

A 1981-1

АКАДЕМИЯ НАУК ЛИТОВСКОЙ ССР

ИНСТИТУТ БОТАНИКИ

ДЕНДРОКЛИМАТОХРОНОЛОГИЧЕСКАЯ ЛАБОРАТОРИЯ

Тема: Изучение хвойных лесов, путем построения дендрохронологических профилей, с целью ретроспективного восстановления изменчивости биологических условий среды в последних столетиях на территории СССР

Руководитель: Зав. лабораторией

Т.Битвинскас

Исполнители: М.Н.С. И.Капraitис

М.Н.С. И.Карпавичюс

Начато: 1979 г.

Окончено: 1981 г.

Кол. листов

Каунас - 1981

LTSR MOKSLŲ AKADEMIJA BOTANIKOS INSTITUTAS

DENDROKLIMATOCHRONOLOGINĖ LABORATORIJA

Temas: Spygliuočių miškų tyrimai sudarant dendrochronologinius profilius tikslu retrospektyviai atstatyti bioekologinių aplinkos sąlygų pakitimus pastaraisiais šimtmečiais TSRS teritorijoje

Mokslinis vadovas: laboratorijos vadovas T. Bitvinskas

Vykdytojai: j.m.b. J. Kairaitis
j.m.b. J. Kerpevičius

Predėta: 1979 m.

Baigta: 1981 m.

Lapų skaičius: 166

Kaunas, 1981

АКАДЕМИЯ НАУК ЛИТОВСКОЙ ССР

УДК

№ гос.регистрации

Инв. №

"Утверждаю"

Директор Института ботаники
АН Литовской ССР

академик А.Н.Меркис

25 декабря 1981 г.

Изучение хвойных лесов, путем построения дендрохронологических профилей, с целью ретроспективного восстановления изменчивости биоэкологических условий среды в последних столетиях на территории СССР

Тема №

Зам.директора по научной
работе, к.б.н.

В.И.Марчоленис
25 декабря 1981 г.

Заведующий лабораторией
дендроклиматохронологии,
к.сх.н.

Т.Т.Битвинскас
25 декабря 1981 г.

Каунас - 1981

Список исполнителей

1.	Т.Битвинскас	ст.н.с.	руководитель
2.	И.Кайрайтис	м.н.с.	исполнитель
3.	И.Карпавичюс	м.н.с.	исполнитель
4.	А.Даукантас	с.и.	исполнитель
5.	Н.Кряучёните	с.и.	исполнитель
6.	А.Семашка	р.г.	исполнитель
7.	В.Бальчионас	с.и.	исполнитель
8.	А.Зокайтис	с.и.	исполнитель
9.	В.Брукштус	с.и.	исполнитель
10.	А.Ступнева	стажер	
11.	А.Барткевичюс	аспирант	

1.0. РЕФЕРАТ

1979–1981 г.г. дендроклиматохронологической лабораторией исполнялась тема "Изучение хвойных лесов путем построения дендрохронологических профилей с целью ретроспективного восстановления изменчивости биоэкологических условий среды в последних столетиях на территории СССР"

Для этой цели использовались дендрошкалы профиля Мурманская область – Закарпатье – 44 пункта исследований (Дендрошкалы профиля опубликованы в первой части сборника Дендроклиматологические шкалы Советского Союза стр.52–78). Материалы профиля служили для получения ряда исследований и выводов Т.Т.Битвинскаса и А.В.Ступневой по изучению связей ширины годичных колец с климатическими факторами и солнечной активностью, а также о закономерностях ритмики радиального прироста сосны в различных участках профиля. На примере изучения профиля предложен математический аппарат изучения дендрохронологической информации для исследований закономерностей изменчивости условий среды (А.В.Ступнева 1981), методика фильтрации временных рядов и дендроклиматохронологии (А.В.Ступнева 1981б), установлены пространственно–временная изменчивость дендрорядов профиля (А.В.Ступнева, Т.Т.Битвинскас 1978, А.Ступнева 1981в), связи радиального прироста с солнечной активностью Т.Битвинскас (1978), В.Дергачев, Г.Кочаров, А.Ступнева 1978).

Завершен сбор дендрохронологического материала по профилю Литва–Дальний Восток (кроме части Читинской и Амурской областей). В экспедициях (И.Кайрайтис, А.Даукантас, А.Бальчюнас и др.) были заложены 74 пункта исследований. На каждом пункте исследований брались от 25 до 75 образцов древесины. Крупной, почти самостоятельной работой, является накопление климатологических данных по данному профилю. Экологические выводы и связи прироста насаждений с факторами среды будут получены в следующем трехлетии. Длина получаемых рядов по профилю от 100 до 400 лет.

Самостоятельной частью исследований является изучение радиоактивного углерода годичных колец и изучение их связей с астрофизическими явлениями.

За последние три года дендроклиматологическая лаборатория накопила новые материалы с Кавказа, Башкирии и Новгорода, предоставила для проблемы "Астрофизические явления и радиоуглерод", материал 1300 образцов точно датированных годовичных кольца. Серия ряда древесины древнего Новгорода предоставлена до 1204 года. Получены выводы о связи содержания ^{14}C в годовичных кольцах и солнечной активности 11-летней и 80-летней цикличности. Установлено увеличение концентрации ^{14}C в годовичных кольцах сосны обыкновенной в эпоху Маудеровского минимума и т.п.

Разработана новая аппаратура для радиоуглеродных исследований, позволяющая учесть большую часть внешних воздействий в эксперименте с более высокой стабильностью и точностью.

Параллельно с задачами повышения точности датирования радиоуглеродным методом решались такие задачи создания сверхдлинных шкал по болотной сосне. Здесь главной задачей является разработка методологии синхронизации дендрохронологических рядов. Для этого разработан комплекс методов позволяющий уточнять и удачно синхронизировать исследуемые образцы древесины:

- 1) Методика расчета годовичных индексов древесины на ЭВМ БЭСМ-6,
- 2) Методика расчета годовичных индексов древесины на ЭВМ Наир-3,
- 3) Методика использования расчета процента сходства дендрохронологических рядов для ЭВМ БЭСМ-6 и Наир-3,
- 4) Методика расчета корреляционных расчетов коэффициентов для ЭВМ БЭСМ-6 и Наир-3,
- 5) Методика определения реперных годов для синхронизации дендрохронологических материалов,
- 6) Применение палеоботанических методов - определение степени разложения ботанического состава, пыльцевого состава торфа.

Эксперимент проведен на материалах торфяника Ужпелю Тирелис и частично - на материалах торфяника Аукштасис Тирас.

Аналогическая работа проводится по ископаемым Сморгонским дубам, хотя работа замедлилась из-за отсутствия новых образцов древесины - разработки песчано-гравийных сморгонских карьеров на рке Нерис (Вилия) в последние годы не дают новой древесины - более глубоких слоев речных залежей древесины нет. В настоящее время наиболее представительных три последних тысячелетия, имеющие почти сплошные перекрывающие ряды дуба, а также пятое и час-

тично шестое тысячелетие.

Проблема глобальных дендроклиматологических исследований поставлена рядом международных дендрохронологических совещаний: Лондонским 1979 г., Ливерпульским 1980 г., Вильнюсским 1981 г. В ряде стран Западной Европы, в Соединенных Штатах Америки, в Литве, в Латвии дендроклиматохронологические исследования уже достаточно развиты и могут служить основой для глобальных пространственно-временных дендроклиматологических и экологических исследований

Т.Битвинскасом были подсчитаны числа необходимых элементарных дендрохронологических исследований для первичных глобальных исследований по Евроазиатскому, Североамериканскому континентам и в частности и отдельным регионам. Пункты исследований были "привязаны" к сети географических долгот и широт. В зависимости от изученности краев, состояния научных групп они дифференцированы.

В лаборатории создана и испытана автоматизированная линия измерения годовых колец. От аналогичных приборов она отличается достаточной точностью измерений (0,01 мм), специализированной направленностью (работает на языке Наир-3), возможностью измерять не только цилиндрики (керна), но и образцы большим диаметром, управляемостью подвижного столика, реверсивностью оригинального счетчика. Также лабораторией осваивается анализатор слоистых структур (создан Красноярским институтом им. Л.В.Киренского СО АН СССР) и рентгенический денситометр (создан объединением "Спектр"). Эти оба прибора позволяют более глубоко изучать структуру древесного кольца и его закономерности, зависящие от изменчивости экологической среды.

2.0. Введение

В теме "Изучение хвойных лесов путем построения дендрохронологических профилей" используем многие материалы собранные лабораторией ДКХ в прошлые этапы работы, часть материалов получены в данном этапе (1979, 1981) и, конечно, дендрохронологическая информация в этих случаях полностью использована быть не могла.

Наиболее широко изучены материалы дендрохронологического профиля Мурманск-Карпаты. Часть интересных выводов, касающихся связей прироста лесных насаждений сосны и солнечной активности показан в отчетах 1976-1978 г.г.^{х)} Тогда был показан закон сокращения внутривидовых ритмов с севера на юг (А.Ступнева). Тогда было доказана необходимость пользоваться статистически обоснованным числом учетных деревьев и о невозможности доверять дендрорядам построенным с отдельных деревьев (И.Карпавичюс).

Была изучена и углублена теория линейности связи радиального прироста с амплитудами солнечной активности (Т.Битвинскас).

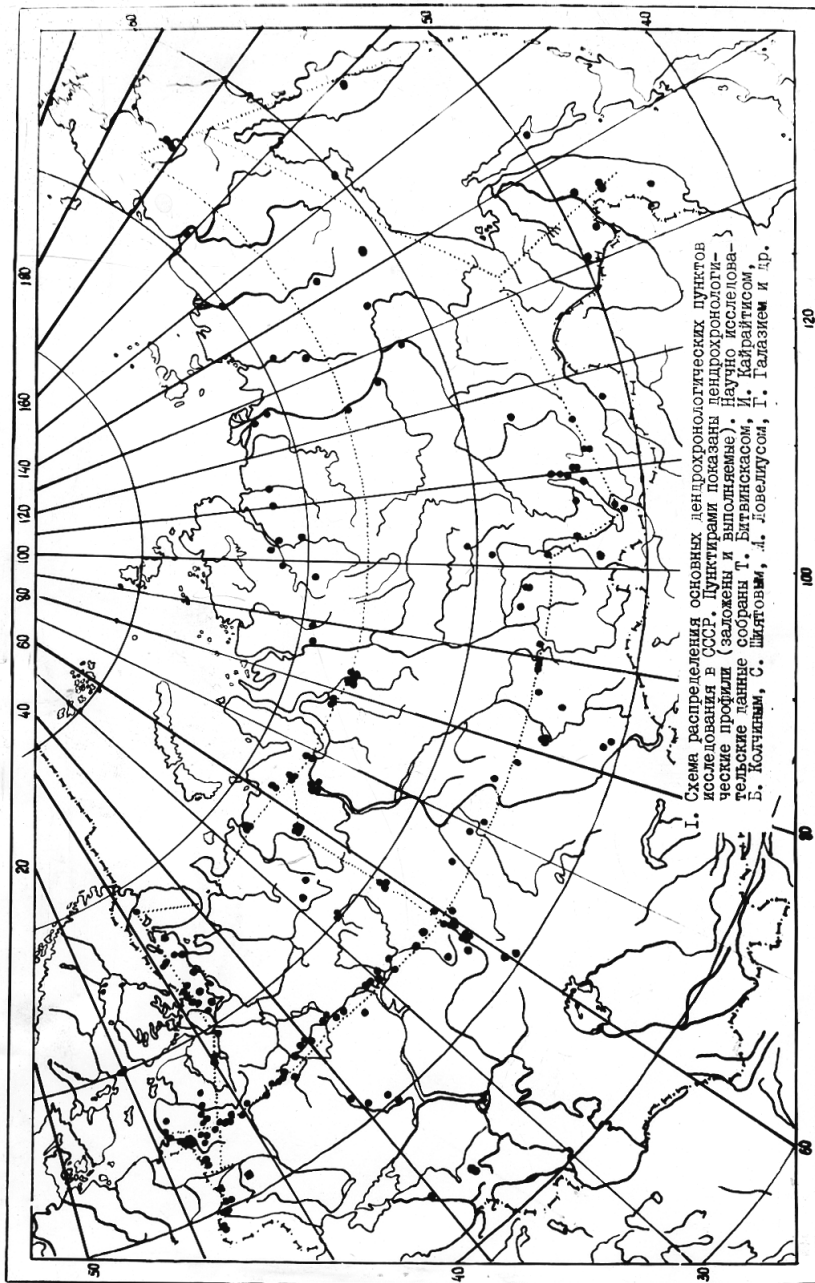
В 1979-81 г.г. разрабатывалась и осваивалась методика обработки дендрохронологических материалов и разрабатывалась методика и конструкция приборов-датчиков снимающих линейную информацию. В итоге построена уникальная линия измерения годовичных колец пригодная для обработки данных ширины годовичных колец на языках ЭВМ Наирри-3 и машинах единой системы (КОИ-7), разработаны методики расчета годовичных индексов радиального прироста, обнаружения корреляционных связей, процента сходства на ЭВМ Наирри-3 и БЭСМ-6, программы построения графиков.

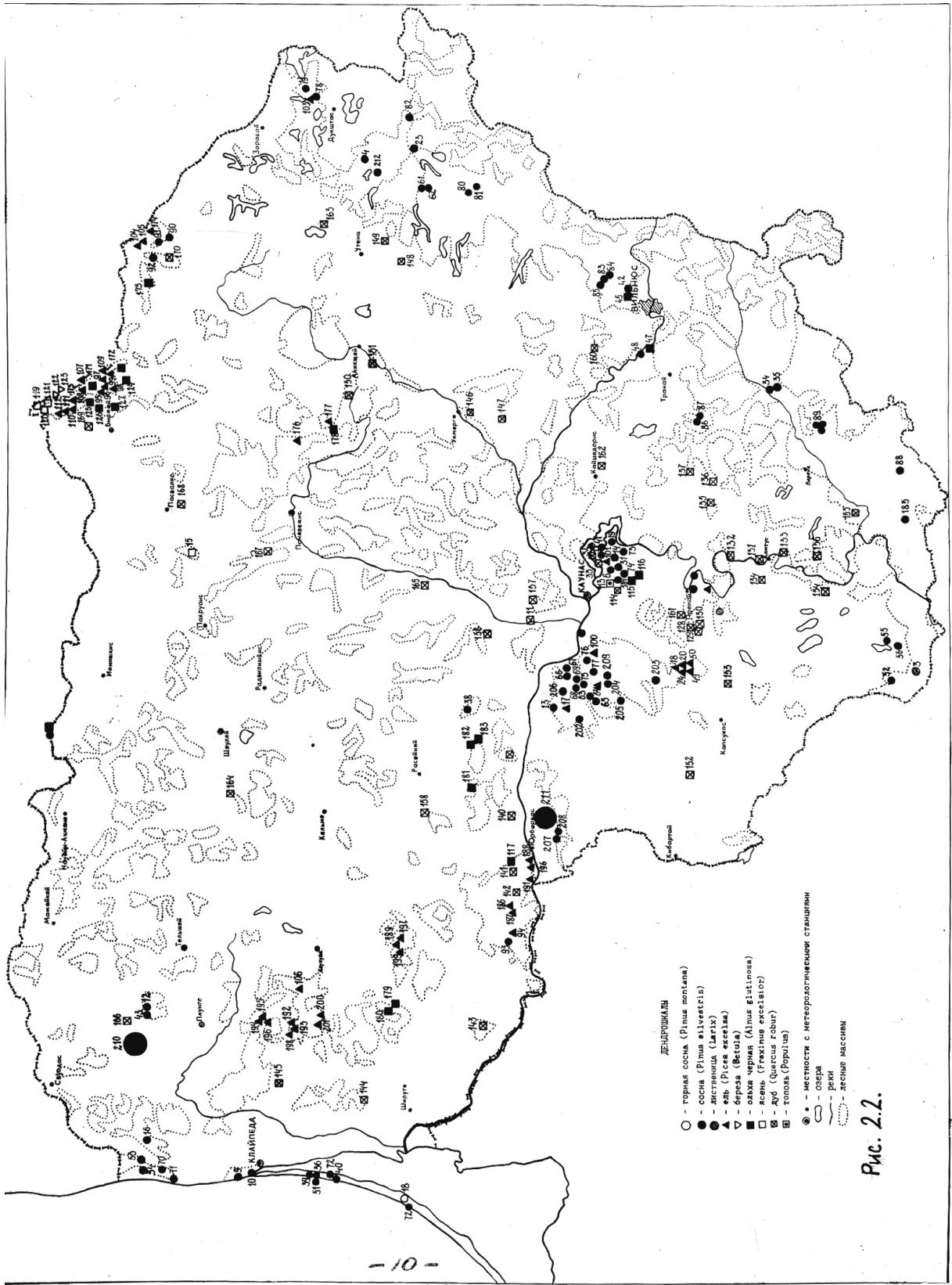
х) Отчет ДКХ

"Разработка методов прогнозирования долгосрочной изменчивости среды дендроклиматохронологическими и радиоуглеродными методами", Каунас, 1978

Цикличность радиального прироста Мурманск-Карпаты была изучена по временным этапам. Таким образом была обнаружена изменчивость цикличности во времени, имеющая практическое значение. Исследования в северо-западном Кавказе, Башкирии и Новгорода, кроме получения информации-материала для радиоуглеродных исследований дают информацию о крупных изменениях в климатических условиях. Например, на Кавказских рядах очень четко выделяется малый ледниковый период. Оправдались и длинные 13 летние усилия лаборатории при изготовлении точно датированных годовых колец древесины. Некоторые итоги показаны в отчете. Были просчитанные основные материалы Торфяника Ужпялко тиражированы на ЭВМ БЭСМ-6. Получены годовые индексы ранней, поздней древесины, графики индексов, просчитанные взаимные корреляции и процента сходства дают возможность строить дендрохронологические сверхдлинные шкалы с использованием нескольких математических приемов.

Региональные закономерности изменчивости годового прироста древесины показаны на материалах торфяника Жувинтас и в закономерностях изменчивости сезонного прироста деревьев в пункте дендроклиматологических и ботанических исследований в д. Вайшноришке.





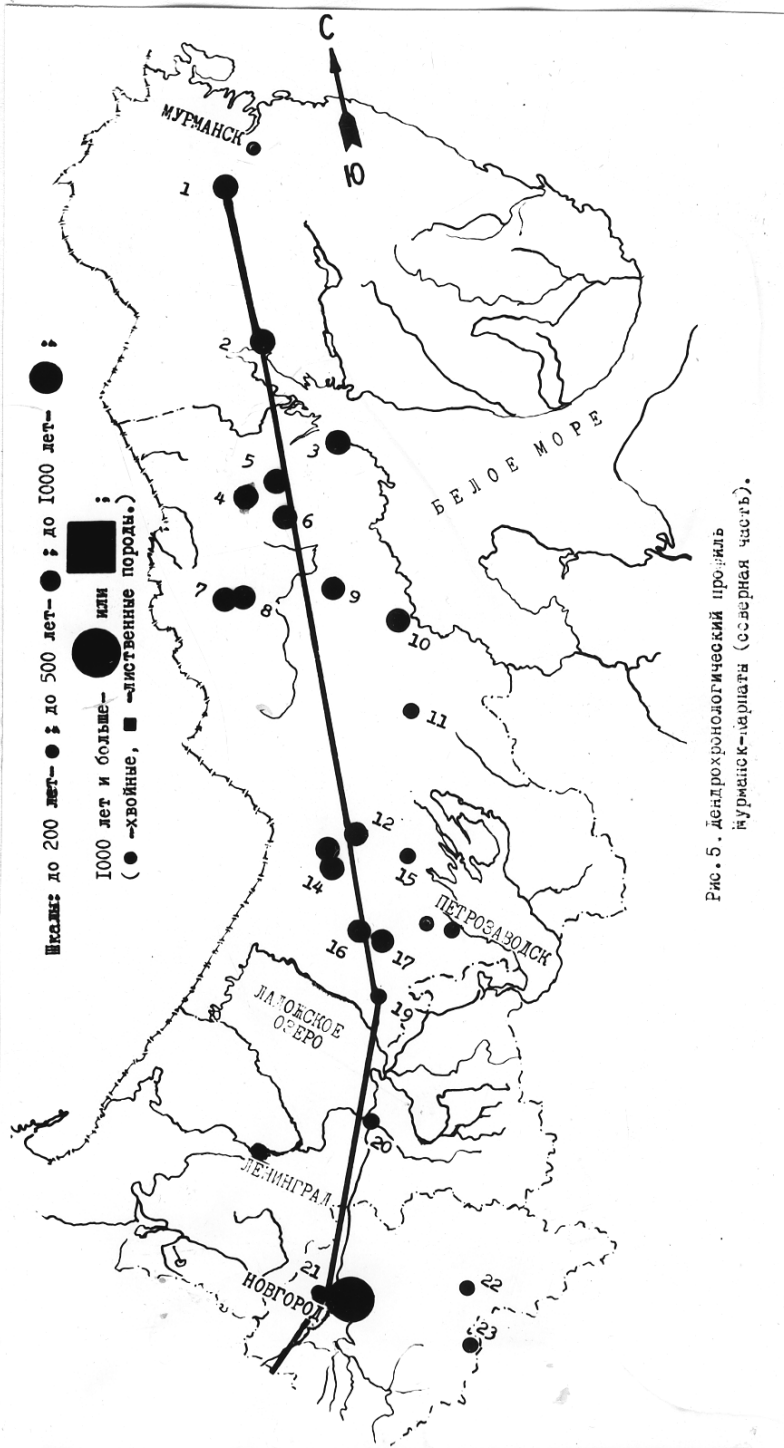


Рис. 5. Дендрохронологический профиль
 Мурманск-Новгород (северная часть).

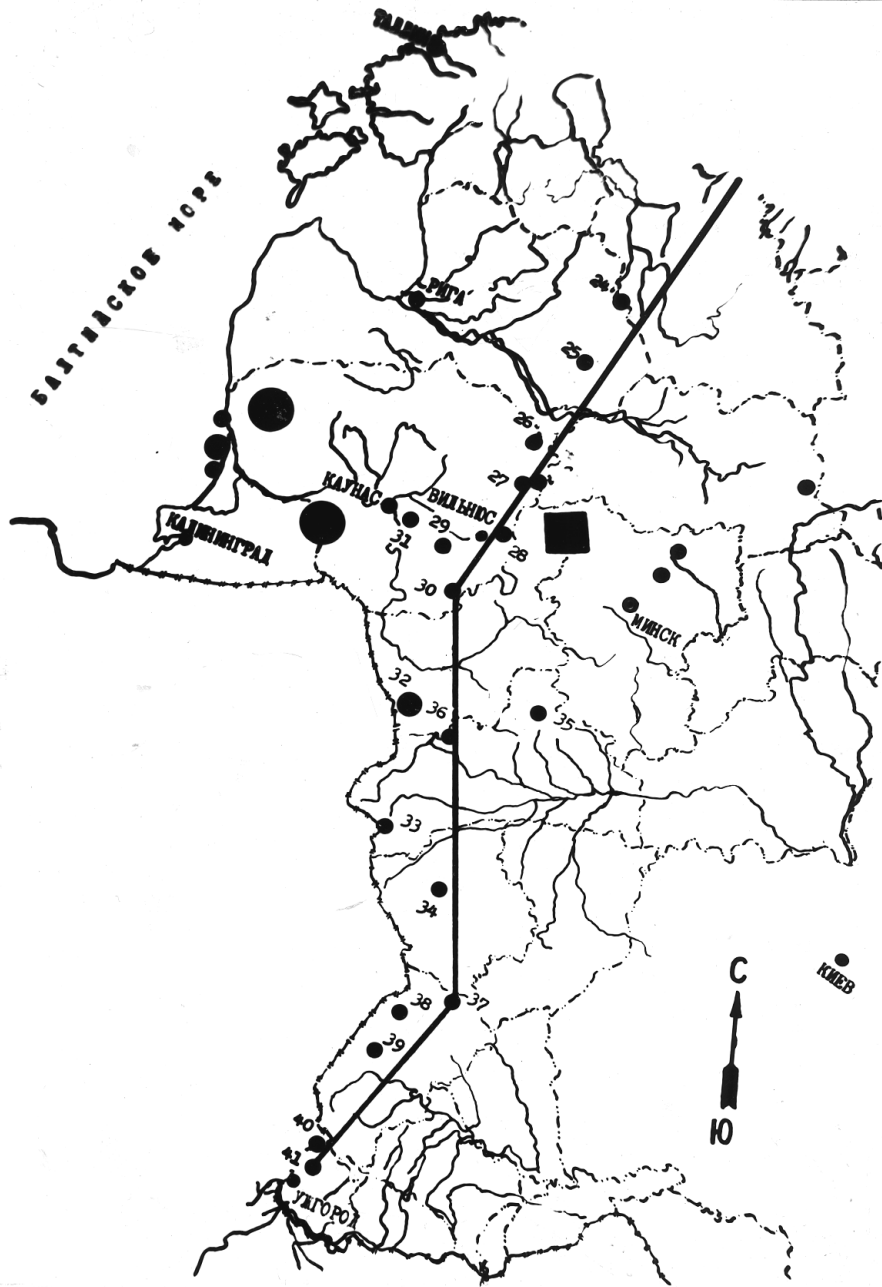


Рис. 2.4. Дендрохронологический профиль
Мурманск-Карпаты /южная часть/

Введение
1.1. Актуальность темы исследования
1.2. Цели и задачи работы
1.3. Методология исследования
1.4. Структура работы

3.0. ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

3.1. Описание экспериментальной установки
3.2. Методика проведения экспериментов
3.3. Результаты экспериментов
3.4. Анализ полученных данных
3.5. Выводы по результатам экспериментов

3.1. МЕТОДИЧЕСКИЕ ОСОБЕННОСТИ ПРОФИЛЬНОГО МЕТОДА ИССЛЕДОВАНИЯ УСЛОВИЙ СРЕДЫ

Семидесятые годы ознаменовались в дендроклиматохронологическом исследовании условий среды активным переходом от анализа отдельных дендрошкал к изучению группы шкал, определенным образом расположенных в пространстве (вдоль определенной широты или долготы, по берегам крупных водоемов и т.д.).

Зачатки пространственных исследований можно найти уже у Дугласа (1928 г.), Шульмана (1956 г.). В СССР профили закладываются Н.В.Ловелиусом, С.Г.Шиятовым, Г.Е. Коминым, Г.Б.Гортинским

Впервые же в СССР работы по закладке профиля (1961 г.) и обоснованию профильного метода в дендроклиматологии (1975 г.) проведены в лаборатории дендроклиматохронологии ин-та Ботаники АН Литовской ССР под руководством Т.Т.Битвинскаса [6].

В качестве основных причин использования профильного метода в дендроклиматологии можно назвать следующие:

1) отсутствие математической модели формирования годичного кольца, и вследствие этого несостоятельность (в большинстве случаев) попыток выделения и изучения тех или иных факторов среды по отдельным сериям годичных колец;

2) сложность и многофакторность воздействия среды на прирост. Профильный же метод выделяет индивидуальную реакцию дерева, локальные и глобальные закономерности прироста. Выбор математического аппарата для анализа дендроклиматического материала определяется свойствами последнего и хорошо обоснован в работе [29]. Это - аппарат изучения случайных процессов.

Методы, при помощи которых изучаются свойства случайных процессов, логично разделить на две группы [2].

1) анализ отдельных реализаций (серий);

2) анализ ансамбля реализаций (всего профиля) при известных статистических свойствах каждой отдельной реализации (серии).

Анализ отдельных реализаций изображен на рис.3.1. Анализ совокупности реализаций показан на рис.3.2.

Дадим краткие пояснения.

АНАЛИЗ ОТДЕЛЬНЫХ РЕАЛИЗАЦИЙ

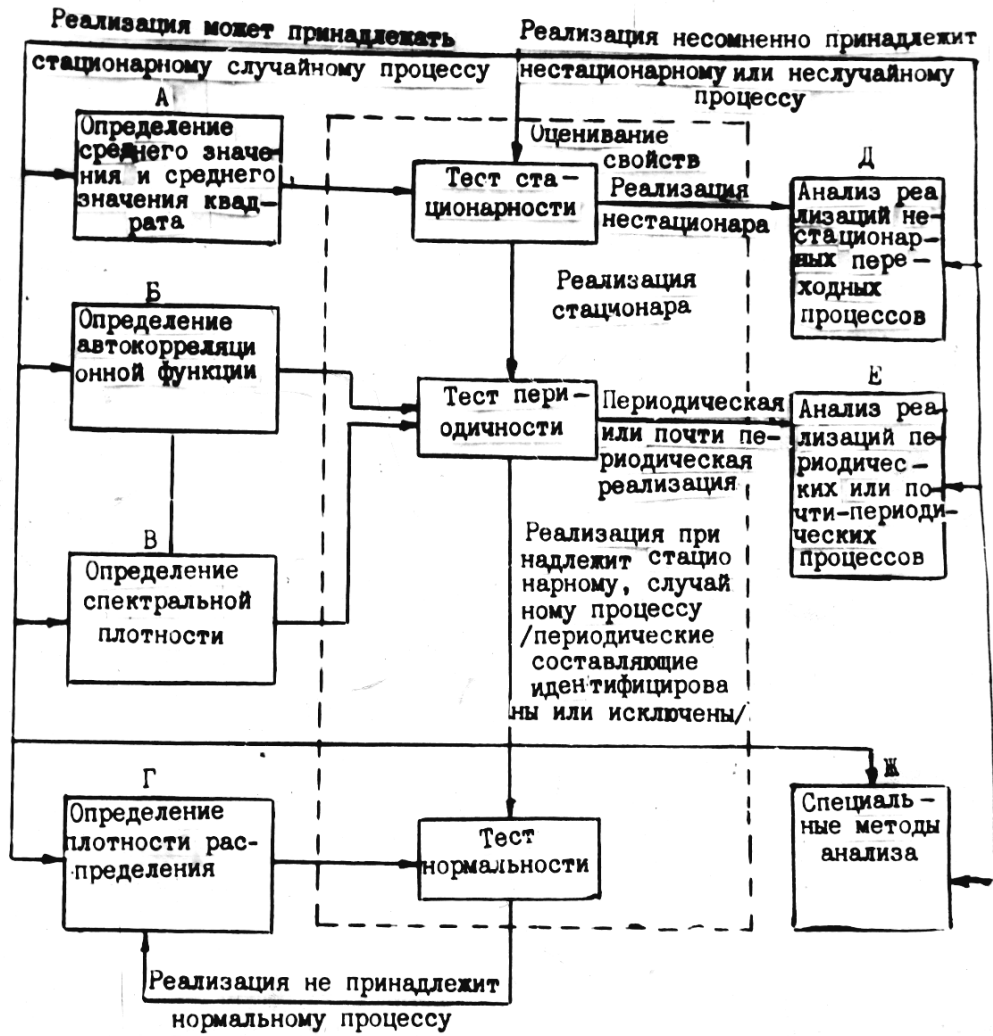


Рис. 3.1.

АНАЛИЗ СОВОКУПНОСТИ РЕАЛИЗАЦИЙ

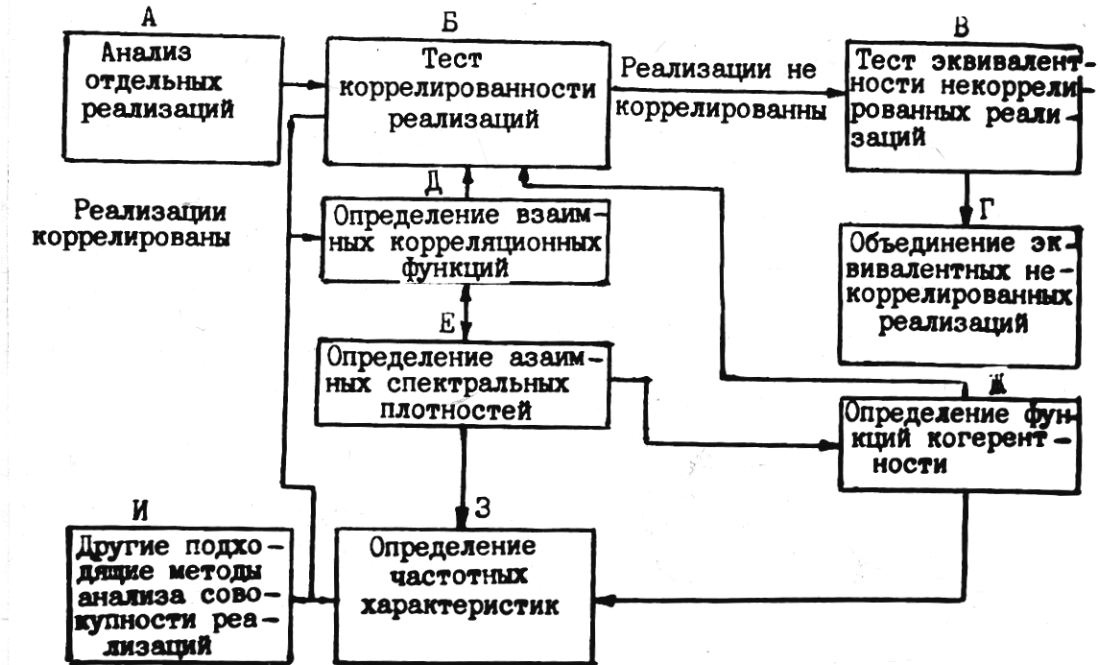


Рис. 3.2.

Параметры отдельной реализации.

Определение среднего и среднего значения квадрата необходимо во всех случаях даже для решения простейших прикладных задач, а также для проверки стационарности и контроля некоторых последующих (вычисление плотности распределения и т.д.) процедур, где также могут быть получены эти параметры (блок А). Вычисление автокорреляционной функции (блок Б) не дает никакой дополнительной информации о процессе по сравнению со спектральной плотностью; но в некоторых случаях автокорреляционная функция дает информацию в более удобном виде (эффективное средство при выделении скрытых периодичностей).

Наиболее важная одномерная характеристика стационарных случайных процессов - спектральная плотность, описывающая частотный состав процесса (блок В).

Последний основной этап рассматриваемой процедуры анализа - определение плотности распределения (блок Г). При анализе процессов плотность распределения часто не измеряют, так как обычно полагают, что все случайные явления подчиняются нормальному закону (наши хронологи весьма близки к нему).

Наиболее полно разработаны методы анализа стационарных случайных процессов. В случае обнаружения нестационарного характера процесса, необходимо использовать специальные методы анализа (блок Д).

Если в результате оценивания основных свойств процесса установлено, что рассматриваемая реализация содержит периодические или почти периодические составляющие, то можно в дальнейшем воспользоваться одним из двух описанных ниже приемов. Во-первых, можно разделить случайную и периодическую части путем фильтрации (учесть возможные искажения) и рассматривать их отдельно (блок Е). Во-вторых, можно совместно анализировать периодическую и случайную части процесса, учитывая присутствие периодической составляющей при интерпретации результатов (спектральной плотности).

Параметры совокупности реализаций.

Блок А содержит соответствующую часть схемы, изображенной

на рис.3.1. На следующем этапе (блок Б) выясняется вопрос о существовании корреляции между отдельными реализациями. В некоторых случаях их некоррелированность с очевидностью следует из эксперимента (например, когда наблюдения над некоторыми физическими явлениями проводились через достаточно большие интервалы времени). В других более или менее очевидна коррелированность отдельных реализаций (если они получены в результате одновременных измерений процессов на входе и выходе некоторой физической системы. Наконец, когда наличие или отсутствие связей не вытекает с очевидностью из элементарных физических или иных соображений, необходимо проведение специального теста.

Если реализации признаны некоррелированными, то следует проверить эквивалентность их статистических свойств (блок В). Нередко встречаются такие случаи, когда различия в результатах, полученных для отдельных реализаций, полностью можно объяснить дисперсией оценок. И, следовательно, никакого физического смысла такой разброс не несет, а лишь приводит к путанице. Объединение же эквивалентных данных до стадии интерпретации позволяет повысить точность полученных оценок.

Вычисление взаимной корреляционной функции полезно тем, что помогает определить задержку по времени между процессами, измеряемыми в двух точках (блок Д).

Наиболее важная часть анализа совместных характеристик совокупности коррелированных реализаций - это вычисление взаимной спектральной плотности (блок Е). Она содержит сведения о линейных зависимостях, которые могут наблюдаться между отдельными реализациями. Приведенные схемы обработки случайных процессов полностью подходят для совокупности реализаций - приростов отдельных деревьев. У нас же, как правило, отдельные реализации представляют собой некую среднюю реализацию, полученную усреднением нормированных на средний прирост по заданному интервалу времени последовательностей прироста отдельных деревьев. В этом случае к описанной схеме необходимо добавить, так называемый, нулевой цикл, состоящий из упрощенного варианта схемы плюс дисперсионный анализ.

Цель анализа реализаций по схеме была описана выше; в двух словах она сводится к выяснению свойств каждой реализации и

затем их взаимоотношений. Целью дисперсионного анализа является выяснение процентного вклада (дисперсии) каждого из заданных факторов, определяющих годовичную изменчивость прироста дерева, в общую вариацию (дисперсию). То есть всю дисперсию около среднего мы раскладываем на частные дисперсии.

$$\sigma_0 = \sigma_1 + \sigma_2 + \sigma_3 + \dots$$

Естественно, что некоторые реализации (деревья) должны отличаться по значениям изучаемых факторов. Так, для изучения влияния на дисперсию фактора возраста, необходимо выбирать деревья двух групп: старые и молодые в равных долях. Для изучения вклада в общую дисперсию фактора изменчивости прироста в разных направлениях от сердцевины, необходимо измерять прирост в двух разных направлениях. Общая для всех деревьев доля вариации, очевидно, характеризует единый для всех деревьев фактор климатический и т.д. Ясно, что желая в дальнейшем изучать по древесным хронологиям изменчивость окружающей среды, мы должны подбирать такие пробные площади, на которых процент вклада в дисперсию климатического фактора возможно больше, а вклад в дисперсию, обусловленный другими факторами возможно меньше.

Продемонстрируем на примере использование схемы (в сокращенном виде). Анализируются пробные площади № 3 и № 4 (профиля Мурманск-Карпаты), расположенные в Северной Карелии неподалеку от берегов Белого моря. Тип леса сосняк бруснично-черничный, общий временной интервал роста 1648-1971 годы:

Анализ отдельных реализаций:

Пробный анализ обоих рядов демонстрирует нестационарность как среднего, так и дисперсии. Переходя к индексам, уничтожаем указанные нестационарности и лишаемся значительной части низкочастотного интервала (подробно об этом в статье о фильтрации 1981 г). Корреляционные и спектральные функции рядов (соответственно а - № 3; б - № 4) приведены на рис. 3.3.а. Уже на этом рисунке просматривается сходство основных закономерностей в рядах.

Взаимоотношения этих рядов видны на рис. 3.3.б где изображены

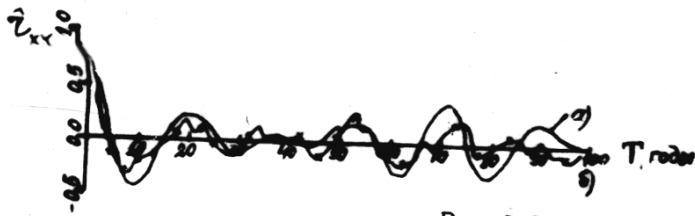


Рис. а. Сверху автокорреляционные функции, снизу - автоспектры пробных площадей №3 /а/ и 4 /б/.

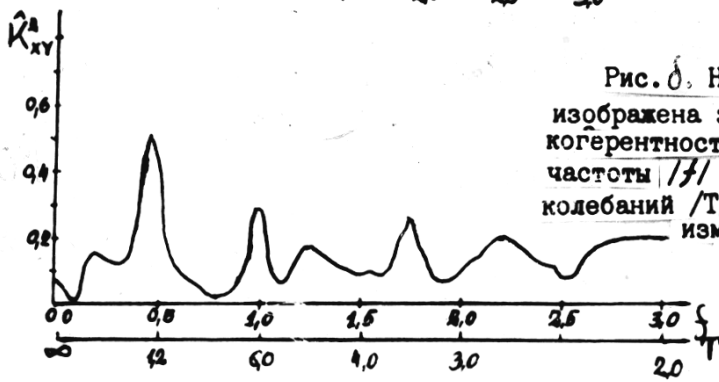


Рис. б. На верхнем рисунке изображена зависимость спектра когерентности (K_{xy}) от основной частоты f или длины периода колебаний T . На нижнем рисунке изменение фазового спектра /запаздывание $-\Phi$ с частотой f ./

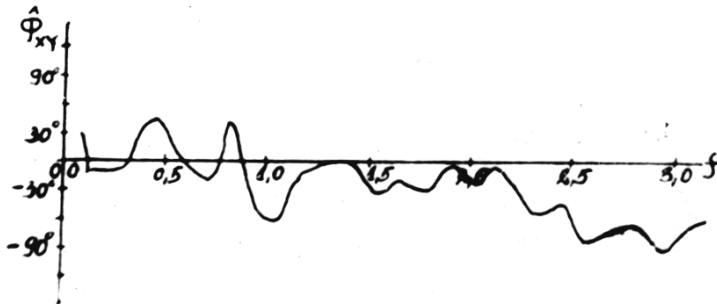


Рис. 3.3.

их спектры когерентности и фазовый.

Значимо обнаруживается связь (на уровне 95%) между рядами для квазигармоник 18-летнего периода. Причем колебания ряда № 4 опережают соответствующие колебания ряда № 3 на 45° или на 5,5 года. Колебания с периодом 6 лет ряда 3 напротив опережают такие же колебания в ряде 4.

Итак, анализ с очевидностью показал связь квазигармоник 18 и 6 лет и некоторый временной сдвиг между ними на соседних пробных площадях. Предложенные схемы могут быть расширены и уточнены в каждом конкретном случае, но в целом следование им позволяет наиболее полно проанализировать как отдельные хронологии, так и их совокупности.

3.1.1. ПРОСТРАНСТВЕННО-ВРЕМЕННАЯ ИЗМЕНЧИВОСТЬ ОСНОВНЫХ СТАТИСТИК ДЕНДРОРЯДОВ ПРОФИЛЯ МУРМАНСК-КАРПАТЫ

В общей схеме анализа дендропрофилей, описанной нами [25] первым этапом называется всестороннее изучение статистик каждого ряда в отдельности. Статистические свойства рядов изучены с помощью корреляционно-спектрального анализа и отражены в [23]. Характерной особенностью большинства геофизических и биологических рядов, как неоднократно отмечалось, является их нестационарность, поэтому необходимым дополнением статистическим свойствам являются динамические, т.е. изучение изменения статистик во времени.

Большинство существующих рядов годовичного радиального прироста древесины в СССР (ширина колец) сравнительно невелики (200-300 лет). Поэтому использование дендрохронологических рядов в качестве индикаторов изменений в окружающей среде (солнечная активность, температура, увлажненность почвы и т.д.) всегда связано с необходимостью удлинения рядов от живых деревьев ископаемой древесиной (из торфяников, со дна рек).

Задача эта очень важна и столь же непроста. Как правило, при верификации пользуются методом реперных точек или вычисляют коэффициенты корреляции, и нередко получают многозначные результаты. Мы видим возможность уточнения и контроля результатов уже имеющихся методов в параллельном применении метода ^{*}скользящих периодограмм. Его назначение - изучение изменений во времени основных статистик рядов.

Сопоставляя основные закономерности изменений статистик в верифицируемых рядах, мы с большей определенностью можем говорить о том, пригодны ли они для создания высоковозрастных дендрошкал и как следует стыковать отдельные хронологии для получения шкалы.

Другое не менее важное основание к применению скользящего периодограмм-анализа, то есть анализа дендрорядов не только цели-

ком, но и на отдельных участках - проверка их на стационарность, что является необходимой предпосылкой применения теории стационарных случайных процессов [9]. В случае принятия гипотезы о нестационарности, оценки спектров (периодограмм) отдельных участков временного ряда дадут представление о характере нестационарности спектра и следовательно помогут в выборе методов дальнейшей обработки ряда.

Кроме того, анализ периодограмм, соответствующих отдельным участкам дендроряда, выявляет те колебания, на которые приходится основная доля изменчивости, и возможное перераспределение мощности между частотами (распад одних периодов, образование других).

Метод скользящих периодограмм позволяет проконтролировать степень и характер искажений различных квази-периодических компонент процедурами вычисления индексов и в случае необходимости внести изменения в сами методики расчета индексов.

И, наконец, сопоставление скользящих периодограмм дендрорядов с аналогичными характеристиками гео- и гелиофизических процессов, возможно, поможет нам лучше понять картину их взаимосвязей.

Все выше сказанное в пользу скользящего периодограмм-анализа позволяет нам считать его одним из основных этапов анализа всех геофизических и биологических рядов.

В программе вычисляются следующие статистические характеристики: 1) периодограмма, 2) спектр, 3) коррелограмма как для всего вводимого ряда, так и его участков продолжительностью 30-100 лет, в зависимости от длины ряда, путем сдвига их последовательно относительно друг друга на 3-10 лет.

Напомним кратко смысл рассчитываемых величин. Искомая дискретная последовательность

$$y_0, y_1, y_2, \dots, y_{k-1}$$

ширины годичных колец дерева всегда может быть разложена в тригонометрический ряд Фурье:

$$y(t) = a_0 + \sum_{p=1}^n (a_p \cos \omega_p t + b_p \sin \omega_p t), \quad t=0, 1, 2, \dots, K-1$$

$$\omega_p = 2\pi p / K \Delta t = 2\pi f;$$

$$n = \text{entier}(K/2)$$

$$p = 1, 2, \dots, n$$

Временный интервал дискретности $\Delta t = 1$ частичный интервал $[0; \pi]$

Коэффициенты разложения определяются формулами:

$$a = \frac{1}{K} \sum_{t=0}^{K-1} y_t$$

$$a_p = \frac{2}{K} \sum_{t=0}^{K-1} y_t \cos \omega_p t$$

$$b_p = \frac{2}{K} \sum_{t=0}^{K-1} y_t \sin \omega_p t$$

Периодограмма определяется как сумма квадратов коэффициентов Фурье, соответствующих колебаниям на определенных частотах.

Оценка периодограммы ряда:

$$\hat{H}(\omega_p) = \frac{K \Delta t}{2\pi} \cdot \frac{a_p^2 + b_p^2}{4};$$

отсюда видно, что колебание с частотой ω_p полностью определяется амплитудой a_p, b_p

Нормируя периодограмму, получаем оценку спектра:

$$\hat{S}(\omega_p) = \frac{4\pi}{K \Delta t} \cdot \sum_{q=1}^p \hat{H}_q / \hat{z}_0, \quad p=1, 2, \dots, n$$

где $\hat{\tau}_0$ - оценка нулевого значения корреляционной функции, которая определяется так:

$$\hat{\tau}_\nu = \begin{cases} \frac{1}{K} \sum_{i=0}^{K-1-|\nu|} y_i y_{i+|\nu|}, & |\nu| \leq K-1 \\ 0, & |\nu| \geq K \end{cases}$$

Для решения перечисленных ранее задач обрабатывалась разного вида информация:

1. Дендроряды абсолютных годовых радиальных приростов для каждой пробной площади профиля;
2. Дендроряды индексов годового радиального прироста для каждой пробной площади;
3. Дендроряды абсолютных годовых радиальных приростов отдельных деревьев первой пробной площади;
4. Временные ряды чисел Вольфа, среднемесячные значения температур.

Начнем с анализа результатов изучения динамики статистик индексных рядов (индексы Т.Битвинскаса).

На рис. 3.3.6 изображена периодограмма и оценки спектральной плотности, полученные сглаживанием периодограммы с помощью цифрового фильтра [?] полиномом степени 3 и интервалом сглаживания, равным 35 точкам. Указанные характеристики рассчитывались для всей длины ряда № 1. Цифрами над пиками отмечены периоды, им соответствующие, и доля дисперсии в процентах, приходящаяся на каждое из этих колебаний. Наиболее мощные из них имеют периоды 37 лет (4,6%), 22 года (11,5%) и 3,7 года (2,8%).

Проведем более подробный анализ и выделим временные интервалы, на которых отмеченные колебания проявляются наиболее заметно.

Оценивались периодограммы на участках ряда продолжительностью 100 лет, путем сдвига их последовательно относительно друга друга на 10 лет. Результат оценивания скользящих периодограмм I ряда приведен в таб. 3.1 в процентах от оценок дисперсий, вычисленных

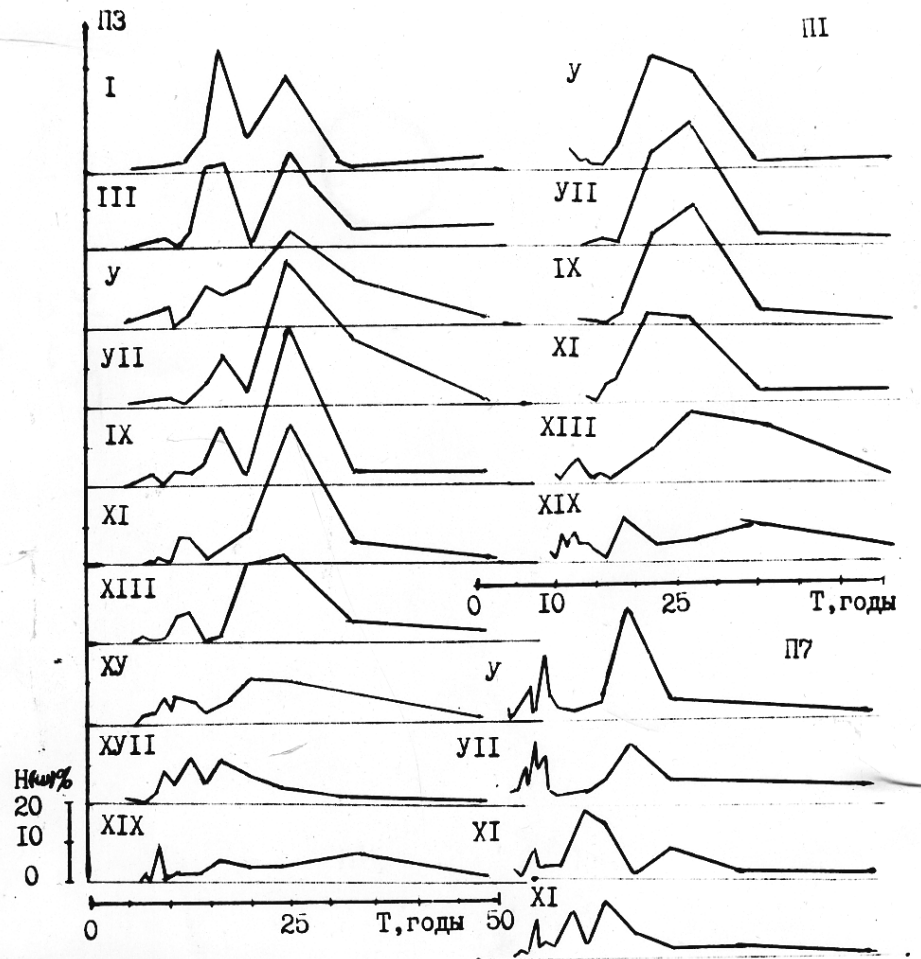
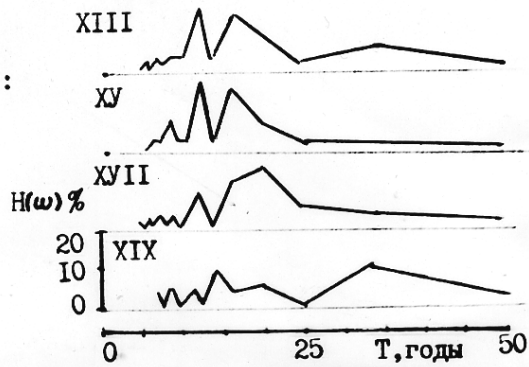


Рис. 3. Изменение по времени
 периодограмм ($H(\omega) \%$)
 / в индексах / пробных площадки:
 №1-Мурманская область;
 №3, 7-Северная Карелия.



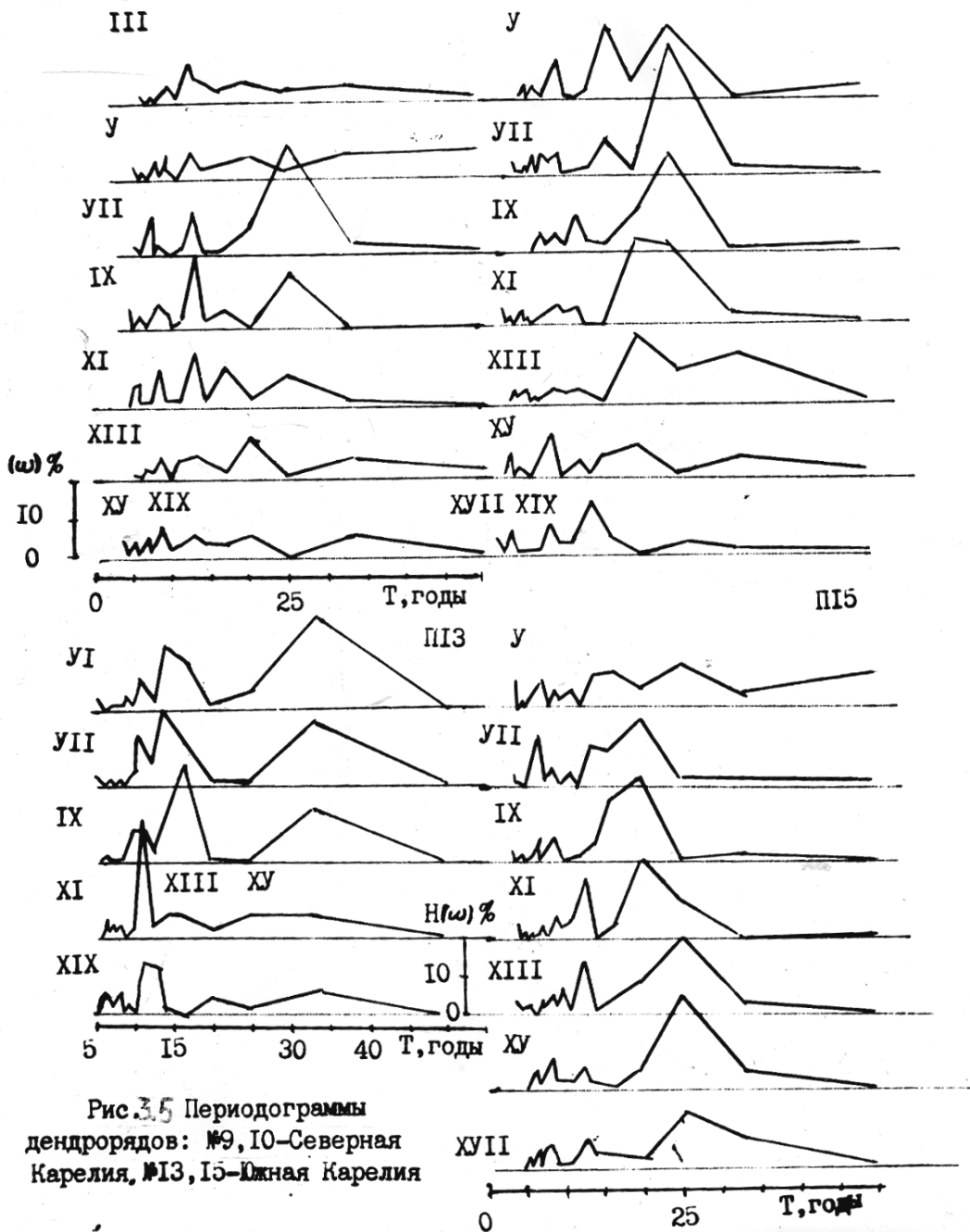


Рис. 3.5 Периодограммы
дендрорядов: №9, 10—Северная
Карелия, №13, 15—Южная Карелия

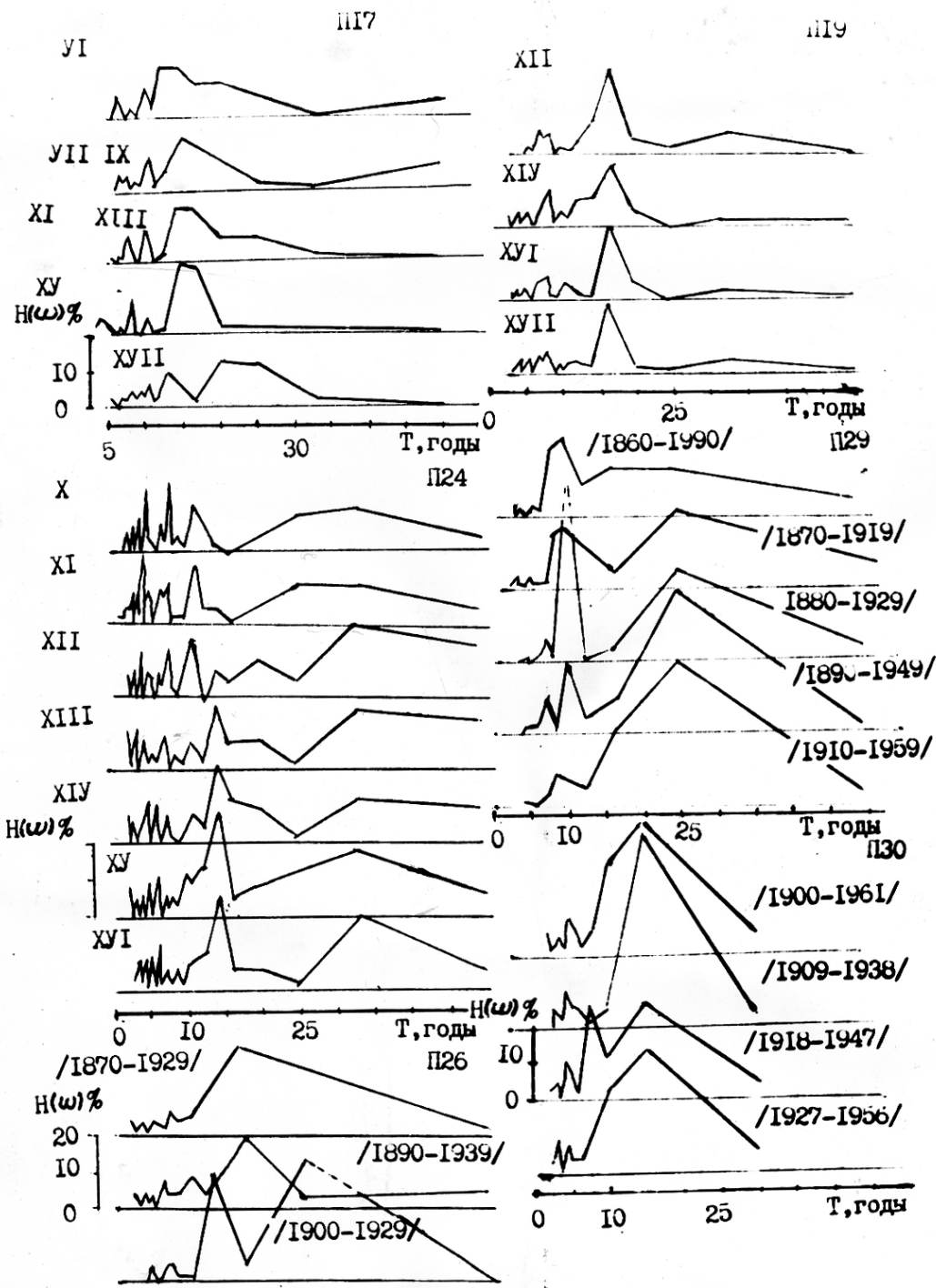


Рис. 36 Периодограммы дендрорядов: №17, 19-Южная Карелия;
 №24-Новгородская обл.; №29-Латвийская ССР; №29, 30-Литовская ССР

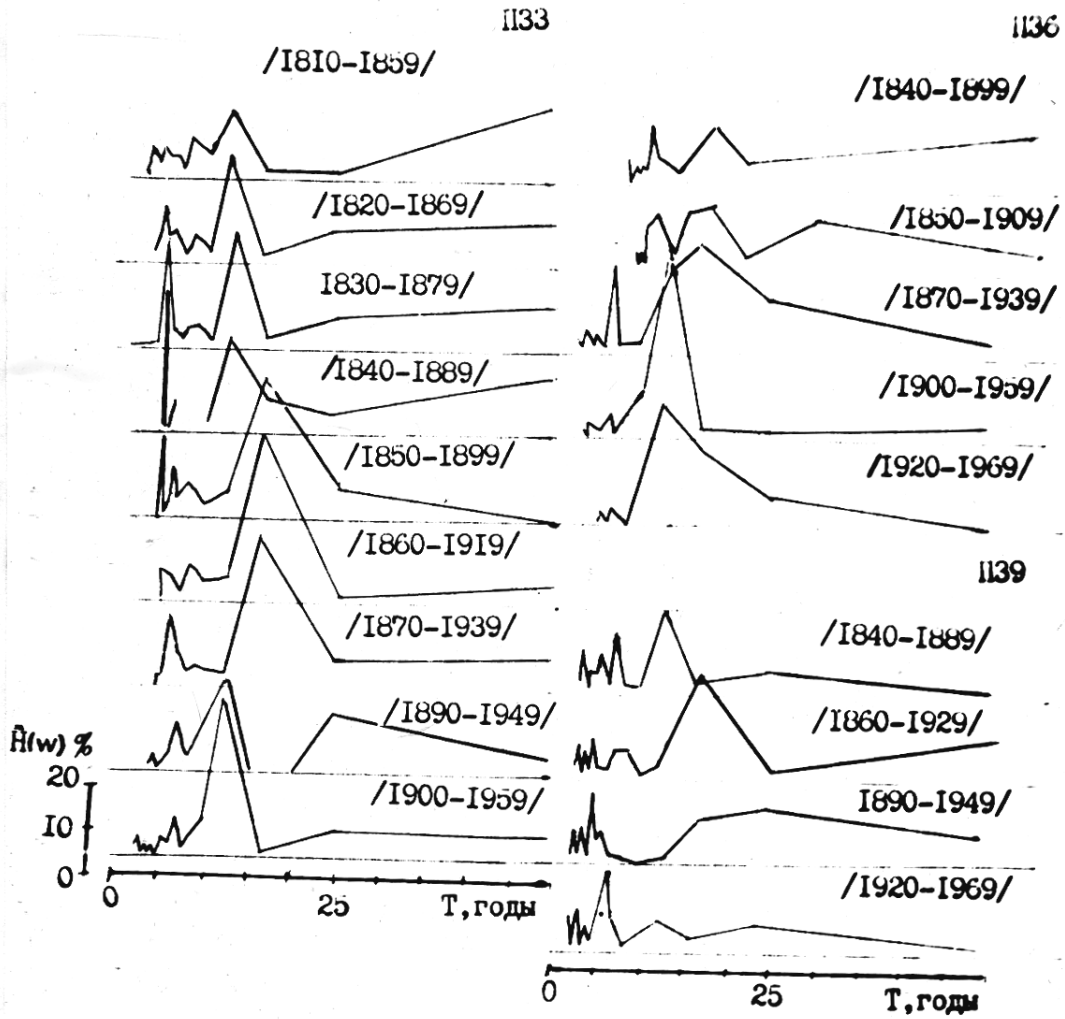


Рис. Периодограммы дендрорядов: №33-Литовская ССР;
 №36-Белорусская ССР; №30-Украинская ССР.

по данным каждого участка. Сумма значений в отдельных столбцах должна равняться 100%, но из-за погрешностей округления она несколько отличается от 100%.

Результаты первых девяти столбцов (1730-1919 гг.) показывают, что наибольшая мощность (до 50-60% дисперсии) соответствует трем точкам периодограммы. Эта особенность обусловлена колебаниями продолжительностью 20-30 лет, причем наиболее выражены колебания более узкого диапазона 20-25 лет. Напомним, что колебания с периодами > 30 лет практически отсутствуют, вследствие их отфильтровки процедурой вычисления индексов [26]. В последних 6 столбцах (1820-1959) энергия распределяется более равномерно по всем частотам, и лишь незначительно превышает уровень белого шума для 33-летних колебаний, кроме того, несколько усиливается амплитуда колебаний с периодами 16,7; 9 и 3,5 года.

Таким образом, в первую половину интервала исследования отчетливо проявляются колебания с периодами 20-25 лет, затем эти колебания распадаются, порождая более быстрые (3,9 лет) и медленные (< 30 лет).

Более наглядны и потому удобны для визуального анализа рисунки временной изменчивости периодограмм (рис. 3.4-3.7). Здесь отобраны наиболее представительные для своего региона пробные площади.

Северная Карелия (№ 3, 7, 9, 10)

Начало интервала исследования (до 1900 г.) характеризуется устойчивыми колебаниями в среднечастотном интервале (15-22 года) с заметным уменьшением амплитуды начиная со второй половины 19 в. Особенность 9 пробной площади заключается в формировании колебаний с $T \sim 25$ лет в интервале 1730-1850 с постепенным возрастанием амплитуды и затем медленным спадом. Типичным для этой группы рядов является появление и устойчивое существование колебаний с периодами меньше 15 лет, а именно 7 и 12 летних, которые, как правило, усиливались в результате распада среднечастотных колебаний и несколько ослабе-

вали в последние 40 лет. Для последних 50 лет характерно усиление низкочастотной составляющей в рядах.

Южная Карелия (№ 13, 15, 17, 19):

В 13 ряде обнаруживаем колебание периода 33 года, быстро убывающее по амплитуде (1740-1850 г.), в то время как компонента с $T \sim 15-20$ лет усиливается. Во второй половине интервала (с 1850 г.) отчетливо проявляются 11-летние колебания.

Ряды 15 и 17 характеризуются постепенным нарастанием к середине интервала колебаний в интервале 15-27 лет, а затем их незначительным спадом на фоне постепенного удлинения волны. При этом оба ряда содержат быстрые колебания (7-12 лет) с уменьшающимся к концу интервала (1850-1950) амплитудами.

Ряд 19 характеризуется устойчивой во времени волной 16 лет и заметной быстрой компонентой. Общим для всей группы является наличие среднечастотных колебаний с несколько ослабевающей к концу интервала амплитудой и высокочастотной компонентой.

Ленинградская и Новгородская обл. (№ 24)

В большинстве своем ряды этого района коротки (60 лет) и скользящий периодограмм анализ в интересующем нас частотном диапазоне к ним не применим. Анализ дендроряда 24 показывает, что дисперсия примерно одинакова на всех частотах, или амплитуды колебаний малы (на уровне шума).

Латвийская и Литовская ССР (№ 26, 29, 30, 33)

Длина рядов уменьшилась, в связи с чем уменьшился с 100 до 50 и 30 скользящий интервал, а шаг сдвига стал 5 и 3, вместо 10.

Основная по проценту забираемой мощности компонента (10-15 лет) также обнаруживает тенденцию распадаться в XX век с образованием быстрых 5-7 лет и медленных (> 20 лет) колебаний, либо (33 ряд) длина ее периода заметно увеличивается.

Белорусская и Украинская ССР (№ 36 , 39)

Большая часть рядов этой группы вследствие малости их длины выпадает из рассмотрения, в оставшихся двух намечается удлинение главной (по дисперсии) волны (12-15 лет в 39 ряду) и заметное усиление быстрых колебаний (5-10 лет).

Результаты распределения дисперсий по частотам отдельных дендрорядов профиля сведены в табл. 3.2. В таблице отобраны характерные для своей группы ряды.

Анализ дендрорядов профиля в абсолютных годовых приростах подтверждает, что на всем временном интервале подавляющая доля дисперсии приходится на низкие частоты ($T \sim 30$ лет), которые не могут быть обстоятельно изучены вследствие недостаточной длины рядов (100-300 лет). Сопоставление скользящих периодограмм абсолютных и индексных выражений годового прироста демонстрирует качественно и количественно искажения, вносимые процедурой вычисления индексов в характеристики отдельных колебательных компонент. В случае, когда интересующие нас колебания искажаются хотя бы и в некоторых временных интервалах значительно, методику вычисления индексов следует признать неудовлетворительной.

Анализ рядов отдельных деревьев одной пробной площади (№ I) позволяет обнаружить индивидуальные особенности роста, детальное изучение которых возможно методом дисперсионного анализа.

Были применены скользящий периодограммный анализ к рядам, характеризующим внешние по отношению к дереву процессы (изменение температуры воздуха, число солнечных пятен), с целью изучения особенностей взаимосвязи между ними.

Анализ среднемесячных значений температуры (ст. Кола, около Мурманска) не показывает каких-либо устойчивых во времени особенностей. Анализ же средневековых чисел Вольфа /и прямое сопоставление характера изменения во времени амплитуды 11-летнего или 22-летних циклов с особенностями профиля не дало положительных результатов.

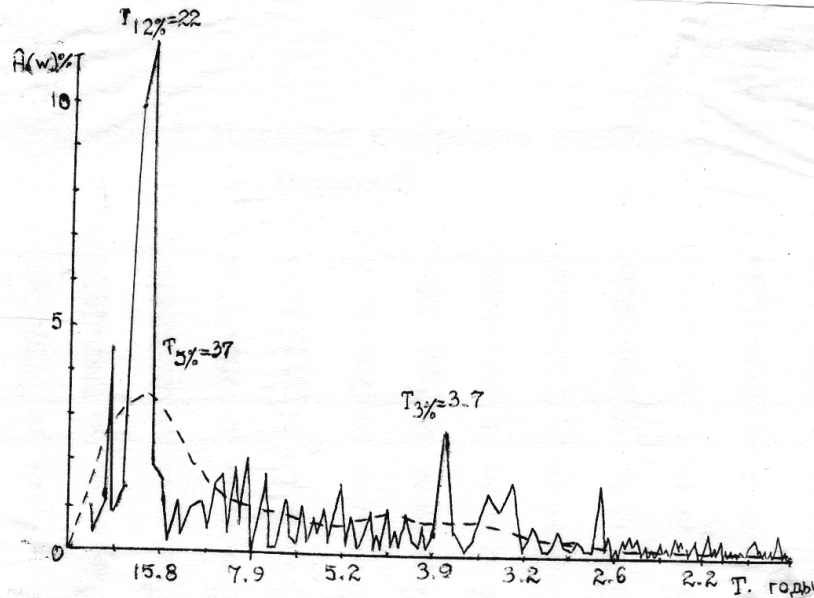


Рис.3.3.в.Периодограмма (сплошная) и оценки спектральной плотности (пунктир) дендроряда I пробной площади (в индексах)

Таблица 3.2. Распределение дисперсий по частотам для периодограмм дендрорядов профиля Мурманск-Карпаты

Номер ряда	Интервал периодов (годы)	Процент дисперсии	Процент дисперсии в интервале 20-25 лет
1	> 20	80	10-40
3	> 20	85	10-45
4	> 20	60	10-40
5 ^x	> 30	60-70	5-15
8 ^x	> 20	60	5-25
10	> 20	60-70	6-16
35	> 20	85	5-14
36	> 20	50	14-35
39	> 30	45-65	< 10

x- тип леса сосняк сфагновый, остальные-сосняк бруснично-черничный

Таблица 3.1.

Периодограммы отдельных интервалов дендроряда № I
(индексы)

Период (лет)	Годы															
	1730-1829	1740-1839	1750-1849	1760-1859	1770-1869	1780-1879	1790-1889	1800-1899	1810-1909	1820-1919	1830-1929	1840-1939	1850-1949	1860-1959	1870-1969	
I	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
100.0	1	0	1	1	1	1	2	1	0	1	2	0	0	0	0	
50.0	2	4	22	2	1	2	2	2	1	1	3	6	3	2	2	
33.3	1	2	3	3	3	2	2	9	14	13	8	5	8	12	8	
25.0	24	30	31	30	30	27	21	17	17	13	9	3	2	4	4	
20.0	28	25	23	24	23	23	22	11	7	4	3	1	2	4	3	
16.7	6	2	1	2	3	5	6	2	2	3	5	11	11	9	9	
14.3	0	1	2	3	2	2	4	2	1	2	2	1	1	0	0	
12.5	2	2	0	0	1	0	1	1	2	4	4	8	7	4	2	
11.1	0	2	2	1	1	1	3	4	5	1	2	1	1	3	4	
10.0	4	2	3	3	2	2	2	4	4	2	1	3	2	4	6	
9.1	1	1	0	1	0	1	0	2	1	5	8	6	8	5	3	
8.3	0	1	1	1	2	2	1	1	1	4	7	4	3	4	6	
7.7	0	0	0	0	1	1	1	1	0	0	1	1	1	0	1	
7.1	0	2	2	1	2	2	3	4	4	4	2	2	2	2	1	
6.7	1	0	0	0	1	1	2	3	2	0	1	2	2	2	2	
6.3	2	1	1	1	0	0	0	1	1	3	2	2	2	1	1	
5.9	1	1	1	1	1	1	1	1	1	0	1	1	1	2	1	
5.6	2	1	1	0	2	3	2	2	1	1	1	1	1	1	1	
5.3	2	3	2	3	2	2	3	2	2	1	1	1	2	2	2	
5.0	0	0	0	0	2	2	1	2	2	3	3	2	2	1	1	
4.8	2	2	2	2	2	2	1	2	0	0	0	0	0	1	1	
4.5	3	3	2	2	1	1	2	3	1	0	0	1	0	0	0	
4.3	1	1	1	0	1	1	3	3	0	1	1	1	3	0	0	
4.2	0	0	0	1	0	0	0	1	3	3	3	2	4	4	2	
4.0	0	0	0	0	0	0	0	1	4	2	2	3	1	2	0	
3.8	1	0	1	1	2	3	2	0	0	0	0	0	0	0	1	
3.7	1	1	2	1	1	0	1	4	3	5	6	7	4	5	4	
3.6	0	0	0	0	0	1	0	2	2	3	2	3	5	4		

Продолжение табл.3.1.1.

I	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
3.4	1	1	1	1	1	1	0	0	0	1	1	0	0	0	0
3.3	1	1	2	1	1	0	1	0	0	0	1	1	3	2	3
3.2	1	1	1	0	0	0	1	0	0	0	2	3	5	5	6
3.1	1	1	2	2	2	3	4	3	3	3	2	1	0	0	1
3.0	1	1	0	0	0	0	0	0	0	0	1	1	0	0	0
2.9	0	0	0	0	0	1	1	1	1	1	1	0	1	0	0
2.8	2	2	2	2	1	1	1	2	4	4	1	2	1	1	2
2.7	2	1	1	2	3	2	2	1	1	2	1	1	1	1	0
2.6	0	0	0	0	0	0	1	0	0	1	4	3	2	3	4
2.5	0	0	0	0	1	0	0	0	1	1	0	0	0	0	1
2.4	1	1	1	1	1	1	1	1	0	1	0	1	0	1	0
2.3	0	0	0	0	0	0	0	1	1	1	0	0	0	1	1
2.2	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0
2.1	0	0	0	0	0	0	0	0	0	1	1	0	1	0	1
2.0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	1

Итак, в результате применения метода скользящих периодограмм при анализе дендрорядов профиля Мурманск-Карпаты были получены следующие результаты:

1. В рядах профиля нарушается стационарность спектральных характеристик, поэтому статистический анализ дендрорядов следует проводить по частям;

2. Основная колебательная компонента (11-22 года) устойчиво существует и в некоторых случаях формируется в 18 веке - первой половине 19 века. В конце 19-первой половине 20 веков среднечастотная колебательная компонента значительно уменьшается по амплитуде, либо совсем распадается;

3. Конец 19 - начало 20 столетий характеризуются появлением и усилением (иногда значительных) быстрых колебаний (5-10 лет);

4. Медленные колебания усиливаются в 20 столетии.

Продолжение таблицы 3.2.

1	2	3	4	5	6
40.	Укр. ССР, Жорнаровский лесокombинаг, Хостинское лесничество	37	Будина	D ₃	1971-1852 (119)
41.	Укр. ССР, Перечинский лесокombинаг, Турье-Решетское лесничество	72	Сосняк черничный	C ₂	1971-1913 (58)

3.1.2. Применение факторного анализа при профильном изучении условий среды в дендрохронологии

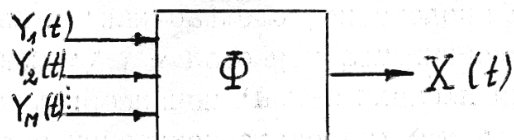
Переход в дендрохронологии от анализа отдельных серий колец деревьев к совокупности подобных серий был следствием осознания того факта, что ширина годичного кольца зависит от многих случайных величин: микроусловий среды, разнообразных климатических параметров (или климатического комплекса), локальных и глобальных геофизических условий и, наконец, астрофизических условий (активность солнца, звезд). Стало ясно, что для выделения и последующего изучения влияния одного из признаков, определяющих прирост, необходимо использовать данные о приросте группы деревьев, объединенных по определенному принципу (пробные площади, профили, сетки).

Естественно было ожидать, что качественный скачок в понимании задач дендроклиматохронологии будет сопровождаться качественным изменением и методов обработки информации, а именно: единичные образцы (деревья) изучаются методами одномерной статистики (корреляционный и регрессионный анализ), группы образцов требуют для своего изучения многомерных методов (множественная регрессия, факторный анализ). К сожалению большой разрыв между математизацией задач и классическим дендроклиматологическим подходом к ним не позволил в широких масштабах осознать необходимость такого перехода и осуществить его. Лишь отдельные авторы использовали многомерные методы для получения адекватной математической модели, часто без глубокого понимания необходимости их применения в решении частных задач, например: типизация экстремальных по метеоусловиям лет прироста /15а/. Надо заметить, правда, что применение многомерной статистики, в частности факторного анализа, связано с предварительным сбором и обработкой (камеральные измерения) образцов древесины с обширных территорий. По полной программе были проведены подобные исследования несколько лет назад в США, лаборатории Фритца /34/.

После обоснования и введения в отечественную практику дендроклиматохронологии Т.Битвинским /6/ профильного метода изучения условий среды появилась возможность и насущная необходимость в использовании многомерных статистических методов при обработке дендроматериала.

3.1.2.1. Формализация задачи.

Рассматривается сложная система - деревья, конечный результат функционирования которой погодичное образование и накопление биомассы (в корнях, в стволе, побегах, листьях, плодах). Наиболее отчетливо годичная биомасса разделяется в стволе на отдельные годичные кольца. Поэтому выходом системы удобно считать последовательность ширины годичных колец - $X(t)$, (где $t = 1, 2, 3, \dots, N$ - годы прироста дерева)



На вход системы подается (вселяет) множество случайных величин $\{Y_j(t)\}$, (где $j = 1, 2, \dots, M$, M -количество входных сигналов), которые влияют на величины $X(t)$, причем характер влияния определяется функцией Φ , законом функционирования системы. Совокупность величин $\{Y_j(t), j = 1, 2, \dots, M\}$ характеризует внешние условия роста дерева. Сюда входят: микроусловия среды (затемненность, ширина кроны, наличие микроэлементов в почве, угол экспозиции, высота места и др.), комплекс климатических условий (температура воздуха и почвы, осадки и т.д.), геофизические характеристики места (магнитное поле) и, наконец, влияние космоса.

Функция Φ во многих случаях может считаться близкой к линейной, во всяком случае, монотонной, а потому может быть достаточно хорошо приближена к прямой линии. Так что в первом приближении считаем Φ - линейной функцией.

Нас интересует, как входная информация $\{Y_j(t), j = 1, 2, \dots, M\}$ системы отражена в выходном сигнале $X(t)$, от входов. Ясно, что для полного отражения всего многообразия входных сигналов потребуется набор входных $\{X_i(t), i = 1, 2, \dots, P\}$. Причем величина P зависит от M и от поставленной задачи.

3.1.2.2. Характеристика статических методов обработки дендрорядов.

Имеется несколько методов решения задач оценок связи между параметрами. Наиболее простые и пригодные для использования -

- корреляционный, регрессионный и факторный анализ. В основе каждого лежат отдельные предпосылки и математические модели, которые определяют границы возможной интерпретации полученных результатов.

Если методы корреляционного и регрессионного анализа - для определения степени связи между наблюдаемыми величинами, то факторный анализ оценивает связь группы наблюдаемых параметров с заранее неизвестным ненаблюдаемым фактором и дает уравнения, определяющие эти факторы.

Корреляционный анализ.

Исследование ограничений и анализ результатов применения корреляционного анализа показал, что этот аппарат обычно не позволяет вскрыть причинно-следственные связи между исследуемыми параметрами. Более целесообразной областью применения корреляционного анализа является применение его при решении задач оценки степени связи между причинно-независимыми величинами (группы входных или выходных параметров).

Регрессионный анализ.

Его целесообразно использовать для раскрытия причинно-следственных связей (оценка входных - выходных зависимостей). Но регрессионный анализ весьма чувствителен к выбору исходных предпосылок. Наиболее существенные ограничения при использовании регрессионного анализа для определения зависимостей входных характеристик сложных систем от входных + необходимость измерения без существенных ошибок входных параметров объекта, включенных в уравнение регрессии, и отсутствие значительных корреляционных связей между всеми входами. Ошибки измерений искажают оцениваемые коэффициенты регрессии. Существенная корреляция между входными параметрами приводит к плохой обусловленности системы нормальных уравнений Гаусса и, следовательно, к ненадежным оценкам регрессионных коэффициентов. Кроме того, эта корреляция усиливает пагубное влияние ошибок измерения на результат. Ясно, что в нашем случае, как и в большинстве других сложных систем, нет возможности измерить без ошибок входные переменные, кроме того, нет сомнения (из общих физических соображений, а также из анализа структуры входных сигналов) в том, что клима-

тические (входные) сигналы связаны между собой, а точнее являются внешним проявлением некоторой климатической сущности, которая определяет тип климата в данной местности.

Для устранения или обхода указанных ограничений регрессионного метода предлагается использовать регрессию факторного анализа, основанную на классической модели факторного анализа. Оба ограничения регрессионного анализа удовлетворяют требованиям факторного анализа, так как коррелированность между переменными — необходимое условие, при котором факторный анализ имеет смысл, а ошибки измерений могут быть выделены как специфические факторы факторной модели. В заключение необходимо указать на еще одно характерное отличие факторного анализа от других статистических методов: он позволяет не только проверять уже сформулированные гипотезы, но и предлагает новые, не следующие с очевидностью из экспериментального материала, а отражающие внутренние ненаблюдаемые свойства системы.

3.1.2.3. Области применения факторного анализа в свете задач дендроклиматохронологии.

1. Сжатие данных.

Факторный анализ имеет дело с большими массивами экспериментального материала, часто трудно обозримыми, а потому не поддающимися непосредственному анализу. Первое, к чему стремится исследователь и что позволяет данный метод — сократить число изучаемых переменных, характеристик системы, без существенной потери информации о свойствах самой системы.

2. Задачи классификации.

Эмпирические методы классификации лесоводов, основанные на анализе большого числа признаков — характеристик отдельных деревьев и групп деревьев — могут быть существенно усилены использованием факторного анализа. В этом случае метод дополняется каким-либо алгоритмом классификации произвольного набора переменных, разрабатываемых теорией распознавания образцов.

3. Нахождение основных скрытых влияний.

(причинный анализ взаимосвязей)

Уже первой грубой анализ связи увеличения (уменьшения) температуры воздуха или осадков с шириной кольца показывает, что

она (связь) неоднозначна, как если бы температура и осадки являлись внешним проявлением некоего скрытого (ненаблюдаемого) климатического фактора, определяющего приrost. А нахождение и интерпретация общих факторов внутреннего поведения системы выдвигает новые гипотезы о порождающем механизме генерирования переменных.

4. Измерение обобщенных показателей.

Об этом уже сказано в предыдущем пункте. Частные измеряемые показатели (температуры, осадки и др.) могут быть лишь внешним выражением реально существующей, но непосредственно неизмеряемой величины — обобщенного (климатического, например, или микроусловий среды) показателя. А тогда изменение этих частных показателей x_i может быть вызвано как обобщенным показателем f , так и другими случайными величинами — e_i . В первом случае, применительно к нашей задаче, произойдет закономерное изменение прироста (под действием обобщенного климатического показателя), а во втором — случайное, т.е. под влиянием случайных (неучтенных, неизучаемых) величин. Модель: $x_i = l_i f + e_i \quad (i = 1, 2, \dots, n)$

5. Структурный анализ случайных процессов.

При анализе дендрорядов (деревьев) одной пробной площади, одного климатического района и т.д. имеется возможность выделения с помощью факторного анализа характерных кривых, отражающих специфические особенности в изменении прироста и соответствующих определенным, наиболее часто встречающимся в изучаемой выборке ситуациям.

Основная модель факторного анализа:

$$\{x(t_i)\}^{(j)} = \overline{\{x(t_i)\}} + \sum_{r=1}^k \{l_r(t_i)\} f_r^{(j)} + \{e(t_i)\}^{(j)}$$

$\{x(t_i)\}^{(j)}$ — j -ая реализация исследуемого параметра в момент времени $t_i \quad (i = 1, 2, \dots, n)$

$l_r(t_i)$ — r -ая характерная кривая;

$f_r^{(j)}$ — вес r -ой характерной кривой в структуре j -ой реализации;

$\{e(t_i)\}^{(j)}$ — последовательность некоррелированных случайных величин;

j - номер отдельной реализации ($j=1,2,\dots,M$)

T - длина реализации;

$\{x(t_i)\}$ - средняя кривая, полученная усреднением по множеству реализаций.

Характерные кривые $l_r(t_i)$ определяют на основании информации, содержащейся в автокорреляционной функции $\chi_{xx}(T)$ причем:

1. Если $\chi_{xx}(T)$ - нестационарна, то $l_r(t_i)$ - отражают тенденции, наиболее часто встречающиеся в структуре реализаций;

2. Если $\chi_{xx}(T)$ - стационарна, то матрица коэффициентов корреляции всех реализаций содержит информацию о наличии характерной колеблемости в структуре реализаций стационарного процесса, а $l_r(t_i)$ - несут информацию о преобладающем спектре частот.

6. Определение связи между переменными.

Об этой области применения факторного анализа мы уже подробно говорили выше, сравнивая методы корреляционного, регрессионного и факторного анализов. Итак, задача определения связей между переменными - лишь одна из множества задач, решаемых столь мощным методом, как факторный анализ.

Остановимся более подробно на применении факторного анализа при изучении взаимосвязей равноправных (с точки зрения причинноследственных связей) переменных - входных или выходных функций системы (1). В этом случае он выступает как альтернатива корреляционного метода. И при изучении причинно-следственных связей выхода от входных функций (2). Тогда он выступает как альтернатива регрессионного анализа.

1. Связь между переменными объясняется наличием общих внутренних факторов, присущих данной системе.

Классическая модель:

$$X_i = a_1 X_1 + \dots + a_{i-1} X_{i-1} + a_{i+1} X_{i+1} + \dots + a_n X_n$$

Используя результаты расчета факторной матрицы исследуемого набора переменных: $X_1, X_2, \dots, X_i, \dots, X_n$

$$X_i = l_{i1} f_1 + l_{i2} f_2 + \dots + l_{ik} f_k + e_i \quad \text{где}$$

$i = 1, 2, \dots, k$ - число общих факторов, а каждый фактор

$$y_r = C_{r1}X_1 + C_{r2}X_2 + \dots + C_{rn}X_n$$

Откуда факторная модель:

$$X_i = \frac{\sum_{r=1}^k \text{liv } C_{r1}}{1 - \sum_{r=1}^k \text{liv } C_{ri}} X_1 + \frac{\sum_{r=1}^k \text{liv } C_{r2}}{1 - \sum_{r=1}^k \text{liv } C_{ri}} X_2 + \dots + \frac{\sum_{r=1}^k \text{liv } C_{rn}}{1 - \sum_{r=1}^k \text{liv } C_{ri}} X_n$$

Далее, используя процедуру вращения, можно добиваться того, чтобы поведение X_i полностью объяснялось одним (или небольшим числом) общим фактором.

Преимущество факторного анализа здесь в том, что он позволяет вскрыть внутренние причины взаимодействия между переменными за счет анализа и интерпретации общих факторов.

2. Анализ взаимосвязи и выходных характеристик.

Часто входные переменные сильно коррелируют и, как правило, измеряются с ошибками, это приводит к некорректности задачи определения уравнения регрессии. Здесь на выручку приходит факторный анализ.

Если в факторную матрицу входных переменных $\{X_i\}$ ввести выходную переменную y , то все рассуждения аналогичны (1), с заменой $X_i \rightarrow y$. Общие факторы здесь содержат информацию о коррелированности входных переменных и их связи с выходом. Специфические факторы берут на себя часть ошибок, поэтому в уравнении регрессии на факторах ошибки не учитываются. Интерпретация общих факторов здесь не допустима.

3.1.2.4 Исходные предпосылки к использованию моделей факторного анализа.

1. Исходный набор наблюдаемых переменных $\{Y\}$ равноправный с точки зрения причинно-следственных связей.
2. Исследуемый набор переменных подчиняется многомерному нормальному распределению или хотя бы удовлетворяет одномерному и двумерному нормальному распределению.
3. Специфические факторы (ошибки) не коррелируют между собой и с общими факторами.

4. Количество общих факторов, наиболее полно определяющих поведение изучаемого явления должно быть невелико.
5. Оператор связи переменных с общим и специфическим факторами должен быть линейным относительно неизвестных коэффициентов (факторных нагрузок).
6. Корреляционная (ковариационная) матрица исходного набора данных устойчива от выборки к выборке.
7. Выборка должна быть представительной, т.е. объем ее определяется из соображений, получения надежных оценок для элементов корреляционной матрицы.
8. В исходных наблюдениях нет автокорреляции (ее наличие не дает возможности использовать имеющиеся приемы проверки статистических гипотез и оценки доверительных интервалов).

Если необходимые требования не выполняются, то полученные результаты — лишь сжатое описание исходной информации в терминах факторного анализа, но в дальнейшем, используя дополнительные методы и приемы можно модель модифицировать, добиваясь выполнения требуемых ограничений.

В нашем случае особое внимание следует уделить выполнению требования коренальности распределения исходных переменных. Во многих случаях эмпирическое распределение ширины колец будет коренальным, там же, где это условие не выполняется, необходимо выяснение причин и устранение их в повторных выборках, либо использовать методические приемы, улучшающие распределение выборок.

3.1.2.5. Модели факторного анализа.

Задачей факторного анализа, как уже отмечалось, является выражение измеряемых переменных X_j в терминах скрытых гипотетических факторов F_i

Простейшей моделью, отличающейся наиболее элегантным описанием и хорошо развитыми методами, является линейная модель. Если даже связь X_j и F_i не является строго линейной, но по крайней мере монотонной, то в этом случае монотонная функция может быть достаточно хорошо аппроксимирована линейной зависимостью. Линейная модель может служить и первым приближением при построении модели, системы.

В рамках линейной модели в зависимости от целей возможны варианты до некоторой степени различные:

1. Метод главных компонент - выделяет и наилучшим образом аппроксимирует максимальные дисперсии.
2. Факторный анализ наилучшим образом аппроксимирует выборочные корреляции.

Модель компонентного анализа (метод главных компонент).

$$X_j = a_{j1}F_1 + a_{j2}F_2 + \dots + a_{jn} \quad [1]$$

$(j = 1, 2, \dots, n)$

$\{F_j\}$ - некоррелированные между собой главные компоненты.

В этой модели не делается никаких гипотез о переменных X , они не обязаны даже быть случайными. Каждая новая компонента дает максимальный вклад в суммарную дисперсию параметров. Обычно на практике оставляют только m , ($m < n$) - компонент, которые отбирают большую часть дисперсии (80-90%), но точная аппроксимация возможна только в случае $m = n$. Переход от измеряемых переменных $\{X_j\}$ к системе главных компонент $\{F_j\}$ осуществляется однозначно на основе корреляционной матрицы R .

Модель классического факторного анализа.

$$X_j = a_{j1}F_1 + a_{j2}F_2 + \dots + a_{jm}F_m + d_jU_j \quad [2]$$

$(j = 1, 2, \dots, n)$

m - число общих факторов ($m \ll n$)

n - число характерных факторов

F_m - общие факторы, учитывающие корреляцию между параметрами, некоррелированы;

U_j - характерные факторы, действуют только в X_j , учитывают оставшуюся, в том числе связанную с различными погрешностями дисперсию.

a_{jm} - нагрузки общих факторов в измеряемых переменных.

В обеих моделях все F_m имеют нулевое среднее и единичную дисперсию.

U_j - независимы между собой и по отношению к F_m

Общее число факторов в последней модели - $(m+n)$ больше n

числа переменных, откуда следует, что как факторы так и их нагрузки не могут быть определены однозначно, а только с точностью до ортогонального вращения, то есть бесконечным числом не противоречащих друг другу способов.

Ортогональное вращение факторов используют для придания решению вида, удобного для интерпретации. Итак, метод главных компонент объясняет полную дисперсию переменных, в то время, как факторный анализ ищет некоторую организацию в данных, объясняя корреляцию между данными.

В зависимости от содержательных особенностей материала, которые позволяют сформулировать рабочие гипотезы и удовлетворить некоторым критериям, осуществляется выбор метода решения корреляционной матрицы и определения факторов и их нагрузок.

5.1.2.6. Выбор метода факторного анализа.

Остановимся несколько более подробно на неопределенности факторных решений. С этой неопределенностью приходится сталкиваться дважды:

на первом этапе, при поиске какого-либо решения, удовлетворяющего модели в статистическом смысле, и на втором этапе, при придании этому решению вида, наиболее удобного с точки зрения интерпретации. Большинство вычислительных процедур факторного анализа дают неоднозначное решение для факторных нагрузок (исключение составляет метод главных компонент).

Метод главных факторов дает единственное решение, но его конечный результат обычно не устраивает специалистов. Поиск содержательного решения/с точки зрения того или иного специалиста/, которое получалось бы из некоторого произвольного факторного решения с помощью вращения, привел к принципу простой структуры. Вращение осей при поиске простой структуры можно понимать как способ уменьшения сложности параметров. Предельным является случай однофакторного решения, когда сложность любого из параметров равна единице.

Методы решения корреляционной матрицы т.е. нахождения факторов могут быть разделены на прямые, основанные на непосредственном анализе корреляционной матрицы, и косвенные, получаемые разного рода преобразованиями прямых решений.

Ряд методов достаточно общи и универсальны. Основное, что требуется от факторного решения - это адекватное описание

взаимосвязей между параметрами, то есть вычисленные коэффициенты корреляции должны хорошо приближать выборочные коэффициенты корреляции. Кроме того, желательно, чтобы факторное решение было как можно более простым и содержательно интерпретируемым.

В настоящее время большинство авторов идут — по стандартной схеме: сначала выборочная корреляционная матрица факторизуется методом главных факторов, затем полученное решение преобразуется к многофакторному виду методом вращения "варимакс". В ряде случаев содержательные особенности изучаемого объекта, позволяющие сформулировать рабочие гипотезы, помогают выбрать и конкретный вид решения. Будем следовать стандартной схеме и опишем кратко метод главных факторов.

Метод главных факторов (МГФ)

МГФ — является развитием фундаментального метода — компонентного анализа. Следовательно нам не обойтись без рассмотрения основных положений компонентного анализа.

В компонентном анализе производится ортогональное преобразование, при котором каждый из p -параметров выражается через главных компонент. Важным свойством компонент является то, что каждая из них по порядку учитывает максимум суммарной дисперсии параметров и не коррелирует со всеми остальными. Сумма дисперсий всех главных компонент равна сумме дисперсий всех исходных параметров.

Поскольку метод связан с суммарной дисперсией параметров, наиболее эффективным он становится в случае, когда все параметры измеряются в одних единицах. Поэтому принято выражать параметры в стандартной форме, при которой дисперсия параметра равна единице (суммарная дисперсия равна p).

Исходным материалом для анализа является корреляционная матрица, на диагонали которой стоят единицы ввиду ее симметричности и положительной определенности все главные компоненты действительны и положительны.

Вообще говоря, под "методом главных факторов" понимают приложения метода главных компонент к редуцированной корреляционной матрице (т.е. к матрице, у которой на главной диагонали вместо единиц стоят значения общностей). Дисперсия и общность параметра \sum_j (в стандартном виде) сведены следующим соотноше-

нием

$$S_j^2 = 1 = a_{j1}^2 + a_{j2}^2 + \dots + a_{jm}^2 + d_j^2$$

/ 3 /

S_j^2 - дисперсия параметра Z_j
 a_{jm}^2 - доли дисперсии параметра Z_j , приходящиеся на соответствующие факторы (F_m).

Тогда общность $h_j^2 = a_{j1}^2 + a_{j2}^2 + \dots + a_{jm}^2$ часть дисперсии, учитываемая факторами ($j=1, 2, \dots, n$), а $d_j^2 = 1 - h_j^2$ характеристность.

В методе главных компонент $d_j^2 = 0 \wedge h_j^2 = 1$ на главной диагонали стоят единицы. В методе главных факторов $h_j^2 < 1$ на d_j^2 , которая показывает, насколько в общих факторах учтена суммарная дисперсия параметра.

В методе главных компонент на первой стадии ищут коэффициенты при первом факторе таким образом, чтобы сумма вкладов фактора в суммарную общность была максимальной. Эта сумма равна:

$$V_1 = a_{11}^2 + a_{21}^2 + \dots + a_{n1}^2$$

/ 4 /

Максимум V_1 должен быть обеспечен при условии, что

$$z_{jk} = \sum_{p=1}^m a_{jp} a_{kp} \quad (j, k = 1, 2, \dots, n)$$

/ 5 /

где $z_{jk} = z_{kj}$, а z_{jj} - общность h_j^2 параметра Z_j

Это условие означает, что выборочные коэффициенты корреляции заменяются вычисленными (z_{jk} - выборочный коэффициент корреляции параметров j, k).

Для того чтобы максимизировать V_1 функцию n переменных a_{j1} связанных дополнительными $\frac{n(n+1)}{2}$ условиями /5/, лучше всего воспользоваться методом множителей Лагранжа.

Составляется функция

$$2T = V_1 - \sum_{j,k=1}^n U_{jk} z_{jk} = V_1 - \sum_{j,k=1}^n \sum_{p=1}^m U_{jk} a_{jp} a_{kp}$$

/ 6 /

где $U_{jk} = U_{kj}$ - множители Лагранжа.

Затем частные производные T по n переменным a_{j1} и остальным a_{jp} ($p \neq 1$) переменным приравняются нулю. Получаются две системы уравнений, после преобразования которых получаем новую систему:

$$\sum_{k=1}^n r_{jk} a_{k1} - \lambda_1 - a_{j1} = 0 \quad / 7 /$$

где
$$\sum_{j=1}^n a_{j1}^2 = \lambda_1 \quad / 8 /$$

Распишем подробно систему:

$$\begin{aligned} (h_1^2 - \lambda) a_{11} + r_{12} a_{21} + r_{13} a_{31} + \dots + r_{1n} a_{n1} &= 0 \\ r_{21} a_{11} + (h_2^2 - \lambda) a_{21} + r_{23} a_{31} + \dots + r_{2n} a_{n1} &= 0 \\ \hline r_{n1} a_{11} + r_{n2} a_{21} + r_{n3} a_{31} + \dots + (h_n^2 - \lambda) a_{n1} &= 0 \quad / 9 / \end{aligned}$$

где параметр λ записан без индекса.

Необходимым и достаточным условием существования нетривиального решения в этой системе и однородных уравнений является равенство нулю детерминанта матрицы коэффициентов этих уравнений, т.е.

$$\begin{vmatrix} h_1^2 - \lambda & r_{12} & r_{13} & \dots & r_{1n} \\ r_{21} & h_2^2 - \lambda & r_{23} & \dots & r_{2n} \\ \hline r_{n1} & r_{n2} & r_{n3} & \dots & h_n^2 - \lambda \end{vmatrix} = 0 \quad / 10 /$$

Получается таким образом полином степени n и относительно λ . Это уравнение называется характеристическим, свойства подобных уравнений хорошо изучены: все корни уравнения действительны.

Они называются "собственными значениями" матрицы коэффициентов. А решение самой системы / 9 / - набор значений a_{kj} , соответствующий каждому собственному значению, называется "собственными векторами матрицы".

Так собственный вектор $\{a_{11}, a_{21}, \dots, a_{n1}\}$, соответствующий наибольшему собственному значению λ_1 , представляет собой коэффициенты a_{j1} при первом факторе F_1 и учитывает максимально возможную долю суммарной общности. Причем максимизируемая функция V_1 равна корню X_1 .

В общем виде задача обычно ставится так:

найти число λ и n -мерный вектор $\bar{q} \neq 0$, такой, при котором

$$\bar{R}\bar{q} = \lambda\bar{q} \quad / 11 /$$

Любое число λ удовлетворяющее этому уравнению, называется собственным значением матрицы \bar{R} , а соответствующий вектор

$q_p = \{a_{1p}, a_{2p}, \dots, a_{np}\}$ и называется собственным вектором матрицы \bar{R} .

(Заметим, кстати, что задача нахождения собственных значений и собственных векторов матрицы является классической в матричной алгебре и хорошо разработана для разных видов матриц. С появлением ЭВМ она эффективно, а главное быстро решается с помощью набора стандартных программ).

Следующая задача состоит в поиске фактора, который учитывал бы максимум остаточной общности. Для этого необходимо вычислить матрицу остаточных коэффициентов корреляции (с учетом только первого фактора). Остаточный коэффициент корреляции:

$$r_{jk} = r_{jk} - a_{j1} a_{k1} = a_{j2} a_{k2} + a_{j3} a_{k3} + \dots + a_{jn} a_{kn} \quad / 12 /$$

В общем виде остаточная матрица запишется так:

$$\bar{R}_1 = \bar{R} - \bar{R}_1 \quad / 13 /$$

где $\tilde{R}_1 = a_1 a_1'$

/ 14 /

есть симметрическая $(n \cdot n)$ матрица коэффициентов корреляции, вычисленных с учетом только первого фактора.

Для определения коэффициентов при втором факторе F_2 необходимо максимизировать функцию

$$V_2 = a_{12}^2 + a_{22}^2 + \dots + a_{n2}^2,$$

/ 15 /

представляющую собой сумму вкладов F_2 в остаточную общность.

Нетрудно доказать, что собственные векторы матрицы \bar{R}_1 идентичны собственным векторам матрицы \bar{R} , как и их собственные значения, за исключением вектора a_1 , собственное значение которого в \bar{R}_1 равно нулю, а в \bar{R} равно λ_1 . Таким образом, максимальное собственное значение матрицы \bar{R}_1 является вторым собственным значением исходной матрицы \bar{R} . Точно так же последующие собственные значения и отвечающие им собственные векторы получаются непосредственно из матрицы \bar{R} . Процесс продолжается вплоть до получения m факторов.

3.1.2.7. Вращение исходного факторного решения - метод "варимакс"

Как уже отмечалось выше, в целях упрощения и получения содержательно интерпретируемых решений, производится ортогональное вращение исходного факторного решения.

Остановимся очень коротко на одном из аналитических методов получения многофакторного ортогонального решения.

Одним из наиболее важных понятий, определяющих выбор из бесконечного многообразия возможных решений некоторого предпочтительного решения, и дающих основание для объективизации принципа простой структуры, является экономия. Мера экономии в задаче вращения можно определить как степень того, насколько упорядочено множество векторов параметров (от совершенно хаотического до идеального, при котором сложность каждого из параметров равна единице).

Существует много вариантов того, каким образом определять

понятие экономии. Например, эффективное и простое определение экономии, как некоторой функции от суммы произведений координат множества коллинеарных точек. Для случая положительных и отрицательных координат, вместо суммы произведений берется сумма квадратов произведений, т.е.

$$\sum_{j=1}^n \sum_{p=1}^m (a_{jp} a_{jq})^2$$

/ 16 /

Для решения общей задачи вращения (в случае ортогональных факторов) вводятся следующие обозначения:

$\bar{A} = (a_{jp})$ - исходная факторная матрица (получена методом главных факторов);

$\bar{B} = (b_{jp})$ - финальная факторная матрица;

$\bar{T} = (t_{qp})$ - ортогональная матрица преобразования.

$$\bar{B} = \bar{A} \cdot \bar{T}$$

/ 17 /

При ортогональном преобразовании \bar{T} матрицы \bar{A} в матрицу \bar{B} , значение общности любого из параметров остается неизменным, т.е.

$$\sum_{p=1}^m b_{jp}^2 = \sum_{p=1}^m a_{jp}^2 = h_j^2 \quad (j = 1, 2, \dots, n) \quad / 18 /$$

Очевидно, квадрат общности также остается постоянным:

$$\left(\sum_{p=1}^m b_{jp}^2 \right)^2 = \sum_{p=1}^m b_{jp}^4 + 2 \sum_{p < q=1}^m b_{jp}^2 b_{jq}^2 = const$$

/ 19 /

Просуммируем это выражение по всем n параметрам:

$$\sum_{j=1}^n \sum_{p=1}^m b_{jp}^4 + 2 \sum_{j=1}^n \sum_{p < q=1}^m b_{jp}^2 b_{jq}^2 = const \quad / 20 /$$

В качестве меры экономии можно принять любой из членов / 20 / или некоторую функцию от них.

Метод "варимакс" делает упор на упрощение столбцов, то есть факторов, факторной матрицы. При этом максимизируется функция

$$V = n \sum_{p=1}^m \sum_{j=1}^n \left(\frac{b_{jp}}{h_j} \right)^2 - \sum_{p=1}^m \left(\sum_{j=1}^n \frac{b_{jp}^2}{h_j^2} \right)^2$$

/ 21 /

Подробное описание алгоритма "варимакс" - вращение приводится ниже при характеристике соответствующей подпрограммы для ЭВМ.

3.1.2.8. Практическое приложение факторного анализа к задачам дендроклиматохронологии

Рассмотрим подробно, какие задачи поможет решить факторный анализ (ϕA) при изучении условий среды профилем методом:

1. Прежде всего, при изучении абсолютных приростов деревьев одной пробной площади ϕA позволит выделить факторы или характерные кривые, типичные для всей совокупности деревьев, затем, возможно групповые факторы, типичные для отдельных групп внутри пробной площади, и, наконец, индивидуальные факторы (сюда может войти возрастная кривая и другие индивидуальные особенности роста дерева).

В результате чего станет ясно, как и на основании какого статистического материала должна быть построена обобщенная хронология пробной площади и какую информацию об особенностях прироста одной пробной площади она несёт.

2. Следующий уровень - изучение изменчивости прироста по профилю, то есть 42 - пробным площадям.

Здесь первый фактор, возможно, явится генеральным и будет характеризовать общую для всего профиля компоненту изменчивости прироста. Далее групповые факторы помогут выделить отдельные районы с различными климатическими или другими географическими особенностями. И, наконец, локальные факторы проинформиру-

ют о характерных кривых каждой пробной площади.

3. Факторный анализ совокупности климатических рядов (среднемесячных температур и осадков) метеостанций, расположенных в районах взятия проб древесины, снабдит нас информацией о факторах или обобщенных климатических показателях и их распределении вдоль изучаемого профиля.

В свою очередь изучение закономерностей распределения в пространстве главных факторов прироста сосны и главных факторов климата позволит установить взаимосвязь приростных характеристик с климатическими уже не на примитивном уровне влияния отдельных месячных климатических показателей, а на уровне одновременного влияния комплекса показателей - факторов, что по существу и наблюдается в природе. Этот этап завершится составлением уравнений регрессии.

1. Ширина кольца как функция климата
2. Климат в зависимости от ширины колец.

Существенным отличием данных уравнений регрессии на факторах от уравнений, построенных при использовании в качестве предикторов непосредственно измеряемых признаков (температуры, осадки, ширина колец) в том, что факторы независимы между собой. Это является необходимым условием применения регрессионного анализа.

3.1.2.9. Описание алгоритма программы FACTO

Блок - схема нулевого уровня программы приведена на странице 9А.

Перейдем к описанию алгоритмов и особенностей отдельных рабочих подпрограмм.

Подпрограмма CORRE

Вычисляет средние, стандартные отклонения, суммы взаимных произведений отклонений от среднего и коэффициенты корреляции.

Входные данные Y_{ij} , где $i=1,2,\dots,n$ - наблюдения, $j=1,2,\dots,m$ - переменные.

Для вычисления этих статистик служат следующие уравнения.

Суммы взаимных произведений отклонений

$$S_{jk} = \sum_{i=1}^n (Y_{ij} - T_j)(Y_{ik} - T_k) - \frac{\sum_{i=1}^n (Y_{ij} - T_j) \cdot \sum_{i=1}^n (Y_{ik} - T_k)}{n}$$

/ 22 /

где $j = 1, 2, \dots, m$; $k = 1, 2, \dots, m$

$$T_j = \frac{\sum_{i=1}^m Y_{ij}}{m}$$

/ 23 /

(Эти временные средние вычитаются из данных в уравнении /22/ для получения центральных моментов).

Средние

$$\bar{Y}_j = \frac{\sum_{i=1}^n Y_{ij}}{n}$$

/ 24 /

где $j = 1, 2, \dots, m$

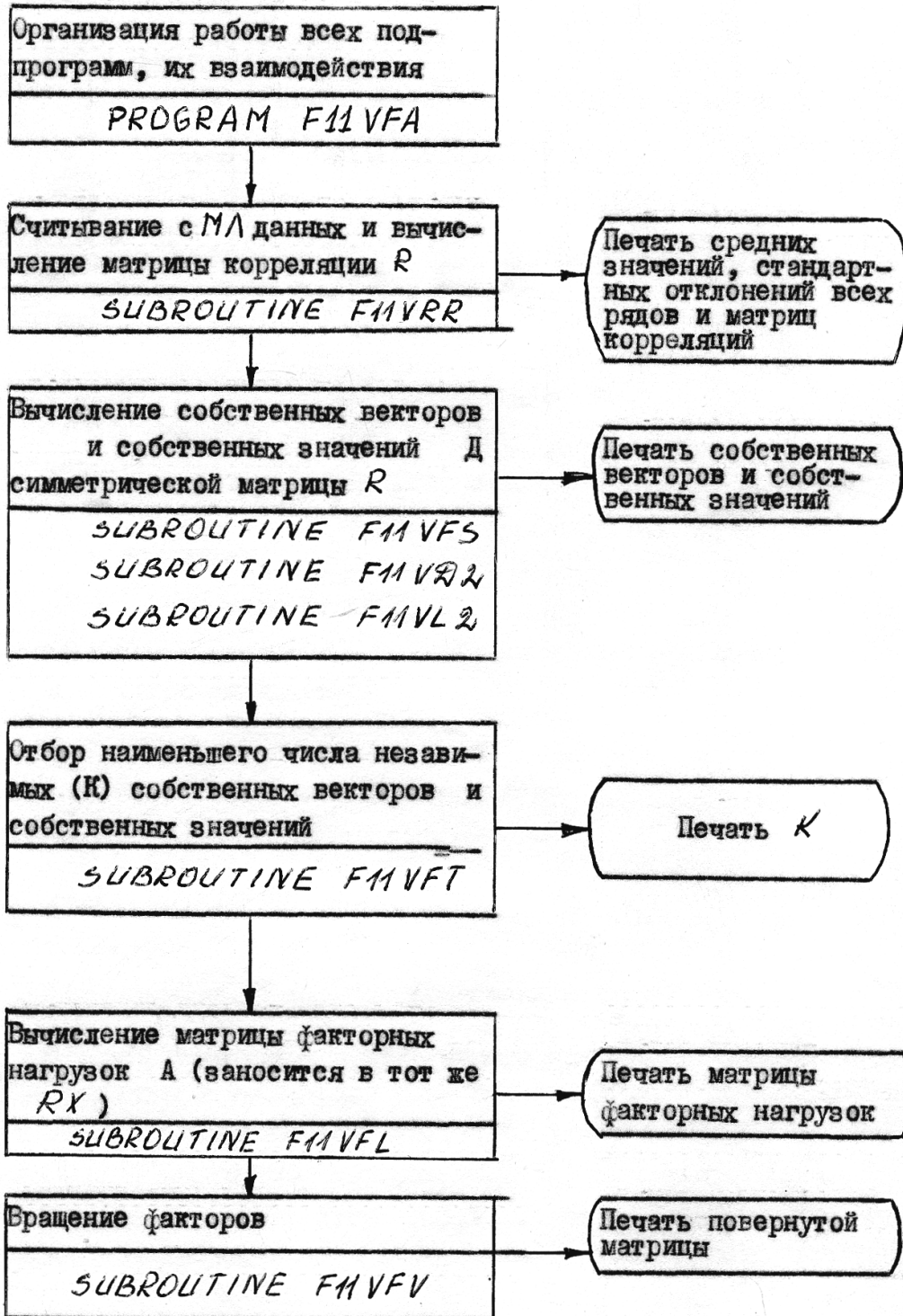
Коэффициенты корреляции

$$r_{jk} = \frac{S_{jk}}{\sqrt{S_{jj}} \cdot \sqrt{S_{kk}}}$$

/ 25 /

где $j = 1, 2, \dots, m$; $k = 1, 2, \dots, m$.

Блок - схема программы ФАСТО



Стандартные отклонения

$$S_j = \frac{\sqrt{S_{jj}}}{\sqrt{n-1}},$$

/ 26 /

где $j = 1, 2, \dots, m$

Вызывается в основной программе FACTO с помощью оператора
CALL CORRRE (X, XBAR, STD, RX, R, D, B, T, CORKO,
N, M, NM, MQV, MM1)

В кодах машины M4030 CORRRE эквивалентно FUVFC

Где:

X - матрица наблюдений;

XBAR - вектор средних значений;

STD - вектор стандартных отклонений;

RX - матрица суммы взаимных произведений, отклонений от среднего;

R, D, B, T - рабочие векторы;

CORKO - полная матрица коэффициентов корреляции;

N - число наблюдений;

M - число переменных;

$$NM = N * M;$$

$$MQV = M * M;$$

$$MM1 = \frac{(M+1) * M}{2}$$

Подпрограмма RR (описание и блок-схема)

С целью сокращения объема памяти при вводе больших массивов информации и вычислении корреляционной матрицы были составлены основная подпрограмма RR (FUVRR) и вспомогательные WM (F11 V WM)

Подпрограммы WM (где $M=1, 2, \dots, N$) - меняется при оформлении записи новых единиц информации) осуществляют запись

на магнитную ленту по определенному закону дендроданных и других экспериментальных материалов.

Подпрограмма RR осуществляет считывание с магнитной ленты по два дендроряда одновременно, вычисление коэффициента корреляции (смотри алгоритм подпрограммы $L\phi RRE$) рядов и занесение его в определенную ячейку памяти, отведенной для корреляционной матрицы. Перебор всех пар рядов, требуемых в задаче, постепенно заполняет матрицу корреляций, которая и является исходной рабочей единицей в факторном анализе. Одновременный же выбор из внешнего запоминающего устройства только двух рядов экономит оперативную память в $\frac{M}{2}$ раз, то есть для $M=100$ (рядов), экономия составит 50 раз. При этом одновременно заметно возрастает время счета задачи.

В программе RR все вычисления производятся с обычной точностью, а на выходе из нее происходит переприсвоение результата массиву с двойной точностью, для перехода в дальнейших подпрограммах к вычислениям с двойной точностью.

Вызов подпрограммы осуществляется в основной программе оператором $CALL FMVRR(M, R)$

где M - общее число рядов,

$R(M, M)$ - матрица коэффициентов корреляции.

Подпрограмма SYNEIG

Вычисляет все собственные значения и собственные вектора симметричной действительной матрицы.

Для решения проблемы собственных значений здесь используется QL - алгоритм со сдвигом, который очень сложен и мало нагляден. Мы приведем здесь другой алгоритм для демонстрации хода рассуждений.

Для заданной симметрической матрицы A порядка N собственные значения образуются на диагональных элементах матрицы. В результате решения также образуется матрица собственных векторов R .

В качестве первого приближения R используется единичная матрица. Вычисляется начальная норма вне диагонали:

$$v_I = \left\{ \sum_{i \neq k} 2A_{ik}^2 \right\}^{1/2}$$

/ 27 /

где V_I - начальная норма;
 A - входная матрица (симметрическая).

Эта норма делится на N и на каждом шаге образует пороговое значение.

Вычисляется конечная норма

$$V_F = \frac{V_I * 10^{-6}}{N} \quad / 28 /$$

Эта конечная норма выбрана достаточно малой с тем, чтобы любой внедиагональный элемент A_{em} был меньше по абсолютной величине V_F , определяющей сходимость процесса.

Образуется индикатор. Этот индикатор в последующем служит для определения, существует ли какой-нибудь внедиагональный элемент, рассчитанный при преобразовании, который больше данного порогового значения.

Проверяются все внедиагональные элементы и преобразования выполняются до уничтожения внедиагонального (ведущего) элемента, как показано на следующих уравнениях:

$$\lambda = -A_{lm} \quad / 29 /$$

$$\mu = \frac{1}{2} (A_{ee} - A_{mm}) \quad / 30 /$$

$$\omega = \text{sign}(\mu) \frac{\lambda}{\sqrt{\lambda^2 + \mu^2}} \quad / 31 /$$

$$\sin \theta = \frac{\omega}{\sqrt{2(1 + \sqrt{1 - \omega^2})}} \quad / 32 /$$

$$\cos \theta = \sqrt{1 - \sin^2 \theta} \quad / 33 /$$

$$B = A_{ie} \cos \theta - A_{im} \sin \theta \quad / 34 /$$

$$C = A_{ie} \sin \theta + A_{im} \cos \theta \quad / 35 /$$

$$B = R_{ie} \cos \theta - R_{im} \sin \theta \quad / 36 /$$

$$R_{im} = R_{ie} \sin \theta + R_{im} \cos \theta \quad / 37 /$$

$$R_{ie} = B \quad / 38 /$$

$$A_{ee} = A_{ie} \cos^2 \theta + A_{im} \sin^2 \theta - 2A_{im} \sin \theta \cdot \cos \theta \quad / 39 /$$

$$A_{mm} = A_{ee} \sin^2 \theta + A_{im} \cos^2 \theta + 2A_{im} \sin \theta \cdot \cos \theta \quad / 40 /$$

$$A_{em} = (A_{ee} - A_{mm}) \sin \theta \cdot \cos \theta + A_{im} (\cos^2 \theta - \sin^2 \theta) \quad / 41 /$$

Эти вычисления повторяются до тех пор, пока ведущие элементы не станут меньше порогового значения,

Вызывается в основной программе FACT ϕ с помощью оператора:

(SYMEIG = FMVFS)

CALL SYMEIG (M, CON, D, RX, 250)

где

D - массив, содержащий собственные значения в порядке убывания;

RX - массив собственных векторов

Подпрограмма TRACE

Эта подпрограмма находит K - число собственных значений, больших либо равных значению заданной константы (CON). Значение CON определяется выбором метода факторного анализа. Заданные собственные значения должны быть расположены в убывающем порядке.

С целью сохранения наименьшего возможного числа независимых размерностей в анализ вводится только те собственные значения, которые больше, либо равны (1).

Вызывается в основной программе FACTO с помощью оператора:

(TRACE = FMVFT) CALL TRACE (M₁, D, 1, K).

Подпрограмма LOAD

Вычисляет коэффициенты каждого фактора:

$$a_{ij} = v_{ij} \sqrt{\lambda_j}$$

где

v_{ij} - элементы нормализованного собственного вектора,

λ_i - соответствующие собственные значения,

$i = 1, 2, \dots, m$ - переменные;

$j = 1, 2, \dots, k$ - оставленные собственные значения.

Вызывается из основной программы FACTO обращением:

(LOAD = FMVFL) CALL LOAD (M, K, D, RX), где

RX - при обращении содержит собственные векторы, расположенные по столбцам.

При выходе из подпрограммы содержит матрицу факторных коэффициентов.

Подпрограмма (вспомогательная) ARRAY

Используется для преобразования одномерного массива в двумерный и наоборот.

Вызывается из основной программы: (ARRAY = FMVAR)

CALL ARRAY (2, M, K, M, M, RY, RX)

где

RX - двумерный массив, а RY - одномерный того же объема.

Подпрограмма VARMX

Эта подпрограмма выполняет ортогональное вращение матрицы факторов размерам $m \times K$, так что

$$\sum_i \left\{ m \sum_j (a_{ij}^2 / h_i^2) - \left[\sum_j (a_{ij}^2 / h_i^2) \right]^2 \right\}$$

/ 42 /

будет максимальным, где $i = 1, 2, \dots, m$ - переменные;

$j = 1, 2, \dots, K$ - факторы;

a_{ij} - нагрузка i -й переменной j -го фактора и

h_i - суммарная нагрузка i -й переменной, определяемая ниже.

Как уже упоминалось выше факторное решение определяется с точностью до ортогонального вращения. Вращение позволяет следовательно привести решение к виду, удобному для интерпретации.

Суммарная нагрузка

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

/ 43 /

где $i = 1, 2, \dots, m$

Нормализованная матрица факторов

$$b_{ij} = a_{ij} / \sqrt{h_i}$$

/ 44 /

где $i = 1, 2, \dots, m$

$j = 1, 2, \dots, K.$

Вариация матрицы факторов

$$V_c = \sum_j \left\{ \left[m \sum_i (b_{ij}^2)^2 - \left(\sum_i b_{ij}^2 \right)^2 \right] / m^2 \right\}$$

/ 45 /

где $C = 1, 2, \dots$ (итерационный цикл).

Проверка на сходимость

Если $V_c - V_{c-1} \leq 10^{-7}$

/ 46 /

для последовательных результатов, то программа прекращает вращение и выполняет расчет по уравнению /69/. В противном случае вращение повторяется, пока проверка на сходимость не будет удовлетворена.

Вращение двух факторов:

Подпрограмма одновременно вращает два нормализованных фактора (b_{ij}). 1 с $2, 1$ с $3, \dots, 1$ с $K, 2$ с $3, \dots, 2$ с $K, \dots, K-1$ с K . Это образует итерационный цикл.

В предположении, что X и Y - вращаемые факторы, причем x - фактор, имеющий меньший индекс или находящийся слева. Вращение этих двух факторов записывается в следующей формуле:

$$\begin{bmatrix} X_1 & Y_1 \\ X_2 & Y_2 \\ \vdots & \vdots \\ X_m & Y_m \end{bmatrix} \cdot \begin{bmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{bmatrix} = \begin{bmatrix} X_1 & Y_1 \\ X_2 & Y_2 \\ \vdots & \vdots \\ X_m & Y_m \end{bmatrix}$$

/ 47 /

где

x_i и y_i - имеющиеся в данный момент нормализованные нагрузки,

а

X_i и Y_i - требуемые нормализованные нагрузки функциями

φ - угла вращения. Шаги вычисления от (а) до

(е) приведены ниже:

а) вычисление NUM и DEN

$$A = \sum_i (x_i + y_i)(x_i - y_i)$$

$$B = 2 \sum_i x_i y_i \quad / 48 /$$

$$C = \sum_i [(x_i + y_i)(x_i - y_i) + 2x_i y_i][(x_i + y_i)(x_i - y_i) - 2x_i y_i]$$

$$D = 4 \sum_i (x_i + y_i)(x_i - y_i) x_i y_i$$

$$NUM = D - 2AB/m$$

$$DEN = C - [(A+B)(A-B)]/m$$

б) Сравнение NUM и DEN

При этом могут возникнуть следующие четыре случая:

$NUM < DEN$ - перейти ниже к (b1);

$NUM > DEN$ - перейти ниже к (b2);

$(NUM + DEN) \geq \varepsilon^*$ - перейти ниже к (b3).

$1 * \varepsilon$ - малый допуск на фактор/

$(NUM + DEN) < \varepsilon$ - перейти к следующему вращению:

$$(b1) \tan 4\theta = |NUM|/|DEN| \quad / 49 /$$

Если $\tan 4\theta < \varepsilon$ и

- 1) REN положительно, перейти к следующему вращению,
- 2) REN отрицательно, положить $\cos \varphi = \sin \varphi = \frac{\sqrt{2}}{2}$
и перейти к пункту (е).

Если $\tan 4\theta \geq \varepsilon$, вычислить:

$$\cos 4\theta = 1 / \sqrt{1 + \tan^2 4\theta} \quad / 50 /$$

$$\sin 4\theta = \tan 4\theta \cdot \cos 4\theta \quad / 51 /$$

и перейти к пункту (Г);

$$(б2) \operatorname{ctn} 4\theta = |NLM| / |DEN| \quad / 52 /$$

Если $\operatorname{ctn} 4\theta < \varepsilon$, положить $\cos 4\theta = 0$ и $\sin 4\theta = 1$.

Перейти к пункту (Г).

Если $\operatorname{ctn} 4\theta \geq \varepsilon$, вычислить:

$$\sin 4\theta = 1 / \sqrt{1 + \operatorname{ctn}^2 4\theta} \quad / 53 /$$

$$\cos 4\theta = \operatorname{ctn} 4\theta \cdot \sin 4\theta \quad / 54 /$$

и перейти к пункту (Г);

$$(б3) \text{ положить } \cos 4\theta = \sin 4\theta = (\sqrt{2})/2$$

и перейти к пункту (Г);

(Г) определение $\cos \theta$ и $\sin \theta$

$$\cos 2\theta = \sqrt{(1 + \cos 4\theta) / 2} \quad / 55 /$$

$$\sin 2\theta = \sin 4\theta / 2 \cos 2\theta \quad / 56 /$$

$$\cos \theta = \sqrt{(1 + \cos 2\theta) / 2}$$

$$\sin \theta = \sin 2\theta / 2 \cos \theta \quad / 57 /$$

/ 58 /

(g) определение $\cos \varphi$ и $\sin \varphi$

(g1) если ΔEN положительно, принять

$$\cos \varphi = \cos \theta$$

/ 59 /

$$\sin \varphi = \sin \theta$$

/ 60 /

и перейти к пункту (g2).

Если ΔEN отрицательно, вычислить

$$\cos \varphi = \frac{\sqrt{2}}{2} \cos \theta + \frac{\sqrt{2}}{2} \sin \theta \quad / 61 /$$

$$\sin \varphi = \left| \frac{\sqrt{2}}{2} \cos \theta - \frac{\sqrt{2}}{2} \sin \theta \right| \quad / 62 /$$

и перейти к пункту (g2);

(g2) если $N \perp M$ положительно, положить

$$\cos \varphi = |\cos \varphi| \quad / 63 /$$

$$\sin \varphi = |\sin \varphi| \quad / 64 /$$

и перейти к пункту (e).

Если N и M отрицательно, положить

$$\cos \varphi = |\cos \varphi| \quad / 65 /$$

$$\sin \varphi = -|\sin \varphi| \quad / 66 /$$

(e) вращение

$$X_i = x_i \cos \varphi + y_i \sin \varphi \quad / 67 /$$

$$Y_i = x_i \sin \varphi + y_i \cos \varphi \quad / 68 /$$

где $i = 1, 2, \dots, m$

После того как один цикл с $K(K-1)/2$ вращениями закончен, подпрограмма возвращается к вычислению вариации для матрицы факторов по уравнению / 45 /.

Денормализация

$$a_{ij} = b_{ij} \cdot h_i \quad / 69 /$$

где $i = 1, 2, \dots, m$;

$j = 1, 2, \dots, K$.

Проверка суммарных весов:

Окончательный суммарный вес

$$f_i^2 = \sum_{j=1}^K a_{ij}^2 \quad / 70 /$$

Разность $d_i = h_i^2 - f_i^2 \quad / 71 /$

где $i = 1, 2, \dots, m$;

Вызывается из основной подпрограммы:

(VAR MX = FUVFV)

CALL VAR MX (M, K, MK, RY, NC, TV, B, T, D, IER)

где $MK = M \cdot K$; RY - при входе исходная матрица факторов,
при выходе из подпрограммы - обращенная матрица факторов;

NC - выводимая переменная, содержит число выполненных циклов,

TV, K, F, D - вспомогательные информативные векторы;
 IER - индикатор ошибок ($IER=0$ - нет ошибок;

$IER=1$ - сходимость не достигается за 50 циклов вращения).

Уменьшение оперативной памяти

Все вычисления необходимо проводить с двойной точностью, то есть сохранять максимальное количество знаков после запятой. Особенно это касается задач матричной алгебры, где для матриц высокого порядка при большинстве операций производится большое число элементарных арифметических действий с малыми величинами. А это означает увеличение как минимум памяти в два раза.

В этой ситуации очень важно применить все возможные приемы для сокращения объема оперативной памяти:

1. Использовать при работе разных подпрограмм одни и те же блоки памяти (*COMMON* - блоки);
2. Внутри подпрограмм операторы *EQUIVALENCE*, позволяющие использовать одни и те же места памяти для разных массивов;
3. Использовать внешние накопители - магнитные ленты и т.д.

С учетом всего сказанного была осуществлена переработка всех подпрограмм и самой программы, а также составлена новая подпрограмма *RR* для считывания рядов парами с магнитной ленты и вычисления коэффициентов корреляции.