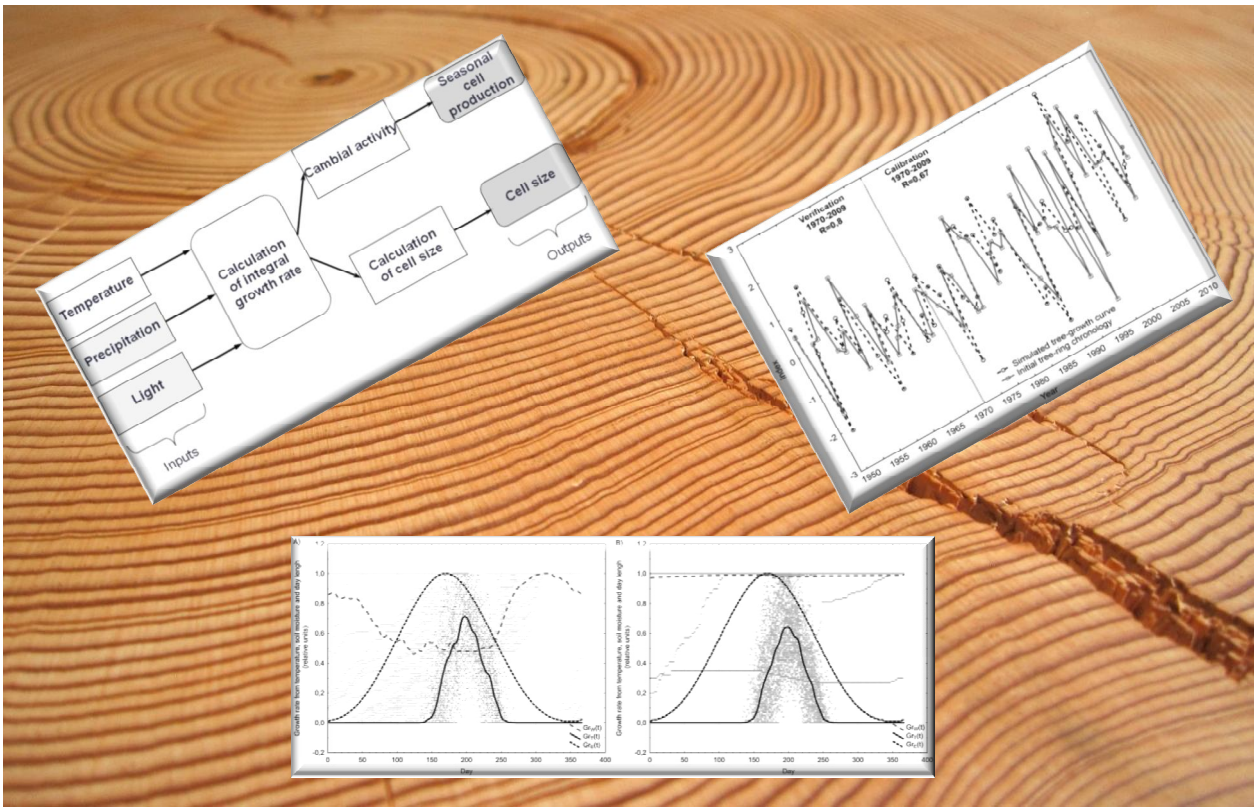
СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ
SIBERIAN FEDERAL UNIVERSITY

Шишов В.В., Тычков И.И., Попкова М.И.

МЕТОДЫ АНАЛИЗА ДЕНДРОКЛИМАТИЧЕСКИХ ДАННЫХ И ИХ ПРИМЕНЕНИЕ ДЛЯ ТЕРРИТОРИИ СИБИРИ.

Учебное пособие



МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

МЕТОДЫ АНАЛИЗА ДЕНДРОКЛИМАТИЧЕСКИХ ДАННЫХ И ИХ ПРИМЕНЕНИЕ ДЛЯ ТЕРРИТОРИИ СИБИРИ.

Учебное пособие

Шишов В.В., Тычков И.И., Попкова М.И.

Красноярск 2015

УДК 51-76, 574.9
ББК 20.1-22.171

Рецензенты:

С.И. Сенашов, д-р физ.-мат. наук, проф., Сибирский государственный аэрокосмический университет им. М.Ф. Решетнева

В.Г. Подопригора, д-р физ.-мат. наук, проф., ФГАОУ ВПО «Сибирский федеральный университет»

Шишов В.В.

МЕТОДЫ АНАЛИЗА ДЕНДРОКЛИМАТИЧЕСКИХ ДАННЫХ И ИХ ПРИМЕНЕНИЕ ДЛЯ ТЕРРИТОРИИ СИБИРИ: УЧЕБНОЕ ПОСОБИЕ/ **В.В.** *Шишов, И.И. Тычков, М.И. Попкова.* ФГАОУ ВПО «Сибирский федеральный университет». – Красноярск, 2015 – 210с.

Книга представляет собой современный обзор математических методов анализа данных, применяемых в дендрохронологических, денроклиматических и дендроэкологических исследованиях. Кроме теоретического обзора, в учебном пособии представлены примеры, описывающие практическое применение данных методов с соответствующей интерпретацией полученных результатов. Предложенные методы по выявлению различных пространственно-временных закономерностей представляют интерес для специалистов, занимающихся анализом пространственно-распределенных систем в биологии, экономике и социологии

Книга предназначена для специалистов, магистратов и аспирантов высших учебных заведений.

Работа выполнена при поддержке гранта Российского научного фонда (проект № 14-14-00219)

ISBN 978-5-906498-15-1

© ФГАОУ ВПО «Сибирский федеральный университет», 2015

© Шишов В.В., Тычков И.И., Попкова М.И., 2015

Оглавление

Предисловие……..... 5

Введение……. 6

П.1 Проблема оценки климатического сигнала на основе дендрохронологических данных……... 9

П.2 Проблемы стандартизации дендрохронологических данных…… 22

П.3 Множественный регрессионный анализ дендроклиматических данных. 30

П.4 Пример использования множественной регрессии при пространственно-временном моделировании прироста древесных растений на территории Сибири……... 43

П.5 Нелинейное оценивание прироста древесных растений. Описание имитационной VS-модели……. 69

П.6 Скользящие ранговые коэффициенты Спирмена, Кендалла, Хубера…... 76

П.7 Пример использования скользящих коэффициентов осредненной корреляции Пирсона и конкордации Кендалла. Сравнительный анализ..... 89

П.8 Алгоритмы теории классификации. D-, S-, X-проблемы……... 99

П.9 Пример использования метода главных компонент при анализе пространственных взаимоотношений между приростом древесных растений и общим содержанием озона в верхних слоях атмосферы…….. 115

П.10 Пример использования скользящего коэффициента конкордации Кендалла, псевдостатистического метода Байеса и метода случайного поиска с адаптацией при выявлении взаимосвязи между интенсивностями проявления Эль-Ниньо и пространственным приростом древесных растений на территории Сибири……. 122

П.11 Пример использования кластерного и дискриминантного анализов при пространственно-временном анализе трендов прироста древесных растений на территории Сибири и Дальнего Востока…….. 156

П.12 Пример использования метода главных компонент и дискриминантного анализа для реконструкции термических характеристик

сезона роста по клеточным измерениям годичных колец…… 176

Заключение…… 190

Литература……...……. 195

Предисловие

В книге дается обзор современных проблем дендрохронологии, дендроклиматологии и дендроэкологии, а также математических методов, позволяющих разрешить некоторые из них. В частности, описывается проблема «дивергенции», возникающая при анализе дендроклиматических данных. Она заключается в несогласованности динамики летней температуры и реального прироста древесных растений после 1960-х годов для регионов, где температура выступает основным фактором, лимитирующим рост деревьев. Приводится множество непротиворечивых гипотез, объясняющих возникновение этой проблемы. Многие гипотезы связаны с методами стандартизации дендрохронологических данных и линейного регрессионного анализа дендроклиматических рядов, которые не способны адекватно описать нелинейные взаимодействия и системные связи между приростом древесных растений и лимитирующими прирост факторами климатической природы. Решение подобных проблем возможно при применении системного подхода к анализу дендроклиматической информации на основе спроектированной интерактивной системы дендроклиматического мониторинга, а также использованием новых методов анализа информации, наряду с традиционно используемыми в дендрохронологии.

В книге, наряду с традиционно используемыми методами анализа данных, описываются новые для дендрохронологии и дендроклиматологии алгоритмы расчета различных скользящих коэффициентов, основанных на известных в литературе и новых ранговых статистиках. Формулируются основные проблемы теории классификации, описываются алгоритмы для их решения.

Практически каждый алгоритм подкреплен примером его применения при решении различных дендроклиматических задач, связанных с пространственно-временным анализом и реконструкцией дендрохронологических данных для территории Сибири на базе рассмотренных скользящих ранговых статистик и классификационных методов.

Введение

Выявление и анализ системных взаимосвязей между элементами различных пространственно-распределенных систем по рядам наблюдений их переменных, контролируемых в конечном множестве точек пространственных координат, является неотъемлемой задачей современной глобальной экологии (Ваганов и др., 1996; 2000; Ваганов, Шиятов, 1999; Мазепа, 1999а; 1999б; Хантемиров, Шиятов, 1999; Шиятов, 1986; Шиятов, Ваганов, 1998; Briffa et al., 1998а, 1998б, 1998в, 2004, 2008; Cook et al., 1987; Fritts et al., 1976). Как правило, такие системы, рассматриваемыми в этой области науки, являются неопределенными, т.е. системами, исследование которых осуществляется в условиях неполной информации о закономерностях их функционирования (Лапко, Лапко, 2007).

Примерами таких систем являются биосфера и, в частности, лесные экосистемы, как важнейшие ее составляющие. Их функционирование и взаимоотношения с другими компонентами Земли изучены недостаточно. Кроме этого, взаимодействие лесных экосистем с другими глобальными системами (например, солнечной иррадиацией) носит нелинейный характер, что значительно усложняет решение задач по выявлению системных взаимоотношений.

Современные состояния биосферы или глобальные биосферные изменения конца XX века являются результатом нелинейных взаимоотношений естественных и антропогенных причин. Эти изменения отражаются, например, в уникальном 30-летнем повышении глобальной приземной температуры в позднем голоцене Земли, современном увеличении концентрации различных парниковых газов, прежде всего, CO_2, увеличении числа и интенсивности различных катастрофических событий (землетрясений, ураганов, штормов, наводнений) в XX век и т.п. (Jones et al., 2001; Jones, 2002; Hulme, Jones, 1994; Kelly et al., 1996; Thorne et al., 2003; Damon, Peristykh, 2005; IV IPCC report, 2007).

С начала 90-х годов XX века в России формируется единая государственная система экологического мониторинга, и в её составе - раздел дендрохронологического и дендроклиматического мониторинга, то есть

«информационная система слежения, оценки и прогноза изменений годичного прироста деревьев и определяющих этот прирост факторов» (Ваганов и др., 1996; 2000; Ваганов, Шиятов, 1999; Шиятов, 1986; Шиятов, Ваганов, 1998).

В настоящее время накоплены уникальные пространственно-распределенные дендроклиматические данные (древесно-кольцевые хронологии и данные метеорологических станций), характеризующие динамическое состояние лесных экосистем на территории Российской Федерации под воздействием факторов климатической природы (Ваганов и др., 1996; 2000; Ваганов, Шиятов, 1999; Мазепа, 1999а; 1999б; Хантемиров, Шиятов, 1999; Шиятов, 1986; Шиятов, Ваганов, 1998; Шишов, 2000; Шишов и др., 2002; Briffa et al., 1998, 2004; Cook et al., 1987; Fritts et al., 1971).

Выбор конкретного подхода к решению задач по выявлению и анализу системных связей между биологическими неопределенными системами и их компонентами обуславливается первичным набором данных, знанием основ функционирования изучаемых систем (например, общих принципов функционирования лесной экосистемы), внутренних и внешних факторов, оказывающие существенное влияние на поведение таких систем.

Современные математические подходы по решению подобных задач в условиях неполной эмпирической информации успешно применяются при анализе аэрокосмической информации, моделировании медико-биологических и социально-экономических процессов (Загоруйко, 1972; 1999; Лапко и др., 2000; Лапко, Лапко, 2007; Лбов, 1981). При этом особое внимание уделяется алгоритмам восстановления стохастических зависимостей, обеспечивающих учет частичных сведений об их виде и данных экспериментальных исследований.

Предлагаемая работа посвящена развитию и теоретическому обоснованию нового математического подхода по анализу неопределённых систем в дендроклиматологии на базе традиционных и новых методов параметрической и непараметрической статистики, а также современных информационных технологий.

Полученные при этом научные результаты создают методическую и математическую основу автоматизации решения подобного рода проблем на основе соответствующих специальных информационных систем обработки информации, которые позволяли бы решать плохо формализованные задачи в полуавтоматическом режиме.

Актуальность темы исследования признана Российским научным фондом и поддержана грантом РНФ (проект № 14-14-00219).

Описанные в книге результаты, связанные с визуализацией и анализом дендроклиматической информации, были использованы Центром климатических исследований университета Восточной Англии (Climatic Research Unit, University of East Anglia, UK) при региональной верификации Германо-Английской глобальной климатической модели GSMs в рамках европейского проекта по моделированию наблюдений и палеоклимату (SO&P) (http://www.cru.uea.ac.uk/cru/projects/soap/).

Предложенные алгоритмы обработки информации использованы для регионального тестирования глобальной атмосферно-океанической циркуляционной модели HadCM3 британского климатического центра Хадли (UK Hadley Centre).

Предложенные методы по выявлению различных пространственно-временных закономерностей представляют интерес для специалистов, занимающихся анализом пространственно-распределенных систем в биологии, экономике и социологии.

П.1 Проблема оценки климатического сигнала на основе дендрохронологических данных.

Основными целями дендрохронологических и дендроклиматических исследований являются:

- выявление значимых статистических и функциональных взаимосвязей между факторами климатической природы и годичным приростом древесных растений (Ваганов и др., 1996; 2000; Ваганов, Шиятов, 1999; Джансеитов, Шишов 1993; 1999; Мазепа, 1999а; 1999б; Сидорова и др., 2007; Хантемиров, Шиятов, 1999; Шиятов, 1986; Шиятов, Ваганов, 1998; Шишов, 2000; Шишов и др., 2002; 2007а; 2007б; Шишов, Кирдянов, 2003; Briffa et. al., 1998, 2004, 2008; Cook et al., 1987; Fritts et. al, 1971; Fritts et al., 1991; Fritts, Shashkin, 1995; Hughes, 1995; Hughes et al., 1999; Cook, Kairiukstis, 1990; Vaganov, Hughes, 2000; Vaganov et al., 2006);
- надежная погодичная количественная оценка прошлых и настоящих климатических изменений на основе длительных календарно-датированных древесно-кольцевых хронологий с целью нахождения общих закономерностей в изменения климата (Ваганов и др., 1996; Мазепа, 1999б; 2000; Хантемиров, 1999, 2000; Хантемиров и др, 1999; Шишов, Кирдянов, 2003; Шиятов и др., 1996; Briffa et. al., 1992, 1995, 1998, 2004, 2008; Fritts, 1969; Fritts et. al., 1971; Hughes et al., 1999; Liu Yu, 2004а, 2004 б; Cook, Kairiukstis, 1990; Naurzbaev et al., 2002; Naurzbaev, Vaganov, 2000);
- краткосрочный прогноз данных изменений в будущее с учетом антропогенной составляющей и анализ влияние данных изменений на лесные экосистемы (Ваганов, Шашкин, 2000; Fritts, Shashkin, 1995; Liu Yu, 2004б; Vaganov et al., 2006).

При этом, древесно-кольцевые хронологии (ДКХ) представляют собой временные ряды (ВР) или случайные функции (СФ) дискретного (натурального) аргумента, которые несут информацию о разнообразных факторах внешней среды, в том числе, и климатической природы. Причем реализации временных рядов могут состоять из значений различных характеристик годичного кольца древесного растения, каждая из которых несет определенную информацию о

специфике роста древесных растений и факторов окружающей среды, влияющих на прирост. В дендроклиматических исследованиях наиболее часто анализируется прирост деревьев, характеризующийся шириной годичного кольца (Ваганов и др., 1996; 2000; Ваганов, Шиятов, 1999; Мазепа, 1999а; 1999б; Хантемиров, Шиятов, 1999; Шиятов, 1986; Шиятов, Ваганов, 1998; Шишов, 2000; Шишов и др., 2002; Briffa et. al., 1998a; 2004; 2008; Cook et al., 1987; Fritts et. al, 1971). В ряде работ была показана возможность использования других независимых характеристик структуры годичных колец в анализе особенностей сезона роста на севере. Ваганов и др. (1999, 2006), Ваганов, Шашкин (2000) выявили высокую положительную связь размеров клеток, толщины клеточной стенки и максимальной плотности древесины лиственницы на севере Евразии с температурой разных промежутков вегетационного периода и, прежде всего, его начала. Связь параметров структуры годичных колец хвойных с климатом и климатическими изменениями анализировалась для различных регионов Земного шара (Ваганов, Пак, 1995; Джансеитов, Шишов, 1999; Терсков и др., 1981; Шишов, 2000; Ваганов и др., 2000; Шишов, Кирдянов, 2003; Yasue et al., 1996, 2000). Очевидно, что использование дополнительных характеристик структуры годичного кольца приводит к получению новой информации об окружающей среде и климате. Отметим, что новые независимые характеристики могут быть получены из уже используемых при помощи математических и статистических преобразований (Briffa et al., 2008).

Интерес к климатическим изменениям вызван неординарными современными изменениями климата Земли (IPCC report, 2007). В планетарном изменении концентрации углекислоты (и других парниковых газов) выделяются два периода: нарастание концентрации с малой скоростью (до 1960-х) и увеличение скорости роста концентрации углекислоты в последние десятилетия (Jones et al., 2001; Jones, 2002; Hulme, Jones, 1994; Kelly et al., 1996; Thorne et al., 2003; Damon, Peristykh, 2005; IPCC report, 2007). Это отражается и в увеличении средней температуры, которое наиболее ярко в глобальном масштабе отражается в Северном полушарии (Jones, Briffa, 1992; Jones et. al, 2001; Jones, 2002; IPCC report, 2007). Этот температурный феномен называется «хоккейной клюшкой»,

основной причиной возникновения которого считается антропогенная составляющая (Jones et al., 2001; Jones, 2002; IPCC report, 2007). Как современные климатические изменения могут влиять на лесные экосистемы? Каковы механизмы такого влияния? Были ли подобные изменения в прошлом и как они отражались на приросте древесных растений? Как подобного рода изменения могут повлиять на развитие лесных экосистем в будущем? Ответы на эти вопросы непросты и неоднозначны.

Пространственный анализ трендов NDVI как индикаторов фотосинтетической активности растительности выявил районы земного шара с увеличением и, наоборот, снижением первичной продуктивности, однако причины возникновения такой активности не ясны (Myneni et al, 1997, 1998; Zhou et al., 2001).

Результаты дендроклиматического анализа, полученные по сети станций дендроклиматического мониторинга для высоких широт севера Евразии с использованием температуро-чувствительных древесно-кольцевых хронологий, свидетельствуют о неоднородности реакции прироста древесных растений на повышение температуры.

Так, пространственный анализ длительных древесно-кольцевых хронологий в средних и высоких широтах Северного полушария на базе множественного регрессионного анализа выявил расхождения в ходе увеличения температуры и реального прироста древесных растений после 1960-х годов для некоторых регионов, в которых температура выступает основным лимитирующим рост фактором (Шишов и др., 2007 а; Briffa et al, 1998; 2004; Vaganov et al., 1999; D'Arrigo et al., 2004, 2007; Vaganov et al., 2006). Это расхождение в литературе получило название «проблемы дивергенции (расхождения)» прироста древесных растений и хода летних температур.

Понимание этой проблемы и ее решение чрезвычайно важно при построении адекватных статистических моделей реконструкции климатических переменных по дендрохронологическим данным (Briffa et al., 1998в; 2008). По сути, эта проблема ставит под сомнение реальность температурных оценок в прошлом (температурных реконструкций) по дендрохронологическим данным

на базе выявленных корреляционных взаимосвязей в настоящем. Так, для последних десятилетий регрессионные климатические модели завышают величины прироста по сравнению с наблюдаемыми значениями (при калибровке на интервалах 1935-1965 или 1935-1975 гг.), а для середины 20-го столетия эти модели занижают величины прироста по сравнению с реальными (при калибровке на интервалах 1965-1998 или 1955-1998 гг.) (Шишов и др., 2007 а). Аналогичные закономерности выявляются и при моделировании годичного прироста древесных растений на базе имитационной VS-модели (Ваганов, Шашкин, 2000; Vaganov et al., 2006).

Возможной причиной возникновения дивергенции может являться температурно-зависимый стресс древесных растений на засуху, особенно ярко выраженный для быстро растущих деревьев (Barber et al., 2000; D'Arrigo et al., 2004). Это заключение было сделано на основе сравнительного анализа дендрохронологических данных с климатическим индексом, являющимся линейной комбинацией температуры и осадков (Barber et al., 2000). Другой гипотезой возникновения проблемы дивергенции является снижение прироста древесных растений при достижении температурой физиологического порога, лимитирующего рост, что в последние десятилетия является высоковероятным событием (D'Arrigo et al., 2004; Hoch, Korner, 2003; Kozlowski et al., 1991; Wilmking et al., 2004). Причем современное потепление в Северном полушарии является беспрецедентным за последние две тысячи лет (IPCC Report, 2007). Используя данные измерений по ширине годичного кольца, максимуму плотности и результатов механистической модели роста Ваганова-Шашкина (Vaganov et al., 2006), была выявлена связь между снижением чувствительности древесных растений и положительным трендом зимних осадков в Субарктической части Сибири в последние десятилетия XX века (Vaganov et al., 1999). Это привело к появлению гипотезы о сдвиге начала сезона роста деревьев к более поздним датам из-за запаздывания таяния снежного покрова на анализируемых территориях и, как следствие, снижением чувствительности деревьев к температурным изменениям (Vaganov et al., 1999). В ряде работ, базирующихся на длительных древесно-кольцевых хронологиях, показано, что

после 1960-х в разных районах Северного полушария отмечается смена лимитирующих рост древесных растений факторов (Jacoby, D'Arrigo, 1995; Lloyd, Fastie, 2002). Еще одной гипотезой расхождения, выдвинутой в последние время, является возрастающий дисбаланс между фотосинтезом и дыханием: рост температуры приводит к более ускоренному росту потерь на дыхание, что и отражается в снижении относительного и абсолютного радиального прироста древесных растений севера Евразии (Шишов и др., 2007 а).

В некоторых случаях, проблема дивергенции может возникать из-за систематической ошибки аппроксимации возрастных трендов «молодых» и «старых» деревьев одной и той же кривой, которая возникает при использовании определенных методов стандартизации (Melvin, 2004).

Но причины могут быть гораздо сложнее и сочетать в себе взаимодействие лимитирующих и ускоряющих рост факторов.

Наличие большого количества непротиворечивых гипотез[1], касающихся одной и той же проблемы, свидетельствует о том, что полная информация, содержащаяся в характеристиках годичного кольца деревьев, не может быть извлечена корректно на базе существующих в дендроклиматологии и дендрохронологии математико-статистических подходов.

Возникает необходимость в развитии математического аппарата, который позволит адекватно оценить влияние внешних факторов на прирост древесных растений.

Более того, проблема «дивергенции» привела к тому, что исследователи начали работу по ревизии существующих дендрохронологических и дендроклиматических данных, методов их обработки (стандартизации) и интерпретации, анализа и получения длительных реконструкций на базе корреляционно-регрессионного анализа (Briffa et al, 1998a; 2004; 2008; D'Arrigo et al., 2007).

В Российской Федерации за последнее десятилетие проведена огромная работа по созданию сети станций дендроклиматического мониторинга (Ваганов

[1] В математической логике известна теорема, которая показывает, что доказать противоречивость гипотезы в рамках теории, в которой она выдвинута, невозможно.

и др., 1996; Ваганов и др., 2000; Ваганов, Шиятов, 2005; Шишов и др., 2002, 2007) (Рис. 1).

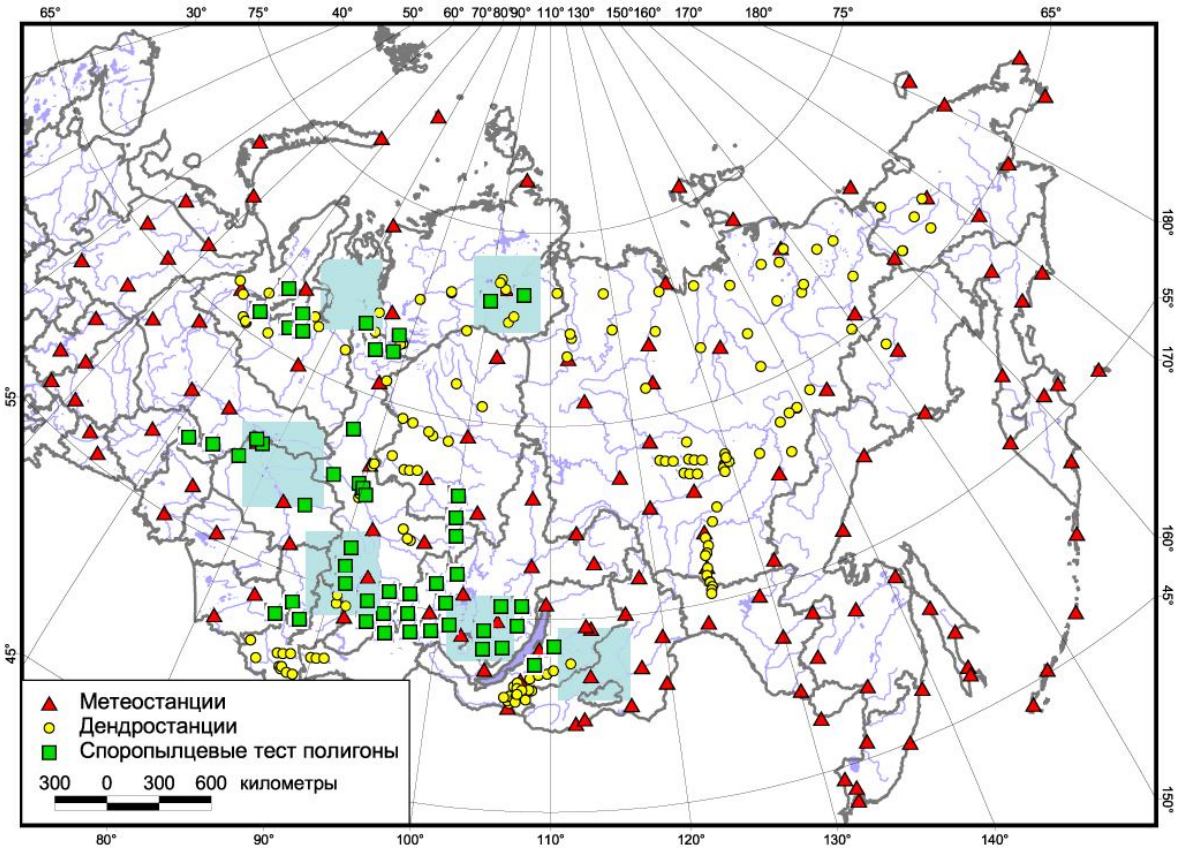


Рис. 1. Пространственное распределение дендрохронологических тест-полигонов (участков), метеорологических станций и споро-пыльцевых тест-полигонов, включенных в систему дендроклиматического мониторинга России

Эта сеть включает более 300 длительных древесно-кольцевых хронологий для основных лесообразующих пород, по которым можно не только получить количественный пространственно-временной отклик в росте древесных растений на климатические изменения, но и построить прогнозные модели реакции древесных растений на прогнозируемые изменения климата.

Все дендрохронологические данные объедены в единый Урало-Сибирский банк данных (Ваганов и др., 1996).

Для каждого дендрохронологического участка имеется краткое описание местоположения (географические координаты, высота над уровнем моря, основная лесообразующая порода и т.п.), а также массив исходных кросс-

датированных измерений по разным характеристикам годичных колец отдельных деревьев. Каждая запись такого массива является временным рядом.

Так как сеть создана в последнее десятилетие и постоянно пополняется, она включает как прошлые, так и текущие изменения прироста древесных растений, чего нет для большинства европейских и северо-американских древесно-кольцевых хронологий, которые были в основном получены в 1960-70-х годах.

Ограниченность по времени древесно-кольцевых хронологий Европы и Северной Америки является важнейшим препятствием к надежному выявлению смены лимитирующих рост факторов в условиях ускоренного увеличения концентрации парниковых газов в атмосфере для этих частей Северного полушария.

Рис. 2. Пространственное распределение метеорологических станций России, имеющих 3-6-часовое разрешение по представленным инструментальным наблюдениям по температуре и осадкам.

Пространственное распределение участков этой сети хорошо согласуется с имеющейся сетью климатических станций, которая включает в себя около 2000 записей инструментальных суточных метеонаблюдений на территории России,

покрывающих последние 30-80 лет XX века (Data documentation…, 2005) (Рис. 2).

Наличие Урало-Сибирского дендрохронологического банка пространственно-распределенных данных, хорошо согласующихся пространственно с имеющимися климатическими показателями на территории России, привело также к необходимости развивать новые системные подходы к анализу дендроклиматических данных.

С этой целью, впервые для РФ спроектирована интерактивная информационная дендроклиматическая система, которая включает в себя обновляемую реляционную базу дендрохронологической и климатической информации на базе SQL технологий, а также функциональное наполнение, объединяющее в себе разнообразное специальное программное обеспечение (ПО) по обработке дендроклиматической информации, написанное на разных операционных платформах (Рис. 3).

В интерактивной системе используется TCL-SQL проект, который предусматривает общий TCL-интерфейс к базам данных. Реализация на начальном этапе состоит в доступе к интерактивной базе данных (технология PostgreSQL[2]), в которой находятся данные измерений. Отметим, что стандартный скрипт-язык TCL[3] (Tool Command Language — «командный язык инструментов») не содержит средств для доступа к PostgreSQL, но этот язык выполнен так, что может расширять свои возможности через подгружаемые новые библиотеки, а следовательно, наличие нужной библиотеки - это необходимое и достаточное условие для того, чтобы использовать в проекте всю мощь SQL (Рис. 4).

---

[2] PostgreSQL — это свободная реляционная объектно-ориентированная система управления базами данных (СУБД), которая является некоммерческой, альтернативой коммерческим СУБД (например, Oracle Database, Microsoft SQL Server, IBM DB2, Informix и СУБД производства Sybase)

[3]TCL, наряду с Perl и Python, стал одним из трёх классических скриптовых языков общего назначения. В TCL можно манипулировать данными всех типов, включая кодами программы, как строками. Это позволяет TCL являться языком с естественной поддержкой метапрограммирования.

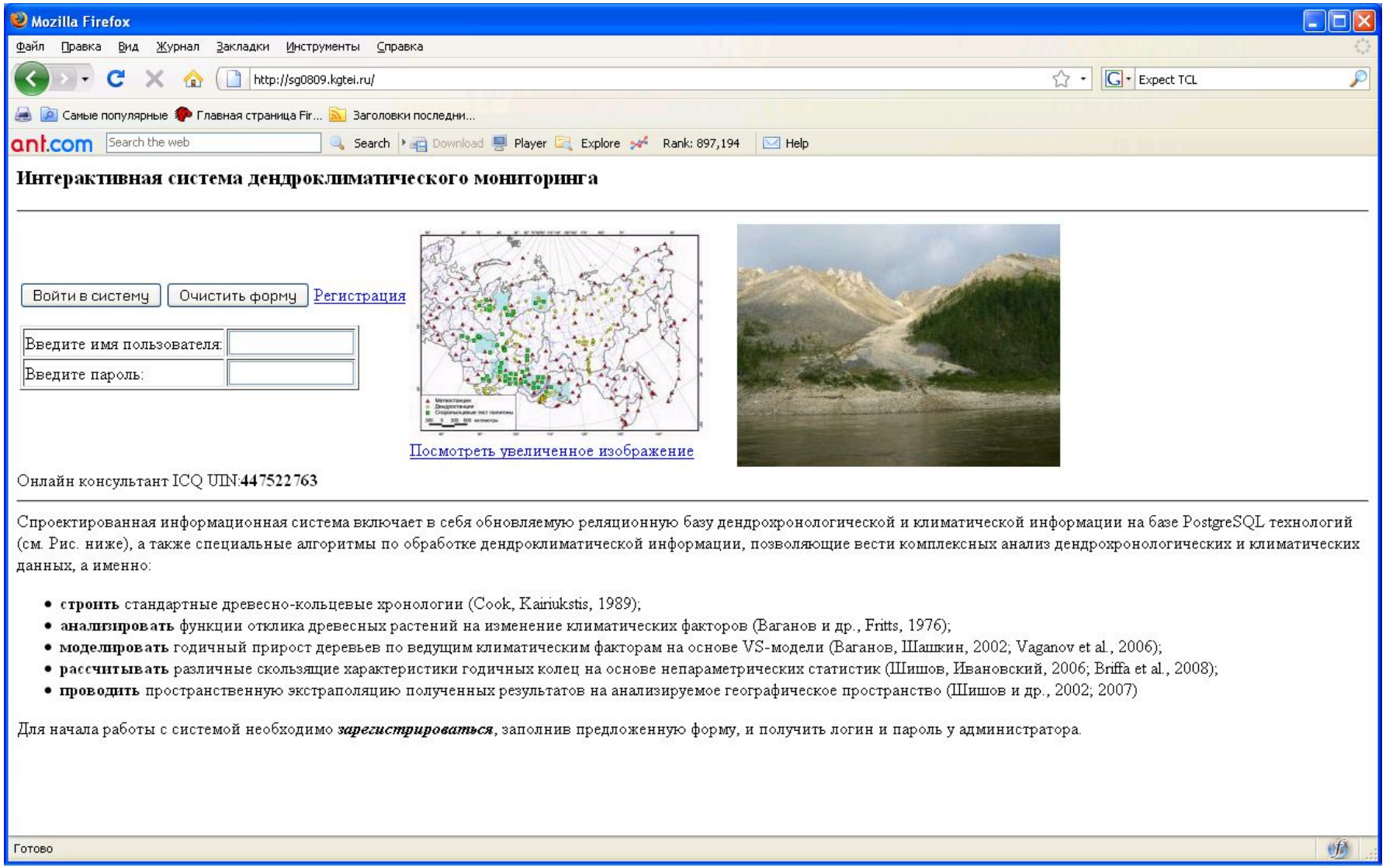


Рис. 3 Главная страница Web-ресурса «Интерактивная система дендроклиматического мониторинга»

Для интегрирования унаследованных приложений (специального прикладного ПО) с разрабатываемой интерактивной системой выбрана технология Expect, как инструмент для автоматизации, тестирования в ОС Unix/Windows (Libes, 1994; 1997; 1998) и являющееся расширением к TCL для неграфических интерактивных унаследованных приложений.

Для данного интерактивного проекта создан домен третьего уровня на базе домена КГТЭИ http://sg0809.kgtei.ru/, через который проброшен туннель по порту 8015 на Windows-машину с сервером приложения. Это позволяет использовать смешанную архитектуру доступа, в результате реализации которой повышается безопасность, а в конечном итоге надежность функционирования системы.

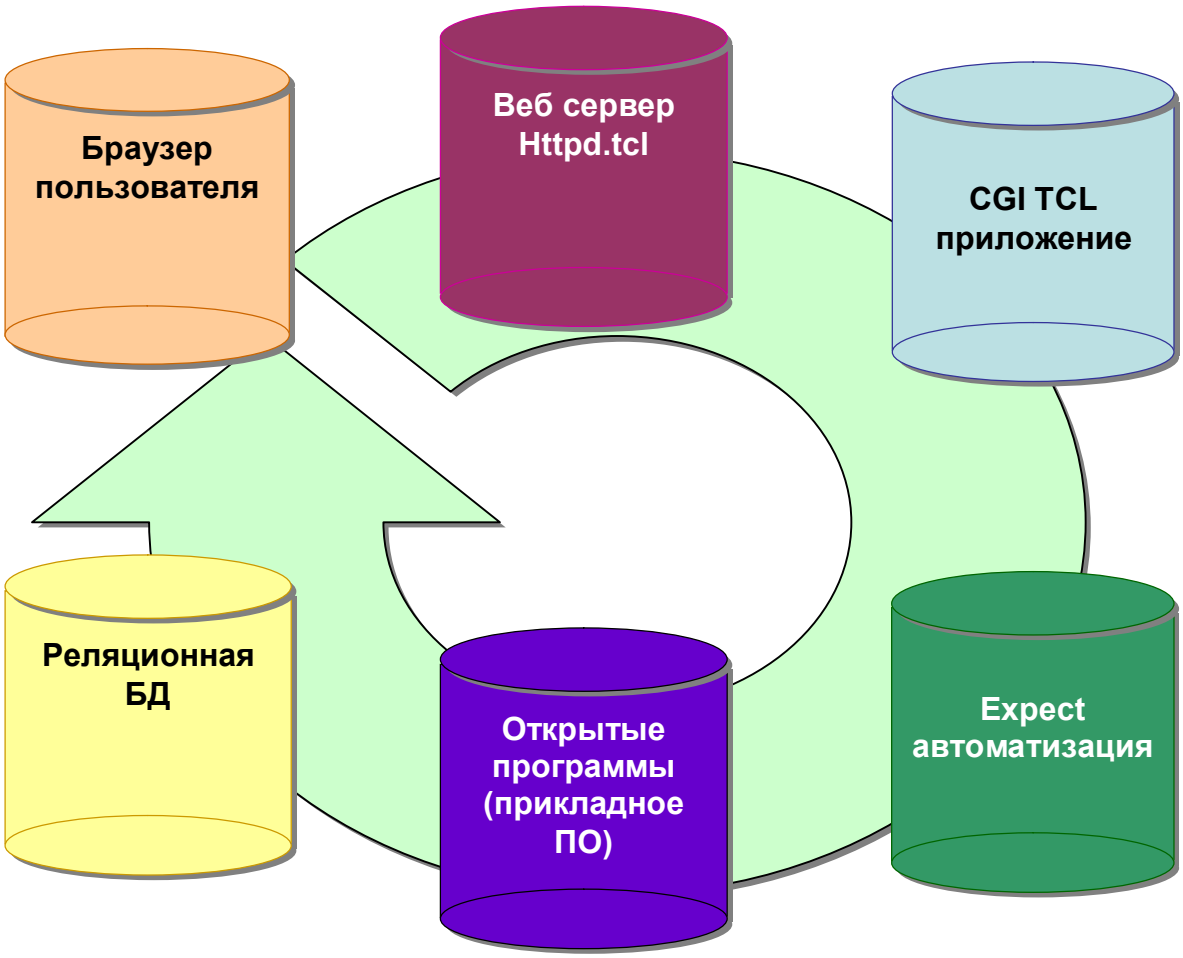


Рис. 4 Общая схема функционирования системы дендроклиматического мониторинга

Спроектированная система использует только сетевое прикладное ПО, которое открыто для общего использования (Open Source). Изначально, идея

Open Source пришла из научной среды. Именно здесь был принят принцип открытого обмена научными идеями и наработками, а, впоследствии, и программным кодом. В данном проекте это позволило использовать многолетние наработки ПО для создания этой интегрированной среды.

В созданной системе осуществляется клиентский доступ к базам данных (Рис. 5), который позволяет выполнять запросы по поиску необходимой дендрохронологической и климатической информации по географическим координатам.

Более того, планируется организовать взаимодействие разработанной системы с уже существующими базами данных, размещенными на других открытых Web-ресурсах (см., например, *http://hurricane.ncdc.noaa.gov/pls/paleo/fm_createpages.treering*) (Рис. 6).

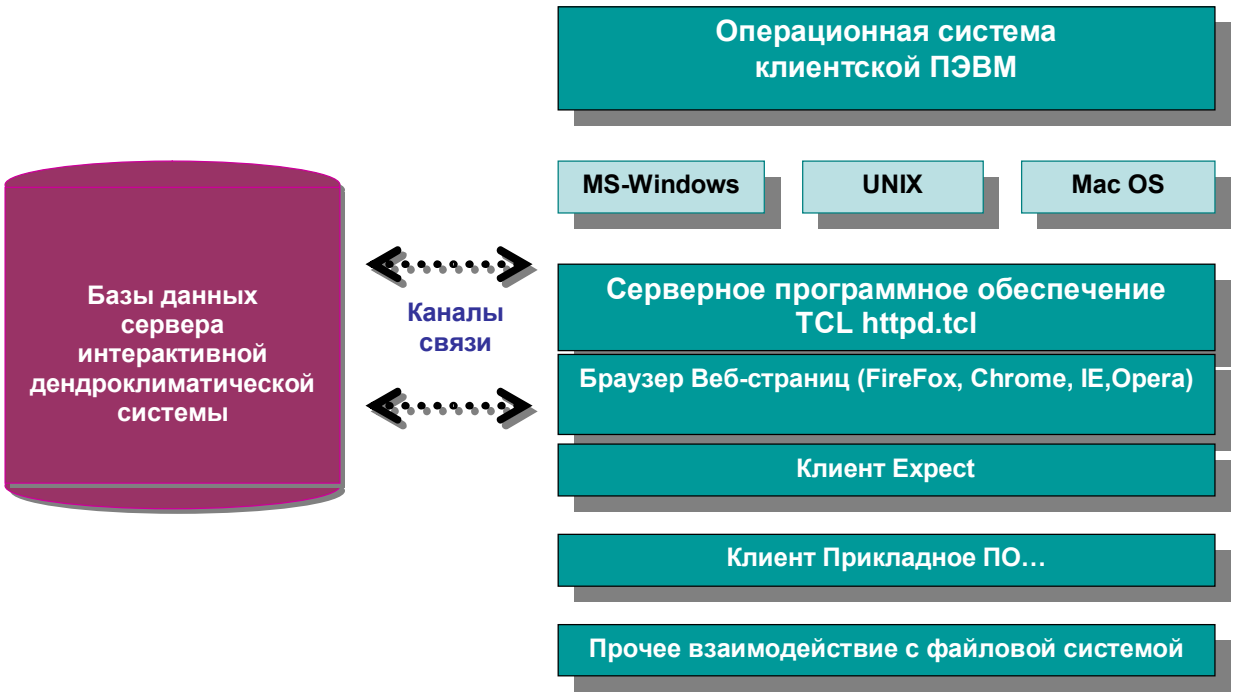


Рис. 5. Схема клиентского доступа к серверу баз данных.

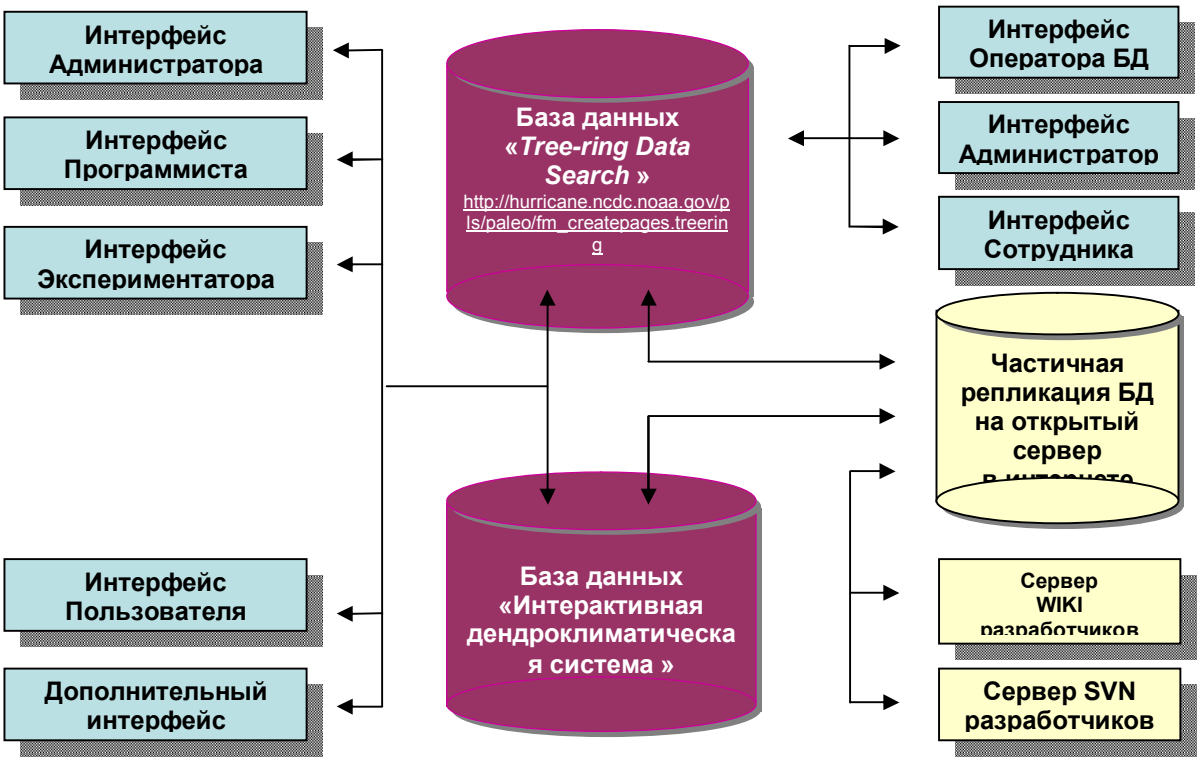


Рис. 6. Общая схема спроектированной системы

В настоящее время, интерактивная система дендроклиматического мониторинга позволяет обрабатывать дендрохронологическую и дендроклиматическую информацию при помощи неграфического прикладного ПО. В самом ближайшем будущем планируется подключение и графических приложений по обработке дендроклиматической информации. Более того, для дальнейшего совершенствования системы планируется частичная репликация опубликованной дендроклиматической информации на один из открытых серверов с целью дальнейшего наполнения функциональных возможностей системы новыми TCL- и другими приложениями, разработанными не только администраторами и системными программистами интерактивной системы, но и добровольными wiki-специалистами при соответствующем продвижении разработанного Web-ресурса (Рис. 6).

Резюмируя все выше сказанное, отметим, что в разработанной интерактивной системе используются спроектированные авторские Web-службы, т.е. программы, доступ к которым осуществляется через протокол HTTP, а обмен данными происходит в формате XML, JSON и REST. В отличие

от обычных динамических библиотек неграфических программ такой подход обладает рядом преимуществ:

- Web-служба находится на серверах, с которыми она взаимодействует. Поэтому в любой момент пользователю доступна самая свежая версия данных и ему не приходится заботиться об обновлениях и вычислительных мощностях, требуемых для выполнения операции;
- инструменты для работы с HTTP и XML есть в любом современном языке программирования, поэтому эта система (веб-служба) переходит в разряд платформонезависимых;
- в качестве сервера приложений выбран TCL Web сервер, который является надежным Web-сервером, а также хорошо стыкуется со скриптами и open-source ПО. Сервер использует "родную" систему ввода-вывода, что позволяет ему без проблем оперировать данными между дисковыми накопителями и сокетами.

Наряду с традиционными методами обработки дендроклиматической информации в интерактивной системе дендроклиматического мониторинга автор предлагает использовать методы и алгоритмы «реляционной математики».

Эти подходы основаны на использовании новых для дендроклиматологии статистических оценок (непараметрических статистик), хорошо разработанных методов теории классификации и спектрального анализа. Непараметрические статистики и их модификация, а также новые методы спектрального анализа позволяют получить новые результаты, касающиеся составляющих изучаемых временных рядов и случайных функций. Методы теории классификации позволяют корректно обобщить информацию об отклике древесных растений на субрегиональном и региональном уровнях с последующей интерполяцией полученных результатов на рассматриваемое географическое пространство.

Эти результаты не только оригинальны, но и ставят вопросы как о качестве анализируемого дендрохронологического и климатического материала, полученных сверхдлительных климатических реконструкций, математико-статистических подходов, используемых в дендроклиматологии, дендрохронологии и дендроэкологии, так и о более глубоком исследовании

механизмов реакции древесных растений на текущие изменения климата и роста концентрации парниковых газов в атмосфере. Это невозможно, в частности, без глубокого понимания методов анализа данных в выделенных областях.

П.2 Проблемы стандартизации дендрохронологических данных.

Древесно-кольцевые хронологии, несмотря на ряд хороших свойств, обладают нестационарностью[4], которая является результатом воздействия на ширину годичного кольца большого количества факторов случайной природы (Ваганов и др., 1996; Ваганов, Шашкин, 2000; Шиятов, 1986; Джансетов, Шишов, 1993; Шишов, 1998; Шишов и др.; 2002; 2007а; Vaganov et al., 2006). В результате в дендрохронологических рядах появляются различного рода возмущения, связанные, в частности, с возрастными особенностями древесных растений и экологией их произрастания (Шиятов, 1986; Ваганов, Шашкин, 2000; Vaganov et al., 2006).

Для избавления от этих возмущений в дендрохронологии используется процедура стандартизации, которая заключается в удаления трендов неклиматической природы, индексации исходных индивидуальных рядов, а затем их усреднения с целью получения древесно-кольцевой хронологии.

Основными методами выделения (аппроксимации) возрастных трендов, которые наиболее часто встречаются в дендрохронологических работах, являются:

I. Детерминистические методы (*a priori* – возрастной тренд G_t известен):

1. $G_t = b_0 + b_1 t$,
2. $G_t = ae^{-bt} + k$ (Fritts, 1976),

---

[4] СФ $X(t)$ называется стационарной (в широком смысле), если:

1. $m_X(t) = m_X = const$,
2. Ковариация СФ $X(t)$ - $K_X(t,t')$ зависит только от расстояния t'-t (Яглом, 1981).

Именно, последние понятие стационарной случайной функции используется в дендрохронологии и дендроклиматологии.

3. $G_t = \dfrac{1}{a + b(t-k)}$, где $k = n/2$, n - длины ряда (для оценки тренда используется гиперболическая функция (Eklund, 1954)),

4. $G_t = at^{-b}$ (Kuusela, Kikki,1963),

5. $G_t = at^b e^{-qt}$ (Waren, 1980; Braker,1981),

6. $G_t = at^{a-1} b^{-a} \exp(-(t/b)^a)$ (Yang et al., 1978),

7. $G_t = b_0 + b_1 t + \ldots + b_p t^p$ (Fritts, 1976; Graybill, 1979);

II. Стохастические методы (*a priori* - тренд неизвестен):

1. $G_t = \sum_{i=-k}^{k} w_i x(t+i) / \sum_{i=-k}^{k} w_i$,

где $x(t+i)$ - значение дендрохронологического ряда в $t+i$ год ($i=-k,\ldots,k$), w_i - весовой коэффициент для $t+i$ года ($i=-k,\ldots,-1,0,1,\ldots,k$). Обычно считается, что w_i имеют нормальный закон распределения (Mitchell et al., 1966: Briffa, 1984). Если все w_i =1, то для убирания избавления от тренда применяется модель скользящего среднего (Ording, 1941; Рудаков, 1951; Битвинскас,1965);

2. Модель скользящего кубического сплайна, который является также симметрическим цифровым фильтром с w_i =1 (Cook, Peters, 1981);

3. $G_t = \alpha \bar{x}(t) + (1-\alpha)\alpha \tilde{x}(t)$,

где $\bar{x}(t) = \alpha(x(t)-1) + (1-\alpha)\bar{x}(t-1)$,

$\tilde{x}(t) = \alpha(x(t) - x(t-1)) + ((1-\alpha)/\alpha)\tilde{x}(t-1)$,

α - вес, который определяет степень сглаживания.

Влияние ширины годичных колец предшествующих лет на ширину кольца в t-год экспоненциально уменьшается как $(1-\alpha)^k$, где k - число колец, предшествующих t-значению. Последний фильтр является несимметрическим (Barefoot et al., 1974).

4. $G_t = \sum_{i=0}^{k} \psi_i x_{t-i}$, где x_{t-i} - значение ширины годичного кольца в t-i год (i=0,...,k). Это авторегрессионая модель k-порядка (АР(k)), которая является частным случаем линейного фильтра (Бокс, Дженкинс,1974).

Эти оценки ростового тренда G_t, фактически, являются аппроксимирующими линиями возрастной кривой дендрохронологического ряда отдельного дерева.

В статистике случайных функций дендрохронологическая процедура индексации называется процедурой центрирования и нормирования, которая описывается следующим соотношением:

$$\overset{\circ}{x}_t = \frac{x_t - \overline{x}}{\sigma_x},$$

где x_i - значение случайной функции (СФ) в t-момент,

$\overline{x}$ - средняя СФ,

σ_x - среднеквадратичное отклонение.

В дендрохронологии показано, что, в большинстве случаев, $\overline{x}$ и σ_x для дендрохронологических рядов коррелированны (Cook, Kairiukstis, 1990), поэтому процедура нормировки («индексации») сводится к следующему уравнению:

$$I_t = \frac{x_t}{G_t},$$

где G_t - выступает в роли оценки «средней» исходной СФ.

Именно индексацией и заканчивается процедура стандартизации.

Существует также и другая процедура стандартизации дендрохронологических рядов - метод коридора, предложенный С.Г. Шиятовым (1986), которая описывается следующим соотношением:

$$I_t = \frac{x_t - x_t^{\min}}{x_t^{\max} - x_t^{\min}} \cdot 100,$$

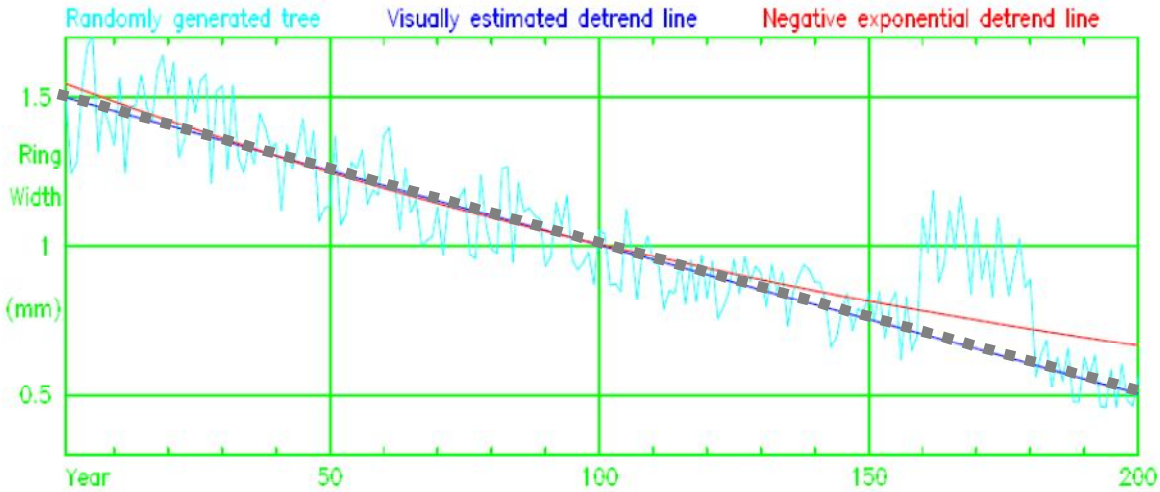
где x_t - значение дендрохронологического ряда в t-момент,

$x_t^{\min}$ - значение так называемой минимальной кривой роста в t-момент,

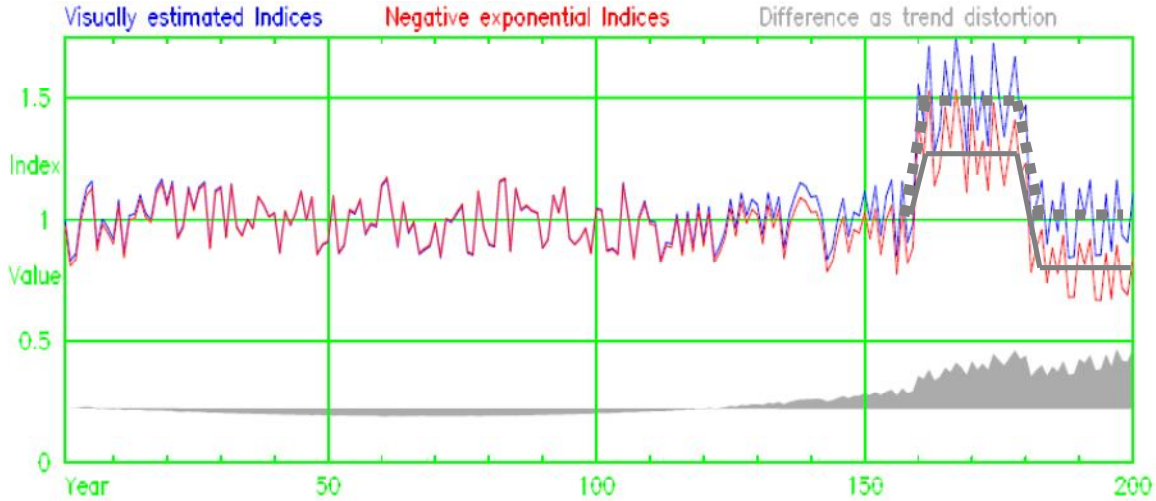
$x_t^{\max}$ - значение максимальной кривой роста в t-момент.

Как и $x^{\min}$, так и $x^{\max}$ находится по специально разработанному алгоритму (Шиятов, 1986).

Большинство из приведенных методов оценки трендов неклиматической природы обладают одной особенностью, которая заключена в искажении (в завышении или занижении) истинных индексных значений дендрохронологических рядов при определенных условиях (Melvin, 2004). Это искажение ярко проявляется на концах исследуемых временных рядов при определенных возмущениях (Melvin, 2004).



А)



Б)

Рис. 7. Случайно сгенерированный возрастной тренд и его оценка при помощи негативной экспоненты (сплошная, тонкая линия) и визуальной оценки (пунктирная линия) А); результат индексации при помощи двух линий тренда Б) (Melvin, 2004).

Так, искусственный всплеск для случайно сгенерированного дендрохронологического ряда за последние 50 лет роста древесного растения (Рис. 7 А), вызванный, например, внешним климатическим воздействием, приводит к занижению индексов прироста для аналогичного периода (Рис. 7 Б) в случае стандартизации негативной экспонентой (Рис. 7 А).

Подобного рода искажения свойственны также для степенных, полиномиальных аппроксимаций степени p ($p \leq 5$) (Fritts, 1976; Graybill, 1979), а также для популярного низкочастотного кубического сплайна (Melvin, 2004).

Такие несоответствия могут существенно изменять как статистическую функцию отклика древесных растений на ведущие климатические факторы (Melvin, 2004), так и, вероятно, параметры настройки VS-модели, которые отвечают за реакцию древесных растений на климатические изменения (Ваганов, Шашкин, 2000; Vaganov et el., 2006).

Как следствие, эти искусственные завышения (занижения) индексных значений могут существенно влиять на качество температурных реконструкций, так как последние 50-70 лет роста древесных растений (период инструментальных климатических измерений) используются для построения (калибровки и верификации) соответствующих регрессионных моделей.

Еще один метод стандартизации дендрохронологических рядов основан на индексации последних к региональной возрастной кривой – RSC-метод (Regional curve standardization) и его модификации CARSC и MARSC (Melvin, 2004). Этот метод является одним из наиболее устойчивым к систематическим концевым отклонениям, которые могут приводить к проблеме дивергенции (Briffa et al., 2001; 1998в; Melvin, 2004). Но существует одно сильное ограничение на его применение: этот метод очень чувствителен к количеству используемых дендрохронологических рядов для получения древесно-кольцевой хронологии. Авторы метода на основе модельного эксперимента рекомендуют для получения региональной возрастной кривой использовать не менее 300 дендрохронологических рядов (Briffa et al., 1992; 1996; 2001; Melvin, 2004).

В этом случае, разница в возрастных кривых между быстрорастущими и медленнорастущими деревьями нивелируется (Рис. 8.).

Отметим, что данный метод стандартизации максимально сохраняет низкочастотные колебания с минимальными вышеупомянутыми завышениями (занижениями), которые, как правило, ассоциируются с внешними воздействиями на прирост древесных растений.

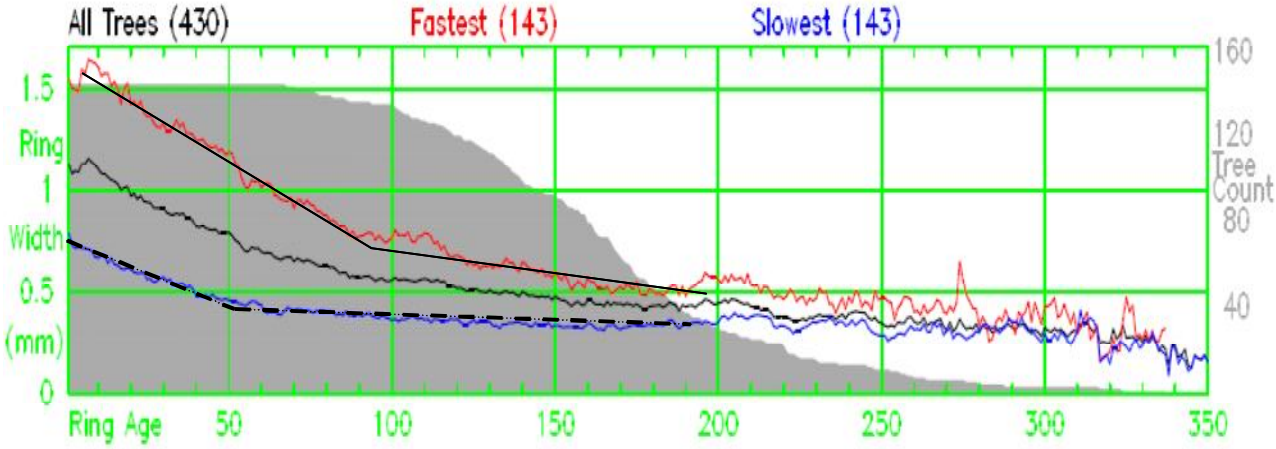


Рис. 8. Возрастные кривые, полученные для быстрорастущих (сплошная линия) и медленнорастущих деревьев (пунктирная линия) (Melvin, 2004).

Перед исследователем всегда возникает вопрос: «Какой же процедурой стандартизации (выделения нужного сигнала из «шума») воспользоваться?».

В дендрохронологии и дендроклиматологии существует несколько критериев для принятия решения в пользу того или иного метода:

1. Максимизация отношения «сигнал к шуму» (Waren, 1980), то есть

$$\max\ SNR = \max \frac{N\bar{r}}{1-\bar{r}},$$

где N - число индексных рядов,

$\bar{r}$ - средняя корреляция между индексными рядами.

Этот критерий пригоден для изучения высокочастотных колебаний, и не всегда применим для низкочастотных, последние из которых могут быть интересны для дендроклиматических исследований (Cook, Kairiukstis, 1990, Briffa et al., 1992; 1996).

2. Критерий «тренд в среднем», который заключается в удалении низкочастотных колебаний дендрохронологических рядов длины n с

частотой меньше, чем 1/*n*, если, *a priori*, неизвестно об аналогичных колебаниях изучаемого природного процесса (Cook, Kairiukstis, 1990).

3. *n*%-критерий, основанный на применении кривой аппроксимаций, длина которой равна определенному проценту от длины дендрохронологического ряда. Например, при использовании скользящего кубического сплайна рекомендуется длину последнего ограничивать 67 % или 75 % от длины ряда (Cook, 1985), что позволяет сохранить низкочастотные составляющие дендрохронологического ряда.

В любом случае, все описанные критерии не всегда дают однозначный ответ о правильности выбора того или иного метода стандартизации, поэтому не последнюю роль в этом процессе играет опыт исследователя.

Возможным подходом к формализации процедуры стандартизации в дендрохронологических данных является изучение структуры анализируемых рядов с целью выделения циклических компонент, трендовых и шумовых составляющих. Выявленные при спектральном анализе тренды могут быть иллюминированы из индивидуальных рядов методами индексации, описанными выше. Более того, появляется возможность выделения региональной специфики роста древесных растений, в частности, выявления региональных возрастных кривых, описанных аналитическими формально-выделенными функциями.

Так как индивидуальные дендрохронологические ряды являются нестационарными, применение традиционного спектрального анализа, широко используемого для исследования обобщенных древесно-кольцевых хронологий (Брукс, Карузерс, 1963, Гренджер, Хатанака, 1972; Андерсон, 1976; Яглом, 1981; Мазепа, 1978), к этим рядам является невозможным. Такие данные могут быть проанализированы при помощи метода множественных сфероидальных последовательностей (Multi-taper method) (Шишов и др., 2007б; Ramachandra, Hamed, 1993; Lees, Park, 1995; Mann, Lees, 1996; Kasatkina et al., 2007; Zanzi et al., 2007), вейвлет анализа (Wavelet analysis) (Ignaccolo et al., 2004; Yiou et al., 2000) и сингулярного спектрального анализа (Singular spectrum analysis) (Liu et al., 2004б; Briffa et al., 2008). В отличие от классического спектрального анализа эти методы не требуют стационарности исходных данных, но, по-прежнему,

остаются чувствительными к высокому уровню «красного» шума, который может полностью изменять спектр изучаемого временного ряда (Mann, Lees, 1996). Наиболее перспективными методами для решения задачи стандартизации являются сингулярный спектральный анализ и его модификация, в которых не делается никаких предположений о природе колебаний (Шишов и др., 2007б), а также новый метод спектрального анализа - Метод эмпирических категориальных коррелограмм (МЭКК) (Шишов, 2009). МЭКК позволяет выявлять все базовые частоты исходного ВР, даже если уровень коллорированного шума является практически максимальным для ВР, используемым, например, в климатологии и дендроклиматологии (Шишов, 2007; Mann, Lees, 1996).

После процедуры стандартизации, как правило, следует либо процедура обычного осреднения индивидуальных индексных рядов для деревьев с определенного дендрохронологического участка, которая носит название процедуры получения обобщенного ряда или древесно-кольцевой хронологии (Шиятов, 1986), либо осреднение индексных рядов отдельных деревьев из разных местообитаний на основании существующих регионализаций (классификаций территорий) (Панюшкина и др., 1996; 1997; Briffa et al., 1992; 1996). Результат осреднения называется, в этом случае, генерализованной хронологией.

И в генерализованных, и обобщенных рядах в общем случае, присутствует неоднородность, связанная с осреднением разного количества индексных рядов на разных временных интервалах. Избавиться от этой неоднородности достаточно тяжело (Schulman, 1956). Эта проблема решается при помощи эмпирико-статистического метода (Шиятов,1986), суть которого сводится к приведению генерализованной или обобщенной хронологии к одному уровню изменчивости.

Из всего вышесказанного становится очевидно, что однозначного ответа на вопрос «Какой метод стандартизации дендрохронологических рядов является наиболее предпочтительным?» не существует. В каждом конкретном случае

необходимо ориентироваться на цели и задачи исследования, а также на методы их решения.

П.3 Множественный регрессионный анализ дендроклиматических данных.

Целью множественного регрессионного анализа является выявление взаимосвязей (в самом простом случае, линейных) между случайной зависимой переменной (результативной переменной или откликом) (например, древесно-кольцевой хронологией) и независимыми переменными или факторами (например, климатическими временными рядами), а затем построение на основании выявленных частных корреляций соответствующих статистических моделей (Бокс, Дженкинс, 1974; Кендалл, Стюарт, 1976; Хеннан, 1974; Халафян, 2007).

Обычно временной ряд (ВР) представляется в виде последовательности типа:

$$x_{t-n}, x_{t-n+1}, \ldots, x_{t-1}, x_t,$$

где x_{t-k} - значение ВР в t-k момент времени ($k = 0,...,n$).

Функциональные связи между элементами ВР определенного типа называются моделями.

Обычно в анализе ВР (множественном регрессионном анализе) для описания той или иной модели применяется операторная форма записи.

Запишем основные операторы:

1) $Bx_t = x_{t-1}$ - оператор «сдвига назад»,
2) $Fx_t = B^{-1}x_t = x_{t+1}$ - оператор «сдвига вперед»,
3) $\nabla x_t = x_t - x_{t-1} = (1-B)x_t$,
4) $Sx_t = \sum_{j=0}^{\infty} x_{t-j} = (1 + B + B^2 + \ldots)x_t = \frac{1}{1-B}x_t = (1-B)^{-1}x_t$.

В анализе временных рядов (Бокс, Дженкинс, 1974) принята следующая классификация моделей для стационарных ВР:

1. Модель скользящего среднего (СС(q)-модель) (линейного фильтра) порядка q.

   Пусть $\ldots, x_{t-1}, x_t$ - изучаемый ВР и $m_x = \mu$,

   $\ldots, a_{t-1}, a_t$ - некоторые фактор или белый шум (ВР), причем m_a=0, $\sigma_a^2 = const$, тогда

$$x_t = \mu + \beta_0 a_t + \beta_1 a_{t-1} + \ldots + \beta_q a_{t-q} = \mu + (\beta_0 + \beta_1 B + \ldots + \beta_q B^q) a_t = \mu + \beta(B,q) a_t,$$

   Окончательно,

$$\overset{\circ}{x}_t = x_t - \mu = \beta(B,q) a_t,$$

   где q - количество последовательных членов ВР a, включенных в модель, называется порядком модели СС.

   В СС-модель могут быть включены несколько, так называемых, независимых ВР (например, k-штук), то есть

$$\ldots, a_{t-1}^1, a_t^1; \ldots, a_{t-1}^2, a_t^2; \ldots; \ldots, a_{t-1}^k, a_t^k,$$

   тогда порядком данной модели будет считаться следующая сумма:

$$q = \sum_{i=1}^{k} q_i,$$

   где q_i - количество последовательных членов ВР a^i (i=1,…,k).

   Если все q_i=1 (i=1,…,k), тогда модель скользящего среднего представляет собой общую модель множественного регрессионного анализа.

2. Модель авторегрессии (АР(p)-модель) порядка p.

   Пусть ВР x и a определены как в п.1, тогда

$$\overset{\circ}{x}_t = x_t - \mu = \phi_1 \overset{\circ}{x}_{t-1} + \ldots + \phi_2 \overset{\circ}{x}_{t-p} + a_t, \text{ следовательно}$$

$$\overset{\circ}{x}_t - \phi_1 \overset{\circ}{x}_{t-1} - \ldots - \phi_2 \overset{\circ}{x}_{t-p} = a_t \text{ или } (1 - \phi_1 B - \ldots \phi_p B^p) \overset{\circ}{x}_t = a_t.$$

   Окончательно,

$$\Phi(B,p) \overset{\circ}{x}_t = a_t \text{ или } \overset{\circ}{x}_t = \Phi(B,p)^{-1} a_t.$$

   Утверждение. Модель авторегрессии является частным случаем модели скользящего среднего.

   Доказательство является практически очевидным.

3. Смешанная модель - авторегрессионная модель скользящего среднего (АРСС(p,q)-модель) порядка (p,q).

$$\overset{\circ}{x}_t = \phi_1 \overset{\circ}{x}_{t-1} + \ldots + \phi_p \overset{\circ}{x}_{t-p} + a_t + \beta_1 a_{t-1} + \ldots + \beta_q a_{t-q}$$

или

$$\Phi(B,p)\overset{\circ}{x}_t = \beta(B,q)a_t$$

Следовательно, $\overset{\circ}{x}_t = \Phi(B,p)^{-1}\beta(B,q)a_t$.

Заметим, что и в этом случае, как и в п.1, в АРСС-модель могут быть включены несколько (например, k-штук) независимых ВР a^i.

4. Авторегрессионная модель проинтегрированного скользящего среднего (АРПСС(p, d, q)-модель) порядка (p, d, q).

$$\Phi(B,p)(1-B)^d x_t = \beta(B,q)a_t \text{ или } (1-B)^d x_t = \Phi(B,p)^{-1}\beta(B,q)a_t .$$

Заметим, что в левой части последнего выражения стоят производные, следовательно, необходимо интегрировать правую часть.

Конечно, после добавления в каждую из описанных выше моделей нескольких независимых ВР (факторов), построение этих моделей значительно усложняется.

Как уже отмечалось выше, целью анализа временных рядов является выявление связей между независимыми факторами и зависимым изучаемым ВР с последующим построением соответствующей линейной модели. Например, статистическое моделирование динамики древесно-кольцевой хронологии при помощи нескольких независимых (некоррелированных) друг от друга климатических факторов (Fritts, 1976). В одномерном случае соответствующая модель имеет вид:

$$x_t = \mu + \beta_0 a_t ,$$

где x - древесно-кольцевая хронология,

a - например, ряд среднемесячных июньских температур.

Решение последнего уравнения представляет собой лучшее линейное приближение в плоскости июньской температуры и дендрохронологического ряда, найденное, например, при помощи метода наименьших квадратов.

Аналогично, в (q+1)-мерной задачи, решение - это лучшее линейное приближение в (q+1)-мерном пространстве отклика q независимыми факторами.

Заметим, что коэффициенты регрессии объясняют независимый вклад каждого фактора в описание динамики зависимого временного ряда и тесно связаны с частными коэффициентами корреляции между откликом и независимыми факторами.

При этом, очень часто возникают следующие вопросы, связанные с проблемой снижения размерности исходной задачи (Айвазян и др., 1989; Айвазян, Мхитарян, 1998; Cook, Kairiukstis, 1990):

1. Какие факторы вносят наибольший вклад в объяснение вариации исходного изучаемого ВР (В данном случае - проблема информативности факторов)?
2. Действительно ли наиболее информативные факторы являются независимыми друг от друга?
3. Является ли незначимая частная корреляция хорошим показателем независимости какого-либо фактора от изучаемого ВР?

Несомненно, эти вопросы являются достаточно сложными, и не существует единого алгоритмического подхода к их разрешению (к каждой конкретной задаче анализа ВР свой подход), при этом могут возникать парадоксальные ситуации:

1. Из попарной независимости некоторого числа случайных величин не следует, вообще говоря, независимость этих случайных величин в совокупности (то есть добавление некоторого независимого фактора в анализ может однозначным образом определять значения изучаемого временного ряда) (*Пример Бернштейна*).
2. Некоррелированность некоторых случайных величин не означает их независимость, а независимость случайных величин означает их некоррелированность, то есть незначимая корреляция (или отсутствие последней) между изучаемым временным рядом и фактором, вообще говоря, не означает их независимость, например, фактор может нести вполне определенную информацию об изучаемом природном явлении.

Эти теоретические утверждения легко доказываются и подтверждают сложность построения линейных множественных статистических моделей, адекватно описывающих анализируемый процесс. Кроме этого, эти модели слабо учитывают тот факт, что почти в каждой распределенной или точечной биологической системе существуют всевозможного рода запаздывания и опережения (Джансеитов и др., 1976; 1978; Джансеитов, Кузьмичев, 1981).

Такие статистические модели широко применяются во многих отраслях науки, в частности, в дендрохронологии и дендроклиматологии, так как аппарат для построения таких моделей хорошо развит (Боровиков, Боровиков, 1997; Дубров и др., 2000; Справочник по прикладной статистики, 1989; Халафян, 2007; Fritts, 1976: Draper, Smith, 1981; Wonnacott , Wonnacott, 1981).

Рассмотрим модели анализа временных рядов, которые наиболее часто применяются в дендрохронологии и дендроклиматологии.

Пусть дан некоторый изучаемый временной ряд x длины n (например, древесно-кольцевая хронология) и m независимых факторов $a^1, a^2,..., a^m$ длины n (например, среднемесячные климатические характеристики). При этом считается, что все члены представленных временных рядов центрированы и нормированы.

Требуется построить адекватную линейную статистическую модель (*a priori* считается, что связь между факторами и изучаемым откликом линейная) вида:

$$x_t = \sum_{i=1}^{m} a_t^i \beta_i + \varepsilon_t ,$$

где ε_t - необъясненная ошибка, имеющий нормальный закон распределения.

В матричной форме последнее уравнение переписывается в следующем виде:

$$X_n = A_n^m B_m + E_n$$

(нижний номер-количество строк, верхний - количество столбцов),

где $B_m = (A_n^{m^T} A_n^m)^{-1} A_n^{m^T} X_n$ (Cook, Kairiukstis, 1990).

При такой оценке регрессионных коэффициентов β_i (i=1,...,m) возникает сразу несколько проблем:

1. Факторы a^1, a^2,..., a^m могут быть коррелированны, что приведёт к проблеме обращения матрицы $A^T A$;
2. Длина общего периода наблюдений n может оказаться недостаточной для надежной оценки регрессионных коэффициентов;
3. Реальные моделируемые взаимоотношения могут носить нелинейный характер (Ваганов, Шашкин, 2000; Vaganov et al., 2006).

Некоторые из этих трудностей можно избежать, например, если использовать в анализе небольшое число достаточно длительных некоррелированных факторов (Hughes et al., 1987).

Для определения значимости модельных коэффициентов можно воспользоваться следующим методом (Cook, Kairiukstis, 1990).

Рассмотрим ковариационную матрицу регрессионных коэффициентов $V_m^m(B)$, которая определяется следующим образом:

$$V_m^m(B) = \sigma_e^2 (A^T A)^{-1},$$

где σ_e^2 - остаточная вариация (дисперсия), то есть часть от дисперсии σ_x^2, которая не объясняется статистической моделью.

По определению, объяснённая дисперсия EV есть:

$$EV = 1 - \frac{\sigma_e^2}{\sigma_x^2}.$$

Чтобы выяснить статистическую значимость коэффициента β_i (i=1,...,m), необходимо сравнить β_i с $\sigma_i^i(\beta_i) \in V_m^m(B_k)$.

Заметим, что в многомерных задачах очень часто стоит вопрос о снижении размерности поставленной задачи, то есть определение наиболее «информативных» факторов.

В дендрохронологии и дендроклиматологии для этой цели применяются различные критерии, в частности, критерий Фишера (Furnival, Wilson, 1974; Draper, Smith, 1981), основанный, например, на максимизации следующего отношения:

$$F = \frac{(\sigma_x^2 - \sigma_e^2)(n - j - 1)}{\sigma_e^2 j},$$

где j - число новых факторов, введённых в исходное уравнение ($j<m$).

В дендроклиматологии при анализе реакции древесных растений на климатические изменения наиболее часто используют расчет множественных регрессионных моделей через, так называемые, ортогональные функции отклика (Fritts, 1976; Cook, Kairiukstis, 1990; Fritts et al., 1991). Особенностью данного методического подхода является то, что перед нахождением адекватной множественной регрессионной модели климатические факторы преобразуются в ортогональные переменные с помощью метода главных компонент.

Пусть $L_m^m(a)$ - диагональная матрица собственных значений и матрица $P_m^m(a)$ - матрица собственных векторов корреляционной матрицы $\frac{(A^T A)}{n}$.

Тогда взаимоотношение между исходными факторами и их главными компонентами задаются следующим соотношением:

$$\frac{(A^T A)}{n} P(a) = P(a) L(a).$$

Так как $P_m^m(a)$ - ортогональная матрица, то верно следующее соотношение:

$$(A^T A)^{-1} = \frac{P(a) L(a)^{-1} P(a)^T}{n}.$$

Заметим, что собственные значения в матрице $L_m^m(a)$ упорядочены в порядке убывания сверху вниз по главной диагонали.

Известно, что собственные значения, близкие к нулю, ведут к большим стандартным ошибкам оценок регрессионных коэффициентов (Cook, Kairiukstis, 1990). Решение этой проблемы состоит, например, в уменьшении числа собственных значений, близких к нулю (то есть их исключение из анализа). Таким образом, задача состоит в нахождении некоторого натурального числа q ($q<m$), которое определяет размерность подматрицы $\tilde{L}_q^q$ матрицы L_m^m.

Эвристическими критериями к нахождению q могут служить следующие правила:

1. Известно, что сумма некоторого числа больших собственных значений $\sum_{i=1}^{k} \lambda_i / k$ объясняет определенную долю общей дисперсии. Считается, что

0.9 - достаточный уровень для построения надежной статистической модели, поэтому q выбирается из соображений, чтобы выполнялось: $\sum_{i=1}^{q} \lambda_i / q \geq 0.9$. В дендрохронологии, однако, существуют примеры (Fritts, Wu, 1986), когда этот уровень составляет 70-75%. Эта «заниженность» часто объясняется спецификой моделируемых объектов.

2. q может быть числом главных компонент с собственными значениями, большими 1.
3. q может быть выбрано на основании так называемого графического подхода (Tatsuoka, 1974). В этом случаи строится зависимость суммы накопленных собственных значений $\sum_{i=1}^{k} \lambda_i$ от их порядка k (k=1,...,m). q в данном случае является порядком того собственного значения, с которого данная кривая выходит, фактически, на постоянный уровень.

Существуют и другие критерии выбора q (Guiot, 1984).

После того, как исследователь определился с q, решение исходного статистического уравнения представляется в виде:

$$B_m = P_m^q(a)\widetilde{L}_q^q(a)^{-1} P_q^m(a)^T A_m^{n\,T} X_n / n .$$

Вариабельность регрессионных коэффициентов (см. выше) оценивается из следующего уравнения:

$$V(B)_m^m = \sigma_s^{\,2} P_m^m(a) L_m^m(a)^{-1} P_m^m(a) / n .$$

В этом случае, собственные значения матрицы $L_m^m(a)$, близкие к нулю, могут приводить к большим вариациям регрессионных коэффициентов.

Ортогональное преобразование климатических факторов устраняет возможную корреляцию между ними и, в дальнейшем, позволяет при расчетах модели регрессии работать с независимыми переменными. Более того, значения коэффициентов модели (функции отклика или response function) дают точную количественную оценку относительного влияния того или иного климатического фактора на изменчивость прироста с учетом многолетней дисперсии и среднего значения. Такой подход позволяет решить проблему мультиколлинеарности исходных климатических переменных.

Для моделирования прироста древесных растений также используется пошаговая регрессия с включением (Шишов и др., 2007а).

Суть метода пошаговой регрессии сводиться к поиску набора наиболее информативных независимых переменных (ежемесячных температур и осадков) из всей совокупности данных (массива климатических переменных) для моделирования зависимого показателя (индекса прироста) на основе F-критерия Фишера. Для этого, выбирается общее критическое значение для F-критерия. Далее, из всего массива независимых переменных находится та, которая при включении в регрессионную модель дает максимальное эмпирическое значение для F-статистики. В случае, если эмпирическое значение превосходит критическое, то переменная включается в модель. На следующем шаге процедура повторяется для новой независимой переменной. Выбор наиболее информативных переменных заканчивается в том случае, когда в исходном массиве данных ни одна из оставшихся независимых переменных не удовлетворяет F-критерию. Такой подход позволяет снизить размерность решаемой задачи, что положительно влияет на качество регрессионной модели, определить наиболее информативные независимые переменные из всего набора данных, а также позволяет частично решить проблему мультиколлениарности исходных независимых переменных.

Отметим, что два описанных регрессионных подхода при комплексном пространственно-временном анализе реакции прироста на климатические изменения дают сравнимые результаты (Шишов и др., 2007а).

Иногда в дендрохронологии и дендроклиматологии применяется канонический корреляционный анализ (Glahn, 1968; Fritts, 1976; Cook, 1985), который отличается от множественного регрессионного анализа тем, что вместо одного результативного признака в левой части модели применяется линейная комбинация результативных признаков (Халафян, 2007).

После построения статистической модели на некотором временном интервале *n* (этот этап в дендроклиматологии называется процедурой калибровки на временном интервале *n*) следует процедура верификации (проверки) качества функционирования модели на другом (независимом)

интервале времени n'. Эта процедура оценивает адекватность работы постулируемой модели на n' при помощи нескольких тестов, например: знакового (Beyer, 1968; Blasing, 1978), PM-теста (Dunn, Clark, 1974), RE-теста (Gordon, 1980), последний из которых основан на статистике, похожей на оценку объяснённой дисперсии (Lorenz, 1977).

Итак, выше описаны основные методы построения статистических моделей, используемые в дендрохронологии и дендроклиматологии.

Классическими книгами по извлечению климатической информации из дендрохронологических рядов с её последующим использованием для реконструкции и прогноза при помощи многомерных линейных статистических моделей считаются работы Х.Фритца (Fritts, 1976; 1991).

Такого рода модели принято, как правило, разделять на функции отклика (response functions), используемые для выявления линейных связей между приростом древесных организмов и внешних, например, климатических факторов, и функции передачи (transfer functions), позволяющие проанализировать связи между определенными внешними факторами и различными дендрохронологическими рядами (Cook, Kairiukstis, 1990). Построение функций передачи считается в дендроклиматологии обратной задачей (Ваганов и др., 1996). Последние модели часто применяются для реконструкции (прогноза в прошлое) некоторых климатических характеристик. Например, значимые корреляции между обобщенными хронологиями и среднемесячными летними температурами, полученные на основании анализа функций отклика по большинству дендрохронологических участков и сопряженных к ним метеостанций в субарктических районах Урала и Сибири, позволили провести реконструкции динамики летних температур, в частности, июнь-июльских температур (Ваганов и др., 1996). При этом, постулируются следующие модели:

1. Если для некоторого региона имелись один обобщенный ряд I и соответствующий ряд температурных наблюдений T, то модель принимает следующий вид:

$$T_t = b_0 + b_1 I_t + \varepsilon_t ,$$

где ε_t - необъяснённая моделью ошибка;

2. Если, для данного региона имелись N обобщенных рядов I^i (i=1,...,N) и соответствующий температурный ряд T, то применяется СС-модель:

$$T_t = b_0 + \sum_{i=1}^{N} b_i I_t^i + \varepsilon_t .$$

Имеется много реальных примеров анализа связей между дендрохронологическими и климатическими рядами с последующей реконструкцией интересующих климатических факторов.

Дендроклиматические реконструкции дают возможность оценить величину и темпы текущих климатических изменений и сравнить их с масштабами подобных процессов в прошлом (Briffa, 2000; Briffa et al., 2001). Поскольку в изменчивости прироста и структуры годичных колец деревьев, произрастающих в экстремальных условиях, наилучшим образом отражается изменчивость основного внешнего фактора, лимитирующего рост, то наиболее пригодным материалом для исследования изменчивости температуры в прошлом являются годичные кольца деревьев северных и высокогорных районов (Fritts, 1976; Шиятов, 1986; Ваганов и др., 1996). В последние годы опубликовано множество реконструкций летней температуры воздуха для отдельных районов Земного шара и всего Северного полушария вцелом (Kelly et al., 1982; Jones, Briffa, 1992; Hughes, 1995, 1999; Mann et al., 1998, Briffa et al., 2001). Большинство дендроклиматических реконструкций основаны на таких интегральных характеристиках как ширина и максимальная плотность поздней зоны годичных колец древесины. Соответственно, была проведена реконструкция суммарной температуры за достаточно продолжительные промежутки года (за май-август (Jacoby, D'Arrigo, 1995, Кирдянов, 1999), май-сентябрь (Briffa et al., 1998a; 1998в)), среднегодовой температуры (Jacoby, D'Arrigo 1989; D'Arrigo, Jacoby, 1995; Naurzbaev, Vaganov, 2000; Naurzbaev et al., 2002). Однако в годы с одинаковой суммарной летней температурой ее динамика во время вегетационного периода может существенно различаться. В связи с этим актуальным становится вопрос о реконструкции климата с более высоким, чем сезон, временным разрешением, что позволило бы выявить и проанализировать

климатические изменения, характерные для более коротких периодов года (декада, месяц), оказывающих наибольшее влияние на растительность того или иного географического региона (Ваганов и др. 1999; Шишов, Кирдянов, 2003). Проведение подобного анализа необходимо также для выяснения причин увеличения продолжительности вегетационного периода в северных широтах, видимого по данным спутниковых наблюдений (Keeling et al., 1996; Myneni et al.,1997; Zhou et al., 2001).

В ряде работ была показана возможность использования характеристик структуры годичных колец годичных колец в анализе особенностей сезона роста на севере. Ваганов и др. (1999), Ваганов, Шашкин (2000) выявили высокую положительную связь размеров клеток, толщины клеточной стенки и плотности древесины лиственницы на севере Евразии с температурой разных промежутков вегетационного периода и, прежде всего, его начала. Анализ связи параметров профиля плотности колец с климатическими переменными показал, что максимальная плотность и средняя плотность поздней древесины на севере Сибири интегрирует условия за продолжительный период сезона роста и года (Schweingruber et al., 1978; Briffa et al., 1995; D'Arrigo et al., 1992; Vaganov et al., 1999), а минимальная плотность и средняя плотность ранней древесины отражает условия лишь нескольких декад (Кирдянов, 1999). Связь параметров структуры годичных колец хвойных с климатом анализировалась для различных регионов Земного шара (Терсков и др., 1981; Шишов, 2000; Ваганов и др., 2000; Yasue et al., 1996, 2000). На основе анализа так называемых «ложных» и «морозобойных» колец, идентифицируемых на основе клеточного анализа, были выявлены, соответственно, годы с малым количеством осадков в начале сезона роста на востоке австрийских Альп (Wimmer et al., 2000), проведена реконструкция условий в годы с заморозками в различные промежутки вегетационного периода на Полярном Урале и на севере Западной Сибири (Гурская, 2002).

С начала 70-х годов XX века для обобщения результатов, полученных для различных регионов Северного полушария, дендрохронологов и дендроклиматологов начали интересовать вопросы пространственно-временной

реконструкции различных климатических характеристик по дендрохронологическим рядам на базе множественных регрессионных линейных статистических моделей. Первая такая попытка была предпринята Х. Фриттсом (Fritts et al., 1971) с целью реконструкции сезонных и годовых климатических условий по дендрохронологическим данным с 65 равномерно распределенных дендрохронологических участков в западной части Северной Америки. Первой наиболее полной и комплексной российской работой в этом направлении можно считать работу Е.А. Ваганова и др. (1996) по реконструкции аномалий температуры июня-июля с 1611 по 1990 г. по дендрохронологическим данным в пределах Урало-Сибирской Субарктики.

Обобщение результатов в этих работах на региональном уровне проводится на первом этапе при помощи поточечного географического сравнения местообитаний древесных растений с последующей пространственной экстраполяцией результатов на втором этапе при помощи различных модификаций Криддинг-метода. Первый, самый трудоемкий этап подобного пространственного анализа можно автоматизировать применением скользящих ранговых статистик, в частности, коэффициента конкордации Кендалла и методов теории классификации. Кроме этого, классификационные методы оказываются эффективным инструментом выявления взаимосвязей между пространственно удаленными процессами и реконструкции климатических характеристик.

Кроме анализа функций отклика и реконструкций климатических характеристик по дендрохронологическим данным на базе множественного регрессионного моделирования, дендрохронологи применяют данный статистический подход при решении различных экологических задач моделирования, связанных с изучением лесной стендовой динамики и моделирования сукцессионных лесных процессов (Till, 1984; Lischke, 2005; Lischke, Löffler, 2006; Lischke et al., 2007; Rickebusch et al., 2007а; 2007б).

П.4 Пример использования множественной регрессии при пространственно-временном моделировании прироста древесных растений на территории Сибири

В работе были использованы 285 древесно-кольцевых хронологий, полученных для обширной территории России (Рис. 9).

Одновременно со стандартизацией индивидуальных кривых и расчетов временных серий индексов радиального прироста, которые аккумулируют в основном климатически обусловленные изменения прироста, была и стандартизация индивидуальных серий при помощи отрицательной экспоненты (http://www.ltrr.arizona.edu/software.html). Этот способ стандартизации обеспечил максимальное сохранение высокочастотных (погодичных) и низкочастотных составляющих в изучаемых временных рядах (Cook, Kairiukstis, 1990; Шишов и др., 2002)

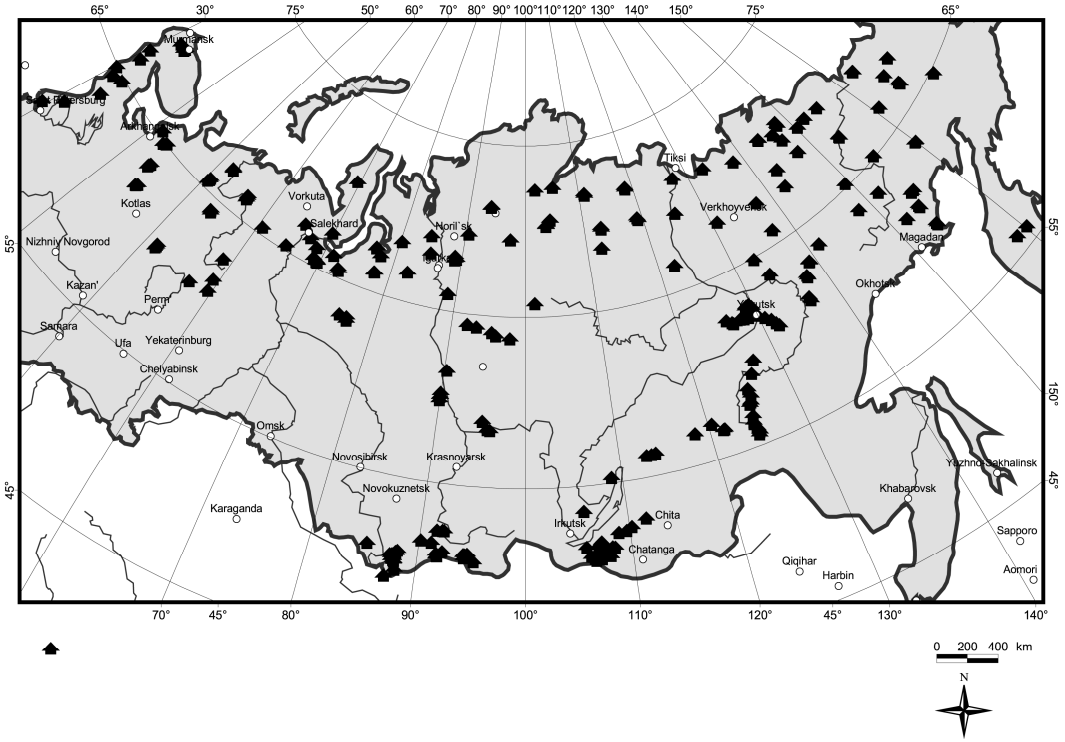


Рис. 9. Карта распределения 285 дендрохронологических участков на территории России

Индексы прироста для отдельного дерева затем усреднялись для всей совокупности деревьев одного участка в соответствии с календарным годом формирования каждого годичного кольца, что в итоге позволяло получить для

каждого участка локальную древесно-кольцевую хронологию, в изменчивости которой содержится сильный климатический сигнал (Шиятов, 1986; Ваганов и др., 1996, Шиятов и др., 1996; Мазепа, 1999б).

Полученные таким образом локальные индексные древесно-кольцевые хронологии становились объектом исследования на следующих этапах работы.

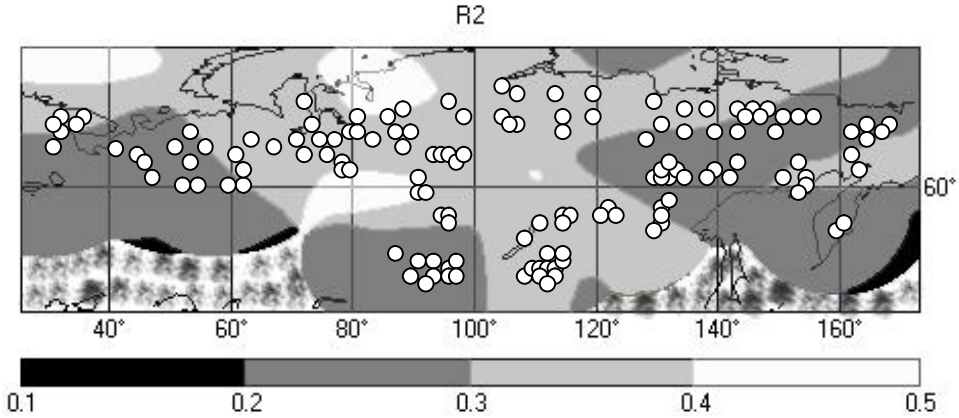
Для сопоставления длительных древесно-кольцевых хронологий с климатическими данными использована база климатических данных высокого пространственного разрешения ($0.5^{o}\times0.5^{o}$), созданная в Центре климатических исследований (Великобритания) (New et al., 2002). В статистическом анализе использовались ежегодные значения средних месячных данных по температуре и осадкам за период нескольких месяцев, предшествующих и текущих сезона роста (сентябрь предыдущего года – август текущего). Для сопоставления древесно-кольцевых хронологий с месячными климатическими данными использовались климатические данные того минимального квадрата ($0.5^{o}\times0.5^{o}$) интерполяционной сетки, в который попадала исследуемая древесно-кольцевая хронология соответственно своим координатам.

Для моделирования прироста древесных растений было использовано два независимых статистических подхода: пошаговая регрессия с включением и традиционная для дендрохронологии множественная регрессия на базе метода главных компонент (Дрейпер, Смит, 1986; Ryan, 1997; Fritts, 1976).

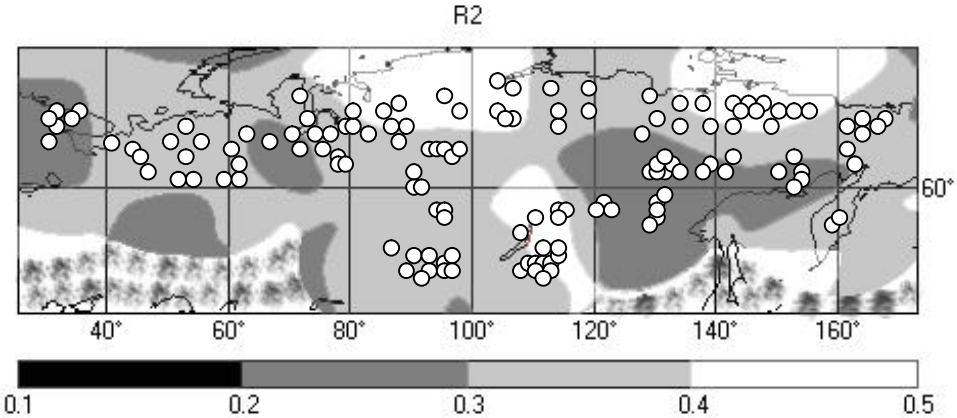
Для анализа соответствия между реальными и модельными временными рядами были рассчитаны следующие характеристики: коэффициент корреляции Пирсона R, средняя сумма разностей между анализируемыми рядами D, средняя интегральная разность между кривыми и среднее Евклидово расстояние E. Средние характеристики были использованы для того, чтобы уменьшить влияние объема выборки. Дело в том, что только 30% анализируемых древесно-кольцевых хронологий имеют длительность до 1998 года и более. Большинство дендрохронологического материала было собрано в период с 1989 по 1994 гг. Следовательно, большинство древесно-кольцевых хронологий не охватывают период в 4-10 лет на интервале с 1989 по 1998 гг. Это приводило к тому, что интегральные характеристики (например, сумма разностей между рядами)

сильно отличались из-за 12-30% недостающих значений для периода верификации моделей. Нормировка этих характеристик к соответствующему объему выборки (длине анализируемого ряда) значительно уменьшало влияние эффекта длительности хронологий.

Пространственные аппроксимации расчетных параметров древесно-кольцевых хронологий были получены на основе метода обратных взвешенных расстояний, который широко используется в ГИС-технологиях, наряду с различными модификациями Криддинг-метода (Высоцкая и др., 2002).



А)



Б)

Рис. 10. Пространственное распределение коэффициентов детерминации, оценивающих качество регрессионных зависимостей прироста деревьев от климатических переменных, для следующих интервалов: А) 1935-1965 гг.; Б) 1965-1998 гг.

Согласно предварительно высказанной гипотезе (о разной скорости возрастания концентрации углекислоты в атмосфере до и после 1965 года)

анализируемый временной интервал с 1935 по 1998 гг. был разбит на две части. На периоде с 1935 по 1965 гг. были построены множественные регрессионные модели зависимости прироста деревьев от температуры и осадков.

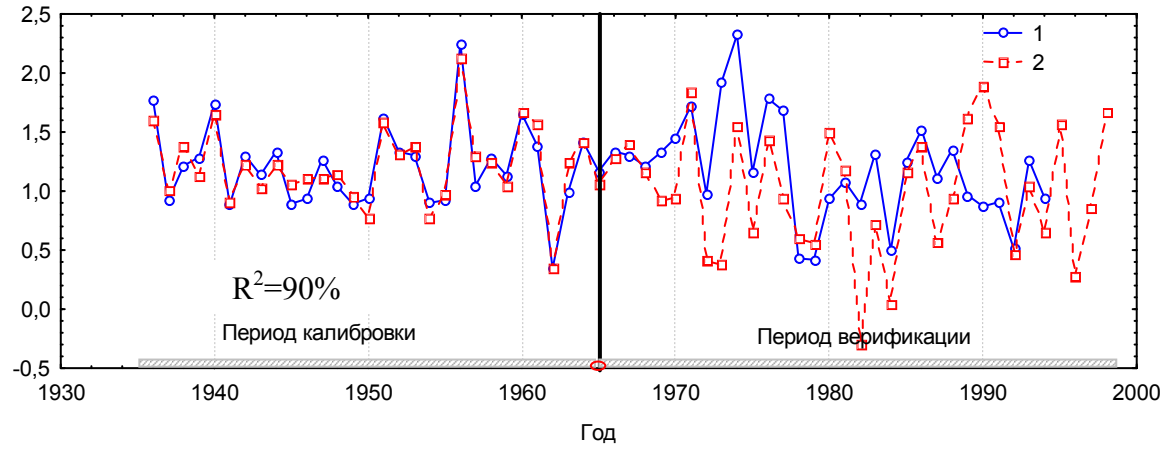
На интервале с 1966 по 1998 гг. вероятный прирост древесных растений (смоделированная кривая прироста) был оценен с помощью соответствующей найденной множественной регрессионной зависимости. Более того, были проведены расчеты для случая, когда калибровочный (1935-1965 гг.) и верификационный периоды (1965-1998 гг.) менялись местами. В этом случае, регрессионные модели строились на интервале с 1965 по 1998 гг., а прогноз (реконструкция) по полученным моделям проводилась для периода 1935-1965 гг.

На рисунке 10 приведены результаты пространственного распределения интегральной оценки качества построенных методом пошаговой регрессии моделей для двух выделенных периодов.

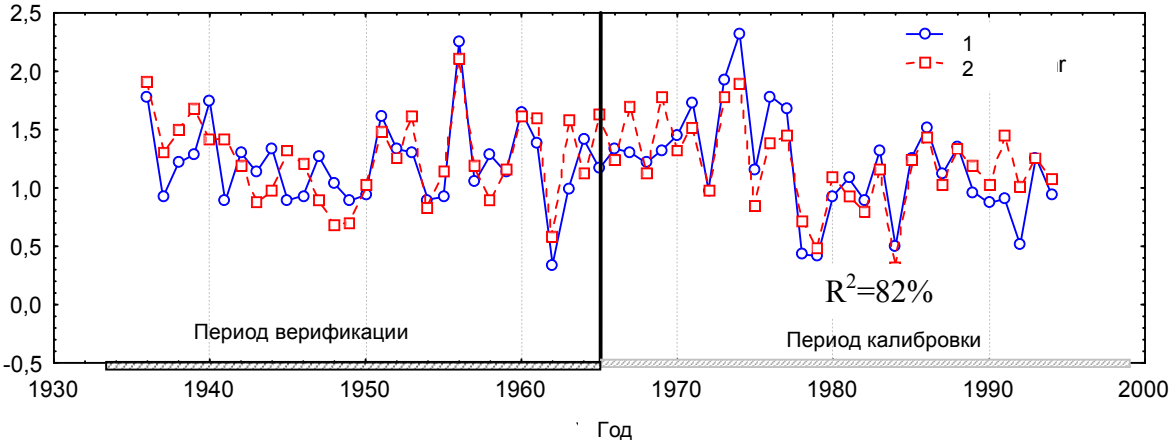
Качество построенных множественных регрессионных моделей как на интервале с 1935 по 1965 гг., так и с 1965 по 1998 гг. является хорошим. Доля климатически обусловленной составляющей для 1965-1998 гг. в среднем объясняет 57% изменчивости прироста и с вероятностью 95% лежит в пределах от 55 до 59% для всей анализируемой территории (Рис. 10 Б). При этом доля выше, чем соответствующий средний коэффициент детерминации (52%), полученный для предыдущего интервала (1935-1965 гг.) (Рис. 10 А).

Для некоторых дендрохронологических участков доля климатически-обусловленной изменчивости достигала 90% (Регионы низовьев р. Индигирка, Таймыра, Эвенкии и Полярного Урала) (Рис. 11).

На интервале с 1966 по 1998 гг. вероятный прирост древесных растений (смоделированная кривая прироста) был оценен при помощи соответствующей найденной множественной регрессионной зависимости. Более того, были проведены расчеты для случая, когда калибровочный (1935-1965 гг.) и верификационный периоды (1965-1998 гг.) менялись местами. В этом случае, регрессионные модели строились на интервале с 1965 по 1998 гг., а прогноз (реконструкция) по полученным моделям проводилась для периода 1935-1965 гг.

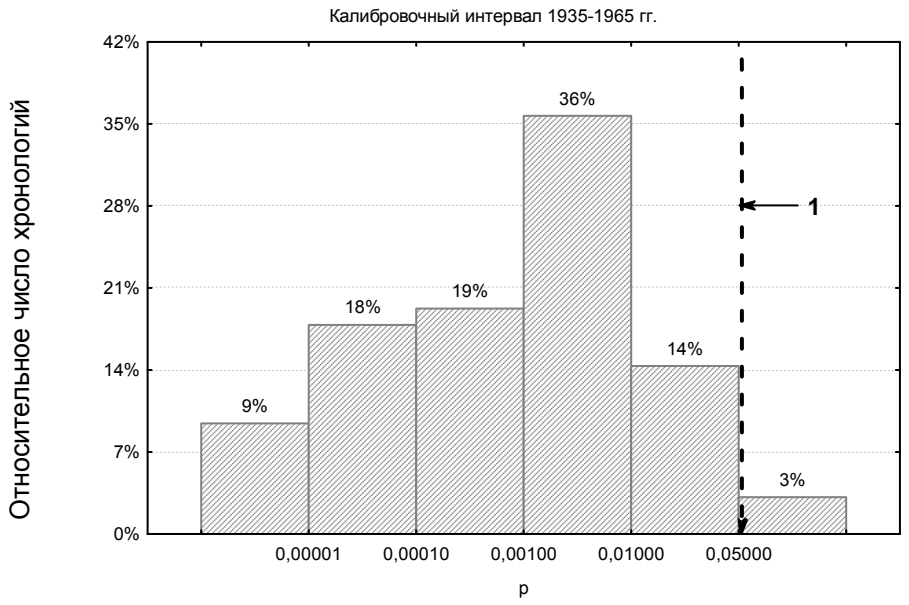


А)

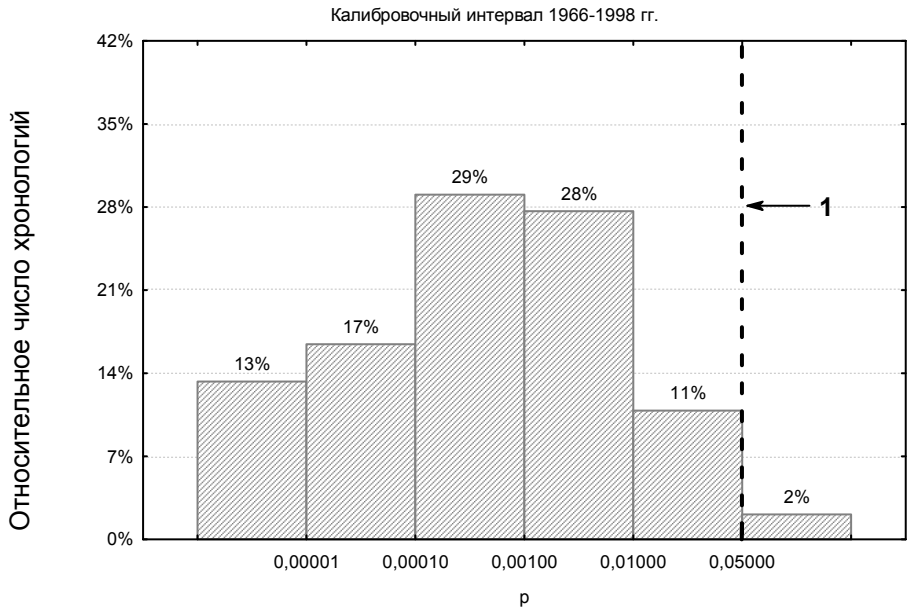


Б)

Рисунок 11. Динамика реальной (1) и смоделированной древесно-кольцевых (2) хронологий, полученных для дендрохронологического участка (70.5° С.Ш., 148.1° В.Д.), находящегося в низовьев р. Индигирка для разных периодов калибровки: А) 1935-1965 гг. (R^2=90%); Б) 1965-1998 гг. (R^2=82%).



А)



Б)

Рис. 12. Гистограмма распределения уровня надежности p, полученного по критерию Фишера для всех анализируемых статистических моделей, полученных для разных калибровочных интервалов: А) 1935-1965 гг. и Б) 1966-1998 гг. (1 – критический уровень надежности в 0.05).

В целом, на обоих периодах калибровки статистических моделей (1935-1965 гг. и 1965-1998), полученных для всех древесно-кольцевых хронологий, только 2% и 3% (соответственно) полученных регрессионных зависимостей не

являются адекватными для выбранного уровня надежности в 0.05 (Рис. 12). Более того, для калибровочного интервала с 1965 по 1998 гг. распределение уровня надежности статистических моделей меняет свой характер. Мода и медиана этого распределения смещается в интервал от 0.0001 до 0.001. Это означает, что в целом качество регрессионных моделей, полученных для этого интервала, улучшается.

Был проведен анализ частоты встречаемости климатических факторов, оказывающих значимое влияние на динамику радиального прироста древесных растений на севере Евразии отдельно для двух выделенных периодов. На рисунке 13 частоты значимых климатических факторов отложены в соответствии со знаком их влияния на прирост – положительным или отрицательным. Нетрудно видеть, что в 40% случаев доминирующее положительное влияние на прирост древесных растений оказывает июньская температура (t6). Это соответствует общепринятым представлениям о лимитирующей роли температуры в изменчивости прироста на севере Евразии (Ваганов и др., 1996; Briffa et. al., 1998а; 1998в; 2008). Однако для первого из исследуемых периодов влияние температуры июля меньше, чем для второго временного периода. Важно также отметить положительное влияние осенних (октябрь) температур на рост древесных растений на севере Евразии. Наиболее существенным отличием двух периодов является то, что для первого периода по сравнению со вторым увеличивается частота встречаемости значимого отрицательно влияния зимних осадков.

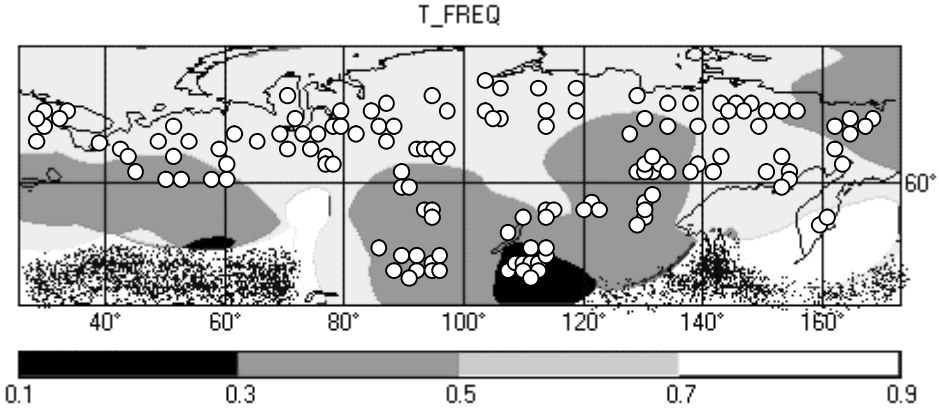
Можно отметить, также, возрастание отрицательного влияния апрельских температур во втором периоде, что свидетельствует о большей зависимости роста древесных растений в исследуемой области от условий, предшествующих сезону роста (начало сезона на севере Евразии отнесено к маю-июню, в зависимости от широты местности) (Ваганов, Шашкин, 2000).

А)

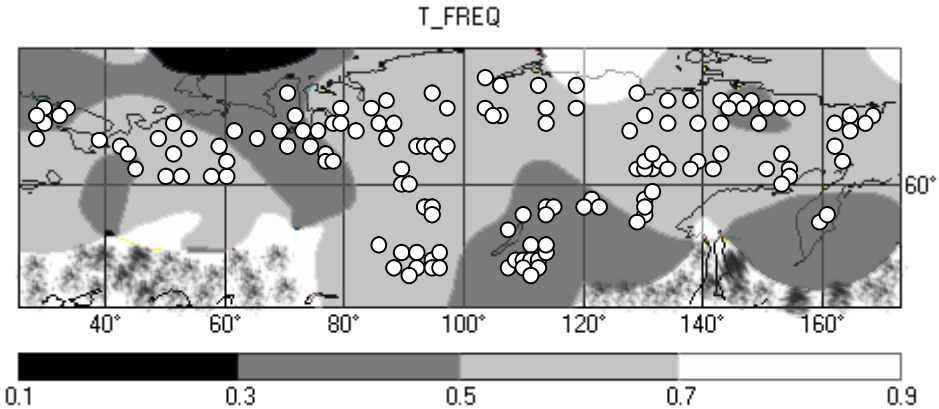
Б)

Рис. 13. Частота встречаемости значимых климатических факторов, влияющих на прирост древесных растений (частоты отложены в соответствии со знаком влияния, положительным или отрицательным) в полученных статистических моделях для различных калибровочных интервалов.

Наиболее температурно-чувствительными оказались регионы циркумполярной области, что не раз отмечалось в целом ряде современных публикаций (Ваганов и др., 1996; Cook, Kairiukstis, 1990; Naurzbaev, Vaganov, 1999; Briffa et. al., 1998, Мазепа, 2000; Хантемиров, Шиятов, 1999; Хантемиров, 2000). Менее чувствительными к температуре оказались хронологии, расположенные в центральных и южных районах Урала, Красноярского края, Иркутской области и Чукотки, где существенное влияние на прирост начинают оказывать осадки (Рис. 14.).



А)



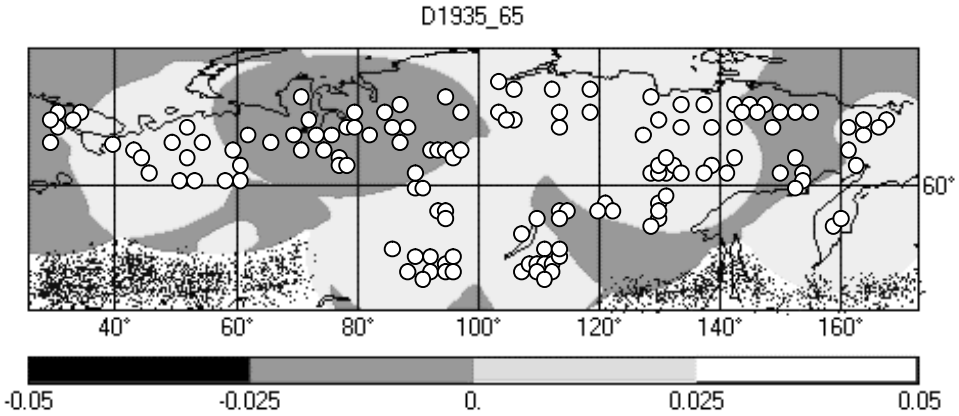
Б)

Рис. 14. Пространственное распределение частоты встречаемости среднемесячных температурных характеристик в полученных статистических зависимостях для двух калибровочных интервалов: А) 1935-1965 гг. и Б) 1965-1998 гг.

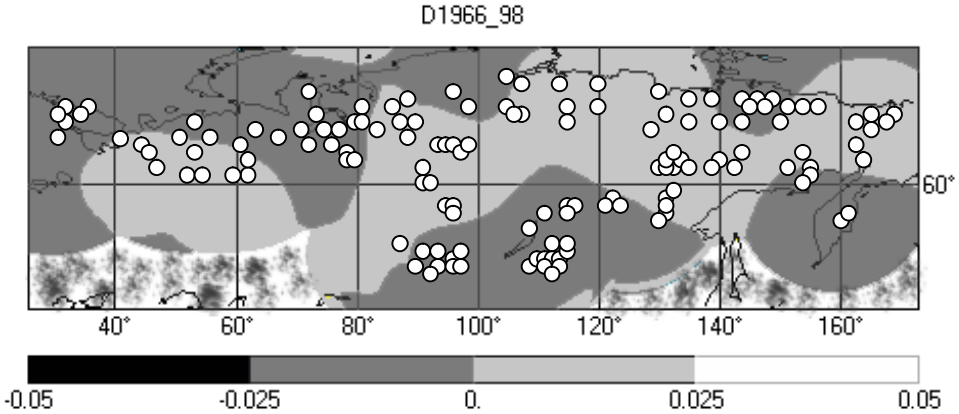
Особо выделяется регион Бурятии, где доминирующее воздействие на прирост древесных растений (до 70% объясненной дисперсии) оказывают, в

основном, весенние и летние осадки текущего года (Андреев и др., 1999). В тоже время, отмечается одна интересная особенность для калибровочного интервала с 1965 по 1998 гг., которая выражается в том, что для регионов Ямала, северной и средней части Западной Сибири происходит смена лимитирующего фактора. В отличие от предыдущего интервала, на прирост существенное влияние начинают оказывать осадки (Мазепа, 1999).

Проведем анализ соответствия между смоделированными индексными рядами и реальными древесно-кольцевыми хронологиями для двух верификационных периодов: с 1965 по 1998 гг. (калибровочный интервал 1935-1965 гг.) и с 1935 по 1965 гг. (калибровочный интервал 1965-1998 гг.).



А)



Б)

Рис. 15. Пространственное распределение средних сумм разностей между реальными и модельными 285 древесно-кольцевыми хронологиями на калибровочных интервалах: А) 1935-1965 гг. и Б) 1965-1998 гг.

На основании множественных регрессионных моделей, рассчитанных для всех древесно-кольцевых хронологий, были получены модельные ряды прироста древесных растений с 1935 по 1965 гг. и с 1965 по 1998 гг. (различные периоды калибровки модели). По аналогии для различных периодов верификации моделей с 1965 по 1998 гг. и с 1935 по 1998 гг. были получены прогнозные значения прироста.

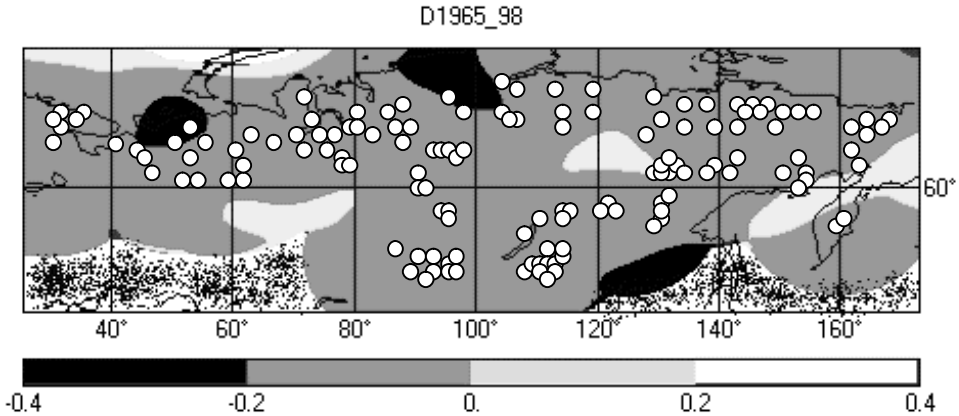
Рассмотрим карты пространственной изменчивости основных статистических характеристик реальных и модельных хронологий на двух калибровочных временных интервалах: А)1935-1965 гг. и Б) 1966-1998 гг (Рис.15).

Пространственное распределение средних сумм разностей указывает на то, что между реальными и модельными рядами древесного прироста не существует значимых различий на обоих периодах калибровки (Рис.15.). Это еще раз указывает на хорошую адекватность статистических моделей.

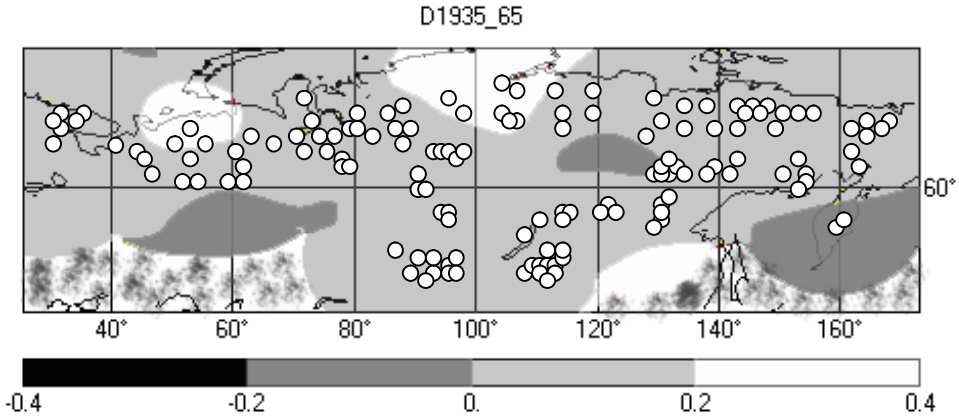
Картина резко меняется при переходе на верификационные интервалы. Между модельными и реальными хронологиями появляются существенные различия (Рис.16).

Различия являются отрицательными практически для всей анализируемой области на периоде верификации моделей с 1965 по 1998 год (соответствующий период калибровки моделей - 1935-1965 гг.) (Рис. 16)). Это означает, что статистическое моделирование приводит на верификационном временном интервале к завышенной оценки прироста древесных растений по сравнению с реально существующими величинами. Напротив, при использовании моделей, полученных на интервале 1966-1998, статистическое моделирование приводит на верификационном временном интервале к заниженной оценке прироста древесных растений по сравнению с реально существующими величинами на периоде верификации с 1935 по 1966 гг. (Рис.16 Б).

Более того, полученные пространственные распределения на соответствующих верификационных интервалах сильно отрицательно коррелированны между собой ($R=-0.86$, $p<<0.01$).



А)



Б)

Рис. 16. Пространственное распределение средних сумм разностей между реальными и модельными 285 древесно-кольцевыми хронологиями на верификационных интервалах: А) 1965-1998 гг. и Б) 1935-1965 гг.

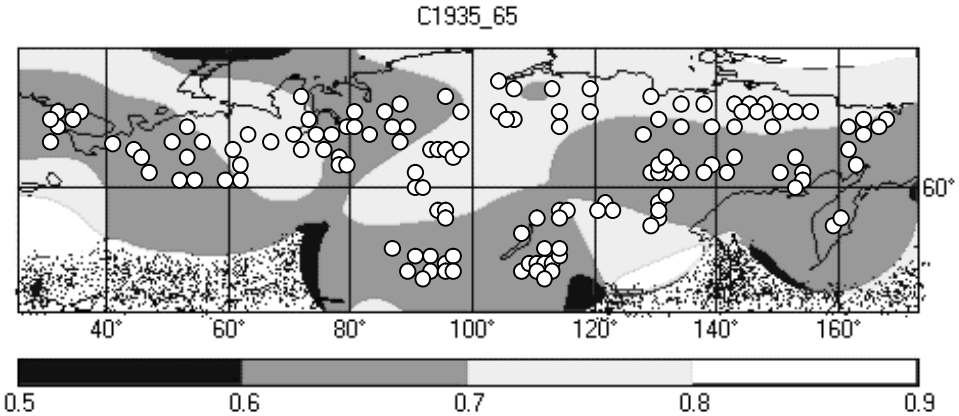
Становится очевидным, что, несмотря на добротность регрессионных моделей, полученных для разных калибровочных интервалов (1935-1965 и 1965-1998 гг.), особенности прироста древесных растений не могут быть также добротно смоделированы на соответствующих калибровочных интервалах времени (1966-1998 и 1935-1965 гг.)

Одной из причин этого, вероятно, могут быть различия в динамике лимитирующих прирост климатических факторов для двух анализируемых интервалов времени. Это приводит к перераспределению факторных нагрузок между климатическими факторами в полученных регрессионных зависимостях,

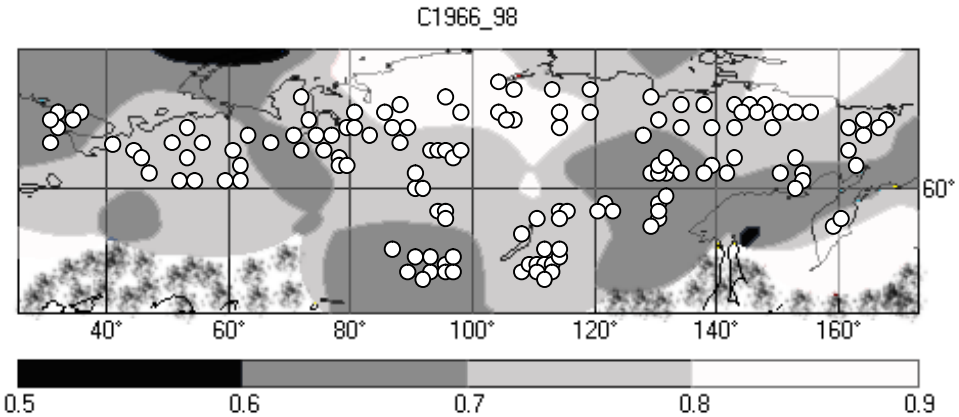
что можно интерпретировать как смену лимитирующих прирост факторов для 1966-1998 гг. по сравнению с предыдущим интервалом (1935-1965 гг.) Последнее подтверждается, в частности, ранее полученным пространственным распределением для регионов Ямала, северной и средней части Западной Сибири.

Аналогичные результаты получаются и при анализе пространственных распределений, полученных для других характеристик..

На периодах калибровки статистических моделей уровень корреляции находится в пределах от 0.6 до 0.8 ($p<0.05$) (Рис. 17).



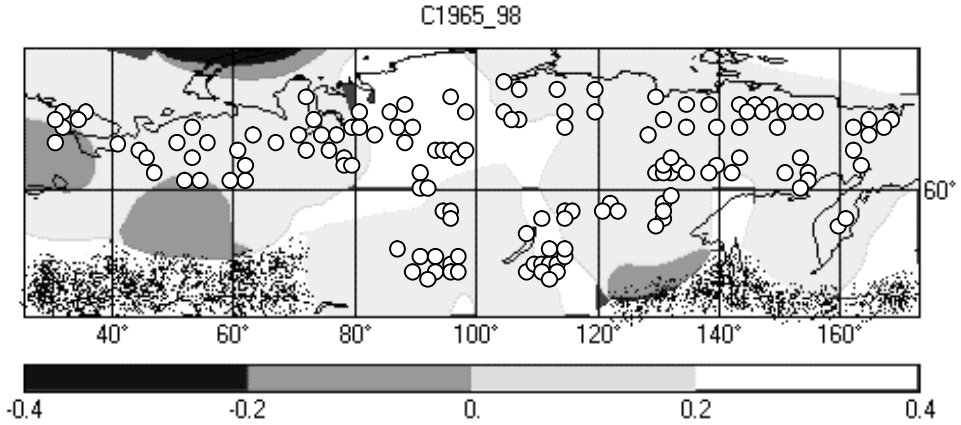
А)



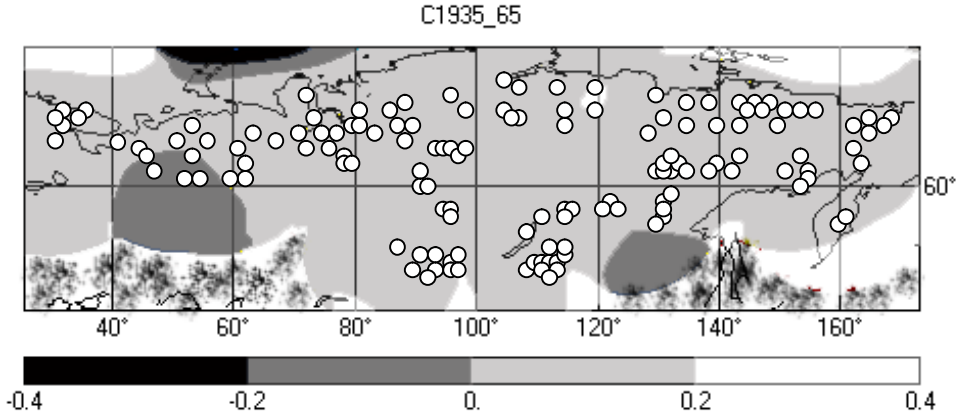
Б)

Рис. 17. Пространственное распределение коэффициентов корреляции между реальными и модельными 285 древесно-кольцевыми хронологиями на калибровочных интервалах: А) 1935-1965 гг. и Б) 1965-1998 гг.

В отличии от этого, модельные и реальные хронологии становятся слабо коррелированными (p>0.05) на соответствующих интервалах верификации моделей (Рис. 18).



А)



Б)

Рис. 18. Пространственное распределение коэффициентов корреляции между реальными и модельными 285 древесно-кольцевыми хронологиями на верификационных интервалах: А) 1965-1998 гг. и Б) 1935-1965 гг.

В тоже время, статистические модели, полученные для калибровочного периода с 1935 по 1965 гг., лучше моделируют прирост (по сравнению с аналогичными моделями, полученными для другого калибровочного интервала 1965-1998 гг.) для большей территории Таймыра севера и центра Красноярского края, небольшой части Иркутской области и Бурятии (Рис. 18 А). При этом,

средний уровень корреляции для выделенных областей лежит в пределах от 0.2 до 0.4 (p=0.06).

Следовательно, снижение чувствительности прироста древесных растений к климатическим факторам во втором периоде (1965-1998 гг.) подтверждается только для локальной части анализируемой территории.

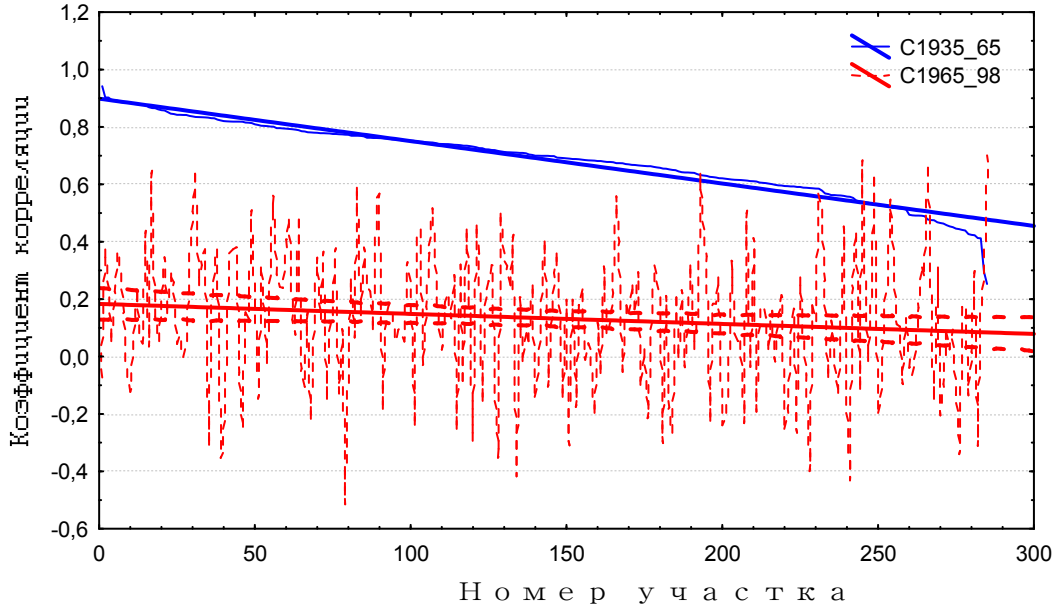
Для оценки зависимости качества статистического прогноза прироста древесных растений на анализируемых интервалах от качества полученных статистических моделей, используем следующую процедуру. Предварительно упорядочим множество коэффициентов корреляции, состоящее из 2-х значений корреляций для каждого из 285 участков, следующим образом. Перенумеруем исходное множество участков (или древесно-кольцевых хронологий) в соответствии с ранжированными в порядке убывания коэффициентов корреляции, полученных для соответствующих калибровочных интервалов: с 1935 по 1965 гг. и с 1965 по 1998 гг.

Отложим по оси абсцисс новые номера участков, а по оси ординат - ранжированные значения корреляции для калибровочных интервалов: 1935-1965 гг. и 1965-1998 гг., а также соответствующие им значения корреляций для верификационных интервалов: 1965-1998 гг. и 1935-1965 гг. (Рис.19).

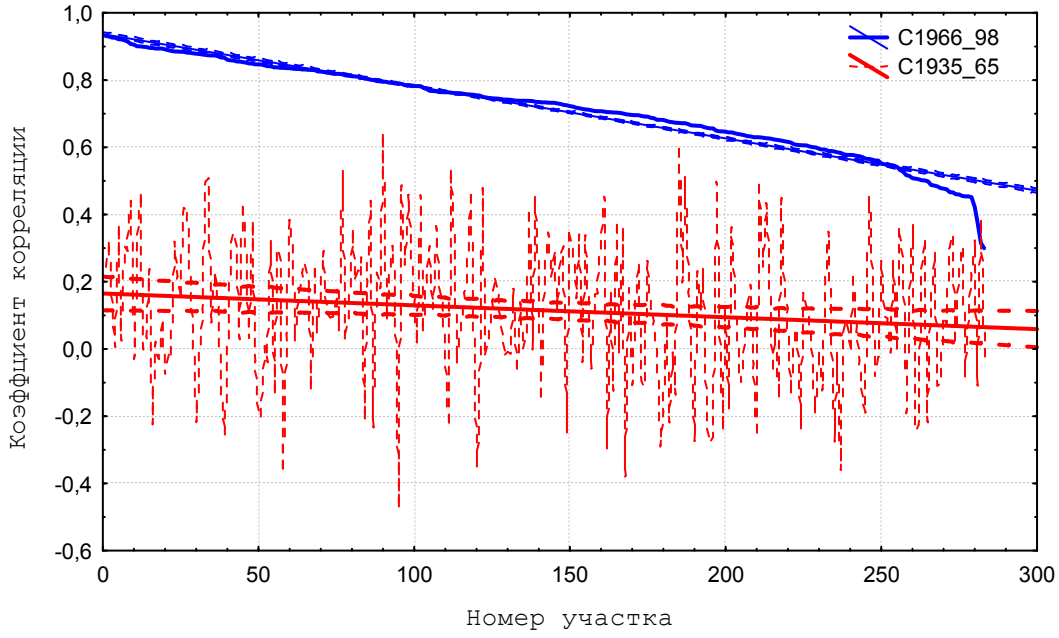
Предположение о том, что качество прогноза будет зависеть от того, насколько добротная статистическая модель используется для этого прогноза, оказывается неверным. Явный тренд, полученный для упорядоченного множества коэффициентов корреляции на периоде калибровки, не сопоставляется с тенденцией для соответствующих коэффициентов корреляции на периоде верификации (или прогноза). Линейная тенденция в последнем случае отсутствует. Последнее указывает на особенность второго периода в 20-м столетии по сравнению с предыдущим 30-летним периодом с точки зрения влияния внешних факторов на прирост древесных растений

Следствием данного анализа является то, что какие бы добротные статистические модели не были получены на периоде калибровки, нельзя напрямую использовать такие модели при реконструкции (или прогнозе) без

доказательств того, что калибровочный интервал охватывает все возможные особенности, которые могут встретиться в прошлом или будущем.



А)



Б)

Рис. 19. Изменчивость коэффициентов корреляции в зависимости от номера дендрохронологического участка для двух временных интервалов: 1935-1965 гг. (коэффициенты ранжированы в порядке убывания) и 1965-1998 гг.

Полученные результаты, связанные с нарушением гетероскедантичности регрессионных моделей, необходимо учитывать при построении длительных

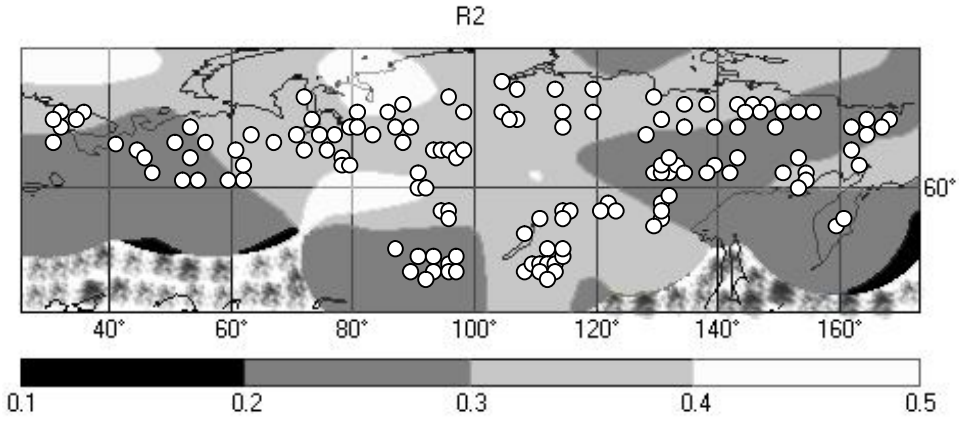
температурных реконструкций, так как не существует гарантий, что подобного рода изменения не имели места в прошлом (Briffa et al., 1998а; 1998б).

Более того, подтвержденное снижение чувствительности древесных растений на температурные изменения для территории России может означать, что оценки концентрации атмосферного CO_2, полученные на базе современных глобальных моделей (carbon-cycle models), могут быть не точны (Cias et al., 1995; Sitch et al., 2003).

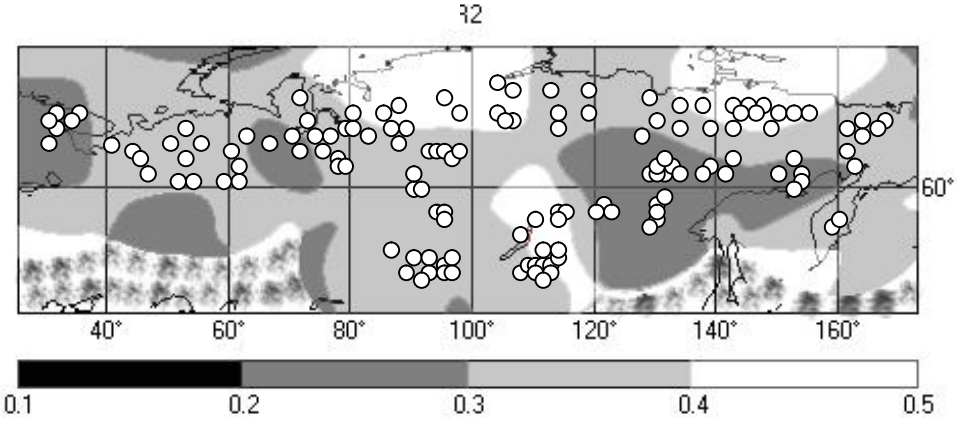
Проведем аналогичный анализ на базе ортогональной множественной регрессии. По структуре эта часть раздела полностью повторяет предыдущую для лучшего сопоставления результатов. Различие связано с некоторым перекрытием первого и второго периодов (первый – 1935-1975 гг., второй – 1955-1998 гг.). Это перекрытие сделано сознательно по двум причинам: 1) улучшить качество моделей на периоде калибровки за счет увеличения числа степеней свободы для соответствующих регрессионных зависимостей; 2) проверить устойчивость в тенденциях моделированных и реальных хронологий, отмеченных ранее на результатах пошаговой регрессии.

Среднемесячные данные температуры и осадков (сентябрь предшествующего года – август текущего) преобразовывались методом главных компонент, и для каждого грида использованы 9 первых главных компонент, описывающих 65-75% изменчивости исходных климатических изменений. Эти главные компоненты затем использованы как независимые переменные при расчетах климатических функций отклика.

Качество построенных множественных регрессионных моделей на двух калибровочных интервалах с 1935 по 1975 гг. и с 1955 по 1995 гг. является несколько ниже, чем для варианта с пошаговой регрессией. Это объясняется тем, что каждая главная компонента включает в себе информацию о 24 анализируемых климатических переменных. Часть из этих климатических переменных не оказывают прямого или опосредованного воздействия на прирост древесных растений, т. е. являются «шумом» при регрессионном моделировании.



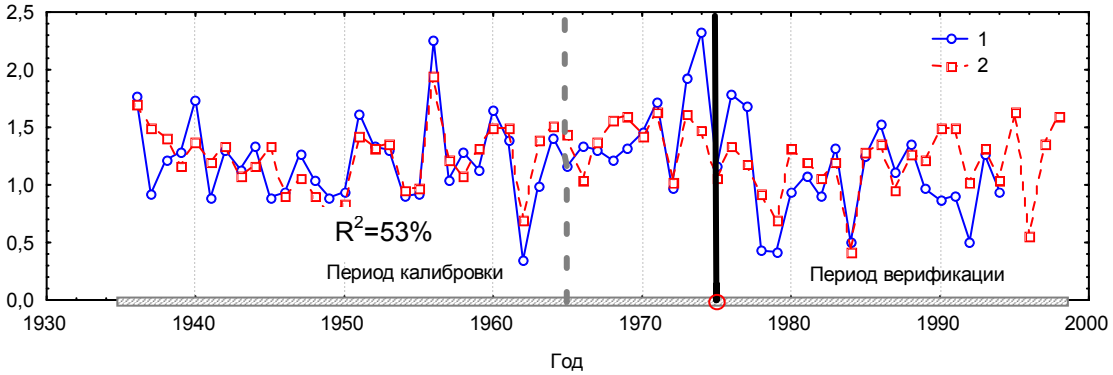
А)



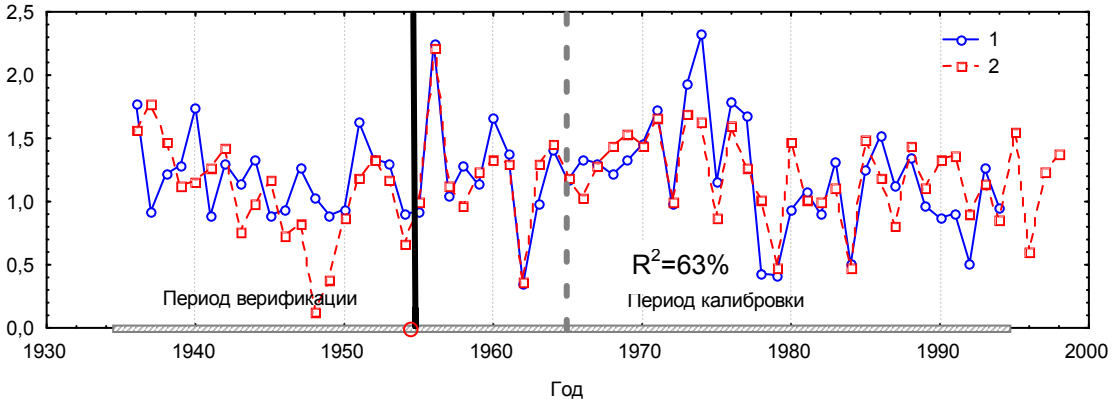
Б)

Рис. 20. Пространственное распределение коэффициентов детерминации, оценивающих качество регрессионных зависимостей прироста деревьев от 9 главных компонент, полученных по климатическим переменным для следующих интервалов: А) 1935-1965 гг.; Б) 1965-1998 гг.

Доля климатически обусловленной составляющей для 1955-1995 гг. в среднем объясняет 34% изменчивости прироста и с вероятностью 95% лежит в пределах от 31 до 36% для всей анализируемой области России. Этот показатель сравним с соответствующим средним коэффициентом детерминации (32%), полученный для предыдущего интервала (1935-1975 гг.) (Рис. 20 А).



А)



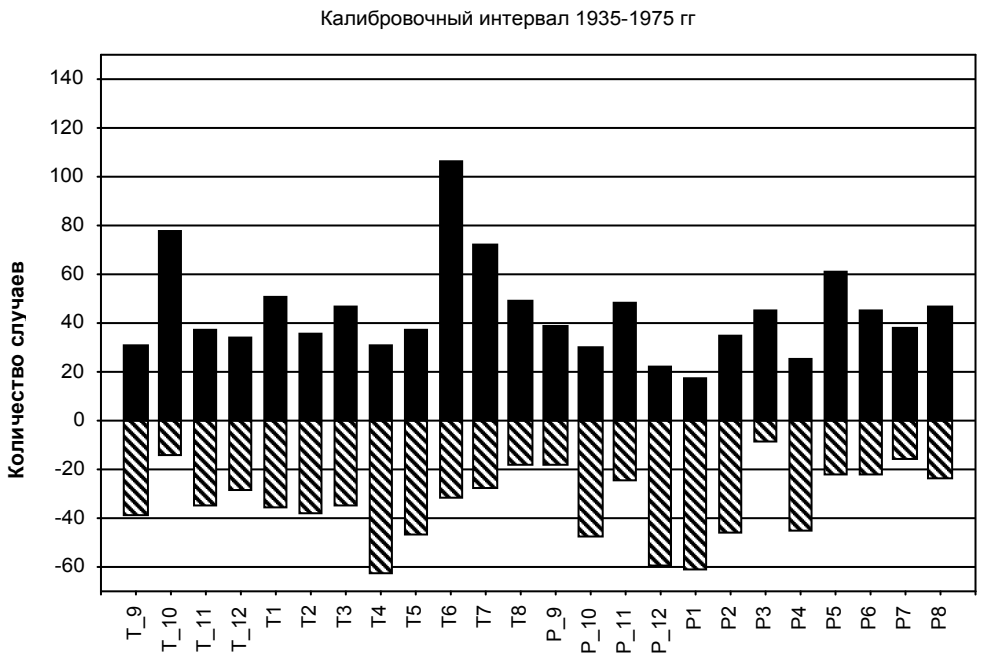
Б)

Рис. 21. Динамика реальной (1) и модельной (2) древесно-кольцевых хронологий, полученных для дендрохронологического участка (70.5° С.Ш., 148.1° В.Д.), находящегося в низовьев р. Индигирка для разных периодов калибровки: А) 1935-1975 гг. (R^2=53%); Б) 1955-1998 гг. (R^2=63%).

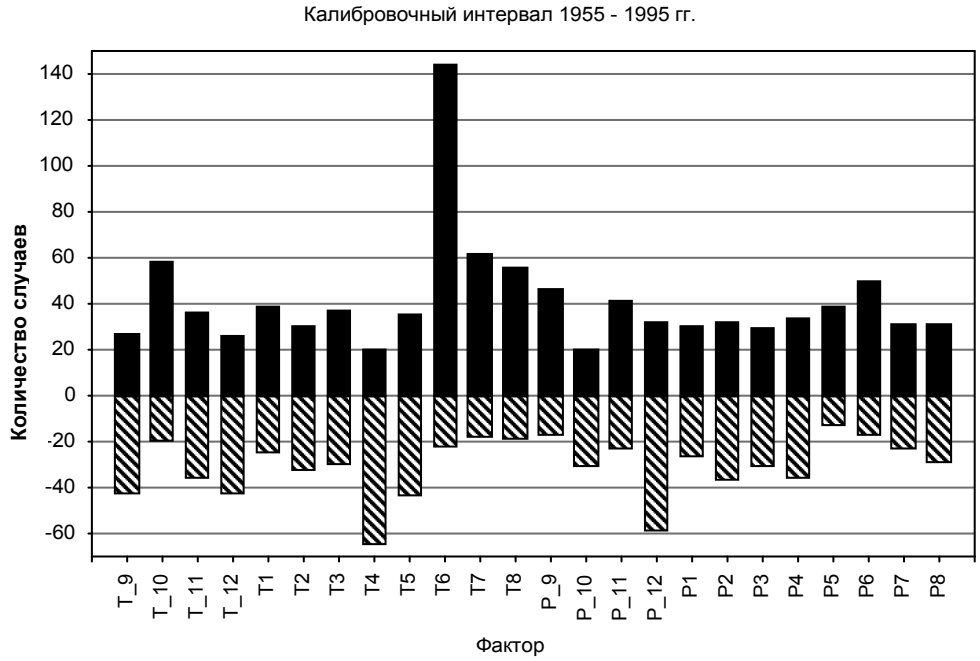
Для некоторых дендрохронологических участков доля климатически-обусловленной изменчивости достигала 70% (Регионы низовьев р. Индигирка, Таймыра, Эвенкии и Полярного Урала) (Рис. 21).

В целом, на обоих калибровочных периодах (1935-1975 гг. и 1955-1995 гг.) большинство статистических моделей, полученных для всех древесно-кольцевых хронологий, адекватны по оценкам достоверности ($p < 0.05$).

Анализ частоты встречаемости климатических факторов, оказывающих значимое влияние на динамику радиального прироста древесных растений на севере Евразии отдельно для двух выделенных периодов, показывает, что в 50% случаев доминирующее положительное влияние на прирост древесных растений оказывает июньская температура (t6) (Рис. 22).



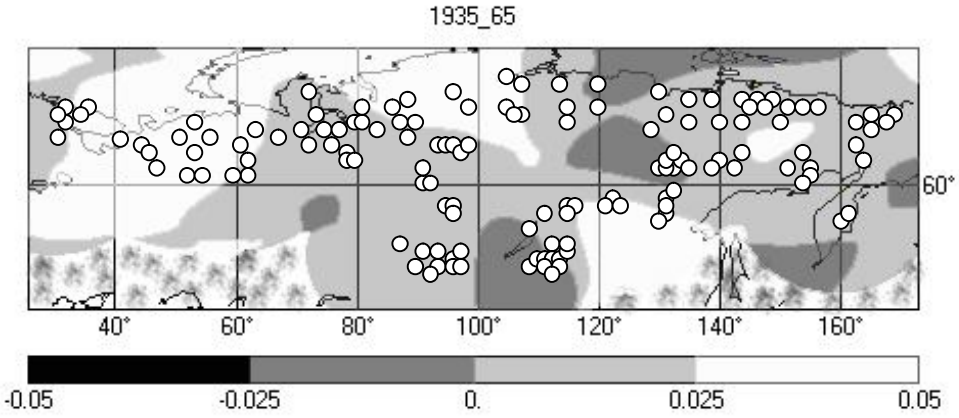
А)



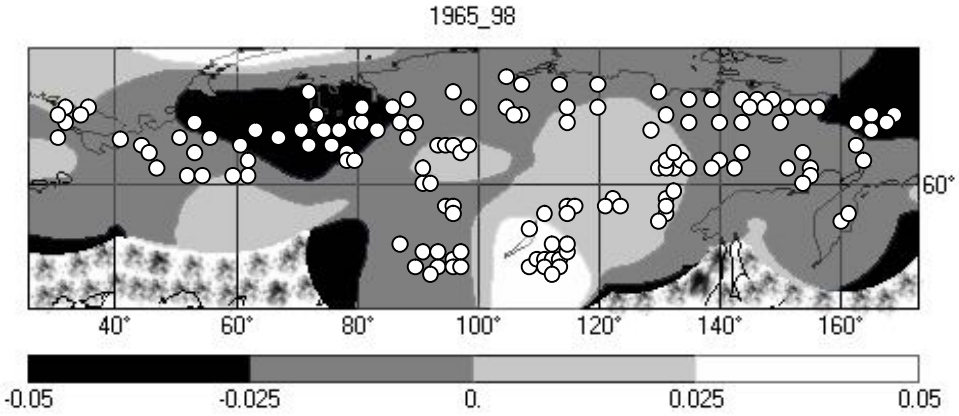
Б)

Рис. 22. Количество появлений климатических переменных, которые оказывают значимое воздействие на прирост древесных растений, во всех полученных статистических моделях для различных калибровочных интервалов.

Однако для первого из исследуемых периодов влияние температуры июня меньше, чем для второго временного периода. Преобладающим влиянием температуры июня во втором периоде результаты расчета функций отклика отличаются от таковых, полученных методом пошаговой регрессии. Положительное влияние осенних (октябрь) температур на рост древесных растений на севере Евразии также подтверждается и в расчетах функций отклика. Данные рисунка 22 подтверждают, что для первого периода увеличивается частота встречаемости значимого отрицательно влияния зимних осадков. А отрицательное влияние апрельских температур и в первом, и во втором периоде приблизительно одинаково.



А)



Б)

Рис. 23. Пространственное распределение средних сумм разностей между реальными и модельными 285 древесно-кольцевыми хронологиями на интервалах: А) 1935-1965 гг. и Б) 1965-1998 гг.

На основании множественных регрессионных моделей, рассчитанных для всех древесно-кольцевых хронологий, для двух калибровочных интервалов были получены модельные ряды прироста древесных растений с 1935 по 1998.

Для анализа соответствия между реальными и модельными временными рядами были использованы следующие характеристики: средняя сумма разностей между анализируемыми рядами D и коэффициент корреляции Пирсона.

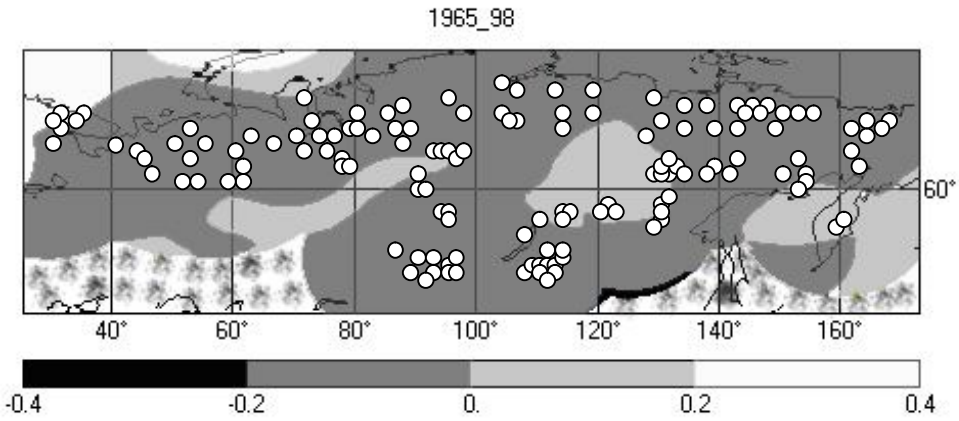
Отметим, что два выделенных периода (1935-1965 и 1965-1998 гг.) на 10 годах перекрывают оба калибровочных интервала. Большинство полученных статистических моделей на разных калибровочных интервалах описывают от 20% до 50% изменчивости радиального прироста деревьев.

Рассмотрим пространственные особенности статистических характеристик моделированных и реальных хронологий на двух временных интервалах: А)1935-1965 гг. и Б) 1965-1998 гг., которые являются частями калибровочных интервалов с 1935 по 1975 гг. и с 1955 по 1995 гг., соответственно (Рис. 23). Как и в предыдущем разделе, эти различия являются незначимыми.

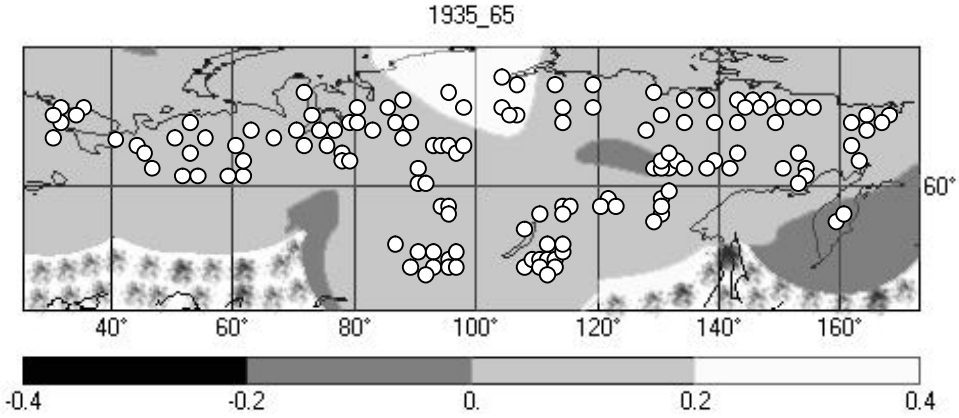
Эти различия отрицательны практически для всей анализируемой территории на периоде с 1965 по 1998 год (соответствующий период калибровки моделей - 1935-1975 гг.) (Рис. 24 А). Это означает, что статистическое моделирование приводит на верификационном временном интервале к завышенной оценки прироста древесных растений по сравнению с реально существующими величинами. Напротив, при использовании моделей, полученных на интервале 1965-1998 гг., статистическое моделирование приводит на верификационном временном интервале к заниженной оценке прироста древесных растений по сравнению с реально существующими величинами на периоде верификации с 1935 по 1965 гг. (Рис. 24 Б).

Эти результаты полностью соответствуют результатам, полученным при пошаговом регрессионном анализе. Более того, полученные пространственные распределения разностей при независимых регрессионных подходах получаются практически идентичными.

Становится очевидным, что, несмотря на различные статистические подходы к моделированию, особенности прироста древесных растений на интервалах времени за пределами калибровочного периода не могут быть также хорошо смоделированы, как и на калибровочных интервалах..



А)



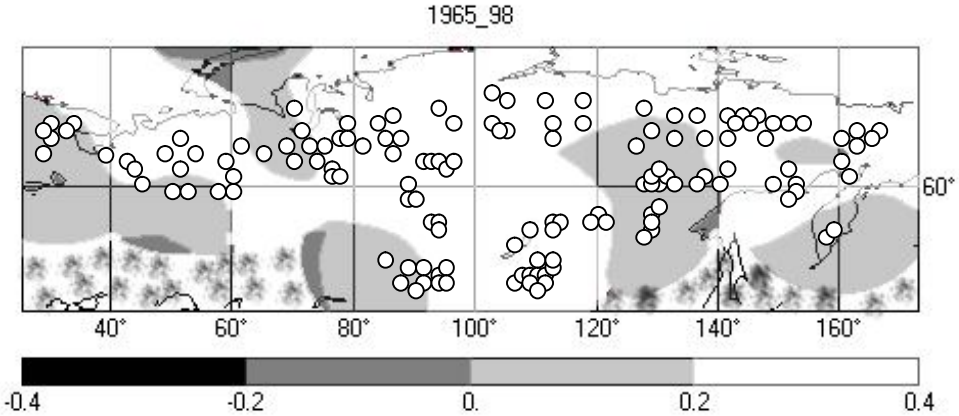
Б)

Рис. 24. Пространственное распределение средних сумм разностей между реальными и модельными 285 древесно-кольцевыми хронологиями на интервалах: А) 1965-1998 гг. и Б) 1935-1965 гг.

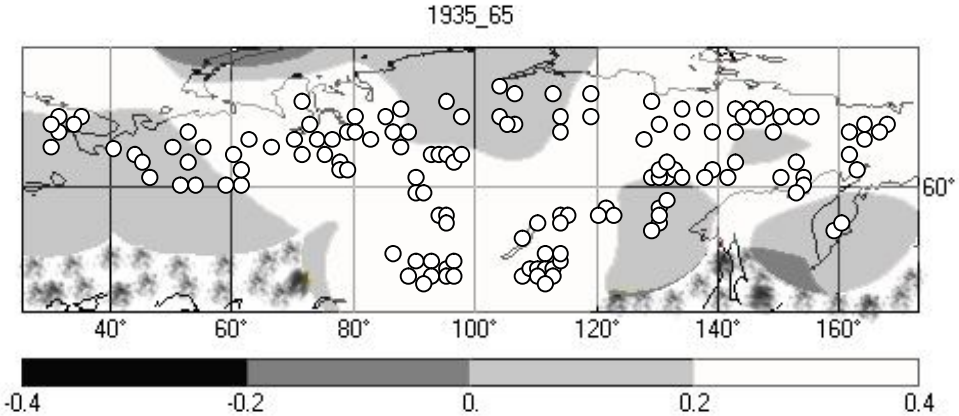
Пространственные распределения, полученные для коэффициентов корреляции, рассчитанных между реальными и модельными кривыми прироста для рассматриваемых временных интервалов, показывают, что уровень корреляции между модельными (калибровочный интервал 1935-1975 гг.) и реальными древесно-кольцевыми хронологиями лежит в пределах от 0.4 до 0.6 ($p<0.05$) для 1935-1965 гг. Аналогичный показатель между модельными (1955-

1995 гг.) и реальными древесно-кольцевыми хронологиями в среднем выше и лежит в диапазоне от 0,4 до 0,7 ($p<0,05$).

В отличие от этого, уровень корреляции между модельными и реальными хронологиями значительно снижается на временных интервалах, которые являются частями верификационных интервалов с 1965 по 1995 и с 1935 по 1965 гг. для соответствующих регрессионных моделей (Рис. 25).



А)



Б)

Рис. 25. Пространственное распределение коэффициентов корреляции между реальными и модельными 285 древесно-кольцевыми хронологиями на временных интервалах: А) 1965-1998 гг. и Б) 1935-1965 гг.

Если сравнить полученные распределения с корреляционными распределениями, полученными при пошаговой регрессии, то можно отметить, что уровень корреляций несколько выше в случае моделей, полученных с

помощью расчета функций отклика. Объясняется это тем, что данные временные интервалы (1965-1998 гг. и 1935-1965 гг.) включают в себя не только «чистые» модельные значения древесно-кольцевых хронологий на периодах верификации моделей (1975-1995 и 1935-1955 гг.), но 11 значений, полученных на соответствующих интервалах калибровки статистических моделей. Именно за счет этих значений смоделированных хронологий уровень корреляции увеличивается и становится значимым ($p<0.05$) практически для всей территории. Более того, если рассматривать пространственные распределения корреляций только для верификационных периодов, а именно, с 1975 по 1998 и с 1935 по 1955 гг., то они становятся высоко коррелированными с пространственными корреляциями, полученными при пошаговой регрессии.

Все результаты, полученные как при пошаговой множественной регрессии, так и при помощи метода расчета функций отклика являются полностью сопоставимыми. Несмотря на некоторые различия в выделении основных значимых факторов для изменчивости прироста древесных растений в разных секторах севера Евразии, оба метода дали одинаковый результат по расхождению в моделированных и реальных кривых динамики прироста. При этом для обоих калибровочных интервалов получены адекватные модели, которые достаточно хорошо описывают климатически обусловленные изменения радиального прироста деревьев. Отмеченные расхождения между расчетными и реальными кривыми прироста устойчивы: для последних десятилетий модели завышают величины прироста по сравнению с наблюдаемыми (при калибровке на интервалах 1935-1965 или 1935-1975 гг.), а для середины 20-го столетия модели занижают величины прироста по сравнению с реальными (при калибровке на интервалах 1965-1998 или 1955-1998 гг.).

Наиболее интересный результат проведения пространственного анализа сети древесно-кольцевых хронологий – отрицательное (для периода 1965-1998 гг.) и положительное (для периода 1935-1965 гг.) расхождения между модельными и реальными хронологиями. При этом, качество моделирования на обоих интервалах калибровки остается практически одинаковым, т.е. нет

оснований утверждать, что древесные растения уменьшают чувствительность к изменениям климата (Briffa et al., 1998в; 1998г). Относительный вклад температуры летних месяцев (июня и июля) также значим на рассмотренных интервалах. Не прослеживается явно и ускоряющее рост влияние повышения концентрации углекислоты. Если бы такое влияние было, то расхождение между реальными и модельными хронологиями для последних десятилетий 20-го столетия было бы положительным по знаку (реальный прирост превышал бы смоделированный).

В ряде последних работ по Аляске было высказано предположение о переходе температуры из зоны ускорения роста в зону лимитирования (D'Arrigo et. al., 2004; Wilmking et. al., 2004). Однако из проведенного анализа наиболее северных хронологий исследуемой территории такого не следует.

Нет сомнения, что в 1965-1998 гг. на прирост древесных растений на территории севера Евразии начинает влиять дополнительный фактор, который, предположительно, связан с возрастанием концентрации углекислоты в атмосфере. Одна из возможных причин, указывающих на такой фактор – различия в реакции у древесных растений на температуру двух основных процессов, участвующих в продукции стволовой древесины: фотосинтеза и дыхания (Schulze et al., 2005). В области увеличения температур от 12° С до 18° С фотосинтез увеличивается более медленно, чем растет дыхание, которое изменяется по положительной экспоненте. Поэтому, даже увеличение фотосинтеза при увеличении температуры и концентрации CO_2 может не компенсироваться ускорением потерь на дыхание (Schulze et al., 2005).

Тогда для периода 1965-1998 гг. с ускоренным ростом CO_2 и значительным ростом температуры реальный прирост древесных растений замедляется вследствие увеличения потерь на дыхание. Более того, увеличивается длительность вегетации древесных растений за счет теплой осени (Jones, Briffa, 1995). При этом на дыхание еще более увеличиваются (Schulze et al., 2005).

Повсеместное на территории Евразии увеличение зимних температур также повышает вероятность увеличения потерь на дыхание.

Трудно говорить, какую роль могут играть другие факторы, отрицательно влияющие на радиальный прирост древесных растений, такие как озон, УФБ-радиация и др. (Зуев, Бондаренко, 2004; Ваганов и др., 2004; Коняев и др., 2005; Laasko, Huffunen, 1998). Также сложно предположить, чтобы они одинаково работали на столь обширной территории, анализируемой нами.

П.5 Нелинейное оценивание прироста древесных растений. Описание имитационной VS-модели.

Как уже отмечалось выше, причины, которые приводят к проблеме «дивергенции», могут быть гораздо сложнее и сочетать взаимодействие лимитирующих и ускоряющих рост факторов. Безусловно, выявление и понимание механизмов взаимодействия факторов, влияющих на прирост древесных растений, невозможно на базе использования только линейных регрессионных моделей. Вызвано это тем, сами регрессионные модели указывают только на наличие линейных корреляций между исследуемыми явлениями и не вскрывают механизмов взаимодействия. Более того, в ряде работ отмечено, что лимитирующие факторы связаны с приростом древесных растений нелинейно (Ваганов, Шашкин, 2000; Kozlowski et al., 1991; Fritts , Shashkin, 1995; Kozlowski, Pallardy, 1997; Hoch, Korner, 2003; Wilmking et al., 2004; Vaganov, 1990; Vaganov et al., 2006; Touchan et al., 2012).

Эти нелинейные взаимодействия стали базой для разработанной имитационной модели сезонного роста и структуры годичных колец хвойных деревьев (VS-модели), основу которой составили физиологические и биофизические механизмы реакции скорости сезонного роста на некоторые основные факторы среды: температуру, доступную для древесных организмов почвенную влагу и освещенность (Ваганов и др., 1996; Ваганов, Шашкин, 2000; Vaganov et al., 2006). «Данная модель позволяет учесть особенности реакции разных видов древесных растений на климатические переменные и учесть особенности рельефа, почв и других особенностей конкретных местообитаний, которые не только влияют на состав лесных экосистем, но и на динамику

основных факторов: температуры, освещенности, влажности почвы» (Ваганов, Шашкин, 2000).

В базовой VS-модели предполагается, что внешние факторы, в первую очередь, оказывают влияние на зону активного деления клеток (камбиальную зону), в результате которого образуются новые клетки (трахеиды) (Ваганов, 1996). Блок-схема работы данной модели представлена на рисунке 26.

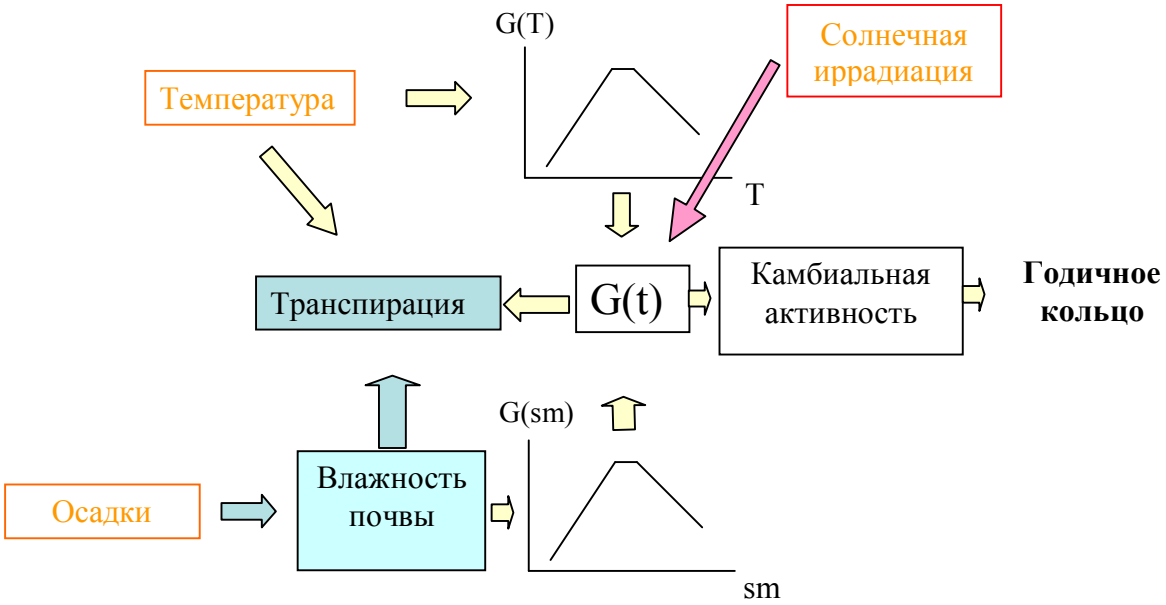


Рис. 26. Блок-схема описания модели Ваганова-Шашкина (VS-модели)

Постулируемая модель имеет двухблочную структуру (Ваганов и др., 1996; Ваганов, Шашкин, 2000; Vaganov et al., 2006; Touchan et al., 2012).

В первом блоке в качестве входных данных выступают три основных фактора, а именно, суточная температура воздуха, суточные осадки и дневная радиация, выходом является скорость роста. При этом соотношение между входными и выходными данными имеет следующий вид:

$$G_t = G_t(E)\min\{G_t(T), G_t(W)\},$$

где G_t - общая скорость роста, которая может изменяться от 0 до 1;

$G_t(E)$ - скорость роста G_t как функция от солнечной радиации E в *t*-день;

$G_t(T)$ - скорость роста G_t от температуры воздуха T в t-день;

$G_t(W)$ - скорость роста G_t от почвенного влагосодержания W в t-день.

Исходя из биологического смысла, функции $G_t(T)$ и $G_t(W)$ должны быть унимодальными, т.е. должны существовать оптимальные значения температуры и влажности почвы, которые максимизируют скорость роста. На основе экспериментальных данных было показано, что наиболее адекватными приближающими функциями являются полиномиальные аппроксимации (Ваганов и др., 1990; Fritts, 1976; Fritts et al., 1991).

Но практическое использование модели показало, что модельные результаты получаются аналогичными, если заменить полиномиальные кривые кусочно-линейной аппроксимацией. В данном случае, кусочно-линейная функция определяется 4 параметрами, а именно: минимумом и максимумом температуры (T1, T4, соответственно), при которых скорость роста равна нулю, и температурами (T2, T3), при которых скорость роста максимальна и равна 1 (Рис. 27 А). Аналогичная ситуация для параметров влажности почвы (W1, W2, W3, W4) (Рис. 27 Б).

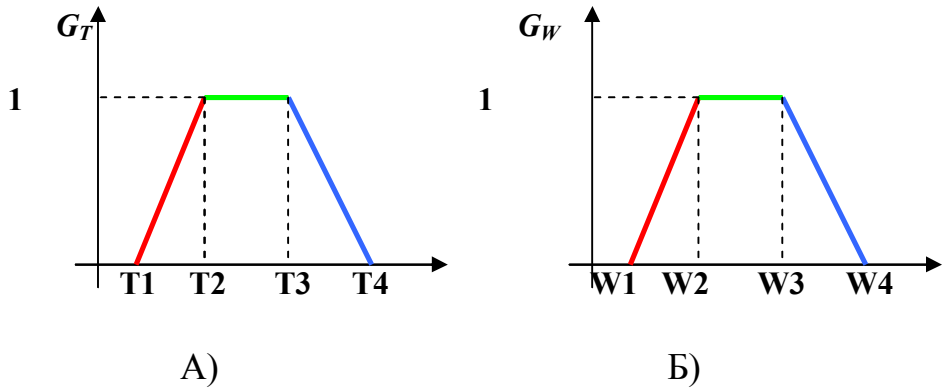


Рис. 27. Типичные кривые роста, используемые в модели.

Основная сложность для оценки G_t связана с корректной оценкой влажности почвы в течение сезона роста на основе ежедневной температуры воздуха и осадков. В данной моделе динамика водной инфильтрации описывается на основе следующего балансового уравнения:

$$\Delta W = f(P) - Tr - q \ \ (\text{or} \ \ W_{t+1} = W_t + f(P_t) - Tr_t - q_t),$$

где ***ΔW*** – приращение водной инфильтрации в почву в единицу времени (в день),

f(P) – количество осадков, попавших в почву,

q – почвенный дренаж,

Tr – испарение.

Известно, что часть осадков задерживается кроной, часть из них удаляется за счет дренажа, а оставшиеся проникают в почву. Обычно, осадки, перехваченные кроной, оцениваются при помощи линейных функций (Ваганов, Шашкин, 2000).

В VS-модели функция ***f(P)*** описывается при помощи следующего уравнения:

$$f(P) = \begin{cases} k_1 P, \ \text{if} \ \ k_1 P \le P_{\max} \\ P_{\max}, \ \text{if} \ \ k_1 P > P_{\max} \end{cases}$$

где $\boldsymbol{P_{max}}$ – максимальная скорость инфильтрации влаги в почву (мм/день),

$\boldsymbol{k_1}$ – модельный параметр.

Дренаж воды из почвы пропорционален водной инфильтрации, а именно:

$$q = rd \circ W,$$

где ***rd*** – скорость водного дренажа из воды.

Водное испарение кроной дерева является результатом разности между водной концентрацией в листьях и окружающим их воздухом, которое описывается при помощи обычного уравнения диффузии. Но существует ряд проблем. Оценка испарения основана на известном уравнении Монтейса и различных его модификациях. Такая оценка требует большого числа точных характеристик энергетического баланса и корректной устичной проводимости. Это невозможно сделать на основе постулируемой модели, так как она направлена на использование инструментальных климатических переменных

(температуры и осадков) как входных параметров. Поэтому в модели используется более простое уравнения для транспирации ***Tr***, а именно:

$$Tr = k_2 Gr(t) e^{k_3 T},$$

где k_2, k_3 – постоянные коэффициенты, ***T*** – температура.

Последнее уравнение соответствует нелинейному увеличению дефицита влажности от роста температуры (Ваганов, Шашкин, 2000).

Для северного предела распространения лесов растительность может использовать только влагу из оттаявшей почвы. В этом случае, скорость почвенного оттаивания *l* пропорциональна температуре воздуха и экспоненциально уменьшается с ростом глубины оттаивания:

$$l_{t+1} = l_t + a_1 T_t e^{-a_2 l_t},$$

где a_1, a_2 – коэффициенты почвенного оттаивания.

Следовательно, $g_W(t)$ можно определить следующим образом:

$$g_W(t) = \frac{l_t}{lr} g_W(t),$$

где *lr* –глубина корней.

Такие оценки оттаивания почвы хорошо разработаны для длительных естественных наблюдений (Поздняков, 1986).

Для других регионов, преполагается, что *l=lr*.

Солнечная иррадиация E оценивается на основе следующего уравнения (Gates, 1980):

$$E = E_0 (h_s \sin\varphi \sin\theta + \cos\theta \sin h_s),$$

где φ - широта метеорологической станции.

θ - угол высоты солнцестояния.

Следовательно, $g_E(t) = E / E_{\max}$.

Известно, если сумма положительных температур превосходит пороговое значение для фиксированного интервала времени, то начинается рост деревьев (Ваганов, Шашкин, 2000; Vaganov et al., 2006; Briffa et al., 2008). В VS-модели

этот факт соответствует параметру *Tg* (пороговое значение температуры для 10 дней).

Все описанные параметры могут быть изменены в компьютерной реализации VS-модели (Ваганов и др., 1996).

Заметим, что интегральная ***Gr(t)*** является входными данными для следующего блока модели.

Во втором блоке рассчитывается продукция клеток за сезон и изменение размеров клеток внутри годичного кольца (входной параметр - «интегральная» скорость роста G_t), при этом, конечно же, всегда может быть вычислена ширина годичного кольца древесного организма. Суть алгоритма расчета сводится к следующему: при увеличении G_t число делящихся клеток в камбии и скорость их деления увеличивается, при уменьшении G_t-наоборот. При этом, если G_t падает ниже некоторого порога, деление клеток в камбии прекращается, рост годичного кольца завершается.

Работа данной модели была апробирована при описании сезонного роста и структуры годичных колец деревьев, произрастающих в полузасушливых и засушливых местообитаниях (Ваганов и др., 1990; Ваганов, Пак, 1995; Ваганов, 1996; Ваганов, Шашкин, 2000; Шашкин, Ваганов, 1993; Fritts , Shashkin, 1995).

В последнее время в описанную выше модель были внесены существенные дополнения, связанные с моделированием системы фотосинтеза древесных организмов (Fritts , Shashkin, 1995; Vaganov et al., 2006).

Отметим, что VS-модель является сложным инструментом, требующим значительного количества параметров модели, которые должны быть адекватно оценены для каждого местообитания древесных растений. Для оптимизации оценки параметров модели был разработан визуальный подход к процессу параметризации модели (VS-осциллограф), который позволяет моделировать рост годичных колец и может легко использоваться исследователями и студентами (Shishov et al., 2015)

Возможно выделить следующие направления использования данной модели (Ваганов, Шашкин, 2000; Vaganov et al., 2006):

- для объяснения корреляций, возникающих между приростом древесных растений и ведущими климатическими факторами;
- для формального описания механизмов влияния климатических факторов на сезонный рост и структуру годичных колец в конкретных условиях местообитания с учетом их микро- и макро-условий ;
- для выявления механизмов влияния случайного специфического фактора (например, пожара, вспышки массового размножения насекомых или точечного антропогенного загрязнения окружающей среды) на сезонный рост древесных растений;.

Введение в данную модель фотосинтетического блока, с большой вероятностью, позволит выделить и проанализировать географические особенности реакции роста древесных растений на повышение концентрации углекислоты в атмосфере на фоне текущих изменений климата и выявить условия произрастания древесных растений там, где это влияние становится значимым в сравнении с влиянием изменений климата.

Это позволит ответить на важнейший вопрос о реакции лесных экосистем бореальной зоны на ожидаемое потепление. «Модель, основанная на эко-физиологических механизмах, рассчитывает не только изменчивость ширины годичных колец, но и показатели структуры (размеры клеток, плотность древесины), то реальным представляется ее применение для оценки прироста древесной массы стволов. Такие данные являются выходом на количественные оценки реального аккумулирования углерода в биомассе стволовой древесины» (Ваганов, Шашкин, 2000).

Отметим, что нелинейные взаимодействия между приростом деревьев и внешними факторами могут быть также учтены при помощи преобразований дендрохронологических характеристик на основе методов непараметрической статистики и пространственно-временных экстраполяций этих данных на базе методов теории классификации, позволяющих выделять общие особенности прироста древесных растений на субрегиональном и региональном уровнях.

П.6 Скользящие ранговые коэффициенты Спирмена, Кендалла, Хубера

Традиционные в статистике случайных функций или временных рядов характеристики, используемые в анализе дендрохронологических рядов, а именно: первые моменты, авто- и взаимно-корреляционные функции, функции спектральных плотностей,– могут оказаться излишне чувствительными, то есть показать изменения математических ожиданий и дисперсий, наличие корреляционных связей и цикличностей там, где их не могло быть по причине, например, воздействия климатических факторов, и наоборот, вследствие воздействия колорированного ”шума”, не обнаружить изменения, связи и цикличности, обусловленные влиянием факторов, носящих глобальный характер (Джансеитов, Шишов, 1993).

Характеристики случайных функций, учитывающие вместо абсолютных величин прироста отношение “больше–меньше” (“лучше–хуже”), а именно: коэффициента ранговой корреляции Спирмена, коэффициента ранговой корреляции Кендалла, коэффициента конкордации или согласованности Кендалла, коэффициента синхронности Хубера, - могут оказаться более эффективными как при изучении динамики древесно-кольцевых хронологий (дендрохронологических рядов), так и при выявлении связей между радиальным приростом древесных растений и динамикой климатических факторов (Джансеитов, Шишов, 1993).

Итак, пусть даны m временных рядов длины n:

$$
\begin{array}{ccccccccc}
x_1(t_0-n+1), & \ldots, & x_1(t-k), & \cdots, & x_1(t), & \ldots, & x_1(t+k), & \ldots, & x_1(t_0), \\
\ldots & \ldots & & \ldots & & \ldots & & \ldots & \ldots \\
x_i(t_0-n+1), & \ldots, & x_i(t-k), & \ldots, & x_i(t), & \ldots, & x_i(t+k), & \ldots, & x_i(t_0), \\
\ldots & \ldots & & \ldots & & \ldots & & \ldots & \ldots \\
x_j(t_0-n+1), & \ldots, & x_j(t-k), & \ldots, & x_j(t), & \ldots, & x_j(t+k), & \ldots, & x_j(t_0), \\
\ldots & \ldots & & \ldots & & \ldots & & \ldots & \ldots \\
x_m(t_0-n+1), & \ldots, & x_m(t-k), & \ldots, & x_m(t), & \ldots, & x_m(t+k), & \ldots, & x_m(t_0).
\end{array}
$$

Скользящий коэффициент Пирсона - это мера линейной зависимости двух временных рядов i и j в момент времени t с шириной окна W=2k+1, оценка которой вычисляется по следующей формуле:

$$r_{ij}(t,W)=\frac{\sum_{p=t-k}^{t+k}(x_i(p)-\bar{x}_i(t))(x_j(p)-\bar{x}_j(t))}{W\sigma_{x_i}^t\sigma_{x_j}^t},$$

где $\bar{x}_i(t)$ - оценка математического ожидания (средняя) ВР x_i в момент времени t,

$\sigma_{x_i}^t$ - средне-квадратичное отклонение ВР x_i в момент времени t,

$\bar{x}_j(t)$ - оценка математического ожидания (средняя) ВР x_j в момент времени t,

$\sigma_{x_j}^t$ - средне-квадратичное отклонение ВР x_j в момент времени t.

При использовании этого коэффициента предполагается, что значения исследуемых временных рядов имеют нормальный закон распределения (это параметрическая статистика, требующая от данных нормального закона распределения).

Наряду с применением в регрессионном анализе, коэффициент Пирсона может быть использован как мера для построения корреляционных плеяд. Суть метода корреляционных плеяд заключается в том, что при определенном выборе порогового значения коэффициента корреляция и определенного «корреляционного» шага выявляется степень линейной взаимосвязи между факторами различной природы, которые представляли собой временные ряды длины, например, W.

Скользящий коэффициент ранговой корреляции Спирмена – мера зависимости двух временных рядов i и j в момент времени t с шириной окна W=2k+1 (Хеттманспергер, 1987). Если ранги i-ряда рассмотрены в естественном порядке Q_p=1,...,W, а R_p – ранг j-ряда, соответствующей той паре ($x_i(p)$, $x_j(p)$), для которой ранг i-ряда равен Q_p, то

$$S_{ij}(t,W)=1-\frac{6S}{W^3-W}\text{ , где } S=\sum_{p=t-k}^{t+k}(R_p-Q_p)^2\text{ ,}$$

или в общем случае, когда ранги *i*-ряда расположены в некотором порядке $R^i_{t-k},\ldots,R^i_{t+k}$, а R^j_p – ранг *j*-ряда, соответствующей той паре ($x_i(p)$, $x_j(p)$), для которой ранг $x_i(p)$ равен R^i_p, то

$$S_{ij}(t,W)=1-\frac{6S}{W^3-W}\ ,\ \text{где}\ S=\sum_{p=t-k}^{t+k}(R^i_p-R^j_p)^2\ .$$

S –коэффициент (Коэффициент Спирмена) изменяется в пределах от [-1, 1].

Если поведение двух временных рядов (например, ширины годичных колец двух различных деревьев в окне *W*) независимо, то $M(S)=0, D(S)=\frac{1}{W-1}$. Таким образом по величине отклонения *S*-коэффициента от нуля можно сделать вывод о независимости или зависимости поведения двух временных рядов в момент *t*.

Для построения соответствующего критерия вычисляется распределение *S*-коэффициентов для независимых признаков. При $4\le W\le 10$ используют специальную таблицу распределения (Таб. 1) (Большев, Смирнов, 1983).

Таблица 1.

Распределение коэффициента Спирмена для независимых признаков при $4\le W\le 10$

W=5		W=7		W=9	
s	Q_s	s	Q_s	s	Q_s
22	0.475	74	0.249	156	0.218
24	0.392	78	0.198	164	0.168
26	0.342	82	0.151	172	0.125
28	0.258	86	0.118	180	0.089
30	0.225	90	0.083	188	0.060
32	0.175	94	0.055	196	0.038
34	0.117	98	0.033	204	0.022
36	0.067	102	0.017	212	0.011
38	0.042	106	0.006	220	0.004
40	0.008	110	0.001	228	0.001

40 112 240

В таблице 1 даны значения вероятностей $Q_s(s;\ W) = P\{S \geq s\}$ для W=5,7,9. Так как распределение суммы S симметрично относительно $(W^3\text{-}W)/6$ и сосредоточено на отрезке $0 \leq s \leq (W^3 - W)/3$, то значения вероятностей $P\{S \leq s\}$ можно вычислить по формуле: $P\{S \leq s\} = Q_s((W^3 - W)/3 - s; W)$.

Величины $(W^3\text{-}W)/3$ указаны в последней строке таблицы.

Пусть, например, при W=9 нужно вычислить нижнее и верхнее значение для S-коэффициента, соответствующее номинальным уровням значимости 0.05. По таблице находим, что $Q_s(196,9)$=0.038 и $Q_s(188,9)$=0.060, поэтому 196 - верхнее критическое значение для S и по формуле S =1-2•196/240=-0.633 - нижнее критическое значение для S-коэффициента. Так как коэффициент Спирмена распределен симметрично относительно 0, то R=0.633 - верхнее критическое значение. При этом истинные уровни значимости равны 0.038, следовательно, критической области $|S(t,W)| \geq 0.633$ при W=9 соответствует вероятность ошибки первого рода 0.076.

При W>10 можно воспользоваться, например, тем, что случайная величина $\sqrt{W-1} \cdot S$ при $W \to \infty$ распределена асимптотически нормально с (0,1). В последнем случае гипотеза независимости отвергается, если $|S| > u_1 - \frac{\alpha}{2} \Big/ \sqrt{W-1}$, где $u_1 - \frac{\alpha}{2}$ есть корень уравнения $\Phi(u) = 1 - \frac{\alpha}{2}$ ($\Phi(u)$ – функция стандартного нормального распределения).

В предположении, что временные ряды имеют совместное нормальное распределение в окне W, можно установить связь S-коэффициента с коэффициентом корреляции Пирсона r, при достаточно больших W, $M(S) \approx \frac{6}{\pi} \arcsin \frac{r}{2}$ следовательно величину $2 \sin \frac{\pi}{6} S$ можно использовать в качестве оценки для r.

Коэффициент ранговой корреляции Спирмена был назван по имени Ч. Спирмена, который использовал его в исследованиях по психологии вместо обычного коэффициента корреляции.

Для вычисления скользящего коэффициента ранговой корреляции Кендалла (K –коэффициента) в момент t с шириной “окна” $W = 2k + 1$ между двумя временными рядами (i -ым и j -ым) необходимо вычислить предварительно (Kendall, 1970; Briffa et al., 2008):

$$R\left(x_i(p), x_i(q)\right) = \begin{cases} +1, & если \quad x_i(q) - x_i(p) > 0, \\ 0, & если \quad x_i(q) - x_i(p) = 0, \\ -1, & если \quad x_i(q) - x_i(p) < 0, \end{cases}$$

$$S_{ij} = \sum_{\substack{p,q=t-k \\ (p<q)}}^{t+k} R(x_i(p), x_i(q)) R(x_j(p), x_j(q)),$$

$$T_i = \sum_{\substack{p,q=t-k \\ (p<q)}}^{t+k} \left(1 - \left|R\left(x_i(p), x_i(q)\right)\right|\right), \quad T_j = \sum_{\substack{p,q=t-k \\ (p<q)}}^{t+k} \left(1 - \left|R\left(x_j(p), x_j(q)\right)\right|\right),$$

и, наконец, сам K –коэффициент

$$K_{ij}(t, W) = \frac{S_{ij}}{\sqrt{0{,}5 \cdot W(W-1) - T_i} \cdot \sqrt{0{,}5 \cdot W(W-1) - T_j}}$$

K –коэффициент (Коэффициент Кендалла) изменяется в пределах от [-1, 1].

Коэффициент ранговой корреляции Кендалла изменяется в пределах от -1 до 1. Коэффициент ранговой корреляции Кендалла применяется для проверки независимости случайных величин, например, несогласованности в приростах 2-х древесно-кольцевых хронологий (Рис. 28).

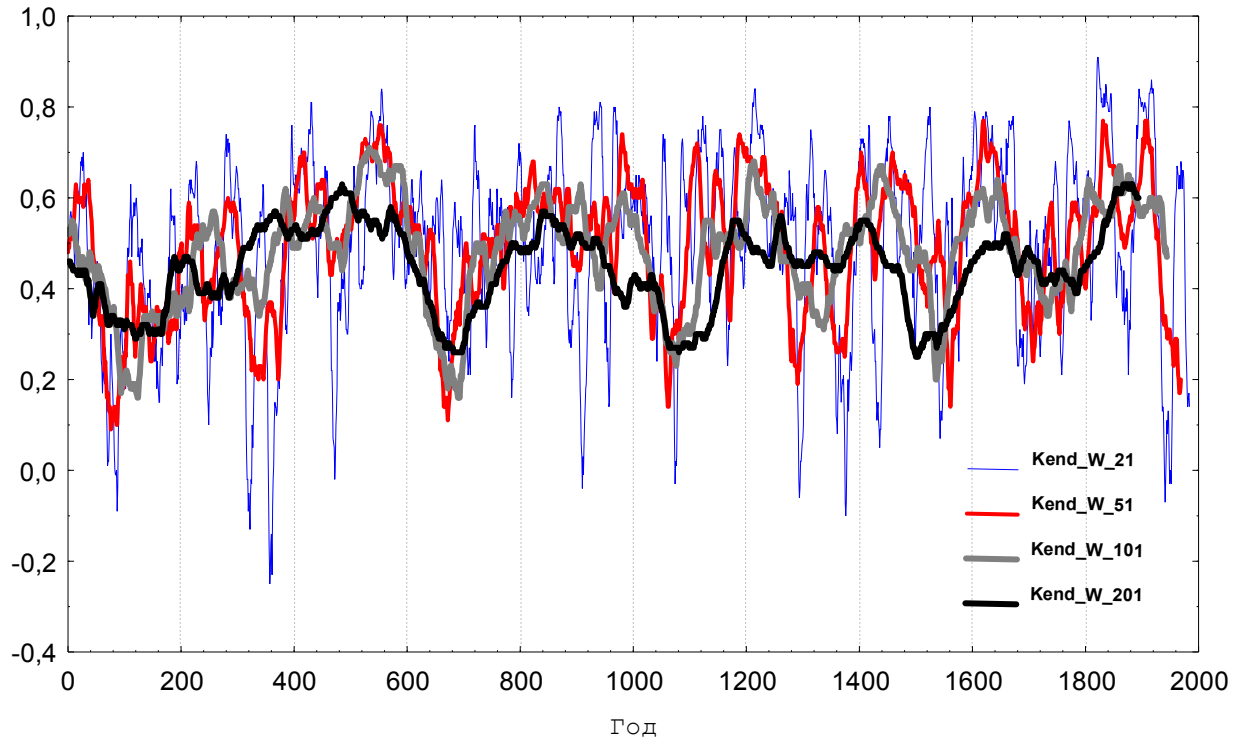


Рис. 28. Пример расчета К-коэффициента с различными окнами W(21, 51, 101 и 201 год), полученных для сверхдлительных древесно-кольцевых RCS хронологий по территории Швеции и Финляндии

Если гипотеза независимости верна, то $M(K)=0$, $D(K)=\dfrac{2(2W+5)}{9W(W-1)}$. При небольшом объеме выборки ($4 \le W \le 10$) проверка статистической гипотезы независимости производится с помощью специальной таблицы (Табл. 2) (Большев, Смирнов, 1983). Эта таблица распределения приведена для нечетных W. В таблице даны вероятности $P\{S_{ij} \ge s\} = Q_K(s,W)$.

При $W>10$ пользуются нормальным приближением для распределения K: если $|K| > u_{\alpha 12}\sqrt{2(2W+5)/9W(W-1)}$, то гипотеза о независимости отвергается, в противоположном случае принимается. Здесь α – уровень значимости, $u_{\alpha 12}$ есть $100 \cdot \dfrac{\alpha}{2}$ – процентная точка нормального распределения (Большев, Смирнов, 1983).

Критерии, основанные на коэффициенте ранговой корреляции Спирмена и на коэффициенте ранговой корреляции Кендалла асимптотически эквивалентны (при W=2 соответствующие ранговые статистики совпадают).

Таблица 2.

Распределения K-коэффициентов при $4 \le W \le 10$

s	W		s	W
	5	9		7
0	0.592	0.540	1	0.500
2	0.408	0.460	3	0.386
4	0.242	0.381	5	0.281
6	0.117	0.306	7	0.191
8	0.042	0.238	9	0.119
10	0.008	0.179	11	0.068
12		0.130	13	0.035
14		0.090	15	0.015
16		0.060	17	0.005
18		0.038	19	0.001
20		0.022	21	0.0002
22		0.012	23	
24		0.006	25	
26		0.002	27	
28		0.001	29	
30		0.0004	31	

Для вычисления скользящего коэффициента синхронности Хубера (H –коэффициента) в момент t с шириной “окна” $W = 2k + 1$ между двумя временными рядами (i-ым и j-ым) необходимо определить для всех $p = t - k, t - k + 1, \ldots, t, \ldots, t + k$

$$r_i(p) = R\big(x_i(p-1), x_i(p)\big) = \begin{cases} +1, \text{ если } x_i(p) - x_i(p-1) > 0, \\ 0, \text{ если } x_i(p) - x_i(p-1) = 0, \\ -1, \text{ если } x_i(p) - x_i(p-1) < 0, \end{cases}$$

и $r_j(p)$, аналогичным образом, а затем, подсчитать число совпадений значений $r_i(p)$ и $r_j(p)$ при $p = t-k,\ldots,t+k$, в случае же равенства одного из сравниваемых значений 0, пару $r_i(p)$, $r_j(p)$ считать совпавшей только при условии совпадении предшествующей пары, а затем полученное общее число совпадений, нормированное к длине окна W будет соответствовать H –коэффициенту $H_{ij}(t,W)$.

Найдем закон распределения этой ранговой статистики (Хеттманспергер, 1987). Будем считать, что два временных ряда $x_i(t)$ и $x_j(t)$ зависимы в момент времени t', если $H_{ij}(t',W) \geq \frac{k}{W}$, где $H_{ij}(t',W) = \sum_{p=t'-k}^{t'+k} \frac{h_p}{W}$, $h_p = \begin{cases} 1, если\ r_i(p) = r_j(p) \\ 0, если\ r_i(p) \neq r_j(p) \end{cases}$ и W – временной интервал, то есть правило таково: отклонить гипотезу H_0: $\Theta = 0$ в пользу альтернативной гипотезы H_A: $\Theta > 0$,если $H \geq \frac{k}{W}$. Само критическое значение $\frac{k}{W}$ определяется таким образом, что $P_{H_0}\left(H \geq \frac{k}{W}\right) = \alpha$, где α – уровень значимости критерия. Ясно, что мы сначала должны найти распределение $H_{ij}(t',W)$ в случае H_0.

При H_0: $\Theta = 0$ $h_{t'-k},\ldots,h_{t'},\ldots,h_{t'+k}$ независимые, одинаково распределенные случайные величины, причем каждая из них – биноминальная с параметрами 1 и $p = \frac{1}{2}$ и обозначается $B(1,1/2)$. Тогда H есть сумма W независимых, одинаково распределенных, как $B(1,1/2)$, случайных величин и потому имеет распределение $B(W,1/2)$. Критическое значение k можно найти по таблице биноминального распределения (Большев, Смирнов, 1983).

Заметим, что распределение H-коэффициента не зависит от распределения $h_{t'-k},\ldots,h_{t'},\ldots,h_{t'+k}$ и, следовательно, критическое значение k можно найти, не зная распределение F элементов h_i. С другой стороны, при выполнении H_A: $\Theta = \Theta' > 0$ случайная величина H имеет $B(W,p)$-распределение, но $p=P(h>0)=1-F(-\Theta')$ зависит от F.

Так как при обеих гипотезах H_{ij} - сумма независимых, одинаково распределенных величин $B(1, p)$ с $D(h_i)<\infty$, то величина

$$\frac{H - M[H]}{\sqrt{D[H]}}$$

имеет приближенное стандартное нормальное распределение с параметрами (0, 1) (согласно центральной предельной теореме), т.е.

$$\frac{H - M[H]}{\sqrt{D[H]}} \xrightarrow{F} Z \approx n(0,1).$$

Тогда при H_0: Θ=0 критическое значение k можно удовлетворительно приблизить по формуле:

$$k = W/2 + Z_\alpha \sqrt{W/2},$$

где Z_α - верхняя α -процентная точка стандартного нормального распределения, т.е. $\alpha = 1 - \Phi(Z_\alpha)$.

Коэффициент Хубера или критерий знаков был впервые применен Арбетнотом в 1710 году при изучении вопросов, связанных с рождаемостью людей. Как видно, критерий знаков довольно прост, но на его базе и на алгоритме его вычисления позднее были разработаны некоторые другие критерии (Хеттманспергер, 1987).

Для вычисления скользящего коэффициента конкордации Кендалла (Кендалл, 1975; Джансеитов, Шишов, 1993; Шишов, 1998; Kendall, 1970; Briffa et al., 2008) (C –коэффициента) в момент t с шириной “окна” $W = 2k + 1$ между всеми m рядами необходимо проранжировать все сегменты временных рядов, используя метод осреднения совпавших рангов, подсчитать суммы рангов для каждого момента $p = t - k, \ldots, t, \ldots, t + k$:

$$\begin{array}{ccccccc}
n_1(t-k), & \ldots, & n_1(p), & \ldots, & n_1(t), & \ldots, & n_1(t+k), \\
\ldots & \ldots & & \ldots & & \ldots & \ldots \\
n_i(t-k), & \ldots, & n_i(p), & \ldots, & n_i(t), & \ldots, & n_i(t+k), \\
\ldots & \ldots & & \ldots & & \ldots & \ldots \\
n_m(t-k), & \ldots, & n_m(p), & \ldots, & n_m(t), & \ldots, & n_m(t+k), \\
\hline
n(t-k), & \ldots, & n(p), & \ldots, & n(t), & \ldots, & n(t+k),
\end{array}$$

где $n(p)=\sum_{i=1}^{m} n_i(p)$ (для любого $p=t-k,\ldots, t+k$).

Далее, необходимо подсчитать S - сумму квадраты отклонений сумм рангов от средней суммы рангов, то есть

$$S = \sum_{p=t-k}^{t+k}\left(n(p)-0{,}5\cdot m(W+1)\right)^2 .$$

Окончательно, C –коэффициент вычисляется по следующей формуле:

$$c_t = C(t,W) = \frac{12S}{m^2\left(W^3-W\right)-m\sum_{T'} T'},$$

где $\sum_t T' = \sum_t \left(t^3-t\right)$ - поправка на одинаковые ранги, которые могут быть во временных рядах на анализируемом сегменте времени [$t-k$, $t+k$].

Заметим, что с увеличением окна скольжения, и, исходя из природы многих естественных временных рядов (когда вероятность появления одинаковых значений достаточно мала), влияние такой поправки на окончательный результат становиться несущественным. Поэтому уже с окном W в 7 и более лет, эту поправкой можно пренебречь.

Скользящие значения коэффициента конкордации Кендалла не рассчитываются для первых и последних k значений исследуемых временных рядов.

Коэффициент конкордации C изменяется в пределах от 0 до 1. Ноль означает полную несогласованность в поведении временных рядов на рассматриваемом временном интервале (Рис. 29). Соответственно, единица - полное совпадение рангов (фактически, динамики) анализируемых временных рядов.

Если нулевую гипотезу H_0 отсутствия конкордации, например, в росте среди деревьев интерпретировать как результат того, что деревья растут случайным образом и их ранги в окне, следовательно, назначаются случайно. Если гипотеза H_0 верна, то $M(C)=1/m, D(C)=\frac{2(m-1)}{m(W-1)}$.

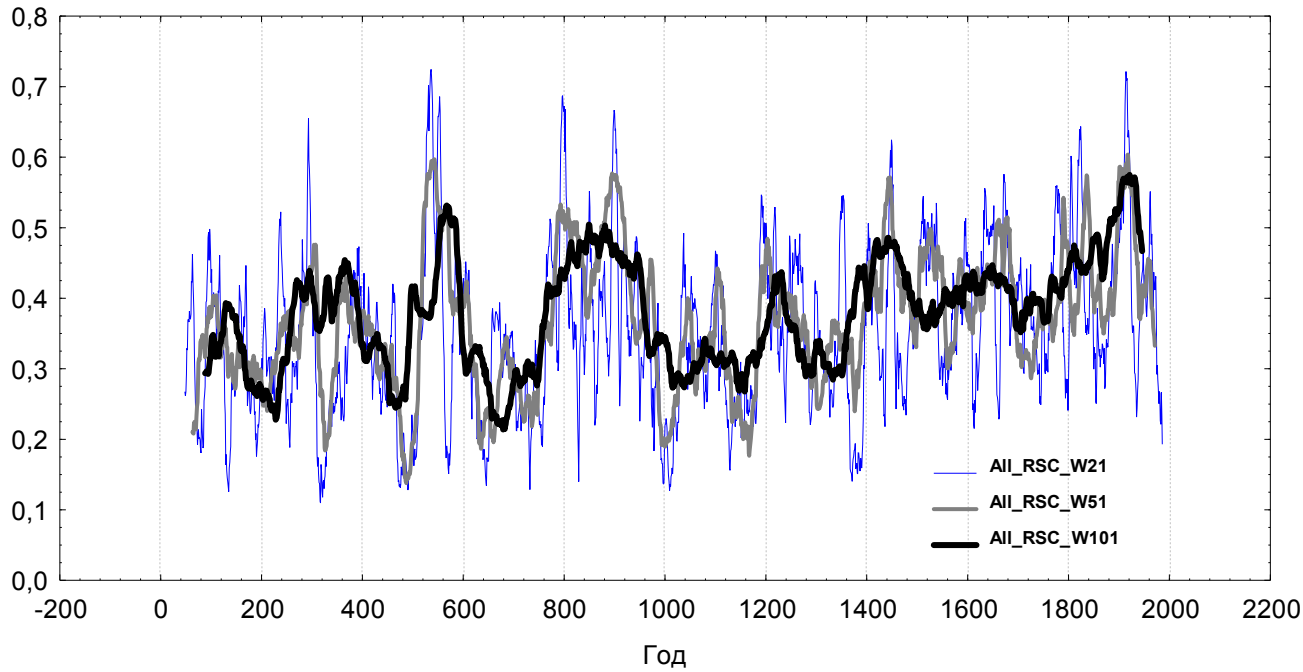


Рис. 29. Пример расчета С-коэффициента с различными окнами W(21, 51, 101 год), полученных для 4-х сверхдлительных древесно-кольцевых RCS хронологий, построенных по территории Швеции, Финляндии, Ямала и Таймыра.

При небольшом объеме выборки ($3 \le W \le 10$ и $3 \le m \le 10$) проверка статистической гипотезы согласованности производится с помощью таблицы точного распределения коэффициента согласованности C (Большев, Смирнов, 1983).

В непредельных случаях можно воспользоваться следующим результатом. Известно, что при длине окна W>7 величина $\chi_r^2 = m(W-1)C$ распределена как статистика χ^2 с W-1 степенями свободы (Kendall, 1970).

Иногда можно использовать аналог коэффициента Хубера для групп временных рядов – групповой коэффициент синхронности (GH –коэффициент) (Джансеитов, Шишов, 1993; Шишов, 1998).

Для вычисления скользящего обобщенного коэффициента согласованности (GH –коэффициента) в момент t с шириной “окна” $W = 2k+1$ между всеми m рядами (1) необходимо, прежде всего, определить $r_i(p)$ по описанной выше

формуле для $i = 1,2,\ldots,m$, и $p = t-k,\ldots,t,\ldots,t+k$, а затем вычислить $r(p)$ по схеме:

$$\begin{array}{ccccccc}
r_1(t-k), & \ldots, & r_1(p), & \ldots, & r_1(t), & \ldots, & r_1(t+k), \\
\ldots & \ldots & & \ldots & & \ldots & \ldots \\
r_i(t-k), & \ldots, & r_i(p), & \ldots, & r_i(t), & \ldots, & r_i(t+k), \\
\ldots & \ldots & & \ldots & & \ldots & \ldots \\
r_m(t-k), & \ldots, & r_m(p), & \ldots, & r_m(t), & \ldots, & r_m(t+k), \\
\hline
r(t-k), & \ldots, & r(p), & \ldots, & r(t), & \ldots, & r(t+k),
\end{array}$$

–где в качестве $r(p)$ берется доля общих значений большинства совпавших $r_i(p)$ ($i = 1,2,\ldots,m$).

Значения скользящего GH –коэффициента вычисляются по формуле:

$$GH(t,W) = \frac{\sum_{p=t-k}^{t+k} r(p)}{W}$$

Все перечисленные выше $S-$, $K-$, $H-$, $C-$, GH –коэффициенты, будучи примененные к двум и более ДХР, характеризуют степень согласованности в поведении этих рядов ($S-$, $K-$ и $H-$ коэффициенты применяются к двум рядам, $C-$и $GH-$ коэффициенты – к большему числу рядов, причем $H-$ и GH –коэффициенты отличаются от $S-$, $K-$ и $C-$ коэффициентов еще большей “загрубленностью”, избавленностью от упоминавшихся выше “шумовых” эффектов абсолютных значений ДХР).

Примененные же к паре временных рядов, один из которых дендрохронологический, а другой – хронологический (ряд последовательных значений моментов времени), коэффициенты Спирмена, Кендалла и Хубера характеризуют улучшение или ухудшение условий для радиального прироста исследуемого дерева на рассматриваемом интервале времени, равном, например, ширине окна $W = 2k+1$ (Джансеитов, Шишов, 1993; Шишов, 1998).

Рассмотрим алгоритм получения средних скользящих частот выбросов для участка (Шишов, 1998).

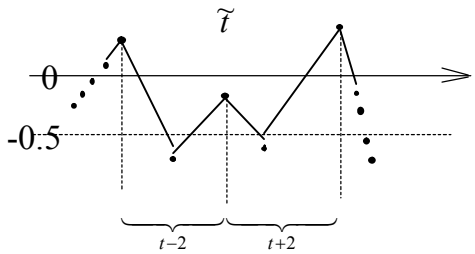
Исходные ДХР для отдельных деревьев переводились в индексные ряды (центрировались и нормировались) по формуле:

$$I_t = \frac{x(t) - \overline{x}}{\sigma_x},$$

где $\overline{x} = \frac{\sum_{t=1}^{n} x(t)}{n}$ – оценка математического ожидания исходного временного ряда,

$\sigma_x = \sqrt{\frac{\sum_{t=1}^{n} \left(x(t) - \overline{x}\right)^2}{n-1}}$ – оценка среднеквадратического отклонения исходного временного ряда, n – длина временного ряда X.

Полученные индексные ряды I отдельных деревьев переводились в так называемые, скользящие частоты выбросов вниз (депрессий роста), так как именно эта характеристика является одной из наиболее информативных при решении основной проблемы – выявление глобальных изменений регулярного характера, с порогом – 0.5σ (в нашем случае $\sigma = 1$) и окном в 21 год по следующему правилу:



(пример приводится для окна W=5 лет, для упрощения) количество выбросов в окне W=5 равно 2, следовательно:

$$d_{\tilde{t}} = \frac{2}{5} = 0.4,$$

где $d_{\tilde{t}}$ – доля (частота) выбросов в $\tilde{t}$ –ый год.

Окончательно, средней частот выбросов вниз является характеристика D, вычисленная по следующей формуле:

$$D_{\tilde{t}} = \frac{\sum_{k=1}^{m} d_{\tilde{t}}^{k}}{m},$$

где m – количество деревьев на участке.

Из всего выше сказанного следует, что рассматриваемую характеристику можно считать знаковой статистикой (Хеттманспергер, 1987), которая лежит в интервале [0, 1] и имеет биномиальный закон распределения.

П.7 Пример использования скользящих коэффициентов осредненной корреляции Пирсона и конкордации Кендалла. Сравнительный анализ.

В этом разделе анализируются скользящие характеристики синхронности временных рядов с точки зрения их устойчивости к шумовым воздействиям различного рода и их ассоциации с внешним периодическим воздействием (Шишов, Ивановский, 2006).

В работе были использованы скользящий коэффициент конкордации Кендалла и средний скользящий коэффициент корреляции Пирсона.

Рассматриваемые характеристики применяются для расчета согласованности в поведении любой совокупности m временных рядов длины N.

Алгоритм расчета скользящего коэффициента конкордации аналогичен алгоритму расчета, предложенного Кендаллом (Kendall, 1970) за исключением следующего. Скользящий коэффициент конкордации C рассчитывается не для всего периода наблюдений N, а для каждого момента времени t, который является центром скользящего окна с длиной $W<N$.

Коэффициент конкордации C изменяется в пределах от 0 до 1 . Ноль означает полную несогласованность в поведении временных рядов на рассматриваемом временном интервале. Соответственно, единица – полная синхронность динамики анализируемых временных рядов.

Для расчета скользящего группового (осредненного) коэффициента корреляции Пирсона, предварительно, необходимо рассчитать скользящие коэффициенты корреляции между всеми парами временных рядов i и j ($\forall$ i ,j = 1..m and (i>j)) в момент времени t с шириной окна $W=2k+1$, которые вычисляются по следующей формуле:

$$r_{ij}(t,W)=\frac{\sum_{p=t-k}^{t+k}(x_i(p)-\bar{x}_i(t))(x_j(p)-\bar{x}_j(t))}{(W-1)\sigma_{x_i}^t\sigma_{x_j}^t},$$

где $\bar{x}_i(t)$ - оценка математического ожидания (средняя) ВР x_i в момент времени t,

$\sigma_{x_i}^t$ - средне-квадратичное отклонение ВР x_i в момент времени t,

$\bar{x}_j(t)$ - оценка математического ожидания (средняя) ВР x_j в момент времени t,

$\sigma_{x_j}^t$ - средне-квадратичное отклонение ВР x_j в момент времени t.

И, окончательно,

$$R(t,W)=\sum_i\sum_j r_{ij}(t,W)\Big/\sum_{f=1}^{m-1}f.$$

Отметим также, что значения скользящих коэффициентов вычислялось только для «полных» окон. Следовательно, скользящие коэффициенты не рассчитывались для начальных (W-1)/2 и конечных (W-1)/2 значений анализируемых временных рядов.

В результате внешнего воздействия или шума на сигнал, появляются отклонения от 1 в рассматриваемых значениях скользящих характеристик. Данное утверждение верно как для коэффициентов конкордации, так и корреляции. Следовательно, чувствительность к шуму той или иной характеристики можно оценить на основании отклонений от 1.

Такой оценкой может быть обычное Евклидово расстояние D между временным рядом, все значения которого постоянны и равны 1 и соответствующим скользящим коэффициентом, т.е.

$$D=\sqrt{\sum_t(1-c_t)^2},$$

где c_t – скользящий коэффициент конкордации или корреляции.

Если бы поведение рассматриваемых рядов было бы абсолютно несогласованным (или случайным), то все c_t=0 и, следовательно, значение D было бы равно квадратному корню от длины исследуемого ряда (В нашем

случае, в зависимости от окна W критические значения D колебались бы в пределах от $38.08=\sqrt{1450}$ до $39.28=\sqrt{1543}$). Наоборот, в случае полной согласованности исследуемых временных рядов, значение D стремилось бы к 0. Таким образом, на основании Евклидова расстояния можно ответить на несколько вопросов. Во-первых, является ли временная согласованность исследуемых рядов случайной или, наоборот, нет. Во-вторых, какая из скользящих характеристик менее подвержена влиянию шума, то есть является более робастной по отношению к шумовым воздействиям различного рода. В-третьих, какая из скользящих характеристик более чувствительна к наличию сигнала. В последнем случае отметим, что это чувствительность будет определяться, прежде всего, величиной окна скольжения W. Очевидно, чем больше величина W, тем больше информации используется для получения той или иной характеристики, и тем больше вероятность того, что скользящая характеристика «поймает» информацию об общем сигнале.

Дополнительным критерием для сравнения изучаемых скользящих характеристик может служить средний процент значимых значений. Алгоритм получения этого среднего процента следующий. При определенных значениях окна скольжения W и количества анализируемых рядов m всегда можно найти критическое значение (при соответствующем уровне значимости p) коэффициента конкордации (или корреляции) на основании известных распределений для соответствующих статистик. Следовательно, для каждого ряда скользящих коэффициентов можно рассчитать процент значимых значений. Таким образом, средний процент значимых коэффициентов конкордации (или корреляции) при определенных значениях W и m можно получить осреднением соответствующих значений процентов, полученных для определенного количества анализируемых скользящих характеристик.

В качестве основных методов для спектрального разложения были использованы метод множественных сфероидальных последовательностей (multiple-taper methods -MTM) (Mann, Lees, 1996; Percival, Walden 1993), сингулярный спектральный анализ (Singular Spectrum Analysis – SSA) (Elsner, Tsonis, 1996) и Вейвлет анализ (Wavelet analysis) (Daubechies, 1992). В отличие

от традиционного преобразования Фурье эти методы могут быть применены к нестационарным временным рядам (Mann, Lees, 1996; Yiou et al., 2000). Более того, эти методы могут эффективно выявлять сигнал в тех случаях, когда уровень шума во временных рядах высокий (Mann, Lees, 1996).

Для сравнительного анализа двух рассмотренных выше скользящих характеристик был использован следующий подход. Было решено проверить, насколько рассмотренные статистические характеристики чувствительны к «шуму» различного рода. Для этого было сгенерированы следующие наборы данных:

1. Выборка 250 временных рядов длины 1550 лет, каждый из которых являлся аддитивной составляющей циклических компонент: $S_t = \sin(2\pi 0.1667t) + \sin(2\pi 0.05t) + \sin(2\pi 0.0333t) + \sin(2\pi 0.01t)$ (где t - время) и красного шума R_t. Циклическая составляющая содержит следующие гармоники в 6, 20, 30 и 100 лет. Напомним, что красный шум обычно описывается при помощи авторегрессионного процесса 1-го порядка AR(1): $R_t=\rho R_{t-1}+\omega_n$, ($n$=0,…,$N$), где ρ - автокорреляционный коэффициент 1-го порядка, ω - гауссовский шум с математическим ожиданием R_0 и дисперсией σ^2 (Allen, Smith, 1994). В качестве ρ было выбрано значение 0.6, которое является, практически, максимальным для различных природных процессов (Mann, Lees, 1996). Такой выбор обусловлен стремлением максимизировать уровень шума в исследуемых рядах.
2. Выборка, представляющая собой 250 временных рядов, каждый из которых представляет собой белый (Гауссовский) шум (N(0,1));
3. Красно-шумовая выборка того же объема.
4. Выборка временных рядов, каждый из которых являлся аддитивной составляющей линейного тренда $Tr_t = 0.002t$ (где t - время) и красного шума с тем же параметром авторегрессии, что и в предыдущем пункте. Модуль величины тренда соответствует среднему тренду в различных временных рядах, полученных в дендроклиматологии.

5. И наконец, множество временных рядов, представляющих собой сумму циклической составляющей S_t, тренда Tr_t и красного шума R_t.

Для каждого из описанных выше наборов были рассчитаны скользящие коэффициенты конкордации и осредненной корреляции с окнами W в 7, 11, 21, 51 и 101 году. Кроме этого, рассматривалась зависимость динамики скользящих характеристик от количества временных рядов *m*, участвующих в расчетах для получения этих коэффициентов. Количество временных рядов варьировалось от 3 до 15.

Для получения достоверной информации о возможных различиях между указанными скользящими коэффициентами было решено получить средний (или обобщенный) скользящий коэффициент для каждой расчетной характеристики при фиксированном *W* и *m*.

При этом, для каждого синтетического набора из 250 временных рядов случайным образом выбиралось *m* рядов. Количество расчетных экспериментов при фиксированных *m* и W было равным 235. После этого, полученные скользящие коэффициенты конкордации (или корреляции) для фиксированных *m* и W усреднялись по времени (Рис. 30).

Начнем сравнительный анализ между скользящими характеристиками с синтетического набора, который является совокупностью циклических составляющих и фонового красного шума (Набор 1).

Во-первых, во всех рассматриваемых сериях в этом множестве присутствует один и тот же гармонический сигнал, причем вклад каждой из регулярных компонент одинаков в общую дисперсию сигнала. И если бы не наличие фонового красного шума, то анализируемые скользящие коэффициенты были бы постоянными и равны 1 вне зависимости от длины окна и количества временных рядов.

Во-вторых, в нашем случае уровень присутствующего шума очень высок и практически является максимальным для различных временных рядов, используемых в дендроклиматологии, климатологии. «Амплитуда» шума такая же, как и у циклических компонент, при этом уровень автокорреляции очень высок (Mann, Lees, 1996) и равен 0.6.

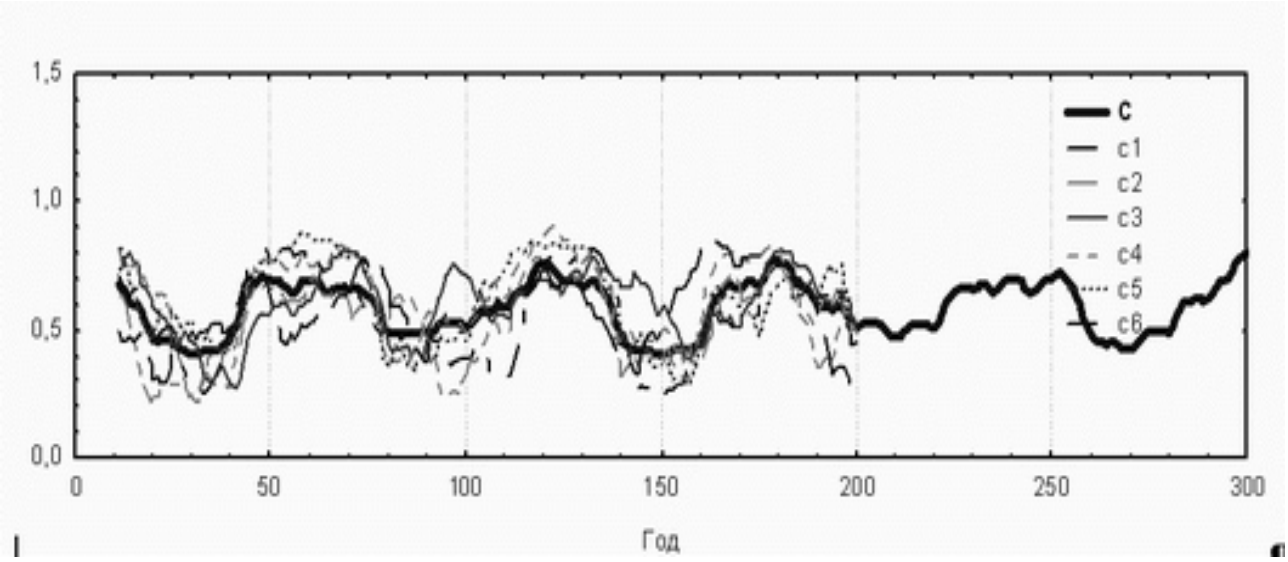


Рис. 30. Пример динамики скользящего коэффициента конкордации, полученного для m=3 и W=21 на интервале в 300 лет. Толстой линией обозначен средний коэффициент конкордации Кендалла, полученный при усреднении 235 скользящих коэффициентов конкордации (7 и них (пунктирные линии) представлены на рисунке).

Этот факт оказывает существенное влияние как на динамику регулярной компоненты (Рис. 31), так и на динамику скользящих характеристик (Рис. 32).

Проанализируем динамику рассматриваемых скользящих коэффициентов.

Выявляется высокая значимая положительная корреляция между скользящими коэффициентами конкордации Кендалла и осредненной корреляции Пирсона.

При этом, в случае, когда в сериях точно присутствует гармонический сигнал, скользящий коэффициент корреляции практически не зависит от количества временных рядов *m*, используемых для его получения (Рис. 32 Б). В то же время, коэффициент конкордации оказывается чувствительным к параметру *m*. Для этого коэффициента можно отметить одно общее свойство. С уменьшением числа *m* среднее значения коэффициента конкордации становиться выше (ближе к 1) при одном и том же значении окна *W* (Рис. 32 А).

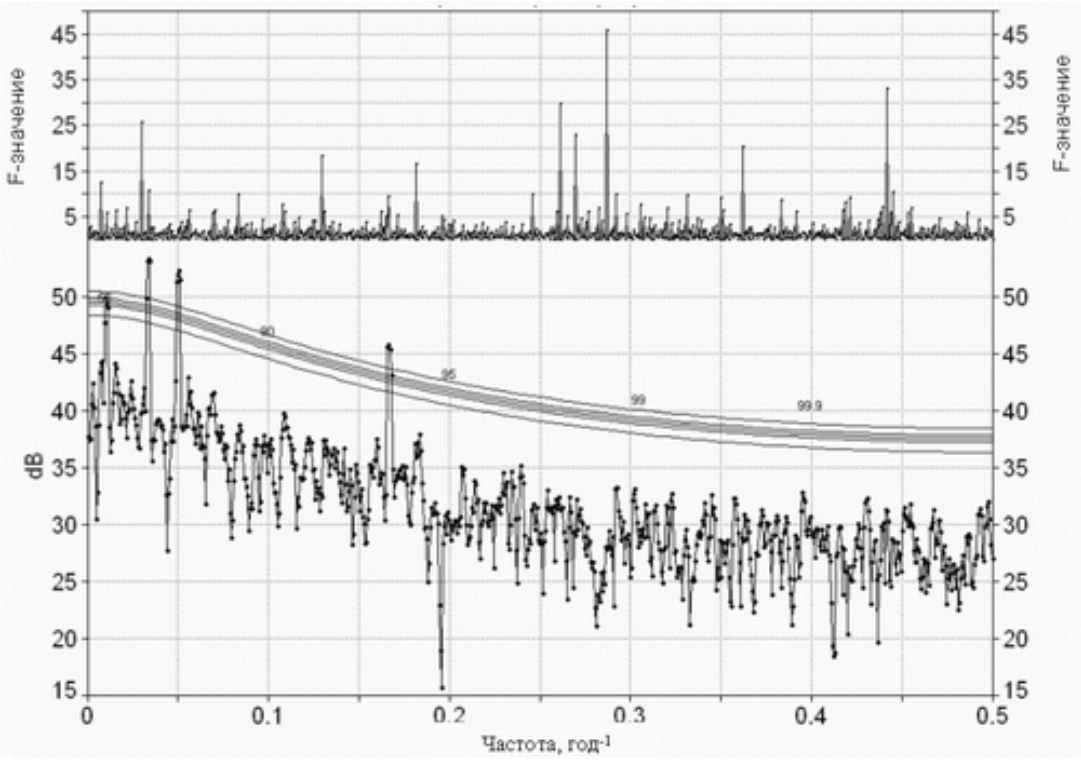


Рис. 31. Спектральное представление одной из 250 смоделированных циклических составляющих (с фоновым красным шумом AR(1)=0.6), полученной при помощи MTM подхода.

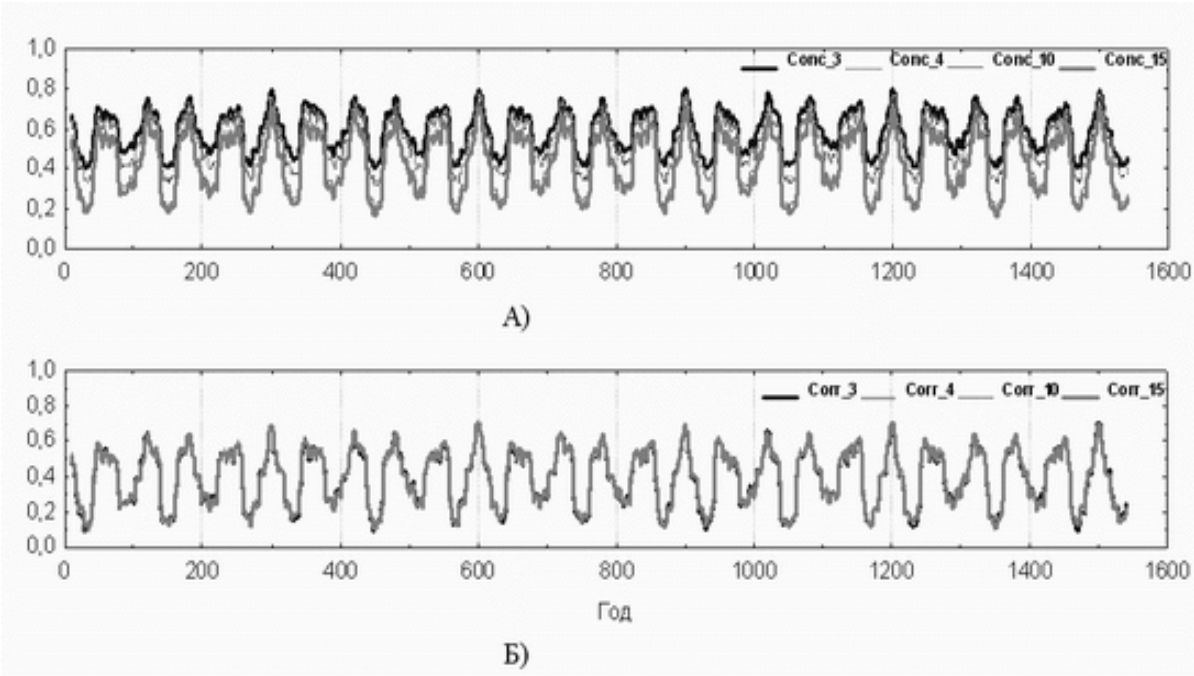


Рис. 32. Средние скользящие коэффициенты конкордации Кендалла (А) и осредненный корреляции Пирсона (Б), полученные для окна W в 21 год и разного количества временных рядов m=3, 4, 10, 15.

Это также подтверждается значениями Евклидова расстояния *D* (Табл. 3) – чем меньше число анализируемых рядов *m*, тем *D* ближе к 0.

Таблица 3.

Статистические характеристики, полученные при осреднении 235 моделируемых скользящих коэффициентов конкордации Кендалла и корреляции Пирсона.

Длина окна *W*	**Число ВР *m***	**Евклидово расстояние для средней конкордации**	**Евклидово расстояние для средней корреляции**	**Средний процент значимых значений коэффициентов конкордации (p<0.05)**	**Стандартное отклонение для предыдущего процента**	**Средний процент значимых значений коэффициентов корреляции (p<0.05)**	**Стандартное отклонение для предыдущего процента**
11	**3**	19.0	27.8	35.3	2.8	12.1	1.6
	4	21.4	27.7	50.2	3.5	9.0	1.5
	10	25.6	27.7	86.3	2.4	4.1	0.9
	15	26.5	27.7	95.3	1.1	2.8	0.7
21	**3**	16.6	24.3	66.4	5.4	45.7	5.1
	4	18.5	24.2	79.7	3.1	45.0	3.9
	10	22.3	24.2	98.2	0.9	44.9	2.6
	15	23.1	24.2	99.8	0.3	45.1	3.3
51	**3**	14.5	21.1	97.6	4.0	92.0	3.6
	4	16.3	21.2	99.6	1.6	94.1	8.5
	10	19.7	21.2	100.0	0.0	98.6	1.6
	15	20.4	21.1	100.0	0.0	99.2	0.5
101	**3**	13.9	20.2	99.7	3.8	99.1	6.6
	4	15.5	20.2	99.9	0.7	99.3	5.3
	10	18.6	20.1	100.0	0.0	100.0	0.0
	15	19.3	20.1	100.0	0.0	100.0	0.0

В частности, средний процент значимых значений для коэффициента конкордации при *W*=11 годам, колеблется в пределах от 35% (при *m*=3) до 95% (при *m*=15) при относительно малых стандартных ошибках процентов (Табл. 3). Такое распределение точно отличается от случайного, что будет показано ниже. Таким образом, скользящий коэффициент конкордации Кендалла может выявлять сигнал в анализируемых «сильно зашумленных» временных рядах даже при небольших значениях окна скольжения W.

При увеличении окна *W* отличия между скользящими характеристиками нивелируются. Но даже при больших значениях окна скользящий коэффициент конкордации продолжает быть характеристикой, которая является более

устойчивой к шуму, так как существуют значимые отличия между абсолютные значениями *D*, рассчитанные для двух множеств реализаций изучаемых скользящих коэффициентов (одно множество – это множество скользящих коэффициентов конкордации Кендалла, другое - множество скользящих коэффициентов корреляции при фиксированной длине окна W, равного 51 или 101 год). Эти различия выявляются на базе дисперсионного и дискриминантного анализа.

Аналогичный анализ синтетического набора временных рядов, состоящего из композиции сигнала, тренда и красного шума (Набор 5) не выявил новых результатов по сравнению с уже описанными.

Анализ других синтетических наборов данных, которые представляли собой временные ряды, состоящие из белого (Набор 2), красного (Набор 3) шумов и линейного тренда под воздействием красного шума (Набор 4), не выявил каких-либо закономерностей в динамике обоих анализируемых скользящих коэффициентов. Все полученные коэффициенты конкордации и корреляции указывают на случайный характер согласованности в изучаемых временных рядах, вне зависимости от их количества и окна скольжения.

Отметим, что даже максимальные значения анализируемых осредненных коэффициентов, полученных для наборов 2-4, являются не значимыми ($p>0.05$).

Динамика скользящих коэффициентов корреляции для наборов 2-4 сильно зависит от количества временных рядов, используемых для его получения. С увеличением числа временных рядов амплитуда колебаний значительно уменьшается. Сравним этот результат с выводом для набора временных рядов 1 и 5 о независимости скользящего коэффициента корреляции от числа временных серий, если они являются комбинаций циклического сигнала и красного шума.

В целом, обнаруживается ряд существенных отличий, которые позволяют однозначно ответить на вопрос о наличии каких-либо закономерностей в динамики согласованности для наборов 2-4. Так, для рассматриваемых синтетических наборов данных средний процент значимых (с уровнем значимости $p<0.05$) значений колебался от 4 до 15% для коэффициента конкордации и от 1 до 5% для коэффициента корреляции. Для набора 1,

характеризующего сигнал, это процент колебался от 36 (при W=11, m=3) до 100% (при W=51) для коэффициента конкордации и от 4 (W=11, m=10; выше уже отмечалось, что осредненный коэффициент корреляции не работает при малых окнах) до 100% (при W=101).

Более того, для синтетических наборов 2, 3, 4 Евклидово расстояние лежит в пределах от 26 до 39 (Заметим, что максимальным Евклидовым расстоянием между 1 и 0 будет, например, 39.23 при длине ряда в 1543 года). В сравнении с синтетическим набором 1 или 5 соответствующее расстояние лежит в пределах от 13 до 25.

На основании проведенного анализа можно сделать следующие заключения:

1. Рассмотренные скользящие коэффициенты конкордации Кендалла и осредненной корреляции Пирсона обладают рядом интересных свойств, связанных с их способностью выявлять сигнал даже в сильно зашумленных временных рядах. Причем коэффициент конкордации может распознавать сигнал даже при меньшем количестве анализируемых временных рядов и меньшей величине окна скольжения по сравнению с осредненным коэффициентом корреляции.
2. Частотный диапазон коэффициента конкордации при различной длине окна скольжения полностью характеризует диапазон исходного сигнала с учетом «новых» пиков спектрограммы, которые образуются в результате нелинейного взаимодействия базовых частот (Шишов и др., 2007б). При этом различные окна скольжения не добавляют «искусственных» пиков.
3. Получено экспериментальное доказательство того, что существуют значимые отличия в динамике скользящих коэффициентов конкордации Кендалла (и осредненной корреляции Пирсона), полученных, с одной стороны, для временных рядов, содержащих циклическую составляющую, с другой, для временных рядов, содержащих только шум, или тренд и шум в совокупности. Эти отличия могут быть выявлены при помощи проверки статистических гипотез на значимость изучаемых коэффициентов и критерия, основанного на сравнении Евклидовых расстояний.

Рассмотренные скользящие характеристики, робастные к шумовым воздействиям различного рода, могут оказаться эффективным инструментом для выявления закономерностей локального и глобального характера при анализе пространственно-распределенных временных рядов различной природы, в частности, дендрохронологических рядов, рядов инструментальных климатических наблюдений и различных климатических реконструкций за последние несколько тысячелетий (Ваганов и др., 1996; Ваганов, Шашкин, 2000; Шишов и др., 2007а; Briffa et al., 1998а; 1998г). В свою очередь, такие закономерности позволяют калибровать различные глобальные климатические модели.

П.8 Алгоритмы теории классификации. D-, S-, X-проблемы.

Пусть $V = \{(x_i, y_i) | x_i, y_i \in \Re \ (i = 1, \ldots, n)\}$, где $\Re$ - множество действительных чисел, является множеством наблюдений (экспериментов). Тогда множество V задает бинарное отношение на $\Re^2$.

Задача сводится к нахождению взаимоотношений (взаимосвязей) между $X = \{x_i | i = 1, \ldots, n\}$ и $Y = \{y_i | i = 1, \ldots, n\}$. Существуют два подхода к решению поставленной задачи:

1. если любому $x_i \in X$ соответствует только один $y_i \in Y$, то ищется некоторый функционал F такой, что:
   a. $Y = F(X)$ - задача прогноза,
   b. $X = F^{-1}(Y)$ - задача управления.

   Замечание: Как правило, в природе одному $x_i \in X$ соответствует несколько $y_i \in Y$, то есть возникает проблема построения функционала F;
2. другой подход основан на анализе именно исходного множества V.

В дальнейшем, нас, в большей степени, будут интересовать различные отношения эквивалентности.

Определение. Отношением эквивалентности S на множестве $\Re$ называется отношение, для которого выполняются следующие условия:

a. S - рефлексивно, то есть:

для любого $x \in \Re$ выполняется $(x, x) \in S$;

b. S - симметрично, то есть:

для любых $x, y \in \Re$ и из того, что $(x, y) \in S$ следует $(y, x) \in S$;

c. S - транзитивно, то есть:

для любых $x, y, z \in \Re$ и из того, что $(x, y) \in S$ и $(y, z) \in S$ следует $(x, z) \in S$.

Схематично некоторое отношение эквивалентности S на $\Re$ можно представить в виде рисунка 33.

Важно отметить, что отношения эквивалентности разбивают некоторое n-мерное пространство на непересекающиеся классы, в чем и состоит задача классификации. Таким образом, некоторая классификация исходных объектов $\omega_1, \ldots, \omega_N$ из некоторой генеральной совокупности Ω, характеризующихся признаками $x_1, \ldots, x_n$ ($x_i \in \Re$)- это задание определенного отношения эквивалентности S в пространстве $\Re^n$.

Сколько существует различных классификаций? Именно столько, сколько существует способов разбиения пространства на непересекающиеся классы. Следовательно, встает проблема с выбором из всех возможных классификаций только тех, которые существуют в природе.

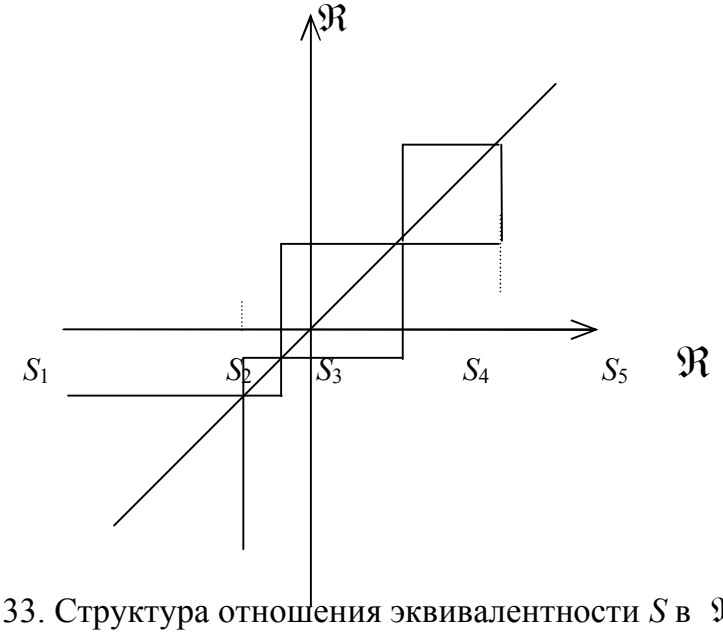


Рис. 33. Структура отношения эквивалентности S в $\Re^2$.

Решение этой и других проблем дает теория классификации (Загоруйко, 1972; 1999; Айвазян и др., 1989; Айвазян, Мхитарян, 1998; Фор , 1989).

Условно все задачи теории классификации можно разбить на три группы:

1. Пусть даны объекты $\omega_1,\ldots,\omega_N$ из некоторой генеральной совокупности Ω, характеризующихся признаками $x_1,\ldots,x_n$ ($x_i \in \Re$), то есть задано пространство признаков *X*. Пусть также существуют классы объектов (т.е. задано некоторое отношение эквивалентности в пространстве *X*).

   Цель: разработать алгоритм, который бы относил некоторый объект ω к нужному классу с наименьшей ошибкой. Следовательно, задача сводится к созданию автомата (программы или машины).

   Это так называемая *D-проблема* (Загоруйко, 1972);
2. Пусть задано пространство признаков $X=\{x_1,...,x_n\}$ и *Z*-обучающая выборка, удовлетворяющая условиям независимости, репрезентативности и так далее.

   Цель: определить оптимальное число классов *K*, а затем определить решающие функции, которые бы с наименьшей ошибкой относили бы некоторый представленный объект к нужному классу.

   Это - *S-проблема*.
3. Пусть даны объекты $\omega_1,\ldots,\omega_N$ из некоторой генеральной совокупности Ω, характеризующиеся признаками $x_1,\ldots,x_n$ ($x_i \in \Re$) *X*. Пусть также существуют классы объектов.

   Цель: оптимизация пространства признаков, то есть: упорядочение признаков по их информативности, уменьшение числа признаков (снижение размерности пространства признаков), замена некоторых признаков на более «доступные».

   Это - *X-проблема*.

Существуют также алгоритмы, которые позволяют разрешать *D*-, *S*-, *X*-проблемы совместно.

В основе практически всех алгоритмов, используемых для решения поставленных проблем, лежит так называемая гипотеза компактности: *объекты из одного и того же класса должны быть расположены «компактно»* (*то есть образовывать группировки*) *в пространстве признаков.* На этой гипотезе основаны многие критерии качества классификации.

Перейдем к более конкретному рассмотрению некоторых алгоритмов, используемых для решения проблем, поставленных выше.

П.8.1. D-проблема и псевдостатистический метод Байеса с модификациями.

Пусть дана выборка объектов $\omega_1,\ldots,\omega_n$ из некоторой генеральной совокупности Ω, характеризующихся признаками $x_1,\ldots,x_p$ ($x_i \in \Re$), т. е. задано пространство признаков *X*. Пусть также существует *k* классов объектов *Y* (т.е. задано некоторое отношение эквивалентности в пространстве *X*).

Тогда *Z*-обучающую выборку можно представить в виде матрицы:

$$\begin{matrix} x_{11} & x_{12} & \cdots & x_{1p} & y_1 \\ x_{21} & x_{22} & \cdots & x_{2p} & y_2 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ x_{n1} & x_{n2} & \cdots & x_{np} & y_n \end{matrix}$$

Значения $x_{i1},\ldots,x_{ip}$ составляют описание $i-$го объекта, а значение y_i указывает на принадлежность его к одному из классов $1,\ldots,k$. Задача состоит в том, чтобы на основании имеющихся данных (обучающей выборке) построить “хорошее” правило предсказания принадлежности наблюдения к классам на основании его описания.

Вероятностная интерпретация задачи классификации с обучением предполагает, что каждое из описаний $x_{i1},\ldots,x_{ip}$ $(i=1,\ldots,n)$ представляет собой наблюдения одной из k $p-$мерных случайных величин, соответствующих классам $1,\ldots,k$ и имеющих плотность вероятности $f_h\left(x_1,\ldots,x_p\right)$ $(h=1,\ldots,k)$. Считаются заданными априорные вероятности q_h $(h=1,\ldots,k)$ того, что очередное наблюдение произведено над $h-$й случайной величиной.

Пользуясь формулой Байеса, можно для любого заданного наблюдения $x_{i1},\ldots,x_{ip}$ вычислить апостериорную вероятность принадлежности его к любому из классов h:

$$P\left(h|x_{i1},\ldots,x_{ip}\right)=\frac{q_h f_h\left(x_{i1,\ldots,}x_{ip}\right)}{\sum_{l=1}^{k} q_l f_l\left(x_{i1},\ldots,x_{ip}\right)}.$$

Можно доказать, что если относить очередное наблюдение к тому классу, для которого максимальна его апостериорная вероятность, то мы получим правило (минимизирующая вероятность ошибочной классификации). Для принятия решений об отнесении наблюдения к классу достаточно сравнить числители $q_h f_h\left(x_{i1},\ldots,x_{ip}\right)$ для $h=1,\ldots k$.

Если предположить, что плотности $f_h\left(x_1,\ldots,x_p\right)$ являются плотностями $p-$ мерных нормальных распределений с равными ковариационными матрицами, то имеем:

$$q_h f_h\left(x_{i1},\ldots,x_{ip}\right)=\frac{q_h}{\sqrt{(2\pi)^p|\Sigma|}}\exp\left(-\frac{1}{2}\left(\vec{x}-\vec{\mu}_h\right)'\Sigma^{-1}\left(\vec{x}-\vec{\mu}_h\right)\right),$$

где

$\Sigma-$ общая для всех классов ковариационная матрица (описание см. ниже),

$\vec{\mu}_h-$ математическое ожидание для распределения $f_h\left(x_1,\ldots,x_p\right)$,

$\vec{x}-$ значение $p-$ мерной случайной величины.

После логарифмирования и отбрасывания общих констант в выражениях для $\ln q_h f_h\left(x_1,\ldots,x_p\right)$ приходим к выражению

$$d_h=\left(\vec{\mu}_h'\Sigma^{-1}\right)\vec{x}-\frac{1}{2}\vec{\mu}_j'\Sigma^{-1}\vec{\mu}_h+\ln q_i, i=1,\ldots k,$$

сравнение которых эквивалентно сравнению апостериорных вероятностей, которые выражаются через d_h следующим образом:

$$P\left(h|x_{i1},\ldots,x_{ip.}\right)=\exp\left\{d_h\right\}\Big/\sum_{l=1}^{k}\exp\left\{d_l\right\}.$$

Для вычисления величины d_i надо знать ковариационную матрицу Σ и векторы математических ожиданий $\vec{\mu}_h$.

Математическое ожидание для каждого класса h оценивается выборочным средним

$$\hat{\vec{\mu}}_j^{(h)} = \sum_{\substack{i=1 \\ y_i=h}}^{n} \frac{x_{ij}}{n_k}, \left(\hat{\vec{\mu}}_j^{(h)} = \left\{ \mu_1^{(h)}, \ldots, \mu_p^{(h)} \right\}' \right),$$

где n_k – число наблюдений, относящихся к классу h $(h = 1, \ldots, k)$,

$$\hat{\vec{\mu}}_j = \sum_{i=1}^{n} \frac{x_{ij}}{n}, \left(\hat{\vec{\mu}} = \left\{ \vec{\mu}_1, \ldots, \vec{\mu}_p \right\} \right)$$

Общая ковариационная матрица $\Sigma = \left\{ \sigma_{lm} \right\}$ оценивается объединенной выборочной внутриклассовой ковариационной матрицей $\hat{\Sigma} = \left\{ \hat{\sigma}_{lm} \right\}$, где

$$\hat{\sigma}_{lm} = \sum_{h=1}^{k} \sum_{\substack{i=1 \\ y_i=h}}^{n} \left(x_{il} - \mu_l^{(h)} \right) \left(x_{im} - \mu_m^{(h)} \right) / (n-k).$$

Выражения

$$\hat{d}_h = \left(\hat{\vec{\mu}}_h \hat{\Sigma}^{-1} \right) \vec{x} - \frac{1}{2} \hat{\vec{\mu}}_h{}' \hat{\Sigma}^{-1} \hat{\vec{\mu}}_h - \ln q_h, \; h = 1, \ldots, k,$$

называются линейными дискриминантами, поскольку представляют собой линейные комбинации наблюдений. Действительно, приведенное выше выражение для $\hat{d}_h$ можно записать в виде:

$$\hat{d}_h = a_{h0} + \sum a_{hj} x_j .$$

Сравнение линейных дискриминантов двух классов h и l можно рассматривать как сравнение с нулем их разности (линейные дискриминантные функции)

$$\hat{l}_{hl} = a_{h0} - a_{l0} + \sum_{j=1}^{p} \left(a_{hj} - a_{ij} \right) x_j .$$

Уравнение $\hat{l}_{hl} = 0$ описывает гиперплоскость в p – мерном пространстве (точку – в одномерном, прямую – в двухмерном, плоскость – в трехмерном и т. д.), делящую все пространство на две части. k дискриминантам $\hat{d}_h$ соответствует $k(k-1)/2$ дискриминантных функций l_{hl} ($h, l = 1, \ldots, k$; $h < l$)

позволяющих делать попарные сравнения принадлежности наблюдения k классам.

Апостериорная вероятность принадлежности к классу h наблюдения оценивается по формуле

$$\hat{P}\left(h|x_{i1},\ldots,x_{ip}\right)=\exp\left\{\hat{d}_h\right\}\Big/\sum_{h=1}^{k}\exp\{\hat{d}_h\}.$$

Оценив апостериорные вероятности для каждого наблюдения из матрицы исходных данных $x_{i1},\ldots,x_{ip}$ ($i=1,\ldots,n$) можно предсказать номер класса y_i^*, к которому должно принадлежать это наблюдение. Кроме того, известен истинный номер класса для этого описания y_i. Соотношение между y_i^* и y_i удобно представлять в виде классификационной таблицы. Строки этой таблицы соответствуют истинным номерам классов, а столбцы – предсказанным. На пересечении строки h и столбца l находится число наблюдений n_{hl} из класса h попавших в класс l. С помощью отношения $n_{hl}/\sum_{l=1}^{k}n_{hl}$ можно оценить вероятность отнесения описания из класса h к классу l (Табл. 4.).

Таблица 4.

Пример классификационной матрицы, полученной при идентификации контрольных 60 объектов при помощи псевдостатистического метода Байеса (см. П.10).

	A	B	Σ
A	25 83.3%	5 16.7%	30 100%
B	0 0%	30 100%	30 100%

Для проверки гипотезы H_0: $\vec{\mu}_1=\ldots=\vec{\mu}_k$ можно использовать статистику

$$D^2 = \sum_{i=1}^{k} n_i (\hat{\bar{\mu}}_i - \hat{\bar{\mu}})' \hat{\Sigma}^{-1} \left(\hat{\bar{\mu}}_i - \hat{\bar{\mu}} \right)$$

называемую обобщенным расстоянием Махаланобиса, которая при больших n_i имеет приближенное распределение χ^2 с $p(k-1)$ степенями свободы

П.8.2. S-проблема и алгоритм семейства «Краб».

Пусть задано пространство признаков $X=\{x_1,...,x_n\}$ и Z-обучающая выборка, удовлетворяющая условиям независимости, репрезентативности и так далее.

Необходимо определить оптимальное число классов K, а затем определить решающие функции, которые бы с наименьшей ошибкой относили бы некоторый представленный объект к нужному классу.

Поставленная задача может быть решена при помощи, например, алгоритма семейства «Краб» (Загоруйко, 1972; Лбов, 1981), критерием же качества классификации может служить F-критерий.

Итак, пусть дана выборка объектов $\omega_1,\ldots,\omega_m$ из некоторой генеральной совокупности Ω, характеризующиеся признаками $x_1,\ldots,x_n$ ($x_i \in \Re$), то есть задано пространство признаков X.

Рассмотрим пространство признаков X.

Необходимо соединить все точки выборки кратчайшей ломанной (известно, что расстояние μ между некоторой точкой M и множеством S определяется, как $\mu(M,S) = \inf_{N \in S}(M,N)$. Так как мы имеем m-точек в пространстве X, то их соединяют (m-1) отрезков ломанной, причем сумма длин всех отрезков ломанной должна быть минимальной (Рис. 34).

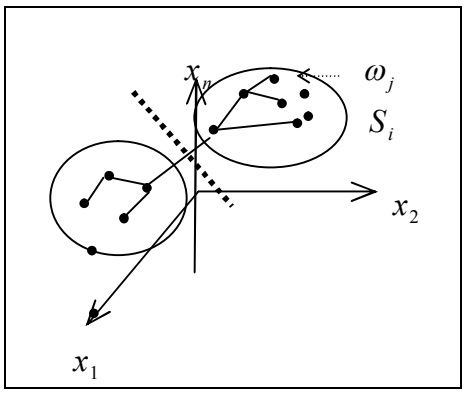


Рис. 34. Схема выделения классов в алгоритме семейства «Краб»

Для того, чтобы получить k классов S_i (i=1,...,k) необходимо разрезать (k-1) отрезков ломанной ($k \le m$). Следовательно, существует $C_{m-1}^{k-1} = \dfrac{(m-1)!}{(k-1)!(m-k)!}$ всевозможных комбинаций разрезания (k-1) отрезков.

Таким образом, необходимо определится с количеством классов k. Для этого используется F-критерий качества классификации, который основан на максимизации следующего выражения:

$$F = \ln \frac{d^a h^b}{(\varphi + \rho)^c (\psi + \lambda)^d},$$

где $d = \dfrac{\sum l_j^{\text{разр}}}{k-1}$ (где $l_j^{\text{разр}}$ - длина j разрезанного отрезка)- мера отдаленности одного класса от другого (чем больше d, тем лучше качество классификации);

$h = k^k \prod_{i=1}^{k} \dfrac{m_i}{m}$ (где m_i- количество объектов в классе S_i (i=1,...,k), причем $m = \sum m_i$) - мера одинаковости численности в классе и $h \in [0,\ 1]$;

$\rho = \dfrac{\sum_{i=1}^{k} \rho_i}{k}$ (где $\rho_i = \dfrac{\sum_{j=1}^{N_i - 1} l_j^{S_i}}{N_j - 1}$ - мера компактности внутри класса S_i) - мера близости внутри классов;

λ - мера одинаковости структуры классов;

a, b, c, d - некоторые весовые коэффициенты;

φ и ψ - некоторые постоянные величины.

Пример:

Пусть $a = b = c = d = 1$, тогда $F = \ln \dfrac{d^a h^b}{(\varphi + \rho)^c (\psi + \lambda)^d}$.

1. Пусть k=2, тогда делаем всевозможные разрезания отрезков кратчайшей ломанной, соединяющей все точки в пространстве X. Количество этих разрезаний равно C_{n-1}^1, при этом для каждого случая считается значение F и находится $F_{\max}$.

2. Пусть k=3, следовательно количество всевозможных разрезаний C_{n-1}^2, аналогично п.1 находим $F_{\max}$ и так далее.

Строится кривая зависимости $F_{\max}$ от *к* - количества введенных классов и ищется глобальный максимум по $F_{\max}$, на основании чего делается заключение о оптимальном количестве классов k (Рис. 35).

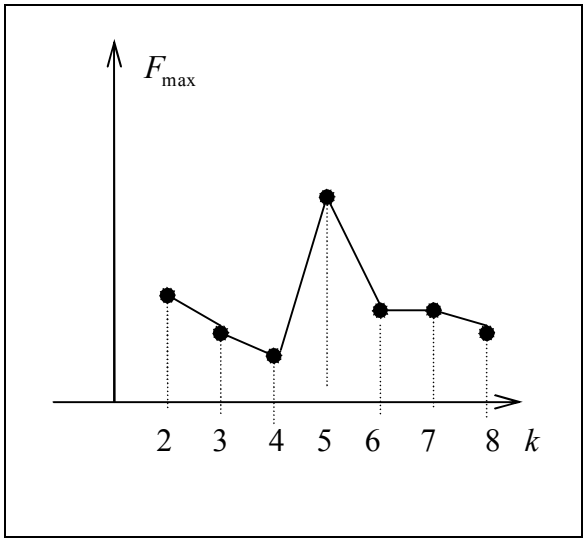


Рис. 35. Кривая зависимости $F_{\max}$ от *к* - количества введенных классов при работе алгоритма семейства «Краб»

Рассмотрим Z-обучающую выборку:

$$\omega_1(x_{11},\dots x_{1n}) \in S_{j_1}$$
$$\omega_2(x_{21},\dots x_{2n}) \in S_{j_2}$$
$$\dots\ \dots\ \dots\ \dots\ \dots$$
$$\omega_m(x_{m1},\dots x_{mn}) \in S_{j_k}$$

Пусть представлен некоторый объект $\omega(x_1,\dots,x_n)$. Необходимо отнести представленный объект к нужному классу (то есть с наименьшей ошибкой).

В алгоритме семейства «Краба» $\omega(x_1,\dots,x_n)$ относят к *j*-классу (*j*=1,...,*k*), если *F*-критерий даёт максимальное значение.

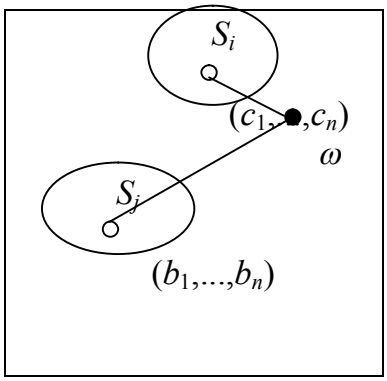


Рис. 36. Схема измерения расстояния между объектом ω и классами S_i, а также S_j.

Другой алгоритм основан на мере, равной евклидовому расстоянию между $\omega(x_1,\dots,x_n)$ и всеми классами (Рис. 36).

Рассмотрим два класса S_i и S_j. Пусть $(c_1,\dots,c_n) \in S_i$ и $(b_1,\dots,b_n) \in S_j$ - координаты центров тяжести для S_i и S_j. В качестве решающей функции предлагается функция вида:

$$F_{ij} = \sum_{r=1}^{n}(c_r - x_r)^2 - \sum_{r=1}^{n}(b_r - x_r)^2$$

При этом, если F_{ij} <0, то скорее всего $\omega(x_1,\dots,x_n)$ относится к *i*-классу.

Процедура повторяется для всех пар S_i и S_j (*i*,*j*=1,...,*k*). После чего, определяется нужный класс.

П.8.3. Х-проблема и метод случайного поиска с адаптацией.

Пусть заданы пространство признаков $X=\{x_1,x_2,\dots,x_n\}$ и классификация, то есть для объектов $\omega_1,\dots,\omega_m$ из генеральной совокупности Ω имеем:

$$\begin{array}{c} \omega_1(x_{11},\dots x_{1n}) \in S_{j_1} \\ \omega_2(x_{21},\dots x_{2n}) \in S_{j_2} \\ \dots\ \dots\ \dots\ \dots\ \ \dots \\ \omega_m(x_{m1},\dots x_{mn}) \in S_{j_k} \end{array}.$$

Требуется упорядочить признаки по информативности, уменьшить число признаков, заменить «дорогие» признаки на более доступные.

Поставленную задачу можно решить, например, при помощи метода случайного поиска с адаптацией, предложенный С.Г. Лбовым (1981).

Поставим в соответствие каждому $x_i \in X$ (i=1,...,n) $\Delta_i \in [0, 1]$ такие, что:

$$\Delta_i = \Delta_j = \frac{1}{n} \text{ (для любых } i, j = 1,\dots,n).$$

1. Выбираем q ($q<n$ (как правило, рекомендуется $q = \frac{2}{3}n$)) признаков $x_{i_1},\dots,x_{i_q}$ из X при помощи датчика случайных чисел. Если $x_{i_j} = x_{i_k}$, то необходимо процедуру выбора повторить, пока $x_{i_j} \neq x_{i_k}$ для любых j, k=1,...,q.
2. Считаем F-критерий качества классификации или средний штраф по формуле (см. выше):

$$F = \ln \frac{d^a h^b}{(\varphi + \rho)^c (\psi + \lambda)^d},$$

где d, h, ρ, λ имеют тот же смысл, что и в алгоритме семейства «Краб».

3. Повторяем п.1 и п.2 l раз ($l<q$) и ищем:

$$\max\{F_1^{(1)},\dots,F_1^{(l)}\} \text{ и } \min\{F_1^{(1)},\dots,F_1^{(l)}\}.$$

Пусть $F_{\max}$ соответствует набору признаков $(x_{i_1},\dots,x_{i_q})$, $F_{\min}$ - $(x_{j_1},\dots,x_{j_q})$.

4. Поощряем $(x_{i_1},\dots,x_{i_q})$ при $F_{\max}$, то есть соответствующие интервалы $(\Delta_{i_1},\dots,\Delta_{i_q})$ увеличиваем на шаг θ, где $\theta = \frac{1}{10n}$, то есть

$$\Delta_{i_k} = \left[\frac{i_k - 1}{n} - \frac{1}{20n}, \frac{i_k}{n} + \frac{1}{20n}\right] (k = 1,\dots,q).$$

Соответственно, штрафуем $(x_{j_1},\ldots,x_{j_q})$ при $F_{\min}$, то есть соответствующие интервалы $(\Delta_{j_1},\ldots,\Delta_{j_q})$ уменьшаем на θ, то есть

$$\Delta_{j_k} = \left[\frac{j_k - 1}{n} + \frac{1}{20n}, \frac{j_k}{n} - \frac{1}{20n}\right]$$

5. Получаем новые интервалы $\Delta_1^{(1)},\ldots,\Delta_n^1$, соответствующие исходным признакам $x_1, x_2, \ldots, x_n$.
6. Повторяем п.1 - п.5 k раз.

Таким образом, получаем:

$$\begin{matrix} \Delta_1^{(1)} & \Delta_2^{(1)} & \ldots & \Delta_n^{(1)} \\ \ldots & \ldots & \ldots & \ldots \\ \Delta_1^{(k-1)} & \Delta_2^{(k-1)} & \ldots & \Delta_2^{(k-1)} \\ \Delta_1^{(k)} & \Delta_2^{(k)} & \ldots & \Delta_n^{(k)} \end{matrix}.$$

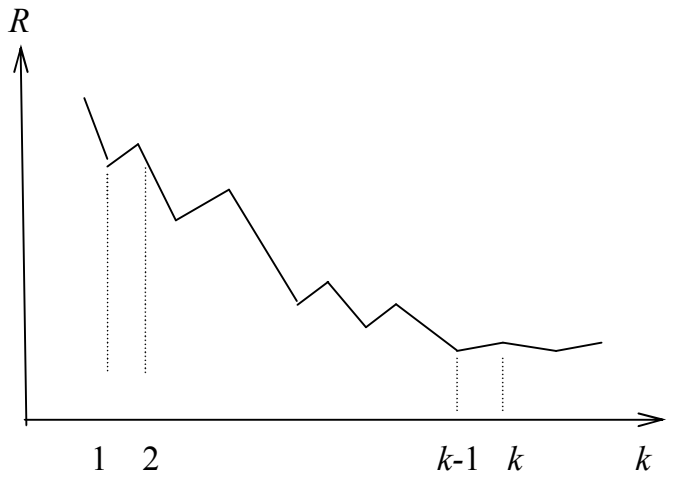


Рис. 37 Схема зависимости *R* от *k*.

Количество повторений k определяется при помощи следующего правила: процесс вычисления продолжается, пока величина $R(k) = \frac{1}{n}\sum_{i=1}^{k}\left|\Delta_i^{(k)} - \Delta_i^{(k-1)}\right|$ практически не станет постоянной (Рис.37).

Таблица 5

Пример использования метода СПА при оценке информативности 21 признака в задаче идентификации событий Эль-Ниньо на основе согласованности в приросте древесных растений.

Номер компонента (признака)	Информативность (в %) при выборе	
	10 компонент	17 компонент
-l0 (10 лет до исследуемого момента)	0,0	3,7
-9 (9 лет до исследуемого момента)	4,2	3,0
-8	15,0	5,4
-7	0,0	7,2
-6	8,3	0,0
-5	0,0	5,6
-4	0,0	6,4
-3	0,0	0,0
-2	12,4	4,5
-1	0,0	0,0
0 (исследуемый момент)	0,0	4,1
1 (1 год после исследуемого момента)	0,0	3,9
2 (2 год после исследуемого момента)	0,0	4,2
3	8,3	7,4.
4	0,0	5,2
5	4,5	5,8
6	0,0	10,4
7	16,8	12,0
8	11,9	0,0
9	16,6	5,9
10 (10 лет после исследуемого момента)	2,0	5,2

Пусть процесс стабилизировался при некотором k, тогда вводится новая величина $p_i = \Delta_i^k$, которая называется информативностью i-исходного признака (i=1,...,n), причем $\sum_i p_i = 1$. Таким образом, все исходные признаки можно упорядочить по информативности, а затем из n признаков выбрать r наиболее информативных, то есть те признаки, которые обладают наибольшими p_i (i=1,...,n) (Таб. 5).

В экологии и климатологии, например, считается, что достаточно выбрать r признаков $(x_{i_1},\ldots,x_{i_r})$, суммарная информативность которых должна превышать 0.75.

Если некоторый признак x_i с высокой информативностью p_i достаточно «дорогой» (например, достаточно трудоемок для измерения), то его заменяют на p более «дешевых» признаков, суммарная информативность которых равна p_i.

П.8.4. XSD - проблема и метод главных компонент.

Пусть $\omega_1,\ldots,\omega_N$ - объекты из генеральной совокупности Ω и $X=\{x_1,...,x_n\}$ - пространство признаков.

Необходимо определится с количеством классов и упорядочить признаки по информативности.

Поставленные задачи можно решить, например, при помощи метода главных компонент (Айвазян и др., 1989), который способствует визуализации выборки объектов без уменьшения количества исходных признаков, следовательно, определиться с количеством классов (S-проблема) и упорядочить признаки по «информативности».

Итак, пусть дана выборка объектов:

$$\begin{cases} \omega_1 = (x_{11},\ldots,x_{1n}) \\ \ldots \quad \ldots \quad \ldots \quad \ldots \\ \omega_N = (x_{N1},\ldots,x_{1N}) \end{cases}.$$

Рассмотрим матрицу ковариации V исходных объектов:

$$V = \begin{Vmatrix} v_{11}, v_{12},\ldots, v_{1n} \\ v_{21}, v_{22},\ldots, v_{2n} \\ \ldots \quad \ldots \quad \ldots \quad \ldots \\ v_{n1}, v_{n2},\ldots, v_{nn} \end{Vmatrix},$$

где $v_{ij} = \dfrac{\sum_{k=1}^{N}(x_{ik}-\overline{x}_i)(x_{jk}-\overline{x}_j)}{N}$ для любого $i, j = 1,\ldots,n$.

Находим собственные значения V, то есть решаем уравнение $|V-\lambda E| = 0$, следовательно, существует n собственных действительных значений $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n \geq 0$, так как матрица V невырождена.

Находим собственные вектора из уравнения:

$$X(V-\lambda E) = 0.$$

Пусть $\bar{f}_i$ - собственный вектор, соответствующий собственному значению λ_i (i=1,...,n).

Каков физический смысл $\bar{f}_i$ и λ_i? λ_1 соответствует максимуму момента инерции l (то есть $\max\frac{\sum(l_i-\bar{l})^2}{N-1}$), l определяет направление $\bar{f}_1$. Затем ищем l_2, ортогональный l, то есть $\bar{f}_2$, ортогональную $\bar{f}_1$ и так далее.

Пусть T ортогональная матрица перехода из одной ортонормированной системы координат $\bar{e}$ в другую - $\bar{f}$, то есть

$$\bar{f} = T\bar{e}$$

Следовательно, старые координаты (признаки) выражаются через новые при помощи системы уравнений:

$$X = YT \quad \text{или} \quad Y = XT^T.$$

Последняя система покоординатно расписывается следующим образом:

$$\begin{cases} y_1 = \tau_{11}x_1 + \ldots + \tau_{n1}x_n \\ \ldots \quad \ldots. \quad \ldots. \\ y_n = \tau_{n11}x_1 + \ldots + \tau_{nn}x_n \end{cases}$$

Таким образом, $(y_1,\ldots,y_n)$ новые координаты (признаки) некоторого объекта ω из генеральной совокупности Ω.

Заметим, что для новых признаков $\hat{v}_{ij} = 0$ (для любых $i, j = 1,\ldots,n$, $i \neq j$).

«Информативность» некоторого нового признака y_i оценивается как значение λ_i.

Таким образом, все исходные признаки можно упорядочить по информативности, а затем из n признаков выбрать r наиболее информативных, то есть те признаки, которые обладают наибольшими λ_i (i=1,...,n).Выбранные признаки также называют главными компонентами. В экологии, например, считается, что достаточно выбрать r признаков (главных компонент) $(y_1,\ldots,y_r)$, суммарная «информативность» которых должна превышать 0.75, то есть

$$\Lambda = \frac{\lambda_1 + \ldots + \lambda_r}{\sum_i \lambda_i}.$$

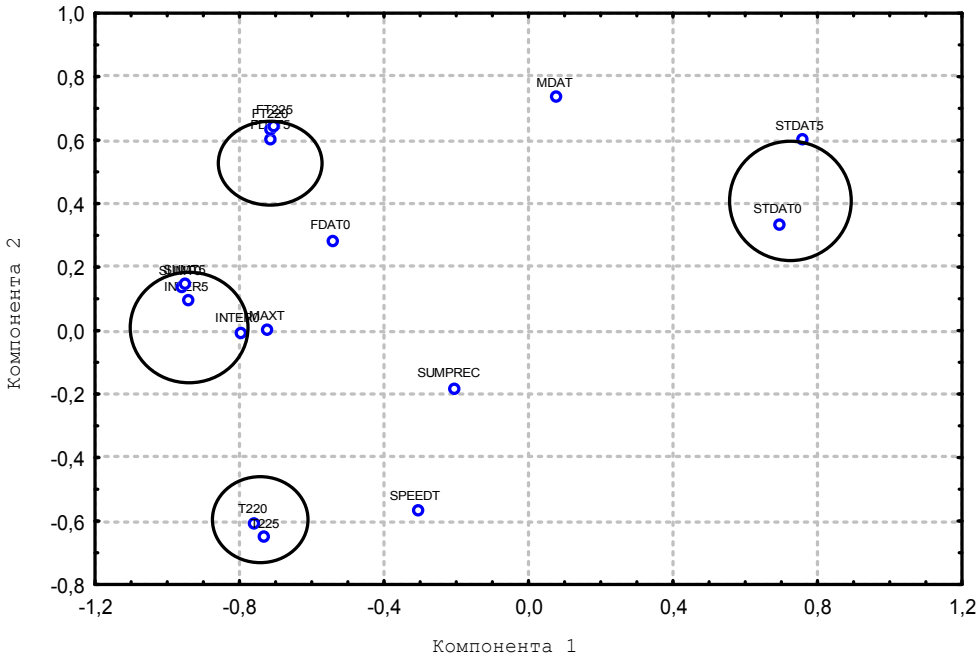


Рис.38. Расположение исходных шестнадцати признаков в плоскости первых двух главных компонент при классификации характеристик сезона роста древесных растений (см. П.12).

Визуализация многомерной классификации, как правило (при удачном выборе исходных признаков), осуществляется при проекции $\omega_1, \ldots, \omega_N$ на первые две главные компоненты y_1 и y_2 (Рис.38).

П.9 Пример использования метода главных компонент при анализе пространственных взаимоотношений между приростом древесных растений и общим содержанием озона в верхних слоях атмосферы

Факт, что температура выступает в роли основного лимитирующего прирост древесных растений фактора в циркумполярной области Северного полушария, является общепринятым (Ваганов и др., 1996, Ваганов, Шашкин, 2000). Об этом свидетельствует, в частности, повсеместная высокая значимая

положительная корреляция в динамике летней температуры и стандартизированными древесно-кольцевыми хронологиями (ДХК). При этом, процент объясненной вариации ДХК при помощи летних температур на периоде инструментальных наблюдений колеблется от 60 до 80 % (Шиятов, 1986; Ваганов и др., 1996). Известно также, что наиболее репрезентативные сверхдлительные хронологии, полученные для Швеции, Ямала, Таймыра и Индигирки являются чувствительными к изменениям глобальной температуры, причем их динамика согласуется в области низких частот (Naurzbaev, Vaganov, 2000).

В то же время, при сравнительном анализе динамики сверхдлительных ДХК с ходом глобальных температур, рассчитанным для различных регионов Северного полушария, наблюдается явное несоответствие в спектре низких частот, которое ярко проявляется в последние 40 лет (с 1960 по 2000 гг.) (Рис. 39). Вероятней всего, на прирост древесных растений начинает воздействовать один из климатообразующих факторов, влияние которого до 1960 года было не существенным.

Ранее была выдвинута гипотеза (Briffa et. al., 1998в), что различия в динамике прироста древесных растений в циркумполярной области и ходом средних температур могут быть объяснены дополнительным воздействием на прирост одного из климатообразующих факторов, например, уменьшением общего содержания озона в верхних слоях атмосферы.

Цель данной работы состоит в проверки выдвинутой гипотезы на базе независимого подхода, основанного на применении факторного анализа с использованием пространственно-распределенных древесно-кольцевых хронологий Урало-Сибирского банка данных (Ваганов и др., 2004).

Наряду с уже упомянутыми сверхдлительными хронологиями, в работе были использованы 94 пространственно-распределенные стандартизированные древесно-кольцевые хронологии, полученные для обширной территории Сибири (Рис. 40).

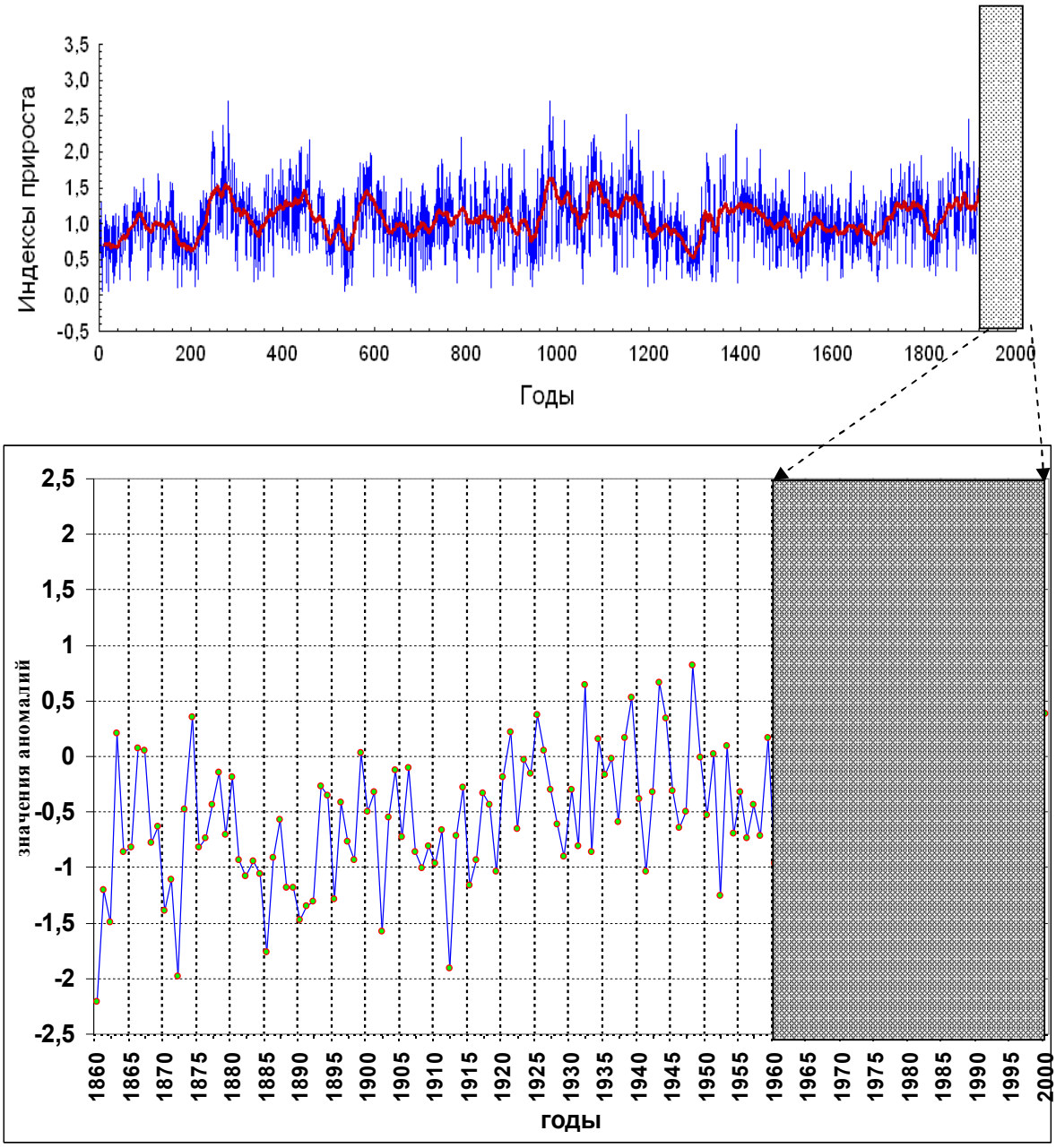


Рис. 39. Динамика индексов прироста древесных растений, полученных для района р. Индигирка, и хода средних температур, рассчитанных для северной части Азии.

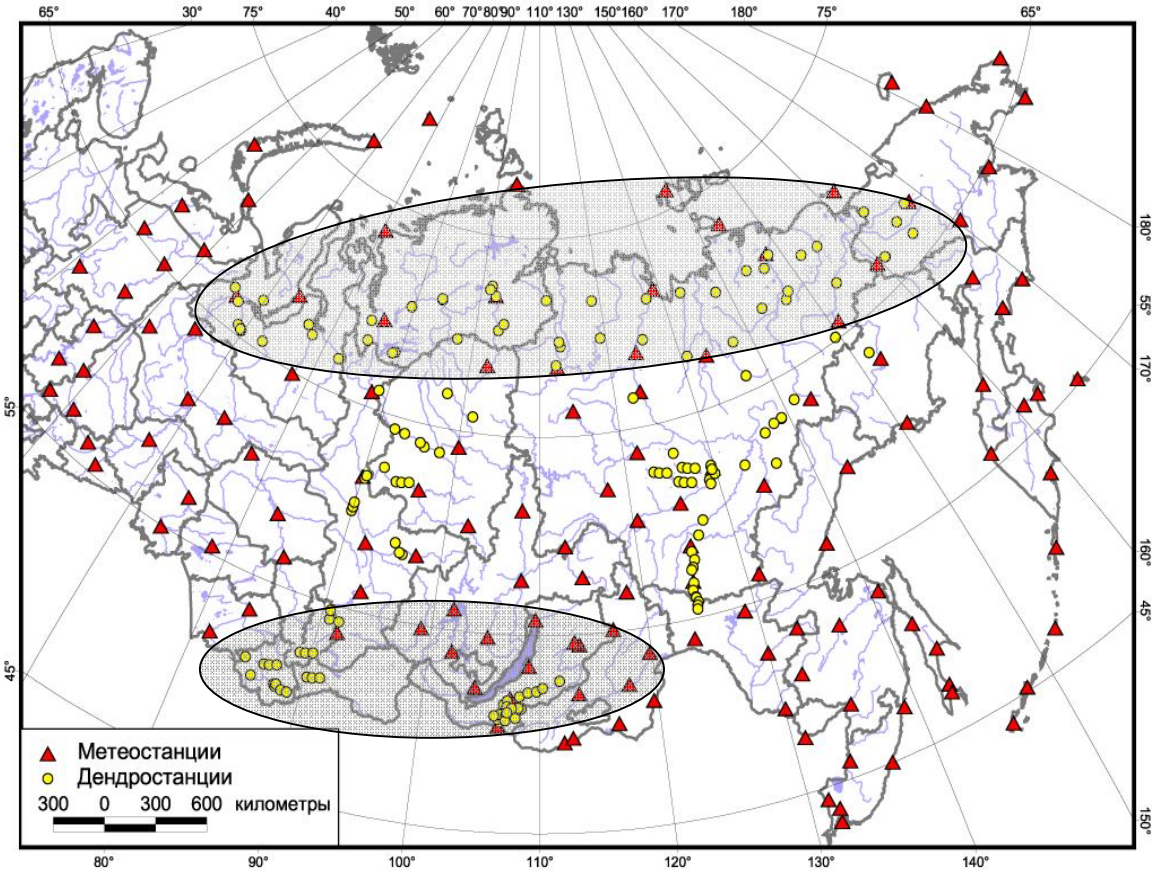


Рис. 40. Карта расположения дендрохронологических тест-полигонов Урало-Сибирского банка данных.

Стандартизация временных рядов велась на основе иллюминирования из исходных данных тренда, связанного с возрастными изменениями древесных растений при помощи негативной экспоненты и линейной регрессии (Шишов и др., 2002; Шишов и др., 2007а).

Дендрохронологические данные сопоставлялись с наиболее репрезентативным рядом наблюдения за общим содержанием озона, полученным в Арозе (Швейцария) (Зуев, 1998; Зуев, Бондаренко, 2001). Несмотря на значительную удаленность сопоставляемых массивов данных, такое сравнение стало возможным на основе широтной модуляции общего содержания озона, которая выражается в высокой значимой пространственной автокорреляции данных по озону, полученных при помощи спутниковой системы TOMS.

В основу анализа пространственно-распределенных данных был положен факторный анализ, а именно метод главных компонент, позволяющий вести эффективную свертку многомерной информации с минимальными потерями (Айвазян и др., 1989).

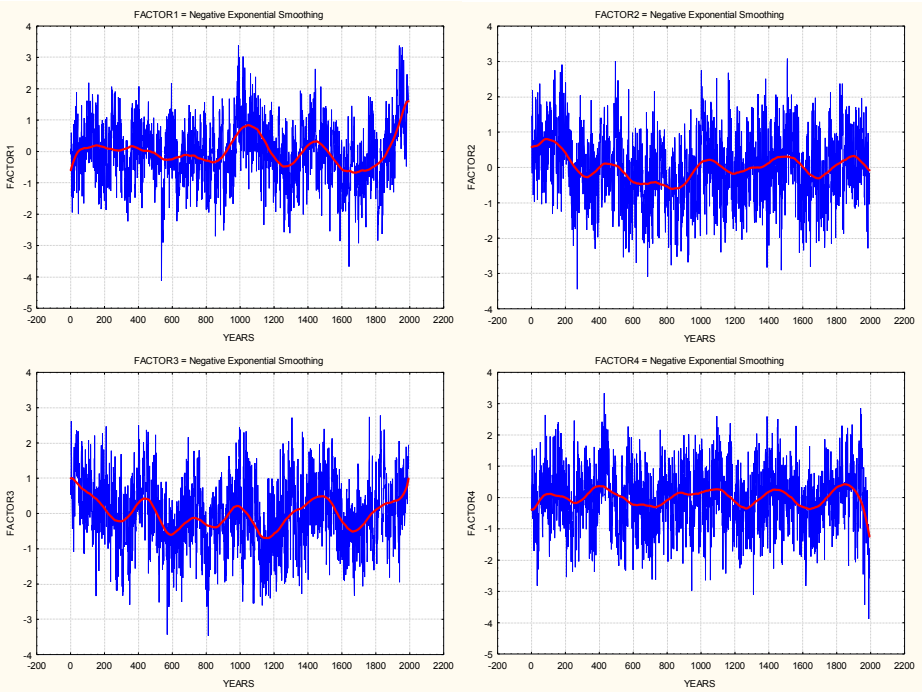


Рис. 41. Главные компоненты (факторы), выделенные по 2000-летним древесно-кольцевым хронологиям.

На первом этапе для 4-х сверхдлительных древесно-кольцевых хронологий были выделены 4 главных компоненты, каждая из которых представляет линейную комбинацию исходных хронологий и является своеобразным аналогом взвешенной «средней» (Рис. 41). Заметим, что при близких факторных нагрузках («весовых» коэффициентов для хронологий) одна из выбранных главных компонент может сильно коррелировать, например, с генерализированной (осредненной) древесно-кольцевой хронологией.

1-ый фактор, на который приходиться 33% общей дисперсии исходной многомерной выборки, положительно коррелирован со всеми ДХК и связан с самым общим климатическим сигналом – глобальной температурой. На 2-ой фактор приходиться 25% общей дисперсии. Интерпретировать эту компоненту

можно как характеристику континентальности рассматриваемых ДХК. 3-ий фактор (23%) связан с западно-восточным трендом в корреляции этой компоненты с исходными древесно-кольцевыми хронологиями. И окончательно, последний выделенный фактор (19%) является характеристикой одного из климатообразующих факторов, который может оказывать эффективное воздействие на прирост древесных растений.

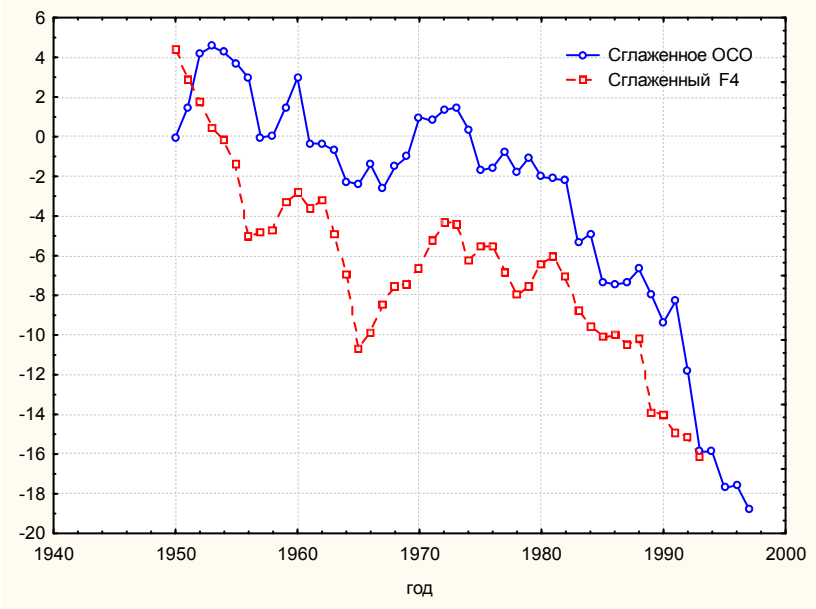


Рис. 42. Динамика сглаженных ОСО и 4-го фактора

Проведенный корреляционный анализ на периоде с 1950 по 1996 гг. выявил наличие значимой положительной связи между данными по общему содержанию озона (ОСО) и последним из выделенных факторов. Причем при экспоненциальном сглаживании анализируемых временных рядов корреляция достигает уровня, равного 0,86 (Рис. 42).

Таким образом, различия в динамике прироста древесных растений в циркумполярной области и ходе средних температур могут быть вполне объяснены увеличением ультрафиолетового жесткого (UF-B) излучения, которое тесно связано с общим содержанием озона (Deluisi et al., 1994).

Для статистического подтверждения полученного результата была использована дендрохронологическая база данных, состоящая из 250 участков, расположенных на территории Сибири.

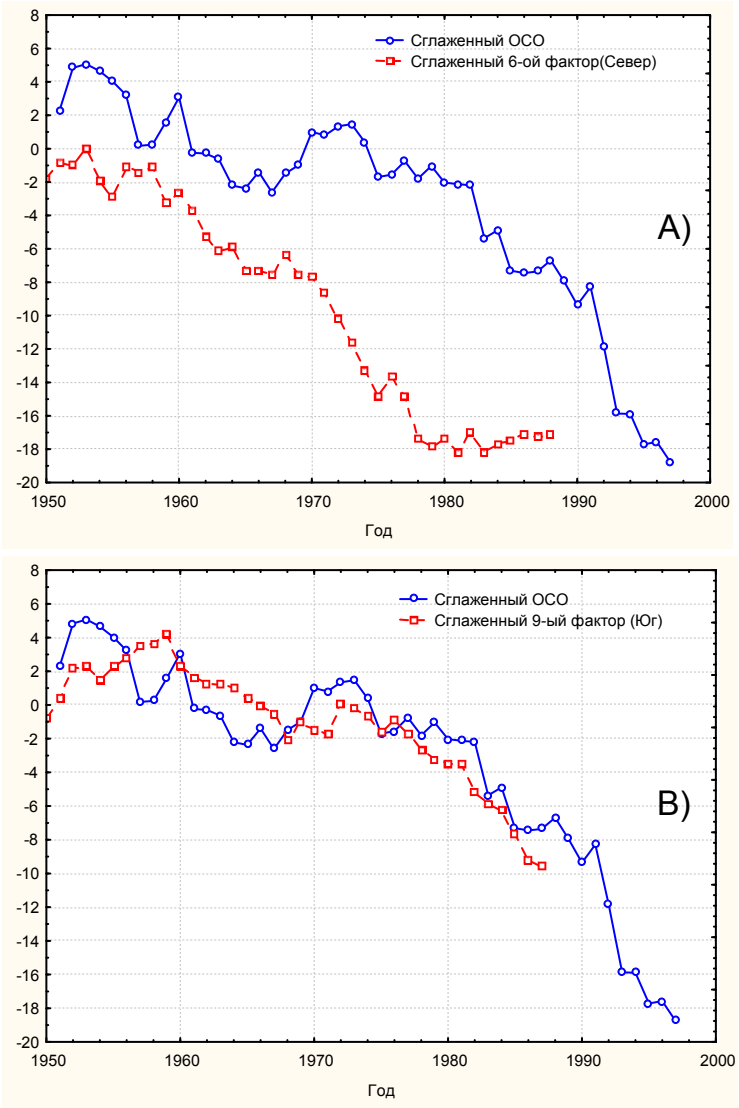


Рис. 43. Динамика сглаженных ОСО, 6-го фактора, полученного для северных ДКХ (А) и 9-го фактора, полученного для южных ДХК.

На основе кластерного анализа было выделено две однородные с точки зрения внешнего климатического сигнала группы древесно-кольцевых хронологий (Рис. 40). В первую группу вошли 51 северных ДКХ, во вторую – 39 южных древесно-кольцевых хронологий. На основе аналогичного подхода на базе факторного и корреляционного анализа для каждой группы хронологий были найдены факторы, которые были положительно коррелированны с ОСО на временном интервале с 1950 по 1996 гг (Рис. 43).

Так для северных хронологий таким фактором оказалась 6-ая компонента, на которую приходилось 5% общей дисперсии. Коэффициент корреляции между

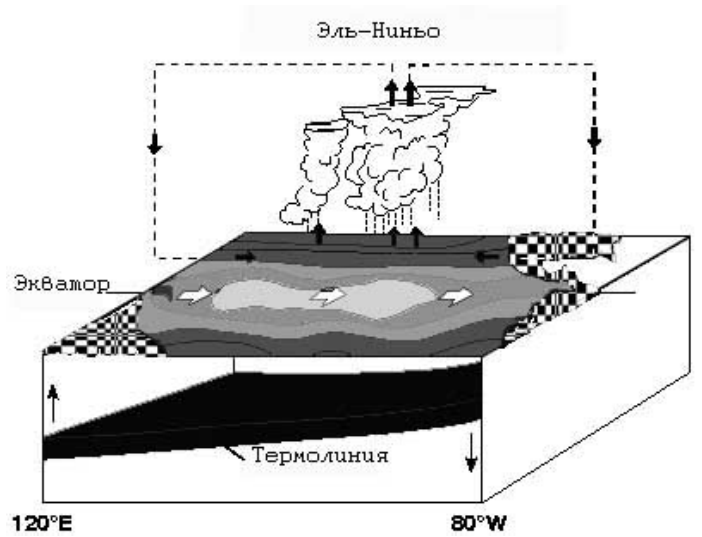
этим фактором и ОСО, предварительно экспоненциально сглаженными, составил 0,76. Для южных хронологий выделенный 9-ый фактор объясняет 2,5% общей изменчивости, при этом коэффициент корреляции между ним и ОСО составил 0,83.

На основе проведенного анализа была установлена статистическая связь между пространственно-распределенным приростом древесных растений в обширных регионах Сибири и динамикой общего содержания озона в верхних слоях атмосферы (Ваганов и др., 2004). При этом обращает на себя внимание тот факт, что выделенный фактор для северных территорий заключает в себе больше информации о древесно-кольцевых хронологиях «северной» группы, чем аналогичный фактор «южной» группы. Мерой информации в данном случае может выступать процент объясненной дисперсии, приходящийся на каждый фактор. Следовательно, высокоширотные ДХК сильнее подвержены влиянию UF-B излучения, отрицательно коррелированного с общим содержанием озона, чем более южные.

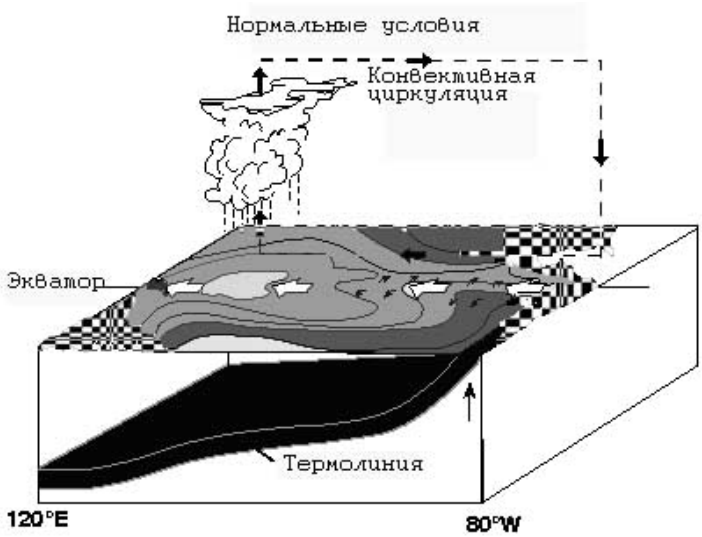
П.10. Пример использования скользящего коэффициента конкордации Кендалла, псевдостатистического метода Байеса и метода случайного поиска с адаптацией при выявлении взаимосвязи между интенсивностями проявления Эль-Ниньо и пространственным приростом древесных растений на территории Сибири.

Термин Эль-Ниньо (El Nino или, далее, EN) определяется, *в узком смысле*, как явление локальное Эквадоро-Перуанское, и *в широком*, как явление глобальное планетарное (Джансеитов, Шишов, 1999; Philander, 1990). Понимание каждого из этих двух толкований чрезвычайно важно при анализе имеющихся данных о времени и силе проявления этого явления в прошлом (Quinn, Neal, 1997).

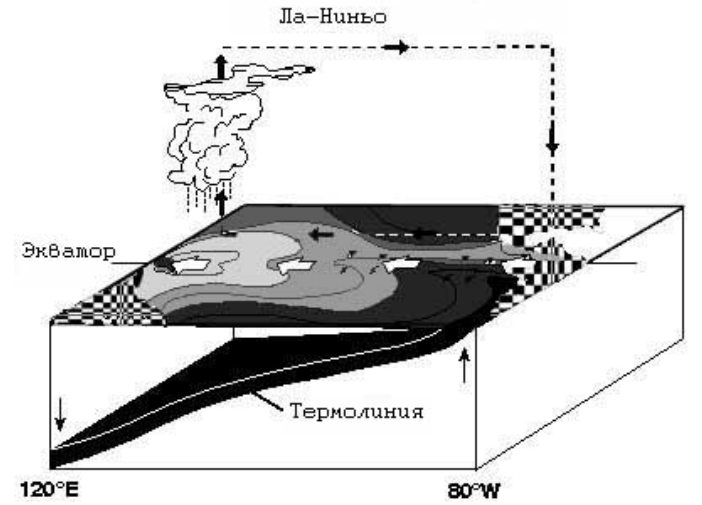
В узком смысле, EN - теплое *сезонное* течение поверхностных вод пониженной солености в восточной части Тихого океана (Рис. 44),



А)



Б)



В)

Рис. 44. Схематическая диаграмма изменения температуры слоев воды, направления течений и ветров в Экваториальной части Тихого океана в годы Эль-Ниньо (А) и Ла-Ниньо (В) по сравнению с нормальными условиями (Б)

которое распространяется летом Южного полушария (декабрь-январь) вдоль берегов Эквадора от экватора до 5-7° ю.ш., в отдельные годы достигая 15° ю.ш (http://www.pmel.noaa.gov/tao/elnino/nino-home.html). По-испански EN означает «Иисус во младенчестве» из-за его «приуроченности» к Рождеству.

Его, в той или иной степени, глубокие (заметные или замеченные) проникновения на Юг осуществляются через различные неравномерные интервалы времени от 2 до 10 лет, а по другим данным от 3 до 7 лет.

Проникновение EN как сезонного течения на Юг вдоль берегов Эквадора и Перу сопровождается: оттеснением от побережья холодного Перуанского прибрежного течения, направленного с Юга на Север, формирующего район исключительно высокой биологической продуктивности вод, особенно из-за сгонного действия юго-восточных пассатных ветров, обеспечивающего подъем глубинных вод, богатых питательными солями (Philander, 1990).

Кроме этого:

- прекращением тонким слоем теплых вод EN поступления кислорода в подповерхностные воды, что губительно действует на планктон и рыб;
- ослаблением действия юго-восточных пассатных ветров, что из-за уменьшения сгонного действия ведет к исчезновению фитопланктона, который может расти только в прохладной воде вблизи освещенной поверхности;
- обилием дождей на побережье Южной Америки, приводящих к многочисленным оползням и наводнениям на обычно засушливом побережье;
- значительным уменьшением количества рыб (анчоус, например, или гибнет, или покидает побережье) и массовой гибелью птиц, обеспечивающих прибрежные почвы, с Востока ограниченных Андами, удобрением (гуано);
- и, как следствие, возникновением бедственного положения у населения, нанесением значительного экономического ущерба прибрежным странам (Philander, 1990; Hayes et al., 1991; McPhaden, 1993).

Температурный баланс, как в прибрежной зоне, так и во всем Тихом океане восстанавливается через несколько месяцев после Рождества, в то время как

формирование этого явления, как будет показано ниже, занимает более длительный период времени. Динамика силы действия EN во времени носит асимметричный характер.

Описанное выше явление особенно резко проявлялось в 1891, 1925, 1941, 1953, 1957-58, 1972-73, 1982-83, 1986-1987, 1991-1992, 1993, 1997-1998 гг..

Среди отмеченных событий, событие EN 1982-1983 годов (EN-82) было первым явлением, носящим исключительный характер:

- потепление воды наступило уже в мае 1982 года, а не в конце года, как бывает обычно;
- после потепления у берегов Южной Америки в декабре на 4° (вместо обычных 1-2°), к апрелю возросло еще на 6° (на некоторых участках у Перуанского побережья температура воды на поверхности достигла 29°), а на глубине более 100 метров - 20°;
- уровень моря у побережья поднялся на 40 см (вместо обычных 15 см) (Quinn, Zopf, 1984; Quiroz, 1983; Rasmusson, Hall, 1983).

Волна стихийных бедствий, включающая действия EN-82 как *локального* прибрежного Эквадоро-Перуанского течения, прокатилась по всему миру, суммарный ущерб по последующим оценкам оказался равным 13 миллиардам долларов. Именно EN-82 в значительной степени заставило говорить об этом явлении как явлении глобальном планетарном (Le Comte, 1983).

Существует ряд публикаций (Le Comte, 1984; Le Comte, D., 1987, Hill, March, 1995), в которых отмечается, что в год развития EN на всей планете меняется климат: теплей становится зима на севере США, в пустынях севера Чили резко увеличивается количество осадков, экстремальные засухи поражают районы Индонезии, Австралии, Индии, Кореи, Китая.

Национальная ассоциация по наблюдению за океаном и атмосферой (NOAA), совместно с агенством CNN, например, сообщила, что климатические изменения, связанные с EN-98, коснулись даже Европейской части и Дальнего Востока России.

Что же из себя представляет EN в широком планетарном смысле? Чтобы это пояснить, воспользуемся схемой (Рис.44), и процитируем описание явления

EN в глобальном смысле, опубликованное на официальном сайте NOAA (http://www.pmel.noaa.gov/tao/elnino/nino-home.html):

«EN, в упрощенном виде можно описать следующим образом: оно прежде всего изменяет направление всех тихоокеанских течений, проходящих вдоль южноамериканского побережья и в зоне экватора. Обычно вода течет вдоль берегов Чили, Перу и Эквадора с юга на север. Вдоль экватора ветры гонят воду на запад в сторону Океании и Новой Гвинеи. Вода, которая приходит с Южного полюса, прохладна. Поскольку сила Кориолиса относит текущие на север воды одновременно от берега и в западном направлении, дополнительно из более глубоких слоев океана к ним примешивается холодная вода. На своем пути вдоль экватора на запад через Тихий океан вода постепенно нагревается. В то время как у берегов Перу ее температура составляет 17°—20° Цельсия, в районе на 13 тысяч километров западнее, около Новой Гвинеи, она в этот же период нагревается до 29°. Однако движение воды в западном направлении приводит не только к повышению температуры поверхностных слоев: напор воды с температурой поверхностного слоя более 20° значительно возрастает в западном направлении. Одновременно возникает также разность уровней моря: поверхность Тихого океана в обычное время слегка поднимается от южноамериканского побережья в сторону Новой Гвинеи…».

«Эль-Ниньо все как бы «выворачивает наизнанку». Вначале перемещение воздушных масс и воды на запад замедляется и, наконец, полностью прекращается. Соответственно не так быстро текут воды вдоль побережья на север, а затем и почти совсем перестают течь. Одновременно это замедление движения приводит к более интенсивному нагреву воды. И в конечном итоге дело доходит даже до изменения этого движения на обратное. Слишком высокий уровень моря у берегов Новой Гвинеи начинает исчезать при постепенном повышении его на восток, в сторону Южной Америки. На западе он понижается примерно на 30 сантиметров, а на востоке, близ Перу, он, наоборот, повышается на 15 сантиметров. Теплые воды, а также ветры перемещаются теперь с запада на восток. Температура воды у берегов Эквадора и Перу повышается, при этом теплый слой становится толще. Вдоль побережья течение несет теплую воду с

экватора в направлении с севера на юг. Прекращается приток холодной воды из глубины ...» (далее следует вспомнить то, что было описано выше о действии EN как сезонном течении вдоль Эквадоро-Перуанского побережья).

«...Одновременно с этим изменяется также система циркуляции атмосферного воздуха над экватором. В то время как обычно в Индонезии барометрическое давление низкое, а в центральной части Тихого океана оно высокое, теперь все становится наоборот. Исследователи даже придумали для этого явления термин «атмосферный эффект качелей». Подобно течениям, протекающим теперь не на запад, а на восток, в воздушных слоях над водой также происходят изменения: в тропосфере на высоте от 10 до 15 километров ветры дуют уже не с запада, а с востока вдоль экватора. Подобным же образом влажные воздушные массы зарождаются теперь на востоке, а не на западе. Соответственно смещаются на восток и осадки.»

Не исключено, что аналогичные глобальные изменения происходят и в Индийском океане между Африкой и Австралией, и в Атлантическом океане между Южной Америкой и Африкой благодаря объединяющему эти три океана холодному Течению западных ветров, охватывающему Антарктиду с Запада на Восток, ветвями которого являются, в частности, Перуанские океаническое и прибрежное течения, а также наличию во всех этих трех океанах теплых Южных пассатных течений (с Востока на Запад) (Philander, 1990).

Эти изменения через Межпассатные противотечения, Северные пассатные течения (Муссонное течение - в Индийском океане), Гольфстрим в Атлантическом и Куросио - в Тихом океане вполне могут переноситься и на Северное полушарие Земли, например, в Россию, соответственно, через Европу и Дальний Восток. Этому переносу могут способствовать и глобальные атмосферные процессы (Швед, 1997*),* тесно взаимосвязанные с гидросферными. Еще в 1924 году английский метеоролог Джильберт Уолкер полагал, что механизмы взаимосвязанностей удаленных друг от друга на значительные расстояния погодных явлений следует искать в воздушных слоях на высотах от 10 до 20 километров (Рис. 45), а там можно проследить влияние взаимодействия постоянно изменяющихся магнитосферы и солнечного ветра

(Рис. 46), за которыми стоят процессы в ядре Земли и Солнца (Пудовкин, 1996; Швед, 1997).

Итак, можно сделать предположение, что EN, в узком смысле, как событие, связанное с проникновением теплого течения вдоль Эквадора и Перу, представляет собой локальное проявление EN в широком планетарном смысле, причем заключительным этапом более длительных глобальных изменений.

Такие проявления EN и их влияние на процессы в северном полушарии свидетельствуют о планетарном масштабе возмущений климатической системы, ассоциируемых с событиями Эль Ниньо разной интенсивности (Арпе и др., 1999; Angell, 1992; Fraedrich, 1990; , Hurrell, 1996). Отметим также, что выявлена связь между полярным вихрем и такими экваториальными характеристиками, как температура поверхности океана и квази-двухлетние осцилляции направления стратосферного экваториального ветра (QBO) (Angell, 1992).

Также показано существование корреляционной связи между изменением зимней температуры в Северном полушарии и динамикой индексов Южной и Северо-Атлантической осцилляций (Hurrell, 1996). Следует отметить, что такие наблюдения, как правило, не опираются на длительные данные и весьма редки.

Эту степень и асимметричность EN в широком смысле необходимо учитывать при оценке достоверности данных, касающихся силы и времени проявления EN (в виде катастроф) в прошлом, особенно основанных на литературных источниках отдельных авторов прошлого, начиная с 1525 года (Quinn, Neal, 1997).

Визуальный анализ данных (Рис. 47), касающихся моментов, замеченных в прошлом проявлений EN, и оценок их интенсивности (силы), показывает, что со временем от настоящего к прошлому уменьшается частота фиксации проявлений EN и уменьшается разброс по силе проявлений (Джансеитов, Шишов, 1999).

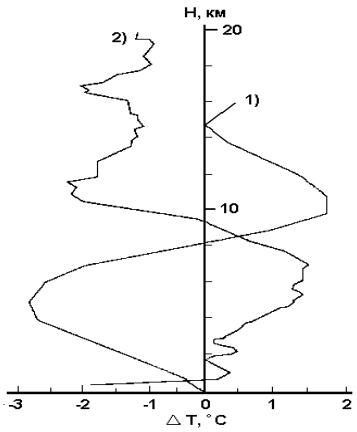


Рис. 45. Высотный профиль отклонений температуры воздуха от спокойного уровня во время всплесков потоков солнечных космических лучей (кривая 1)) и форбуш-понижений (кривая 2).

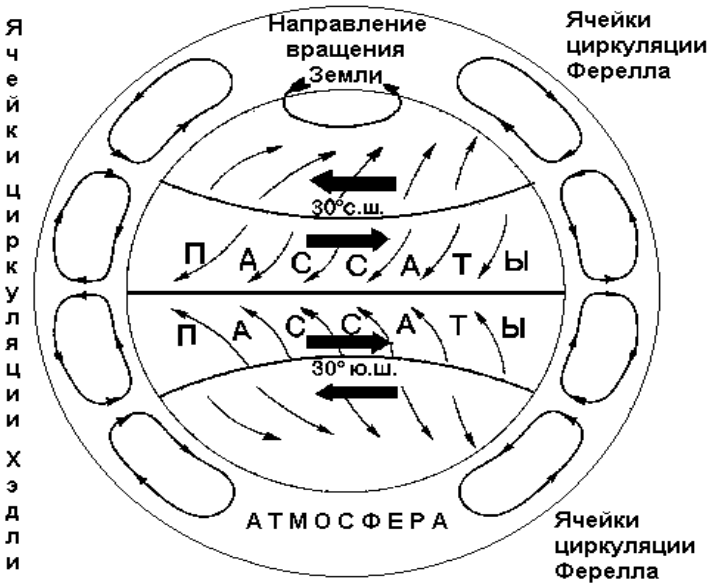


Рис. 46. Система конвективных ячеек межширотной циркуляции воздуха в атмосфере Земли, соответствующие ячейкам приповерхностные ветры (тонкие незамкнутые стрелки) и направление силы трения атмосферы о поверхность планеты (толстые стрелки).

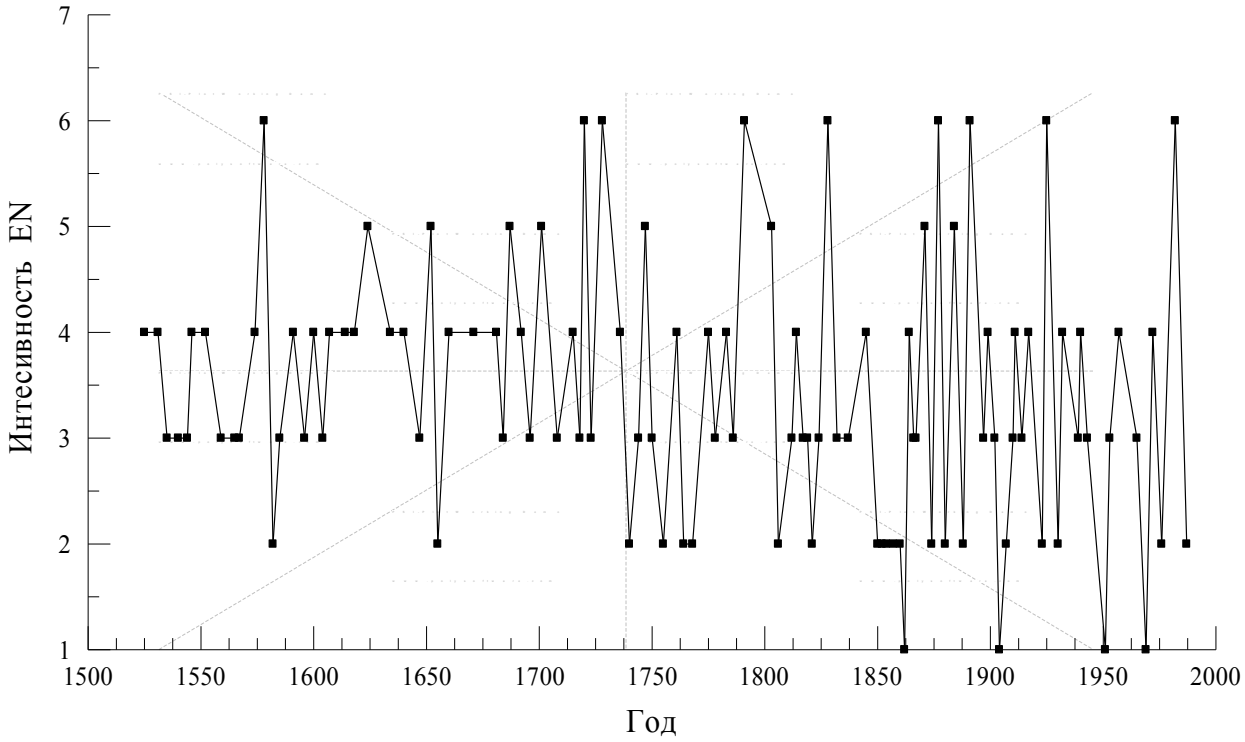


Рис. 47. Динамика интенсивности проявлений событий Эль-Ниньо

Рассмотрим частоты проявления EN по пятидесятилетиям и столетиям. Если предположить, что число проявления EN на определенном интервале времени представляет собой случайную величину, имеющую биномиальный закон распределения, то вероятности наступления события EN должны находится в указанных доверительных интервалах с указанными уровнями надежности *p* (Дудин-Барковский, Смирнов, 1955).

В нашем случае, вероятности наступления события EN по 50-летиям (Рис. 48 (1)) существенно различаются. Например, доверительный интервал, соответствующий 50-летию с центром в 1875 году не содержит частоты наступления события EN в интервалах с центрами в 1575, 1625, 1675, 1725, 1775 гг., что равносильно наступлению некоторого события с вероятностью 0.05^5. Аналогичная ситуация наблюдается и в случаи столетних интервалов (Рис. 48 (2)): доверительный интервал, соответствующий столетию с центром в 1900 году не содержит частоты наступления события EN в интервалах с центрами в 1600, 1700, 1800 гг., что равносильно наступлению некоторого события с вероятностью 0.05^3.

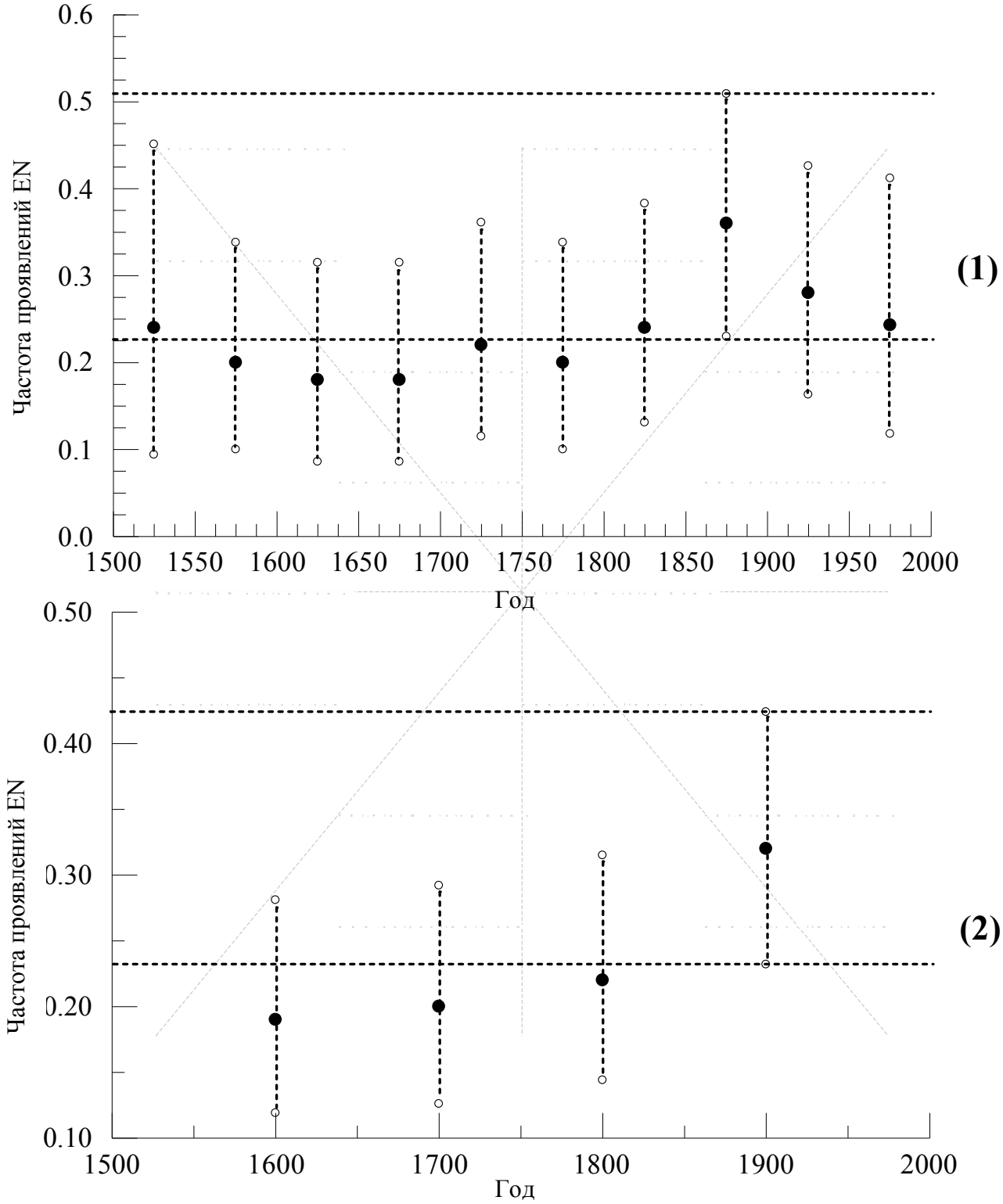


Рис. 48. Частоты проявления EN по 50-летиям (1) и 100-летиям (2) с указанием доверительных интервалов с $p<0.05$ для вероятности наступления события EN

Следовательно, если судить по данным (Рис. 47), гипотеза H_0 о постоянстве вероятности наступления события EN в рассматриваемом интервале с 1525 по

1987 гг. отвергается, т.е. либо не все события EN фиксировались, либо анализируемый процесс является нестационарным. Наблюдается рост вероятности наступлений EN.

Проанализируем оценки средних квадратических отклонений числа наступлений события EN в пределах столетних отрезков времени с доверительными интервалами (Рис. 49 (2)), в предположении, что эти числа наступлений имеют биномиальный закон распределения.

Расчеты показывают, что доверительный интервал, соответствующий столетнему интервалу с центром в 1900 году, не содержит оценки средних квадратических отклонений столетних интервалов с центрами в 1600, 1700, 1800 гг., что подтверждает вышеизложенное утверждение (Дудин-Барковский, Смирнов, 1955).

Рассмотрим показатели силы проявления EN (Рис. 49), которые имеют следующую градацию: M- - умеренное EN со знаком «-» (принимаемое нами за 1), M - умеренное (2), M+ - умеренное с «+» (3), S - сильное (4), S+ - сильное с «+» (5) и VS - очень сильное (6) (как правило, качественные значения показателей равноудалены друг от друга). Встречающиеся в данных 4 значения M/S, 1 - M/S+, 1 - M-/M+ отнесем, соответственно, к M+, S и M (Quinn, Neal, 1997).

Средние значения показателей силы проявления EN по столетиям с указанием их доверительных интервалов в предположении, что эти значения представляют собой сгруппированные случайные величины, имеющие нормальный закон распределения (Рис. 49 (1)), показывают, что гипотеза H_0 о постоянстве средней силы проявления события EN в рассматриваемом интервале с 1525 по 1987 гг. отвергается, т.к. доверительный интервал значения показателя средней силы проявления EN, соответствующего столетию с центром в 1900 году, не содержит среднее значение показателей, соответствующих столетиям с центрами в 1600, 1700 гг. (Дудин-Барковский, Смирнов, 1955).

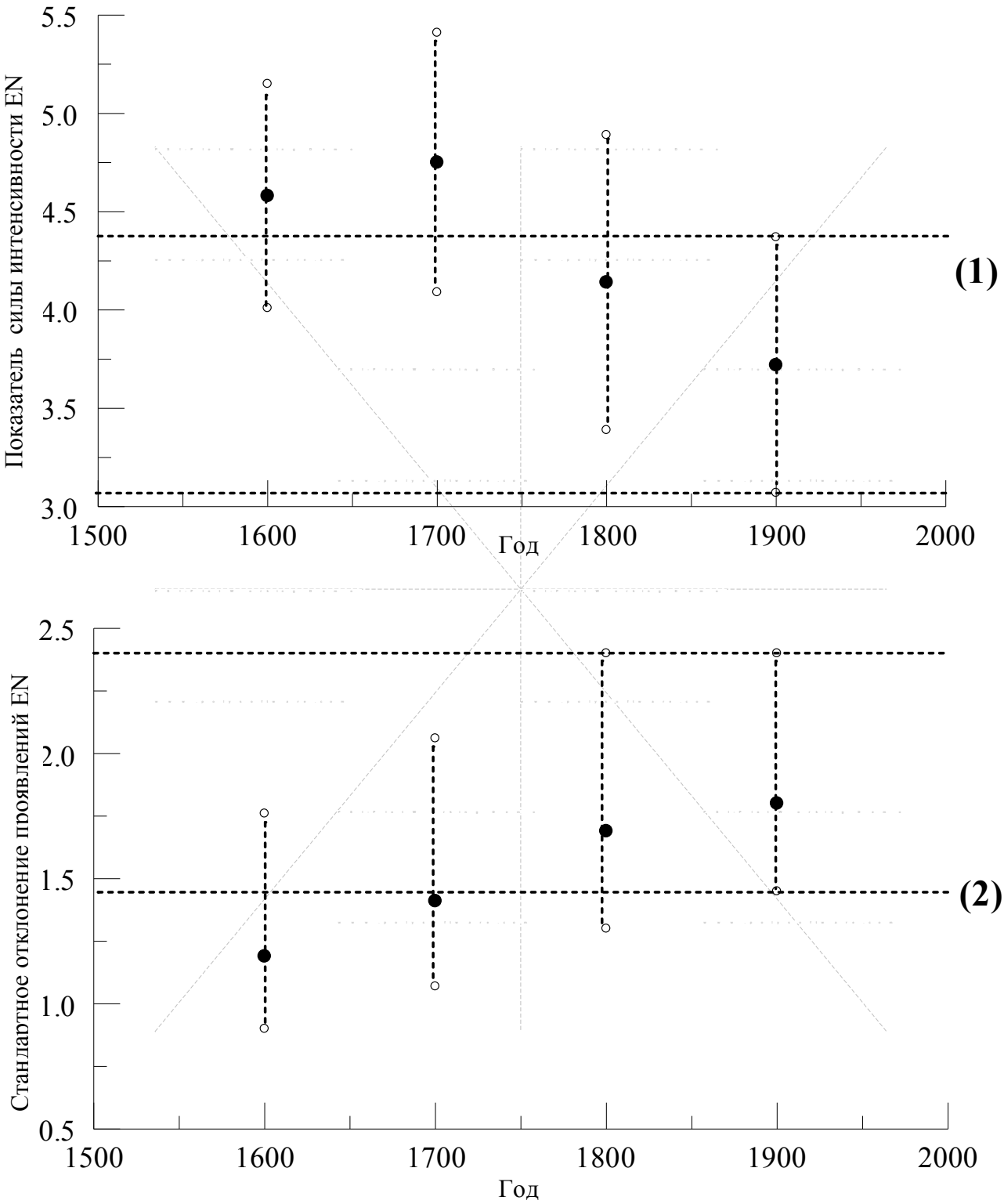


Рис. 49. Средние значения показателей силы проявления EN (1) и оценки среднеквадратических отклонений числа наступлений событий EN (2) по столетиям с указанием доверительных интервалов ($p<0,05$)

Таким образом, либо в последние столетия стали больше фиксировать слабые проявления EN и точней оценивать эти проявления EN, чем в первые

столетия, в которых также, из-за меньшей достоверности, влияние EN занижалось (Рис. 50, 51), либо идет объективное уменьшение силы проявления EN (Джансеитов, Шишов, 1999).

Рассмотрим средние значения интервалов между проявлениями EN разной интенсивности по столетиям. На рисунке 51 к 1600 и 1900 годам отнесены интервалы от 1525 до 1650 гг. и от 1850 по 1987 гг. При построении каждой кривой учитывалось число проявлений EN, сила которых не меньше указанного порога. Например, кривая (6) учитывает все проявления EN , а кривая (1) - только проявления очень высокой интенсивности (VS).

Убывание кривой (1) указывает на увеличение числа фиксаций проявлений очень сильных EN, убывание кривых (5), (6) указывает на увеличение общего числа фиксаций проявлений EN. Средние величины интервалов между фиксациями по столетиям меняется от 3.4 до 5.1 (Рис. 51).

Сопоставление кривых (Рис. 50) динамики интенсивности замеченных проявлений EN и числа авторов, отмечающих характер интенсивности проявлений, показывает существование достаточно высокой значимой корреляции между ними ($\rho = 0.52$). Кстати, число авторов со временем растет по экспоненциальному закону $y = Ae^{\alpha t}$ (A=0.066, α=0.0025).

Заметим, что между числом авторов и достоверностью данных по интенсивности проявлений EN также существует высокая значимая корреляция ($\rho = 0.64$)

Рассмотрим доказательство статистической связи между сильными проявлениями событий Эль-Ниньо и изменениями летней температуры в высоких широтах северного полушария для обширного района Субарктики Сибири. При этом, не делается каких-либо дополнительных предположений о существовании циклических колебаний в проявлениях EN различной интенсивности (Шишов, 2000).

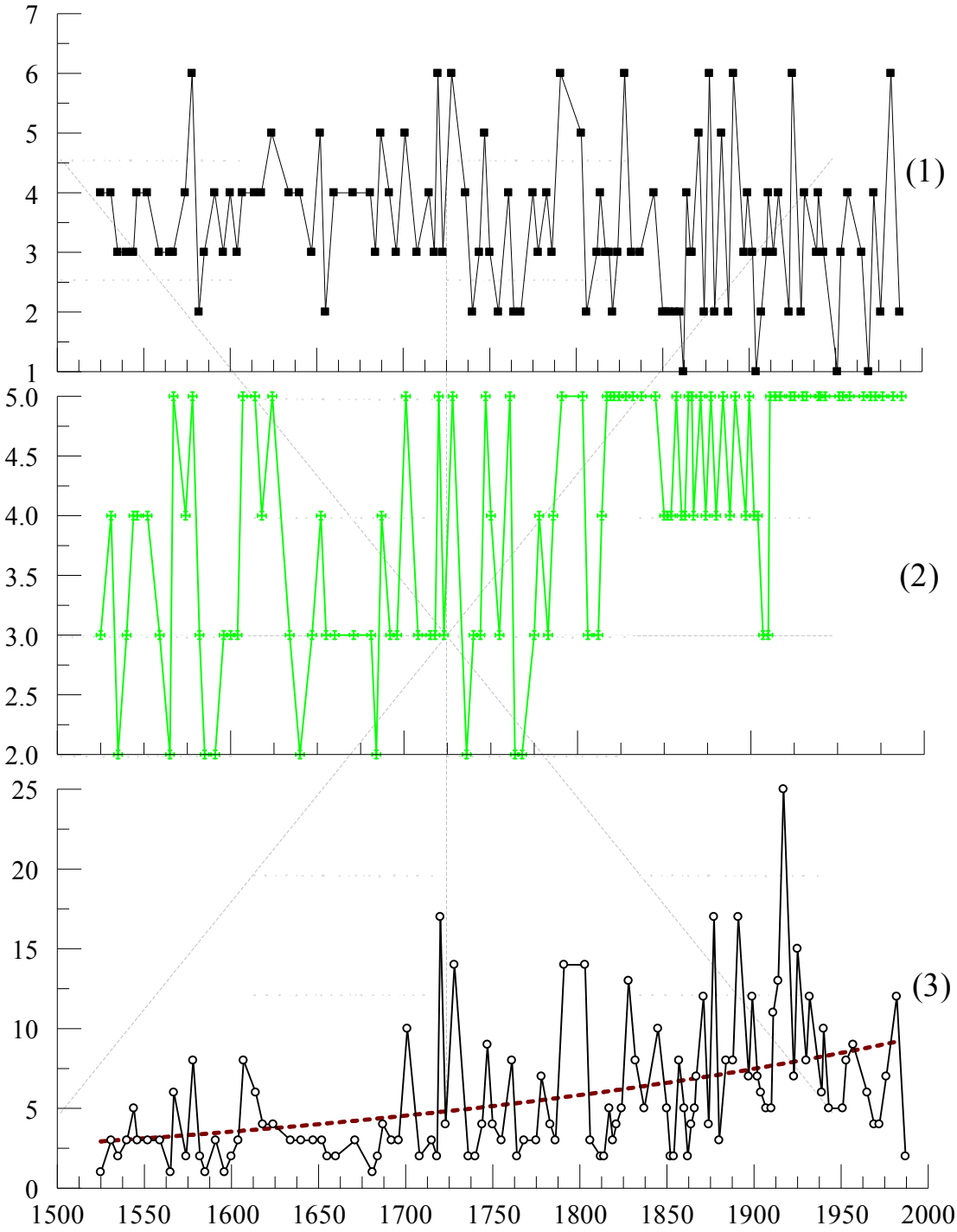


Рис. 50. Динамика интенсивности замеченных проявлений EN (1), достоверность проявлений(2), количество авторов, отмечающих характер проявлений (3)

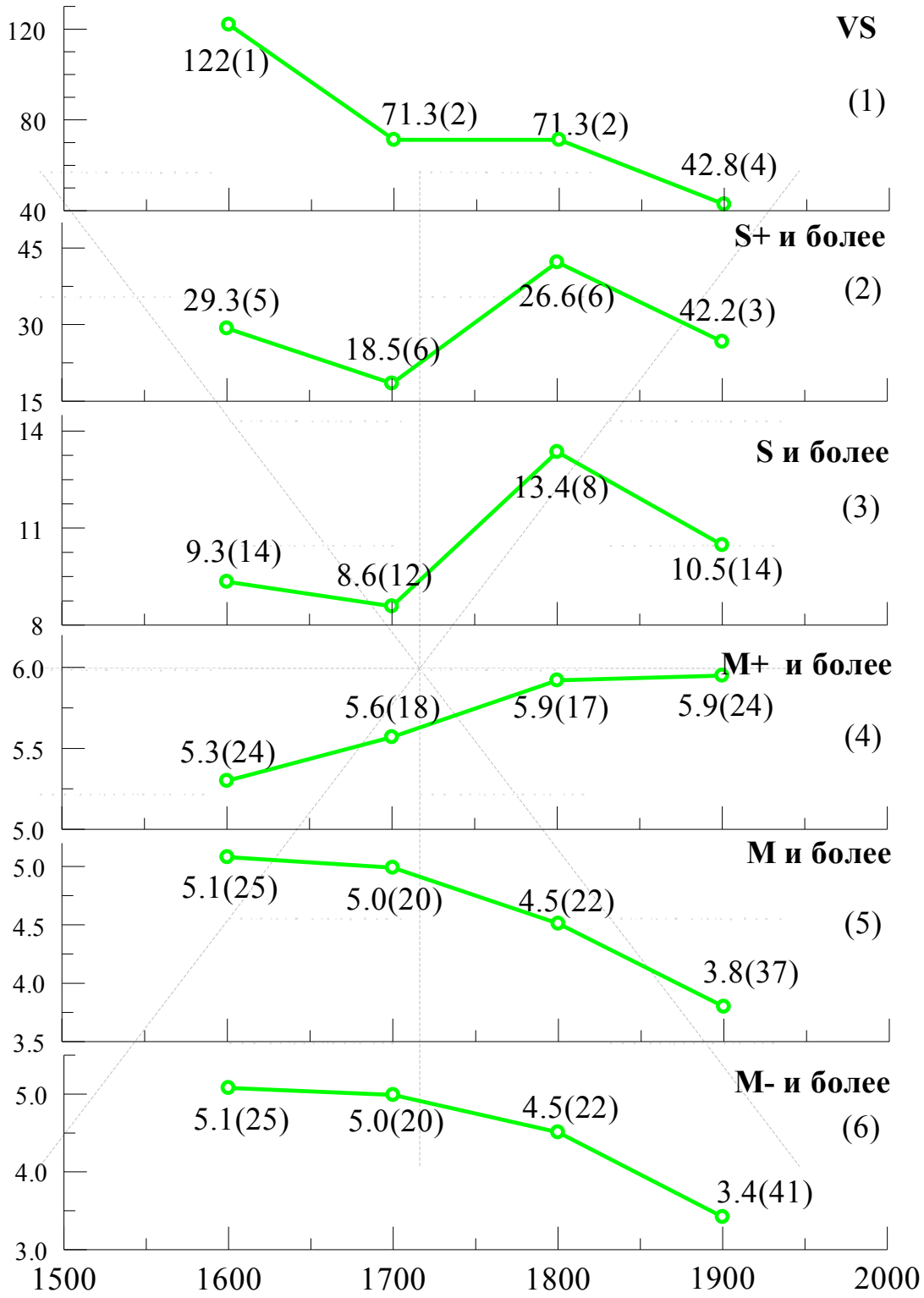


Рис. 51 Средние значения интервалов времени между проявлениями EN различной интенсивности по столетиям (в скобках указано число проявлений)

В работе были использованы опубликованные данные о проявлениях (1525-1982 гг.) Эль-Ниньо разной интенсивности (Quinn, Neal, 1997). Для сопоставления использованы 33 обобщенные древесно-кольцевые хронологии сети дендрохронологических тест-полигонов (станций) в Субарктике Сибири, относительная изменчивость ширины годичных колец, которая на 50-70% определяется изменениями летней температуры (Ваганов и др., 1996; Briffa et.al, 1998a). Сеть станций организована в две широтных полосы: северная - соответствует полярной границе распространения древесной растительности, и более южная - на 200 км южнее (Рис. 52).

При анализе материала использованы два основных метода. Первый - «метод наложенных эпох», в котором величины радиального прироста в обобщенных древесно-кольцевых хронологиях брались во временном окне с центром, который соответствовал году проявления Эль Ниньо определенной интенсивности, и затем усреднялись по годам и участкам (Fritts, 1976). Анализировались изменения индексов прироста за 10 лет до и после года проявления Эль Ниньо. Второй метод – псевдостатистический метод Байеса (Фор, 1989). Этот классификационный метод с обучением был использован для идентификации проявлений Эль Ниньо разной интенсивности по совокупности изменений прироста деревьев на разных участках широтного трансекта.

Напомним, что статистический анализ однородности исторических сведений по проявлениям Эль Ниньо показал: гипотезы о постоянстве вероятности наступления Эль Ниньо и постоянстве его средней интенсивности (по классификации Куина (Quinn) и др., 1997) за период с 1525 до наших дней отвергаются. Либо не все события Эль Ниньо зафиксированы в истории, либо сам процесс не стационарен, т.е. число проявлений EN увеличивается со временем.

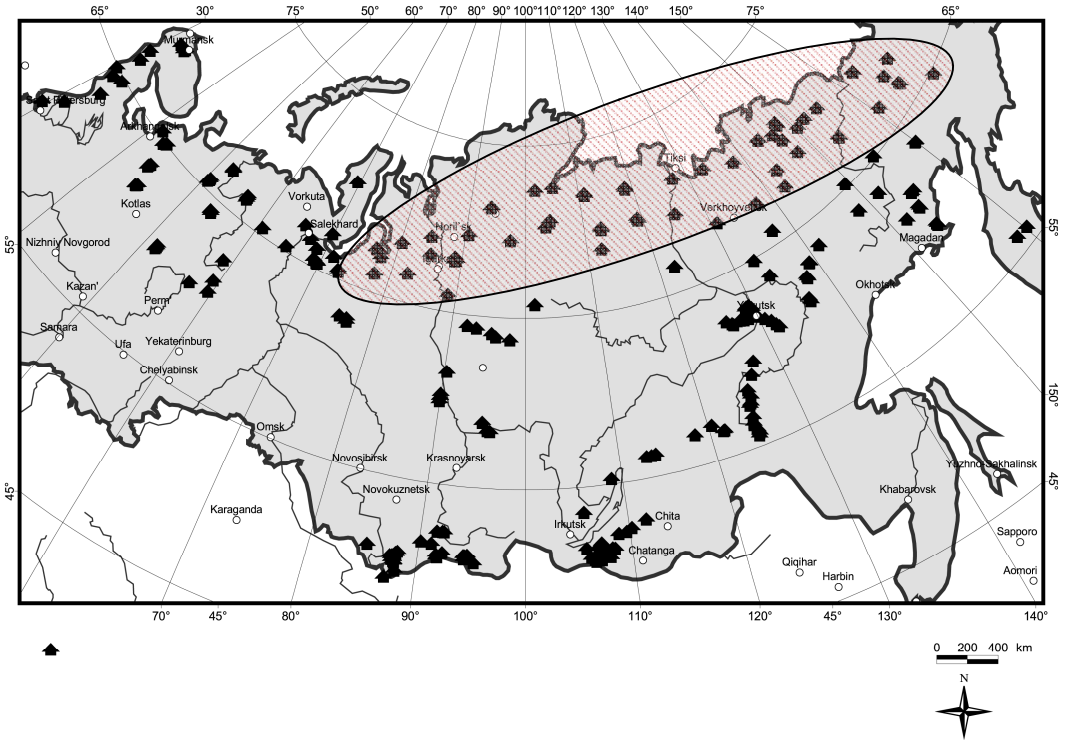


Рис. 52. Карта распределения 285 дендрохронологических станций на территории России. Выделенная (заштрихованная) область на карте представляет собой район исследования

Статистическим критериям однородности удовлетворяет период с 1800 по 1982 год. Этот временной интервал и был взят за основу при анализе (Шишов, Джансеитов, 1999).

Усредненные по годам и участкам кривые изменчивости относительного прироста деревьев получены методом наложенных эпох для проявлений Эль-Ниньо разной интенсивности (Рис. 53). Для сильных и средних интенсивностей проявлений Эль-Ниньо в динамике изменения радиального прироста деревьев до и после события не отмечается каких-либо существенных изменений, значимо отличных от среднего уровня. Для сильнейших и суммы сильнейших и очень сильных проявлений Эль-Ниньо такие значимые отличия присутствуют. Так, прирост за два предшествующих года и особенно за два года после сильнейших проявлений Эль-Ниньо значимо ниже среднего уровня для всей территории Субарктики Сибири.

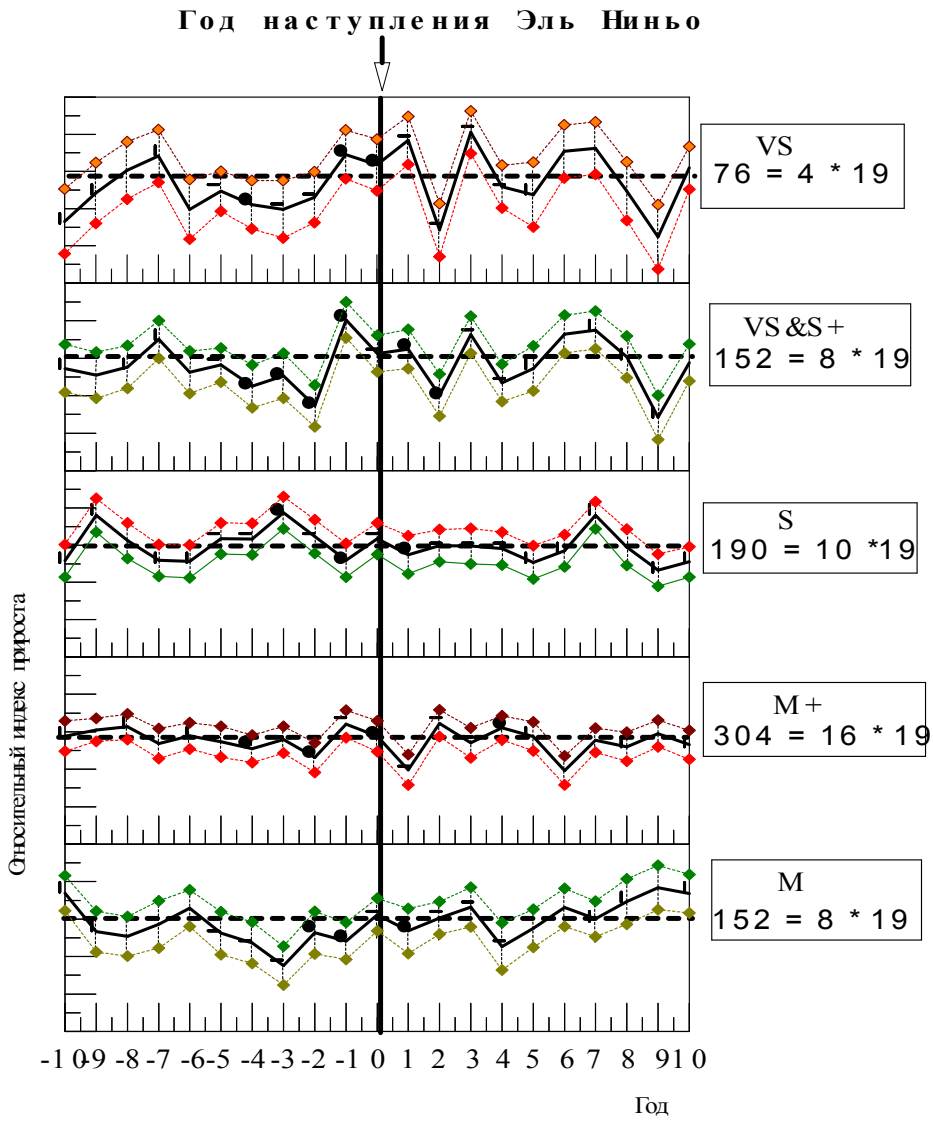


Рис. 53. Характерные средние кривые прироста деревьев, полученные методом наложенных эпох, для следующих проявлений Эль Ниньо: VS - сильнейшие, VS и S+ - сумма сильнейших и очень сильных, S - сильные, M+ - выше среднего, M - средние.

По-видимому, это является следствием некоторого понижения температуры в Субарктики Сибири. Отметим также, на значимые отличия в анализируемых

кривых прироста, возникающие за 8-10 лет и после проявлений Эль-Ниньо анализируемых интенсивностей. Ниже будет показано, что эти различия являются не случайными и обусловлены циклическими колебаниями в проявлениях Эль-Ниньо.

Таким образом, на уровне усредненных величин отмечаются значимые различия между сильнейшими и очень сильными проявлениями Эль-Ниньо, а также их средними проявлениями.

Рассмотрим результаты классификации событий Эль-Ниньо по их интенсивности методом Байеса. Для обучающей выборки, которая включала 19 участков из отобранных 33, равномерно распределенных внутри широтного транссекта, решалась задача разделения сильнейших и очень сильных (VS и S+) от средних (M) проявлений Эль Ниньо. При этом каждое наблюдение (год проявления Эль Ниньо определенной интенсивности) характеризовалось 21-мерным вектором (21 последовательными значениями индексов прироста деревьев обобщенных хронологий для каждого участка за 10 лет до и после события Эль Ниньо). Результаты классификации показали, что из 133 объектов (7 лет х 19 участков) в случае сильнейших и очень сильных проявлений Эль Ниньо правильно классифицируются 127 (или 95,5%), в случае средних - 125 (94%). Среднее качество идентификации или правильного распознавания объектов составляет 95%. Проверка работы классификации на независимой выборке (7 лет и 14 других участков) дала следующие результаты: идентификация сильнейших и очень сильных проявлений Эль-Ниньо составила 84 из 98 (или 85,7%), средних - 83 из 98 (84,7%). Среднее качество идентификации составило 85%.

Отметим также, что при помощи предложенного классификационного подхода можно показать оправданность использования 21-мерных векторов.

Рассмотрим конкретные примеры.

При построении обучающей выборки были использованы стандартные индексные хронологии с 12 самых восточных участков (Рис. 52). В итоге было получено 72 вектора (объекта) различной длины (в зависимости от окна выборки – W = 3, 5, 7, 11, 15, 21, 31), 36 из которых соответствуют точкам максимума и 36 - точкам минимума действия EN.

В качестве контрольной выборки были использованы стандартные индексные хронологии с 9 более западных участков (Рис. 52). В итоге было получено 54 вектора (объекта), 27 из которых соответствуют точкам максимума и 27 - точкам минимума.

Если рассмотреть трехмерные векторы (объекты), то для обучающей выборки качество идентификации 72 объектов характеризуется таблицей:

	A	B	Σ
A	25 69%	11 31%	36 100%
B	15 42%	21 58%	36 100%

со средним значением качества идентификации – 64%.

Качество идентификации контрольной выборки из 54 объектов - таблицей:

	A	B	Σ
A	13 48%	14 52%	27 100%
B	12 44%	15 56%	27 100%

-со средним значением качества - 51,9 % (практически случайная идентификация).

В этом случае решающее правило отнесения объектов к нужному классу по методу Байеса с наименьшей ошибкой работает некачественно.

По мере увеличения W качество классификации улучшается и достигает лучших значений (объекты – 21-мерные вектора), затем постепенно ухудшается, например, в случае W=31.

Так для W=21 соответствующие таблицы и средние значения таковы:

	A	B	Σ
A	36 100%	0 0%	36 100%
B	0 0%	36 100%	36 100%

	A	B	Σ
A	21 78%	6 22%	27 100%
B	0 0%	27 100%	27 100%

- 100% (для обучающей выборки) - 89% (для контрольной выборки).

Для W=31 соответствующие таблицы и средние значения таковы:

	A	B	Σ
A	36 100%	0 0%	36 100%
B	0 0%	36 0%	36 100%

	A	B	Σ
A	11 41%	16 59%	27 100%
B	0 0%	27 100%	27 100%

- 100% (для обучающей выборки) - 70,5% (для контрольной выборки).

Таким образом, 21-мерные объекты при решении данных классификационных задач являются наиболее подходящими.

Решение других классификационных задач по разделению сильнейших и очень сильных проявлений Эль Ниньо от сильных и средних (S, M+, M, M-) показало, что первые идентифицируются по независимой выборке на 89-92%, тогда как вторые - на 61-82%. Таким образом, по данным изменчивости относительного прироста деревьев хорошо идентифицируются годы сильнейших и очень сильных проявлений Эль-Ниньо.

Заключительным этапом статистического анализа с использованием классификационного метода Байеса было использование данных прироста

деревьев на отобранных 33 участках для выявления лет сильнейших и очень сильных проявлений Эль-Ниньо с одной стороны и средних - с другой для периода с 1800 по 1982 гг. Для этого полученный алгоритм классификации применялся к каждому году для указанного временного интервала, и оценивался процент идентификации соответствующего события. Результаты выявления соответствующих событий в рамках алгоритма классификации сравнимы с историческими данными (Рис. 54).

Таким образом, рассмотренные выше результаты можно рассматривать как доказательство наличия статистической связи между очень сильными проявлениями Эль-Ниньо и изменением летней температуры в субарктической области Сибири северного полушария. При этом наиболее значительные изменения летней температуры в субарктике (похолодания) имеют опережение или запаздывание на два года по сравнению с сильнейшими и очень сильными проявлениями Эль-Ниньо.

Наблюдения за циркуляцией в средней и верхней стратосфере отмечают квази-двухлетний (2,2 года) цикл переноса воздушных масс от севера к экватору и обратно (Holton, Tan, 1982). Если рассматривать Эль-Ниньо как сильнейшее термическое возмущение в экваториальной части, то его реакция в субарктической зоне Сибири с временной задержкой (или опережением) в 2 года хорошо согласуется с квази-двухлетним циклом. К.Бриффа (K. Briffa) с соавторами (1998б) нашли статистически значимую связь между крупными вулканическими извержениями и резкими изменениями летней температуры в субарктической области Евразии.

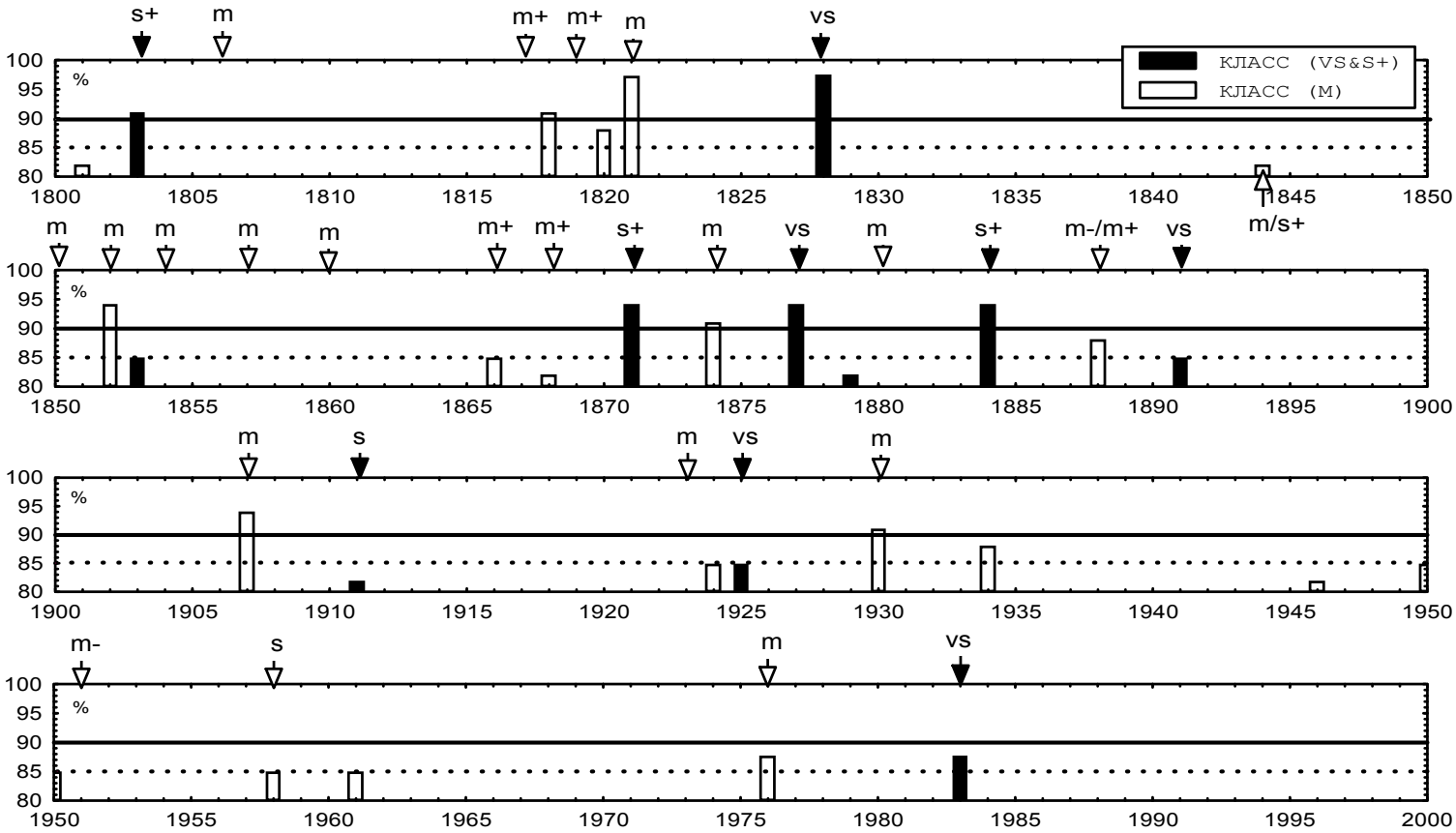


Рис. 54.Пример качества идентификации объектов (лет) при помощи решающей функции, полученной на обучающей выборке. Эта выборка построена по двум классам проявлений Эль Ниньо: (VS и S+) и (M). Стрелками указаны исторические данные. В целом, среднее качество идентификации составило 93% (с учетом лет, не имеющих проявлений Эль Ниньо с указанными выше интенсивностями).

Это еще один механизм квази-периодических вариаций летней температуры в субарктической области, имеющий глобальный характер. Выявление таких механизмов существенно дополняет известные гео- и гелио-физические факторы изменчивости глобального климата Земли.

Выше было сделано предположение о том, что в проявлениях Эль-Ниньо нет циклических закономерностей в изменениях интенсивностей этого природного явления.

Проведем независимый анализ, предполагая, что интенсивности этого глобального явления протекают неслучайно, а изменяются по определенному закону, который может быть определен функционально.

Предположим, что EN, как глобальное явление, протекает во времени непрерывно. Тогда интервалы времени от 2 до 10 лет между проявлениями (фиксациями) EN можно считать как отсутствие наблюдений из-за интенсивностей EN, меньших M-, поэтому имеет смысл заполнить пробелы в кривой динамики интенсивности проявлений EN нулями. В результате мы получим временной ряд или случайную функцию с дискретным аргументом $I_t = I(t)$ (Рис. 55).

Результаты автокорреляционного анализа I_t показывают наличие значимых положительных корреляций при сдвигах на 7-8, 11-13, 26-36, 40-60 , 70-130 лет и отрицательной - при сдвиге на 1-2 года. Цикличности с указанными интервалами известны в литературе. Отрицательная корреляция при сдвиге на 1-2 года может быть объяснена упомянутой выше асимметричностью явления EN.

Практически аналогичные периоды за исключением вековых и сверхвековых цикличностей выявляются при МЭКК анализе (Шишов, 2009). При этом можно отметить хорошее совпадение значимых периодов как для всего периода фиксаций событий EN (с 1525 по 1987 гг.), так и для интервала с 1800 по 1987 (Рис. 56). Это еще раз указывает на хорошую применимость данного метода спектрального анализа при анализе нестационарных данных.

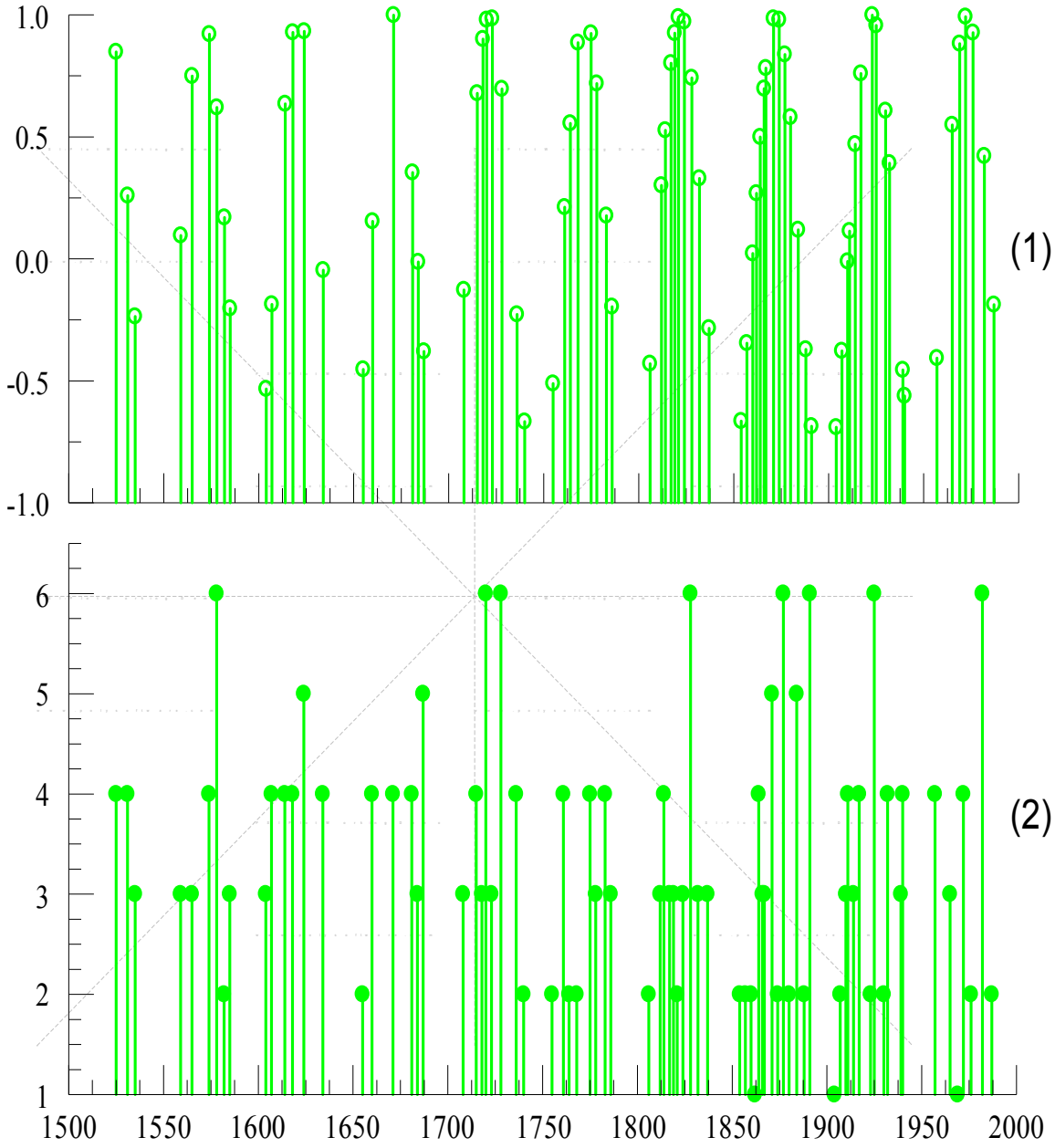
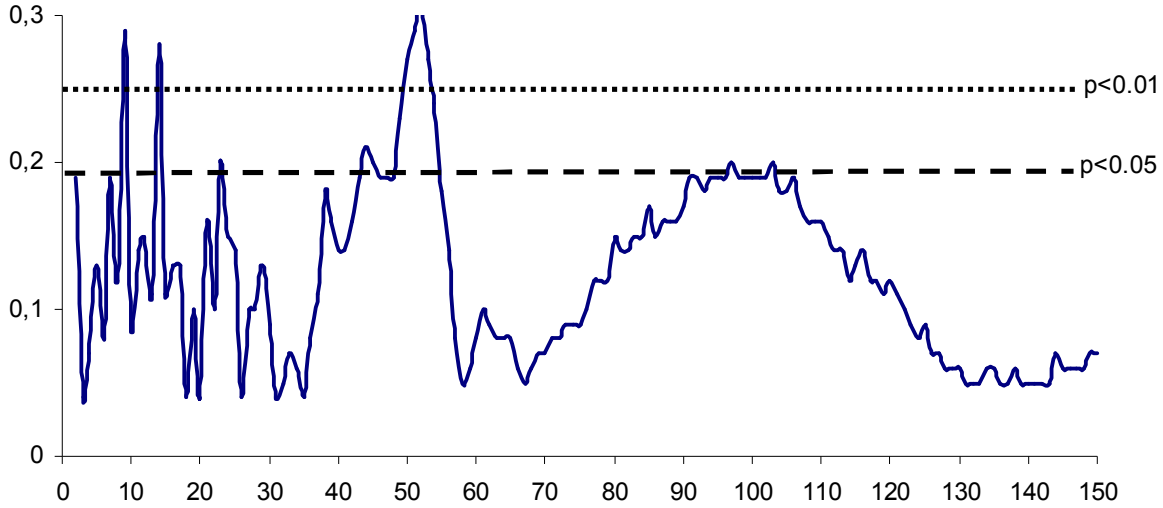
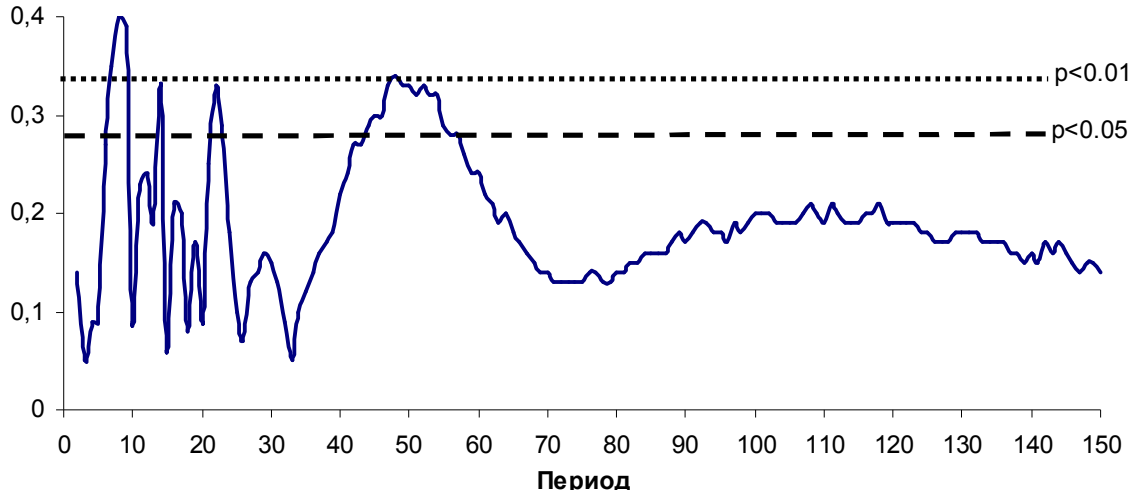


Рис. 55. Динамика преобразованной согласно данным о силе проявлений EN (см. текст) синусоиды sin0.126t (1) и данных о силе EN (2), соответствующих области определения кривой (1)



А)



Б)

Рис. 56. МЭКК-коррелограмма для циклов, выявленных при анализе исходных данных по интенсивности событий Эль-Ниньо для всего периода фиксаций событий EN (А) и с 1800 по 1987 гг. (Б). Пунктирные линии указывают на уровни значимости коэффициентов корреляции.

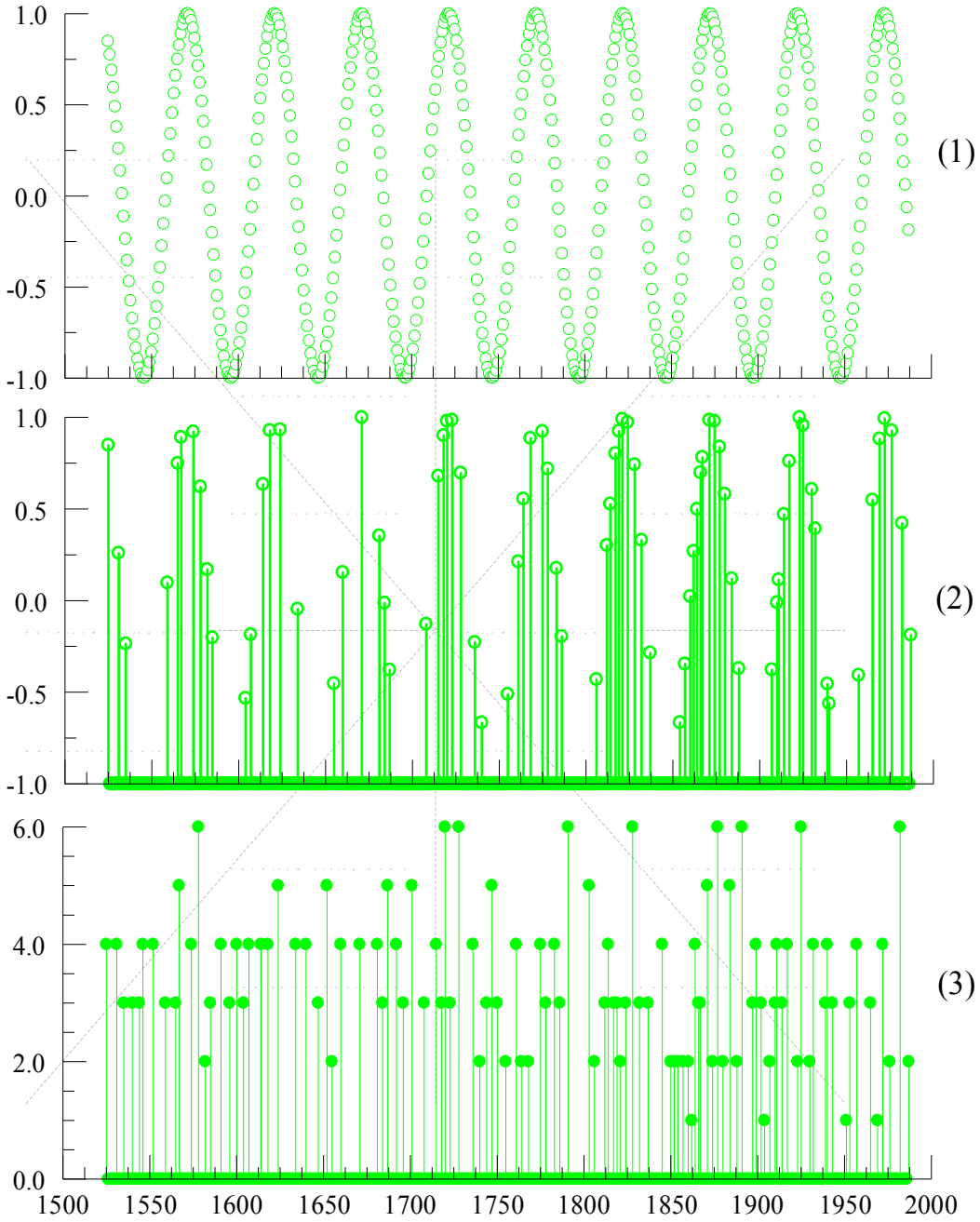


Рис. 57. Динамика синусоиды sin0.126t (1), кривой (2) – производной от 1, кривой (3) – I(t)

Рассмотрим более подробно цикличность в 40-60 лет. Возьмем синусоиду с периодом в 50 лет $S_t = \sin 0.126t$, выбросим значения функции S_t меньше или равных -0.67 и значения в точках t, соответствующих моментам времени отсутствия фиксаций явления EN (моментам нулевых значений I_t), полученную функцию обозначим через $\hat{S}_t$. Пусть $\hat{I}_t$ - функция, полученная из I_t

выбрасыванием значений, соответствующих моментам времени *t*, для которых $\hat{S}_t$ не определена. Будем воспринимать динамику EN во времени как синусоиду, полученную из $\hat{S}_t$ сдвигом вдоль оси ординат и растяжением (Рис. 57). Корреляция между значениями $\hat{I}_t$ и $\hat{S}_t$ оказалась значимой с достоверностью *P*=0.97. Последнее позволяет нам считать, что *в динамике интенсивностей фиксированных проявлений EN значимо присутствует с P=0.97 цикличность с периодом в 50 лет.*

Наличие этой цикличности подтверждают и данные, изображенные на рисунке 57, где кривая (1) - $S_t = \sin 0.126t$, кривая (2) - $\widetilde{S}_t$ получена из S_t приравниванием к -1 тех значений в точках *t*, где $S_t < -0.67$, или где I_t равна 0. Оценка корреляции между I_t и $\widetilde{S}_t$ равна 0.72, т.е. является значимой с достоверностью большей, чем *P*=0.999. Кстати, значимая оценка корреляции между S_t и $\widetilde{S}_t$, полученной из S_t, - всего 0.32, то есть $\widetilde{S}_t$ ближе к I_t, чем к $S_t = \sin 0.126t$, притом, что значимой корреляции между I_t и S_t не существует ($\rho = 0.03$) (последнее не противоречит выявленной выше цикличности у I_t в пределах 40-60 лет).

Проанализируем дендрохронологическими данными с 1790 по 1990 гг., с участков расположенных вдоль северной границы распространения лесов Сибири и Дальнего Востока при условии существования циклических колебаний в проявлениях событий Эль-Ниньо (Джансеитов, Шишов, 1999).

Интервал 1790-1990 гг. соответствует интервалу наиболее достоверных данных по EN и интервалу с наиболее полными дендрохронологическими данными на исследуемых участках.

Рассмотрим самые восточные 11 участков.

Путем скользящего сглаживания кривой центрированных значений интенсивности фиксированных проявлений EN с предварительным переносом нулей I_t на уровень среднего значения 3.44 с окном в 11 и 21 год были выявлены точки максимумов : 1825, 1875, 1925 гг., - и точки минимумов: 1855, 1905, 1950 гг. сглаженных кривых (1), (2) (Рис. 58) на интервале с 1800 по 1975 год,

что соответствует окрестностям точек максимумов и минимумов синусоиды $S_t = \sin 0.126t$ и значениям интенсивности VS и M-, соответственно.

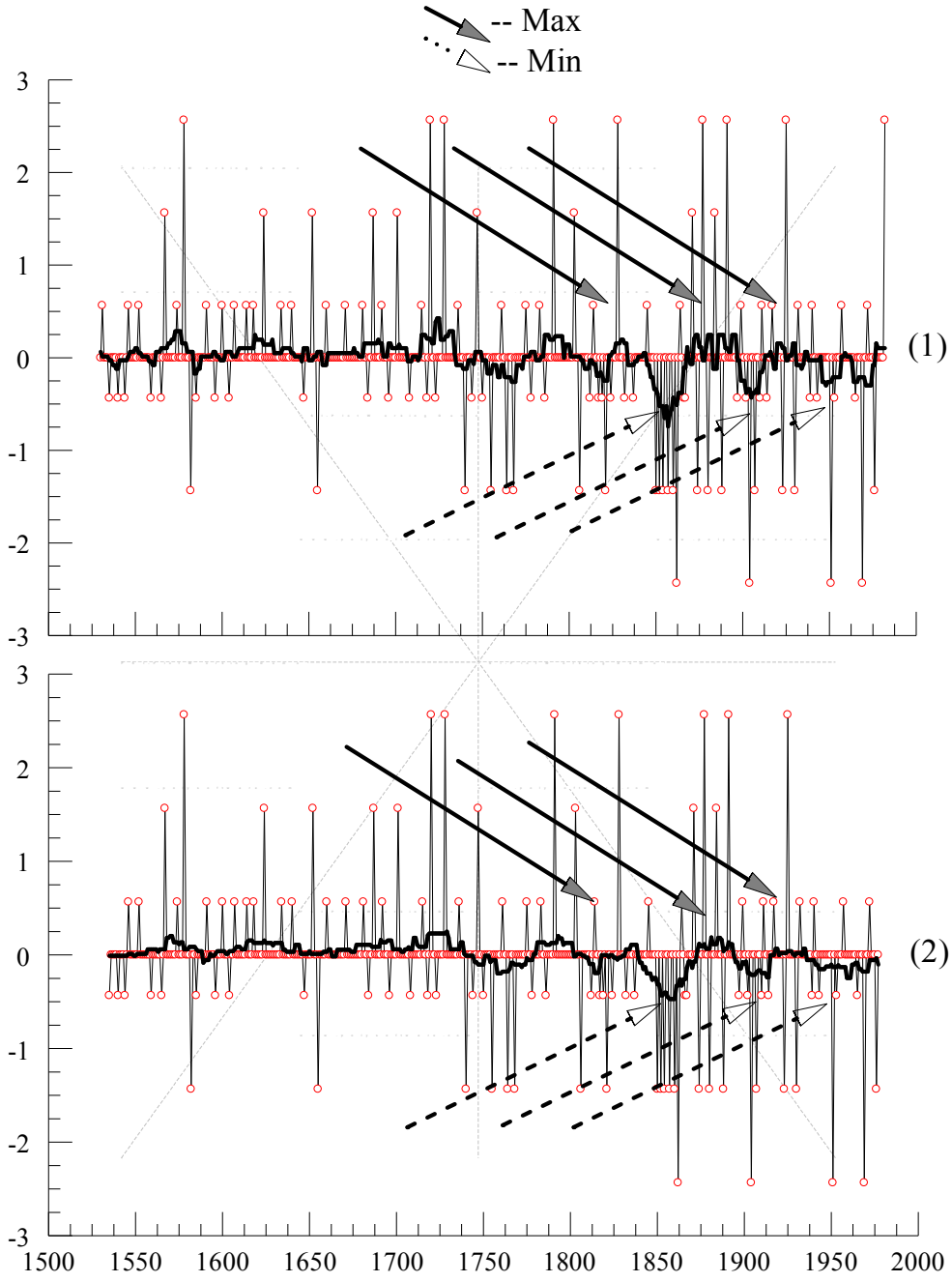


Рис. 58. Кривые скользящих средних с окном в W=11 (1) и W=21 год (2) центрированных данных по интенсивности EN с указанием точек максимумов (Max) и минимумов (Min) на интервале 1800-1975 гг.

Отметим, что коэффициент согласованности (конкордации) Кендалла, вычисляемый с окном в 21 год, имеет высокую отрицательную корреляцию с условиями произрастания (Джансеитов, Шишов, 1993; Шишов и др., 2007б).

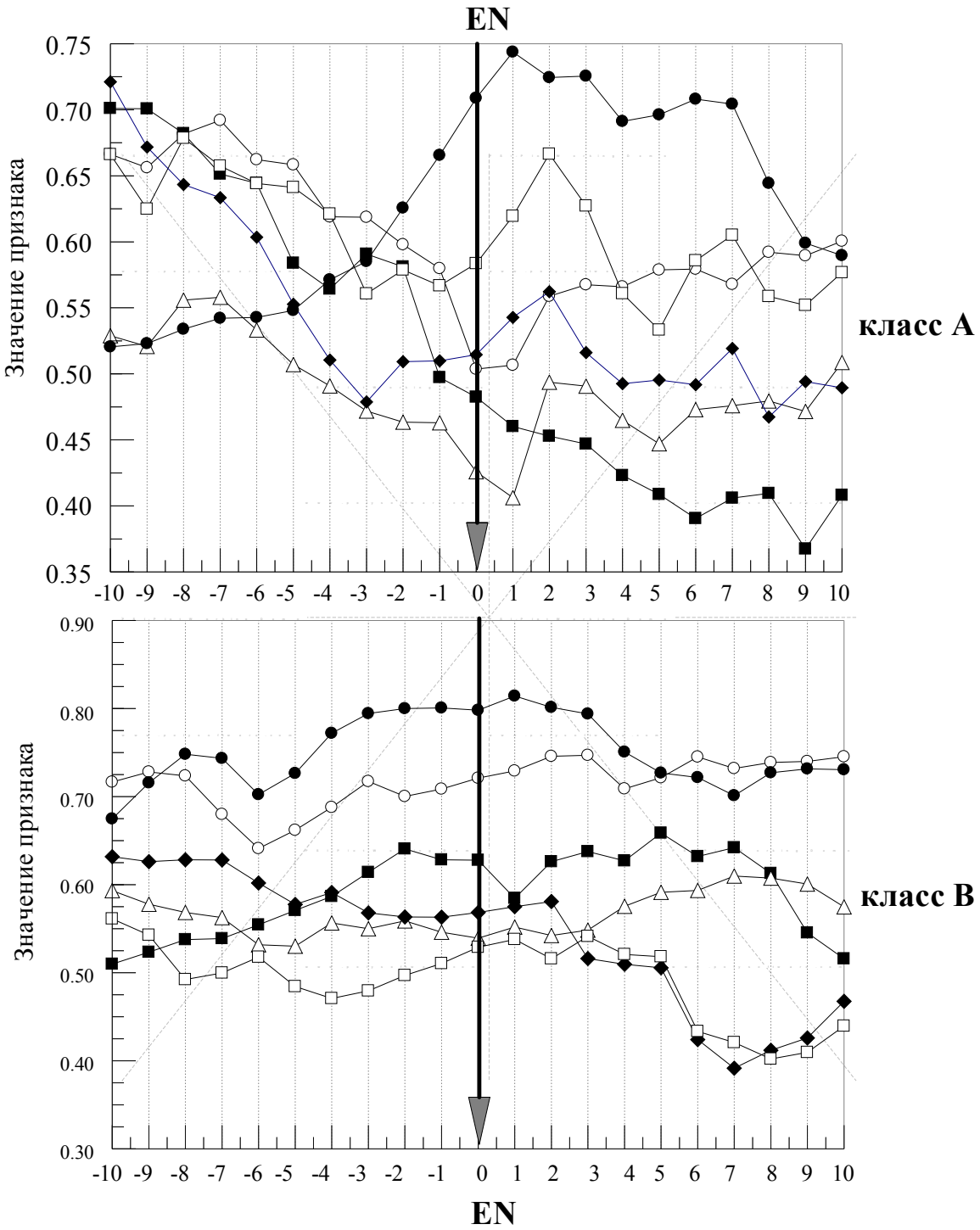


Рис. 59. Примеры типичных 21-мерных объектов из классов A (максимумов проявления EN) и B (минимумов). Компонента 0 соответствует моменту проявления EN той или иной сглаженной интенсивности.

В предположении, что проявления EN сказываются на условиях произрастания деревьев на выбранных участках, было решено для каждого указанного выше года *t* (точек экстремумов) и каждого участка рассмотреть 21-мерный вектор, составленный из 21 значения коэффициентов конкордации,

вычисленных для моментов времени $t-10, t-9, \ldots, t, \ldots, t+10$. В итоге было получено 66 векторов (объектов), 33 из которых соответствуют точкам максимумов и 33 - точкам минимумов, т.е. два класса объектов: класс максимумов (класс А) и класс минимумов (класс В) (Рис. 59).

В результате применения псевдостатистического метода идентификации Байеса была получена решающая функция, представляющая собой отношение двух функций плотностей распределения классов А и В (отношение правдоподобия Байеса) - выборок двух 21-одномерных случайных величин с предполагаемыми 21-мерными нормальными законами распределения.

Качество идентификации 66 объектов (так называемой обучающей выборки) при помощи найденной решающей функции характеризуется таблицей:

	A	B	Σ
A	33 100%	0 0%	33 100%
B	0 0%	33 100%	33 100%

- матрица порядка 2×2 вероятностей ошибок $P = \left\| p_{ij} \right\|$, где p_{ij} - вероятность отнесения объекта из i-го класса в j-ый класс является единичной, т.е. идентификация оказалась идеальной (100-процентной).

Для проверки полученного алгоритма идентификации были взяты дополнительно 60 объектов (21-мерных векторов), соответствующих 10-ти более западным участкам и 6-ти найденным выше точкам максимумов и минимумов (контрольная выборка).

Качество идентификации контрольных 60 объектов при помощи найденной выше решающей функции характеризуется таблицей:

	A	B	Σ
A	25 83.3%	5 16.7%	30 100%
B	0 0%	30 100%	30 100%

- т.е. средняя вероятность правильной идентификации объектов контрольной выборки объема 60 равна 91.65 % (т.е. $(p_{11} + p_{22})/2 = (83.3 + 100)/2$).

Если осуществить произвольное случайное разбиение обучающей выборки объемом 66 на два класса, при помощи метода Байеса найти решающую функцию и идентифицировать нашу контрольную выборку объемом 60, то качество идентификации для каждого такого разбиения характеризуется таблицами вида:

	A	B	Σ
A	4 13.3%	26 86.7%	30 100%
B	1 3.3%	29 96.7%	30 100%

- т.е. для данного одного из случайных разбиений среднее качество идентификации объектов контрольной выборки оказалась равной 55 % (идентификация практически отсутствует).

В данном разделе в качестве обучающей выборки были взяты объекты, соответствующие «восточным» участкам России, а контрольной – объекты, соответствующие более западным. Рассмотрим вариант, когда обе выборки осуществляются равномерно по имеющимся участкам всей циркумполярной области. Обучающая выборка состоит из 120 объектов (21-мерных векторов), соответствующих 20 участкам под номерами 1, 2а, 7, 8, 10, 11, 14, 15, 16, 17, 20, 24, 26, 30, 34, 35, 36, 37, 45, 46 и прежним 6-ти точкам максимумов и минимумов проявления EN.

Качество идентификации 120 объектов (обучающей выборки) при помощи найденной решающей функции характеризуется таблицей:

	A	B	Σ
A	57 95%	3 5%	60 100%
B	0 0%	60 100%	60 100%

- среднее качество идентификации 97,5%.

При идентификации 102 объектов контрольной выборки, отвечающих 17 участкам под номерами 4, 5, 6, 9, 12, 13, 14а, 18, 19, 23, 25, 29, 30а, 34а, 35а, 43, 44 и прежним точкам максимума и минимума, качество оказалось следующим:

	A	B	Σ
A	46 90,2%	5 9,8%	51 100%
B	4 7,8%	47 92,2%	51 100%

- т.е. среднее качество идентификации – 91,2%.

Рассмотрим вопрос о возможности реконструкции степени проявления EN по дендрохронологическим данным более ранних лет на базе использования решающих функций, полученных по обучающей выборке объектов замечания 3.

Качество идентификации контрольной выборки из 76 объектов (21-мерных векторов), соответствующих 19 дендрохронологическим участкам двум точкам максимума проявления EN (1730, 1785 гг.) и двум точкам минимума (1760, 1805 гг.), характеризуется таблицей:

	A	B	Σ
A	25 65,9%	13 34,1%	38 100%
B	7 17,8%	31 82,2%	51 100%

- среднее качество идентификации 74%. Более низкое значение качества идентификации можно объяснить, например, доказанной в п.2.1 низкой надежностью наблюдений проявления EN с 1525 по 1800 гг.

В заключение оценим информативность значений компонент 21-мерных векторов (при их идентификации), используя метод Г.С. Лбова (1981) случайного поиска с адаптацией – СПА (Таб. 6).

Таблица 6.

Оценка информативности 21 признака по методу СПА

Номер компонента	Информативность (в %) при выборе	
	10 компонент	17 компонент
-l0 (10 год до и наступления события EN)	0,0	3,7
-9 (9 лет до наступления события EN	4,2	3,0
-8	15,0	5,4
-7	0,0	7,2
-6	8,3	0,0
-5	0,0	5,6
-4	0,0	6,4
-3	0,0	0,0
-2	12,4	4,5
-1	0,0	0,0
0 (год наступления события EN)	0,0	4,1
1 (1 год после наступления события EN)	0,0	3,9
2 (2 год после наступления события EN	0,0	4,2
3	8,3	7,4.
4	0,0	5,2
5	4,5	5,8
6	0,0	10,4
7	16,8	12,0
8	11,9	0,0
9	16,6	5,9
10 (10 год после наступления события EN)	2,0	5,2

Количество случайным образом выбранных признаков (компонент) в методе СПА было равным в одном случае 10, в другом – 17. Величина «поощрения» и «наказания» на каждом шаге алгоритма, состоящим из 40

случайных выборок указанного выше числа компонент, дающих соответственно максимальное и минимальное значение F-критерия, была равной 0,0005.

Результаты, представленные в таблице 6, показывают информативность использованных признаков (значений коэффициентов конкордации) за 2 года до и 3 года после проявлений событий Эль-Ниньо, что полностью согласуется с результатом, полученным методом наложенных эпох при анализе стандартных древесно-кольцевых хронологий.

Особо можно отметить информативность значений коэффициентов конкордации за 8-9 лет до и после наступлений событий Эль-Ниньо. Это является еще одним обоснованием использования 21-мерных векторов при изучении 50-летних проявлений EN.

Таким образом, изложенные выше результаты можно рассматривать как доказательство наличия статистической связи между проявлениями Эль Ниньо различной интенсивности, условиями роста древесных растений, а следовательно, и изменением летней температуры в субарктической области Сибири северного полушария.

П.11. Пример использования кластерного и дискриминантного анализов при пространственно-временном анализе трендов прироста древесных растений на территории Сибири и Дальнего Востока.

Для расчёта трендов прироста был задействован Урало-Сибирский дендрохронологический банк данных. Для оценки трендов использовались 154 длительных древесно-кольцевых хронологии, полученные для сети дендроклиматических станций Сибири (Шишов и др., 2002) (Рис. 60).

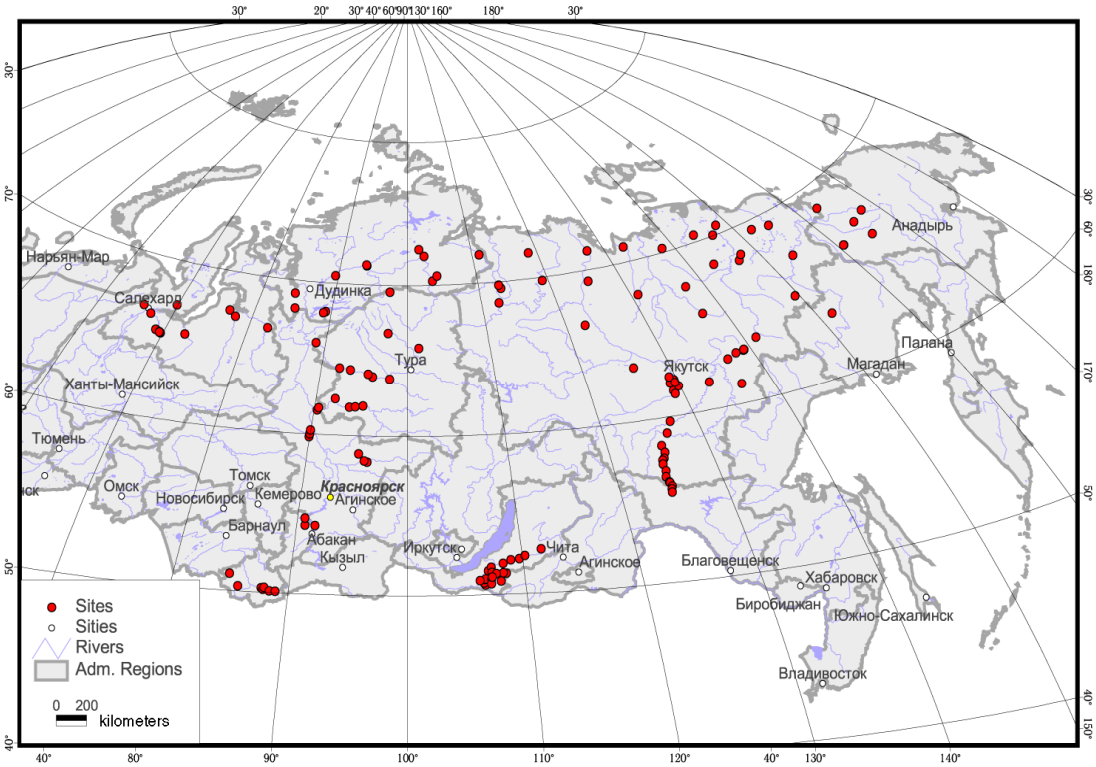


Рис. 60 Расположение дендрохронологических участков, использованных для расчета трендов прироста.

Все хронологии получены для хвойных видов деревьев: по лиственнице -101, по сосне - 26, по ели - 27.

Так как в рядах изменчивости абсолютных величин прироста деревьев содержатся самые различные неклиматические сигналы (возрастные изменения, влияние почвенно–грунтовых условий, конкурентные взаимоотношения, воздействие различных катастрофических факторов и др.), то в дендроклиматологии разработана специальная методика, позволяющая исключать или, по крайней мере, сильно снижать влияние неклиматических факторов при помощи стандартизации или индексирования исходных временных рядов абсолютного прироста древесных растений. Это даёт возможность сопоставлять изменчивость прироста у хронологий, полученных для различных пород деревьев, условий местообитания и районов (Fritts, 1976; Cook, Kairiukstis, 1990).

Длительность 60-ти хронологий достигала 1998-2000 гг., остальные заканчивались 1991-1995 гг. При стандартизации (устранении влияния возраста на ширину годичных колец) использован единый для всех индивидуальных

серий измерений прироста метод аппроксимации возрастных изменений негативной экспонентой, реализованный в дендрохронологическом пакете программ (DPL – Dendrohronology program library) (Cook, Kairiukstis, 1990). Как уже отмечалось выше, этот метод стандартизации может искажать динамику индивидуальных серий прироста, особенно в последние годы роста дерева. Поэтому, предварительно, на основе визуального анализа, были отобраны только те дендрохронологические ряды, для которых такой способ аппроксимации возрастного тренда был наиболее оптимальным (т.е. максимизирующий коэффициент детерминации) из всех используемых в дендрохронологии (Шишов и др., 2002; Шишов, Рубцов, 2006).

В основу оценки трендов прироста древесных растений положен метод наименьших квадратов. В качестве оценочной функции для значений стандартизированного дендрохронологического ряда y_t использована функция вида:

$$\hat{y}_t = at + b$$

Задача сводится к нахождению неизвестных параметров a и b на основе минимизации следующего функционала:

$$F = \sum_{t=1}^{n} (y_t - \hat{y}_t) = \sum_{t=1}^{n} (y_t - at - b)^2$$

Параметры a и b рассчитывались для различных скользящих временных интервалов (временных окон) стандартизированных хронологий за последние 250 лет (1750-2000 гг.). Отметим, что данные параметры тренда приписывались центру (году) скользящего временного интервала. Выбор 250-летнего интервала обусловлен тем, что 95% анализируемых древесно-кольцевых хронологий покрывают данный промежуток времени. На рисунке 61 изображен пример линейного тренда, выделенного для древесно–кольцевой хронологии на временном интервале с 1925 по 1945 гг.

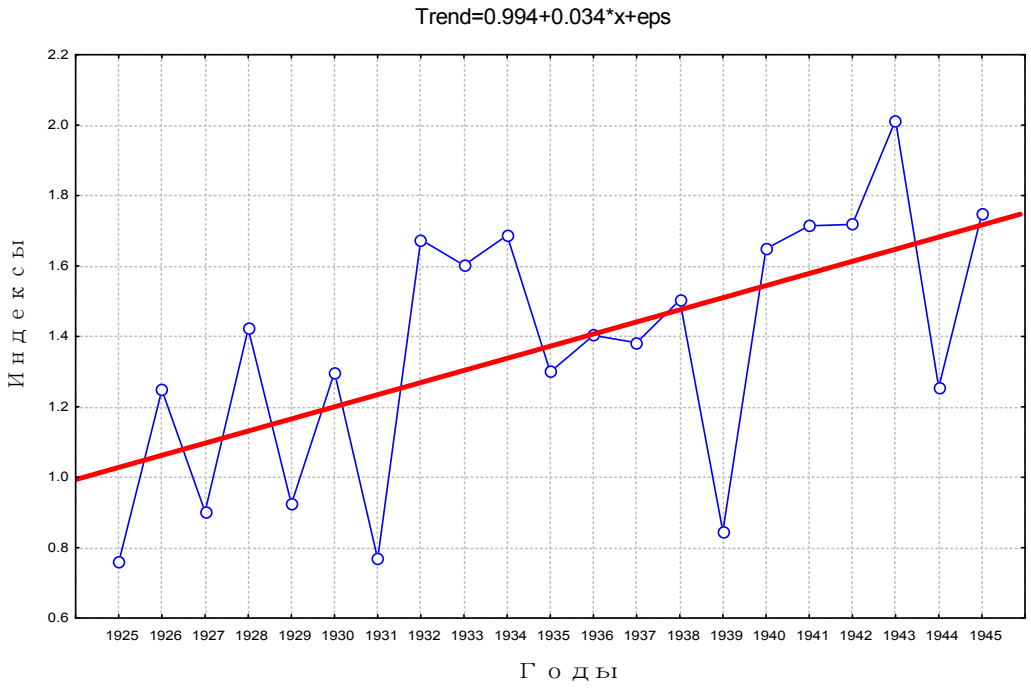


Рис. 61. Пример типичного 19-летнего тренда древесно–кольцевой хронологии

Программа оценки трендов прироста была написана в среде разработки (на языке программирования) Delphi 5 (Рис. 62).

На вход программы подаётся файл, содержащий стандартизированные осреднённые хронологии, полученные с помощью дендрохронологического пакета программ (Таб. 7).

Заголовок входного файла (первая строка) содержит названия древесно – кольцевых хронологий. Они совпадают с названиями всех *.rwl файлов, участвующих в анализе. В первом столбце находятся годы хронологий. В нашем случае – с 1300 по 1999 гг. Каждый последующий столбец - это обобщённая древесно-кольцевая хронология, построенная для соответствующего дендрохронологического участка, обозначенного соответствующим уникальным кодом в дендрохронологическом банке данных. Значение 9,000 обозначает данных

Таблица 7.

Формат входного файла, состоящего из древесно-кольцевых хронологий

YEARS	CHA	KHE	SID	SOL	.....	.....	RAS	KAM

1300	9,000	9,000	9,000	9,000	.....	.....	0,603	9,000
1301	9,000	9,000	9,000	9,000	.....	.....	0,906	9,000
1302	9,000	9,000	9,000	9,000	.....	.....	1,337	9,000
1303	9,000	9,000	9,000	9,000	.....	.....	1,456	9,000
.....	.....	.....	.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....	.....	.....	.....
1762	1,409	9,000	1,235	1,239	.....	.....	1,530	0,448
1763	1,289	9,000	0,988	1,162	.....	.....	1,369	0,804
1764	1,196	0,737	0,874	0,776	.....	.....	1,450	0,580
1765	1,222	0,900	0,977	1,039	.....	.....	1,283	0,943
.....	.....	.....	.....	.....	.....	.....	.....	.....
.....	.....	.....	.....	.....	.....	.....	.....	.....
1996	0,536	0,819	0,717	0,886	.....	.....	1,030	0,69
1997	0,697	1,087	0,728	9,000	.....	.....	1,179	0,62
1998	0,876	0,865	9,000	9,000	.....	.....	9,000	0,77
1999	9,000	9,000	9,000	9,000	.....	.....	9,000	0,96

В выходном файле каждый год (центр соответствующего окна скольжения) расчёта представлен двумя строками – в первой строке – значениями коэффициентов *a*, во второй - значениями *b* по всем участкам (Таб. 8). Число -999 означает отсутствие данных в году на данном участке.

Таблица 8.

Формат выходного файла с коэффициентами a и b

		CHA	KHE	SID	.....
1781	a	-999,0	-999,0	-999,0	.....
1781	b	-999,0	-999,0	-999,0	.....
1782	a	0,040	0,041	0,032	.....
1782	b	1,003	0,945	2,015	.....
1783	a	0,012	0,014	0,016	.....
1783	b	3,750	6,548	1,280	.....
.....	.....	.....	.....	.....	.....

Программа рассчитывает также оценки коэффициентов корреляции между годами (объектами), которые характеризовались 154 значениями параметра *a*. (Таб. 9). Это необходимо для оценки автокорреляционной структуры, заложенной в алгоритм расчета. Отметим, что именно тангенс угла наклона прямой (параметр *a*) является наиболее информативной характеристикой линейного тренда.

Таблица 9.

Формат выходного файла с коэффициентами корреляции

	1750	1751	1752	1753	1754	1755	1756	1757	.....
1750	1.00	0.96	0.89	0.77	0.72	0.66	0.56	0.42	.....
1751	0.96	1.00	0.97	0.87	0.82	0.75	0.64	0.46	.....
1752	0.89	0.97	1.00	0.94	0.87	0.82	0.71	0.51	.....
1753	0.77	0.87	0.94	1.00	0.96	0.93	0.85	0.67	.....
1754	0.72	0.82	0.87	0.96	1.00	0.98	0.92	0.78	.....
1755	0.66	0.75	0.82	0.93	0.98	1.00	0.96	0.84	.....
1756	0.56	0.64	0.71	0.85	0.92	0.96	1.00	0.92	.....
1757	0.42	0.46	0.51	0.67	0.78	0.84	0.92	1.00	.....
.....	.....	.....	.....	.....	.....	.....	.....	.....	.....

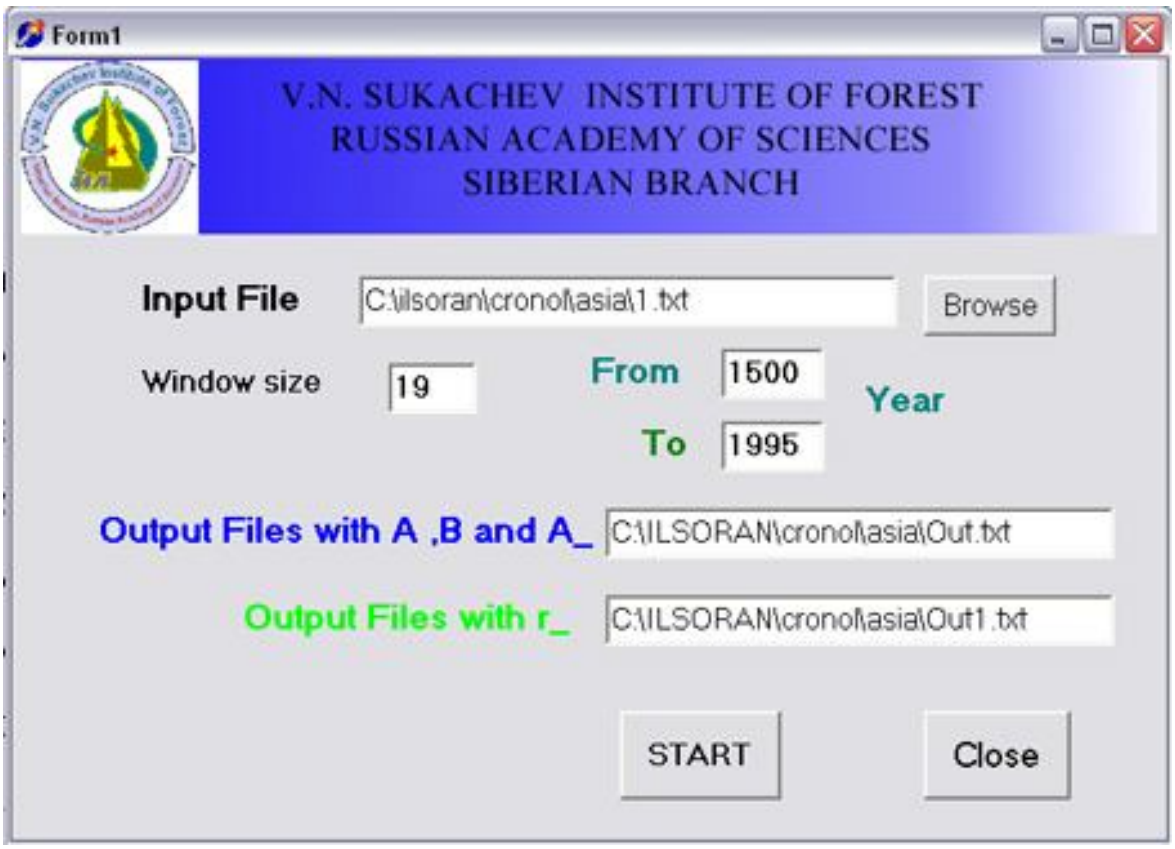


Рис. 62. Интерфейс программы расчёта трендов

Остановимся более подробно на интерфейсе программы расчета трендов прироста древесно-кольцевых хронологий (Рис. 62). К основным опциям программы относятся следующие:

Input File – Входной файл данных с хронологиями можно задать вручную, указав логический диск, путь и имя файла с расширением в строке ввода, либо воспользоваться стандартным диалоговым окном Windows открытия файла при помощи кнопки **Browse** .

Window size – размер окна.

From ___Year – начальный год расчёта

To ___Year – конечный год расчёта

Output Files with A, B and A_ - указывается логический диск, путь и имя (с расширением) выходного файла с коэффициентами a, b и a – средним

Output Files with r_ - указывается логический диск, путь и имя (с расширением) выходного файла с коэффициентами кореляции годов.

Кнопка **Start** – запуск процесса расчёта

Кнопка **Close** – выход из программы

Рассмотрим пример расчета «скользящих» трендовых параметров *a* и *b*. Допустим, начальный год расчёта 1763, размер окна равен 3. Следовательно, в расчёте 3-летнего тренда некоторой древесно-кольцевой хронологии будут взяты 3 значения индексов прироста y_t, соответствующие следующим годам *t*: 1762, 1763, 1764. Далее, центральный год расчёта увеличивается на 1. После чего процедура расчета повторяется до года расчёта, заданного пользователем.

Таким образом, тренды прироста были получены для всех 154 участков и для каждого 19-летнего интервала с 1750 по 1990 гг. со сдвигом на 1 год, то есть был использован аналог «скользящей» характеристики с окном в 19 лет. Выбор окна в 19 лет был обусловлен двумя причинами. Во-первых, необходимо было охватить временной интервал с 1981 по 2000 год (центр интервала - 1990 год), для которого построена карта трендов NDVI с целью дальнейшего сопоставления этих данных с картами трендов, полученными по дендрохронологическим данным. Во-вторых, известно, что прирост древесных растений на большей части территории Сибири лимитируется изменениями температуры (Ваганов и др., 1996; Ваганов, Шиятов,1999; Coок, Kairiukstis, 1990). Именно, в температурных данных присутствуют 20-30 летние периоды похолодания и потепления, особенно ярко выраженные в XX веке.

Анализ корреляционной матрицы (Таб. 10), построенной для всех лет анализируемого временного интервала, начиная с 1750 года, выявил существование блоков 5×5 клеток, «нанизанных» на главную диагональ, с коэффициентами корреляции близкими к 1.

Этот результат свидетельствует о наличии высокой автокорреляционной связи 5-го порядка между годами (центров 19-летних скользящих интервалов), для которых были рассчитаны тренды.

В связи с этим, было решено из 240 лет (240 19-летних интервалов) включить в анализ каждый 5 год (интервал), начиная с 1750, т.е. 1750 (соответствующий интервал - 1741-1759), 1755 (1746-1764), 1760 (1751-1769), 1765 (1756-1774),..., 1900 (1891-1909), 1905 (1896-1914), 1910 (1901-1919),..., 1990 (1981-1999). Это было сделано с целью сокращения анализируемых объектов (лет) без потери информации. Всего в дальнейшем анализе было использовано 49 лет.

Таблица 10.

Часть корреляционной матрицы, построенной для всех лет анализируемого временного интервала, начиная с 1750 года

	Y1750	Y1751	Y1752	Y1753	Y1754	Y1755	Y1756	Y1757	Y1758
Y1750	1,00	0,98	0,94	0,84	0,80	0,75	0,62	0,46	…
Y1751	0,98	1,00	0,98	0,88	0,86	0,80	0,67	0,48	…
Y1752	0,94	0,98	1,00	0,95	0,92	0,86	0,73	0,55	…
Y1753	0,84	0,88	0,95	1,00	0,97	0,93	0,83	0,67	…
Y1754	0,80	0,86	0,92	0,97	1,00	0,98	0,89	0,75	…
Y1755	0,75	0,80	0,86	0,93	0,98	1,00	0,95	0,84	…
Y1756	0,62	0,67	0,73	0,83	0,89	0,95	1,00	0,92	…
Y1757	0,46	0,48	0,55	0,67	0,75	0,84	0,92	1,00	…
Y1758	…	…	…	…	…	…	…	…	…

Следующая задача заключалась в нахождении такого вида нелинейной регрессионной функции ***F***:

$$a_{tk} = F_t(Lat_k, Lon_k, Alt_k),$$

где $\boldsymbol{a_t}$ – значения тренда в t-год для k-участка (k=1,…,154),

$\boldsymbol{Lat_k}$– значения широты для k-участка,

$\boldsymbol{Lon_k}$ – значения долготы для k-участка,

$\boldsymbol{Alt_k}$ – значения высоты над уровнем моря для k-участка, -

которая адекватно описывала пространственное распределение трендов для всех выделенных временных интервалов.

В результате постепенного усложнения вида искомой функции $\boldsymbol{F}$, путем введения дополнительных нелинейных слагаемых при помощи нелинейной покомпонентной регрессии, заключающейся в нелинейном преобразовании независимых переменных, искомый оператор принял следующий вид:

$$Trend_{tk} = b_1 Lat_k + b_2 Lon_k + b_3 Alt_k + b_4 Lat_k^{\ 3} + b_5 Lon_k^{\ 3} + b_6 Alt_k^{\ 3} + const$$

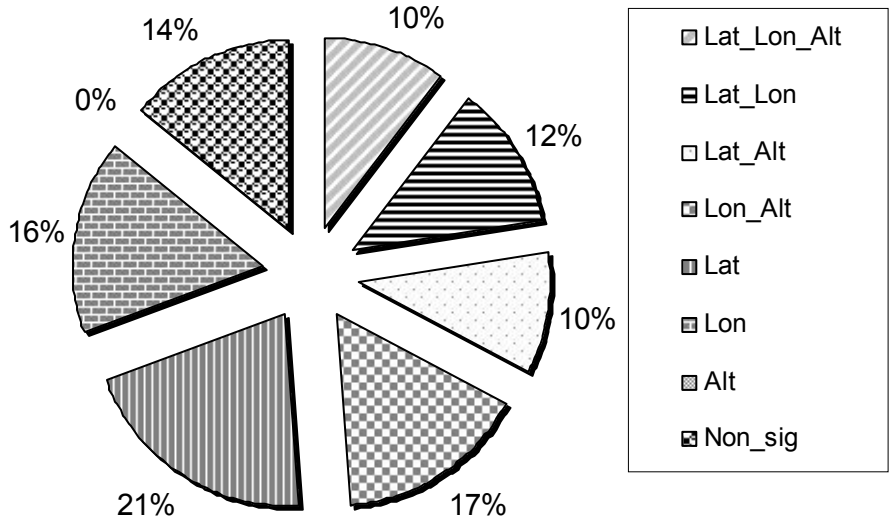


Рис. 63. Распределение значимости независимых переменных по годам (100% составляет 49 выбранных лет (интервалов)).

Такие уравнения были получены для всех 49 анализируемых интервалов с целью дальнейшего построения пространственных карт изменчивости прироста деревьев за последние 250 лет.

Анализ частных коэффициентов корреляции выявил наличие изменения значимости независимых переменных, используемых в уравнении, со временем. Так, широта, долгота, высота и их нелинейные преобразования (Lat_Lon_Alt) в совокупности являются значимыми переменными в 10,2% случаев, широта и долгота (Lat_Lon) – в 12,2%, широта и высота (Lat_Alt) – в 10,2%, долгота и высота (Lon_Alt) – в 16,3%, отдельно широта (Lat) – в 20,4%, долгота (Lon) – в 16,3% (Рис. 63). Ни в одном уравнении высота над уровнем моря (*Alt*) не является значимой переменной для описания пространственной изменчивости трендов прироста. Это объясняется тем, что локальная температура изменяется с увеличением высоты линейно. С другой стороны, температура является лимитирующим фактором для большинства местообитаний древесных растений в Сибири (см., например, Ваганов и др., 1996; Шишов и др., 2007а). Следовательно, увеличение или уменьшение высоты не отражается на временных трендах прироста.

Полученное распределение значимости свидетельствует о значительных пространственно – временных перераспределениях основных тенденций прироста древесных растений на территории Сибири за последние 250 лет, которые, в большей степени, зависят от широты произрастания древесных организмов и в меньшей - от долготы.

Сделаем предположение о том, что подобные регрессионные уравнения, задаваемые 7 коэффициентами и описывающие зависимость трендов прироста от географических координат, порождают при экстраполяции исходных данных подобные пространственно-временные карты.

В связи с этим рассматривается задача классификации пространственно-распределенных трендов прироста древесных растений для всех 49 рассматриваемых временных интервалов на основе полученных 7-мерных объектов (лет или временных интервалов), каждый из которых представляет собой набор из 7 коэффициентов регрессионного уравнения (Таб. 11). Для

разделения объектов на классы был использован кластерный анализ (Халафян, 2007).

Таблица 11.

Пример используемых для классификации 8 типичных объектов (лет).

Цветом выделены ячейки со значимыми регрессионными коэффициентами

Факторы	Объекты классификации							
	1955	1960	1965	1970	1975	1980	1985	1990
Const	0,303673	0,400201	0,510706	-0,34912	-0,80833	-0,13016	0,421043	0,426427
b_1 (LAT)	-0,00521	-0,00874	-0,0128	0,006905	0,018531	0,003924	-0,00744	-0,00814
b_2 (LON)	-0,0011	-2,2E-05	0,000893	0,000569	0,000136	-0,00046	-0,00112	-0,00096
b_3 (ALT)	3,89E-06	-2,2E-05	-3,8E-05	5,46E-07	2,45E-05	1,94E-05	2,75E-05	3,05E-05
b_4 (LAT^3)	4,47E-07	6,2E-07	8,55E-07	-5,6E-07	-1,4E-06	-2,5E-07	5,89E-07	6,99E-07
b_5 (LON^3)	1,95E-08	-1,4E-09	-1,1E-08	-9,2E-09	-1,2E-08	1,97E-09	1,98E-08	1,23E-08
b_6 (ALT^3)	-1,9E-12	2,21E-12	3,53E-12	2,15E-12	-1,2E-12	-2E-12	-1,5E-11	-1,5E-11

Для избавления от проблемы абсолютной размерности признаков (коэффициентов b_i), значения коэффициентов были стандартизированы на основе следующего соотношения:

$$\hat{b}_k^t = \frac{b_k^t - \overline{b}_k}{\sigma_k},$$

где $\hat{b}_k^t$ - стандартизированное значение ***k***-признака ($k := 1 \div 7$) для ***t***-объекта ($t := 1 \div 49$);

b_k^t - исходное значение ***k***-признака ($k := 1 \div 7$) для ***t***-объекта ($t := 1 \div 49$);

$\overline{b}_k$ - средние значение ***k***-коэффициента ($k := 1 \div 7$);

σ_k - среднеквадратическое отклонение ***k***-коэффициента ($k := 1 \div 7$).

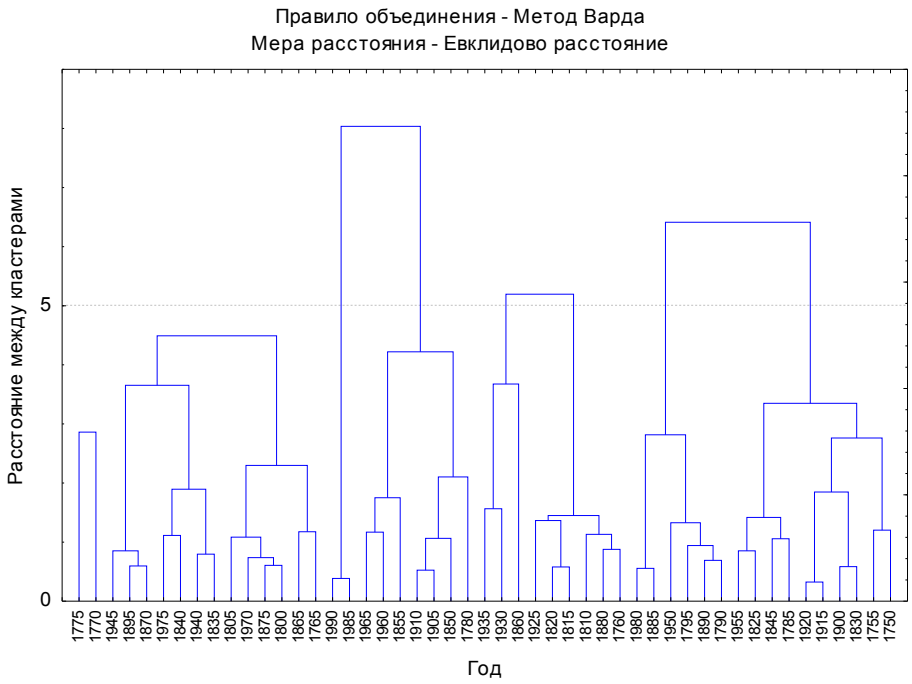


Рис. 64. Дерево кластеризации для 49 анализируемых объектов

Объекты классифицировались (группировались) при помощи двух методов кластерного анализа: метода построения деревьев кластеризации, который применяется, в основном, для нахождения количества классов и примерной структуры классов (Рис. 64), и метода K-средних, применяющийся для непосредственного выделения классов (Таб. 12).

В результате проведение кластерного анализа исходная совокупность 49 лет (19-летних интервалов) была разбита на 6 классов. В первый класс попало 6 объектов, во второй – 5, в третий – 2, в четвертый - 12, в пятый – 15 и в шестой – 9 (Табл. 12). В основе теории классификации лежит так называемая гипотеза компактности: объекты из одного и того же класса должны быть расположены «компактно» (то есть образовывать группы) в пространстве признаков. На этой гипотезе основаны многие критерии качества классификации. Таким образом, попадание объектов в один класс означает их близость в пределах класса, а в разные классы - достаточную удаленность в 7-мерном пространстве признаков, каковыми являются коэффициенты нелинейного регрессионного уравнения.

Таблица 12.

Классификация объектов с выделением 6 классов. D – Евклидово расстояние между центроидом группы (класса) и объектом.

Класс	Объекты														
1	1780	1850	1905	1910	1985	1990									
D	0,36	0,58	0,54	0,39	0,73	0,74									
2	1855	1925	1930	1960	1965										
D	0,37	0,39	0,73	0,5	0,47										
3	1770	1775													
D	0,58	0,58													
4	1760	1765	1800	1805	1810	1815	1860	1865	1875	1880	1935	1970			
D	0,39	0,54	0,33	0,33	0,37	0,5	0,99	0,59	0,15	0,39	0,69	0,28			
5	1750	1755	1785	1790	1795	1820	1825	1830	1845	1890	1900	1915	1920	1950	1955
D	0,69	0,55	0,46	0,57	0,44	0,48	0,2	0,48	0,45	0,43	0,45	0,28	0,3	054	0,38
6	1835	1840	1870	1885	1895	1940	1945	1975	1980						
D	0,24	0,55	0,46	0,45	0,45	0,19	0,32	0,53	0,43						

Проверка качества разбиения выборки на классы проводилась при помощи линейной модификации псевдостатистического метода Байеса – дискриминантного анализа (Фор, 1989). Напомним, что суть этого метода сводится к построению по имеющимся объектам (в нашем случае – 49 годам) и их принадлежности к разным классам (6 классам), алгоритма, который с наименьшей ошибкой относил бы вновь предъявляемый объект к нужному классу.

Для оценки качества идентификации были использованы многофакторный критерий Фишера, матрицы ошибок и статистика Уилкса, которая может принимать значения от 0 (отличное разделение) до 1 (нет дискриминации).

На основе применения вышеописанного алгоритма были получены результаты, свидетельствующие об отличном разделении объектов на классы. Так среднее качество правильной идентификации составляет на обучающей выборке 100 %, то есть построенный алгоритм не ошибается на обучающей выборке (Таб. 13).

Таблица 13.

Матрица ошибок, оценивающая качество идентификации объектов обучающей выборки.

Исходная классификация	Процент корректной идентификации	Предсказанная идентификация					
		Класс 1 p=,12245	Класс 2 p=,10204	Класс 3 p=,04082	Класс 4 p=,24490	Класс 5 p=,30612	Класс 6 p=,18367
Класс 1	100	6	0	0	0	0	0
Класс 2	100	0	5	0	0	0	0
Класс 3	100	0	0	2	0	0	0
Класс 4	100	0	0	0	12	0	0
Класс 5	100	0	0	0	0	15	0
Класс 6	100	0	0	0	0	0	9
Всего	100	6	5	2	12	15	9

О хорошем качестве идентификации говорят и следующие расчетные значимые (p-уровень < 0,0001) значения: статистики Уилкса λ_W=0,009 и критерия Фишера F= 11,4 (p<0.0001).

Проанализируем более подробно первый выделенный класс, к которому принадлежит временной интервал 1981-1999 гг. с центром в 1990 г. Именно для этого временного интервала существуют наблюдения по спутниковым данным NDVI и проведена целая серия исследований с целью изучения поглощения солнечной радиации фитомассой и исследования вегетационной активности растительности (Myneni et al., 1997; Myneni et al., 1998; Zhou et al., 2001).

Рассмотрим процедуру экстраполяции данных на основе полученных нелинейных функций для территории Сибири и получения карт. Для этого были использованы пакеты программ ERDAS Imagine и ARC VIEW GIS – для построения карт.

ERDAS Imagine версии 8.2 задействован для пространственной экстраполяции трендов как функции географических координат. В работе была использована растровая цифровая карта высот GTOPL30 для оценки высоты над уровнем моря, на которую была наложена векторная карта (1 : 2 500 000) лесных территорий Сибири, в качестве маски, для отсечения нелесных территорий (Шишов и др., 2002; Шишов, Рубцов, 2006).

На рисунке 65 изображён алгоритм экстраполяции трендов, созданный с помощью модуля Erdas Modeler. Кругами обозначены функции, а “ромбами” – растровые объекты. На вход алгоритма подаётся растровая карта рельефа Евразии (**n2_asia_dem2**). Функции **MAPX** и **MAPY** вычисляют с растровой карты координаты по X и по Y, соответственно, и хранят элементы растра в переменных **n6_memory** и **n4_memory**. Координаты по Z непосредственно поступают на вход результирующей функции, которая обозначена на данном рисунке числом **0,639704**. Эта функция вычисляет пространственное - временное распределение трендов прироста по заданным коэффициентам для широты, долготы и высоты.

На выходе результирующей функции получается растровый слой, который конвертируется в программу Map Info, где на него накладывается карта лесных территорий, раскрашивается и создаётся легенда.

В результате применения вышеописанной процедуры были построены карты трендов прироста для всех временных интервалов 1-ого класса (Шишов, Рубцов, 2006).

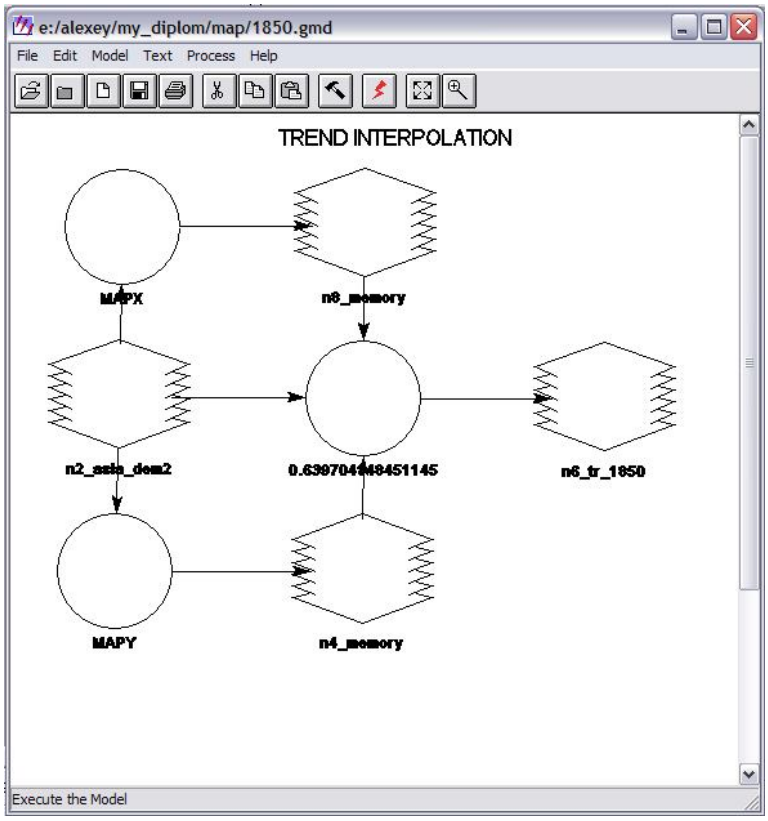


Рис. 65. Алгоритм построения карт трендов

Отметим, что полных количественных аналогов пространственно-временной изменчивости прироста древесных растений для последнего временного интервала с 1981 по 1999 годы не существует, за исключением соседнего к нему интервала с 1976 по 1994 годы. В последнем случае, евклидово расстояние между объектами D = 0,38, корреляция между картами R=0,98 ($p<0.001$).

Качественными аналогами можно считать все остальные пространственно-временные карты, соответствующие объектам 1-ого класса. Так, например, для карты пространственной динамики трендов прироста, соответствующей интервалу 1901-1919 гг. Евклидово расстояние до современного пространственного образца (1981-1999 гг.) D равно 2,68, а коэффициент корреляции между ними $R_{p<0.001}$ =0,54, соответствующие характеристики для 1771-1789 гг. следующие: D= 2,12, $R_{p<0.001}$ =0,79.

Таким образом, можно констатировать, что 1 класс является качественно однородным, то есть все объекты этого класса сохраняют одинаковые качественные (не количественные) пространственные тенденции. А именно, можно проследить общее постепенное уменьшение значение трендов прироста с Запада на Восток и более высокие значения трендов в южных широтах исследуемой территории.

При этом отмечается самая значительная количественная изменчивость прироста древесных растений за последние 25 лет XX века, которая, вероятней всего, связана с самым значительным потеплением за последние 600 лет (Tett et al., 1999).

Проанализируем пространственные соответствия между картой трендов NDVI (Zhou et al., 2001) и картой трендов прироста древесных растений, полученных для временного интервала с 1981 по 1999 годы.

Рассмотрим более подробно данные NDVI, используемые в анализе.

Установленная на спутнике NOAA аппаратура регистрирует восходящее излучение от земной поверхности в красном (0,58 – 0,68 мкм), ближнем инфракрасном (0,725 – 1,0 мкм) и некоторых других областях спектра (Кашкин, 2000). Сравнивая между собой яркости f_1 и f_2 изображения растительности в

красном и ближнем ИК участках, можно судить о состоянии растений в данном районе. Для этого используют дифференцированный вегетационный индекс DVI $= f_2 - f_1$. Однако DVI очень чувствителен к взаимному расположению источника освещения (Солнца), спутника и растений. Нормализованный дифференцированный вегетационный индекс в значительной степени лишён этого недостатка:

$$NDVI = \frac{f_2 - f_1}{f_1 + f_2};$$

Это связано с тем, что f_1 и f_2 изменяются приблизительно пропорционально при изменении взаимного расположения Солнца, спутника и растений.

Для растений в нормальном состоянии NDVI близок к 0,6 – 0,65; низкие значения (0,3 – 0,4) свидетельствуют о недостатке влаги или о заболевании растения. Индексы NDVI представляют собой прямые наблюдения за поглощённой активной фотосинтетической радиацией, являющейся показателем потенциального фотосинтеза, который определяет продуктивность древесных растений. Реальная же величина фотосинтеза и продуктивности зависит не только от физических свойств экосистемы поглощать солнечную энергию, но и от целого ряда характеристик (температуры, влажности воздуха, почвы и т. д.) Поэтому реальный фотосинтез всегда будет ниже, чем измеряемый по NDVI. Тем не менее, величины NDVI и фотосинтеза зависят от роста и развития листовой поверхности деревьев, размеры и свойства которой зависят от текущих климатических условий внешней среды.

Предварительно, карта трендов прироста древесных растений была подвергнута обработке, заключающейся в выделении 6 классов для имеющихся значений трендов. В первый класс попали все значения трендов, меньших -0,015, во второй попали все значения трендов, лежащих в пределах [-0,015; -0,005), в третий - [-0,005; 0), в четвертый - [0; 0,005), в пятый - [0,005; 0,02) и, наконец, в шестой класс попали все значения трендов, больших 0,02. Предварительная обработка делалась для того, чтобы существовало соответствие по значениям между сравниваемыми картами по NDVI и трендам прироста.

Для сопоставления пространственных образцов был применен корреляционный анализ. Для этого была использована пространственная маска, которая содержала 238 точек, общих для сравниваемых карт. В результате сопоставления двух рядов пространственных точек выявлена хорошая согласованность между данными NDVI и трендами прироста древесных растений (Рис. 66).

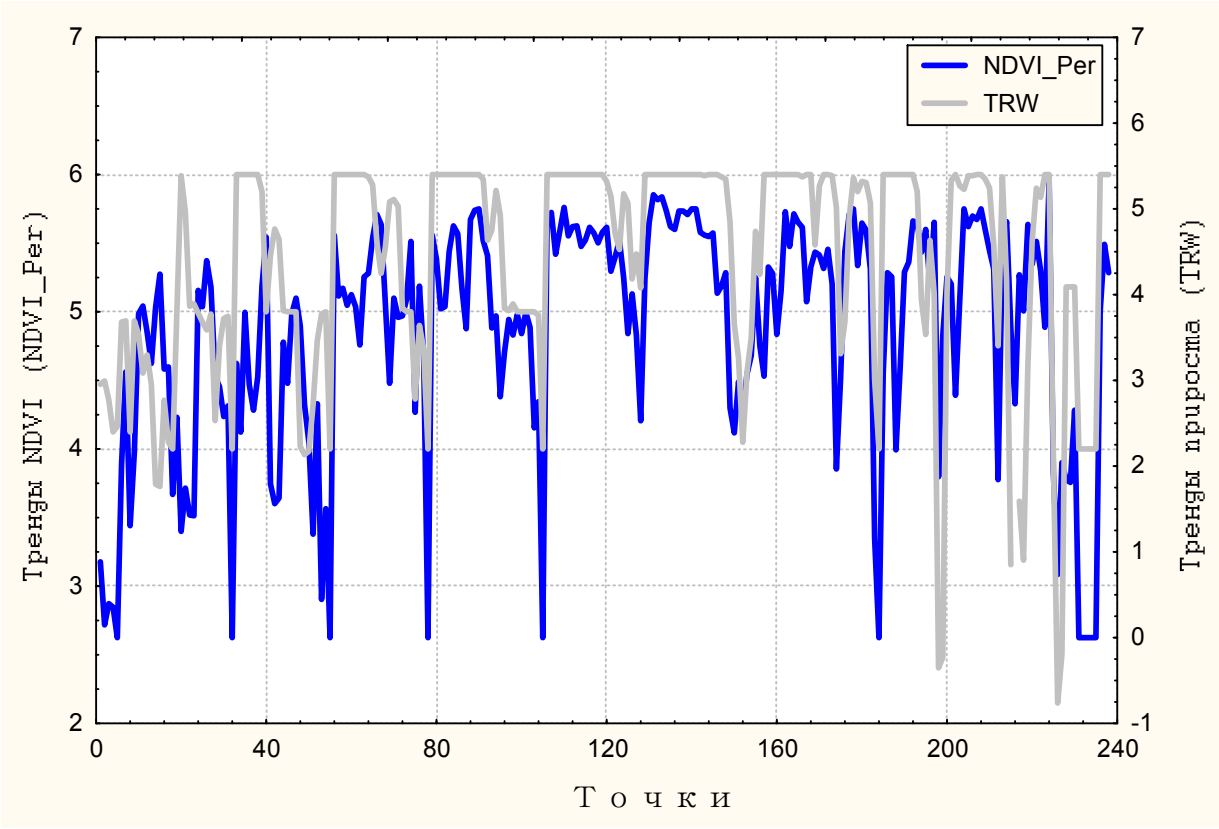


Рис. 66. Пространственная изменчивость трендов NDVI и трендов прироста древесных растений, полученная для территории Сибири

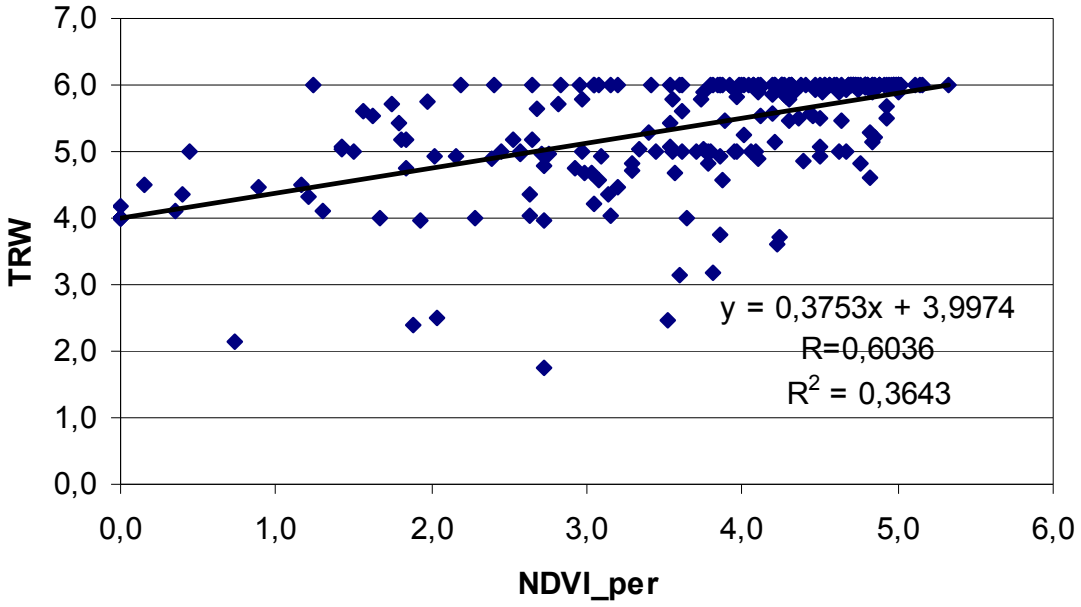


Рис. 67. Оценка корреляционной связи между рядами трендов NDVI (**NDVI_per**) и трендов прироста древесных растений (**TRW**).

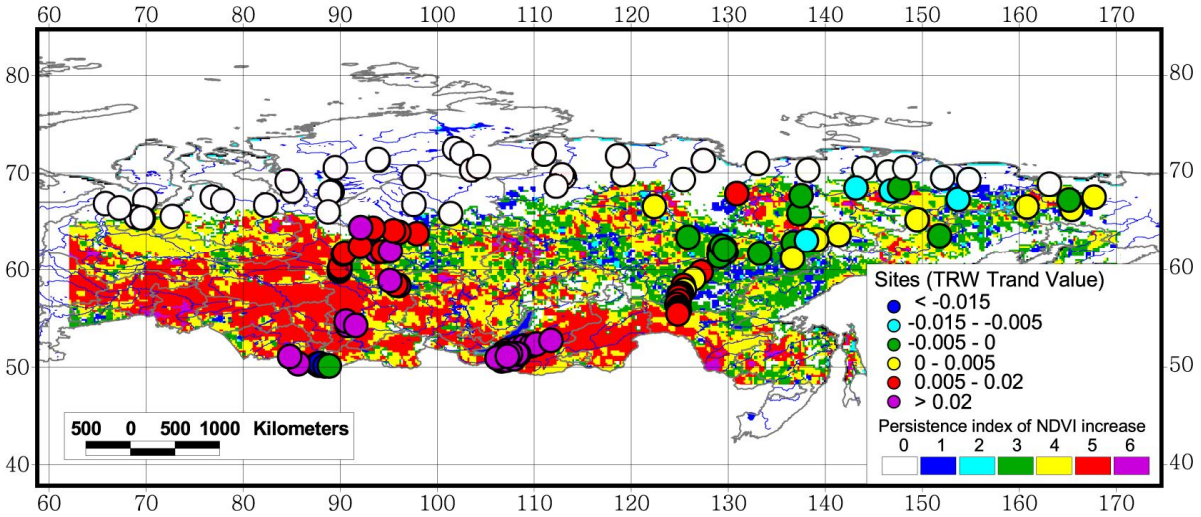


Рис. 68. Карта трендов NDVI, совмещенная с пространственными значениями трендов прироста древесных растений

Оценка коэффициента корреляции между исследуемыми рядами является значимой с вероятностью 0,999 и равна 0,6 (Рис. 67).

При совмещении карты трендов NDVI с данными, полученными по приросту, выявляется хорошая пространственная согласованность в

изменчивости этих показателей (Рис. 68). Проявляется общая тенденция к уменьшению значений с Запада на Восток, а также более высокие значения трендов в южных широтах исследуемой территории.

Независимый дендроклиматический анализ показывает, что большинство использованных в анализе древесно-кольцевых хронологий имеют значимую связь с изменениями летней температуры. Поэтому полученные карты трендов прироста можно интерпретировать как региональные тренды температуры, что для условий Сибири не противоречит связи общей продуктивности растительности с изменениями температуры. В таком случае проведенный анализ выявляет важную особенность. Прогнозы глобальных климатических моделей при удвоении углекислоты в атмосфере предполагают более значительный разогрев субарктических областей (Wuethrich, 1999). Однако для последних десятилетий, когда согласно ряду теоретических оценок вклад антропогенной составляющей в повышение температуры является значимым и ощутимым (Tett et al., 1999), для территории Сибири наблюдается увеличение радиального прироста деревьев и увеличение температуры в средних широтах. Это, по нашему мнению, не согласуется с данными прогнозов глобальных климатических моделей. Возможны два объяснения наблюдаемым пространственным неоднородностям трендов радиального прироста деревьев в связи с изменениями температуры. Первое - это то, что теоретические оценки вклада парниковых газов в увеличение глобальной температуры завышены и пока фиксируемые изменения температуры обусловлены природными явлениями (Corti et al., 1999). Второе - это то, что глобальные модели недостаточно точно работают на региональном уровне, вплоть до значительных территорий, каковой является Сибирь. Оба объяснения требуют более адекватной разработки и более тщательной верификации глобальных климатических моделей. Полученные результаты позволяют надеяться, что развиваемая сеть дендроклиматических станций по всей планете может оказаться весьма полезным инструментом настройки и верификации климатических моделей как на глобальном, так и региональном уровнях.

П.12 Пример использования метода главных компонент и дискриминантного анализа для реконструкции термических характеристик сезона роста по клеточным измерениям годичных колец.

Дендроклиматические реконструкции дают возможность оценить величину и темпы текущих климатических изменений и сравнить их с масштабами подобных процессов в прошлом (Briffa et al., 2001).

В данном разделе предложен новый подход к реконструкции погодных условий произрастания лиственницы на северной границе леса на основе данных клеточной структуры годичных колец.

Были использованы ежедневные климатические данные по температуре и осадкам с метеостанции Чокурдах (70° 62' С.Ш.; 147° 88' В.Д). Длительность наблюдений составила 45 лет (с 1945 по 1989 гг.).

Исходные температурные данные сглаживались при помощи метода скользящих средних с окном в 43 дня. Подобного рода сглаживание проводилось с целью избавления от «случайных» температурных флуктуаций. При дальнейшем увеличении окна внутренняя корреляционная структура полученных характеристик по годам не меняется.

Для каждого года были рассчитаны 16 климатических характеристик (Табл. 14), что позволило охарактеризовать каждый сезон роста 16-мерным вектором.

Климатические параметры сопоставлялись с данными по клеточной структуре годичных колец лиственницы (*Larix cajanderi*, Mayr.), произрастающей в районе нижнего течения р. Индигирки (69^{0} 27'-70^{0} 30' с.ш., 148^{0} 07'-150^{0} 27' в.д.). Сбор образцов древесины (диски и керны) проводился на 4 участках, два из которых находятся на северной границе леса, а два других участка – на верхней границе распространения древесной растительности.

Измерение параметров структуры колец были выполнены на специализированной установке анализа изображений (Image-System). Для каждого абсолютно датированного годичного кольца у клеток 5-и случайно выбранных радиальных рядов измерялись величина

внутриклеточного просвета L и толщина двойной клеточной стенки 2W (Рис. 69).

Таблица 14.

Климатические характеристики, рассчитанные по ежедневным данным температуры и осадков по метеостанции в районе нижнего течения реки Индигирка

№	Название характеристики	Обозначение
1.	Дата перехода через 0° в начале сезона роста	STDAT0
2.	Дата перехода через 5°	STDAT5
3.	Дата перехода в конце сезона через 5°	FDAT0
4.	Дата перехода в конце сезона через 0°	FDAT5
5.	Длительность сезона от 0° до 0°	INTER0
6.	Длительность сезона от 5° до 5°	INTER5
7.	Максимум температуры	MAXT
8.	Дата достижения максимума температуры	MDAT
9.	Сумма температур больше 0°	SUMT0
10.	Сумма температур больше 5°	SUMT5
11.	Сумма температур выше 0° до 22 июня	T220
12.	Сумма температур выше 5° до 22 июня	T225
13.	Сумма температур от 22 июня до перехода через 0° в конце сезона	FT220
14.	Сумма температур от 22 июня до перехода через 5° в конце сезона	FT225
15.	Скорость подъема температуры	SPEEDT
16.	Сумма осадков в течение сезона роста	SUMPREC

Эти данные использовались для расчета радиального размера S и толщины клеточной стенки W каждой трахеиды. Данные были получены по 5 деревьям для каждого года за период с 1437 г. по 1993 г.

Количество клеток в годичных кольцах, сформировавшихся у одного дерева в разные годы, и в кольцах разных деревьев в один и тот же год может существенно различаться. При этом форма трахеидограмм (зависимость размера трахеиды от положения этой трахеиды в радиальном ряду годичного кольца),

полученных для годичных колец одного года формирования, но разных деревьев, одинаковая (Ваганов и др., 1985). Поэтому для сравнения особенностей анатомической структуры древесных колец использовалась процедура нормировки числа клеток в каждом измеренном ряде к некоторому стандартному числу (Vaganov, 1990). Нормировка «сжимает» или «растягивает» исходную трахеидограмму по числу клеток, оставляя без изменения размеры клеток. В нашем случае нормировка проводилась к 11 клеткам – среднему числу трахеид, сформированных разными деревьями в разные период с 1945 по 1989 гг. (период, обеспеченный инструментальными наблюдениями).

Так как каждая клетка нормированной трахеидограммы характеризовалась двумя параметрами, то в результате усреднения данных по радиальным сериям и модельным деревьям были получены 22 временных ряда, характеризующие погодичное изменение структуры колец: 11 хронологий по радиальным размерам трахеид (S1, S2,…, S11) и 11 хронологий по толщине клеточной стенки (W1, W2,…, W11).

Для учета влияния погодных условий нынешнего года на рост и структуру кольца в следующем году в анализ были включены параметры клеточной структуры колец, сформированных в следующем году (S1N, S21,…, S111, W11, W21,…, W111). В итоге было получено 44 признака, которые позволили охарактеризовать условия каждого сезона роста за период измерений.

Для анализа совокупности многомерных данных и реконструкции были использованы метод главных компонент (Айвазян и др., 1989), дисперсионный анализ и линейная модификация псевдостатистического метода Байеса - дискриминантный анализ (Фор, 1989; Халафян, 2007).

Корреляционный анализ полученных шестнадцати климатических характеристик на основе метода главных компонент позволил визуализировать 16-мерную выборку в пространстве первых трех главных компонент. На первые две главных компоненты приходится 72,21% объясненной дисперсии, а на первые три - 81,27% (Таб. 15).

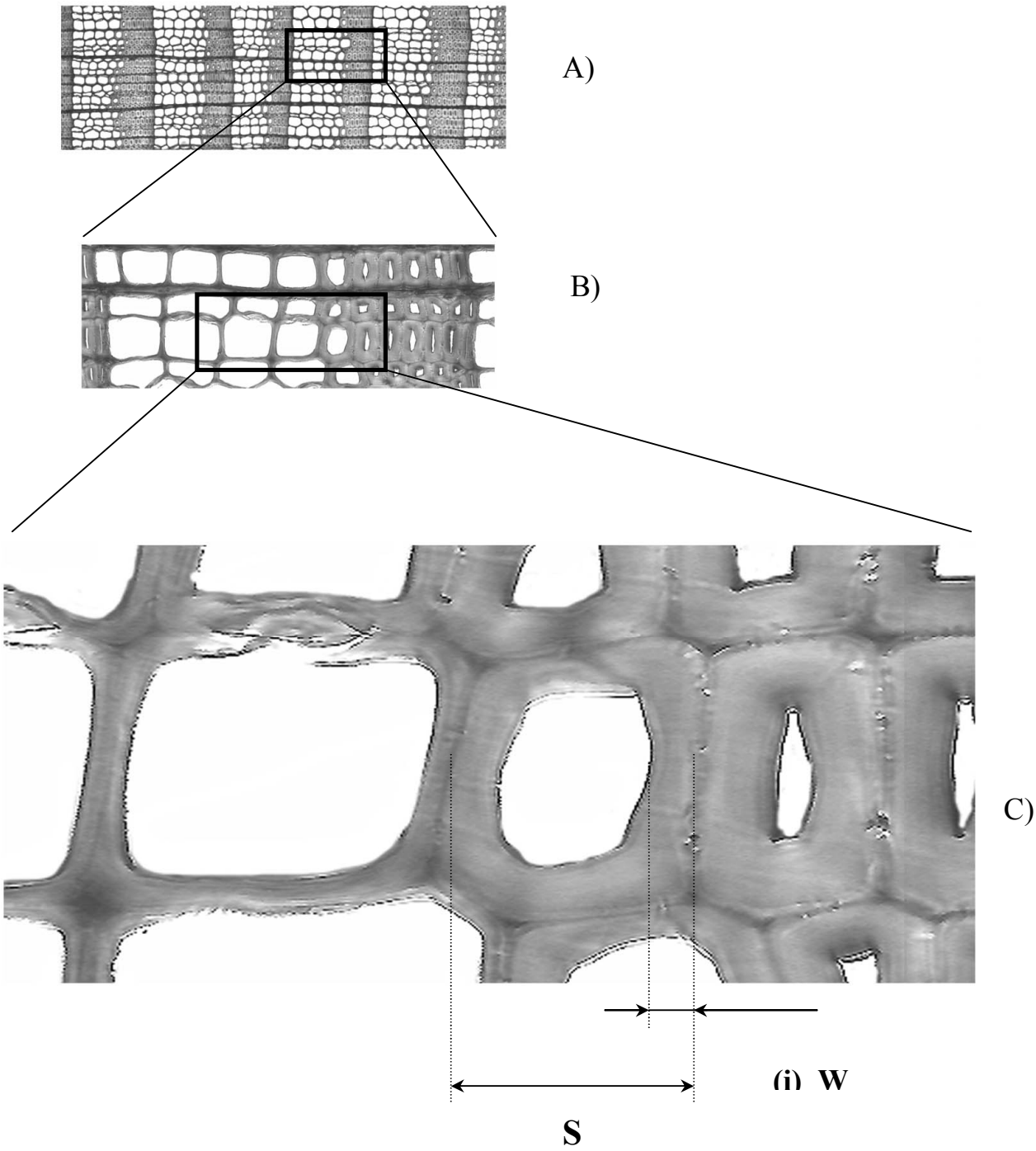


Рис. 69. Микрофотография образца древесины, показывающая структуру годичных колец (A), увеличенное изображение одного годичного кольца (B), характеристики, используемые при работе (C)

Таким образом, исходная многомерная выборка может быть адекватно отображена в пространстве двух или трех главных компонент (Рис.70).

На основе корреляционных связей климатических параметров из шестнадцати исходных характеристик были выделены шесть наиболее информативных:

1. STDAT0 (коэффициент корреляции с STDAT5 – 0,81)
2. SUMT0 (коэффициент корреляции с SUMT5 – 0.99, INTER0 – 0,69, INTER5 – 0,89);
3. MAXT.
4. MDAT;
5. T220 (коэффициент корреляции с T225 – 0,99);
6. SUMPREC

Таблица 15

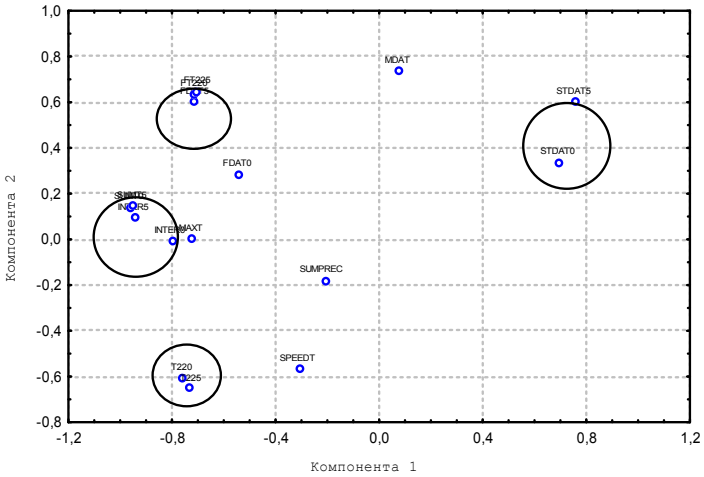
Собственные значения и процент объясненной дисперсии для первых 5 главных компонент.

№ компоненты	Собственные значения	% объясненной дисперсии	Накопленные собственные значения	Накопленный % объясненной дисперсии
1	8,05	50,29	8,05	50,29
2	3,51	21,92	11,55	72,21
3	1,45	9,06	13,00	81,27
4	1,36	8,49	14,36	89,76
5	0,64	3,99	15,00	93.75

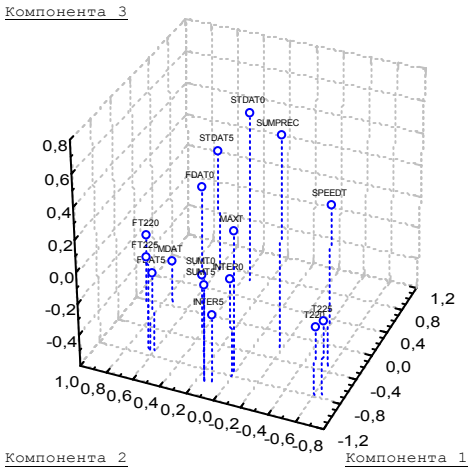
Анализ расположения этих признаков в плоскости первых двух главных компонент (Рис.70) указывает на их «независимость» и «информативность».

Относительный вклад каждой из 6 выбранных климатических характеристик в изменчивость первых двух главных компонент позволяет интерпретировать их следующим образом:

1) увеличение значений первой главной компоненты свидетельствует о более низкой накопленной за вегетационный период температуре для дней с температурой выше 0° С, более низкой максимальной температуре, более низкой суммарной температуре, накопленной до 22 июня (дня летнего солнцестояния);



А)



Б)

Рис.70. А) Расположение исходных шестнадцати признаков в плоскости первых двух главных компонент; Б) Расположение исходных признаков в плоскости первых трех главных компонент.

2) более высокие значения второй главной компоненты являются показателем более поздней даты наступления летней максимальной температуры, более низкой скорости подъема температуры сезона роста и более высокой накопленной температуры после 22 июня.

Рассмотрим пример идентификации сезонов роста на основе двух климатических характеристик: даты перехода температуры через 0^0C (STDAT0) и накопленной температуры для дней с T>0^0C до 22 июня (T220). Эти характеристики вносят большой вклад в суммарную дисперсию климатических переменных: для 1-ой главной компоненты весовые коэффициенты для STDAT0 и T220 равны, соответственно, 0,63 и –0,78, для 2-ой – 0,15 и 0,48, для 3-ей – 0,63 и 0,37. Эти характеристики сильно отрицательно коррелируют между собой (R=-0,82). Кроме того, рассматриваемые показатели полностью определяют начало сезона роста: STDAT0 напрямую связана с датой начала камбиальной активности, а T220 является интегральной характеристикой хода положительных температур до дня летнего солнцестояния (т.е. интервала сезона с максимальной освещенностью и подъемом температуры).

Введем классификацию (разбиение на классы) сезонов роста следующим образом. Рассмотрим эллипс рассеивания объектов (сезонов роста) в 2-мерном фазовом пространстве (ось абсцисс – STDAT0, ось ординат – T220) (Рис.71): 1-ый класс - 7 (или 16%) лет в верхней части эллипса (раннее начало сезона роста, высокая накопленная температура); 2-ой класс – 30 (или 66%) лет в средней части эллипса (умеренное начало сезона роста, умеренная накопленная температура), 3-ий класс – 8 (или 18%) лет в нижней части эллипса (позднее начало сезона роста, низкая накопленная температура).

Можно отметить значительную разницу в динамике температур выделенных классов в начале и середине вегетационного периода (середина мая, июнь, июль) (Рис.72). В тоже время, наблюдается сходная динамика в августе - начале сентября. Тем не менее, различия между классами значимы (критерий Фишера при проведении однофакторного дисперсионного анализа равен 3,23 ($p<0,05$)), что подтверждает корректность введенной классификации.

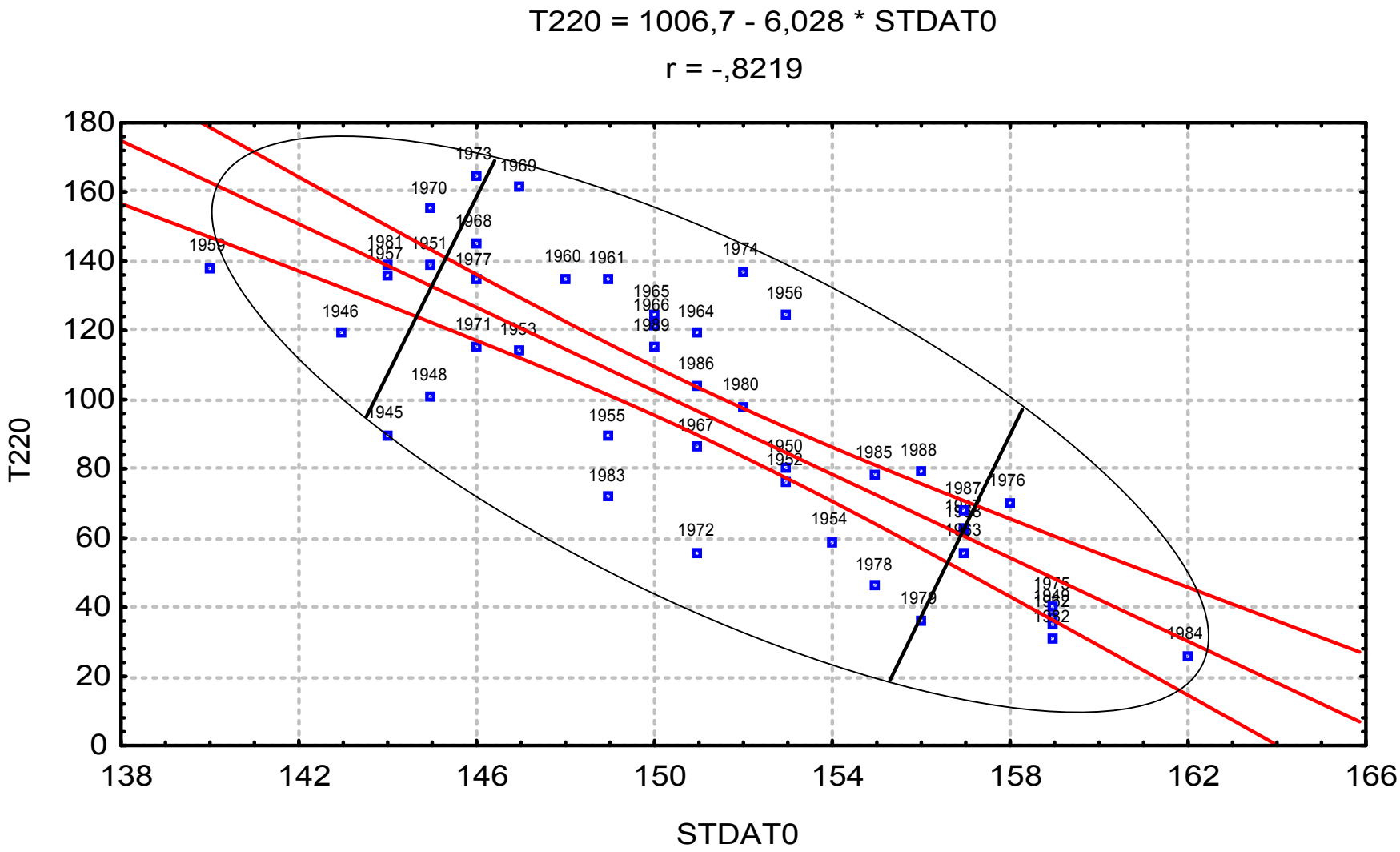


Рис. 71. Классификация 45 объектов (лет) в фазовой плоскости (0_STDAT_T220).

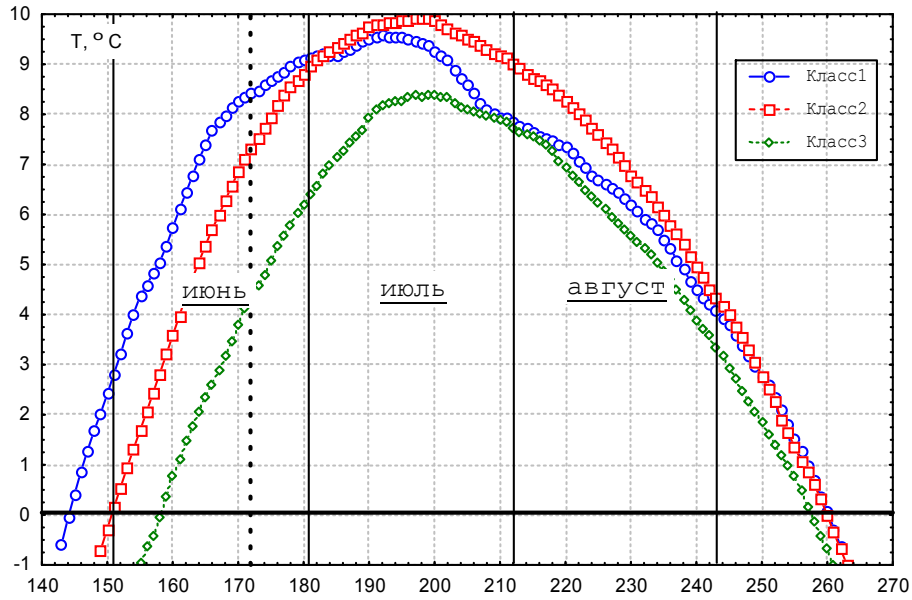


Рис. 72. Динамика средних сглаженных температурных кривых для 1, 2 и 3 классов.

Таблица 16.

Классификационная таблица, оценивающая качество идентификации объектов обучающей выборки.

Исходная классификация	Процент корректной идентификации	Предсказанная классификация		
		Класс 1	Класс 2	Класс 3
		p=0,16	p=0,66	p=0,18
Класс 1	100	7	0	0
Класс 2	100	0	30	0
Класс 3	75	0	2	6
Итого	96	7	32	6

Оценим качество проведенного разбиения объектов (сезоны роста) на выделенные классы, охарактеризовав сезоны уже различными измерениями клеточной структуры годичных колец древесных растений. Для этого объекты

(сезоны роста) охарактеризуем наиболее информативными из рассматриваемых 44 клеточных признаков при помощи пошагового дискриминантного анализа с включением.

В результате было отобрано 14 наиболее информативных характеристик клеточной структуры: S1, S2, S9, S10, S11, S111, W2, W3, W7, W8, W9, W10, W101, W111. То есть, для того чтобы правильно распознавать объекты в предложенной классификации достаточно охарактеризовать их 14-мерными векторами.

Среднее качество правильной идентификации составляет на обучающей выборке 96%. О хорошем качестве идентификации говорят и следующие расчетные и значимые ($p<0,001$) значения: $\lambda_W=0,17$, $F= 2,8$.

Рассмотрим канонические корреляции и канонические корни для рассматриваемой классификации (Рис. 73). Корень 1 в основном дискриминирует между классом 2 и объединением классов 1 и 3, тогда как Корень 2 разделяет классы 1 и 3. Расположение объектов по классам свидетельствует, что разбиение исходной совокупности лет по клеточным измерениям является достаточным для хорошей идентификации объектов (Рис.73).

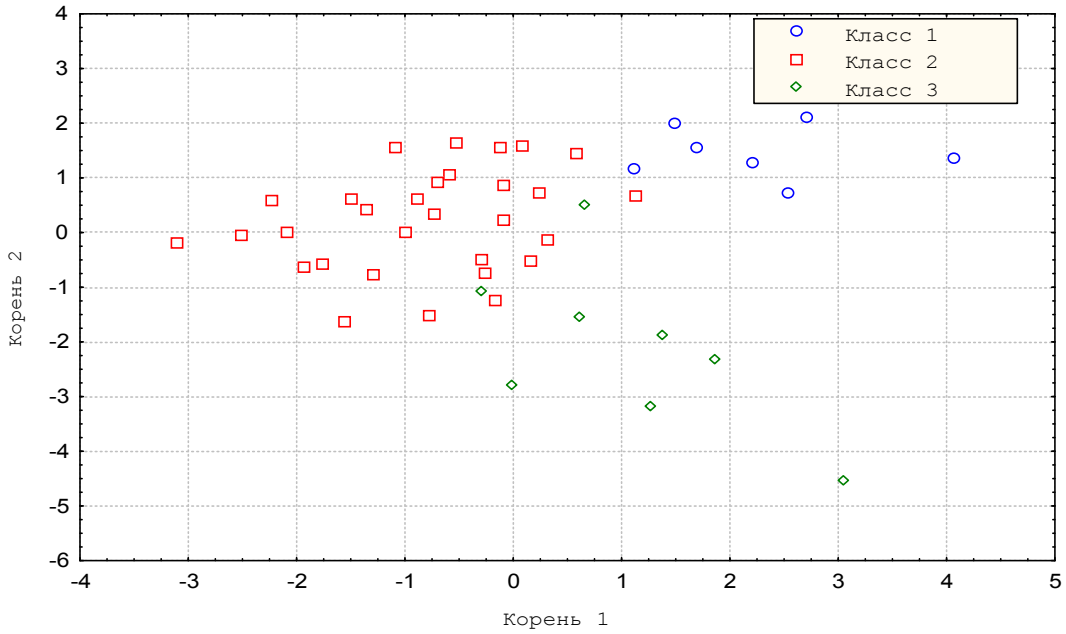


Рис. 73. Расположение объектов в плоскости первых 2-х канонических корней

Качество идентификации объектов (сезонов роста) было оценено при помощи классификационной матрицы – матрицы ошибок (Табл. 16).

Таблица 16.

Коэффициенты для рассчитанных классификационных функций.

Признаки	Класс1	Класс2	Класс3
	p=,16	p=,66	p=,18
W8	-224,17	-190,20	-406,55
S111	213,27	268,62	283,60
W7	178,32	182,20	302,95
S9	28,50	54,06	50,18
S2	150,22	180,83	158,36
S1	-104,05	-136,53	-115,54
W2	970,28	773,28	738,18
W9	-134,15	-268,73	-119,21
W3	-112,32	146,18	120,58
S11	-325,36	-420,59	-416,44
S10	582,04	655,74	676,63
W10	-48,02	8,57	-65,03
W101	477,49	505,15	422,59
W111	-494,40	-554,94	-484,97
Constant	-490,81	-532,53	-511,59

На основании полученных трех решающих (классификационных) функций, которые представляют собой линейную комбинацию произведений исходных признаков на соответствующие коэффициенты (Таб.16), была проведена реконструкция условий сезонов роста по классам за период с 1437 по 1993 г. (Рис. 74).

Отметим, что предъявляемый 14-мерный объект приписывается к тому классу, для которого решающая функция максимальна.

На основе анализа 21-летних скользящих частот появления классов можно сделать вывод о том, что за последние 600 лет наиболее неординарным интервалом был 50-летний период с 1700 по 1750 гг (Рис. 74а). Этот временной интервал характеризуется максимальной частотой появления самых неблагоприятных сезонов роста (с поздним началом и минимальной накопленной температурой до 22 июня). В целом, динамика появления классов сезонов роста совпадает с известными периодами похолоданий и потеплений на анализируемом 600-летнем временном интервале. Поскольку 1 класс соответствует наиболее благоприятным условиям роста (раннему началу сезона роста и максимальной раннелетней температуре), а 3 класс – позднему началу и минимальной накопленной температуре, то проведенная реконструкция отрицательно коррелирует с реконструкциями летней и раннелетней температуры, полученной при помощи традиционного множественного регрессионного анализа (Hughes et al., 1999; Панюшкина, Наурзбаев, 2000). Наиболее высокая корреляционная связь (от -0,30 до -0,87 ($p<0,05$)) на 50-летних интервалах отмечается для среднелетней температуры за последние 300 лет (с 1700 по 1996 гг.) (Рис. 74). Для реконструкции раннелетней температуры значимые корреляции выявляются на меньшем числе интервалов: 1800-1850 ($R_{ранлет}$=-0,74 ($p < 0,01$)), 1900-1950 ($R_{ранлет}$=-0,82 ($p < 0,01$)) и 1950-1996 ($R_{ранлет}$=-0,69 ($p < 0,01$)). Это позволяет сделать вывод о том, что реконструированные классы условий роста более тесно связаны с интегральными летними температурными характеристиками.

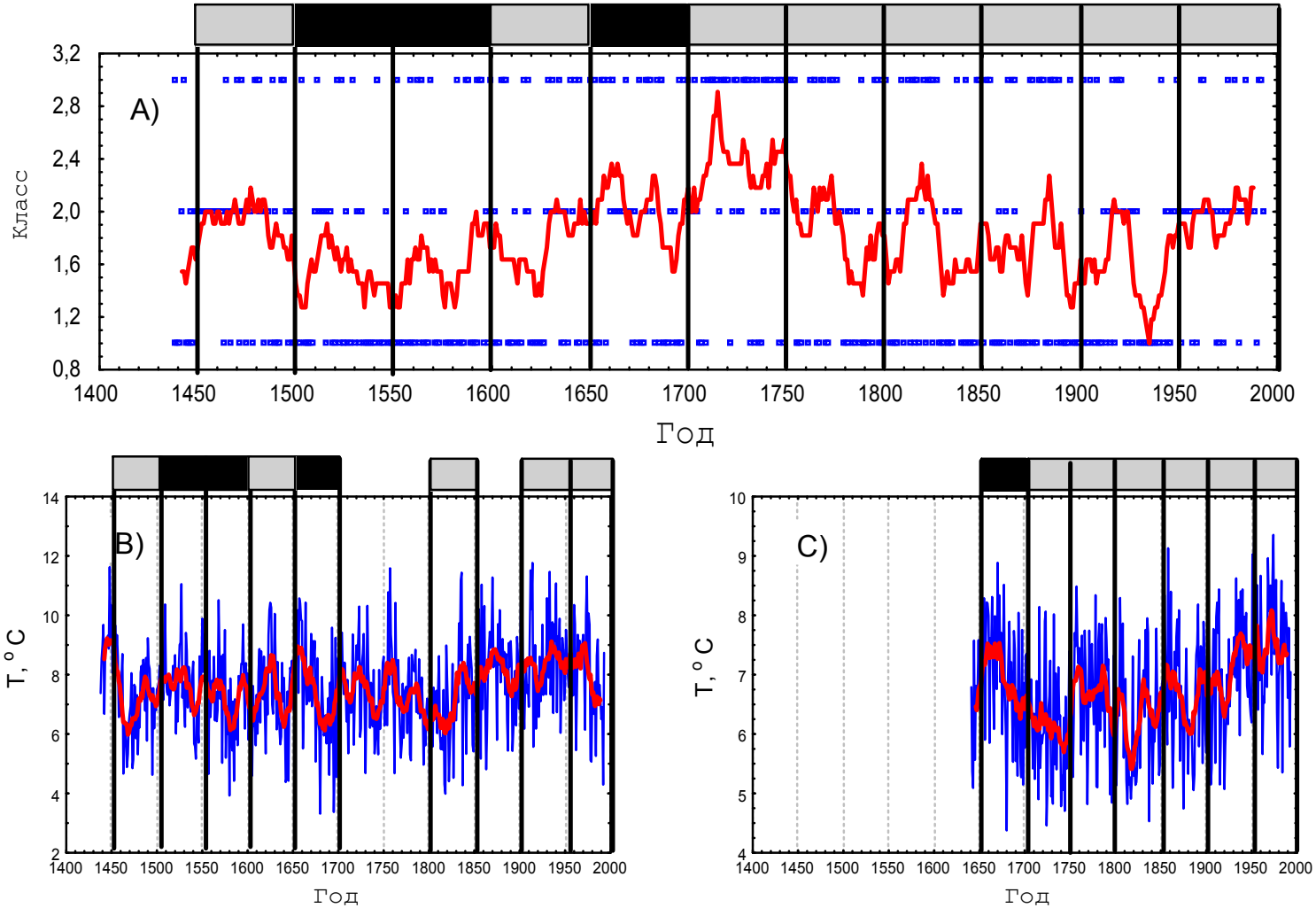


Рис. 74. Климатические реконструкции, полученные по клеточным данным: A) Реконструкция классов сезонов роста; B) Летняя температура; C) Раннелетняя температура

Обратная связь в анализируемых реконструкциях отмечается в XVII веке. Так, на интервале с 1650 по 1700 гг. корреляция между реконструкцией сезонов роста и реконструируемыми температурными характеристиками становится положительной ($R_{лето}$=0,47, $R_{ранлет}$=0,47, $p < 0,01$). Т.е. годы с высокой раннелетней температурой в реконструкции Hughes et al. (1999) попадают в третий, самый неблагоприятный для роста годичных колец, класс по введенной классификации.

Такая связь между реконструированными рядами может быть обусловлена смещением времени перехода температуры через 0° С (соответственно, начала сезона роста) на более поздние даты и/или медленным ростом температуры в период до 22 июня. Подобная динамика температуры может быть обусловлена неординарными климатическими условиями, вызванными вековыми экстремумами солнечной активности (Шпорера и Маундера) (Адаменко, 1976).

Ваганов и др. (Vaganov et al., 1999) показали, что дата начала вегетационного периода и раннелетняя температура являются климатическими параметрами, в наибольшей степени определяющими прирост и структуру годичных колец лиственницы на севере Евразии. На основе дискриминантного анализа были выделены характеристики клеточной структуры колец, которые являются наиболее чувствительными к тому, когда начинается сезон роста и как быстро растет температура в его начале.

Эти данные хорошо согласуются с ранее полученными результатами других авторов. Так, радиальные параметры первых трахеид в кольце зависят от даты схода снежного покрова и, соответственно, даты начала вегетационного периода. Также в изменчивости размеров клеток отражаются термические условия самого начала сезона роста (Vaganov et al., 1999). Размеры клеток и толщина клеточных стенок поздней древесины, определяющие максимальную плотность колец, являются интегральными характеристиками и аккумулируют информацию об изменчивости температуры за наиболее продолжительный промежуток вегетационного периода, в т.ч. и его начала (Yasue et al. 1996, Vaganov et al., 1999). Необходимо отметить также, что погодные условия

текущего года оказывают влияние на структуру кольца, формирующегося на следующий год.

Проведенное сравнение реконструкции с данными других авторов показывает правомерность нового подхода к восстановлению условий в прошлом. Подобным образом могут быть реконструированы условия для любых комбинаций информативных климатических переменных, т.е. становится возможным получить более полную и подробную информацию о климатических изменениях. Предложенный подход позволяет с высоким уровнем достоверности идентифицировать годы с экстремальными условиями.

Заключение

Данная работа посвящена проектированию, разработке и применению статистических оценок, методов, комплексов программ и ГИС технологий в решении плохо формализируемых проблем глобальной экологии. Такой подход позволяет решать сложные задачи дендроклиматологии в полуавтоматическом режиме с минимальными затратами по времени. Подобные подходы могут стать функциональной основой систем дендроклиматического мониторинга России при решении задач, связанных с выявлением и анализом системных связей и закономерностей в приросте древесных растений в связи с меняющимся климатом на территории РФ.

Автором работы уделяется большое внимание новым для дендрохронологии и дендроклиматологии методам и статистическим оценкам на базе непараметрических статистик. К таким оценкам, в частности, относятся модификации известных ранговых статистик Спирмена, Кендалла, Хубера (Kendall, 1970). Модификация состояла в использовании для расчета парных коэффициентов Спирмена, Кендалла, Хубера, множественного коэффициента конкордации Кендалла, обобщенного Хубера скользящего окна (Джансеитов, Шишов, 1993; Шишов, 1998; Шишов, Ивановский, 2006; Шишов и др., 2007б; Briffa et al., 2008). Это позволяет прослеживать динамику распределения значений этих коэффициентов для различных окон скольжения непрерывно на всем периоде наблюдений. В работе использованы также скользящие частоты

выбросов вниз (депрессий роста) и скользящие тренды прироста древесных растений, предложенные автором (Шишов, Рубцов, 2006; Шишов и др., 2007а). Автором показано на примере коэффициента конкордации Кендалла, что скользящие ранговые коэффициенты корреляции являются более устойчивыми к шумовым воздействиям различного рода в сравнении с коэффициентом корреляции Пирсона (NBar) (Шишов, Ивановский, 2006). На основе вычислительного эксперимента показано, что скользящие коэффициенты конкордации Кендалла и осредненной корреляции Пирсона обладают рядом интересных свойств, связанных с их способностью выявлять сигнал даже в сильно зашумленных временных рядах. Причем коэффициент конкордации может распознавать сигнал даже при меньшем количестве анализируемых временных рядов и меньшей величине окна скольжения по сравнению с осредненным коэффициентом корреляции, т.е. коэффициент конкордации является более робастной характеристикой.

Для расчета различных скользящих характеристик автором разработано прикладное программное обеспечение, которое является функциональной составляющей спроектированной интерактивной системы мониторинга. Это позволяет автоматизировать процесс получения новой дендроклиматической информации для системы дендроклиматического мониторинга Урала, Сибири и Дальнего Востока.

В работе дано детальное описание алгоритмов теории классификации: псевдостатистического метода Байеса, его линейной модификации - дискриминантного анализа, метода случайного поиска с адаптацией Лбова, алгоритма семейства «Краб», метода главных компонент, а также возможности их использования в дендроклиматологии для решения разнообразных задач пространственно-временного анализа.

Проведенный анализ позволяет сделать следующие выводы:

1. Описана проблема «дивергенции», которая заключается в несогласованности динамики летней температуры и реального прироста древесных растений после 1960-х годов для регионов, где

температура выступает основным фактором, лимитирующим рост деревьев.

2. Дан анализ возможных причин возникновения «дивергенции».
3. Описана структура, свойства и функциональные возможности спроектированной интерактивной системы дендроклиматического мониторинга, позволяющей в полуавтоматическом режиме анализировать сложные системные проблемы дендроклиматологии, в частности, проблему дивергенции.
4. Показано, что эта проблема может быть связана с методами стандартизации дендрохронологических данных, а также линейным регрессионным анализом дендроклиматических рядов. Линейные корреляции не способны адекватно описать нелинейные взаимодействия между приростом древесных растений и лимитирующими прирост факторами климатической природы.
5. Нелинейные взаимодействия могут быть описаны при помощи разработанной имитационной VS-модели роста.
6. Нелинейные взаимодействия между приростом древесных растений и лимитирующими прирост факторами климатической природы могут быть описаны на базе новых статистических характеристик структуры годичного кольца и новых ранговых статистик, а также новых методов пространственно-временного и спектрального анализа данных.
7. Предложены алгоритмы расчета скользящих коэффициентов корреляции Пирсона, Спирмена, Кендалла, синхронности Хубера, конкордации Кендалла, группового коэффициента Хубера, а также средних скользящих частот выбросов. Эти реляционные характеристики являются устойчивыми к различного рода шумовым воздействиям на ширину годичного кольца древесных растений.
8. На основе вычислительного эксперимента показано, что скользящие коэффициенты конкордации Кендалла и осредненной корреляции Пирсона обладают рядом интересных свойств, связанных с их способностью выявлять сигнал даже в сильно зашумленных временных

рядах. Причем коэффициент конкордации может распознавать сигнал даже при меньшем количестве анализируемых временных рядов и меньшей величине окна скольжения по сравнению с осредненным коэффициентом корреляции, т.е. коэффициент конкордации является более робастной характеристикой.

9. Описаны алгоритмы теории классификации: псевдостатистический метод Байеса, его линейная модификация (дискриминантный анализ), алгоритм семейства «Краб», метод случайного поиска с адаптацией и метод главных компонент. Приведены практические примеры их использования в дендроклиматологии.
10. На основе применения метода главных компонент установлена статистическая связь между пространственно-распределенным приростом древесных растений в обширных регионах Сибири и динамикой общего содержания озона в верхних слоях атмосферы.
11. На основе скользящего коэффициента конкордции Кендалла, метода случайного поиска с адаптацией и псевдостатистического метода Байеса получен алгоритм идентификации событий Эль-Ниньо различной интенсивности по древесно-кольцевым хронологиям, равномерно-распределенным на территории севера Западной, Центральной и Восточной Сибири. Предложен надежный алгоритм реконструкции событий EN интенсивности VS, S+ и M по дендроклиматическим данным.
12. На основе методов теории классификации и GIS технологий выявлены пространственно-временные закономерности изменений прироста древесных растений в прошлом, которые являются качественно (не количественно) согласованными с современной картой трендов прироста деревьев за последние 20 лет XX века. А именно, можно проследить общее постепенное уменьшение значений трендов прироста с Запада на Восток и более высокие значения трендов в южных широтах исследуемой территории. В то же время, отмечается самая сильная количественная изменчивость прироста древесных

растений за последние 25 лет XX века, которая, вероятней всего, связана с самым значительным потеплением за последние 600 лет. При совмещении карты трендов NDVI с данными, полученными по приросту, выявляется хорошая пространственная согласованность в изменчивости этих показателей. Проявляется общая тенденция к уменьшению значений с Запада на Восток, а также более высокие значения трендов в южных широтах исследуемой территории.

13. Предложено правило идентификации сезонов роста на основе линейных дискриминантных функций, полученным по клеточным характеристикам годичного кольца древесных растений. Проведена реконструкция условий начала сезонов роста за период с 1438 по 1989 гг. Предложенный подход позволяет с высоким уровнем достоверности идентифицировать годы с экстремальными условиями.

Таким образом, использование новых для дендроклиматологии ранговых статистик и методов теории классификации позволило получить ряд новых результатов в дендроклиматическом мониторинге и избежать некоторые ограничения, свойственные традиционным статистическим методам, используемых при анализе древесно-кольцевых хронологий.

Литература

1. Адаменко В.Н. Изменчивость увлажнения за последние 5000 лет на основании анализа косвенных показателей / В.Н. Адаменко // ДАН СССР. – 1976. – Т. 228, №2. – С. 463.
2. Айвазян С.А. Прикладная статистика. Классификация и снижение размерности: справочное издание / С.А. Айвазян, В.М. Бухштайер, И.С. Енюков, Л.Д. Мешалкин. − М.: Финансы и статистика, 1989. − 607 с.
3. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики./ С.А. Айвазян, В.С. Мхитарян – М.:ЮНИТИ, 1998.
4. Андерсон Т. Статистический анализ временных рядов / Т. Андерсон: перевод с англ. И.Г Журбенко, В.П. Носко; под ред. Ю.К. Беляева. – М.: Мир, 1976. – 756 с.
5. Андреев С.Г. Регистрация годичными кольцами сосны многолетних колебаний атмосферных осадков, стока Селенги и уровня озера Байкал / С.Г. Андреев, Е.А. Ваганов, М.М. Наурзбаев, А.К. Тулохонов // ДАН. − 1999. − Т. 368, № 3. − С. 400-403.
6. Арпе К. Анализ и моделирование изменений гидрологического режима в бассейне Каспийского моря / К. Арпе, Л. Бенгтссон, Г.С. Голицын, И.И. Мохов, В.А. Семенов, П.В. Спорышев // ДАН. −1999. − Т. 366, №2. −С. 248-252.
7. Битвинскас Т.Т. К вопросу об изучении связи колебаний климата и прироста насаждений / Т.Т. Битвинскас // Докл. ТСХА. – 1965. – Т. 103. – С.385-390.
8. Бокс Дж. Анализ временных рядов / Дж. Бокс, Г. Дженкинс: перевод с англ. – М.: Мир, 1974. – Вып. 1, 2. – 406 с.
9. Большев Л.Н. Таблицы математической статистики / Л.Н. Большев, Н.В. Смирнов. – М.: Статистика, 1983. – 416 с.
10. Боровиков В.П. Statistica. Статистический анализ и обработка данных в среде Windows. / В.П. Боровиков, И.П. Боровиков. – М.:Филинъ, 1997.
11. Брукс К. Применение статистических методов в метеорологии / К. Брукс, Н. Карузерс. – Л.: Гидрометеоиздат, 1963. – 405 с.
12. Ваганов Е.А. Гистометрический анализ роста древесных растений / Е.А. Ваганов, А.В. Шашкин, И.В. Свидерская, Л.Г. Высоцкая. – Новосибирск: Наука, 1985. − 102 с.
13. Ваганов Е.А. Погодные условия и структура годичного кольца деревьев: имитационная модель трахеидограммы / Е.А. Ваганов, И.В. Свидерская, Е.Н. Кондратьева // Лесоведение. – 1990. – № 2. – С.37-45.
14. Ваганов Е.А. Динамика сезонного роста годичных колец сосны густоцветной и сосны жесткой на Корейском полуострове / Е.А. Ваганов, В.К. Пак // Лесоведение. – 1995. – № 2. – С.31-41.

15. Ваганов Е.А. Дендроклиматические исследования в Урало-Сибирской Субарктике / Е.А. Ваганов, С.Г. Шиятов, В.С. Мазепа. − Новосибирск: Наука, 1996. − 244 с.

16. Ваганов Е.А. Механизмы и имитационная модель формирования структуры годичных колец у хвойных / Е.А. Ваганов // Лесоведение. − 1996. − № 1. − С. 3-15.

17. Ваганов Е.А. Роль дендроклиматических и дендрогидрологических исследований в решении глобальных и региональных экологических проблем (на примере Азиатской части России) / Е.А. Ваганов, С.Г. Шиятов // Сиб. экол. ж. − 1999. − Т. 6, № 2. − С. 3-17.

18. Ваганов Е.А. Длительные климатические изменения в арктической области Северного полушария / Е.А. Ваганов, К.А. Бриффа, М.М. Наурзбаев, Ф.Г. Швейнгрубер, С.Г. Шиятов, В.В. Шишов // ДАН. − 2000. − Т.375, №1. − С.103-106.

19. Ваганов Е.А. Рост и структура годичных колец хвойных / Е.А. Ваганов, А.В. Шашкин. – Новосибирск: Наука, 2000. − 232 с.

20. Ваганов Е.А. Свидетели “средневекового потепления” климата / Е.А. Ваганов, М.М Наурзбаев, М.К. Хьюс // Природа. − 2000. − №12. − С. 53-56.

21. Ваганов Е.А. Пространственные взаимоотношения между приростом древесных растений и общим содержанием озона в верхних слоях атмосферы / Е.А. Ваганов, В.В. Шишов, О.В. Сидорова: под ред. В.В. Зуев // Современные достижения в исследованиях окружающей среды и экологии. - Томск: STT, 2004. − С. 210-216.

22. Ваганов Е.А. Дендроклиматические и дендроэкологические исследования в Северной Евразии / Е.А. Ваганов, С.Г. Шиятов // Лесоведение. – 2005. − №4. − С. 18-27.

23. Высоцкая Г.С. Пространственное распределение трендов климатических параметров (XX век) / Г.С. Высоцкая, А.И. Дмитриев., Л.Ф. Ноженкова, В.В. Шишов // Основные закономерности глобальных и региональных изменений климата и природной среды в позднем Кайнозое Сибири. - Новосибирск: Изд. Института археологии и этнографии СО РАН, 2002. − Вып. 1. − С. 83-87.

24. Гренджер К. Спектральный анализ временных рядов в экономике / К. Гренджер, М. Хатанака. – М.: Статистика, 1972. – 309 с.

25. Гурская М.А. Морозобойные повреждения ксилемы хвойных деревьев в лесотундре Западной Сибири и Полярного Урала: автореф. дисс….. канд. биол. наук: 03.00.16 / М.А. Гурская; Институт экологии растений и животных УрОРАН. – Екатеринбург, 2002. − 24 с.

26. Джансеитов К.К. Конкуренция и периодичность процесса прироста леса / К.К. Джансеитов, В.В. Кузьмичев, Ю.В. Кибардин // ДАН СССР. – 1976. – Т.226 – С.695-697.

27. Джансеитов К.К. Пространственные и временная изменчивость процесса прироста леса / К.К. Джансеитов, В.В. Кузьмичев, В.П. Черкашин // ДАН СССР. – 1978. – Т.239. – С.245-249.

28. Джансеитов К.К. Формирование мозаичной структуры древостоя / К.К. Джансеитов, В.В. Кузьмичев // Исследование динамики роста организмов. – Новосибирск: Наука, 1981. – С.78-86.

29. Джансетов К.К. Пространственно-временной анализ дендрохронологических рядов на основе отношений «лучше-хуже», «больше-меньше» / К.К. Джансеитов, В.В. Шишов: препринт ИЛ СО РАН. – Красноярск, 1993. – 29 с.

30. Джансеитов К.К. «Эль-Ниньо» − погодные качели планеты / К.К. Джансеитов, В.В. Шишов: препринт ИЛ СО РАН. – Красноярск: ИЛ СО АН, 1999. − 29 с.

31. Дрейпер Н. Прикладной регрессионный анализ: В 2-х кн. / Н. Дрейпер, Г. Смит. – М.: Финансы и статистика, 1986. – 392 с.

32. Дубров А.М. Многомерные статистические методы: Учебник./ А.М. Дубров, В.С. Мхитарян, Л.И. Трошин. – М.: Финансы и статистика, 2000.

33. Дунин-Барковский И.В. Теория вероятностей и математическая статистика в технике / И.В. Дунин-Барковский, Н.В. Смирнов. – М.: Технико-теоретическая литература, 1955. − 556 с.

34. Загоруйко Н.Г. Методы распознавания и их применение / Н.Г. Загруйко. – М.: Сов. радио, 1972. − 214 с.

35. Загоруйко Н.Г. Методы распознавания и их применение. / Н.Г. Загруйко. – Новосибирск: Изд-во Ин-та Математики, 1999. – 270 с.

36. Зуев В. В. Поведение озонового слоя Земли: возможный вариант развития / В.В. Зуев. // Оптика атмосферы и океана. − 1998. − Т.11, №12. − С. 1356-1357.

37. Зуев В.В. Взаимосвязь долгопериодной изменчивости озонового слоя атмосферы с обусловленной УФ-В воздействием изменчивостью плотности древесины / В.В. Зуев, С.Л. Бондаренко // Оптика атмосферы и океана. − 2001. − Т. 14, №. 12. − С.1-4.

38. Кашкин В.Б. Автоматизированная обработка изображений. Космические средства контроля окружающей среды: Учебное пособие для студентов специальностей 011903 – "Геоинформационные системы", 320700 – "Защита окружающей среды и рацинальное использование природных ресурсов", 071900 – "Информационные системы" / В.Б. Кашкин В.Б. – Красноярск: ИПЦ КГТУ, 2000. – 199с.

39. Кендалл М. Ранговые корреляции / М. Кендалл: пер.с англ. – М.: Статистика, 1975. − 216 с.

40. Кендалл М. Многомерный статистический анализ и временные ряды / М. Кендалл, С. Стюарт С.: перевод с англ. Э.Л. Пресмана, В.И Готоря; под ред. А.Н. Колмогорова, Ю.В. Прохорова. – М.: Наука, 1976. – 736 с.

41. Кирдянов А.В. Сравнительный анализ роста и структуры годичных колец хвойных в лесотундре в северной и средней тайге Средней Сибири: автореф. дисс….. канд. биол. наук: 03.00.16 / А.В. Кирдянов; Институт леса им. В.Н. Сукачева СО РАН. − Красноярск, 1999. − 23 с.

42. Коняев К.В. Зависимость продуктивности растительного покрова от радиационного баланса и потока скрытого тепла в Северной Евразии / К.В. Коняев, А.Н. Золотокрылин, В.В. Виноградова, Т.Б. Титкова // Исследования Земли из космоса. − 2005. − №2. − С. 13-19.

43. Лапко А.В. Непараметрические системы классификации./ А.В. Лапко, В.А. Лапко, М.И. Соколов, С.В. Ченцов. – Новосибирск: Наука, 2000. – 240 с.

44. Лапко А.В. Непараметрические системы обработки неоднородной информации./ А.В. Лапко, В.А. Лапко – Новосибирск: Наука, 2007. – 174 с.

45. Лбов Г.С. Методы обработки разнотипных экспериментальных данных./ Г.С. Лбов – Новосибирск: Наука, 1981. – 160 с

46. Мазепа В.С. Математико–статистические модели дендрохронологических рядов: автореф. дисс….. канд. физ.–мат. наук / В.С. Мазепа. – Свердловск, 1978. – 39 с.

47. Мазепа В.С. Влияние осадков на динамику радиального прироста хвойных в субарктических районах Евразии / В.С. Мазепа // Лесоведение. – 1999а. − № 6. − С. 15-22.

48. Мазепа В.С. Погодичная реконструкция средней летней температуры воздуха на севере Западной Сибири с 1690 года на основе данных о радиальном приросте деревьев / В.С. Мазепа // Сиб. экол. журн. − 1999б. − № 2. − С. 175-183.

49. Мазепа В.С. Дендроклиматическая реконструкция летней температуры воздуха с 1690 года в Субарктических районах Сибири / В.С. Мазепа // Проблемы экологического мониторинга и моделирования экосистем. − СПб.: Гидрометеоиздат, 2000. − Т. 17. − С. 170-187.

50. Панюшкина И.П., Ваганов Е.А., Шишов В.В. Статистический анализ изменчивости прироста лиственницы на севере Средней Сибири / И.П. Панюшкина, Е.А. Ваганов, В.В. Шишов // Географ. и прир. ресурсы. – 1996. – № 4. – С.93-102.

51. Панюшкина И.П. Дендроклиматический анализ прироста лиственницы на севере Средней Сибири / И.П. Панюшкина, Е.А. Ваганов, В.В. Шишов // Географ. и прир. ресурсы. – 1997. – № 1. – С.80-90.

52. Пудовкин М.И. Влияние солнечной активности на состояние нижней атмосферы и погоду. / М.И. Пудовкин // Соросовский образоват. журн. − 1996. − № 10. − С. 106-113.

53. Рудаков В.Е. Метод изучения влияния колебания климата на толщину годичных колец деревьев / В.Е. Рудаков // Докл. АН АрмССР. – 1951. – Т.13. – № 3. – С. 75-79.

54. Сидорова О.В. Региональные особенности радиального прироста лиственницы на севере средней Сибири по 1000-летним древесно-кольцевым хронлогиям. / О.В. Сидорова О.В., М.М. Наурзбаев, М.К. Хьюс, В.В. Шишов// Экология. – 2007. – № 2. – С. 99-103

55. Справочник по прикладной статистике: в 2 т. / перевод с анг.; под ред. Э.Ллойда, У. Ледермана, Ю.Н. Тюрина – М.: Финансы и статистика, 1989. – Т. 1 – 510 с.

56. Терсков И.А. К методике реконструкции погодных условий по динамике роста и структуре годичных колец древесных растений / И.А. Терсков, Е.А. Ваганов, И.В. Свидерская И.В. // Пространственно-временная структура лесных биогеоценозов. − Новосибирск: Наука, 1981. − С. 13-26.

57. Фор А. Восприятие и распознавание образов / А. Фор: под ред. Г.П. Катыса; пер. с фр. А.В. Серединского. − М.: Машиностроение, 1989. − 276 с.

58. Халафян А.А. Statistica 6. Статистический анализ данных. : Учебник./ А.А. Халафян. – М.: Бином, 2007. – 512 с.

59. Хантемиров Р.М. Дендроклиматический потенциал можжевельника сибирского / Р.М. Хантемиров, С.Г. Шиятов, Л.А. Горланова // Лесоведение. − 1999. − № 6. − С. 34-39.

60. Хантемиров Р.М. Древесно-кольцевая реконструкция летних температур на севере Западной Сибири за последние 3248 лет / Р.М. Хантемиров // Сиб. экол. журн. − 1999. − № 2. − С. 185-191.

61. Хантемиров Р. М. Основные этапы развития древесной растительности на Ямале в голоцене / Р.М. Хантемиров, С.Г. Шиятов // Экология. − 1999. − № 3. − С. 163-169.

62. Хантемиров Р.М. 4309-летняя хронология для Ямала и ее использование для реконструкции истории климатических изменений на севере Западной Сибири / Р.М. Хантемиров // Проблемы экологического мониторинга и моделирования экосистем. − СПб.: Гидрометеоиздат, 2000. − Т. 17. − С. 287-301.

63. Хеннан Э. Многомерные временные ряды / Э. Хеннан: перевод с англ. А.С. Холево; под ред. Ю.А. Розанова. – М.: Москва, 1974. – 575 с.

64. Хеттманспергер Т. Статистические выводы, основанные на рангах / Т. Хеттманспергер: перевод с англ.; предисл. Ю.Н. Тюрина, Д.С. Шмерлинга. – М.: Финансы и статистика, 1987. – 334с.

65. Шашкин А.В. Имитационная модель климатической изменчивости хвойных (на примере роста сосны в степной зоне) / А.В. Шашкин, Е.А. Ваганов // Экология. – 1993. – № 5. – С.34-43.

66. Швед Г.М. Циркуляция атмосферы / Г.М. Швед // Соросовский образоват. журн. − 1997. − № 3. − С. 75-81.

67. Шишов В.В. Пространственно-временной анализ дендрохронологичесих рядов методами реляционной математики: автореф. дисс….. канд. физ.-мат. наук / Шишов В.В. – Красноярск: ИЛ СО РАН, 1998. – 22 с.

68. Шишов В.В. Статистическая связь между проявлениями Эль Ниньо и летней температурой в Субарктике Сибири / В.В. Шишов // ДАН. − 2000. − Т.375. − №5. − С. 676-679.

69. Шишов В.В. Статистические модели прогноза интенсивности Эль-Ниньо по дендроклиматическим данным Субарктической части Евразии / В.В. Шишов // Проблемы реконструкции климата и природной среды голоцена и плейстоцена Сибири. Под ред. Е.А. Ваганова и др. – Новосибирск: Издательство Института археологии и этнографии СО РАН, 2000. – Вып.2. – С. 458-466.

70. Шишов В.В. Пространственная изменчивость прироста древесных растений на территории Сибири в последнем столетии / В.В. Шишов, Е.А. Ваганов, М.К. Хьюз, М.А. Корец // ДАН. − 2002. − Т.387, №5. − С. 690-693.

71. Шишов В.В. Реконструкция термических характеристик сезона роста по клеточным измерениям годичных колец с помощью методов теории классификации / В.В. Шишов, А.В. Кирдянов // Проблемы экологического мониторинга и моделирования экосистем. – СПб.: Гидрометеоиздат. – 2003. – Т. XIX. – С. 200-216.

72. Шишов В.В. Сравнительный анализ скользящих коэффициентов при анализе временных рядов / В.В. Шишов, А.Б.Ивановский // Вестник СибГАУ. – Красноярск: Изд. СибГАУ. – 2006. – Вып. 2(9). – С. 29-33.

73. Шишов В.В. Пространственные модели отклика характеристик продуктивности лесной растительности на региональные и глобальные изменения климата./ В. В. Шишов , А.В. Рубцов / В кн.: «Биоразнообразие и динамика экосистем: информационные технологии и моделирование»/ Под ред. Шумного В.К. и др. – Новосибирск: Из-тво СО РАН (Интеграционные проекты СО РАН). – 2006. – С. 547-557

74. Шишов В.В. Анализ изменчивости радиального прироста древесных растений на территории севера Евразии в последние десятилетия / В.В. Шишов, М.М. Наурзбаев, Е.А.Ваганов, А.Б.Ивановский, М.А.Корец // Известия РАН. Сер. географ. − 2007а. - № 3. – С. 49-59.

75. Шишов В.В. Анализ согласованности в приросте древесных растений по сверхдлительным древесно-кольцевым хронологиям севера Евразии / В.В. Шишов, К. Бриффа, Т. Мелвин, М.М. Наурзбаев // Проблемы экологического мониторинга и моделирования экосистем. −СПб.: Гидрометеоиздат. − 2007б. − Т. XXI. – С. 203-222.

76. Шишов В.В. Метод эмпирических категориальных коррелограмм и его применение. / В.В. Шишов// Вестник СибГАУ. – Красноярск: Изд. СибГАУ.. – 2009. – С.33-38

77. Шиятов С.Г. Дендрохронология верхней границы леса на Урале / С.Г. Шиятов. – М.: Наука, 1986. – 136 с.

78. Шиятов С.Г. Методическая основа организации системы дендроклиматического мониторинга в лесах азиатской части России / С.Г. Шиятов, Е.А. Ваганов // Сиб. экол. журн. –1998. – Т. 5, № 1. – С. 31-38.

79. Яглом А.М. Корреляционная теория стационарных случайных функций (с примерами из метеорологии) / А.М. Яглом. – Л.: Гидрометеоиздат, 1981. – 287 с.

80. Allen M.R. Investigating the origins and significance of low-frequency modes of climate variability / M.R. Allen, L.A. Smith // Geophys. Res. Lett. – 1994. – V. 21. – P. 883-886.

81. Angell J.K. Relation between 300 mb North Polar vortex and equatorial SST, QBO, and sunspot number and record contraction of the vortex in 1988-89 / J.K. Angell // J. Climate. – 1992. – V.5. – P. 22-29.

82. Barber V. Reduced growth of Alaska white spruce in the twentieth century from temperature-inducted drought stress./ V. Barber, G. Juday, B. Finney// Nature. – 2000. – V. 405. – P. 668-772.

83. Barefoot A.C. Developing a Dendrochronology for Winchester, England |/ A.C. Barefoot, L.B. Woodhouse, W.L. Hafley, E.H. Wilson // J of the Institute of Wood Science. – 1974. – № 6. – P.34-40.

84. Beyer W.H. Handbook of tables for probability and statistics / W.H. Beyer. – 2nd edit. – Cleveland: Chemical Rubber Co., USA, 1968. – 139 p.

85. Blasing T.J. Time series and multivariate analysis in Paleoclimatology / T.J. Blasing: H.H. Shugart eds. // Time series and ecological processes: SIAMS–SIMS conference series. – Society for industrial and applied mathematics, Philadelphia, USA, 1978. – P. 211-226.

86. Braker O.U. Der Alterstrend bei Jahrringdichten und Jachrringbreiten von Nadelholzern un sein Ausgleich / O.U. Braker // Mitt. Forstl. Bundes–Vers.–Anst. Wien. – 1981. – V. 142. – P. 75-105.

87. Briffa K.R. Tree-climate relationships and dendroclimatological reconstruction in the British Isles / K.R. Briffa: Ph.D. Dissertation, University of East Anglia, Norwich, England, UK, 1984. – 280 p.

88. Briffa K. R. 1992 Fennoscandian summers from AD-500—temperature-changes on short and long timescales./ K.R. Briffa, P.D. Jones, T.S. Bartholin, D. Eckstein, F.H. Schweingruber, W. Karlen, P. Zetterberg, M. Eronen. // Clim. Dynam. – 1992. – V.7. – P. 111–119.

89. Briffa. K.R. Unusual twentieth-century summer warmth in a 1,000-year temperature record from Siberia / K.R. Briffa, P.D. Jones, F.H. Schweingruber, S.G. Shiyatov, E.R. Cook // Nature. – 1995. – V. 376. – P. 156-159.

90. Briffa, K. R. Tree-ring variables as proxy-climate indicators: Problems with low frequency signals./ K.R. Briffa, P.D. Jones, F.H. Schweingruber, W. Karlen, S.G. Shiyatov/ In book "Climatic Variations and Forcing Mechanisms of the Last 2000 Years"/ Eds. Jones P. D., Bradley R. S., Jouzel J. - Berlin: Springer-Verlag. – 1996. - P. 9-41.

91. Briffa K.R. Trees tell of past climates: but are they speaking less clearly today? / K.R. Briffa, F.H. Schweingruber, P.D. Jones, T.J. Osborn, I.C. Harris, S.G. Shiyatov, E.A. Vaganov, H. Grudd // Phil. Trans. R. Soc. London. – 1998 а. – V. 353. – P. 65-73.

92. Briffa, K.R. Influence of volcanic eruptions on Northern Hemisphere summer temperature over the past 600 years / K.R. Briffa, P.D. Jones, F.H. Schweingruber, T.J. Osborn // Nature. – 1998 б. – V. 393. – P. 450-455.

93. Briffa K.R. Reduced decadal thermal response in recent northern tree growth / K.R. Briffa, F.H. Schweingruber, P.D. Jones, T.J. Osborn, S.G. Shiyatov, E.A. Vaganov // Nature. – 1998 в. – V. 391. – P. 678-682.

94. Briffa K. 2000. Annual climate variability in the Holocene: interpreting the message from ancient trees./ K. Briffa // Quat. Sci. Rev. – 2000. – V.19. – P. 87–105.

95. Briffa K.R. Low-frequency temperature variations from the northern tree-ring-density network / K.R. Briffa, T.J. Osborn, F.H. Schweingruber, I.C. Harris, P.D. Jones, S.G. Shiyatov, E.A. Vaganov // J. Geophys. Res. – 2001. – V. 106. – P. 2929-2941.

96. Briffa, K.R. Large-scale temperature inferences from tree rings: a review / K.R. Briffa, T.J. Osborn, F.H. Schweingruber // Global and Planetary Change. – 2004. – V. 40. – P. 11-26.

97. Briffa K.R. Trends in recent Temperature and Radial Tree Growth spanning 2000 years across Northwest Eurasia./ K.R. Briffa, V.V. Shishov, T.M. Melvin, E.A. Vaganov, H. Grudd, R.M. Hantemirov, M. Eronen, M.M. Naurzbaev// Philosophical Transactions of the Royal Society of London Series B-Biological sciences. –. 2008. – Special Issue. – doi:10.1098/rstb.2007.2199. P. 1-14

98. Cias P. A large northern hemisphere terrestrial CO_2 sink indicated by the ^{13}C/^{12}C ratio of atmospheric CO_2 / P. Cias, P. Tans, M. Trolier, J.W.C. White, R.J. Francey // Science. – 1995. – V. 269. – P.1098-1109.

99. Cook E.R. The smoothing spline: a new approach to standartizing forest interior tree–ring width series for dendroclimatic studies / E.R. Cook, K. Peters // Tree–Ring Bull. – 1981. –V. 41. – P.45-53.

100. Cook E.R. A time series analysis approach to tree–ring standartization / E.R. Cook: Ph.D. Dissertation, University of Arizona, Tucson, AZ, USA, 1985. – 240 p.

101. Cook E.R. Forest decline: Modeling the effect of climate in tree rings / E.R. Cook, A.J. Johnson, T.J. Blasing // Tree physiol. – 1987. – № 3. – P.27-40.

102. Cook E.R. Methods of Dendrochronology. Applications in the Environmental Sciences / E.R. Cook, L.A. Kairiukstis. – Dordrecht, Boston, London: Kluwer Acad. Publ., 1990. – 394 p.

103. Corti S. Signature of recent climate change in frequencies of natural atmospheric circulation regimes / S. Corti, F. Molteni, T.N. Palmer // Nature. – 1999. – V. 398, № 6730. – P.799-802.

104. Data documentation for dataset 9290c. Global Synoptic Climatology Network. The former USSR. – National Climatic Data Center. – 2005. – Version 1.0. – 17p.

105. D'Arrigo R.D. 1992. Dendroclimatic evidence from northern North America. / R.D. D'Arrigo, G.C. Jacoby // Climate since A.D. 1500. - London, Routledge, 1992. – P. 296-311.

106. D'Arrigo R.D. Thresholds for warming-induced growth decline at elevational tree line in the Yukon Territory, Canada / R.D. D'Arrigo, R.K. Kaufmann, N. Davi, G.C. Jacoby, C. Laskowski, R.B. Myneni, P. Cherubini // Global Biogeochem. Cycles. – 2004. – V.18. – GB3021.

107. D'Arrigo R. On the "Divergence Problem" in northern forests: a review of tree-ring evidence and possible causes./ R. D'Arrigo, R. Wilson, B. Liepert, P. Cherubini// Global and Planetary Change. – 2007.

108. Damon P.E. Solar forcing of global temperature change since AD 1400 / P.E. Damon, A.N. Peristykh // Climatic Change. – 2005. – V. 68. – P. 101-111.

109. Daubechies I. Ten lectures on wavelets / I. Daubechies // Society for industrial and applied mathematics. 1992. – 357 p.

110. DeLuisi J.J. Northern and middle-latitude ozone profile features and trends observed by SBUV and Umkehr, 1979–1990 / J.J. DeLuisi, C.L. Mateer, D. Theisen, P.K. Bhartia, D. Longenecker, B. Chu // J. Geophys. Res. – 1994. – V. 99. – P. 18901-18908.

111. Draper N.R. Applied regression analysis / N.R. Draper, H. Smith. – New York: 2 edition: John Wiley and Sons, 1981. – 423 p.

112. Dunn O.J. Applied statistics: Analysis of variance and regression / O.J. Dunn, V.A. Clark. – New York: John Wiley and Sons, 1974. – p. 252-305.

113. Eklund B. Variation in the widths of the annual rings of pine and spruce due to climatic conditions in Northern Sweden during the 1900–1944 / B. Eklund // Medd. Statens Skogsforskingsinstitut. – 1954. – V. 44 – P. 1-150.

114. Elsner J.B. Singular Spectrum Analysis / J.B. Elsner, A.A. Tsonis. – London: Plenum Press, 1996. – 274 p.

115. Fraedrich K. European Grosswetter during warm and cold extremes of the El Nino/Southem Oscillation / K. Fraedrich // Internat. J. Climatology. – 1990. – V. 10. – P. 21-32.

116. Fritts H.C. Bristlecone Pine in the White Mountains of California: Growth and Ring Width Characteristics / H.C. Fritts // Papers of the Laboratory of Tree–Ring Res., 4. Tucson: University of Arizona Press: AZ, USA, 1969. – 44 p.

117. Fritts H.C. Multivariate techniques for specified tree-growth and climate relationships and for reconstructing anomalies in paleoclimate / H.C. Fritts, T.J. Blasing, B.P. Hayden, J.E. Kutzbach // J App. Meteorol. – 1971. – V. 10. – P.845-864.

118. Fritts H.C. Tree-rings and climate / H.C. Fritts. – London; New York; San Francisco: Acad. Press, 1976. – 576 p.

119. Fritts H.C. A comparison between response function analysis and other regression techniques / H.C. Fritts, X. Wu // Tree–ring Bull. – 1986. – V. 46. – P.31-46.

120. Fritts H.C. Climatic variation and tree-ring structure in conifers: empirical and mechanistic models of tree-ring width, number of cell, cell size, cell-wall thickness and wood density / H.C. Fritts, E.A. Vaganov, I.V. Sviderskaya, A.V. Shashkin // Climate Res. – 1991. – №1. – P. 97-116.

121. Fritts H.C. Modeling tree-ring structure as related to temperature, precipitation, and day length / H.C. Fritts, A.V. Shashkin: T.E. Lewis ed. // Tree rings as indicators of ecosystem health. – Boca Raton, Ann Arbor, London, Tokyo: CRC Press, 1995. – P.17-59.

122. Furnival G.M. Regression by leaps and bounds / G.M. Furnival, R.W. Wilson // Technometrisc . – 1974. – V. 16. – P. 499-511.

123. Glahn H.R. Canonical correlation and relationship to discriminant analysis and multiple regression / H.R. Glahn // Atmospheric sc. – 1968. – V. 25. – P. 25-31.

124. Gordon G.A. Verification tests for dendroclimatical reconstructions / G.A. Gordon // Technical Notes. – Northern Hemisphere Reconstruction Group, Laboratory of tree–ring research, University of Arizona, Tucson, AZ, USA, 1980. – P. 65-78.

125. Graybill d.A. Revised computer programs for tree–ring research / D.A. Graybill // Tree–ring bull. – 1979. – V. 39. – P. 77-82.

126. Guiot J. Deux methodes d'utilisation de l'epaisseur des cernes ligneux pour la reconstruction des parametres climatiques anciens, l'exemple de leur application dans le domaine Alpin / J. Guiot // Paleogeography, Paleoclimatology, Paleoecology . – 1984. – V. 45. – P. 347-368.

127. Hayes S. TOGA-TAO: A moored array for real-time measurements in the tropical Pacific Ocean [Электронный ресурс] / S.P. Hayes, L.J. Mangum, J. Picaut, A. Sumi, K. Takeuchi // Bull. Am. Meteorol. Soc. – 1991. – V. 72. – P. 339-347. – Режим доступа: http://www.pmel.noaa.gov

128. Hill D.K. March, 1995. Pacific Warming Unsettles Ecosystems / D.K. Hill // Science. – 1995. – V 267. – P. 1911-1912.

129. Hoch G. The carbon charging of pines at the climatic treeline: a global

comparison./ G. Hoch, C. Korner// Oecologia . – 2003. – V. 135. – P. 10–21.

130. Hughes M.K. July–August temperature at Edinbourgh between 1721 and 1975 from tree–ring density and width data / M.K. Hughes, F. Schweingruber, D. Cartwrigth, P.M. Kelly // Nature. – 1987. – V. 308. – P.341-344.

131. Hulme M. Global climate change in the instrumental period / M. Hulme, P.D. Jones // Environmental Pollution. – 1994. – V. 83. – P. 23-36.

132. Hughes M.K. Tree-rings and the challenge of global change research / M.K. Hughes // Tree-Rings. From the past to the future: Proceedings of workshop on Asian and Pacific dendrochronology. – Tsucuba, Forestry and Forest Prod. Res. Publ., 1995. – P. 1-7.

133. Hughes M.K. Twentieth-century summer warmth in northern Yakutia in a 600-year context / M.K. Hughes, E.A. Vaganov, S.G. Shiyatov, R. Touchan, G. Funkhouser // The Holocene. – 1999. – V. 9. – P. 629-634.

134. Hurrell J.W. Influence of variations in extratropical wintertime teleconnections on Northern Hemisphere temperature / J.W. Hurrell // Geophys. Res. Lett. – 1996. – V. 23. – P. 665-668.

135. Ignaccolo M. Scaling in non-stationary time series. (I) / M. Ignaccolo, P. Allegrini, P. Grigolini, P. Hamilton and B. J. West // Physica A: Statistical and Theoretical Physics, 2004. – V. 336. – P. 595-622

136. IPCC. Climate change 2007: the physical science basis. Summary for policymakers. Contribution of working group I to the fourth assessment report of the Intergovernmental Panel on Climate Change. – 2007. - http://ipcc-wg1.ucar.edu/

137. Jacoby G.C. Reconstructed Northern Hemisphere annual temperature since 1671 based on high-latitude tree-ring data from North America / G.C. Jacoby, R.D. D'Arrigo // Climatic Change. – 1989. – V. 14. – P. 39-59.

138. Jacoby G.C. Tree-ring width and density evidence of climatic and potential forest change in Alaska / G.C. Jacoby, R.D. D'Arrigo // Global Biochem. Cycles. – 1995. – V. 9. – P. 227-234.

139. Jones P.D. Global surface air temperature variations during the twentieth century: Pt 1. Spatial, temporal and seasonal details / P.D. Jones, K.R. Briffa // The Holocene. – 1992. – V. 2. – P. 174-188.

140. Jones P.D. The evolution of climate over the last Millennium / P.D. Jones, T.J. Osborn, K.R. Briffa // Science. – 2001. – V. 292. – P. 662-667.

141. Jones P.D. Greenhouse effect and climate data / P.D. Jones // Encyclopedia of Physical Science and Technology. London: Academic Press, 2002. – V. 7. – P. 87-106.

142. Kasatkina E. Stardust component in tree rings / E. Kasatkina, O. Shumilov, N.V. Lukina, M. Krapiec, G. Jacoby // Dendrochronologia, 2007 – V. 24 – P.131–135.

143. Keeling C.D. Increased activity of northern vegetation inferred from atmospheric CO_2 measurements / C.D. Keeling, J.F.S. Chin, T.P. Whorf // Nature. 1996. – V. 382. – P. 146-149.

144. Kelly P.M. Variations in surface air temperature: Part 2. Arctic regions, 1881-1980 / P.M. Kelly, P.D. Jones, S.B. Sear, B.S.G. Cherry, R.K. Tavakol // Monthly Weather Rev. – 1982. – V. 110. – P. 71-83.

145. Kelly P.M. The spatial response of the climate system to explosive volcanic eruptions / P.M. Kelly, P.D. Jones, J. Pengqun // Internat. Journal of Climatol. – 1996. – V. 16. – P. P. 537-550.

146. Kendall M. Rank correlation methods / M. Kendall. – London: Griffin, 1970. – 170 p.

147. Kozlowski T. The physiological ecology of woody plants./ T. Kozlowski, P. Karmer, S. Pallardy / Academic Press, California. – 1991.

148. Kozlowski TT. Growth control in woody plants./ T.T. Kozlowski, S. G. Pallardy – Academic Press, San Diego. – 1997. – 641 p.

149. Kuusela J. Multiple regression of increment percentages on other characteristics of Scotch pine stands / J. Kuusela, P. Kikki. – Finland Society of Forestry: Helsinki, Finland, 1963. – 32p.

150. Laasko K. Effect of ultraviolet-B radiation (UV-B) on conifers / K. Laasko, s. Huffunen // Environmental Pollution. – 1998. – V. 99. – P. 313-328.

151. Le Comte D. World Weather 1982 / D. Le Comte // Weatherwise. – 1983. – V. 36. – P. 14-17.

152. Le Comte D. Worldwide Extreme Floods and Droughts/ D. Le Comte // Weatherwise. – 1984. – V. 37. – P. 8-18.

153. Le Comte D. Highlights Around the World - Water, Water Almost Everywhere... /D. Le Comte // Weatherwise. – 1987. – V. 40. – P. 9-10.

154. Lees J.M. Multiple-taper spectral analysis: a stand alone C-Subroutine / J.M. Lees, J. Park // Computers & Geosciences. – 1995. – V. 21. – P. 199-236.

155. Libes D. Exploring Expect / D. Libes. - O'Reilly, 1994. – 599 c.

156. Libes D. "Expect", Tcl/Tk Extensions. ed. M.Harrison. / D. Libes. - O'Reilly & Associates Inc., 1997.

157. Libes D. Tcl/Tk-based Agents for Mail and News Notification - or - A Tale of Two Biffs / D. Libes. - West Sussex, England: Software - Practice & Experience, John Wiley & Sons, 1998

158. Lischke H. Modeling tree species migration in the Alps during the Holocene: What creates complexity? / H. Lischke // Ecological Complexity, 2005. – V. 2. – P. 159-174.

159. Lischke H. Intra-specific density dependence is required to maintain species diversity in spatio-temporal forest simulations with reproduction / H. Lischke, T. J. Löffler // Ecological Modelling, 2006. – V. 198. – P. 341-361.

160. Lischke H. TreeMig: A forest-landscape model for simulating spatio-temporal patterns from stand to landscape scale / H. Lischke, N. E. Zimmermann, J. Bolliger, S.e Rickebusch, T. J. Löffler // Ecological Modelling, 2007. – V. 199. – P. 409-420.

161. Liu Yu. Reconstruction of May-July precipitation in the north Helan Mountain, Inner Mongolia since A.D. 1726 from tree-ring late-wood widths / Y. Liu, J. Shi, V. Shishov, E. Vaganov, Y. Yang, Q. Cai, J. Sun, L. Wang, I. Djanseitov // Chinese Science Bulletin. – 2004a. – V. 49. – P. 405-409.

162. Liu Yu. The forecast of seasonal precipitation trend at the north Helan Mountain and Baiyinaobao regions, Inner Mongolia for the next 20 years / Y. Liu, V. Shishov, S. Jiangfeng, E. Vaganov, S. Junyan, C. Qiufang, I. Djanseitov, A. Zhisheng // Chinese Science Bulletin. – 2004б. – V. 49. – P. 410-415.

163. Lloyd A.H. Spatial and temporal variability in tree growth and climate response of treeline trees in Alaska / A.H. Lloyd, C.L. Fastie // Climatic Change. – 2002. – № 52. – P. 481-509.

164. Lorenz E.N. An experiment in nonlinear statistical weather forestcasting / E.N. Lorenz // Monthly Weather Rev. – 1977. – V. 105. – P.590-602.

165. Mann M.E. Robust Estimation of Background Noise and Signal Detection in Climatic Time Series / M.E. Mann, J. Lees // Climatic Change. – 1996. – V. 33. – P. 409-445.

166. Mann M.E. Global scale temperature patterns and climate forcing over the past six centuries / M.E. Mann, R.S. Bradley, M.K. Hughes // Nature. – 1998. – V. 392. – P. 779-788.

167. McPhaden, M.J., 1993: TOGA-TAO and the 1991-93 El Niño-Southern Oscillation Event [Электронный ресурс] / M.J. McPhaden // Oceanography. – 1993. – V. 6. – P. 36-44. – Режим доступа: http://www.pmel.noaa.gov/pubs/docs/mcph1401.html

168. Melvin T. Historical growth rates and changing climatic sensitivity of boreal conifers./ T. Melvin / Ph.D. thesis. – Climatic Research Unit, University ofEast Anglia, Norwich, UK. – 2004. – 220 p.

169. Mitchell J.M. Climate change / J.M. Mitchell, B. Dzerdzeevskii, H. Flohn, W.L. Hofmeyr, H.H. Lamb, K.N. Rao, C.C. Wallen // Technical Note. – Geneva: Word Meteorological Organization, 1966. – № 79. – 79 p.

170. Myneni R.B. Increase plant growth in the northern high latitudes from 1981-1991 / R.B. Myneni, C.D. Keeling, C.J. Tucker, G. Asrar, R.R. Nemani // Nature. – 1997. – V.386. – P. 698-702.

171. Myneni R.B., Tucker C.J., Asrar G., Keeling C.D. Interannual variations in satellite-sensed vegetation index data from 1981 to 1991 // J. Geophys. Res. – 1998. – V.103. – P. 6145-6160.

172. Naurzbaev M.M. Variation in early summer and annual temperature in the East Taymir and Putoran (Siberia) over the last two millennia inferred from tree-rings

/ M.M. Naurzbaev, E.A. Vaganov // J.Geophys. Res. – 2000. – V. 105 – P. 7317-7327.

173. Naurzbaev, M.M. Summer temperatures in eastern Taimyr inferred from a 2427-year late-Holocene tree-ring chronology and earlier floating series./ M.M. Naurzbaev, E.A. Vaganov, O. V. Sidorova, F.H. Schweingruber // Holocene. – 2002. – V.12. – P. 727-736.

174. New M. A high-resolution data set of surface climate over global land areas / M. New, D. Lister, M. Hulme, I. Makin // Climate Research. – 2002. – V. 21. – P. 229-233.

175. Ording A. Arringnalyser pa gran og furu / A. Ording // Medd. Norske skogsforsokv., 1941. – V. 7. – № 25. – P. 102-354.

176. Percival D.B. Spectral Analysis for Physical Applications / D.B. Percival, A.T. Walden. – Cambridge: Cambridge University Press, 1993. – 583 p.

177. Philander, S.G.H. El Niño, La Niña and the Southern Oscillation / S.G.H. Philander. – San Diego: Academic Press, 1990. – 289 p.

178. Quinn, W. H. The Unusual Intensity of the 1982-1983 ENSO Event / W.H. Quinn, D.O. Zopf // Tropical Ocean-Atmosphere Newslett. – 1984. - № 26 – P. 17-20.

179. Quinn W.H. El-Nino occurrences over the past four and half centuries / W.H. Quinn, V.T. Neal // Journal of Geoph. Res. – 1997. – Vol. 92. – P. 14449-14461.

180. Quiroz R.S. The Climate of the El Nino Winter of 1982-1983. A Season of Extraordinary Climatic Anomalies / R.S. Quiroz // Monthly Weather Review. – 1983. – V. 111. – P. 1685-1706.

181. Rickebusch S. Combining probabilistic land-use change and tree population dynamics modelling to simulate responses in mountain forests / S. Rickebusch, M. Gellrich, H. Lischke, A. Guisan, N. E. Zimmermann// Ecological Modelling, 2007a. – V. 209. – P. 157-168.

182. Rickebusch S. Understanding the low-temperature limitations to forest growth through calibration of a forest dynamics model with tree-ring data / Rickebusch, H. Lischke, H. Bugmann, A. Guisan, N. E. Zimmermann // Forest Ecology and Management, 2007б. – V.246. – P. 251-263

183. Ramachandra R. A. Multi-taper method of analysis of periodicities in hydrologic data / R.A. Ramachandra, K. Hamed // J. Hydrology. – 2003. – V. 279. – P. 125-143.

184. Rasmusson E.M. August 1983. El Nino: The Great Equatorial Pacific Ocean Warming Event of 1982-1983 / E.M. Rasmusson, J. M. Hall // Weatherwise, 1983. – V. 36. – P. 166-175.

185. Ryan T.P. Modern Regression Methods / T.P. Ryan. - New York: Wiley. 1997. – 229 p.

186. Schulman E. Dendroclimatic changes in semiariad America / E. Schulman. – Tucson: University Arizona press, 1956. – 142 p.

187. Schulze E.-D. Plant Ecology / E.-D. Schulze, E. Beck, K. Muller-Hohenstein. – Berlin: Springer, 2005. – 702 p.

188. Schweingruber F.H. The X–ray technique as applied to dendroclimatology / F.H. Schweingruber, H.C. Fritts, O.U. Braker, L.G. Drew, E. Schaer // Tree–ring Bull. – 1978. – V. 38. – P.61-91.

189. Shishov V.V. VS-oscilloscope: a new tool to parameterize tree radial growth based on climate conditions. / V.V. Shishov, I. I. Tychkov, M. I. Popkova,V. A. Ilyin, M. V. Bryukhanova , A V. Kirdyanov // Dendrochronologia. – 2015. - DOI: http://dx.doi.org/doi:10.1016/j.dendro.2015.10.001

190. Sitch S. Evaluation of ecosystem dynamics, plant geography and terrestrial carbon cycling in the LPJ dynamic global vegetation model / S. Sitch, B. Smith, I.C. Prentice, A. Arneth, A. Bondeau, W. Cramer, J.O. Kaplan, S. Levis, W. Lucht, M.T. Sykes, K. Thonicke, S. Venevsky // Global Change Biol. – 2003. – V. 9. – P. 161-185.

191. Tatsuoka M. Multivariate analysis: Techniques for educational psychological research / M. Tatsuoka. –New York: John Wiley and Sons, 1974. – 310 p.

192. Tett S.F.B. Causes of twentieth century temperature change / S.F.B.Tett, P.S. Stott, M.R. Allen, W.J. Ingram, J.F.B. Mitchell // Nature. – 1999. – V. 399. – №6736. – P.569-572.

193. Thorne P.W. Probable causes of late twentieth century tropospheric temperature trends / P.W. Thorne, P.D. Jones, S.F.B. Tett, M.R. Allen, D.E. Parker, P.A. Stott, G.S. Jones, T.J. Osborn, T.D. Davies // Climate Dyn. – 2003. – V. 21. – P. 573-591.

194. Till C. A synthesis of response functions from eight cedar forests located in Northern Africa / C. Till // Dendrochronologia. – 1984. – № 2. – P.73-82.

195. Touchan R., Shishov V.V., Meko D.M., Nouiri I., Grachev A. 2012. Process based model sheds light on climate signal of Mediterranean tree-ring width. Biogeosceinces, V. 9: 965–972, doi:10.5194/bg-9-965-2012

196. Vaganov E.A. The tracheidogramm method in tree-ring analysis and its application /E.A. Vaganov // Methods of Dendrochronology. - Eds. by E.Cook, L. Kairiukstis. – Dordrecht: Cluwer Acad. Publ., 1990. – P. 63-75.

197. Vaganov E.A. Influence of snowfall and melt timing on tree growth in subarctic Eurasia / E.A. Vaganov, M.K. Hughes, A.V. Kirdyanov, F.H. Schweingruber, P.P. Silkin // Nature. – 1999. – V. 400. – P. 149-151.

198. Vaganov E.A. Tree rings and global carbon cycle /E.A. Vaganov, M.K. Hughes // Проблемы экологического мониторинга и моделирования экосистем. – 2000. – Т. 17. – P. 36-48.

199. Vaganov E.A. Growth Dynamics of Conifer Tree Rings: Images of Past and Future./ E.A. Vaganov, M.K. Hughes, A.V. Shashkin// Ecological studies 183: Springer. – 2006. – 368 p.

200. Waren W.G. On removing the growth trend from dendrochronological Data / W.G. Waren // Tree–Ring Bull. – 1980. – V. 40. – P.35-44.

201. Wilmking M. Recent climate warming forces contrasting growth responses of white spruce at treeline in Alaska through temperature thresholds /M. Wilmking, G.P. Juday, V.A. Barber, H.S.J. Zald // Global Change Biol. – 2004. – V.10. – P.1724-1736.

202. Wimmer R. Use of false rings in Austrian pine to reconstruct early growing season precipitation / R. Wimmer, G. Strumia, F. Holawe // Canad. .J. Forest Res. – 2000. – № 30. – P. 1691-1697.

203. Wonnacott H. Regression: A second course in statistics / H. Wonnacott, J. Wonnacott. – New York: John Wiley and Sons, 1981. – P. 200-211.

204. Wuethrich B. Climate chnage: new center gives Japan an Arctic Toehold / B. Wuethrich // Science. – 1999. – V. 285. – № 5435. – P. 1827.

205. Zanzi A. **S**pectral analysis on mountain pine tree-ring chronologies. / A. Zanzi, M. Pelfini, G. Muttoni, M. Santilli, G. Leonelli // Dendrochronologia, 2007 – V. 24 – P.145–154.

206. Yang R.S. The potential of Weibull–Type functions as flexible growth curves / R.S. Yang, A. Kozak, J.H.G. Smith // Canadian J Forest Res. – 1978. – № 8. – P.424-431.

207. Yasue K. The effect of climate on the variation in Maximum density of *Picea glehnii* Mast. and associated changes in tracheid dimensions / K. Yasue, R. Funada, K. Fukazawa, J. Ohtani // Dendrochronologia. – 1996. – №. 14. – P. 89-97.

208. Yasue K. The effect of tracheid dimensions on variations in maximum density of *Picea glehnii* and relationships to climatic factors / K. Yasue, R. Funada, O. Kobayashi, J. Ohtani // Trees. – 2000. – № 14. – P. 223-229.

209. Yiou P. Data-adaptive wavelets and multi-scale singular-spectrum analysis / P. Yiou, D. Sornette, M. Ghil // Physica D. – 2000. – V. 142. – P. 254-290.

210. Zhou L. Variations in Northern Vegetation Activity Inferred from Satellite Data of Vegetation Index During 1981 to 1999 / L. Zhou, C.J. Tucker, R.K. Kaufmann, D. Stayback, N.V. Shabanov, R.B. Myneni // J. Geophys. Res. – 2001. – № 106 (D17). – P 20069-20083.