнозами непосредственно исходных рядов океанологических характеристик с помощью одномерных моделей в отношении как эффективности, так и заблаговременности.

Если первый результат (применение одномерных стохастических моделей) достаточно известен и подтвержден практикой прогнозирования [65, 146, 151], то второй — позволяет по-новому взглянуть на возможности прогнозирования обобщенных параметров. Действительно, в результате анализа проведенных прогнозов стало понятно, что данный подход позволяет: а) прогнозировать многомерный процесс без увеличения мерности прогностической модели; б) увеличивать заблаговременность прогноза не за счет экстраполяции модельных решений, а за счет выявленной устойчивости базисных функций, т. е. учитывать связность полей или временных рядов процесса.

Прогнозирование пространственного распределения сестона в Норвежском море посредством построения вектора классовых состояний отражает современные тенденции в гидробиологии: прогнозировать конкретные величины биомассы сестона пока не представляется возможным из-за сложной структуры исходной информации. При этом прогностическое решение становится многовариантным. Здесь важно отметить, что без проведения диагноза различными методами МСА в принципе было невозможно добиться приемлемого прогностического результата.

Таким образом, эффективность прогнозирования во многом определяется комплексностью подхода как в диагностическом, так и в прогностическом блоках (см. рис. В.5).

ЗАКЛЮЧЕНИЕ

Завершая изложение результатов статистического моделирования, хотелось бы еще раз подчеркнуть, что множественность параметров и неоднозначность их взаимосвязей, характеризующих состояние океана, предопределила методологический подход к исследованию океанологических процессов, в основе которого лежит аппарат многомерного статистического анализа. Здесь читатель правомерен задать вопросы: «Являются ли методы МСА универсальными в выявлении скрытых структурных особенностей процессов?» и «Чего мы достигли?».

Многолетний отечественный и зарубежный опыт применения методов МСА в геофизических исследованиях, активное и успешное его внедрение в других областях науки позволяют на первый вопрос ответить утвердительно. Этот вывод следует и из результатов настоящей работы. Одним из положительных эффектов проведенных с помощью методов МСА исследований является переход от традиционного трехмерного информационного пространства к M -мерному пространству состояний, размерность которого определяется числом переменных в векторе наблюдений. При этом исследователю открывается все «многоцветие» многочисленных связей между переменными, которое затем представляется четкими образами через главные компоненты, факторы, классы, типы, существенно облегчающие интерпретацию сложных физических механизмов.

В большой степени многостороннему анализу океанологических процессов способствовало применение различных методов МСА. С одной стороны, это вызвано пониманием того, что описание бесконечного многообразия состояний океанологических процессов невозможно посредством некоторой «наилучшей» модели, так как различные типы моделей призваны выполнять различные функции и предназначены для разных целей. С другой стороны, чтобы взглянуть на процессы «под разным углом зрения», как бы вращая куб данных, необходимо применение различных методов МСА: степень описания изменчивости процесса конкретным методом и статистическая структура остатков, не описываемая им, определяют необходимость применения другого метода.

Реализация данного подхода позволила отойти от исследования связей между отдельными элементами и перейти к комплексному анализу процессов в целом. В этом, на наш взгляд, состоит развернутый ответ на первый вопрос, который раскрывает суть предлагаемой методологии исследований океанологических процессов.

Результаты, полученные в контексте вышеизложенной методологии, конечно, не исчерпывают всех возможностей применения метода МСА в океанологии. Но они позволяют, по нашему мнению, в полной мере ответить на второй поставленный вопрос. Во-первых, стало понятно, что исследуемый процесс с помощью одного метода полностью не описать — это особенно наглядно прослеживается в многомерных океанологических процессах, где оценки однородности, стационарности, устойчивости связей носят вероятностный характер. Неопределенность взаимосвязей и изменчивости наблюдаемых параметров можно минимизировать, если использовать идею агрегирования нескольких методов МСА.

Во-вторых, обобщенные параметры (компоненты, факторы, классы, типы), отражающие результаты статистического диагноза, приобретают ключевое значение при прогнозировании крупномасштабных составляющих многомерных процессов. Включая, например, общие факторы в качестве входных параметров одномерных стохастических моделей можно прогнозировать эволюцию многомерного процесса (поле элемента во времени, временной ход нескольких элементов).

Нами проведены массовые прогностические расчеты по трем океанологическим элементам: температуре воды, ледовитости и биомассы зоопланктона по различным регионам Атлантического и Северного ледовитого океанов. Примеры показывают высокую эффективность подобных прогнозов. Эти результаты вселяют определенную надежду и оптимизм: перспектива такого подхода очевидна, особенно в связи с ближайшими задачами исследования глобальных изменений климата.

В-третьих, нами реализована идея комплексации нескольких прогностических моделей в океанологической практике. Этот путь оказался плодотворным — в значительной степени уменьшаются неопределенности статистического описания эволюции процесса и, следовательно, увеличивается эффективность прогнозов и их статистическая предсказуемость.

Несомненно, мы понимаем, что окончательное суждение о перспективности данного подхода требует не только значительного расширения списка прогнозируемых и верифицируемых процессов, но главным образом оценок специалистов, применяющих в своей практике в той или иной степени статистические методы многомерного анализа. Изложенные в книге результаты статистического диагноза и прогноза океанологических процессов могут показаться читателю излишне сложными для практической реализации. Нам думается, для этого нет серьезных оснований, так как к настоящему времени созданы универсальные пакеты прикладных программ для ЭВМ различных классов, которые в значительной степени облегчают вычислительный аспект предлагаемой методологии.

СПИСОК ЛИТЕРАТУРЫ

1. Абрамов Р. В., Гущин О. А. Исследование трендов поля т. п. о. на основе статистической модели.—Деп. в ВИНТИ, 2.09.85, № 6439—85.
2. Адаменко В. Н., Богданов А. Л. Классификация крупных озер мира на основе оценки составляющих их массо-энергообмена/Гидрология Байкала и других водоемов.—Новосибирск: Наука, 1984.—С. 5—14.
3. Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Прикладная статистика. Основы моделирования и первичная обработка данных.—М.: Финансы и статистика, 1983.—471 с.
4. Александров В. В., Горский Н. Д. Алгоритмы и программы структурного метода обработки данных.—Л.: Наука, 1983.—208 с.
5. Алексеев А. П. Полярный фронт в Норвежском и Гренландском морях//Труды ПИНРО, 1959.—Вып. 11.—С. 60—73.
6. Алексеев Г. В., Мурзин А. И. Разложение по Е. О. Ф. как основа численного представления сплоченности льда и прогноза его перераспределения//Труды ААНИИ, 1966.—Вып. 279.—С. 105—112.
7. Алексеев Г. В., Николаев Ю. В., Романцов В. А. Норвежская энергоактивная зона//Итоги науки и техники. Атмосфера, океан, космос — Программа «Разрезы».—М.: Наука, 1985.—Т. 5.—С. 45—62.
8. Андрукович Н. Ф. Некоторые свойства метода главных компонент/Многомерный статистический анализ в социально-экономических исследованиях.—М.: Наука, 1974.—С. 189—228.
9. Андрущенко В. И. Способ вероятностного прогноза непериодических колебаний уровня моря//Труды ААНИИ, 1985.—Вып. 389.—С. 89—93.
10. Асмус В. В. и др. Программный комплекс кластеризации многозначных данных//Исследование Земли из космоса, 1988.—№ 3.—С. 86—94.
11. Асмус В. В. и др. Кластерный анализ самолетных радиолокационных изображений ледяного покрова//Труды ГосНИИЦИПР, 1986.—Вып. 25.—С. 64—72.
12. Афффи А., Эйзен С. Статистический анализ. Подход с использованием ЭВМ.—М.: Мир, 1982.—488 с.
13. Багров Н. А. Объединение нескольких прогнозов//Метеорология и гидрология, 1982.—№ 8.
14. Багров Н. А. О комплексном методе прогнозов//Метеорология и гидрология, 1962.—№ 4.
15. Багров Н. А. Преобразование и отбор предсказателей в корреляционном анализе//Труды Гидрометцентра СССР, 1970.—Вып. 64.—С. 3—23.
16. Багров Н. А. Индекс аналогичности векторных полей//Труды ЦИП, 1963.—Вып. 123.
17. Багров Н. А. Статистическая теория опознавания в гидрометеорологии//Труды Гидрометцентра СССР, 1983.—Вып. 244.—С. 51—62.
18. Багров Н. А. О канонической корреляции векторов//Труды Гидрометцентра СССР, 1980.—Вып. 226.—С. 3—10.
19. Багров Н. А. О некоторых особенностях корреляционного анализа и их применения к прогнозам погоды//Метеорология и гидрология, 1968.—№ 1.
20. Багров Н. А., Мерцалова Н. И. О тепловом взаимодействии океана и атмосферы//Труды Гидрометцентра СССР, 1970.—Вып. 64.—С. 24—34.
21. Багров Н. А., Кожевникова Н. Н. Объективный анализ температуры поверхности океана на Северном полушарии//Метеорология и гидрология, 1981.—№ 12.—С. 69—76.

22. Батырева О. В. Сравнение канонических переменных с главными компонентами при учете влияния Северной Атлантики на температуру воздуха на ЕТС//Труды Гидрометцентра СССР, 1983.— Вып. 244.— С. 62—77.
23. Белкин И. М. О методике объективного выделения водных масс//Труды ВНИИГМИ—МЦД, 1984.— Вып. 102.— С. 146—156.
24. Белкин И. М. Морфолого-статистическая объективная классификация вертикальных профилей гидрофизических параметров//Докл. АН СССР, 1986.— № 3.— С. 707—711.
25. Белкин И. М. Методы анализа вертикальных профилей гидрофизических параметров//Труды ВНИИГМИ—МЦД, 1981.— Вып. 90.— С. 60—67.
26. Белкин И. М., Семенов Е. В. Сплайновая интерполяция вертикальных профилей геофизических параметров. Обзор проблемы и методов интерполяции//Труды ВНИИГМИ—МЦД, 1984. Вып. 102.— С. 135—145.
27. Беляев В. И. Обработка и теоретический анализ океанологических наблюдений.— Киев: Наукова думка, 1973.— 295 с.
28. Беляев В. С., Бунимович Л. А. Методика классификации участков записей мелкомасштабных флуктуаций гидрофизических полей//Океанология, 1983.— 23.— № 5.— С. 882—886.
29. Бирман Б. А., Ларин Д. А., Позднякова Ф. Г. Классификация полей аномалий теплового баланса северной части Атлантического океана//Труды ВНИИГМИ—МЦД, 1986.— № 122.— С. 3—42.
30. Благущ П. Факторный анализ с обобщениями.— М.: Финансы и статистика, 1989.— 248 с.
31. Божков А. Т., Гудзь П. К., Сперанский Д. А. Структура и энтальпия вод в районе океанской станции «С» в Северной Атлантике//Труды ГОИН, 1983.— Т. 170.— С. 32—40.
32. Божков А. Т., Буб А. Ф., Вольф С. В. Некоторые результаты исследования водных масс с помощью метода главных компонент//Проблемы Арктики и Антарктики, 1978.— № 53.— С. 27—35.
33. Бокс Дж., Дженкинс Т. Анализ временных рядов: прогноз и управление, вып. 1.— М.: Мир, 1974.— 406 с.
34. Болч Б. У., Хуань К. Д. Многомерные статистические методы для экономики.— М.: Финансы и статистика, 1979.— 341 с.
35. Борисенков Е. П. Результаты и перспективы развития физико-статистических методов сверхдолгосрочных и климатических прогнозов и их интерпретация//Труды V Всес. совещ. по применению статистических методов в метеорологии, 1987.— С. 3—12.
36. Брусиловский П. М. Коллективы предикторов в экологическом прогнозировании.— Изд. Саратовского гос. ун-та, 1987.— 104 с.
37. Булгаков Н. П., Ильин Ю. П., Рыбалка В. А. Использование главных компонент для анализа структуры верхнего слоя Черного моря.— 1983.— 14 с. МГИ. Деп. ВИНТИ № 187—84 деп. 4 янв. 1984.
38. Булгаков Н. П., Ломакин П. Д., Рыбалка В. А. Исследование гидрофизической структуры бассейна Карибского моря с применением кластер-анализа//Докл. АН УССР, 1988.— № 12.— С. 12—15.
39. Вайновский П. А. О факторном анализе вертикального распределения плотности воды//Труды ЛГМИ.— 1984.— Вып. 87.
40. Вайновский П. А., Кондратьев А. В. Стохастическое моделирование термического режима верхнего слоя Норвежского моря//Гидрометеорологические процессы в промысловых районах Северной и Южной Атлантики.— Л.: Изд. ЛПИ, 1987.— С. 80—85.
41. Вайновский П. А., Масловский М. И. Крупномасштабные фронтальные зоны Норвежского моря.— В кн.: Исследование южной части Норвежского моря. М.: Гидрометеиздат.— 1989.— С. 30—34.
42. Вайновский П. А., Титов Ю. Э., Малек Н. В. О вероятностном прогнозировании температуры воды в океане (байесовский подход)//Труды ЛГМИ, 1990. Вып. 106.
43. Вайновский П. А., Титов Ю. Э., Трущина Т. В. Результаты комплексирования статистических моделей прогноза//Труды ЛГМИ, 1989.— Вып. 105.— С. 32—36.

44. Вайновский П. А., Титов Ю. Э. Методические рекомендации по статистико-вероятностному прогнозированию океанологических характеристик. — Мурманск, 1989. — 93 с.
45. Вайновский П. А., Титов Ю. Э. Практические аспекты применения многомерного анализа. — 1986. — 17 с. Рук. деп. ВИНТИ 11.12.1986 № 8462—В86.
46. Вайновский П. А., Титов Ю. Э. О структуре термохалинных полей деятельного слоя Норвежской энергоактивной зоны. В сб.: Исследование крупномасштабного взаимодействия океана и атмосферы в Северной Атлантике. — Л.: Изд. ЛПИ. — 1986. — С. 113—121.
47. Василенко В. М., Мирабель А. П., Кузин В. И. О пространственной скорости течений в Саргассовом море по данным ПОЛИМОДЕ// Океанология, 1980. — Т. 20. — № 3. — С. 432—440.
48. Василенко В. М., Мирабель А. П. О параметризации вертикальной структуры течений в тропической Атлантике с помощью статистически ортогональных функций//Океанология, 1976. — Т. 16. — № 2.
49. Василенко В. М., Мирабель А. П. О вертикальной структуре течений в океане в различных диапазонах частот//Изв. АН СССР. ФАО. — 1977. — Т. 13, № 3. — С. 328—331.
50. Вительс Л. А. Многолетние изменения повторяемости форм атмосферной циркуляции и их преобразование в связи с солнечной активностью// Труды ГГО, 1960. — Вып. 90. — С. 95—115.
51. Воробьев В. Н., Смирнов Н. П., Титов Ю. Э. О пространственной и временной структуре энthalпии вод Северо-Европейского бассейна. — В сб.: Исследование взаимодействия океана и атмосферы в энергоактивных зонах. — Л.: Изд. ЛПИ, 1985. — С. 3—15.
52. Воробьев В. Н., Титов Ю. Э. Сезонная изменчивость энthalпии вод и адвекции тепла в северной части Атлантического океана//Научн. тр. (межвузовские): Исследование крупномасштабного взаимодействия океана и атмосферы в Северной Атлантике. — Л.: Изд-во ЛПИ, 1986.
53. Вязилов Е. Д. Применение методов объективного анализа для восстановления океанографических полей//Труды ВНИИГМИ—МЦД, 1981. — № 80. — С. 50—60.
54. Гаврилин Б. Л., Мирабель А. П. Методы параметризации крупномасштабных океанологических процессов//Океанология, 1973. — Т. 3. — № 6.
55. Галеркин Л. И. Двумерная статистика термохалинных полей и водные массы Северной Атлантики//Труды ВНИИГМИ—МЦД, 1984. — Вып. 102. — С. 108—134.
56. Галеркин Л. И. и др. Опыт машинной классификации кривых вертикального распределения температуры воды, солёности, плотности, скорости звука и их вертикальных градиентов//Труды ВНИИГМИ—МЦД. — 1978. — Вып. 45. — С. 126—138.
57. Гандин Л. С. Объективный анализ метеорологических полей. — Л.: Гидрометеиздат, 1963. — 287 с.
58. Гандин Л. С. О применении метода канонических корреляций в метеорологии//Труды ГГО, 1967. — Вып. 208.
59. Гасюков П. С. О совместном разложении полей температуры воды и воздуха по статистическим ортогональным функциям//Труды ВНИРО. — 1976. — Вып. 114. — С. 64—72.
60. Глаголева М. Г. и др. Об аналитическом представлении полей температуры воды в Тихом океане//Труды Гидрометцентра СССР, 1980. — № 229. — С. 52—56.
61. Глаголева М. Г., Скриптунова Л. И. Прогноз температуры воды в океане. — Л.: Гидрометеиздат. — 1979. — 167 с.
62. Глобальный климат/Под ред. Дж. Т. Хотона. — Л.: Гидрометеиздат, 1987. — 551 с.
63. Гольдман Р. С., Ильичев В. И. Методы прогнозирования промышленных скоплений в океане на основе анализа океанической изменчивости// Доклады АН СССР, 1983. — Т. 267. — № 3. — С. 737—739.
64. Груза Г. В. Опыт применения главных компонент в технике корреляционного прогноза//Труды САНИГМИ, 1969. — Вып. 40 (55). — С. 19—36.

65. Груза Г. В., Ранькова Э. Я. Вероятностные метеорологические прогнозы.—Л.: Гидрометеониздат.—1983.—271 с.
66. Даценко Н. М. О расчете естественных составляющих для метеорологических полей, заданных наблюдениями в большом числе точек//Труды Гидрометцентра СССР, 1978. Вып. 210.—С. 48—59.
67. Даценко Н. М. Анализ естественных составляющих поля температуры водной поверхности океанов северного полушария//Труды Гидрометцентра СССР, 1987.—Вып. 278.—С. 108—126.
68. Даценко Н. М., Перфилов В. И., Сонечкин Д. М. Методика расчета естественных составляющих метеорологических полей.—Изв. АН СССР. ФАО.—1983.—Т. 19.—№ 4.—С. 348—356.
69. Драган Я. П., Рожков В. А., Яворский И. Н. Методы вероятностного анализа ритмики океанологических процессов.—Л.: Гидрометеониздат, 1987.—319 с.
70. Дрейпер Н., Смит Г. Прикладной регрессионный анализ.—М.: Статистика, 1973.—392 с.
71. Дрыгина И. А., Смирнов Н. П., Саруханян Э. И. Об использовании компонентного анализа при разработке методики долгосрочного прогноза ледовитости арктических морей//Труды ААНИИ, 1977, вып. № 341.—С. 119—141.
72. Дубров А. М., Мазуров М. Е. Анализ статистических связей в одномерном и многомерном анализе.—М.: Финансы и статистика, 1980.—150 с.
73. Дубровский С. А. Прикладной многомерный статистический анализ.—М.: Финансы и статистика, 1982.—216 с.
74. Жамбю М. Иерархический кластерный анализ и соответствия.—М.: Финансы и статистика, 1988.—342 с.
75. Жуковская В. М., Мучник И. Б. Факторный анализ в экономических исследованиях.—М.: Статистика, 1976.—151 с.
76. Жуковский Е. Е., Гудновский А. Ф. Методы оптимального использования метеорологической информации при принятии решений.—Л.: Гидрометеониздат, 1978.—52 с.
77. Загоруйко Н. Д. и др. Пакет прикладных программ ОТЭКС.—М.: Финансы и статистика, 1986.
78. Елисеева И. И., Рукавишников В. О. Логика прикладного статистического анализа.—М.: Финансы и статистика, 1982.—191 с.
79. Елкина В. Н. и др. Применение методов таксономии в океанологии//Экспресс инф. ЦНИИТЭИРХ, 1973.—Сер. 9.—Вып. 2.—С. 1—35.
80. Енюков И. С. Методы, алгоритмы и программы многомерного статистического анализа.—М.: Финансы и статистика, 1986.—232 с.
81. Иберла К. Факторный анализ.—М.: Статистика.—1980.—398 с.
82. Иванова Е. В., Незлин Н. П. Выделение ассоциаций планктонных фораминифер в современных осадках Мирового океана при помощи факторного анализа//Океанология, 1986.—Т. 26.—Вып. 1.—С. 94—99.
83. Имас Л. И. Обзор работ по дискриминантному анализу//Труды САРНИГМИ, 1976.—Вып. 3 (112).—С. 3—27.
84. Имас Л. И. Численный эксперимент прогноза сильных похолоданий в Ташкенте методом дискриминантного анализа//Труды САРНИГМИ, 1971.—Вып. 58 (73).
85. Имас Л. И., Шмакова В. С. Дискриминантный анализ и прогноз//Труды САРНИГМИ, 1976.—Вып. 22 (103).
86. Иереског К. Г., Клован Д., Реймент Р. А. Геологический факторный анализ.—Л.: Недра, 1980.—223 с.
87. Калашников П. А. Первичная обработка гидрологической информации: физико-химические свойства морской воды. Л.: Гидрометеониздат, 1985.—152 с.
88. Карандашева Т. Г. Сезонная изменчивость термической стратификации вод Норвежского и Гренландского морей//Метеорология и гидрология, 1985.—№ 3.—С. 83—89.
89. Карнаухов А. А. О возможности определения термической и соленостной структуры водных масс по одному или двум параметрам на поверх-

ности.— В кн.: Гидрофизические исследования в северной части Тихого океана.— Владивосток, 1982.— С. 37—44.

90. Карпеев Г. А. О некоторых особенностях разложения метеорологических полей по системам ортогональных функций//Труды Гидрометцентра СССР, 1970.— Вып. 64.— С. 69—80.

91. Карпеев Г. А. Статистическая ортогонализация случайных векторов и прогноз случайных полей//Труды Гидрометцентра СССР, 1970.— Вып. 64.— С. 81—90.

92. Карпова И. П., Кузина Н. А. Об использовании показателя оптимальности в классификации в задачах выделения водных масс//Исследование и освоение Мирового океана. Л.: ЛПИ, 1982.— Вып. 77.— С. 103—108.

93. Кашьяп А., Рао Р. Построение динамических стохастических моделей по экспериментальным данным.— М.: Наука, 1983.— 323 с.

94. Кендалл М. Д., Стьюарт А. Многомерный статистический анализ и временные ряды.— М.: Наука, 1976.— 736 с.

95. Ким Дж.-и др. Факторный, дискриминантный и кластерный анализ.— М.: Финансы и статистика.— 1989.— 215 с.

96. Киткин П. А., Колесник Ю. А., Краснопольский Ю. Я. Вероятностные методы решения задач промысловства и оптимального использования биоресурсов.— Владивосток, 1984.— 127 с.

97. Классификация и кластер/Под ред. Дж. Райзина.— М.: Мир, 1980.

98. Клепиков В. В., Лукин В. В. Выделение водных масс и фронтов в Южном океане по совокупности признаков с помощью оптимального цикла//Информ. бюлл. САЭ, 1972.— № 83.— С. 52—56.

99. Клепиков В. В., Смирнов Н. П., Божков А. Т. Выделение и исследование водных масс в Индийском океане с помощью разложения по естественным ортогональным функциям//Вестник ЛГУ, 1974.— 24.— С. 72—81.

100. Климатолого-статистическое исследование термогалинных полей Северной Атлантики/Под ред. А. С. Моница. М.: Гидрометеоиздат, 1982.— 164 с.

101. Клюквин Л. Н. О методике интерполяции океанографических характеристик с помощью ЭВМ//Труды ААНИИ, 1977.— Т. 342.— С. 64—79.

102. Ковалев Е. Г., Николаев Ю. В., Прямыков С. А. Автоматизация поиска оптимальных предсказателей и построения прогностических схем//Труды ААНИИ, 1981.— Т. 372.— С. 44—52.

103. Колесник Ю. Д., Волох Ф. Ф. Применение факторного анализа для изучения распределения, поведения и прогнозирования улова япономорской сардины//Итоги исследований по вопросам рационального использования и охраны биоресурсов Сахалина и Курильских о-вов.— Южно-Сахалинск, 1981.— С. 37—39.

104. Колмогоров А. Н. Теория вероятностей и математическая статистика.— М.: Наука, 1986.— 534 с.

105. Кравчук М. А., Саруханян Э. И. Изучение 14-месячных колебаний уровня Мирового океана с помощью метода главных компонент//Труды ААНИИ, 1983.— Т. 371.— С. 21—30.

106. Кротова В. А., Мезенцова Т. И., Шимараев Н. Н. Эмпирические модели вертикального распределения температуры и плотности воды в Балтийском море глубже 300 м. В кн.: Гидрология Байкала и других водоемов.— Новосибирск: Наука, 1984.— С. 31—40.

107. Куприянова Т. П. Принципы и методы физико-географического районирования с применением ЭВМ.— Казань, 1977.— 125 с.

108. Левасту Т., Ларкинд Г. Морская промысловая система. Количественная оценка параметров и регулирование рыболовства.— М.: Агропромиздат, 1987.— 165 с.

109. Лемешко Е. М. О методе комплексных главных компонент. МГИ АН УССР, Севастополь, 1987.— 19 с. Деп. ВИНТИ 06.02.87. № 873—В87.

110. Лифшиц А. В., Хромов В. М. Исследование взаимодействий в водных экосистемах с помощью информационно-логического анализа//Журн. общей биологии, 1980.— Т. 41.— № 1.— С. 49—55.

111. Ломакин А. Ф., Покудов В. В. Естественные ортогональные функции температуры поверхностного слоя северной части Тихого океана//Труды ДВНИИ, 1984.— Вып. 111.— С. 50—57.
112. Лопатухин Л. И., Рожков В. А., Румянцева С. А. Применение метода главных компонент для районирования морей и океанов по режимным характеристикам//Труды ЛГМИ, 1984.— Вып. 87.— С. 10—19.
113. Максимов И. В. Геофизические силы и воды океана.— Л.: Гидрометеоиздат, 1970.— 448 с.
114. Мещерская А. В., Гирская Э. И. Об интерпретации форм естественных ортогональных функций//Труды ГГО, 1974.— Вып. 298.— С. 90—96.
115. Мещерская А. В. и др. Естественные составляющие метеорологических полей.— Л.: Гидрометеиздат, 1970.— 160 с.
116. Минский Д. С., Чижевский А. М. Быстрая байесовская классификация многозональных изображений//Исследование Земли из космоса, 1982.— № 1.— С. 52—59.
117. Мирвис В. М. К вопросу об объективизации проведения границ в задачах прикладной классификации климатов//Труды ГГО, 1975.— Вып. 330.— С. 156—170.
118. Миронов Е. У. Представление положения кромки льда посредством разложения на естественные ортогональные составляющие//Труды ААНИИ, 1981.— Т. 384.— С. 45—52.
119. Монин А. С., Каменкович В. И., Корт В. Г. Изменчивость Мирового океана.— Л.: Гидрометеиздат, 1974.— 264 с.
120. Монин А. С. Прогноз погоды как задача физики.— М.: Наука, 1969.— 184 с.
121. Мордвинов В. И., Репинская Р. П. Разложение случайных процессов на косоугольные составляющие и возможности его применения в метеорологии//Труды ГГО, 1988.— Вып. 516.— С. 121—127.
122. Мостеллер Ф., Тьюки Дж. Анализ данных и регрессия.— М.: Финансы и статистика.— Вып. 1.— 1981. Вып. 2.— 1982.
123. Налимов В. В. Теория эксперимента.— М.: Наука, 1971.— 207 с.
124. Никифоров С. Г., Шпайхер А. О. Закономерности формирования крупномасштабных аномалий режима Северного Ледовитого океана.— Л.: Гидрометеиздат, 1980.— 269 с.
125. Николаев Ю. В. Классификация гидрометеорологических процессов с помощью ЭВМ.— Л.: Гидрометеиздат, 1976.— 35 с.
126. Николаев Ю. В. Модификация метода разложения по естественным ортогональным составляющим//Труды ААНИИ, 1975.— Т. 321.
127. Николаев Ю. В. Применение дискриминантного анализа для долгосрочных прогнозов льдов арктических морей//Труды ААНИИ, 1976.— Т. 320.— С. 4—26.
128. Николаев Ю. В. Применение метода главных компонент в изучение многолетних колебаний ледовитости арктических морей//Труды ААНИИ, 1973.— Т. 307.
129. Николаев Ю. В., Ковалев Е. Г. К вопросу применения дискриминантного анализа для долгосрочного прогноза ледовитости//Труды ААНИИ, 1971.— Т. 303.— С. 14—26.
130. Николаев Ю. В., Ленин А. И. Применение теории распознавания образов в океанологии//Киев: Наукова Думка, 1973.— 106 с.
131. Николаев Ю. В., Смирнов Н. П. Опыт применения дискриминантного анализа для прогноза годового стока Верхней Волги//Информ. Бюлл. ИБВВ АН СССР, 1972.— № 12.— С. 67—73.
132. Николаев Ю. В., Саруханян Э. И., Смирнов Н. П. Применение дискриминантного анализа при определении различий между классами//Информ. бюлл. ИБВВ АН СССР.— 1970.— № 8.— С. 18—25.
133. Обухов А. М. Статистическое описание непрерывных полей//Труды Геофизического ин-та АН СССР, 1954.— Вып. 24 (151).— С. 3—42.
134. Оганесян В. Б. Каноническая корреляция температурных полей атмосферы и океана//Метеорология и гидрология.— 1978.— № 2.— С. 42—51.

135. Орлов Н. Ф. Метод долгосрочного прогноза среднемесячной ледовитости районов северо-западной части Атлантического океана//Труды ААНИИ, 1983.— Т. 385.— С. 46—49.
136. Островский А. Г., Питербарг Л. И. О статистической предсказуемости температуры поверхности океана//Изв. АН СССР. ФАО.— 1986.— № 4.
137. Островский А. Г., Питербарг Л. И. Динамико-стохастическая модель долгопериодных колебаний уровня моря//Океанология, 1988.— Т. 28.— № 6.— С. 918—924.
138. Перфилов В. И. О собственных значениях выборочной корреляционной матрицы метеорологических полей//Труды Гидрометцентра СССР, 1982.— Вып. 243.— С. 114—119.
139. Песаран М., Слейтер Л. Динамическая регрессия. — М.: Финансы и статистика, 1984.— 310 с.
140. Петрович М. Л. Регрессионный анализ и его математическое обеспечение на ЕС ЭВМ.— М.: Финансы и статистика, 1982.— 199 с.
141. Перевозчикова О. И. Пространственно-временной анализ изменения теплосодержания деятельного слоя Северной Атлантики//Труды Гидрометцентра СССР, 1985.— Вып. 269.— С. 89—100.
142. Питербарг Л. И. Динамика и прогноз крупномасштабных аномалий температуры поверхности океана.— Л.: Гидрометеиздат, 1989.— 198 с.
143. Питербарг Л. И., Островский А. Г. Динамическая стохастическая модель долгопериодной изменчивости температуры поверхности океана//Докл. АН СССР.— 1984.— Т. 276.— № 6.— С. 1469—1470.
144. Плюта В. Сравнительный многомерный анализ в экономическом моделировании.— М.: Финансы и статистика.— 1989.— 175 с.
145. Поздынин В. Д. Возможности байесовских методов в обработке результатов гидрофизических измерений//Океанология, 1983.— Т. 23.— № 4.— С. 671—676.
146. Поляк И. И. Многомерные статистические модели климата.— Л.: Гидрометеиздат, 1989.— 183 с.
147. Поляков В. М. Применение факторного анализа при обработке данных дистанционного зондирования окружающей среды.— В кн.: Статистические методы обработки данных дистанционного зондирования окружающей среды. Труды I Межведом. совещ.— М.: Изд. ИРЭ АН СССР, 1983.— С. 109—112.
148. Поляков Г. Г., Хайбуллина Л. С. О прогнозировании полей с использованием канонических корреляций//Труды ЗапСибНИИ, 1984.— Вып. 66.— С. 85—99.
149. Покровский Б. Н., Мартышенко С. Н. Применение метода многомерной статистической классификации при изучении гидрофизических полей.— В кн.: Автоматизация океанологических исследований.— Владивосток, 1981.— С. 103—106.
150. Праги У. Факторный анализ как метод решения проблемы классификации в географии.— В кн.: Математико-географическое моделирование территориальных систем.— Казань: Изд-во КГУ, 1984.— С. 47—54.
151. Привальский В. Е. Климатическая изменчивость (стохастические модели, предсказуемость, спектр).— М.: Наука, 1985.— 182 с.
152. Прошутинский А. Ю., Уранов Е. Н. Комплексный метод прогноза сгонно-нагонных колебаний уровня на устьевом взморье Енисея в зимний период — заблаговременностью 2—3 суток//Труды ААНИИ, 1985.— Т. 389.— С. 78—85.
153. Рабочая книга по прогнозированию.— М.: Мысль, 1982.— 423 с.
154. Репинская Р. П., Скляренко В. Л., Смирнов Н. П. Прогноз поля давления экстраполированием параметров разложения по е. о. ф.//Труды ЛГМИ, 1975.— Вып. 53.— С. 124—133.
155. Рожков В. А., Трапезников Ю. А. К вопросу о построении вероятностных моделей океанологических процессов//Труды ГОИН, 1973.— Вып. 169.— С. 46—60.

156. Романов Л. Н., Шарифуллин Н. К. Об уточнении границ классов при альтернативном прогнозировании//Метеорология и гидрология, 1983.— № 4.— С. 40—47.
157. Романов Л. Н. и др. О выборе параметров в регрессионных моделях с помощью скользящего контроля//Тр. ЗапСибНИИ.— 1980.— Вып. 46.
158. Романов Л. Н. и др. О выборе параметров для построения регрессионных моделей//Метеорология и гидрология, 1980.— № 7.
159. Романцов В. А., Буб А. Ф. Некоторые результаты многомерного анализа водных масс Норвежского и Гренландского морей//Труды ААНИИ, 1977.— Т. 342.— С. 24—37.
160. Романцов В. А., Смирнов Н. П. О тепловом состоянии вод Норвежского моря//Труды ААНИИ, 1983.— Т. 382.— С. 101.
161. Русанов В. П. О классификации вод Арктического бассейна по гидрохимическим признакам//Труды ААНИИ, 1984.— Т. 368.— С. 5—21.
162. Саруханян Э. И., Смирнов Н. П. Водные массы и циркуляция Южного океана.— Л.: Гидрометеиздат, 1986.— 288 с.
163. Себер Дж. Линейный регрессионный анализ.— М.: Статистика, 1980.— 456 с.
164. Семенов Ю. Л. Факторный анализ общего природного процесса, формирующего гидролого-гидрохимический режим восточной части Каспия.— В кн.: Вопросы промысловой океанологии Мирового океана.— Мурманск, 1977.— С. 100—101.
165. Семенов Ю. Л. Эмпирическая модель гидрохимических условий восточной части среднего и Южного Каспия на основе факторного анализа//Гидрохимические материалы, 1986.— Т. 94.— С. 90—103.
166. Семовский С. В., Белкин И. М. Климатолого-статистическая модель восстановления вертикальных профилей гидрофизических параметров//Докл. АН СССР, 1986.— Т. 287, № 1.— С. 208—210.
167. Симонов А. И. и др. Химико-океанографическое районирование и гидрохимический режим Мирового океана//Докл. АН СССР, 1988.— Т. 299, № 4.— С. 981—985.
168. Скляренок В. Л., Смирнов Н. П. О применении многомерного анализа в гидрологии//Труды ИБВВ АН СССР, 1974.— Вып. 26 (29).— С. 180—206.
169. Скляренок В. Л., Смирнов Н. П. О формальных и неформальных аспектах применения многомерного анализа в гидрометеорологии//Труды ААНИИ, 1977.— Т. 342.— С. 5—23.
170. Смирнов Н. П., Вайновский П. А. Исследование структуры гидрофизических полей методом факторного анализа//Метеорология и гидрология, 1989. № 1.— С. 101—103.
171. Смирнов Н. П., Кондратьев А. В., Вайновский П. А. Статистическое моделирование и предсказуемость теплового состояния верхнего слоя Норвежской ЭАЗО/Тез. докл. III Съезда советских океанологов. Физика и химия океанов.— Л.: Гидрометеиздат, 1987.— С. 68—70.
172. Смирнов Н. П., Скляренок В. Л. Классификация в гидрологии как задача теории распознавания//Труды ИБВВ АН СССР, 1974.— Вып. 26 (29).— С. 166—180.
173. Смирнов Н. П., Скляренок В. Л. Методы многомерного статистического анализа в гидрологических исследованиях.— Л.: Изд. ЛГУ, 1986.— 190 с.
174. Смирнова Н. Ф. Пространственно-временное распределение зоопланктона в Норвежском и Баренцевом морях//Экология, биологическая продуктивность и проблемы марикультуры Баренцева моря. Тез. докл. II Всесоюзн. конф.— Мурманск, 1988.— С. 83—84.
175. Спичкин В. А., Кириллов А. А. Определение средней сезонной характеристики ледовых условий методом главных компонент//Труды ААНИИ, 1977.— 346.— С. 135—142.
176. Статистические методы анализа климатологической информации. Факторный анализ//Обзорная информ. ВНИИГМИ—МЦД, 1984.— Вып. 2.— 36 с.
177. Статистические методы для ЭВМ/Пер. с англ.— М.: Наука, 1986.— 464 с.

178. Степанов В. И. Многомерный анализ водных масс северо-западной части Тихого океана методом потенциальных функций//Труды ВНИИГМИ—МЦД, 1984.— 101.— С. 27—40.
179. Степанов В. Н. Океаносфера. М.: Мысль, 1983.— 270 с.
180. Тер-Мкртчян. О применении дискриминантного анализа для улучшения статистических прогнозов по методу множественной регрессии//Труды Гидрометцентра, 1970.— Вып. 64.— С. 130—139.
181. Титов Ю. Э. Об использовании метода разложения рядов гидрометеорологических элементов по естественным ортогональным функциям времени//Научн. труды (межвузовские): Исследование и освоение Мирового океана.— Л.: Изд-во ЛПИ, 1981.— Вып. 77.— С. 46—50.
182. Трофимов А. М. Математические методы и физической географии.— Казань: Изд-во КГУ.— 1977.
183. Трусенков С. Г. Климатическая пространственная структура полей температуры поверхности северной части Тихого океана/Препринт ИАПУ ДВНЦ АН СССР.— Владивосток, 1986.— 30 с.
184. Трусенков С. Т. Эмпирические ортогональные функции температуры поверхности северной части Тихого океана/Препринт ИАПУ ДВНЦ АН СССР.— Владивосток, 1988.— 24 с.
185. Трусенков С. Т., Трусенкова О. О. Анализ геофизических полей на основе многомерных статистических методов/Препр. ИАПУ ДВНЦ АН СССР.— Владивосток, 1984.— 45 с.
196. Тьюки Дж. Анализ результатов наблюдений: разведочный анализ.— М.: Мир, 1981.— 693 с.
197. Угрюмов А. И. Тепловой режим океана и долгосрочные прогнозы погоды.— Л.: Гидрометеиздат, 1970.— 176 с.
198. Урбах В. Ю. Дискриминантный анализ: основные идеи и приложения//Статистические методы классификации, 1969.— Вып. 1.— С. 79—174.
199. Федоров К. Н. Физическая природа и структура океанических фронтов.— Л.: Гидрометеиздат, 1983.— 296 с.
200. Фёрстер Э., Рёнци Б. Методы корреляционного и регрессионного анализа.— М.: Наука, 1983.— 302 с.
201. Филюшкин Ю. Б. Представление полей температуры и солёности с помощью статистически ортогональных функций//Океанология, 1984.— Т. 24.— Вып. 5.— С. 756—763.
202. Фортус М. И. Метод эмпирических ортогональных функций и его применение в метеорологии//Метеорология и гидрология, 1980.— № 4.— С. 113—119.
203. Хайбуллина Л. С., Еремин В. В., Аксенова Е. Н. Анализ статистической структуры поля температуры воздуха на основе метода канонических корреляций//Труды ЗапСибНИИ, 1984.— Вып. 64.— С. 23—27.
204. Харман Г. Современный факторный анализ.— М.: Статистика, 1972.— 486 с.
205. Хаустов А. П. Факторная модель формирования подземного стока в бассейне оз. Байкал.— В кн.: Гидрология Байкала и других водоемов.— Новосибирск: Наука, 1984. С. 129—138.
206. Хей Дж. Введение в методы байесовского статистического вывода.— М.: Финансы и статистика, 1987.— 335 с.
207. Цехоцкая Л. К., Злобин В. С. Опыт применения метода разложения океанологических полей по е. о. ф. к анализу водных масс Баренцева моря//Труды ПИНРО, 1977.— Вып. 38.— С. 45—57.
208. Шайхмейстер В. А. Реализация на ЭВМ некоторых алгоритмов по дискриминантному анализу//Труды ГГО, 1983.— Вып. 308.
209. Ширяев Е. Е. Новые методы картографического отображения и анализа геоинформации с применением ЭВМ.— М.: Недра, 1977.— 183 с.
210. Эфрон Б. Нетрадиционные методы многомерного статистического анализа.— М.: Финансы и статистика.— 1988.— 262 с.
211. Яглом А. М. Корреляционная теория стационарных случайных функций.— Л.: Гидрометеиздат, 1981.— 280 с.
212. Яковлев В. Н. Гидрометеорологическое обеспечение океанического рыболовства.— М.: Пищевая промышленность, 1976.— 229 с.

213. Яковлев В. Н. Классификация гидрометеорологических полей Северной Атлантики.— В сб.: Океанографические основы биологической продуктивности океана.— Калининград, 1983.— С. 141—142.
214. Яковлев А. А. О расчете естественных ортогональных составляющих при малых выборках//Изв. АН СССР, ФАО, 1979.— № 3.— С. 338—40.
215. Яковлев В. Н., Альтман Ю. С. Классификация гидрометеорологических элементов в море Скотия.— В сб.: Океанологические факторы в промысловом прогнозировании.— Калининград, 1984.— С. 61—66.
216. Яковлев В. Н., Альтман Ю. С. Методические рекомендации по прогнозированию промыслово-океанологических характеристик некоторыми статистико-вероятностными методами.— Калининград: АтлантНИРО, 1985.— 94 с.
217. Яковлев В. Н., Альтман Ю. С. Типизация гидрометеорологических элементов в Северной Атлантике. В сб.: Океанологические основы формирования биологической продуктивности Северной Атлантики.— Калининград, 1981.— С. 3—40.
218. Barnett T. P. Statistical relations between ocean—atmosphere fluctuations in the tropical Pacific//J. Phys. Ocean, 1982.— V. 11.— N 8.— P. 1043—1057.
219. Barnett T. P. Significance tests for regression model//J. Phys. Oceanogr, 1981.— V. 11.— N 8.— P. 1238—1245.
220. Barring L. Regionalization of daily rainfall in Kenia by means of common factor analysis//J. Climatol, 1988.— V. 8.— N 4. P. 371—389.
221. Bennet A. F., Buldol W. P. Ocean data assimilation and the Kalman filter: spatial regularity//J. Phys. Ocean, 1987.— V. 17.— N. 10.— P. 1583—1601.
222. Borkowski M. R., Goulet T. Comparison of methods for interpolation oceanographic data//Deep Sea Res, 1971.— V. 18.— N 2.— P. 269—274.
223. Christenson W. J., Bryson R. A. An investigation of the potential of component analysis for weather classification//Mon. Wea. Rev., 1966.— V. 94.— P. 697—704.
224. Coulter R. F. Application of the bayes decision rules/Proc. 17th Int. Symp. Remote sens. Environ. Amsterdam, 1983.— V. 2.— P. 589—597.
225. Davis R. E. Techniques for statistical analysis and prediction of geophysical fluid systems//Geophys. astrophys. Fluid Dynamics, 1977.— V. 8.— P. 245—277.
226. Doeffler R. Factor analysis in ocean colour interpretation//Oceanogr. Spac. Proc. COSPAR/SCOR/IUCRAM Symp.— New York, 1981.— P. 339—345.
227. Duckstein H., Bobec B., Bogardi I. Bayesian forecasting of hydrologic variables under changing climatology/The Influence of climate change and climat variability on the hydrologic regime and water resources.— IAHS Publ., 1987.— N 168.— P. 301—311.
228. Fleming H. E., Hill M. L. An objective procedure for detecting and correcting errors in geophysic data//J. Geoph. Res., 1982.— C. 87, N 9.— P. 7312—7324.
229. Flos J. Ordination and cluster analysis applied to oceanographic data//Estuarine and coast. Marine Scia., 1980.— 11.— N 4.— P. 393—406.
230. Glahn H. R. Canonical correlation and its relationship to discriminant analysis and multiple regression//J. Atmos. Sci., 1988.— V. 25.— P. 2—23.
231. Graham N., Michaelson J., Barnett F. P. An investigation of the El-Niño—South Oscillation cycle with statistical models//J. Geoph. Res., 1987.— C 92.— N 13.— P. 14251—14270.
232. Gonella J. A rotary-component method for analyzing meteorological and oceanographic vector time series//Deep-Sea Res., 1972.— V. 19.— P. 833—846.
233. Hannes G. Factor analysis of coastal air pressure and water temperature//J. Appl. Meteor, 1976.— V. 15.— P. 120—126.
234. Hasselmann K. Stochastic climate models. Part. 1./Tellus, 1976.— V. 28. P. 473—485.
235. Horel J. D. Complex principal component analysis: theory and examples//J. Clim. Appl. Meteor., 1984.— 23.— P. 1660—1673.

236. Horel J. D., Wallace J. M. Planetary scale atmospheric phenomenon associated with the southern oscillation//*Mon. Wea. Rev.*, 1981.— V. 109.— P. 813—829.
237. Hsiung J., Newell R. The principal nonseasonal modes of variation of Global sea surface temperature//*J. Phys. Ocean.*, 1983.— V. 13, N 10.— P. 1957—1967.
238. Johnston C. M., Lemke P., Barnett T. P. Linear prediction of sea ice anomalies//*J. Geoph. Res.*, 1985.— V. 90.— N D3.— P. 5665—5675.
239. Johnston R. J. Regarding the delimitation of regions according to climatic fluctuation//*Arch. Met. Geoph. Bioclim.*, 1981.— Ser. B.— V. 29, N 3.
240. Karl T. R., Koscielny A. J., Diaz H. F. Potential errors in the application of principal components analysis to geophysical data//*J. Appl. Meteor.*, 1982.— V. 21.— P. 1183.
241. Kutzbach J. E. Empirical eigenvectors of sea-level pressure surface temperature and precipitation complexes over North America//*J. Appl. Meteor.*, 1976.— Vol. 6, N 5.— P. 791—802.
242. Li Fengi, Su Yusong, Yu Zuxiang. Application of cluster analysis method to modified water masses in shallow sea//*Acta ocean. sin.*, 1984.— 3, N 4.— P. 451—461.
243. Lopatukhin L. J., Rozhkov V. A. The Baltic sea zoning by means of wind and wave probability characteristics//*Proceed. of XIII Conf. of the Baltic Oceanogr.*— Helsinki, 1982.— V. 2.— P. 522—534.
244. Lorenz E. Empirical orthogonal functions and statistical weather prediction//*Statistical Forecasting Project.*— Scie. Rep. N 1, MIT.— 1956.— 48 p.
245. Makriakis S., Winkler R. L. Averages of Forecasts: Some empirical results//*Management Sci.*, 1983.— Vol. 29.— N 9.— P. 987—996.
246. Maves J. L. Application d'un modele chaine Markov ala prevision du temps a longue echeance//*Meteorology and Hydrology*, 1985.— N 1—2.— P. 41—48.
247. McBoyle G. R. Factor analysis approach to a climatic classification of Europe//*Climatol. Bull.*— 1972.— N. 12.— P. 1—11.
248. Michaelsen J. The estimation of artificial skill in forecast models//*Ninth Conf. on Probabil. and Statist.*, Amer. Meteor. Soc.— Boston.— 1985.— P. 247—251.
249. Michaelsen J. Long—period variability in North Pacific sea surface temperature anomalies//*J. Phys. Oceanogr.*— 1982.— 12, N 7.— P. 694—703.
250. Miller R. N. Toward the application of the Kalman filter to regional open ocean modeling//*J. of Physic Oceanogr.*— Vol. 16, 1986.— P. 72—86.
251. Morin G., Fortin J. P., Sochanska W., Lardeau J. P. Use of principal component analysis to identify homogeneous precipitation stations for optimal interpolation//*Water Resources Res.*— 1979.— Vol. 15.— P. 1841.
252. Murthy H. A., Haykin S. Bayesian classification of surface—based ice—radar images//*IEEE J. Ocean. Eng.*— 1957, 12, N 3.— P. 493—502.
253. Neumann J. M. Canonical correlation applications//*U. S. Air Weather Service, Techn. Rep.*, 1977, N 7.— P. 6.1—6.14.
254. Nichol D. G. Autonomous Extraction of an Eddy-like structure from inferred images of the Ocean//*IEEE Transact. on Gloscie, and Remote sens.*— 1987.— V. GE25, N 1.— P. 28—34.
255. Nielsen P. B. Orthogonal representation of oceanographic data//*Rept. Inst. fis. oceanogr. Kopenhagen. Univ.*— 1980.— 42.— P. 129—144.
256. North G. R., et al. Sampling errors in the estimation of empirical orthogonal functions//*Mon. Wea. Rev.*— 1982.— 110.— P. 72—82.
257. Perry A. H. Filtering climatic anomalies fields using principal component analysis//*Transact. Inst. Brit. Geographes.*— 1970.— 50.— P. 55—72.
258. Preisendorfer R. W., Barnett T. P. Significance tests for empirical orthogonal functions.— Fifth conference on probability and statistics in atmospheric sciences, Las Vegas, Nevada, November 15.18, 1977.— P. 169—172.
259. Preisendorfer R. W., Mobley C. D., Barnett T. R. The principal discriminant method of prediction: theory and evaluation//*Journ. Geoph. Res. D.*, 1988.— Vol. 93, N 9.— P. 10815—10830.

260. Qiu D., Zhou S., Zi C. Application of cluster analysis method in determining water masses of the Huanghai sea//Acta ocean. sin. 1985.— Vol. 4, N 3.— P. 337—348.
261. Reynolds R. W., Gemmill W. H. An objective monthly mean sea surface temperature analysis//Tropical ocean—atmos. news—lett. 1984, N 23.— P. 4—5.
262. Richman M. B. Rotation of principal components//J. Climatol.— 1986.— 6, N 3.— P. 293—335.
263. Richman M. B. Specification of complex modes of circulation with T-mode factor analysis/Second Internat. Met. on Stat. Climatol. Lisboa, Portugal, 1983, P. 5.1.1—5.1.8.
264. Rinnie J., Jarvenoja S. A. A rapid method of computing empirical orthogonal functions from large dataset//Mon. wea. Rev.— 1986.— 114, N 12.— P. 2571—2577.
265. Rothrock D. A., Thomas D. R., Thorndike A. S. Principal component analysis of satellite passive microwave data over sea ice//J. Geoph. Res.— 1988.— C 93, N 3.— P. 2321—2332.
266. Servain J., Legler D. M. Empirical orthogonal function analysis of tropical Atlantic sea surface temperature and wind stress 1964—1979//J. Geoph. Res.— 1986.— 91.— P. 14181—14191.
267. Solow A. R. A Bayesian approach to statistical inference about climate change//J. of climate Amer. Meteorol. Society, 1988.— Vol. 1, P. 512—521.
268. Stundev M. The potential for expert system development in meteorology//Research Associate Artificial Intelligence Branch GEOMET Technologies, Inc. Germontowa, Maryland, 1987. P. 1—2.
269. Time series in the time domain. Handbook of statistics vol. 5//Hannan J., Krishnaiah P. R., Rao M. M. Amsterdam, North—Holland, 1985.— 490 p.
270. Wallace J. M., Dickinson R. E. Empirical orthogonal representation of time series in frequency domain//J. Appl. Meteor.— 1972.— 11.— P. 887—892.
271. Weare B. C. Interannual variation in net heating at the surface of the tropical Pacific Ocean//J. Phys. Ocean.— 1983.— V. 13, N 5.— P. 873—885.
272. Weare B. C. Empirical orthogonal analysis of Atlantic Ocean surface temperature//Quart. J. Roy. Meteorol. Soc.— 1977, 103, N 437.— P. 467—478.
273. White W. E. West—ward propagation of short—term climatic anomalies in the western North—Pacific ocean from 1964 to 1974//J. Marine Res.— 1981.— V. 41, N 1.— P. 113—125.
274. Weare B. C., Nasstrom J. S. Examples of extend empirical orthogonal function analysis//Mon. Wea. Rev.— 1982.— 110.— P. 481—485.
275. Weare B. C., Navato A. R., Newell R. E. Empirical orthogonal analysis of Pacific ocean surface temperature//J. Phys. oceanogr.— 1976.— 6.— P. 671—678.
276. Wieuwer N. Extrapolation, interpolation, and smoothing of stationary time series//Cambridge, Mass: MIT Press, 1949.— 163 p.
277. Worthington L. V. The Norwegian sea as a mediterranean basin//Deep Sea Res.— 1970.— 17.— P. 77—84.
278. Xie Simei, et al. Characteristics on interannual variations of North Pacific sea surface temperature and its relation to East Asia Climate//Acta oceanol. sin.— 1987.— 6.— N 1.— P. 69—79.
279. Yao Dirong. The nonlinear discriminant and stepwise nonlinear discriminant analyses//Adv. Atmos. Sci.— 1988.— 5.— N 1.— P. 27—31.

ОГЛАВЛЕНИЕ

Предисловие	3
Введение	7
Глава 1. Описание взаимосвязей процессов	31
1.1. Множественный регрессионный анализ. Множественная линейная регрессия	—
1.2. Метод канонического корреляционного анализа	41
Глава 2. Описание структуры процессов	47
2.1. Метод главных компонент	—
2.2. Метод факторного анализа	55
Глава 3. Типизация процессов	67
3.1. Метод автоматической классификации	—
3.2. Метод дискриминантного анализа	78
Глава 4. Прогностическое содержание методов МСА	87
4.1. Предсказание структуры процессов	—
4.2. Предсказание классов состояния процессов	91
4.3. Комплексование прогностических моделей	95
4.4. Статистическая предсказуемость океанологических процессов	97
Глава 5. Результаты статистического диагноза океанологических процессов	104
5.1. Диагноз тепловых полей Северной Атлантики	—
5.2. Диагноз структуры вод Норвежского моря	120
5.3. Диагноз структуры водных масс моря Беллинсгаузена	138
Глава 6. Результаты статистического прогнозирования океанологических процессов	148
6.1. Прогнозирование тепловых полей Северной Атлантики	—
6.2. Прогнозирование ледовых условий в Баренцевом море	170
6.3. Прогнозирование пространственного распределения сестона в Норвежском море	178
Заключение	185
Список литературы	187

Монография

Смирнов Николай Павлович
Вайновский Павел Антонович
Титов Юрий Эдуардович

**СТАТИСТИЧЕСКИЙ ДИАГНОЗ И ПРОГНОЗ
ОКЕАНОЛОГИЧЕСКИХ ПРОЦЕССОВ**

Редактор О. В. Лапина. Художник А. В. Васильев. Художественный редактор Б. А. Бураков.
Технический редактор Н. В. Морозова. Корректор Э. Э. Белякова.

ИБ № 1961

Сдано в набор 05.02.92. Подписано в печать 01.06.92. Формат 60×90^{1/16}. Бумага офсетная.
Гарнитура литературная. Печать высокая. Печ. л. 12,5. Кр.-отт. 12,5. Уч.-изд. л. 13,76.
Тираж 400 экз. Индекс ОЛ-99. Заказ № 20.
Гидрометеиздат, 199397, Санкт-Петербург, ул. Беринга, 38.

Типография № 8 ордена Трудового Красного Знамени ГПО «Техническая книга»
Мининформпечати РФ 190000, г. Санкт-Петербург, Прачечный переулок, 6.

