

Федеральное государственное
бюджетное научное учреждение
«Федеральный исследовательский центр
Институт прикладной физики
Российской академии наук»

на правах рукописи

Селезнев Алексей Федорович

Развитие методов эмпирической реконструкции
распределённых систем с внешними воздействиями
и их приложение к исследованию динамики явления
Эль-Ниньо

1.3.4 – радиофизика

ДИССЕРТАЦИЯ

на соискание ученой степени
кандидата физико-математических наук

Научный руководитель
кандидат физико-математических наук
Мухин Дмитрий Николаевич

Нижний Новгород – 2022

Оглавление

Введение	4
1 Построение и оптимизация эмпирических моделей	14
1.1 Введение	14
1.2 Стохастическая модель оператора эволюции	16
1.2.1 Общий вид модели	16
1.2.2 Типы параметризаций модели	17
1.3 Байесов подход к оптимизации сложности модели оператора эволюции	18
1.3.1 Критерий оптимальности на основе байесовой обоснованности	18
1.3.2 Алгоритм построения оптимальной модели	21
1.3.3 Оценка статистической значимости на основе суррогатных данных	23
1.4 Пример: эмпирическая модель динамики потребительской активности общества в условиях пандемии COVID-19	24
1.4.1 Метод главных компонент	24
1.4.2 Построение эмпирической модели и анализ ее динамики	26
1.5 Заключение	38
2 Преодоление весеннего барьера при внутригодовом прогнозе изменчивости явления Эль-Ниньо с помощью эмпирических моделей	40
2.1 Введение	40
2.2 Атмосферный предиктор явления Эль-Ниньо	41
2.2.1 Совместный корреляционный анализ данных ТПО и атмосферных данных	41
2.2.2 Построение эмпирической модели динамики ТПО и анализ ее предсказательной способности	48
2.3 Предиктор явления Эль-Ниньо на основе данных теплосодержания верхнего слоя тропического Тихого океана	52

2.3.1	ЭОФ-анализ данных ТПО и теплосодержания верхнего слоя тропического Тихого океана	53
2.3.2	Построение совместной эмпирической модели динамики ТПО и теплосодержания	57
2.4	Заключение	65
3	Исследование межгодовой и вековой изменчивости явления Эль-Ниньо на основе эмпирических моделей	67
3.1	Введение	67
3.2	Структура связей между компонентами ЭНЮК на межгодовых масштабах	68
3.2.1	Межгодовые зависимости между аномалиями ТПО и теплосодержания верхнего слоя тропического Тихого океана	68
3.2.2	Эволюция характера межгодовых зависимостей в системе ЭНЮК с начала 20 века	73
3.3	Отклик ЭНЮК на вариации солнечной активности на столетних масштабах	78
3.3.1	Механизмы влияния солнечной активности на климат	78
3.3.2	Построение эмпирической модели и анализ ее динамики	80
3.4	Заключение	84
	Заключение	86
	Приложение	91
	Список публикаций по теме диссертации	93
	Литература	98

Введение

Актуальность работы. Задачи построения математических моделей различных процессов, происходящих в природе, возникают во многих областях современной науки. При решении этих задач исследуемый процесс часто рассматривается как последовательность состояний, порожденная динамической системой (ДС). При этом предполагается, что состояние в данный момент времени может быть определено как совокупность некоторых параметров (координат), и существует закон, определяющий эволюцию состояния с течением времени. Таким образом, построение модели исследуемого процесса сводится к реконструкции ДС, которая подразумевает введение пространства координат, однозначно определяющих ее состояние (фазовое пространство), и задание оператора эволюции [1]. Широко распространенным подходом к реконструкции ДС является построение их моделей в виде уравнений «из первых принципов»¹, которые формулируются на основе известных физических законов. Альтернативным подходом является так называемое эмпирическое моделирование ДС на основе наблюдаемых данных. В общем случае при эмпирическом подходе *не предполагается наличие какой-либо информации об устройстве исследуемой системы*, которая рассматривается как «черный ящик». В ситуации, когда уравнения, лежащие в основе исследуемой ДС неизвестны, либо не могут быть применены достаточно эффективно², эмпирические методы моделирования выходят на первый план.

Возможность реконструкции ДС напрямую по производимому ей сигналу была осознана к началу 80-х годов прошлого века. В работе [2] было выдвинуто предположение о том, что фазовый портрет многомерной ДС может быть восстановлен даже по скалярному временному ряду, если в качестве остальных координат вектора состояния использовать тот же самый ряд, сдвинутый по времени с некоторой задержкой нужное количество раз (метод координат с задержками). При этом было показано, что в получаемом таким образом фазовом пространстве можно коррект-

¹Уравнения движения среды, отдельных частиц, силовых полей, переноса излучения, химической кинетики, тепло и массопереноса и пр.

²Такая ситуация характерна для большинства природных систем - климата, живых систем, а также социо-экономических систем.

но оценить некоторые характеристики аттрактора исследуемой ДС. В дальнейшем эти предположения получили математическое подтверждение в целом ряде работ, наиболее известной из которых является статья Такенса [3]. Доказанные Такенсом теоремы обосновывают возможность построения топологически эквивалентной копии (вложения) аттрактора ДС – компактного d -мерного многообразия – в пространстве последовательных по времени значений произвольной (гладкой) скалярной функции ее фазовых переменных (функции измерений), достаточная размерность которого составляет $2d + 1$. Позже эти теоремы были обобщены для случаев, когда ДС определена на множестве дробной размерности [4], неавтономных и стохастических ДС [5], ДС с бесконечной размерностью фазового пространства [6], с учетом возможной погрешности в функции измерений [7] и ее многомерности [8]. Первые попытки глобальной реконструкции³ уравнений ДС по наблюдаемому скалярному временному ряду были предприняты в [10, 11]. К настоящему времени опубликовано огромное количество работ, посвященных разработке и приложениям различных подходов к реконструкции уравнений динамики системы по наблюдаемым временным рядам (см., например, работы [12–17]). Были созданы и успешно применены методы, эффективные для систем, принадлежащих важным частным классам: систем с известным оператором эволюции и скрытыми переменными [18–22], неавтономных систем [23–26], систем с запаздыванием [27], ансамблей связанных систем, в том числе, с запаздыванием (см., например, [28, 29]), с разреженными матрицами связей [30–32], нейроосцилляторов [33, 34] и др. [35]. В целом, стоит отметить, что задача реконструкции ДС по наблюдаемым временным рядам является некорректной обратной задачей и не существует универсального подхода к решению этой задачи в общем случае.

Настоящая диссертационная работа направлена на развитие и применение методов эмпирической реконструкции ДС в наиболее общей ситуации, характерной, в частности, для климатической системы Земли, когда наблюдаемые пространственно распределенные данные представляют собой мультимасштабный процесс, протекающий в присутствии внешних воздействий (эмиссии парниковых газов и вариации солнечной активности являются типичными примерами таких воздействий в случае климата Земли). При этом отсутствует информация как о размерности оператора эволюции, лежащего в основе данных, так и об уравнениях динамики. В такой ситуации первый необходимый шаг построения эмпирической модели – реконструкция фазового

³Методы глобальной реконструкции ДС предполагают построение оператора эволюции, действующего во всей области фазового пространства, восстанавливаемой по наблюдаемому временному ряду. Существуют также методы локальной реконструкции ДС, в которых для каждой точки фазового пространства строится свой оператор эволюции (см., например, [9]). В данной работе развиваются методы глобальной реконструкции ДС.

пространства по наблюдаемым данным – становится нетривиальным. Так, в случае климатических данных размерность временного ряда равна произведению числа узлов пространственной сетки на Земном шаре, в которых производится измерение, и числа измеряемых характеристик; для современных данных высокого разрешения эта величина может достигать десятков тысяч. Ясно, что, с одной стороны, такой массив данных содержит больше информации об исследуемой системе, чем каждый из составляющих его скалярных временных рядов. С другой стороны, максимальная размерность восстанавливаемого по данным фазового пространства (например методом координат с задержками), в котором можно получить статистически обоснованную аппроксимацию оператора эволюции, сильно ограничена из-за конечной длины временного ряда. В этой связи чрезвычайно важной является задача эффективного уменьшения (редукции) размерности наблюдаемых данных с целью выделения относительно небольшого числа переменных, содержащих информацию о ключевых свойствах наблюдаемой динамики. Как правило процедура редукции заключается в том, что исходные данные высокой размерности проецируются на многообразие существенно меньшей размерности, динамика на котором определяется процессами на тех пространственно-временных масштабах, которые наиболее хорошо представлены в наблюдаемом временном ряде. Таким образом, построение оператора эволюции по пространственно распределенным временным рядам (например по временным рядам, измеряемым в различных узлах пространственной сетки на Земном шаре) происходит в низкоразмерном пространстве, получаемом в ходе редукции исходных данных высокой размерности. При этом новые переменные, получаемые в ходе редукции данных, определяют, вообще говоря, лишь низкоразмерную проекцию полного фазового пространства ДС. В такой ситуации восстановление «настоящего» однозначного оператора эволюции ДС, действующего в исходном высокоразмерном фазовом пространстве становится невозможным. Кроме того, наблюдаемые временные ряды всегда измеряются с конечным шагом по времени, что приводит к потере информации о процессах с сопоставимыми (с шагом по времени) временными масштабами. Одним из распространенных способов описания неоднозначности восстанавливаемого по данным оператора эволюции и учета влияния плохо разрешенных процессов является построение моделей оператора эволюции в виде стохастической ДС [26, 36]. На данный момент существует большое количество работ, в которых предложены различные параметризации таких моделей. В некоторых из этих работ случайный оператор эволюции представляется в виде суммы линейной детерминированной части и стохастической части в виде гауссова случайного процесса, распределенного в пространстве [37–40]. Также были предложены различные нелинейные параметризации, в частности полиномы [41–45] и искусственные нейронные сети [46–50]. Резюмируя вышесказанное можно заклю-

чить, что успешное решение задачи реконструкции ДС по многомерным временным рядам может быть достигнуто, во-первых, за счет удачного выбора пространства фазовых переменных, получаемых в процессе редукции наблюдаемых высокоразмерных данных, и, во-вторых, за счет оптимальной параметризации оператора эволюции, действующего в этом пространстве.

В диссертации концепция эмпирической стохастической модели, развитая в работах [26, 47, 51], применяется в случае высокоразмерных пространственно-распределенных данных. В качестве основного инструмента редукции данных используется метод главных компонент [52], который позволяет представить наблюдаемый вектор данных в виде линейной комбинации ортогональных векторов (эмпирических ортогональных функций – ЭОФ [53]). Получаемые при этом коэффициенты разложения – главные компоненты – являются некоррелирующими друг с другом временными рядами и могут быть использованы в качестве фазовых переменных при построении эмпирической модели. Фактически, при такой процедуре, исходный высокоразмерный временной ряд линейно отображается (проецируется) в пространство меньшей размерности. Рассматриваются различные функциональные формы модели, включающие как стандартные линейные, так и нелинейные параметризации, допускающие, в том числе, учет внешних воздействий⁴ (форсингов) для анализа отклика исследуемой системы на них. Особое внимание уделяется вопросу об определении оптимальной сложности⁵ эмпирической модели, поскольку выбор той или иной параметризации модели по имеющейся выборке данных, а также введение в модель дополнительных факторов (форсингов) должны быть статистически обоснованы. Для поиска ответа на данный вопрос используются байесов формализм решения обратных задач и методы проверки статистических гипотез, суть которых подробно изложена в соответствующем разделе диссертации.

В качестве основного объекта приложения развиваемых в настоящей диссертационной работе методов эмпирического моделирования выступает явление Эль-Ниньо-Южное Колебание (ЭНЮК или просто колебание Эль-Ниньо), которое является наиболее существенной составляющей изменчивости климата Земли на межгодовых масштабах [54–56]. Проявляясь как сильные аномалии (отклонения от среднеклиматической нормы) поверхностной температуры воды на огромной площади в центральной и восточной части экваториального Тихого океана, колебание Эль-Ниньо оказывает влияние на погодные условия по всему Земному шару [57–59]. Собственно Эль-Ниньо традиционно называется теплой фазой колебания, соответствующая положительным

⁴Под ДС с внешним воздействием понимается система, оператор эволюции которой зависит от параметров, изменяющихся с течением времени независимо от динамики системы.

⁵Сложность модели определяется общим количеством ее параметров.

аномалиям поверхностной температуры, а противоположная ей холодная фаза называется Ла-Нинья. Как правило, сильные события Эль-Ниньо/Ла-Нинья происходят каждые 3-8 лет и продолжаются в течение одного года. При этом пик аномалий поверхностной температуры обычно приходится на зимние месяцы Северного полушария. В основе современных представлений о динамике ЭНЮК лежит несколько ключевых концептуальных элементов. Один из них – гипотеза Бьеркнеса [60] о положительной обратной связи между океаном и атмосферой, которая предполагает, что ослабление пассатов в ответ на повышение температуры поверхности океана (ТПО) приводит к еще большему повышению ТПО в восточной и центральной частях тропического Тихого океана. Идеи Бьеркнеса получили дальнейшее развитие в работах Виртки [61, 62], который предположил, что накопление теплых подповерхностных вод в экваториальной части Тихого океана является предпосылкой для возникновения теплой фазы ЭНЮК – Эль-Ниньо. Усиление пассатных ветров способствует накоплению теплых вод в западной части тихоокеанского бассейна, создавая перепад высоты уровня моря в восточном направлении. Избыточное количество теплой воды создает благоприятные условия для запуска обратной связи Бьеркнеса, приводящей к ослаблению пассатов за счет повышения ТПО, что способствует переносу накопленных теплых вод с запада на восток и наступлению Эль-Ниньо. На основе гипотез Бьеркнеса и Виртки были созданы теоретические концепции, объясняющие ярко выраженную циклическую природу ЭНЮК. Согласно так называемой теории «перезаряжающегося осциллятора» (англ. – recharge oscillator, в русскоязычной литературе упоминается как теория «загрузки-разгрузки» [63]) меридиональный перенос подповерхностных вод из высоких широт в сторону экватора или наоборот, обусловленный антициклонической (циклонической) завихренностью напряжения ветра (известный как Свердруповский перенос) определяет динамику глубины залегания термоклина (и, как следствие, теплосодержания) в экваториальном Тихом океане и обеспечивает переключение между фазами Эль-Ниньо и Ла-Нинья [64, 65]. Альтернативная теория ЭНЮК основана на моделях в виде осцилляторов с запаздыванием [66, 67], в которых подчеркнута роль экваториальных волн – переносчиков аномалий глубины залегания термоклина. Распространяясь в разных направлениях и отражаясь от западной и восточной границ Тихого океана, эти волны обеспечивают квазипериодические колебания поверхностной температуры, соответствующие разным фазам ЭНЮК. Физические процессы, лежащие в основе обеих упомянутых теорий, имеют место в моделях ЭНЮК промежуточной сложности – совместных моделях атмосферы и океана на основе уравнений мелкой воды [68–70]. Регулярные колебания, предсказываемые концептуальными моделями, не наблюдаются в реальности, поскольку эти модели не учитывают стохастические процессы – еще один важнейший фактор, определяющий динамику ЭНЮК [71–75].

Как правило, эти процессы ассоциируются с мелкомасштабным воздействием атмосферы, проявляющимся в виде аномалий зонального ветра в тропическом Тихом океане (например, всплески западных ветров [74, 76]). В качестве драйверов таких аномалий могут выступать колебания Маддена-Джулиана [77–79] или крупномасштабные флуктуации атмосферы в субтропических широтах [80, 81].

Продолжающееся последние десятилетия накопление качественных данных измерений различных климатических характеристик открывает широкие возможности как для эмпирической проверки существующих концепций динамики ЭНЮК, так и для построения прогностических моделей. Эмпирические или статистические модели становятся эффективным инструментом для внутригодового прогноза ЭНЮК; в этом отношении они могут конкурировать с динамическими моделями, основанными на уравнениях из первых принципов [82]. Центральной проблемой, возникающей при внутригодовом прогнозе ЭНЮК как с помощью эмпирических, так и с помощью динамических моделей, является наличие так называемого весеннего барьера предсказуемости, который существенно ограничивает горизонт прогноза климатической динамики в тропическом Тихом океане [83–86]. Многие эмпирические модели [40, 43] основаны исключительно на данных ТПО, корректный прогноз которой является основной целью при прогностическом моделировании ЭНЮК [82]. Существование весеннего барьера в таких моделях обусловлено наблюдаемым резким снижением автокорреляций в данных ТПО, возникающим в конце весны – начале лета. Опираясь на современные теоретические концепции динамики ЭНЮК, многие исследования направлены на поиск дополнительных атмосферных и океанических предикторов, которые могут помочь в преодолении весеннего барьера при эмпирическом прогнозе ТПО. Были предложены предикторы, основанные на данных теплосодержания океана [63, 87–89], а также атмосферных данных [87, 89, 90]. Тем не менее до сих пор не существует общепринятого подхода как к получению статистически обоснованных предикторов по данным наблюдений, так и к учету этих предикторов в эмпирических прогностических моделях. К другим актуальным вопросам, касающимся динамики ЭНЮК, которые активно исследуются эмпирическими методами, можно отнести влияние на характеристики ЭНЮК (интенсивность, локализацию и пр.) естественной декадной изменчивости климата, тренда глобального потепления, вариаций солнечной активности [91–102].

Основной целью диссертации является развитие методов эмпирической реконструкции распределенных систем, эволюция которых происходит в присутствии внешних воздействий, и приложение этих методов для построения эффективных моделей динамики явления Эль-Ниньо в широком диапазоне временных масштабов. Для достижения поставленной цели в диссертации решались следующие **задачи**:

1. Разработка метода построения эмпирической модели, сочетающего редукцию наблюдаемых данных высокой размерности и реконструкцию оператора эволюции в редуцированном пространстве новых переменных; разработка методов оценки статистической обоснованности выводов, сделанных на основе эмпирической модели.
2. Поиск эффективных атмосферных и океанических предикторов, позволяющих снизить весенний барьер при эмпирическом прогнозе поверхностной температуры экваториального Тихого океана; построение на основе найденных предикторов прогностической модели внутригодовой изменчивости явления Эль-Ниньо и анализ ее предсказательной способности.
3. Анализ связей между компонентами явления Эль-Ниньо на межгодовых масштабах; построение эмпирической модели изменчивости явления Эль-Ниньо на вековых масштабах и исследование с ее помощью связи температуры поверхности экваториального Тихого океана с сигналом солнечной активности.

Структура диссертации. Диссертация состоит из введения, трех глав, заключения, приложения, списка публикаций по теме диссертации и списка цитируемой литературы. Общий объем диссертации 113 страниц, включая 30 рисунков и список литературы из 160 наименований.

В **первой главе** диссертационной работы излагается метод построения стохастической модели оператора эволюции по многомерному временному ряду наблюдаемых данных: разбирается метод главных компонент, который используется в качестве основного инструмента предварительной редукции данных, формулируется байесов подход к обучению и оптимизации сложности эмпирической модели, обсуждается численный алгоритм его реализации. Рассматриваются различные функциональные формы модели, поясняются смысл каждого из структурных параметров, соответствующих выбранной функциональной форме, а также ограничения, накладываемые на ее параметры с целью регуляризации решения обратной задачи реконструкции. Отдельный раздел главы посвящен описанию методов, используемых для оценки статистической обоснованности результатов, полученных на основе эмпирической модели.

Возможности развиваемых методов реконструкции демонстрируются на примере построения эмпирических прогностических моделей потребительской активности общества в условиях пандемии COVID-19 в крупнейших городах Российской Федерации по временным рядам от банка Tinkoff, охватывающим интервал с января 2019 по июль 2021 года. Важнейшей задачей при построении прогностической модели

является правильный учет отклика активности общества на различные ограничения, вводимые властями. Для решения этой задачи была проанализирована база данных Oxford COVID-19 Government Response Tracker, составленная исследователями из Оксфордского университета, которая представляет собой набор индексов, характеризующих правительственную политику, связанную с закрытием мест массового скопления людей и другими мерами сдерживания эпидемии. Показывается, что учет в эмпирической модели информации об интенсивности ограничительных мер является обоснованным в рамках предложенного байесова подхода к определению оптимальности модели и статистически значимо улучшает ее прогностические способности для некоторых категорий потребительской активности.

Во **второй главе** диссертации решается задача построения эмпирической модели для внутригодового прогноза изменчивости явления Эль-Ниньо. Проводится анализ пространственно-распределенных данных давления на уровне моря и теплосодержания верхнего слоя океана на предмет выявления прогностических переменных (предикторов), позволяющих снизить весенний барьер предсказуемости климатической динамики в тропической части Тихого океана. По данным давления на уровне моря рассчитывается индекс, отражающий состояние субтропической атмосферной циркуляции в зимне-весенний сезон (февраль-март), который имеет статистически значимые длинные корреляции с колебанием Эль-Ниньо в предстоящий период с июня текущего года по май следующего года. Устанавливается, что найденный индекс может быть связан с влиянием на динамику Эль-Ниньо крупномасштабного паттерна атмосферной циркуляции, охватывающего среднеширотную часть северного Тихого океана. В ходе анализа данных теплосодержания океана выявляется индекс, который может характеризовать динамику подповерхностных теплых вод в центральной и западной части тропического Тихого океана, которая, согласно современным концепциям, является ключевой составляющей цикла колебания Эль-Ниньо. Показывается, что найденный индекс имеет четкий фазовый сдвиг относительно традиционных индексов явления Эль-Ниньо, основанных на данных ТПО, и, в отличие от них, характеризуется зимним барьером предсказуемости. Демонстрируется, что учет найденных предикторов в эмпирической модели Эль-Ниньо, описывающей эволюцию температуры поверхности экваториальной части Тихого океана, позволяет добиться статистически значимого улучшения ее прогностической способности на масштабах от нескольких месяцев до одного года, в том числе при прогнозе через весенний барьер. При этом, индекс, выявленный по данным давления на уровне моря, играет роль атмосферного форсинга, влияющего на долгосрочное состояние системы, а индекс, найденный по данным теплосодержания, является динамической переменной системы, дополняющей ТПО-переменную.

В **третьей главе** исследуется динамика ЭНЮК на больших (межгодовых и вековых) временных масштабах. На основе предикторов, найденных в предыдущей главе, анализируются связи между компонентами Эль-Ниньо на межгодовых масштабах. Показывается, что структура этих связей меняется с течением времени от линейной в начале 20 века к нелинейной в современном климате. Вековая изменчивость ЭНЮК исследуется путем построения эмпирической модели по временным рядам индексов явления Эль-Ниньо Nino 3.4 (основанным на трех различных базах данных), охватывающим интервал 850 лет с шагом один год. Показывается, что использование сигнала, характеризующего вариации солнечной активности в комбинации с сигналом, отражающим содержание углекислого газа в атмосфере в качестве двумерного форсинга модели является статистически обоснованным. С помощью построенной модели подтверждается связь солнечной активности с индексом ЭНЮК на масштабах порядка 200 лет.

В **Заключении** перечислены **основные результаты диссертации**, сформулированы **положения, выносимые на защиту**, а также **научная новизна и практическая ценность** работы.

В **Приложении** изложен подробный вывод ценовых функций, используемых для оптимизации сложности и нахождения неизвестных параметров эмпирической модели в рамках используемого в диссертации байесова подхода.

Апробация представленных в работе результатов. Основные результаты работы докладывались на семинарах и конкурсах молодых ученых в Институте прикладной физики РАН, а также на российских и международных научных конференциях: Сессии РАН по нелинейной динамике (Москва, 2016, 2018), Всероссийской конференции молодых ученых «Состав атмосферы. Атмосферное электричество. Климатические процессы» (Нижний Новгород, 2016, Борок, 2017), Всероссийской научной школе «Нелинейные волны» (Нижний Новгород, 2016, 2018), международных симпозиумах «Забабахинские научные чтения» (Снежинск, 2021), «Analysis of dynamic networks and data driven modelling of the climate» (Potsdam, Germany, 2015), «Science of the future» (Казань, 2016), «Frontiers of nonlinear physics» (Нижний Новгород – Санкт-Петербург, 2016), «Topical problems of nonlinear wave physics» (Москва – Санкт-Петербург, 2017), «European Geosciences Union General Assembly» (Vienna, Austria, 2016, 2018, 2021), «American Geophysical Union Fall Meeting» (New Orleans, USA, 2017), «Fourth International Conference on Earth System Modelling» (Hamburg, Germany, 2017), «JpGU-AGU Joint Meeting» (Makuhari Messe, Japan, 2017), «Dynamics Days» (MD – Maryland, USA, 2017), «IUGG Conference on mathematical geophysics» (Нижний Новгород, 2018).

Результаты, представленные в диссертации, рекомендованы в Отчеты РАН за 2018

и 2021 годы. Проведенные исследования были поддержаны грантом Правительства РФ «Новые подходы к исследованию климатических процессов и прогнозу экстремальных явлений» (14.Z50.31.0033), а также грантами РНФ (16-12-10198, 18-12-00231, 19-42-04121) и РФФИ (19-02-00502 А). Эмпирическая модель потребительской активности общества в условиях пандемии COVID-19 построена в ходе выполнения работ в рамках Соглашения # 075-11-2020-011 от 19.10.2020 с Министерством науки и высшего образования Российской Федерации по реализации научного проекта «Моделирование эпидемий вирусных инфекций». По теме диссертации опубликовано 33 работы, в том числе 6 статей в рецензируемых научных журналах из списка ВАК [А1–А6], 4 свидетельства о регистрации программы для ЭВМ [А7–А10], 2 работы в сборниках трудов конференций [А11, А12] и 22 тезисов и абстрактов [А13–А33].

Глава 1

Построение и оптимизация эмпирических моделей

1.1 Введение

Построение математической модели динамической системы (ДС) по наблюдаемому временному ряду (эмпирическое моделирование) предполагает два ключевых шага: (i) определение по временному ряду набора переменных, образующих ее фазовое пространство (фазовые переменные) и (ii) задание оператора эволюции состояний в этом пространстве. В наиболее общей ситуации, характерной для природных систем, у исследователя отсутствует информация о размерности фазового пространства реконструируемой ДС. При этом наблюдаемые временные ряды всегда ограничены по времени, что не позволяет получить статистически обоснованную оценку размерности фазового пространства¹ в том случае, если они порождены ДС с большим числом степеней свободы. Таким образом, при эмпирической реконструкции высокоразмерных ДС (таких как климатическая система) построение полного набора фазовых переменных по наблюдаемому временному ряду конечной длины технически невозможно и модель оператора эволюции неизбежно строится в некотором подпространстве (проекции) полного фазового пространства исследуемой ДС. Ясно, что в общем случае однозначного (детерминированного) оператора эволюции в таком подпространстве не существует. Для учета такого рода неоднозначности можно использовать модели оператора эволюции в виде случайной (стохастической) ДС, способ построения которых излагается в разделе 1.2 данной главы.

Задача восстановления модели оператора эволюции по наблюдаемой ограниченной

¹Например с использованием хорошо известных методов оценки размерностей аттракторов ДС по временным рядам [103, 104].

выборке данных является, вообще говоря, некорректно поставленной. Распространенный подход к решению подобных задач основан на построении так называемой ценовой функции, которая определяет меру соответствия выбранной параметризации модели наблюдаемым данным. Неизвестные параметры модели при таком подходе находятся путем оптимизации по ним выбранной ценовой функции. В качестве примеров реализаций данного подхода, которые широко применяются на практике, можно привести метод наименьших квадратов [105, 106] и его различные обобщения [107, 108]. Можно показать, что ценовые функции, лежащие в основе упомянутых методов, могут быть получены² в рамках более общего байесова подхода к решению обратных задач, который подробно изложен в разделе 1.3 и будет использоваться нами для определения оптимальной модели оператора эволюции по имеющейся выборке данных. В этом же разделе обсуждаются методы оценки статистической значимости результатов, получаемых с помощью эмпирической модели.

Возвращаясь к задаче построения фазовых переменных отметим, что в случае многомерного временного ряда она тесно связана с задачей эффективного уменьшения его размерности для нахождения относительно небольшого числа переменных, которые содержат основную информацию о динамике ДС на интересующих исследователя временных масштабах. Пространство, образуемое этими переменными, часто трактуется как некое «главное многообразие» [110], которое «встроено» в многомерное пространство исходных данных. Дальнейшее построение модели оператора эволюции происходит на главном многообразии, которое, в таком случае, и является ее фазовым пространством, восстанавливаемым по наблюдаемым данным. В разделе 1.4 данной главы обсуждается метод главных компонент, идея которого состоит в аппроксимации данных линейными многообразиями меньшей размерности и который будет использоваться нами для анализа многомерных данных с целью построения фазовых переменных и моделей оператора эволюции по ним. В этом же разделе возможности развитого эмпирического подхода к реконструкции ДС демонстрируются на примере построения модели по данным потребительской активности общества в условиях пандемии COVID-19.

Материалы данной главы изложены в статьях [A1–A3] и публикациях [A11, A13–A29, A32, A33].

²При соответствующих предположениях о дефекте модели [109], величине шума измерений и т.д.

1.2 Стохастическая модель оператора эволюции

1.2.1 Общий вид модели

Пусть имеющийся временной ряд данных $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_N)$ представляет собой набор векторов $\mathbf{x}_n \in \mathbb{R}^D$, каждый из которых содержит измерения некоторой физической величины в один из N эквидистантных моментов времени t_1, \dots, t_N . В случае климатических данных величина D может характеризовать число точек пространственной сетки на Земном шаре, в которых известны значения наблюдаемой величины. Исходя из предположения о том, что данные \mathbf{X} порождены динамической системой (ДС) будем строить по ним модель оператора эволюции в виде стохастического процесса [26, 47, 51]:

$$\mathbf{y}_n = \mathbf{f}(\mathbf{y}_{n-1}, \dots, \mathbf{y}_{n-l}) + \widehat{\mathbf{g}} \cdot \boldsymbol{\xi}_n. \quad (1.1)$$

Здесь $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_N)$, $\mathbf{y}_n \in \mathbb{R}^d$ – фазовые переменные модели оператора эволюции исследуемой ДС, которые сконструированы по данным \mathbf{X} . Без ограничения общности будем предполагать, что $\langle y_k \rangle = \frac{1}{N} \sum_{n=1}^N y_{k,n} = 0$ и $\langle y_k \rangle^2 = \frac{1}{N} \sum_{n=1}^N (y_{k,n} - \langle y_k \rangle)^2 = 1$ (здесь $y_{k,n}$ – k -я компонента вектора \mathbf{y}_n , $k = 1, \dots, d$). Детерминированная часть модели \mathbf{f} представляет собой дискретное отображение, которое зависит от l последовательных состояний в фазовом пространстве. Стохастическая часть модели описывает вклад процессов с малым временем автокорреляции, которые плохо разрешены в наблюдаемых данных (например, процессов, масштабы которых близки к шагу дискретизации данных по времени). Она выражается в виде произведения нижнетреугольной $d \times d$ матрицы $\widehat{\mathbf{g}}$ и вектора $\boldsymbol{\xi}_n \in \mathbb{R}^d$, компоненты которого представляют собой реализации гауссовых случайных некоррелированных (в пространстве и времени) процессов с нулевыми средними и единичными дисперсиями. Случайное возмущение, определяемое стохастической частью, распределено³ в фазовом пространстве модели и характеризуется ковариационной матрицей $\widehat{\mathbf{g}}\widehat{\mathbf{g}}^T$. Параметры функций \mathbf{f} и $\widehat{\mathbf{g}}$ неизвестны *априори* и оцениваются в процессе оптимизации ценовой функции, получаемой на основе байесова подхода (см. раздел 1.3).

³Заметим, что матричная функция $\widehat{\mathbf{g}}$, которая определяет амплитуду стохастической части модели, может явно зависеть от состояния в фазовом пространстве. Такое представление может быть более эффективным при реконструкции оператора эволюции в ситуации, когда в восстановленном по наблюдаемым данным фазовом пространстве присутствуют области с разной степенью неоднозначности зависимости текущего состояния от предыдущих [26]. В диссертации будет использоваться модель с постоянной амплитудой стохастической части.

1.2.2 Типы параметризаций модели

В диссертации будут использоваться два типа параметризации детерминированной части модели (1.1).

1. Линейная параметризация:

$$\mathbf{f}(\mathbf{z}_n) = \mathbf{A}_n \mathbf{z}_n + \mathbf{B}_n \mathbf{c}_n, \quad (1.2)$$

где вектор $\mathbf{z}_n \in \mathbb{R}^{ld}$ содержит компоненты векторов $\mathbf{y}_{n-1}, \dots, \mathbf{y}_{n-l}$, вектор $\mathbf{c}_n \in \mathbb{R}^{d_c}$ характеризует возможное внешнее воздействие (предполагается, что временной ряд этого воздействия известен) на реконструируемую ДС, \mathbf{A}_n и \mathbf{B}_n матрицы коэффициентов размеров $d \times ld$ и $d \times d_c$ соответственно, которые могут зависеть от момента времени t_n .

2. Нелинейная параметризация в виде трехслойного персептрона (вид искусственных нейронных сетей – ИНС). Структура персептрона устроена таким образом, что сигнал со входного слоя сначала подается на внутренний слой, который представлен нейронами с нелинейной функцией активации (в данной работе используется гиперболический тангенс), а затем на выходном слое формируется линейная комбинация этих нейронов:

$$\mathbf{f}(\mathbf{z}_n) = \sum_{i=1}^m \alpha_i \tanh(\boldsymbol{\omega}_i^T \mathbf{z}_n + \boldsymbol{\delta}_i^T \mathbf{c}_n + \gamma_i). \quad (1.3)$$

Здесь $\alpha_i \in \mathbb{R}$, $\boldsymbol{\omega}_i \in \mathbb{R}^{ld}$, $\boldsymbol{\delta}_i \in \mathbb{R}^{d_c}$, $\gamma_i \in \mathbb{R}$ – коэффициенты. Функция вида (1.3) способна аппроксимировать произвольную нелинейную зависимость за счет увеличения числа нейронов m [111].

Функции (1.2)-(1.3) характеризуются набором структурных параметров или гиперпараметров. В случае линейной параметризации (1.2) таким гиперпараметром⁴ является глубина памяти l . В случае нелинейной параметризации (1.3) к нему добавляется еще один гиперпараметр – m , задающий число нейронов в скрытом слое персептрона. Значения гиперпараметров определяют число параметров модели, а следовательно – ее сложность и аппроксимационные способности. Ясно, что за счет увеличения числа параметров можно добиться точной аппроксимации моделью наблюдаемого временного ряда. Однако практическая ценность такой модели будет невелика, поскольку она стремится описать лишь конкретную выборку, а не динамические свойства реконструируемой системы. Таким образом, вопрос о выборе

⁴Конкретный вид зависимости коэффициентов функции (1.2) от времени может характеризоваться своим набором гиперпараметров.

оптимальной сложности модели по имеющейся выборке данных является ключевым при реконструкции неизвестной ДС с помощью предложенной модели. Ответ на этот вопрос в данной работе дается с помощью байесова критерия оптимальности, который обсуждается в следующем разделе главы.

1.3 Байесов подход к оптимизации сложности модели оператора эволюции

1.3.1 Критерий оптимальности на основе байесовой обоснованности

Выражение для байесовой обоснованности

В данном разделе обсуждается байесов подход, который мы используем для оптимизации сложности стохастической модели (1.1). В рамках данного подхода задача поиска оптимальной эмпирической модели сводится к одной из известных задач математической статистики – определению неизвестных параметров распределения по имеющейся выборке данных. Предполагается, что оптимальная модель, построенная по данным, представляет собой разумный баланс между «слишком простой» моделью, плохо описывающей данные, и «слишком сложной» моделью, которая содержит неоправданно большое количество параметров и имеет тенденцию к тому, чтобы максимально точно описать конкретную выборку, а не захватить законы, лежащие в основе наблюдаемой динамики. Пусть $\mathbf{H} = \{H_1, H_2, \dots, H_i, \dots\}$ – множество возможных гипотез о сложности модели. В случае стохастической модели (1.1) каждая гипотеза H_i определяется конкретной комбинацией гиперпараметров, соответствующих ее выбранной параметризации. Согласно теореме Байеса условная вероятность $P(H_i|\mathbf{Y})$ того, что данная модель H_i лежит в основе временного ряда \mathbf{Y} выражается как:

$$P(H_i|\mathbf{Y}) = \frac{P(\mathbf{Y}|H_i)d\mathbf{Y} \cdot P(H_i)}{\sum_i P(\mathbf{Y}|H_i)d\mathbf{Y} \cdot P(H_i)}. \quad (1.4)$$

Условная плотность вероятности (ПВ) $P(\mathbf{Y}|H_i)$ называется байесовой обоснованностью модели H_i [112], которая характеризует вероятность того, что временной ряд \mathbf{Y} принадлежит ансамблю всевозможных временных рядов, который может быть получен с помощью модели H_i ; $P(H_i)$ – априорная вероятность модели H_i . Знаменатель в (1.4) является нормировочным множителем, который не зависит от H_i . Предполагая, что все модели из \mathbf{H} априори равновероятны, выражение (1.4) можно переписать как $P(H_i|\mathbf{Y}) = \alpha P(\mathbf{Y}|H_i)$, где α не зависит от H_i . Определим байесов

критерий оптимальности модели

$$L = -\ln P(\mathbf{Y}|H_i), \quad (1.5)$$

минимизация которого приводит к максимизации ПВ $P(\mathbf{Y}|H_i)$. Критерий оптимальности (1.5) имеет понятную интерпретацию. Если модель H_i слишком проста, то множество наборов данных, которые она способна воспроизвести, сильно ограничено. Это означает, что наблюдаемые данные \mathbf{Y} почти наверняка лежат на хвосте распределения $P(\mathbf{Y}|H_i)$, т.е. вероятность того, что они могут быть воспроизведены такой моделью, мала. Напротив, слишком сложная модель из-за большого количества параметров способна с одинаковой вероятностью воспроизвести всевозможные варианты данных, в том числе и данные \mathbf{Y} , поэтому вероятность воспроизвести именно эти данные для такой модели также мала. Таким образом, согласно критерию (1.5) модель оптимальной сложности – это не слишком простая и не слишком сложная модель.

Выражение для байесовой обоснованности $P(\mathbf{Y}|H_i)$ может быть получено путем интегрирования соответствующей функции правдоподобия $P(\mathbf{Y}|\boldsymbol{\mu}_f, \boldsymbol{\mu}_{\hat{\mathbf{g}}}, H_i)$ по априорной вероятностной мере $P(\boldsymbol{\mu}_f, \boldsymbol{\mu}_{\hat{\mathbf{g}}}|H_i)d\boldsymbol{\mu}_f d\boldsymbol{\mu}_{\hat{\mathbf{g}}}$, введенной в пространстве параметров модели [112]:

$$P(\mathbf{Y}|H_i) = \int P(\mathbf{Y}|\boldsymbol{\mu}_f, \boldsymbol{\mu}_{\hat{\mathbf{g}}}, H_i) \cdot P(\boldsymbol{\mu}_f, \boldsymbol{\mu}_{\hat{\mathbf{g}}}|H_i)d\boldsymbol{\mu}_f d\boldsymbol{\mu}_{\hat{\mathbf{g}}}. \quad (1.6)$$

Здесь векторы $\boldsymbol{\mu}_f, \boldsymbol{\mu}_{\hat{\mathbf{g}}}$ содержат параметры детерминированной и стохастической части модели (1.1) соответственно.

Функция правдоподобия модели и априорные ограничения на ее параметры

С учетом того, что стохастическая часть модели (1.1) представляет собой дельта-коррелированный во времени гауссов процесс с амплитудой $\hat{\mathbf{g}}$, функция правдоподобия выражается как⁵:

$$P(\mathbf{Y}|\boldsymbol{\mu}_f, \boldsymbol{\mu}_{\hat{\mathbf{g}}}, H_i) = \prod_{n=1}^l P_{\mathcal{N}}(\mathbf{y}_n, \hat{\mathbf{I}}) \times \prod_{n=l+1}^N P_{\mathcal{N}}(\mathbf{y}_n - \mathbf{f}(\mathbf{y}_{n-1}, \dots, \mathbf{y}_{n-l}), \hat{\mathbf{g}}\hat{\mathbf{g}}^T). \quad (1.7)$$

Здесь $P_{\mathcal{N}}(\mathbf{u}, \hat{\Sigma}) := \frac{1}{\sqrt{(2\pi)^d |\hat{\Sigma}|}} \exp\left(-\frac{1}{2}\mathbf{u}^T \hat{\Sigma}^{-1} \mathbf{u}\right)$ – гауссова ПВ, $\mathbf{u} \in \mathbb{R}^d$, $\hat{\Sigma}$ – симметричная $d \times d$ матрица. Первая группа сомножителей в (1.7) определяет вклад в правдоподобие первых l векторов данных $\mathbf{y}_1, \dots, \mathbf{y}_l$ (начальное состояние модели), которые не воспроизводятся в рамках закона (1.1) и трактуются в рамках байесова подхода как

⁵Подробный вывод функции правдоподобия модели приводится в Приложении.

случайные величины, распределенные по нормальному закону с нулевыми средними и единичными дисперсиями⁶ (см. подробнее [51]).

Априорная ПВ $P(\boldsymbol{\mu}_f, \boldsymbol{\mu}_g | H_i)$ ограничивает область изменения параметров эмпирической модели. Удачный выбор априорной ПВ может предотвратить возможное вырождение в пространстве параметров для выбранной параметризации модели и, тем самым, обеспечить эффективную регуляризацию при решении обратной задачи эмпирической реконструкции. В качестве априорных ПВ параметров функций (1.2)-(1.3), задающих детерминированную часть модели, нами используются гауссовы распределения с нулевыми средними и своими дисперсиями для каждого из параметров. При этом все параметры предполагаются априори взаимно независимыми, поэтому итоговая априорная ПВ является произведением гауссовых ПВ каждого параметра. Дисперсии каждой из гауссовых ПВ, фактически, задают диапазон изменения параметров в функциях (1.2)-(1.3) и, тем самым, определяют поведение этих функций. Из каких соображений выбираются значения дисперсий для каждой группы параметров? В случае нелинейной параметризации (1.3) этот выбор, отчасти, продиктован следующим обстоятельством. Легко видеть, что когда аргумент гиперболического тангенса принимает большие значения (положительные или отрицательные), его значение асимптотически стремится к 1 и -1 соответственно, т.е. нейроны теряют чувствительность к изменению сигнала на входе. Это так называемые области насыщения, в которых градиент гиперболического тангенса стремится к 0 и обучение ИНС становится неэффективным. Аналогичные эффекты возникают и при использовании других типов функции активации нейрона [113]. Для того, чтобы частично избежать подобных проблем, можно ввести дисперсии гауссовых априорных распределений на параметры $\boldsymbol{\omega}$, $\boldsymbol{\delta}$ и γ функции (1.3) таким образом, чтобы они ограничивали область изменения аргументов гиперболических тангенсов вдали от областей насыщения. В диссертации они вводятся следующим образом: $\sigma_\omega^2 = \frac{1}{d \cdot l}$, $\sigma_\delta^2 = \frac{1}{d_c}$, $\sigma_\gamma^2 = 1$. Дисперсии параметров $\boldsymbol{\alpha}$ выбираются равными $\sigma_\alpha^2 = \frac{1}{m}$ для того, чтобы ограничить диапазон изменения выходного сигнала функции (1.3) с учетом нормировки данных \mathbf{Y} . Выбор дисперсий гауссовых априорных ПВ на параметры в случае линейной параметризации будет обсуждаться в следующей главе, где будет определен конкретный вид зависимости коэффициентов функции (1.2) от времени.

Задача поиска параметров стохастической части модели является хорошо обусловленной и не требует существенной регуляризации, поскольку максимизация правдоподобия (1.7) по $\hat{\mathbf{g}}$ при фиксированных параметрах $\boldsymbol{\mu}_f$ сразу приводит к ответу: $\hat{\mathbf{g}}\hat{\mathbf{g}}^T = \frac{1}{N-l} \sum_{n=l+1}^N (\mathbf{y}_n - \mathbf{f}(\mathbf{y}_{n-1}, \dots, \mathbf{y}_{n-l})) \cdot (\mathbf{y}_n - \mathbf{f}(\mathbf{y}_{n-1}, \dots, \mathbf{y}_{n-l}))^T$. Поэтому в качестве

⁶С учетом того, что ряд \mathbf{Y} центрирован и нормирован (см. раздел 1.2.1).

априорной ПВ для параметров стохастической части используется неинформативное константное априорное распределение.

1.3.2 Алгоритм построения оптимальной модели

Вычисление байесовой обоснованности

Точное вычисление интеграла (1.6) с правдоподобием (1.7) для рассматриваемых параметризаций модели (1.2)-(1.3) не представляется возможным по причине очень сложного устройства подынтегрального выражения. Для его приближенного вычисления нами используется метод Лапласа, идея которого состоит в оценке значения интеграла в окрестности максимума подынтегрального выражения. Обозначим минус логарифм подынтегрального выражения в (1.6) как $\Psi_{H_i}(\boldsymbol{\mu}_f, \boldsymbol{\mu}_{\hat{g}})$. Тогда само подынтегральное выражение можно переписать в тождественном виде:

$$P(\mathbf{Y}|\boldsymbol{\mu}_f, \boldsymbol{\mu}_{\hat{g}}, H_i) \cdot P(\boldsymbol{\mu}_f, \boldsymbol{\mu}_{\hat{g}}|H_i) = \exp(-\Psi_{H_i}(\boldsymbol{\mu}_f, \boldsymbol{\mu}_{\hat{g}})). \quad (1.8)$$

Поскольку выражение в показателе экспоненты пропорционально большому числу $\frac{(N-l)d}{2}$, $N \gg l$ (см. Приложение), можно взять интеграл (1.6) с учетом выражения (1.8) методом Лапласа, аналогично тому, как это делалось в работах [51, 114]. Для этого $\Psi_{H_i}(\boldsymbol{\mu}_f, \boldsymbol{\mu}_{\hat{g}})$ раскладывается в ряд Тейлора второго порядка в окрестности своего минимума⁷. В результате получается следующее выражение для байесова критерия оптимальности (1.5):

$$L = -\ln P(\mathbf{Y}|H_i) = \Psi_{H_i}(\overline{\boldsymbol{\mu}}_f, \overline{\boldsymbol{\mu}}_{\hat{g}}) + \frac{1}{2} \ln \left[\frac{1}{(2\pi)^M} |\nabla \nabla^T \Psi_{H_i}(\overline{\boldsymbol{\mu}}_f, \overline{\boldsymbol{\mu}}_{\hat{g}})| \right]. \quad (1.9)$$

Здесь $\Psi_{H_i}(\overline{\boldsymbol{\mu}}_f, \overline{\boldsymbol{\mu}}_{\hat{g}})$ — значение функции в точке минимума, M — полное число параметров модели, собранных в векторах $\boldsymbol{\mu}_f, \boldsymbol{\mu}_{\hat{g}}$, $\nabla \nabla^T \Psi_{H_i}(\overline{\boldsymbol{\mu}}_f, \overline{\boldsymbol{\mu}}_{\hat{g}})$ — $M \times M$ матрица вторых производных (гессиан) функции в точке минимума. Выражение (1.9) для байесова критерия оптимальности будет использоваться нами в качестве ценовой функции при построении эмпирической модели. Чем обусловлена эффективность выбранной ценовой функции? Как видно из выражения (1.9), его поведение определяется двумя слагаемыми. Первое слагаемое $\Psi_{H_i}(\overline{\boldsymbol{\mu}}_f, \overline{\boldsymbol{\mu}}_{\hat{g}})$ характеризует ошибку описания наблюдаемой выборки данных \mathbf{Y} моделью. Оно убывает с ростом сложности

⁷Вообще говоря, функция $\Psi_{H_i}(\boldsymbol{\mu}_f, \boldsymbol{\mu}_{\hat{g}})$ имеет не единственный минимум, однако предполагается, что вычисление интеграла происходит в ее наименьшем локальном минимуме, который вносит определяющий вклад в интеграл.

(количества параметров) модели. Второе слагаемое⁸ $\frac{1}{2} \ln \left[\frac{1}{(2\pi)^M} |\nabla \nabla^T \Psi_{H_i}(\overline{\mu}_f, \overline{\mu}_g)| \right]$, напротив, растет при увеличении количества параметров модели. Таким образом, первое слагаемое отсекает или «штрафует» слишком простые модели H_i , которые плохо соответствуют наблюдаемым данным, а второе слагаемое штрафует слишком сложные или переобученные модели. Минимум ценовой функции достигается за счет баланса этих членов и, таким образом, соответствует наиболее оптимальной модели.

Построение оптимальной модели

Поиск минимума ценовой функции (1.9) для построения оптимальной модели включает следующие основные шаги:

1. Генерация случайной комбинации гиперпараметров H_i и начальных условий параметров μ_f, μ_g в соответствии с априорной плотностью вероятности $P(\mu_f, \mu_g | H_i)$.
2. Обучение модели, которое состоит в поиске минимума функции $\Psi_{H_i}(\mu_f, \mu_g)$ при заданных гиперпараметрах H_i .
3. Расчет ценовой функции (1.9) в найденной точке $(\overline{\mu}_f, \overline{\mu}_g)$.
4. Многократное повторение шагов 1 - 3 для того, чтобы определить гиперпараметры H_i и соответствующие им параметры $(\overline{\mu}_f, \overline{\mu}_g)$, которые соответствуют наименьшему из локальных минимумов ценовой функции (1.9).

Для минимизации функции $\Psi_{H_i}(\mu_f, \mu_g)$ в данной работе используется комбинация двух известных численных методов. Первый из них – алгоритм градиентного спуска – приводит значения параметров функции в окрестность минимума. После этого применяется квазиньютоновский метод Бroyдена–Флетчера–Гольдфарба–Шанно [116], который обеспечивает окончательную сходимость к минимуму.

Вычислительная сложность представленного алгоритма связана с двумя основными проблемами: (1) большим количеством локальных минимумов ценовой функции при конкретных гиперпараметрах и (2) большим количеством всевозможных комбинаций гиперпараметров, которые требуется перебрать. Тем не менее, шаги 1-3 алгоритма выполняются независимо для различных комбинаций гиперпараметров H_i , что позволяет эффективно ускорить алгоритм за счет использования параллельных вычислений.

⁸В асимптотическом пределе $N \gg M$ это слагаемое эффективно оценивается как $\frac{1}{2} M \ln N$. В этом случае критерий оптимальности (1.9) аналогичен хорошо известному критерию Шварца, или байесовскому информационному критерию [115].

Критерием остановки алгоритма, как правило, служит повторяемость результата: число итераций на шаге 4 считается достаточным, когда несколько наиболее «хороших» моделей из насчитанного ансамбля (т.е. моделей, соответствующих минимальным значениям ценовой функции) начинают иметь близкие значения гиперпараметров и демонстрировать одинаковые динамические свойства.

1.3.3 Оценка статистической значимости на основе суррогатных данных

Основной целью при эмпирической реконструкции неизвестной динамической системы является построение такой модели, которая будет эффективна при решении практических задач. В рамках байесова подхода удается определить критерий для нахождения оптимальной модели по имеющейся выборке данных. Тем не менее, сам по себе факт, что конкретная модель H_i является более оптимальной по отношению к другой модели H_j с точки зрения критерия (1.5), вообще говоря, не означает, что H_i окажется более эффективной на практике. Рассмотрим следующий пример. Пусть необходимо определить, помогает ли учет некоторого внешнего сигнала (форсинга) \mathbf{c} в эмпирической модели, построенной по данным \mathbf{Y} , лучше прогнозировать эти данные? Ясно, что нельзя дать положительный ответ на этот вопрос только на основании того, что модель с учетом форсинга является более оптимальной с точки зрения критерия (1.5) – необходимо напрямую сравнить предсказательные способности моделей с учетом и без учета форсинга. При этом, в типичном случае данных \mathbf{Y} ограниченной протяженности, необходимо оценивать статистическую обоснованность полученных результатов. В диссертации для этого используется распространенный подход к проверке статистических гипотез на основе построения суррогатных данных [117]. Поясним суть этого подхода на рассмотренном выше примере. Пусть H_i и H_j – модели с форсингом и без соответственно, построенные по данным \mathbf{Y} . Требуется проверить, имеет ли модель H_i лучшую предсказательную способность по сравнению с моделью H_j . Для этого, с помощью модели *без форсинга* H_j производится ансамбль из K суррогатных временных рядов $\{\bar{\mathbf{Y}}_1, \bar{\mathbf{Y}}_2, \dots, \bar{\mathbf{Y}}_K\}$, причем длина каждого суррогата равна длине временного ряда данных \mathbf{Y} . В том случае, если модель H_j является достаточно «хорошей», временные ряды из суррогатного ансамбля корректно отражают некоторые свойства наблюдаемого временного ряда \mathbf{Y} (например, форму спектра). При этом, в суррогатных рядах *специально обеспечено отсутствие информации о сигнале форсинга*. Далее, по каждому из суррогатных временных рядов

рассчитывается предсказательная способность \mathbf{PS}^9 (для определенности будем считать, что большее значение \mathbf{PS} соответствует лучшей предсказательной способности) модели с форсингом H_i . Таким образом, получается статистический ансамбль из K значений \mathbf{PS} и постулируется нулевая гипотеза о том, что модель с учетом форсинга обладает *не лучшей* предсказательной способностью, по сравнению с моделью без учета этого форсинга. На следующем этапе задается так называемая критическая область, т.е. область значений критерия \mathbf{PS} , при которых нулевая гипотеза отвергается (в рассматриваемом примере, когда большее значение \mathbf{PS} соответствует лучшей предсказательной способности, критическая область называется правосторонней). В качестве границы критической области можно выбрать α -квантиль полученного распределения значений характеристики \mathbf{PS} , т.е. такое число \mathbf{PS}_α , что вероятность $P\{\mathbf{PS} < \mathbf{PS}_\alpha\} = \alpha$. Если величина \mathbf{PS} модели с форсингом, рассчитанная *по наблюдаемым данным* превышает значение \mathbf{PS}_α , то нулевая гипотеза отвергается на уровне значимости $p = 1 - \alpha$ и можно сделать вывод о том, что учет форсинга в эмпирической модели позволяет улучшить ее предсказательную способность. Представленный подход к оценке статистической значимости будет широко использоваться нами в дальнейшем.

1.4 Пример: эмпирическая модель динамики потребительской активности общества в условиях пандемии COVID-19

1.4.1 Метод главных компонент

Специфика реконструкции динамических систем по многомерным массивам данных состоит в том, что, с одной стороны, такие данные содержат важную информацию об исследуемой системе, которую необходимо учитывать при построении эмпирической модели. С другой стороны, как уже было отмечено, построение модели оператора эволюции в пространстве слишком большой размерности не представляется возможным. Для преодоления указанных трудностей при построении эмпирической модели в диссертационной работе применяется предварительная *линейная* редукция многомерных наблюдаемых данных на основе известного метода главных компонент [52]. Основная идея данного метода состоит в том, что исходные данные высокой размерности $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_N)$ (здесь используются те же обозначения, что и в разделе

⁹Конкретные примеры мер предсказательной способности эмпирической модели приведены в следующем разделе главы.

1.2.1) линейно проектируются на направления, задаваемые собственными векторами ковариационной матрицы $\mathbf{C}_X = \frac{1}{N} \sum_{n=1}^N \mathbf{x}_n \mathbf{x}_n^T$ ¹⁰:

$$\bar{\mathbf{p}}_n = \bar{\mathbf{V}}^T \mathbf{x}_n. \quad (1.10)$$

Получаемые таким образом новые переменные $\mathbf{P} = (\bar{\mathbf{p}}_1, \dots, \bar{\mathbf{p}}_N)$, $\bar{\mathbf{p}}_n \in \mathbb{R}^D$ называются главными компонентами (ГК) и представляют собой некоррелирующие во времени временные ряды, т.е. их матрица ковариаций $\mathbf{C}_P = \frac{1}{N} \sum_{n=1}^N \bar{\mathbf{p}}_n \bar{\mathbf{p}}_n^T$ диагональна. Столбцы $D \times D$ матрицы $\bar{\mathbf{V}}$ представляют собой собственные вектора матрицы \mathbf{C}_X и называются эмпирическими ортогональными функциями (ЭОФ) [53]. Они упорядочены по убыванию соответствующих собственных значений $\lambda_1, \dots, \lambda_D$ матрицы \mathbf{C}_X .

Соотношение (1.10) задает ортогональный поворот исходных данных \mathbf{X} , который направлен на то, чтобы временной ряд каждой из получаемых ГК $\{\bar{p}_{k,1}, \dots, \bar{p}_{k,N}\}$ ($\bar{p}_{k,n}$ – k -я компонента вектора $\bar{\mathbf{p}}_n$, $k = 1, \dots, D$) захватывал как можно большую часть полной вариации¹¹ данных, которая совпадает с соответствующим собственным значением $\lambda_k = \frac{1}{N} \sum_{n=1}^N (\bar{p}_{k,n} - \langle \bar{p}_k \rangle)^2$. Редукция размерности данных с помощью ЭОФ-разложения заключается в том, что после поворота данных в соответствии с (1.10) для дальнейшего анализа используются лишь первые $d < D$ ГК, которые захватывают основную часть вариации наблюдаемых данных:

$$\mathbf{p}_n = \mathbf{V}^T \mathbf{x}_n. \quad (1.11)$$

Здесь $D \times d$ матрица \mathbf{V} содержит d ведущих ЭОФ, $\mathbf{p}_n \in \mathbb{R}^d$ – соответствующие им ГК. Обратное преобразование ГК в физическое пространство данных определяется как:

$$\mathbf{x}_n = \mathbf{V} \mathbf{p}_n + \mathbf{r}_n, \quad (1.12)$$

где $\mathbf{r}_n = \mathbf{V}' \mathbf{p}'_n$ – остаток, который описывает часть вариации данных не захваченную ведущими ГК, \mathbf{V}' – $D \times D - d$ матрица, столбцы которой – остаточные ЭОФ, $\mathbf{p}'_n \in \mathbb{R}^{D-d}$ – соответствующие ГК.

Таким образом, построение эмпирической модели (1.1) происходит в d -мерном фазовом пространстве ведущих ГК, описывающих процессы, которые вносят основной энерговклад (в смысле захваченной вариации) в данные на рассматриваемых временных масштабах. При этом остаточные ГК могут содержать информацию о процессах

¹⁰Предполагается, что данные имеют нулевое среднее по времени, т.е. $\langle x_k \rangle = \frac{1}{N} \sum_{n=1}^N x_{k,n} = 0$, где $x_{k,n}$ – k -я компонента вектора \mathbf{x}_n , $k = 1, \dots, D$.

¹¹Под вариацией временного ряда здесь и далее понимается средний квадрат отклонения принимаемых им значений от его среднего значения. Под вариацией многомерного временного ряда понимается сумма вариаций составляющих его временных рядов.

с малым масштабом автокорреляции в наблюдаемой динамике, учет которой может быть важен с точки зрения краткосрочного прогнозирования траектории моделируемой системы. Иными словами, прогноз модели (1.1) в физическом пространстве данных \mathbf{X} может быть улучшен за счет учета динамики остаточных ГК, которые не входят в ее фазовое пространство. В диссертации мы строим простую дополнительную модель остаточных ГК $\{p'_{k,1}, \dots, p'_{k,N}\}$ ($k = 1, \dots, D - d$) в виде авторегрессионного процесса первого порядка AR(1):

$$p'_{k,n} = b_k p'_{k,n-1} + \sigma_k \cdot \eta_{k,n}. \quad (1.13)$$

Здесь $\eta_{k,n}$ — реализации дельта-коррелированных гауссовых процессов с нулевым средним и единичной дисперсией, b_k и σ_k — параметры, оцениваемые методом наименьших квадратов. Фактически, при таком описании остаточные ГК моделируются как независимые красные шумы. Добавление модели остаточных ГК в схему прогнозирования направлено на улучшение предсказательной способности эмпирической модели на масштабах сопоставимых с временем автокорреляции процессов, которые не захвачены ведущими ГК.

1.4.2 Построение эмпирической модели и анализ ее динамики

В данном разделе описаны результаты, полученные при построении эмпирической модели потребительской активности общества в условиях ограничений, вводимых властями. В качестве данных использовались временные ряды потребительской активности от банка Tinkoff на интервале с января 2019 года по июль 2021 года в 6 крупнейших городах европейской и азиатской частей Российской Федерации – Москве, Нижнем Новгороде, Калининграде, Екатеринбурге, Новосибирске, Хабаровске. Важнейшим аспектом при построении эмпирической модели является исследование с ее помощью отклика активности общества на различные ограничения, вводимые властями. Нами была проанализирована база данных Oxford COVID-19 Government Response Tracker [118], составленная исследователями из Оксфордского университета, которая представляет собой набор индексов, характеризующих правительственную политику, связанную с закрытием мест массового скопления людей и другими мерами сдерживания эпидемии, политику в области здравоохранения и экономическую политику для более чем 180 стран, а также субнациональных юрисдикций нескольких стран начиная с 1 января 2020 года. Эта база данных позволяет изучать влияние ограничительных мер, вводимых правительствами стран мира, на распространение пандемии COVID-19.

Описание данных

Для каждого из рассматриваемых городов Российской Федерации (РФ) данные X от банка Tinkoff представляют собой временные ряды индексов потребительской активности¹² в $D = 32$ категориях, измеряемые с временным шагом один день. Измеряемой величиной являются изменения (в процентах) потребительской активности в конкретной категории по отношению к ее среднему значению (своему для каждого города) за 2019 год. В исходных данных ярко выражен недельный цикл – сигнал с характерным периодом равным одной неделе. Этот сигнал не представляет интереса для моделирования, поэтому исходные временные ряды категорий потребительской активности были приведены к недельному шагу по времени путем их усреднения в скользящем гауссовом окне с соответствующей характерной шириной. Динамика временных рядов потребительской активности качественно совпадает для всех рассматриваемых городов. В частности, прослеживается резкий спад активности в отдельных категориях, который приходится на март 2020 года, когда в РФ в ответ на стремительный рост числа заражений начали вводиться серьезные ограничительные меры по сдерживанию пандемии COVID-19. В качестве примера на рисунке 1.1 показаны временные ряды индексов потребительской активности в городе Москве. На рисунке 1.2 показаны индексы из базы данных Oxford COVID-19 Government Response Tracker, характеризующие строгость различных ограничительных мер, вводимых в РФ начиная с января 2020 года. Бирюзовым цветом на данном графике показан обобщенный по всем типам ограничительных мер индекс, который в данном случае совпадает с главной компонентой, получаемой с помощью ЭОФ-разложения многокомпонентного ряда индексов ограничений различных типов, представленных на рисунке. Каждый из индексов измеряется по собственной шкале в баллах представляющих целое положительное число начиная с 0. Обобщенный индекс изменяется в пределах от 0 до 100, причем каждое значение представляет собой рациональное число. Данный индекс использовался нами в качестве заданного внешнего воздействия (форсинга) при построении эмпирической модели. Для этого временной ряд индекса с шагом неделя был синхронизован по времени с временными рядами потребительской активности. При этом значения индекса до 1 января 2020 года (до начала пандемии COVID-19) экстраполировались нулями. Длина исследуемых данных потребительской активности и индекса ограничений на интервале с 1 января 2019 года по 4 июля 2021 года составляет $N = 131$ неделю.

¹²Индексы учитывают данные о финансовых операциях в онлайн и офлайн более 25 млн клиентов экосистемы Тинькофф во всех регионах России. Данные были любезно предоставлены коллегами из института технической физики имени академика Е. И. Забабахина (Снежинск). Более подробное описание данных приведено на сайте банка Tinkoff: <https://www.tinkoff.ru/>.

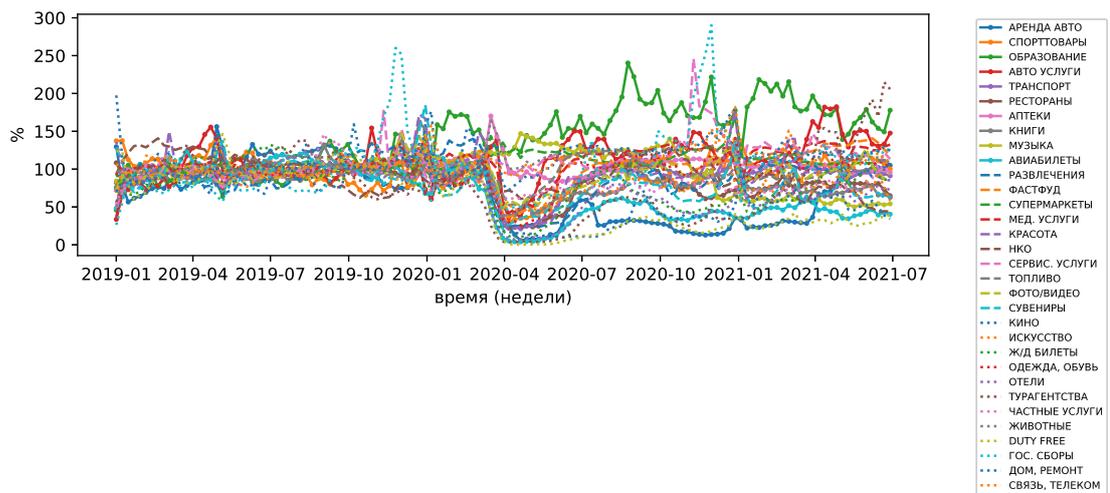


Рис. 1.1: Временные ряды индексов потребительской активности для различных категорий в Москве.

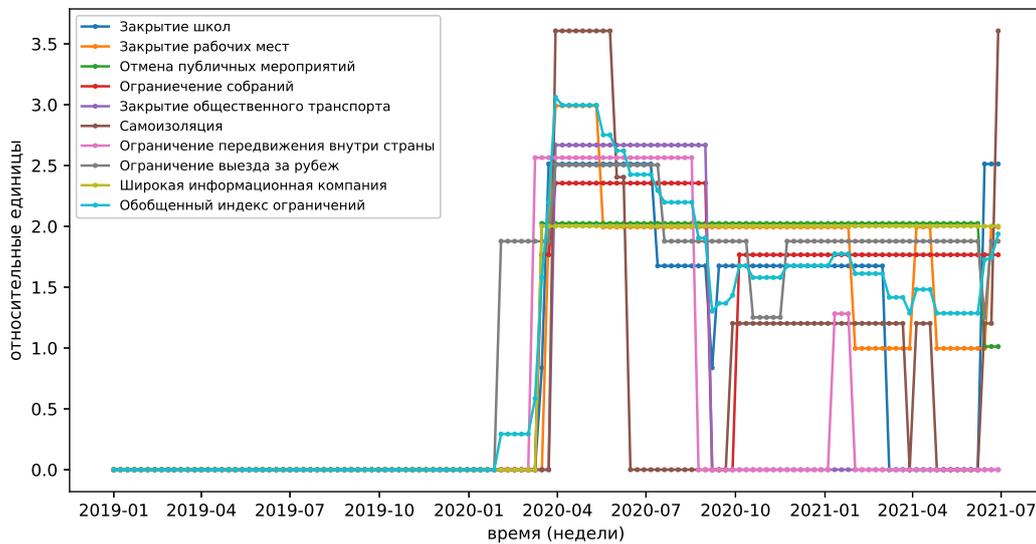


Рис. 1.2: Временные ряды индексов ограничений. Бирюзовым показан обобщенный индекс, используемый в качестве форсинга в эмпирической модели. Ряды приведены к единичной дисперсии путем их деления на свое среднеквадратичное отклонение.

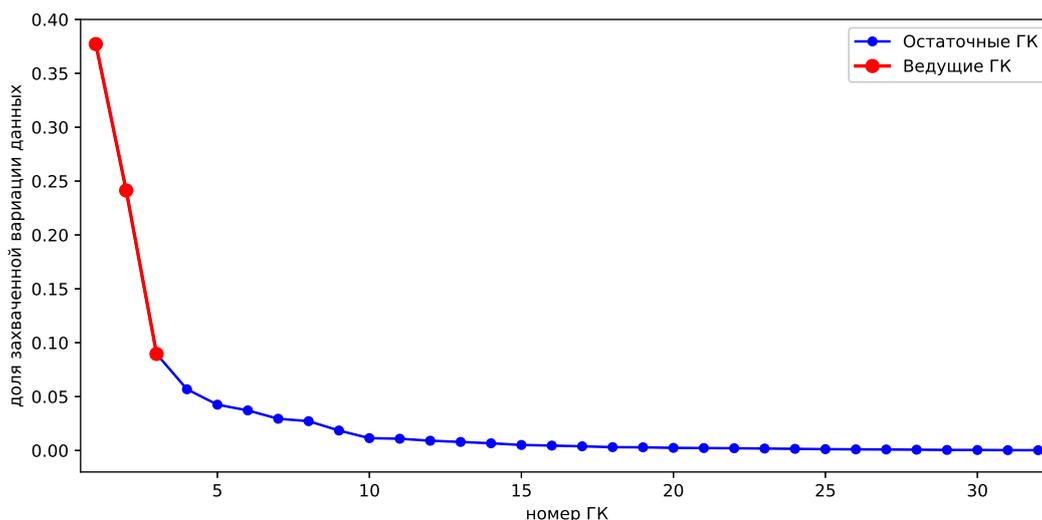


Рис. 1.3: Доли вариации данных $\frac{\lambda_k}{\sum_{i=1}^D \lambda_i}$, $k = 1, 2, \dots, D$, захваченные главными компонентами, в зависимости от номера компоненты. Красным показаны доли вариации, захваченные тремя ведущими главными компонентами.

Результаты моделирования

Для построения фазовых переменных эмпирической модели оператора эволюции по данным потребительской активности использовался метод главных компонент, описанный в разделе (1.4.1). На рисунке 1.3 показана доля вариации данных активности в Москве, захваченная главными компонентами (ГК) в зависимости от номера компоненты. Как видно из рисунка, три ведущих ГК захватывают суммарно около 71% вариации данных. При этом доля захваченной вариации начинает слабо отличаться для ГК начиная с четвертой. В этой связи, стохастическая модель с памятью (1.1) строилась в фазовом пространстве трех ведущих ГК, временные ряды которых (в случае города Москвы) показаны на рисунке 1.4, в то время как для остальных ГК строились независимые AR(1) модели (1.13). Заметим, что в паттернах трех ведущих ЭОФ, которые показаны на рисунке 1.5, прослеживается достаточно четкое разделение по группам категорий потребительской активности. Так в первой ЭОФ наиболее сильно представлены категории, связанные со сферой передвижений и в, частности, путешествий – «аренда авто», «транспорт», «Duty Free», «Авиа- и ж/д билеты», «отели», «туризм». Во второй ЭОФ выделяются категории массового потребления – «супермаркеты», «медицинские услуги», «авто услуги», «животные» и т.д. В третьей ЭОФ наиболее сильно по отношению к двум первым представлены такие

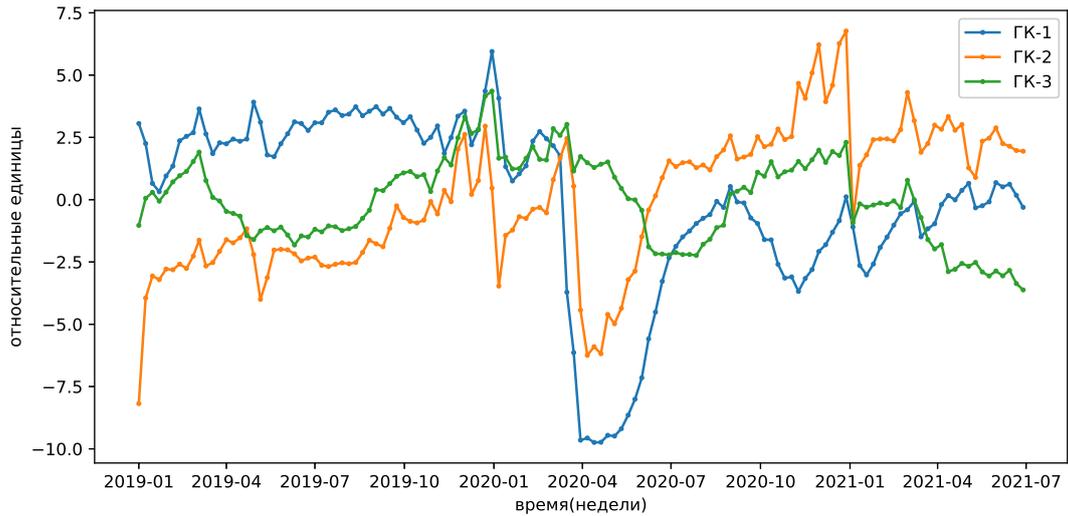


Рис. 1.4: Временные ряды трех ведущих главных компонент, полученные в результате ЭОФ-разложения данных потребительской активности в Москве.

категории как «книги» и «музыка». В ходе ЭОФ-анализа данных потребительской активности по другим городам были получены качественно аналогичные результаты – во всех случаях три ведущих ГК описывают около 70 % полной вариации данных, поэтому эмпирическая модель в этих городах строилась таким же образом.

Учитывая сильно ограниченную (2.5 года с шагом 1 неделя) протяженность имеющихся в наличии временных рядов, а также фактическое отсутствие априорной информации об устройстве исследуемой системы очевидно, что реконструкция достаточно сложных закономерностей (например статистически обоснованных нелинейных законов) по имеющимся данным невозможна. Поэтому при построении модели (1.1) в качестве функции, задающей ее детерминированную часть, была выбрана наиболее простая линейная параметризация (1.2) с постоянными (не зависящими от времени) коэффициентами. Таким образом, итоговая модель ведущих ГК потребительской активности представляет собой трехмерную авторегрессионную модель порядка l . Оптимальная глубина памяти модели l , а также параметры ее детерминированной и стохастической части заранее неизвестны и находятся в процессе обучения модели путем оптимизации ценовой функции (1.9). После нахождения оптимальной модели становится возможным рассчитать прогнозную оценку будущих значений потребительской активности во всех категориях¹³, а также построить доверительный

¹³Имея прогностическую оценку в пространстве ГК, ее всегда можно пересчитать в исходное 32-мерное пространство категорий с использованием соотношения (1.12).

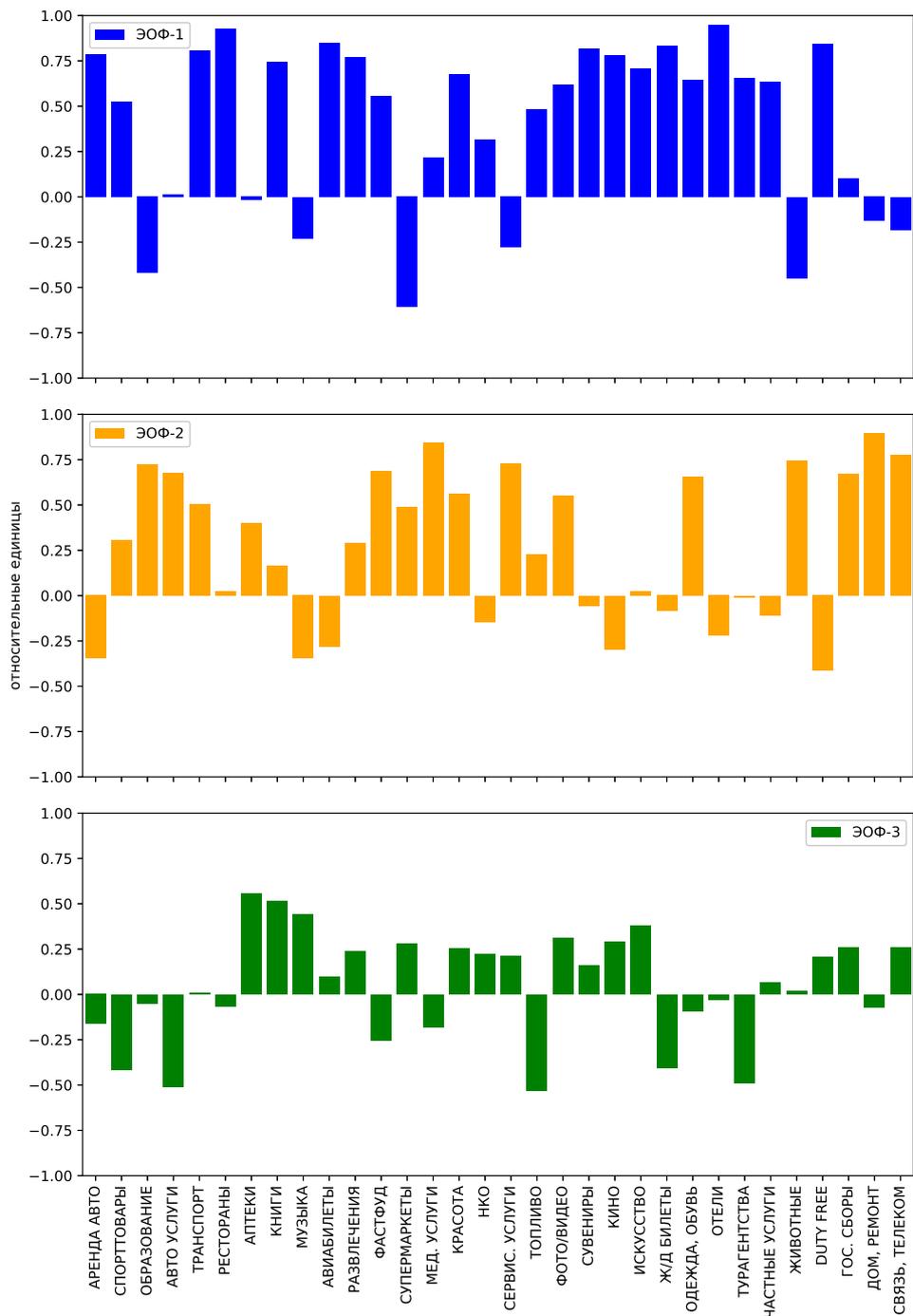


Рис. 1.5: Три ведущих ЭОФ данных потребительской активности в Москве. Перед применением процедуры ЭОФ-разложения временные ряды данных нормировались путем деления на свое среднеквадратичное отклонение.

интервал для этой оценки с заданным уровнем вероятности. Поскольку модель (1.1)-(1.13) – это случайная динамическая система, прогнозируемые ею значения являются случайными величинами. При прогнозировании на некоторое число шагов вперед модель итерировалась соответствующее количество раз с некоторыми реализациями своей стохастической части. При этом в качестве прогностических оценок берутся медианные значения распределений, вычисленных по ансамблю большого числа реализаций прогноза модели при разных шумовых реализациях. По тому же ансамблю оцениваются доверительные интервалы прогностических оценок.

Эмпирическая модель была обучена по данным потребительской активности в Москве, Нижнем Новгороде, Калининграде, Екатеринбурге, Новосибирске, Хабаровске. Было получено, что для всех рассматриваемых городов включение обобщенного индекса ограничений в эмпирическую модель (1.1)-(1.13) в качестве внешнего воздействия является статистически обоснованным с точки зрения критерия оптимальности (1.9) и улучшает прогностические свойства модели. В качестве меры предсказательной способности эмпирической модели нами использовались две характеристики. Первая – квадратный корень из среднеквадратичной ошибки прогноза (КСКО) – характеризует отличия между истинным и прогнозируемым значениями интересующей переменной x внутри интервала обучения модели:

$$e_j = \sqrt{\frac{\sum_{n=l}^{N-j} (x_{n+j} - \bar{x}_{n,j})^2}{N - j - l + 1}}. \quad (1.14)$$

Здесь j – дальность прогноза в неделях, x_{n+j} – истинное значение прогнозируемой переменной в момент времени t_{n+j} , $\bar{x}_{n,j}$ – значение переменной, предсказанное моделью, начиная с момента времени t_n . КСКО (1.14) для каждой из 32 категорий потребительской активности в городе Москве (горизонтальная ось) для различной дальности прогноза (вертикальная ось) показана на верхней панели рисунка 1.6. Еще одна характеристика предсказательной способности модели – коэффициент корреляции Пирсона между переменной и ее прогнозом:

$$r_j = \frac{\sum_{n=l}^{N-j} \Delta x_{n+j} \cdot \Delta \bar{x}_{n,j}}{\sqrt{\sum_{n=l}^{N-j} (\Delta x_{n+j})^2 \cdot \sum_{n=l}^{N-j} (\Delta \bar{x}_{n,j})^2}}, \quad (1.15)$$

где Δx_{n+j} и $\Delta \bar{x}_{n,j}$ – отклонения x_{n+j} и $\bar{x}_{n,j}$ от их средних значений. Меры (1.14)-(1.15) дополняют друг друга: в то время как КСКО характеризует количественное соответствие между реальными и прогнозируемыми значениями, коэффициент корреляции отражает способность модели предсказывать качественное поведение интересующей

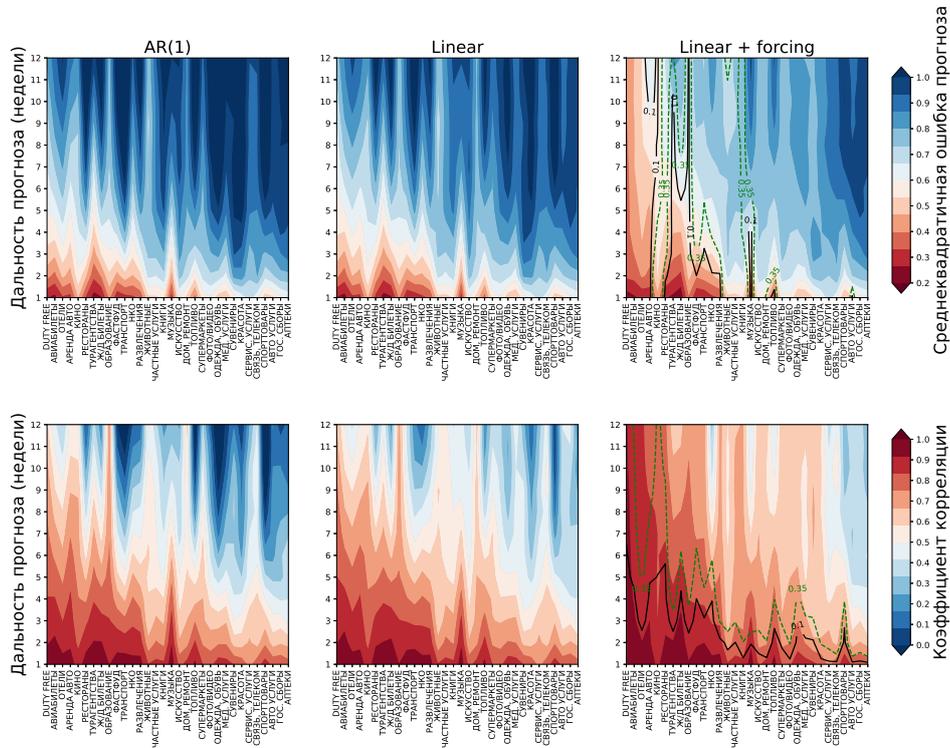


Рис. 1.6: Предсказательная способность различных эмпирических моделей динамики потребительской активности в городе Москве. Слева направо: независимые AR(1) модели для каждой из ГК, оптимальная линейная модель (1.1)-(1.13) без учета форсинга в виде индекса ограничений, оптимальная линейная модель (1.1)-(1.13) с учетом этого форсинга. Зеленая и черная линии на ограничивают области статистически значимого улучшения прогноза модели с форсингом по отношению к модели без форсинга на уровнях значимости $p = 0.35$ и $p = 0.1$ соответственно. При оценке границ областей значимости в случае КСКО задавалась левосторонняя критическая область, в случае коэффициента корреляции – правосторонняя.

переменной (убыль либо рост, смена знака производной и т.д.). Слева на рисунке 1.6 изображены меры предсказательной способности простейшей эмпирической модели, когда динамика каждой из ГК данных потребительской активности описывается *независимо* с помощью модели AR(1). Фактически, это соответствует описанию динамики потребительской активности с помощью модели красного шума. На средних панелях показана предсказательная способность модели (1.1)-(1.13) с оптимальной глубиной памяти $l = 1$, но без учета форсинга в виде индекса ограничений. Из рисунка видны улучшения предсказательной способности данной модели по сравнению с моделью красного шума, особенно при достаточно большой дальности прогноза. Справа на рисунке показана предсказательная способность модели (1.1)-(1.13) с учетом индекса ограничений и оптимальной глубиной памяти $l = 2$. Там же нанесены контуры статистически значимого (уровни значимости подписаны на контурах) улучшения прогноза по сравнению с моделью без учета ограничений (центральные панели). Для определения областей значимости использовался статистический тест, представленный в разделе (1.3.3). С помощью оптимальной модели без форсинга (1.1)-(1.13) с $l = 1$ был сгенерирован ансамбль из $K = 1000$ суррогатов трех ведущих ГК. Каждый суррогат представляет собой стохастическую реализацию модели, запущенной со случайной начальной точки. Длина каждого суррогатного временного ряда равна длине временного ряда данных потребительской активности $N = 131$ неделя. Модель с форсингом (1.1)-(1.13) и оптимальной длиной памяти $l = 2$ обучалась по каждому из временных рядов суррогатного ансамбля, после чего на них вычислялись меры (1.14)-(1.15) в физическом пространстве категорий потребительской активности. Используя полученный ансамбль суррогатных значений меры предсказательной способности, можно определить области, в которых значения меры предсказательной способности оптимальной модели с форсингом, построенной по реальным данным, лежат на хвосте распределения суррогатных значений. Таким образом, нулевая гипотеза, которая отвергается в данном статистическом тесте, предполагает, что модель, учитывающая информацию об ограничительных мерах, обладает не лучшей предсказательной способностью, чем модель не учитывающая эту информацию. Как видно из рисунка 1.6, наиболее значимые улучшения предсказательной способности эмпирической модели, учитывающей индекс ограничительных мер происходят при прогнозе категорий, связанных со сферой передвижений. Именно эти категории наиболее сильно представлены в первой ЭОФ данных потребительской активности (см. рисунок 1.5).

На рисунке 1.7 на примере Москвы показаны траектория системы в пространстве трех ГК данных потребительской активности в сравнении с не привязанными к данным траекториями (длинным случайным реализациям), полученными путем итерирования моделей (1.1)-(1.13) с форсингом (внизу) и без (вверху). Видно, что

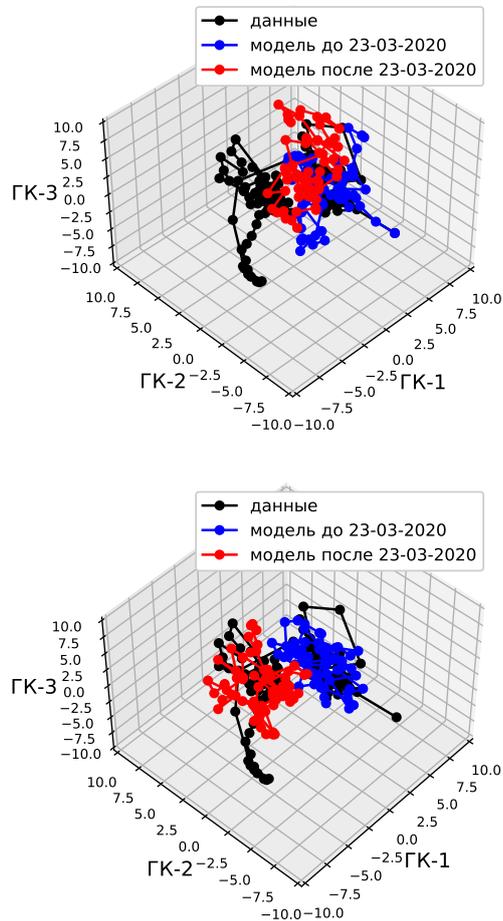


Рис. 1.7: Фазовые траектории системы (черный) и эмпирических моделей (красный и синий) в пространстве трех ведущих главных компонент. Синий цвет соответствует периоду времени до начала пандемии, красный — после. Верху — оптимальная линейная модель (1.1)-(1.13) без учета форсинга в виде индекса ограничений, внизу — оптимальная линейная модель (1.1)-(1.13) с учетом этого форсинга.

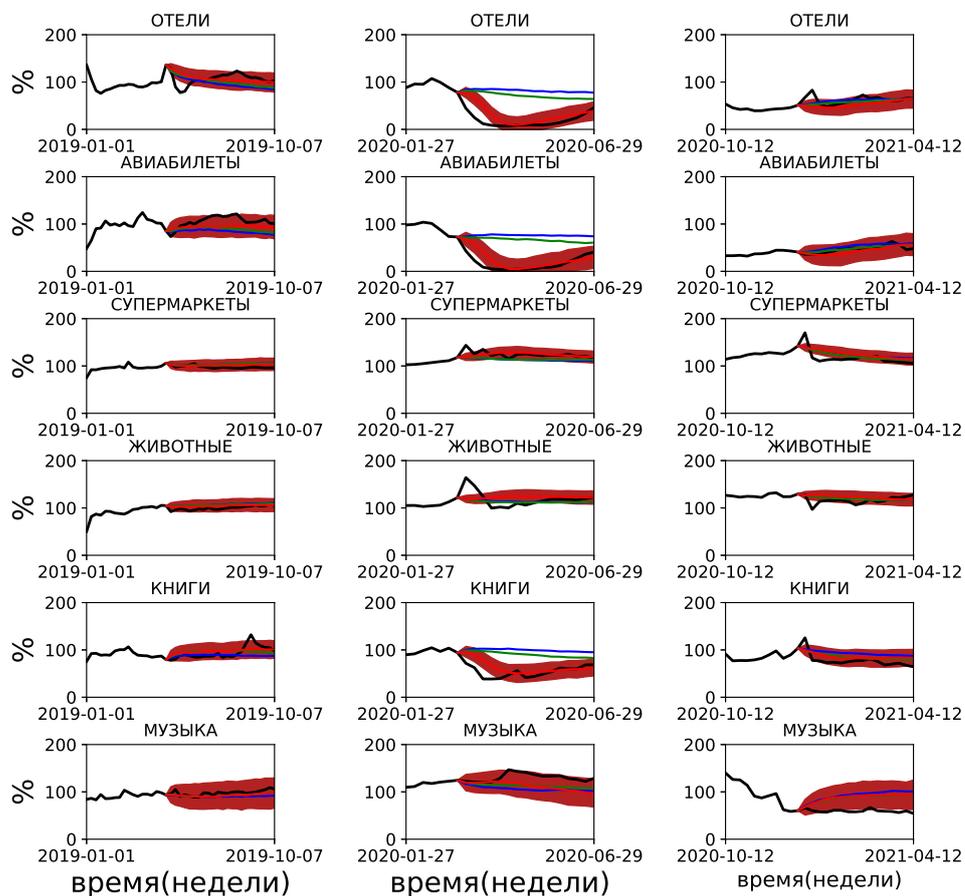


Рис. 1.8: Ретроспективный прогноз некоторых категорий потребительской активности в Москве. Черная линия — данные, красная линия — оптимальная модель с форсингом, цветом закрашен доверительный 95% интервал прогностической оценки. Синяя и зеленая линии — AR(1) модель и оптимальная модель без форсинга, соответственно.

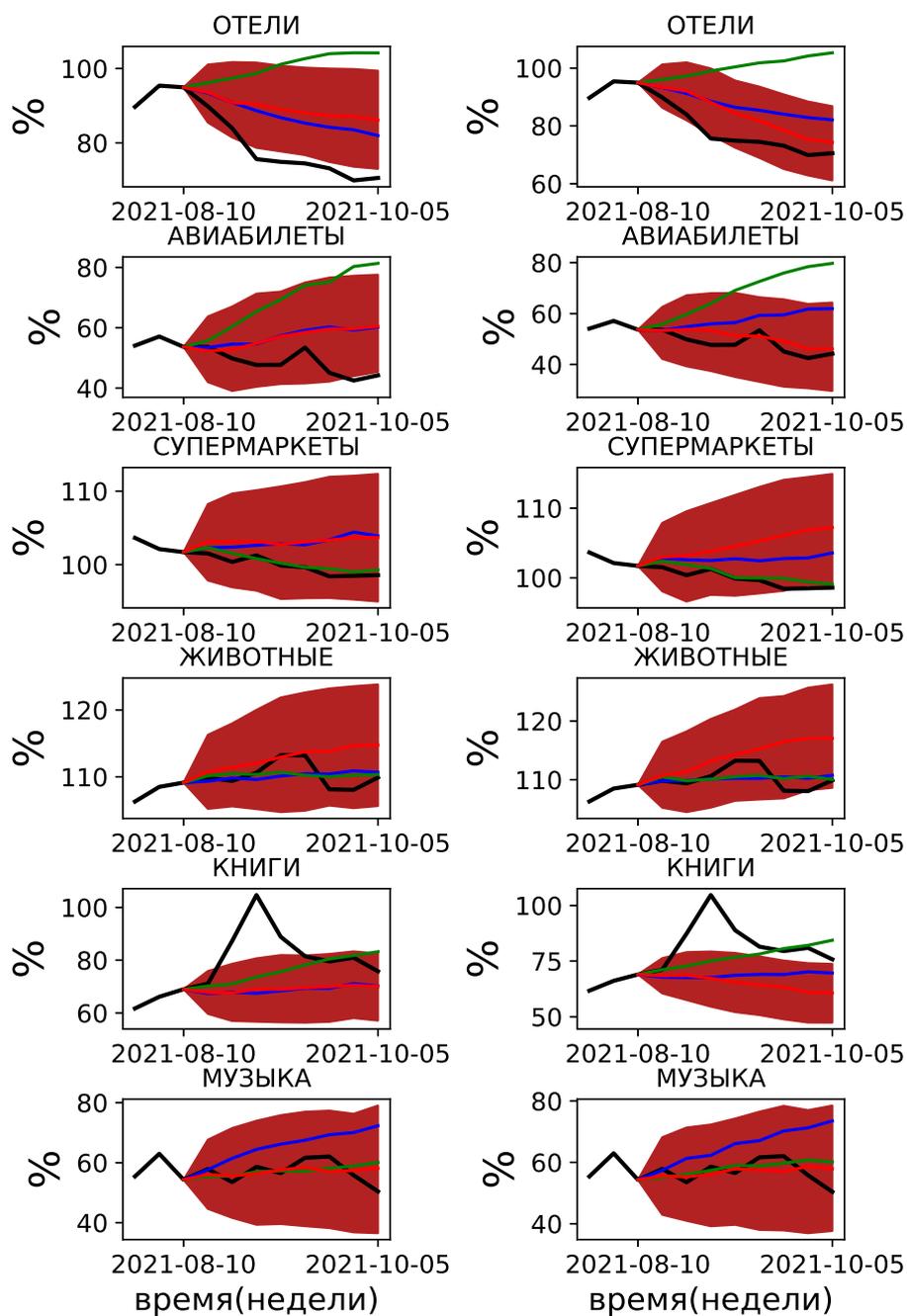


Рис. 1.9: Реальный прогноз некоторых категорий потребительской активности в Москве с 10.08.2021 по 09.10.2021 с помощью различных моделей. Прогноз оптимальной модели с форсингом на левой панели выполнен при фиксированном значении форсинга, соответствующем 10.08.2021, на правой панели – при реально изменяющемся форсинге.

траектория системы посещает две разнесенные области в фазовом пространстве трех ГК, причем момент перехода системы из одной области в другую совпадает с началом пандемии COVID-19 в Российской Федерации. При этом, из двух рассматриваемых эмпирических моделей только модель с форсингом воспроизводит данный переход, т.е. внешнее воздействие в виде ограничений существенно влияет на собственную динамику модели, «сдвигая» ее фазовую траекторию в правильном направлении. На рисунке 1.8 показаны примеры ретроспективных прогнозов (т.е. прогнозов на том же интервале, по которому обучалась модель) в некоторых категориях потребительской активности с трех различных моментов времени для города Москвы. Линиями разного цвета показаны различные модели (красная линия – оптимальная модель (1.1)-(1.13) с учетом форсинга), цветами закрашены 95% доверительные интервалы прогностической оценки оптимальной модели с форсингом. Центральная панель захватывает интервал времени, включающий март 2020 года, на котором произошел резкий рост числа заболевших COVID-19 в Российской Федерации, сопровождаемый введением властями строгих ограничений, направленных на сдерживание распространения эпидемии. Видно, что некоторые категории активности в этот период претерпевали существенный спад, который верно прогнозируется только оптимальной моделью (1.1)-(1.13), учитывающей ограничения. Наконец на рисунке 1.9 показаны примеры реальных (за пределами интервала обучения модели) прогнозов эволюции потребительской активности с 10.08.2021 по 09.10.2021 в Москве. Как видно из рисунков, в целом, оптимальная эмпирическая модель (1.1)-(1.13) с учетом форсинга корректно прогнозирует имеющиеся в данных тенденции. Для остальных городов были получены качественно аналогичные результаты.

1.5 Заключение

В главе был рассмотрен подход к построению моделей оператора эволюции динамических систем (ДС) по многомерным данным наблюдений. Предложенная в рамках данного подхода эмпирическая модель представляет собой дискретную стохастическую ДС, оператор эволюции которой задается в виде суммы детерминированной функции и гауссова случайного процесса. При этом, детерминированная часть модели направлена на реконструкцию динамики системы на тех временных масштабах, которые наиболее хорошо представлены в наблюдаемых данных, в то время как ее стохастическая часть описывает влияние мелкомасштабных процессов с малым временем автокорреляции. Такая форма модели позволяет смягчить ограничения, свойственные полностью детерминированным моделям, которые связаны с фундаментальной невозможностью восстановления полного однозначного фазового пространства вы-

сокоразмерной ДС по конечному временному ряду данных наблюдений. Важной особенностью предложенного подхода к реконструкции ДС является оптимизация сложности эмпирической модели на основе специальной ценовой функции, получаемой в рамках вероятностного байесова подхода. При таком подходе удастся определить как оптимальные гиперпараметры, задающие сложность выбранной параметризации модели, так и значения ее параметров в зависимости от длины имеющихся данных. В качестве инструмента для построения фазовых переменных эмпирической модели предложен метод главных компонент, идея которого состоит в эффективной аппроксимации многомерных данных линейным многообразием меньшей размерности.

Возможности подхода продемонстрированы на примере построения эмпирической модели потребительской активности населения Российской Федерации в условиях пандемии COVID-19. Данный пример отражает наиболее интересную с точки зрения эмпирического моделирования ситуацию, когда априорные знания об исследуемой системе, а соответственно и модели этой системы из первых принципов, фактически, отсутствуют. В процессе реконструкции было установлено, что в основе динамики системы лежит более сложная, чем тривиальный красношумовой процесс, динамическая модель. Показано, что учет в эмпирической динамической модели информации об интенсивности ограничительных мер, вводимых властями, является обоснованным с точки зрения выбранной ценовой функции – байесова критерия оптимальности и статистически значимо улучшает ее предсказательные способности. При этом наиболее восприимчивыми к влиянию ограничений оказываются категории активности, связанные со сферой передвижений. Построенная прогностическая модель, потенциально, может быть использована в качестве дополнительного блока в моделях распространения эпидемий, которые учитывают информацию о мобильности населения (например, модель ВНИИТФ [119]) и могут использоваться в интересах властей в качестве инструмента для выработки рекомендаций по введению ограничений (при имеющихся предположениях о том, какие ограничения будут действовать в период предсказания потребительской активности) для достижения целевых эпидемических показателей.

В следующих главах развитые здесь методы будут применены для построения эмпирических моделей явления Эль-Ниньо – одной из ключевых мод собственной изменчивости климата Земли.

Глава 2

Преодоление весеннего барьера при внутригодовом прогнозе изменчивости явления Эль-Ниньо с помощью эмпирических моделей

2.1 Введение

В этой главе диссертации разработанные методы эмпирического моделирования применяются для построения прогностической модели явления Эль-Ниньо–Южное Колебание (ЭНЮК) на внутригодовых масштабах. При этом наибольший интерес представляет прогноз температуры поверхности тропической части Тихого океана (ТПО), по которой определяются наиболее известные индексы ЭНЮК, характеризующие интенсивность, локализацию и разнообразие типов событий Эль-Ниньо/Ла-Нинья [120, 121]. Решение задачи эффективного внутригодового прогноза ЭНЮК существенно осложняется вследствие существования так называемого весеннего барьера предсказуемости климатической системы в тропиках [83–86], который проявляется в данных ТПО в виде резкого снижения автокорреляций в конце весны – начале лета. В главе демонстрируется, что обоснованный учет в эмпирической модели дополнительных (помимо ТПО) атмосферных и океанических предикторов, которые находятся исходя из современных представлений о физике явления Эль-Ниньо, позволяет частично преодолеть проблему весеннего барьера и существенно улучшить предсказательные способности модели.

Представленные в данной главе результаты опубликованы в статьях [А4, А6] и тезисах [А31].

2.2 Атмосферный предиктор явления Эль-Ниньо

В работах [122, 123] был описан механизм, согласно которому внутритропические процессы атмосферной циркуляции могут оказывать долгосрочное воздействие на динамику ЭНЮК посредством так называемого «следа ТПО». Такой след, обуславливаемый сложившимся распределением потоков тепла у поверхности океана весной, сохраняется до лета и индуцирует аномалии пассатных ветров – важнейшей составляющей динамики ЭНЮК. В работе [124] было выдвинуто предположение о том, что атмосферные переменные – предикторы ЭНЮК, отражающие данный механизм, должны учитываться при прогнозе ЭНЮК наряду с общепринятыми предикторами, характеризующими состояние океана. Данный раздел главы посвящен поиску атмосферных предикторов ЭНЮК на основе анализа пространственно распределенных данных давления на уровне моря. В результате определяется индекс, отражающий состояние субтропической атмосферной циркуляции в зимне-весенний сезон (февраль-март), который имеет статистически значимые длинные корреляции с циклом ЭНЮК в предстоящий период с июня текущего года по май следующего года. Показывается, что учет найденного предиктора в эмпирической модели ЭНЮК, описывающей эволюцию ТПО, позволяет добиться статистически значимого улучшения ее прогностической способности на масштабах от нескольких месяцев до одного года, в том числе при прогнозе через весенний барьер.

2.2.1 Совместный корреляционный анализ данных ТПО и атмосферных данных

Описание используемых данных и их предобработка

Для поиска атмосферного предиктора явления Эль-Ниньо исследовался многомерный временной ряд давления на уровне моря (ДУМ) из базы данных реанализа NCEP/NCAR [125] с шагом по времени 1 месяц, представленный на пространственной сетке, покрывающей Земной шар с шагом $1^\circ \times 1^\circ$ (144 точки по широте и 73 – по долготе). В качестве характеристики динамики явления Эль-Ниньо использовался традиционный индекс Nino3.4¹, рассчитанный по многомерному временному ряду ТПО из базы данных Extended Reconstructed SST (ERSST dataset version 5, пространственная сетка $2^\circ \times 2^\circ$, шаг по времени 1 месяц) [126]. Данные рассматривались на временном интервале с января 1960 года по декабрь 2019 года. В ходе предварительной обработки исходных временных рядов ДУМ и ТПО были получены временные

¹Данный индекс определяют как аномалии ТПО, усредненные в центральном регионе экваториального Тихого океана 5° с.ш. - 5° ю.ш., 120° з.д. - 170° з.д. [120].

ряды их аномалий – т.е. отклонений от среднегодового значения. Процедура получения аномалий состоит в вычитании из временных рядов данных в каждой точке пространственной сетки их усредненных (своих для каждого месяца года) значений на рассматриваемом интервале времени. Кроме того, из полученных временных рядов аномалий ДУМ в каждой пространственной точке вычитался линейный тренд, а из временных рядов аномалий ТПО – линейная регрессия на сигнал, характеризующий концентрацию CO_2 в атмосфере².

Корреляционный анализ

Нами были проанализированы корреляции между взятыми с шагом 1 год (в разные месяцы года) временным рядом индекса Nino3.4 и временными рядами аномалий ДУМ в каждой точке пространственной сетки. Наибольший интерес при этом представляют лаговые корреляции между временными рядами ДУМ в зимние и весенние месяцы года (перед весенним барьером) и временным рядом индекса Nino3.4 в последующие месяцы (после весеннего барьера). Карта таких статистически значимых корреляций изображена на рисунке 2.1. Для оценки значимости нами использовался статистический тест на основе генерации суррогатных данных, идея которого была описана в разделе 1.3.3 первой главы диссертации. Были рассмотрены две модификации данного теста, отличающихся способом построения статистического ансамбля значений корреляций. В первой из них для временных рядов аномалий ДУМ в каждой точке пространственной сетки производился набор из 10000 суррогатных временных рядов с использованием модели AR(1) (1.13). Затем, по полученному в каждой пространственной точке ансамблю суррогатов рассчитывался ансамбль значений их корреляций с индексом Nino3.4, с которыми сравнивались значения корреляций, рассчитанные по реальным данным. Нулевая гипотеза, которая отвергается в данной модификации теста, предполагает, что значения корреляций между временным рядом индекса Nino3.4 и временным рядом аномалий ДУМ *в данной* пространственной точке не превышают (по абсолютной величине) значения корреляций этого индекса с суррогатными временными рядами аномалий ДУМ *в этой же* пространственной точке. Значимые корреляции (на уровне значимости 0.1), соответствующие этой модификации теста, показаны цветом на рисунке 2.1.

Другая модификация теста учитывает пространственные свойства данных ДУМ. Соответствующий суррогатный ансамбль получался следующим образом. Сначала выполнялось ЭОФ разложение (см. раздел 1.4.1 первой главы диссертации) данных

²Временные ряды концентрации CO_2 в атмосфере, предоставляемые Лабораторией глобального мониторинга NOAA, доступны по адресу <https://gml.noaa.gov/aftp/products/trends/co2/>.

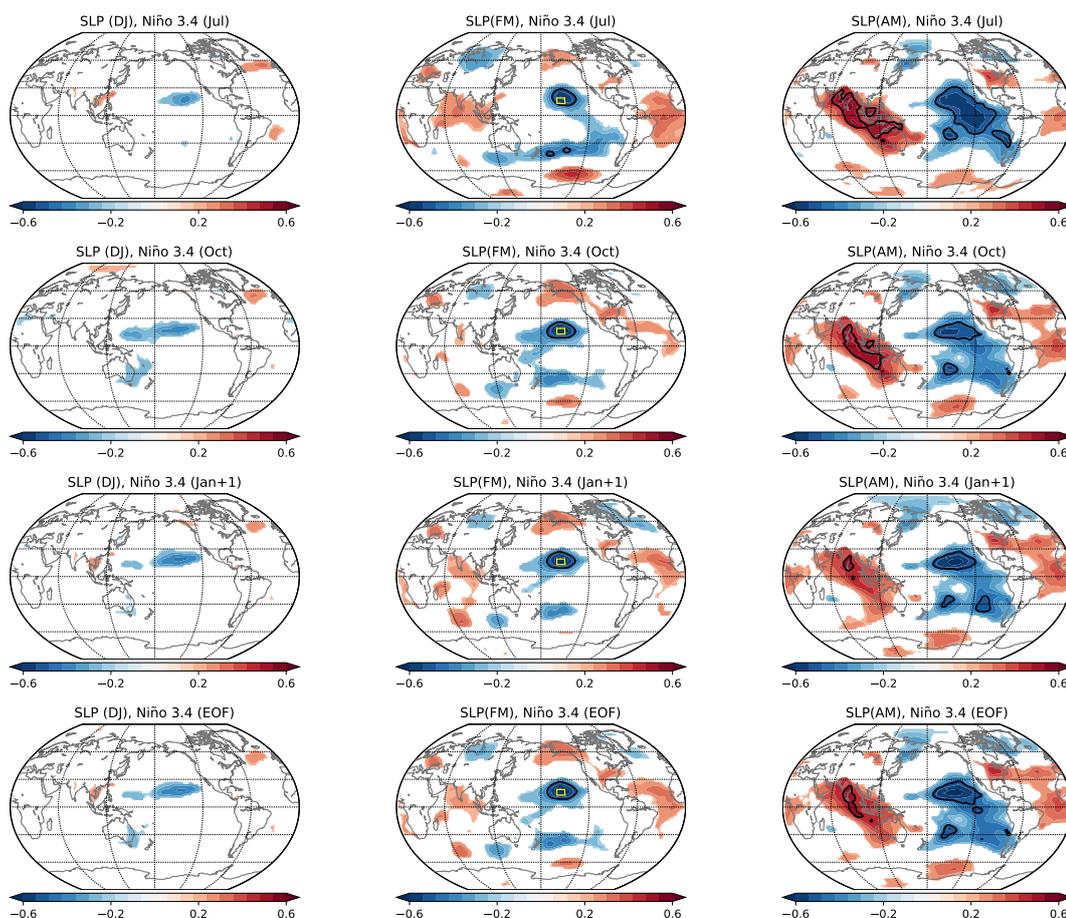


Рис. 2.1: Корреляции сезонных средних аномалий ДУМ за декабрь-январь (DJ), февраль-март (FM) и апрель-май (AM) с индексом Niño3.4 в последующие месяцы. На трех верхних панелях показаны корреляции с индексом Niño3.4 в июле, октябре и январе. На нижней панели показаны корреляции со всем годовым циклом ЭНЮК, который следует после сезонов, в которых рассчитываются значения ДУМ (см. основной текст). Цветом показаны только статистически значимые корреляции (уровень значимости 0.1) в соответствии с суррогатным тестом, примененным к абсолютным значениям корреляций независимо в каждой точке пространственной сетки. Черные контуры ограничивают области значимости (на уровне 0.1), соответствующие более сильному тесту, основанному на пространственно распределенных суррогатах аномалий ДУМ (см. основной текст). Желтыми прямоугольниками отмечена область, по которой рассчитывается индекс HI.

ДУМ, в результате которого они представлялись в виде (1.12). Затем с помощью модели AR(1) генерировался ансамбль из 10000 суррогатов для каждой из полученных в ходе разложения главных компонент (ГК). Суррогатные ГК пересчитывались в физическое пространство данных согласно (1.12) при неизменном базисе ЭОФ. Получаемые таким образом суррогаты отражают информацию не только о временных (как в случае первой модификации теста), но и пространственных свойствах данных ДУМ. Далее рассчитывались корреляции между индексом Niño3.4 и полученными многомерными суррогатными временными рядами ДУМ. При этом, в отличие от первого случая, учитывались только максимальные (по всем точкам пространства) абсолютные значения корреляций, из которых и формировался статистический ансамбль из 10000 значений суррогатных корреляций, с которыми сравнивались значения корреляций, рассчитанные по реальным данным. Таким образом, нулевая гипотеза, которая отвергается в данной модификации теста, является более сильной: она предполагает, что значения корреляций между временным рядом индекса Niño3.4 и временным рядом аномалий ДУМ в *данной* точке пространства не превышают значения корреляций этого индекса с суррогатными временными рядами аномалий ДУМ (сохраняющими пространственно-временные свойства исходных данных) *хотя бы в какой-то* точке пространства. Другими словами, опровергается возможность случайного возникновения в произвольной точке Земного шара полученного значения корреляции. Соответствующие области значимых на уровне 0.1 корреляций ограничены черными контурами на рисунке 2.1. Как видно из рисунка 2.1, временные ряды аномалий ДУМ в центральной части Тихого океана в зимне-весенний сезон (усредненные за февраль-март и апрель-май) имеют максимальные лаговые корреляции с индексом Niño3.4 в последующие месяцы года. При этом в регионе Гавайских островов наблюдается область ДУМ (наиболее выделенная в феврале-марте), которая устойчиво коррелирует с индексом Niño3.4 на протяжении всех последующих месяцев с июня по май следующего года.

Для того, чтобы обобщить карты корреляций индекса Niño3.4 в разные месяцы, рассмотрим также корреляции значений ДУМ в зимне-весенний период со всем предстоящим циклом ЭНЮК. Согласно работе [127], сигнал ЭНЮК в индексе Niño3.4 хорошо описывается первой (старшей) ЭОФ, рассчитанной по $D = 12$ мерному набору данных, составленному из неперекрывающихся последовательных 12-месячных фрагментов ряда индекса, начинающихся в июне³. Соответствующая главная компонента представляет собой временной ряд с шагом один год, каждая точка которого является проекцией (см. формулу (1.10) в разделе 1.4.1 первой главы диссертации)

³Данный аспект будет обсуждаться более подробно в следующей главе диссертации.

12-месячного фрагмента ряда индекса Nino3.4 с июня по май на этот ЭОФ и, фактически, характеризует амплитуду годового цикла ЭНЮК. Для проанализированного нами набора данных старшая ЭОФ описывает около 88% вариации индекса Nino3.4. Как видно из нижней панели рисунка 2.1, существует выделенный регион в районе Гавайев, в котором достигаются наиболее значимые корреляции между усредненными за февраль-март аномалиями ДУМ и предстоящим циклом ЭНЮК. На основании этого мы определяем потенциальный атмосферный предиктор явления Эль-Ниньо – Гавайский индекс (Hawaiian index – HI) как среднее за февраль-март значение аномалий ДУМ в регионе 13° с.ш. - 19° с.ш., 150° з.д. - 160° з.д. Как видно из рисунка 2.2 (а), определяемый таким образом годовой ряд HI значимо коррелирует с индексом Nino3.4 во все месяцы цикла ЭНЮК. При этом временной ряд амплитуд годового цикла ЭНЮК (ГК старшей ЭОФ) вносит основной вклад в эту корреляцию – абсолютное значение коэффициента корреляции между ним и временным рядом HI составляет 0.66. Из рисунка 2.2 (б) видно, что высокие значения корреляции обеспечиваются, главным образом, за счет умеренных и сильных событий Эль-Ниньо/Ла-Нинья.

Потенциальную пользу от найденного атмосферного предиктора ЭНЮК можно оценить на основе простой авторегрессионной модели индекса Nino3.4:

$$x_{ni} = a_1^i x_{ni}^{-1} + a_2^i x_{ni}^{-2} + \dots + a_l^i x_{ni}^{-l} + b^i h_n + \xi_{ni}. \quad (2.1)$$

Здесь x_{ni} – значение индекса Nino3.4 в i -й месяц n -го цикла ЭНЮК, x_{ni}^{-j} – значение этого индекса за j месяцев до x_{ni} , h_n значение индекса HI перед началом n -го цикла ЭНЮК, l – временной лаг, определяющий порядок авторегрессии. В таких обозначениях индекс i принимает значения от 1 до 12, причем значение $i = 1$ соответствует июню – начальному месяцу цикла ЭНЮК. В рамках данной модели значение индекса Nino3.4 в определенный месяц предсказывается на основе значений этого индекса в l предыдущих месяцев, а также значения индекса HI, вычисляемого один раз в год путем усреднения аномалий ДУМ в феврале–марте и остающегося постоянным в течение последующего цикла ЭНЮК с июня по май. Связанная с наличием весеннего барьера потеря автокорреляций отчетливо проявляется в виде ярко выраженных июньских пиков среднеквадратичной ошибки прогноза индекса Nino3.4 на один месяц ($\langle \xi_{ni}^2 \rangle_n$ в случае моделей без учета форсинга в виде сигнала HI ($b^i = 0$) (рисунок 2.2 (с))). Увеличение лага модели свыше 2 месяцев фактически не уменьшает ошибку прогноза для всех месяцев цикла ЭНЮК, однако учет форсинга HI в модели с лагом 2 приводит к существенному уменьшению величины пика среднеквадратичной ошибки прогноза в июне, вызванного весенним барьером. Таким образом, весенний барьер при прогнозе индекса Nino3.4 существенно понижается, когда прогностическая модель строится не только на основе предыдущих значений этого индекса, но и учитывает

также атмосферный форсинг в виде кусочно-постоянного сигнала NI , который определяет правильное смещение выхода модели, обеспечивая более корректный прогноз событий Эль-Ниньо/Ла-Нинья.

Атмосферный предиктор как часть крупномасштабного паттерна ДУМ в средних северных широтах тихоокеанского региона

Нами была предпринята попытка установить связь найденного предиктора ЭНЮК с известными атмосферными процессами в средних широтах. Для этого были проанализированы композитные пространственные распределения ДУМ, которые получаются путем усреднения значений аномалий ДУМ в феврале-марте по тем годам, в которых начиналось развитие достаточно сильных событий Эль-Ниньо/Ла-Нинья (по классификации <https://ggweather.com/enso/oni.htm>, отмечены звездочками на рисунке 2.2 б). Рассматриваемый нами интервал времени включает в себя 13 таких Эль-Ниньо, начавшихся в 1963, 1965, 1968, 1972, 1982, 1986, 1987, 1991, 1994, 1997, 2002, 2009, 2015 годах, и 10 Ла-Нинья, начавшихся в 1970, 1973, 1975, 1988, 1995, 1998, 1999, 2007, 2010, 2011 годах. По двум этим подвыборкам лет и определялись композитные распределения ДУМ, соответствующие Эль-Ниньо и Ла-Нинья, которые изображены на рисунке 2.3. Статистическая значимость полученных пространственных структур ДУМ устанавливалась путем опровержения нулевой гипотезы о том, что такие же структуры могут быть получены путем усреднения подвыборок аномалий ДУМ такого же размера (суррогатные ансамбли состоят из 10 000 подвыборок длины 13 для Эль-Ниньо, и длины 10 для Ла-Нинья) в случайно выбранные (т.е. не обязательно относящиеся к годам сильных событий Эль-Ниньо/Ла-Нинья) годы на рассматриваемом интервале времени. Контуры на рисунке 2.3 ограничивают области, где абсолютные значения полученных композитов превышают соответствующие значения 0.95 (уровень значимости 0.05) и 0.65 (уровень значимости 0.35) квантилей распределений, полученных путем усреднения аномалий ДУМ по суррогатным подвыборкам. Дипольная структура композитных распределений ДУМ в годы сильных событий Эль-Ниньо и Ла-Нинья очень близка к структуре так называемой Тихоокеанско-североамериканской дальней связи (North Pacific Oscillation – NPO) в ее отрицательной и положительной фазе, соответственно [128]. Это согласуется с выводами работы [122], где было показано, что зимне-весенний паттерн NPO способствует возникновению специфического распределения аномалий ТПО в субтропиках (след ТПО), которое сохраняется до лета и может влиять на изменчивость ЭНЮК, вызывая зональные аномалии ветров вдоль экватора. Как следует из рисунка 2.3, найденная NPO-подобная структура, по-видимому, не симметрична по отношению

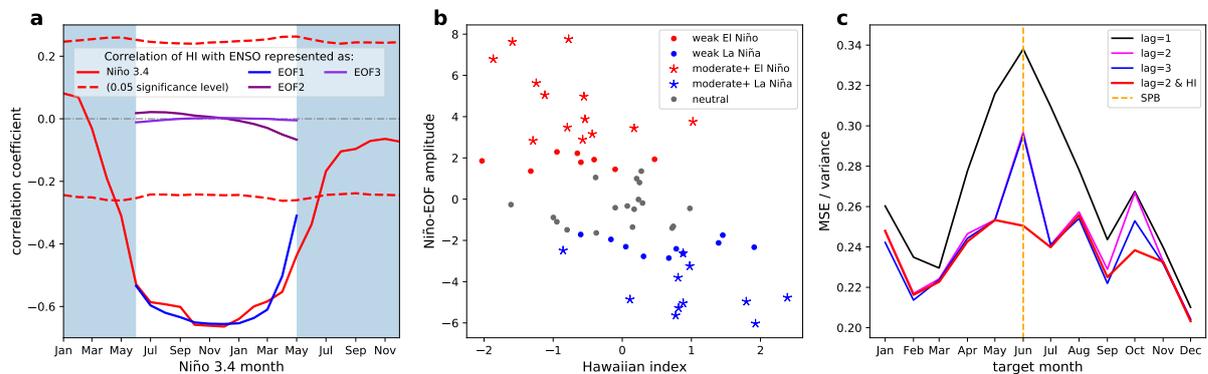


Рис. 2.2: (а) Корреляции индекса HI с индексом Niño3.4 в различные месяцы года (сплошная красная кривая). Красные пунктирные линии ограничивают области значимых на уровне 0.05 значений корреляции, рассчитанные на основе AR(1)-суррогатного теста. Предстоящий по отношению к индексу HI цикл ЭНЮК показан на белом фоне. Синими и фиолетовыми кривыми (см. легенду) показаны корреляции индекса HI с тремя ведущими ГК, соответствующих ЭОФ (первая из которых определяет цикл ЭНЮК, см. текст), рассчитанным по временному ряду, составленному из неперекрывающихся последовательных 12-месячных фрагментов ряда индекса Niño3.4, начинающихся в июне. Синими и фиолетовыми кривыми (см. легенду) показаны вклады этих ЭОФ в общую корреляцию. (б) Амплитуды циклов ЭНЮК в зависимости от индекса HI. Амплитуды циклов, соответствующие сильным (умеренным), слабым и нейтральным событиям Эль-Ниньо/Ла-Нинья (по классификации <https://ggweather.com/enso/oni.htm>), отмечены звездочками, цветными кружками и черными кружками соответственно. Красным цветом обозначена фаза Эль-Ниньо, синим – Ла-Нинья, черным – нейтральная фаза. (с) Среднеквадратические ошибки (СКО) прогноза индекса Niño3.4 на один месяц с помощью линейной авторегрессионной модели (2.1) в зависимости от месяца года. СКО модели без учета форсинга HI ($b^i = 0$) порядка $l = 1, 2, 3$ показаны черным, розовым и синим цветами соответственно. СКО модели с учетом форсинга HI и $l = 2$ показаны красным цветом.

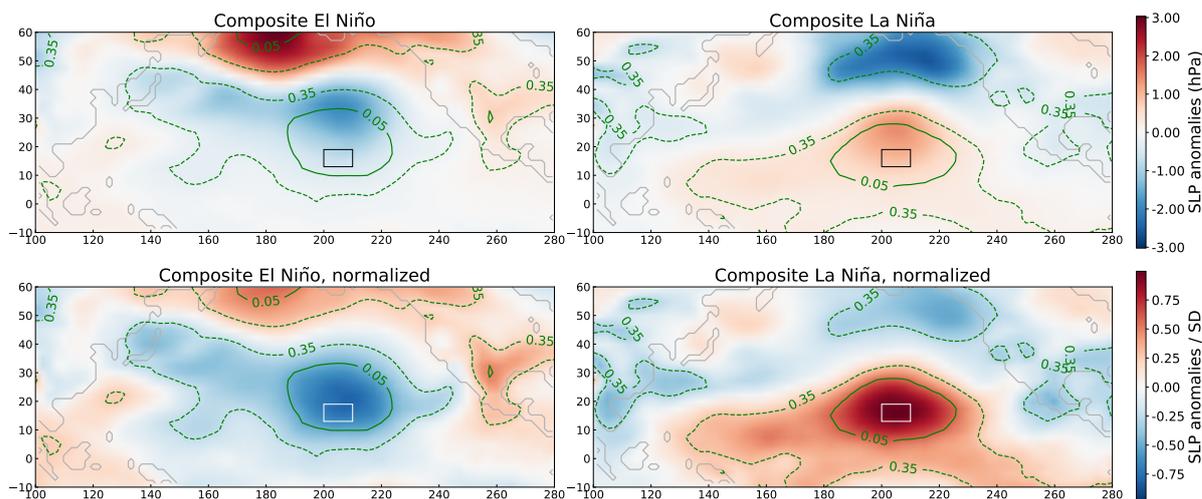


Рис. 2.3: Композитные пространственные распределения аномалий ДУМ, предшествующие событиям Эль-Ниньо (левый столбец) и Ла-Нинья (правый столбец). На верхней панели показаны абсолютные значения аномалий ДУМ, на нижней панели – значения, нормированные на среднеквадратичное отклонение аномалий ДУМ в каждой точке пространственной сетки. Зеленые контуры ограничивают области статистической значимости на уровнях 0.05 и 0.35. Прямоугольники ограничивают область, по которой рассчитывается индекс НИ.

к фазе ЭНЮК: композит, соответствующий фазе Эль-Ниньо не является простой инверсией композита, соответствующего фазе Ла-Нинья и наоборот. В частности, в фазе Ла-Нинья северная часть диполя NPO сильнее смещена на восток, а его южная часть – в тропические широты. При этом наиболее значимая область каждого из композитов содержит в себе область, где ранее был определен индекс НИ. На основании этого можно сделать вывод о том, что найденный индекс НИ может характеризовать линейную составляющую взаимодействия между зимне-весенним паттерном NPO и циклом ЭНЮК посредством механизма следа ТПО, описанного в [122].

2.2.2 Построение эмпирической модели динамики ТПО и анализ ее предсказательной способности

В данном разделе главы будут построены оптимальные стохастические модели индекса Niño3.4 вида (1.1) с учетом и без учета форсинга НИ, а также исследованы предсказательные способности этих моделей. Детерминированная часть модели задавалась в виде линейной функции (1.2), коэффициенты которой являлись периодическими функциями с периодом $T = 12$ месяцев и представлялись в виде дискретного

ряда Фурье:

$$\begin{aligned}\mathbf{A}_n &= \mathbf{A}^0 + \sum_{k=1}^q [\mathbf{A}_c^k \cos \frac{2\pi k}{12}n + \mathbf{A}_s^k \sin \frac{2\pi k}{12}n], \\ \mathbf{B}_n &= \mathbf{B}^0 + \sum_{k=1}^q [\mathbf{B}_c^k \cos \frac{2\pi k}{12}n + \mathbf{B}_s^k \sin \frac{2\pi k}{12}n].\end{aligned}\quad (2.2)$$

Здесь q принимает целые значения от 1 до 6 ($\mathbf{A}_s^6, \mathbf{B}_s^6 = \mathbf{0}$ по определению; случай $q = 0$ соответствует обычной линейной модели с постоянными коэффициентами $\mathbf{A}_n = \mathbf{A}^0, \mathbf{B}_n = \mathbf{B}^0$) и определяет возможную зависимость модели от различных гармоник сезонного хода. Явный учет зависимости коэффициентов модели от времени года обусловлен тем обстоятельством, что простое вычитание сезонного хода при построении аномалий ТПО, описанное в предыдущем разделе, вообще говоря, не удаляет его полностью, поскольку его вклад в динамику ЭНЮК может быть не аддитивным и приводить к нетривиальным эффектам⁴. Эмпирические модели (1.1) с периодическими коэффициентами (2.2) с учетом и без учета форсинга НИ обучались по скалярному ряду индекса Nino3.4 длины $N = 720$ месяцев (январь 1960 года – декабрь 2019 года) путем оптимизации ценовой функции (1.9). Дисперсии априорных распределений на коэффициенты моделей задавались как $\sigma_{\mathbf{A}}^2 = \frac{1}{(q+1) \cdot l}$ (без учета форсинга) и $\sigma_{\mathbf{A}, \mathbf{B}}^2 = \frac{1}{(q+1) \cdot (l+1)}$ (с учетом форсинга) для того, чтобы ограничить выход моделей в рамках диапазона изменения индекса Nino3.4 (см. подробнее раздел 1.3 первой главы диссертации). В процессе обучения были найдены оптимальные значения гиперпараметров – $l = 2$ и $q = 1$ – которые оказались одинаковыми для обеих моделей⁵. При этом учет в модели форсинга НИ оказался статистически обоснованным с точки зрения критерия (1.9).

Сравнивались прогностические способности следующих эмпирических моделей: (i) AR(2) – модель с глубиной памяти $l = 2$, но без учета сезонного хода ($q = 0$, постоянные коэффициенты) и форсинга, (ii) AR(2) periodic – модель с $l = 2$ и $q = 1$, но без учета форсинга, (iii) AR(2) periodic + NI forcing – оптимальная модель с $l = 2$ и $q = 1$ с учетом форсинга. Как и в разделе 1.4.2 первой главы диссертации, в качестве мер прогностической способности использовались квадратный корень из среднеквадратичной ошибки прогноза (КСКО) (1.14) и коэффициент корреляции Пирсона (1.15), которые показаны для каждого прогнозируемого месяца года в зависимости от заблаговременности его прогнозирования на рисунках 2.4 (а) для трех рассматриваемых моделей.

Отметим, что заблаговременность прогноза (дальность прогнозирования) данного месяца года i с помощью модели, в которую входит форсинг НИ, ограничена, поскольку

⁴В частности, в следующей главе диссертации будет подробно обсуждаться эффект, связанный с наличием сезонных циклов в аномалиях характеристик ЭНЮК.

⁵Как видно из (2.2), в случае $q = 1$ коэффициенты модели являются гармоническими функциями с периодом 12 месяцев.

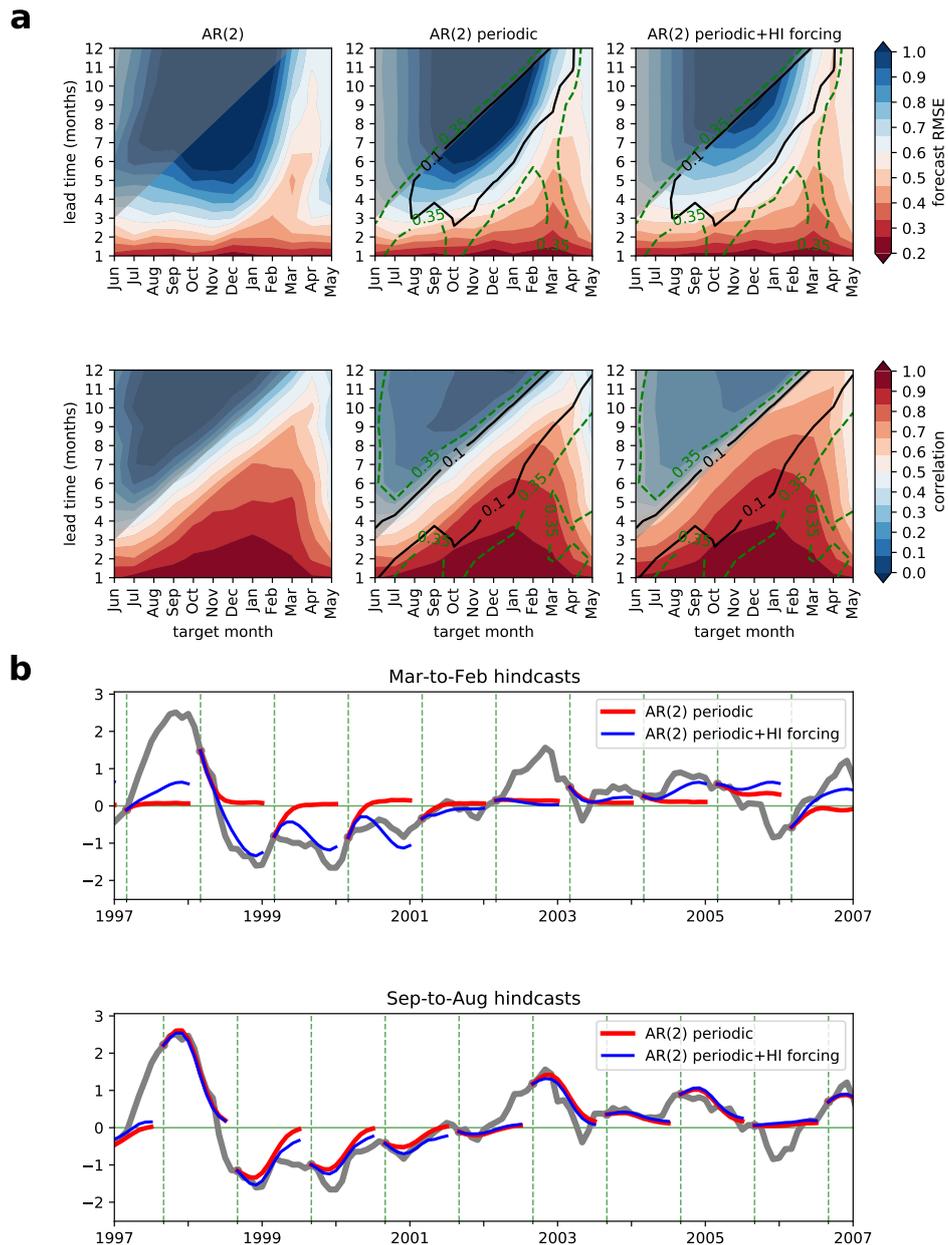


Рис. 2.4: (а) Предсказательная способность различных моделей индекса Niño3.4 по мерам КСКО (нормированной на среднеквадратичное отклонение индекса Niño3.4, верхняя панель) и коэффициента корреляции (нижняя панель) в плоскостях прогнозируемый месяц – дальность прогнозирования. Слева направо: модель без учета сезонной зависимости коэффициентов и форсинга HI, модель с сезонной зависимостью коэффициентов, но без учета форсинга HI, оптимальная модель с учетом обоих факторов. Линиями нанесены уровни значимого улучшения прогноза оптимальной модели, по отношению к модели, учитывающей только сезонную зависимость коэффициентов. (б) Примеры ретроспективных прогнозов (медианы большого ансамбля прогнозных реализаций) моделей с форсингом и без для двух разных сезонов.

значение форсинга, которое используется при прогнозе, рассчитывается один раз в год в феврале-марте. Так, если прогноз начинается в январе текущего года, известное на этот момент значение форсинга NI (рассчитанное в феврале-марте предыдущего года) можно использовать лишь при прогнозах месяцев до ближайшего мая; далее необходимо использовать новое значение NI , которое еще неизвестно. Области (в плоскостях прогнозируемый месяц-заблаговременность прогноза), в которых невозможен прогноз моделью, учитывающей форсинг, затемнены на рисунках 2.4 (а). Наиболее очевидным вариантом прогнозирования в этих областях является использование оптимальной модели без форсинга (т.е. модели $AR(2)$ periodic) для продолжения прогноза модели с форсингом.

Как видно на из рисунков 2.4 (а), предсказательная способность модели улучшается при учете как сезонной зависимости ее коэффициентов, так и форсинга NI . В частности, модель $AR(2)$ periodic (с учетом только сезонного хода) демонстрирует заметно лучшую прогностическую способность на дальности прогноза до 6–7 месяцев, по сравнению с моделью $AR(2)$ с постоянными коэффициентами. Добавление форсинга NI в модель позволяет существенно улучшить дальние прогнозы на времена предсказания более 4 месяцев (там, где этот форсинг доступен). При этом наиболее значимые улучшения предсказательной способности по мере КСКО (верхние панели рисунка 2.4 (а)) происходят для целевых месяцев с августа по март, а по коэффициенту корреляции – для всех месяцев цикла ЭНЮК с июня по май (нижние панели рисунка 2.4 (а)). Это согласуется с результатами, полученными в предыдущем разделе, согласно которым аномалии ДУМ в феврале-марте в регионе NI содержат полезную информацию обо всем предстоящем цикле ЭНЮК. Области значимого улучшения прогноза с помощью оптимальной модели (в которой учтен и форсинг и сезонная зависимость коэффициентов) оценивались аналогично тому, как это делалось в разделе 1.4.2 первой главы диссертации при анализе прогностической способности эмпирической модели потребительской активности общества. На рисунке 2.4 (б) показаны примеры ретроспективных прогнозов эмпирических моделей с форсингом и без для разных сезонов. Видно, что в том случае, если прогноз начинается накануне наступления весеннего барьера (весной), модель без учета форсинга имеет тенденцию к неинформативному предсказанию нейтральных (нулевых) значений индекса $Nino3.4$ после весеннего барьера, в то время как модель с учетом форсинга правильно захватывает тенденции к развитию положительных (Эль-Ниньо) либо отрицательных (Ла-Нинья) значений индекса. Напротив, прогнозы, которые начинаются далеко от весеннего барьера (осенью) фактически одинаковы для обеих моделей.

2.3 Предиктор явления Эль-Ниньо на основе данных теплосодержания верхнего слоя тропического Тихого океана

В рамках современных концепций изменчивость теплосодержания верхнего слоя тропического Тихого океана рассматривается как один из ключевых элементов динамики ЭНЮК [129]. Тот факт, что величина аномалий ТПО в ходе событий Эль-Ниньо/Ла-Нинья напрямую связана с величиной зонально усредненных аномалий теплосодержания, убедительно подтверждается эмпирически [88]. В этой связи многие исследования направлены на поиск эффективных предикторов ЭНЮК по данным, характеризующим теплосодержание верхнего слоя тропического Тихого океана. Часто такие предикторы определяются путем обнаружения существенных лаговых корреляций между временными рядами индексов ЭНЮК на основе ТПО, которые необходимо прогнозировать, и индексами, отражающими изменчивость теплосодержания. При этом, найденные предикторы, как правило, подаются в прогностическую модель ТПО как некоторые *заданные* сигналы (например, как факторы регрессии [63, 87, 89]), а не как динамические переменные, что затрудняет использование таких моделей для «настоящего» прогноза за пределы интервала, на котором доступны данные наблюдений, поскольку требует экстраполяции предикторов в будущее⁶. В данном разделе проводится анализ пространственно-распределенных данных ТПО и теплосодержания верхнего слоя тропического Тихого океана на предмет выявления эффективных переменных для эмпирического прогноза изменчивости колебания Эль-Ниньо. По данным ТПО выявляется индекс, который описывает связанную с ЭНЮК изменчивость в тропической части Тихого океана и имеет высокую корреляцию с традиционным индексом Niño3.4. По данным теплосодержания определяется сигнал, который может характеризовать динамику подповерхностных теплых вод в центральной и западной части Тихого океана, которая, является важнейшей составляющей цикла колебания Эль-Ниньо. Строится эмпирическая модель явления ЭНЮК, которая учитывает найденные индексы как равноценные динамические переменные. Демонстрируется, что такая совместная модель имеет существенно лучшую предсказательную способность, по сравнению с моделью, основанной только на данных ТПО.

⁶Такая проблема возникала в предыдущем разделе главы.

2.3.1 ЭОФ-анализ данных ТПО и теплосодержания верхнего слоя тропического Тихого океана

Описание используемых данных и их предобработка

При построении эмпирической модели ЭНЮК использовались пространственно распределенные временные ряды ТПО из базы данных ERSST dataset version 5 (их подробное описание было дано в предыдущем разделе), а также временные ряды теплосодержания верхнего (0-300 метров) слоя океана (ТО), представленные на пространственной сетке с разрешением $1^\circ \times 1^\circ$ [130]⁷. Временные ряды с шагом 1 месяц исследовались на интервале с января 1960 по декабрь 2020 года (общая длина $N = 732$ месяца) в тропической части Тихого океана 10° с.ш. - 10° ю.ш., 80° з.д. - 120° в.д., где наиболее сильно представлена изменчивость, связанная с ЭНЮК. Как и в предыдущем разделе, по исходным данным были получены их аномалии путем вычитания помесечных средних, рассчитанных на рассматриваемом интервале времени, и линейной регрессии на временной ряд концентрации CO_2 в атмосфере в каждой точке пространственной сетки.

ЭОФ анализ

К полученным данным аномалий ТПО и ТО было применено ЭОФ-разложение (1.12). На рисунке 2.5 показаны пространственные распределения (паттерны), соответствующие двум главным ЭОФ, рассчитанным по данным ТПО (верхняя панель) и ТО (нижняя панель). В обоих случаях они описывают основную часть (более 70 %) общей вариации данных. Пространственный паттерн первой ЭОФ ТПО (рисунок 2.5 (a)) имеет характерную форму «языка», вытянутого с востока на запад в тропической части Тихого океана. Во многих работах [75, 120, 131] отмечалось, что именно первая ЭОФ описывает наиболее существенную часть связанной с ЭНЮК изменчивости ТПО в тропической части Тихого океана, а соответствующая ей главная компонента (ГК) сильно коррелирует с индексом Niño3.4. Вторая ЭОФ ТПО (рисунок 2.5 (b)) описывает существенно меньшую часть вариации данных (доли захваченных вариаций указаны на подписях к рисункам) по сравнению с первой, однако ее учет позволяет описать разнообразие типов ЭНЮК, связанное с различной пространственной локализацией максимумов аномалий ТПО, которые могут возникать во время событий Эль-Ниньо (в частности, разделяют восточно-тихоокеанское или каноническое Эль-Ниньо и центрально-тихоокеанское Эль-Ниньо или Эль-Ниньо Модоки [121]).

⁷Данные, предоставленные китайским Институтом физики атмосферы доступны по ссылке: <https://pan.cstcloud.cn/s/sloceeVQjo>.

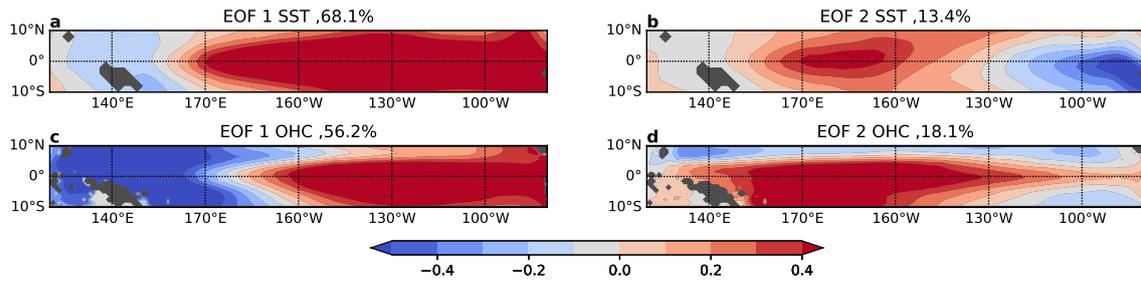


Рис. 2.5: Пространственные распределения, соответствующие двум главным ЭОФ аномалий ТПО (верхняя панель, цветовая шкала в $^{\circ}C$) и ТО (нижняя панель, цветовая шкала в $10^9 J/m^2$). Доли захваченной вариации в процентах показаны сверху для каждой ЭОФ.

Для того, чтобы проинтерпретировать главные ЭОФ, полученные по данным аномалий ТО, мы рассмотрели попарные зависимости соответствующих им ГК от ГК аномалий ТПО (рис. 2.6). Из рисунка 2.6 (b) видно, что ГК первых ЭОФ аномалий ТПО (ГК-1 ТПО) и ТО (ГК-1 ТО) сильно коррелируют (абсолютные значения корреляций указаны на подписях к рисункам), т.е. их зависимость близка к линейной. При этом структуры пространственных паттернов ЭОФ, соответствующих этим ГК и показанные на рисунках 2.5 (a), (c), также очень похожи. Качественно другие зависимости были выявлены между ГК-1 ТПО и ГК второй ЭОФ ТО (ГК-2 ТО) (рисунок 2.6 (c)), а также ГК-1 и ГК-2 ТО (рисунок 2.6 (d)). Эти зависимости демонстрируют характерную циклическую структуру, которая указывает на фазовый сдвиг между соответствующими временными рядами ГК, которые практически не коррелируют. Сильные события Эль-Ниньо/Ла-Нинья проявляются в данных зависимостях в виде круговых орбит (циклов) большого радиуса. В остальных плоскостях на рисунке 2.6 не наблюдается каких-либо характерных зависимостей. В ходе анализа лаговых корреляций было получено, что максимальные абсолютные значения корреляций между ГК-1 ТПО и ГК-2 ТО достигаются при их относительном сдвиге на 5-9 месяцев (рисунок 2.7). Пространственный паттерн второй ЭОФ аномалий ТО (рис. 2.6 (d)) существенно отличается от соответствующего паттерна аномалий ТПО (рис. 2.6 (b)) и описывает изменчивость в центральной и западной части тропического Тихого океана, которая может быть связана с накоплением и разгрузкой подповерхностных теплых вод накануне и во время событий Эль-Ниньо [132–134]. Отметим, что данные результаты о взаимосвязях между аномалиями ТПО и ТО в тропической части Тихого океана согласуются с результатами, полученными в работах [88, 133] при исследовании

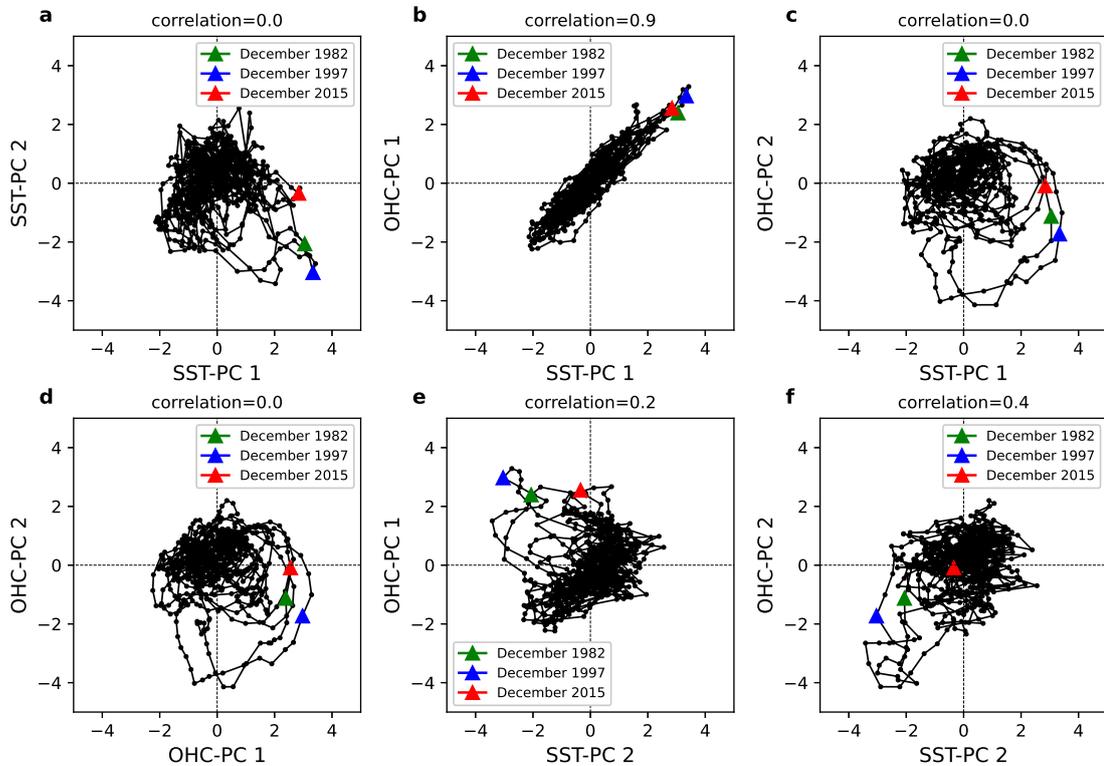


Рис. 2.6: Зависимости между ведущими ГК аномалий ТПО и ТО. (а) ГК-1 и ГК-2 ТПО, (b) ГК-1 ТПО и ГК-1 ТО, (c) ГК-1 ТПО и ГК-2 ТО, (d) ГК-1 и ГК-2 ТО, (e) ГК-2 ТПО и ГК-1 ТО, (f) ГК-2 ТПО и ГК-2 ТО. Цветные метки соответствуют декаблям тех лет, когда происходили экстремально сильные события Эль-Ниньо (согласно классификации <https://ggweather.com/enso/oni.htm>). Временные ряды всех ГК нормированы на свое среднеквадратичное отклонение. Абсолютные значения коэффициента корреляции Пирсона между временными рядами ГК показаны в заголовках к рисункам.

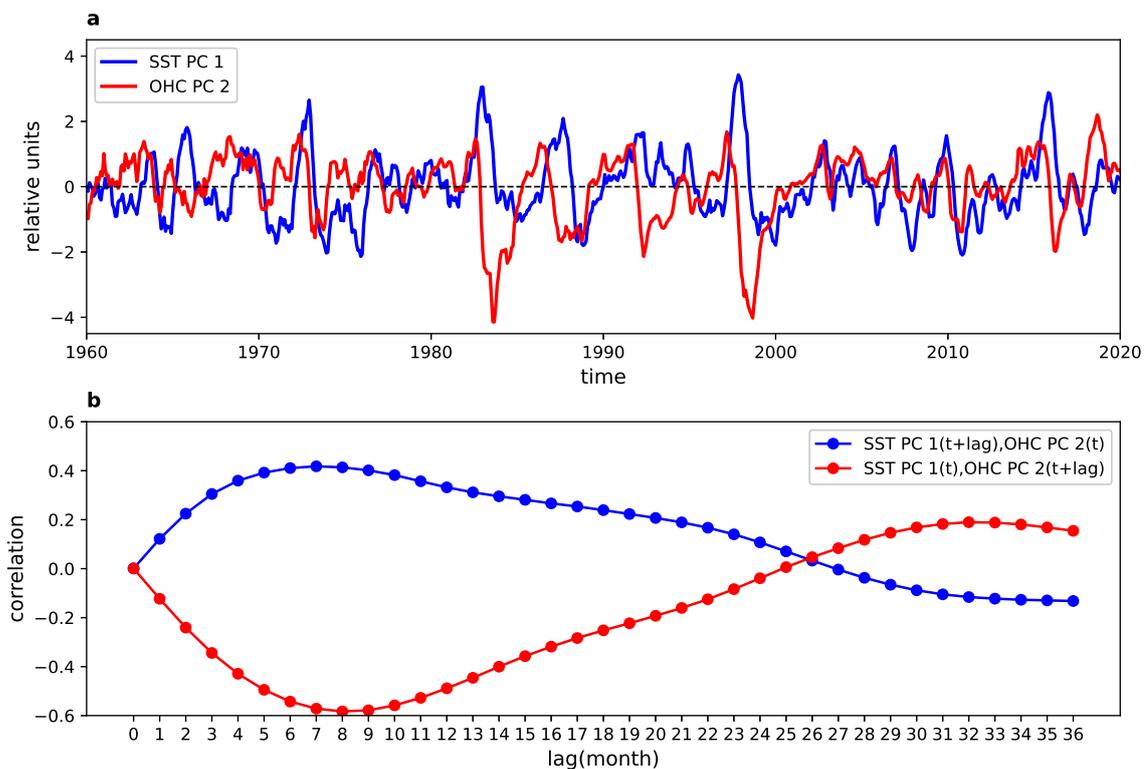


Рис. 2.7: (а) Временные ряды ГК-1 ТПО (синяя линия) ГК-2 ТО (красная линия), нормированные на свои среднеквадратичные отклонения. (б) Лаговые корреляции между этими временными рядами в случае, когда ТПО опережает ТО (синяя кривая) и наоборот (красная кривая).

данных, характеризующих объем воды выше изотермы 20 °C в тропической части Тихого океана.

2.3.2 Построение совместной эмпирической модели динамики ТПО и теплосодержания

Результаты, полученные в предыдущем разделе в ходе анализа данных ТПО и ТО, имеют понятную интерпретацию в рамках одной из наиболее известных концепций динамики ЭНЮК на основе «перезаряжающегося» осциллятора [65, 135]. В базовом варианте этой теории [135] динамика ЭНЮК описывается моделью в виде классического двумерного линейного осциллятора с затуханием, фазовыми переменными которого (играющими роль «координаты» и «скорости») являются ТПО в восточной части Тихого океана и зонально усредненная глубина залегания термоклина. Выявленные нами циклические связи между временными рядами ГК-1 ТПО и ГК-2 ТО (рисунок 2.6 (с)) и структуры соответствующих им пространственных паттернов (рисунки 2.5 (а), (d)) полностью согласуются с положениями данной теории. В этой связи, временные ряды ГК-1 ТПО⁸ и ГК-2 ТО использовались нами в качестве равноценных фазовых переменных при построении эмпирической модели (1.1). Такая совместная модель динамики ТПО и ТО отражает современные представления о физике ЭНЮК, и, в отличие от концептуальных моделей, не имеющих привязки к реальным данным, может использоваться для прогнозирования.

Для того, чтобы продемонстрировать важность совместного учета ТО и ТПО при эмпирическом прогнозе динамики ЭНЮК, мы сравнивали предсказательную способность следующих моделей:

1. Независимые линейные модели (1.2)-(2.2) ТПО и ТО, построенные отдельно по временному ряду ГК-1 ТПО (модель L-ТПО) и временному ряду ГК-2 ТО (модель L-ТО) соответственно;
2. Совместная линейная модель (1.2)-(2.2) ТПО и ТО, построенная по временным рядам ГК-1 ТПО и ГК-2 ТО (модель L-ТПО+ТО);
3. Совместная нелинейная модель (1.3)⁹ (модель NL-ТПО+ТО).

⁸ГК-1 ТПО и ГК-1 ТО содержат идентичную информацию о динамике ЭНЮК (см. рисунки 2.5 (а), (с) – 2.6 (b)). При построении эмпирической модели нами использовалась ГК-1 ТПО, поскольку прогноз ТПО и связанных с ним индексов (например Nino3.4) представляет наибольший практический интерес.

⁹Сезонная зависимость учитывается в данной параметризации модели путем добавления на вход функции (1.3) двумерного гармонического сигнала с периодом 12 месяцев: $\mathbf{c}_n = \left(\cos \frac{2\pi}{12}n, \sin \frac{2\pi}{12}n \right)$.

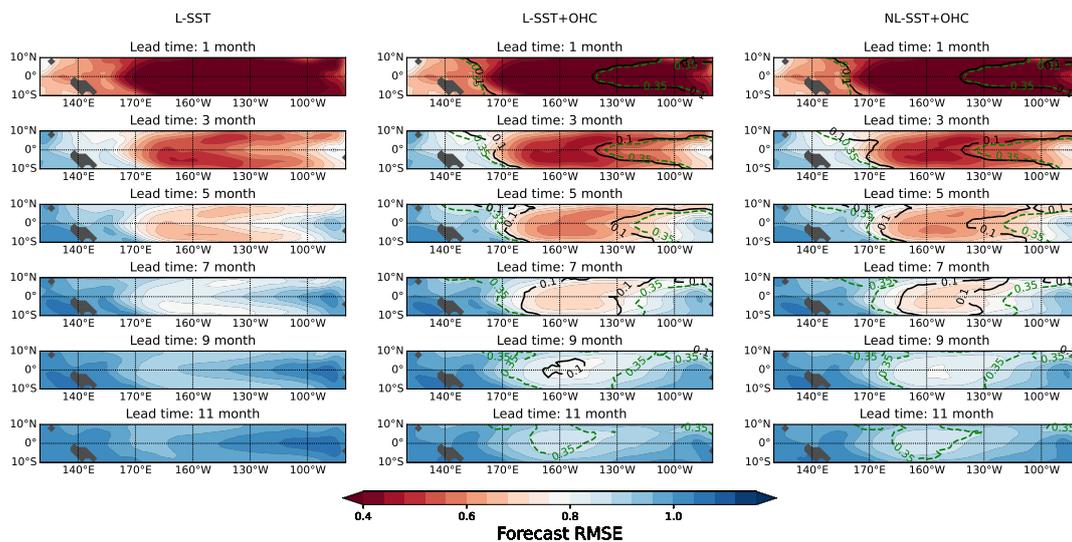


Рис. 2.8: Пространственные распределения КСКО прогноза аномалий ТПО (нормированной на стандартное отклонение временного ряда аномалий ТПО в каждой точке пространственной сетки) для трех эмпирических моделей при различной дальности прогноза (указана в месяцах в заголовке каждой карты). Слева направо: линейная модель, основанная только на данных ТПО, совместная линейная модель, основанная на данных ТПО и ТО, совместная нелинейная модель. Контурные карты в центральном и правом столбце ограничивают области статистически значимого уменьшения ошибки прогноза совместных моделей, по отношению к ошибке прогноза модели, основанной только на ТПО.

Наблюдаемые временные ряды аномалий ТПО и ТО были разделены на обучающую (январь 1960 года – декабрь 2007 года) и контрольную (январь 2008 года – декабрь 2020 года) выборки. При построении моделей задействовалась только обучающая выборка длиной $N = 576$ месяцев. Каждая из моделей была обучена по временным рядам ГК-1 ТПО и/или ГК-2 ТО путем оптимизации ценовой функции (1.9). Оптимальные значения гиперпараметров, найденные в процессе обучения, равны: $l = 2$, $q = 1$ для моделей L-ТПО и L-ТПО+ТО, $l = 3$, $q = 1$ для модели L-ТО и $l = 2$, $m = 5$ для модели NL-ТПО+ТО. На рисунке 2.8 показаны пространственные распределения (карты) КСКО прогноза ТПО, рассчитанные для трех рассматриваемых моделей на обучающей выборке. Данные карты были получены в соответствии с концепцией, изложенной в разделе 1.4.1 первой главы диссертации, согласно которой динамика каждой из остаточных ГК аномалий ТПО (начиная со второй) описывалась авторегрессионной моделью первого порядка AR(1) (1.13). Прогноз в пространстве ГК (ГК-1 с помощью рассматриваемых моделей и остаточных ГК с помощью модели AR(1)) пересчитывался в физическое пространство согласно (1.12), после чего рассчитывалась ошибка прогноза модели в данной точке пространственной сетки. Как видно из рисунка 2.8, наиболее качественный прогноз аномалий ТПО с помощью каждой из трех рассматриваемых моделей достигается в центральной части тропического Тихого океана. При этом совместные модели L-ТПО+ТО и NL-ТПО+ТО имеют значительно меньшую ошибку прогноза в этой области, по сравнению моделью L-ТПО (без учета ТО) для любой дальности прогноза вплоть до 12 месяцев. Области статистически значимого улучшения прогноза совместных моделей L-ТПО+ТО и NL-ТПО+ТО по отношению к модели L-ТПО оценивались на основе суррогатных тестов аналогично тому, как это делалось в предыдущих разделах диссертации. Из рисунка 2.8 следует, что области значимого улучшения прогноза на временах прогнозирования до 12 месяцев локализованы в центральной части тропического Тихого океана, включающей, в том числе, регион 5° с.ш. - 5° ю.ш., 120° з.д. - 170° з.д., в котором определяется ключевой индекс явления Эль-Ниньо Niño3.4.

На рисунке 2.9 представлены помесячные (сезонные) меры предсказательной способности рассматриваемых эмпирических моделей в случае индекса Niño3.4. Верхняя панель рисунка соответствует КСКО (1.14), нижняя панель – коэффициенту корреляции (1.15). Как видно из рисунка, для всех трех моделей характерно ухудшение качества прогноза индекса Niño3.4 в летние месяцы года, связанное с весенним барьером предсказуемости. Тем не менее, совместные модели L-ТПО+ТО и NL-ТПО+ТО демонстрируют значительно лучшую предсказательную способность по сравнению с моделью L-ТПО для всех целевых месяцев года с января по декабрь и на дальности от 4 до 12 месяцев по обоим мерам. На рисунке 2.10 представлены сезонные меры

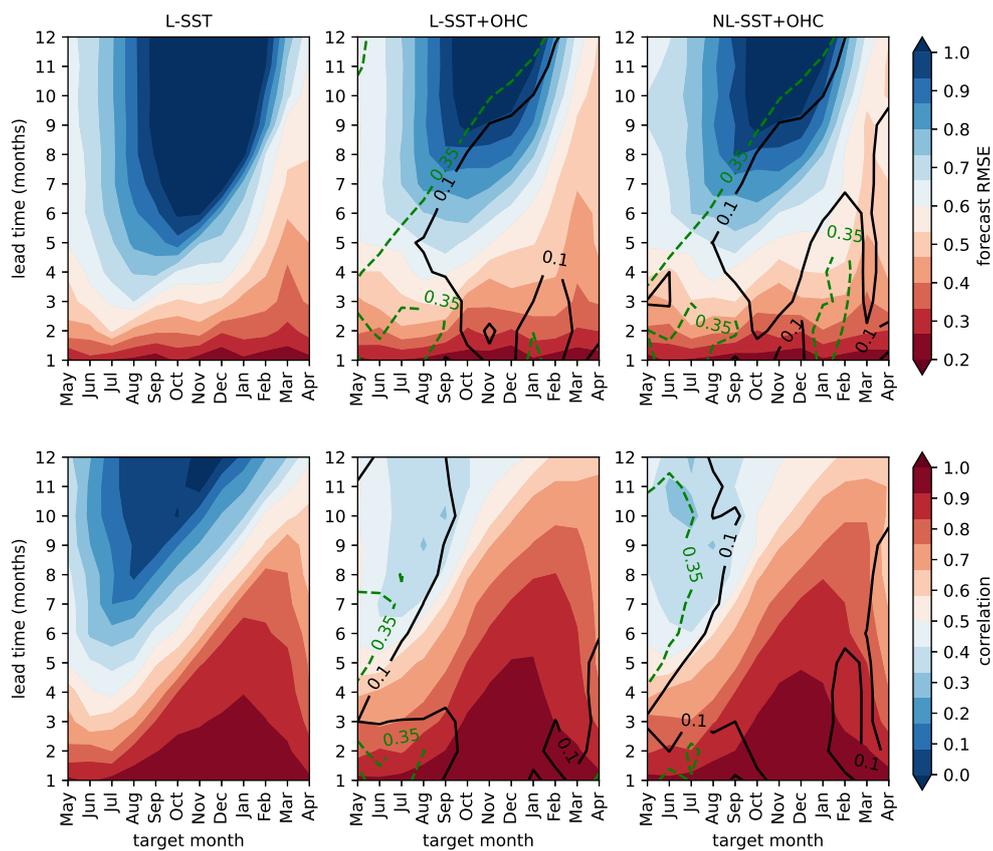


Рис. 2.9: Сезонная зависимость предсказательной способности моделей (в плоскостях прогнозируемый месяц-заблаговременность прогноза) для индекса Niño3.4. Верхняя панель – КСКО, нижняя – коэффициент корреляции. Порядок столбцов, соответствующих каждой модели, тот же, что и на предыдущем рисунке.

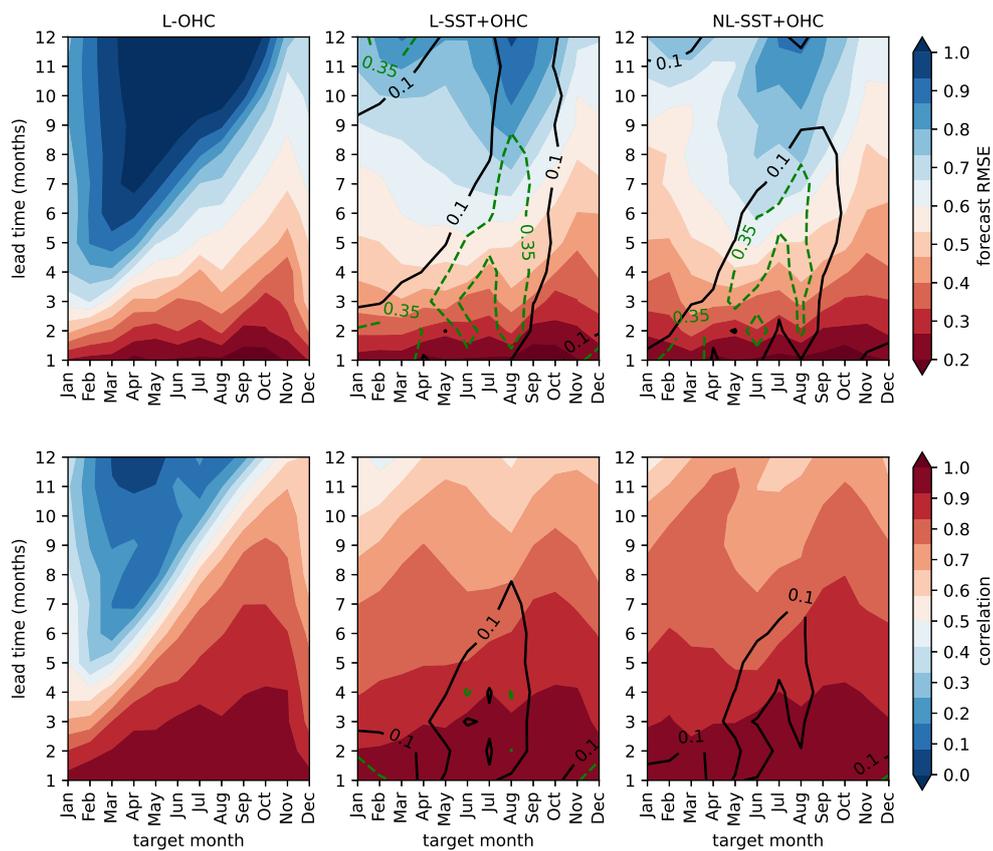


Рис. 2.10: Сезонная зависимость предсказательной способности моделей для индекса аномалий теплосодержания (ГК-2 ТО). Верхняя панель – КСКО, нижняя – коэффициент корреляции. Слева направо: модель, основанная только на данных теплосодержания, совместные линейная и нелинейная модели.

предсказательной способности в случае ГК-2 ТО, которая играет роль индекса, характеризующего динамику подповерхностных теплых вод (накопление и разгрузку) в тропической части Тихого океана. Видно, что данный индекс также характеризуется барьером предсказуемости, который отличается от весеннего барьера в данных ТПО – в данном случае ухудшение прогноза характерно для зимних целевых месяцев (январь-февраль-март). Таким образом, барьеры предсказуемости температуры поверхности и теплосодержания тропического Тихого океана сдвинуты друг относительно друга на 5-7 месяцев, что согласуется с рисунком 2.7, который демонстрирует эквивалентный фазовый сдвиг между соответствующими главными компонентами. Совместные модели L-ТПО+ТО и NL-ТПО+ТО показывают еще более значительное (чем в случае индекса ТПО Nino3.4) улучшение предсказательной способности индекса теплосодержания по сравнению с индивидуальной моделью L-ТО, фактически нивелируя сезонный «зимний» барьер предсказуемости.

Резюмируя полученные результаты можно сделать вывод о том, что совместные модели L-ТПО+ТО и NL-ТПО+ТО обладают существенно лучшей предсказательной способностью по сравнению с индивидуальными моделями L-ТПО и L-ТО при прогнозе изменчивости как температуры поверхности, так и теплосодержания верхнего слоя тропического Тихого океана. В тоже время, прогностические способности совместных линейной (L-ТПО+ТО) и нелинейной (NL-ТПО+ТО) моделей между собой практически не отличаются, что указывает на то, что учет нелинейности в эмпирических прогностических моделях ЭНЮК не играет существенной роли на внутригодовых масштабах.

Проанализируем способности индивидуальной (L-ТПО) и совместной (L-ТПО+ТО) линейных моделей к предсказанию конкретных событий Эль-Ниньо/Ла-Нинья на примере прогноза индекса Nino3.4. На рисунке 2.11 показаны примеры прогнозов сильных (по классификации <https://ggweather.com/enso/oni.htm>) событий Эль-Ниньо 1972-73, 1982-83 и 1997-98 годов, и Ла-Нинья 1973-74, 1998-99 и 1999-00 годов, принадлежащих обучающей выборке. Слева показаны прогнозы через весенний барьер, стартующие в марте, справа – прогнозы, стартующие в сентябре далеко от весеннего барьера. Стоит отметить, что аномалии ТПО в ходе сильных событий Эль-Ниньо/Ла-Нинья, как правило, начинают развиваться в конце весны – начале лета Северного полушария, достигая своего пика предстоящей зимой. В этой связи именно весенние прогнозы (т.е. прогнозы событий до того, как они начались) представляют наибольший практический интерес. Из рисунка видно, что модель L-ТПО, не учитывающая информацию о динамике ТО, имеет тенденцию к неинформативному нейтральному прогнозу индекса Nino3.4 через весенний барьер. При этом, совместная модель L-ТПО+ТО, стартуя с нейтральных (почти нулевых) значений индекса Nino3.4 в марте, корректно прогно-

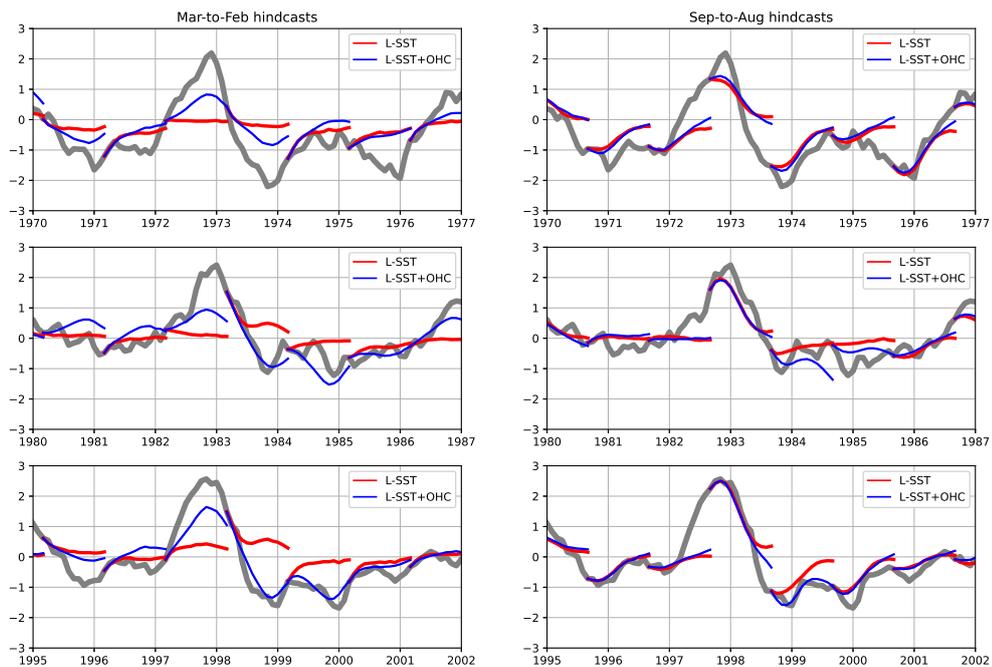


Рис. 2.11: Примеры ретроспективных (на обучающей выборке) прогнозов событий Эль-Ниньо/Ла-Нинья (на основе индекса Niño3.4) с помощью линейной модели, основанной только на данных ТПО (красные кривые) и линейной совместной модели (синие кривые). Левый столбец – прогнозы, начинающиеся с марта, правый – с сентября. Временной ряд индекса Niño3.4 обозначен серым цветом.

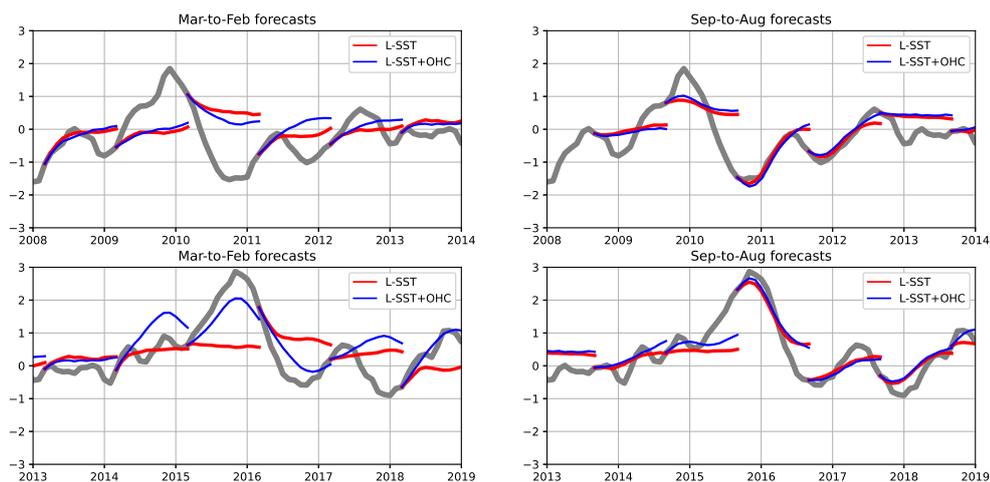


Рис. 2.12: Примеры реальных (на тестовой выборке) прогнозов событий Эль-Ниньо/Ла-Нинья.

зирует тенденции к развитию всех рассматриваемых событий Эль-Ниньо/Ла-Нинья. Осенние прогнозы совместной и индивидуальных моделей, стартующие в разгаре развития событий, почти не отличаются. Контрольная выборка содержит одно умеренное и одно сильное события Эль-Ниньо 2009-10 и 2015-16 годов соответственно, а также несколько событий Ла-Нинья разной интенсивности. Как видно из рисунка 3.1, и в этом случае совместная модель L-ТПО+ТО демонстрирует более корректный прогноз событий через весенний барьер, верно предсказывая сильное Эль-Ниньо 2015-16 годов и последовавшее за ним Ла-Нинья 2016-17 годов. В тоже время, совместная модель не показала преимуществ при прогнозе умеренного Эль-Ниньо 2009-10 годов. Данное обстоятельство может быть связано с тем, что в отличие от остальных рассмотренных Эль-Ниньо, это событие относят к центрально-тихоокеанскому типу, при прогнозе которого роль предикторов на основе теплосодержания снижается [136]. Заметим также, что совместная модель, в отличие от индивидуальной, предсказала несостоявшееся Эль-Ниньо 2014-15 года, которое также ложно предсказывалось многими известными моделями ЭНЮК исходя из сильных положительных аномалий теплосодержания в западной части Тихого океана в начале 2014 года [74]. Данный результат свидетельствует о том, что увеличение теплосодержания в экваториальной части Тихого океана является необходимым, но не достаточным условием для возникновения Эль-Ниньо. Иными словами, существуют дополнительные факторы, которые могут как способствовать, так и препятствовать процессу «разрядки» накопленного в океане

тепла и дальнейшему увеличению поверхностной температуры. В первую очередь к таким факторам относятся флуктуации зонального ветра (всплески восточных и западных ветров), которые могут обуславливаться как внутрисезонными процессами в тропической атмосфере (например колебаниями Маддена-Джулиана [59, 74, 137]), так и крупномасштабными атмосферными процессами в средних широтах, которые обсуждались в предыдущем разделе главы.

2.4 Заключение

В главе решалась задача построения эмпирической модели для эффективного внутригодового прогноза изменчивости явления Эль-Ниньо – Южное Колебание (ЭНЮК), который существенно осложнен из-за существования весеннего барьера предсказуемости климатической динамики в тропической части Тихого океана. Был проведен анализ пространственно-распределенных данных температуры поверхности и теплосодержания верхнего (0-300 м) слоя тропического Тихого океана, а также данных давления на уровне моря (ДУМ) на предмет выявления прогностических переменных (предикторов), позволяющих преодолеть весенний барьер предсказуемости при эмпирическом прогнозе явления Эль-Ниньо. По данным ДУМ определен индекс, который отражает состояние субтропической атмосферной циркуляции в зимне-весенний сезон (февраль-март) и имеет статистически значимые длинные корреляции с колебанием Эль-Ниньо в предстоящий период с июня текущего года по май следующего года. Установлено, что найденный индекс может описывать влияние на динамику ЭНЮК крупномасштабной пространственной структуры аномалий ДУМ в средних широтах северного Тихого океана. Показано, что учет найденного предиктора в линейной эмпирической модели индекса явления Эль-Ниньо Nino3.4 позволяет добиться статистически значимого улучшения ее предсказательной способности на масштабах от нескольких месяцев до одного года, в том числе при прогнозе через весенний барьер.

В ходе анализа данных температуры поверхности океана (ТПО) выявлен индекс, который описывает связанную с Эль-Ниньо изменчивость тропического Тихого океана и имеет высокую корреляцию с традиционным индексом Nino3.4. По данным теплосодержания океана (ТО) определен индекс, который может характеризовать процессы накопления и разгрузки подповерхностных теплых вод в центральной и западной части Тихого океана, которые, согласно современным концепциям, являются ключевой составляющей цикла колебания Эль-Ниньо. В ходе анализа лаговых корреляций выявлен фазовый сдвиг найденных индексов по отношению друг к другу на 5-9 месяцев. Установлено, что, в отличие от традиционных индексов Эль-Ниньо на основе

ТПО, найденный индекс ТО характеризуется зимним барьером предсказуемости. Проанализированы предсказательные способности индивидуальных (основанных только на данных ТПО либо ТО) и совместных эмпирических моделей, как линейных так и нелинейных. Показано, что совместные модели, учитывающие найденные индексы как равноценные динамические переменные, имеют существенно лучшую предсказательную способность по сравнению с моделями основанными только на ТПО (ТО) и позволяют преодолеть сезонные (весенний в индексе ТПО и зимний в индексе ТО) барьеры предсказуемости. При этом прогностические способности совместных линейной и нелинейной моделей практически не отличаются, что может свидетельствовать о том, что учет нелинейности в эмпирических прогностических моделях ЭНЮК не играет роли на внутригодовых масштабах. Тем не менее, в следующей главе диссертации будет продемонстрировано, что нелинейные связи между компонентами ЭНЮК значимо проявляют себя на более длинных, межгодовых временных масштабах, и корректно воспроизводятся только нелинейной моделью.

Таким образом, в данной главе были найдены эффективные атмосферные и океанические предикторы, которые могут использоваться в современных эмпирических моделях для внутригодового прогноза связанной с ЭНЮК изменчивости ТПО. Атмосферный предиктор на основе ДУМ, при этом, играет роль заданного форсинга, который влияет на амплитуду годового цикла ЭНЮК с июня по май, а океанический предиктор на основе ТО является динамической переменной, дополняющей ТПО-переменную. Нами не исследовался вопрос об относительном вкладе этих предикторов в прогноз тех или иных типов событий Эль-Ниньо/Ла-Нинья. Данный вопрос является частью более общего вопроса о взаимосвязи внетропических процессов крупномасштабной атмосферной изменчивости с межгодовой изменчивостью в тропических широтах, исследование которого выходит за рамки диссертации.

Следующая глава диссертации посвящена эмпирическому исследованию динамики ЭНЮК на межгодовых и вековых временных масштабах.

Глава 3

Исследование межгодовой и вековой изменчивости явления Эль-Ниньо на основе эмпирических моделей

3.1 Введение

Данная глава диссертации посвящена эмпирическому исследованию низкочастотной – межгодовой и вековой – изменчивости явления Эль-Ниньо – Южное Колебание (ЭНЮК). Для анализа изменчивости на таких масштабах необходимо задействовать ряды данных, которые захватывают динамику ЭНЮК на достаточно протяженном интервале времени в прошлом. Считается [138], что глобальные (т.е. покрывающие большую часть поверхности Земли) измерения исторических приземных данных (температуры поверхности океана, давления на уровне моря и т.д.) являются достаточно надежными приблизительно с 1950 года. Данные на глобальной пространственной сетке в более ранний период (на протяжении около 100 лет до 1950 года) восстанавливаются путем экстраполяции современных данных в прошлое с учетом доступных в тот период наблюдений (в первую очередь корабельных измерений). Результаты, получаемые при такой процедуре, могут существенно отличаться в зависимости от используемых методов экстраполяции [139]. Для восстановления отдельных характеристик ЭНЮК в еще более далеком прошлом используются различные косвенные характеристики климата (так называемые климатические прокси), такие как тропические кораллы, годовые кольца деревьев, ледяные керны, донные отложения. Реконструкция соответствующих временных рядов с использованием прокси-данных также сопряжена с большими техническими трудностями и неопределенностью получаемых результатов [100]. В этой связи, ключевым аспектом при построении эмпирических

моделей ЭНЮК по временным рядам его характеристик в достаточно далеком прошлом является проверка устойчивости получаемых результатов при использовании различных типов данных. В разделе 3.2 данной главы на основе предикторов, полученных в предыдущей главе на основе анализа пространственно распределенных данных температуры поверхности (ТПО) и теплосодержания верхнего слоя (ТО) тропического Тихого океана, исследуется структура связей между этими ключевыми компонентами ЭНЮК на межгодовых масштабах, а также эволюция этих связей с начала 20 века до настоящего времени. В разделе 3.3 строится эмпирическая модель изменчивости ЭНЮК на масштабах десятилетий по временным рядам индексов Niño3.4 протяженностью 850 лет, получаемым на основе климатических прокси. На основе модели исследуется отклик ТПО в экваториальной части Тихого океана на солнечную активность. В обоих случаях демонстрируется, что получаемые результаты совпадают для различных типов используемых данных.

Результаты, представленные в главе, опубликованы в статьях [A5, A6] и тезисах [A30].

3.2 Структура связей между компонентами ЭНЮК на межгодовых масштабах

3.2.1 Межгодовые зависимости между аномалиями ТПО и теплосодержания верхнего слоя тропического Тихого океана

Обсуждавшаяся в предыдущей главе проблема весеннего барьера предсказуемости тесно связана с ярко выраженной сезонной цикличностью в изменчивости климатических характеристик тропического Тихого океана [140]. Так, аномалии ТПО в ходе сильных событий ЭНЮК, как правило, начинают расти в конце весны-начале лета (май-июнь), достигают своего пика предстоящей зимой, а к следующей весне (т.е. через год после начала роста) вновь возвращаются к нейтральным значениям [141]. В работе [127] было показано, что сезонный цикл аномалий ТПО в тропическом Тихом океане описывается первой ЭОФ (далее – первой временной ЭОФ), рассчитанной по набору данных, составленному из неперекрывающихся последовательных 12-месячных сегментов¹ временного ряда индекса Niño3.4, начинающихся в июне. Соответствующая

¹В данном случае эти сегменты играют роль последовательных векторов данных x_n , к которым применяется ЭОФ-разложение (1.12). Ясно, что результат такого разложения будет зависеть от стартового месяца сегментов.

этой ЭОФ главная компонента (временной ряд с шагом один год) характеризует амплитуду цикла. Данный результат указывает на то, что динамика аномалий ТПО в различные месяцы года внутри цикла с июня по май сильно скоррелирована, и, следовательно, предсказуема. Весенний барьер при таком описании трактуется как нарушение этой скоррелированности при переходе между последовательными циклами. В данном разделе мы применяем аналогичный анализ для выделения сезонных циклов по временным рядам ГК-1 ТПО и ГК-2 ТО на интервале с 1960 по 2020 год, которые, как было показано (см. раздел 2.3), содержат наиболее важную информацию о динамике ЭНЮК. На рисунке 3.1 (а) показаны доли вариации, захваченные первой временной ЭОФ, рассчитанной по данным, составленным из 12-месячных сегментов ГК-1 ТПО, в зависимости от стартового месяца сегмента. Видно, что наибольшая часть вариации (около 86 %) сосредоточена в первой ЭОФ в том случае, когда временной ряд ГК-1 ТПО разбивается на сегменты, начинающиеся в мае и июне, что согласуется с результатами работы [127] для индекса Nino3.4. Соответствующий сезонный цикл (временной ЭОФ) показан на рисунке 3.1 (с). Видно, что пик цикла аномалий ТПО ожидаемо приходится на зимние (декабрь-январь) месяцы Северного полушария. Результаты, полученные при применении аналогичной процедуры к временному ряду ГК-2 ТО, приведены на рисунках 3.1 (b) и (d). Как видно из рисунков, доля вариации, захваченная первой ЭОФ в случае, когда временной ряд ГК-2 ТО разбивается на сегменты, начинающиеся в декабре и январе, составляет 88 % и существенно превышает долю захваченной вариации, соответствующую другим начальным месяцам сегментов. Таким образом, динамика аномалий теплосодержания верхнего слоя тропического Тихого океана также характеризуется ярко выраженным сезонным циклом, который сдвинут относительно цикла ТПО на 6-8 месяцев и, в отличие от него, достигает своего пика в конце лета-начале осени Северного полушария. Данный результат объясняет наличие зимнего барьера предсказуемости в данных теплосодержания, обнаруженного в предыдущей главе диссертации.

Способны ли эмпирические модели ЭНЮК, построенные нами в предыдущей главе, воспроизвести найденные сезонные циклы в данных ТПО и ТО? Для ответа на этот вопрос, мы применили описанный выше анализ к временным рядам ГК-1 ТПО и ГК-2 ТО, произведенным совместными линейной L-ТПО+ТО и нелинейной NL-ТПО+ТО моделями. При этом, с помощью обеих моделей были сгенерированы ансамбли из 1000 стохастических реализаций (такой же длины, как и временные ряды данных), по которым были оценены 90% доверительные интервалы для первой ЭОФ и доли захваченной ей вариации в зависимости от месяца года. Результаты показаны на рисунке 3.1 синим и красным цветом соответственно. Видно, что обе модели корректно воспроизводят сезонные циклы двух ключевых переменных ЭНЮК.

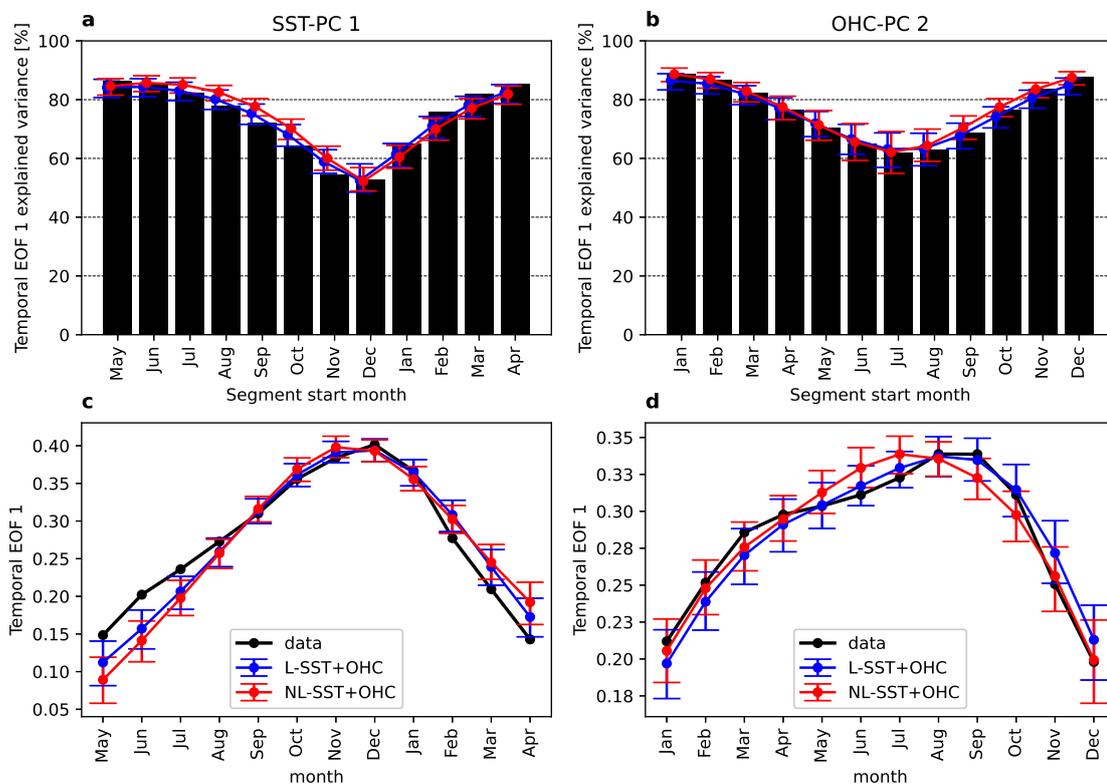


Рис. 3.1: Сезонные циклы в динамике ЭНЮК. Верхняя панель: доли вариации ГК-1 ТПО (a), и ГК-2 ТО (b), описываемые главной временной ЭОФ в зависимости от начального месяца сегмента (см. основной текст). Нижняя панель: главная временная ЭОФ ГК-1 ТПО (c) и ГК-2 ТО (d). На каждом из рисунков черным цветом показаны характеристики, полученные по данным. Синим и красным показаны средние значения характеристик и их 90% доверительные интервалы, полученные по ансамблям из 1000 реализаций моделей L-ТПО+ТО и NL-ТПО+ТО соответственно.

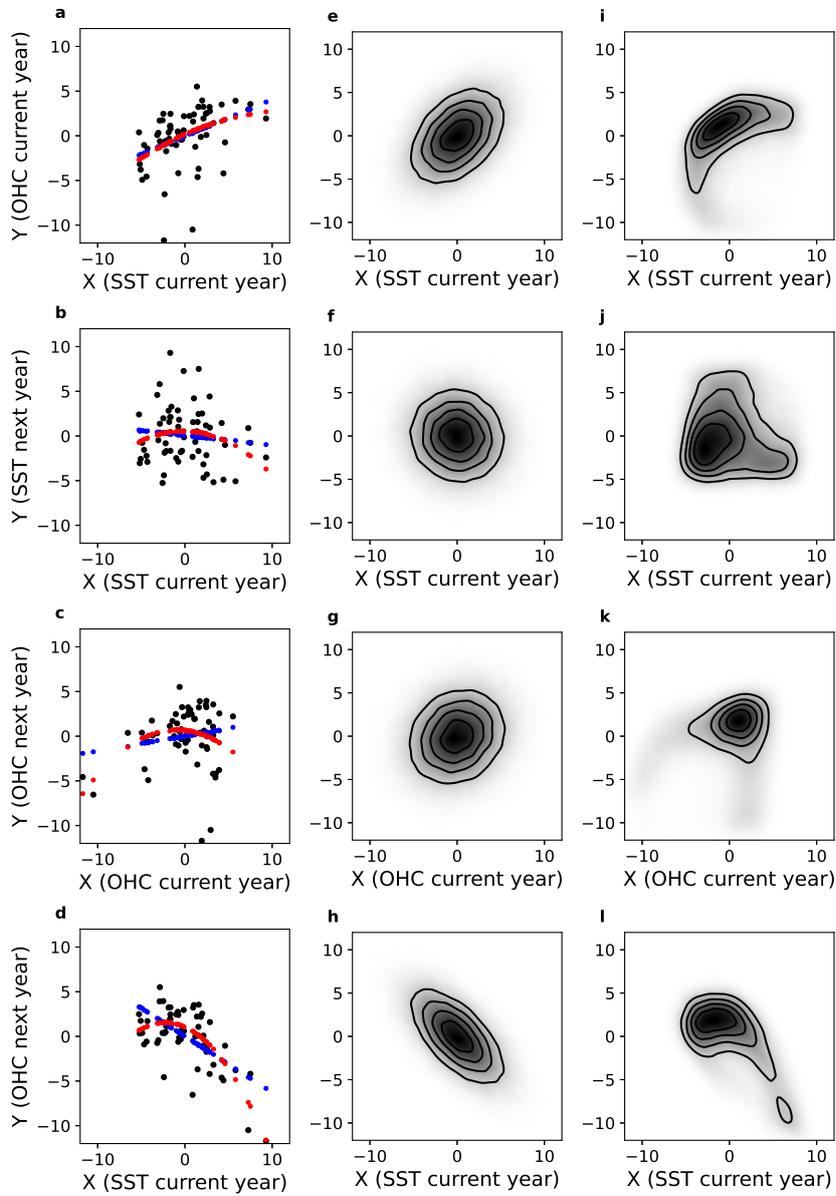


Рис. 3.2: Левая колонка: зависимости между амплитудами сезонных циклов ТПО и ТО, полученные по данным (черные точки) и их аппроксимации на основе линейной (синяя кривая) и квадратичной (красная кривая) регрессионных моделей. Центральная и правая колонки: плотности вероятности, рассчитанные в тех же плоскостях по ансамблям из 1000 реализаций моделей L-ТПО+ТО и NL-ТПО+ТО соответственно.

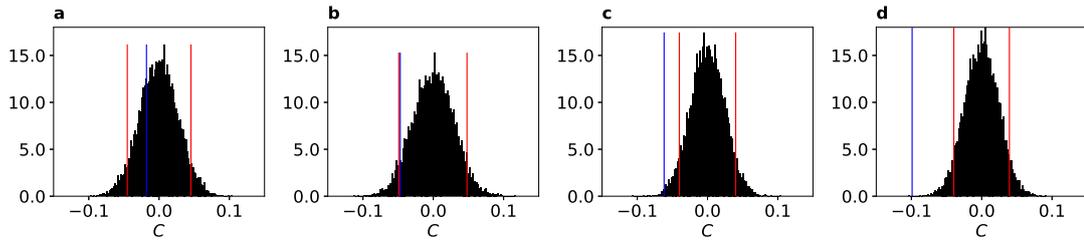


Рис. 3.3: Проверка статистической значимости квадратичной аппроксимации зависимостей, показанных на рисунке 3.2. На каждом из рисунков слева направо показаны плотности распределения суррогатных значений коэффициента C (см. текст), соответствующие определенной плоскости на рисунках 3.2 (a)-(d). Синими линиями обозначены значения коэффициента C квадратичной модели, подогнанной к данным. Красными линиями отмечены 10-й и 90-й процентиля суррогатных распределений.

Проанализируем межгодовые зависимости амплитуд сезонных циклов ТПО и ТО², показанные на рисунках 3.2 (a)-(d). Для того чтобы определить вид этих зависимостей, мы строили линейную $Y = B \cdot X + A + \varepsilon$ и квадратичную $Y = C \cdot X^2 + B \cdot X + A + \varepsilon$ регрессионные модели в каждой из рассматриваемых плоскостей. Коэффициенты регрессии A , B и C и дисперсия ошибки аппроксимации ε оценивались с помощью традиционного метода наименьших квадратов. Результаты регрессии для обеих моделей показаны на рисунках 3.2 (a)-(d) красным и синим цветом. Как видно из рисунков, различия между линейной и нелинейной регрессиями наиболее отчетливо видны в плоскостях (b) и (d). Проверка статистической значимости квадратичной аппроксимации осуществлялась на основе суррогатного теста, идея которого была описана в разделе 1.3.3 первой главы диссертации. Нулевая гипотеза, которая опровергалась в ходе данного теста, предполагает, что полученное по данным значение коэффициента C при квадратичной степени регрессионной модели могло быть получено по выборке такой же длины, но с заведомо линейной зависимостью между предиктором X и предиктантом Y . Суррогатный ансамбль из 1000 случайных выборок предиктанта Y при заданном предикторе X , соответствующий данной нулевой гипотезе, получался в каждой из рассматриваемых плоскостей с помощью линейной регрессионной модели, подогнанной к исходной выборке данных. По каждой паре X, Y из суррогатного ансамбля строилась квадратичная регрессионная модель. Распределения значений коэффициента C , получаемые по суррогатным ансамблям в каждой плоскости, показаны на рисунке 3.3. Значения коэффициента C , получаемые по данным, отмечены

²Напомним, что эти амплитуды описываются временными рядами с шагом один год, каждая точка которых является проекцией (см. соотношение (1.11)) на ведущий ЭОФ 12-месячных фрагментов с июня по май в случае ГК-1 ТПО и с января по декабрь в случае ГК-2 ТО.

на этих рисунках синими линиями. Видно, что в плоскостях (b) и (d) значение коэффициента C , полученное по данным, лежит на хвосте суррогатного распределения. В этих случаях квадратичная аппроксимация является статистически значимой на уровне 0.1. Таким образом, мы выявили нелинейные зависимости между амплитудами сезонных циклов аномалий ТПО и ТО (конкретнее – зависимости между амплитудой цикла ТПО в текущем году и цикла ТПО в последующем году (рисунки 3.2–3.3 (b)) и между амплитудой цикла ТПО в текущем году и цикла ТО в последующем году (рисунки 3.2–3.3 (d))) на межгодовых масштабах.

Также, мы проанализировали способности построенных в предыдущей главе эмпирических моделей к воспроизведению установленных межгодовых зависимостей между компонентами ЭНЮК. На рисунках 3.2 (e)-(l) показаны плотности вероятности (ПВ) амплитуд сезонных циклов ТПО и ТО, рассчитанные по ансамблям из 1000 временных рядов, произведенным каждой из моделей. Линейная модель L-ТПО+ТО (рисунки 3.2 (e)-(h)) по построению не способна отразить нелинейные закономерности и демонстрирует квази-гауссовы ПВ в каждой из плоскостей. В то же время, нелинейная модель NL-ТПО+ТО характеризуется существенно негауссовыми ПВ в каждой из плоскостей (рисунки 3.2 (i)-(l)), подтверждая межгодовые нелинейные связи между компонентами ЭНЮК. Таким образом, несмотря на то, что обе модели ЭНЮК демонстрируют одинаковую предсказательную способность на внутригодовых масштабах (см. раздел 2.3), нелинейная модель является более адекватной при описании межгодовой динамики.

3.2.2 Эволюция характера межгодовых зависимостей в системе ЭНЮК с начала 20 века

В настоящее время достоверно установлено, что характеристики ЭНЮК (период, локализация, амплитуда, структура пространственных паттернов) изменяются на масштабах десятилетий (см., например, обзор [138] и ссылки там). Так, в работе [142] было показано, что доминирующий период и амплитуда ЭНЮК увеличились с конца 70-х годов прошлого века. Там же было выдвинуто предположение о связи этих изменений с известным климатическим сдвигом 1976–77 годов [143, 144]. В работе [145] отмечался общий тренд к увеличению числа экстремальных событий Эль-Ниньо во второй половине 20 века. В данном разделе нами предпринята попытка проследить изменения характеристик ЭНЮК с начала 20 века в контексте эволюции характера связей между его компонентами, которые были обнаружены в предыдущем разделе. Поскольку данные, характеризующие теплосодержание океана, доступны только с середины 20 века, мы концентрируемся на анализе данных ТПО. Мы рассмотрели временные ряды

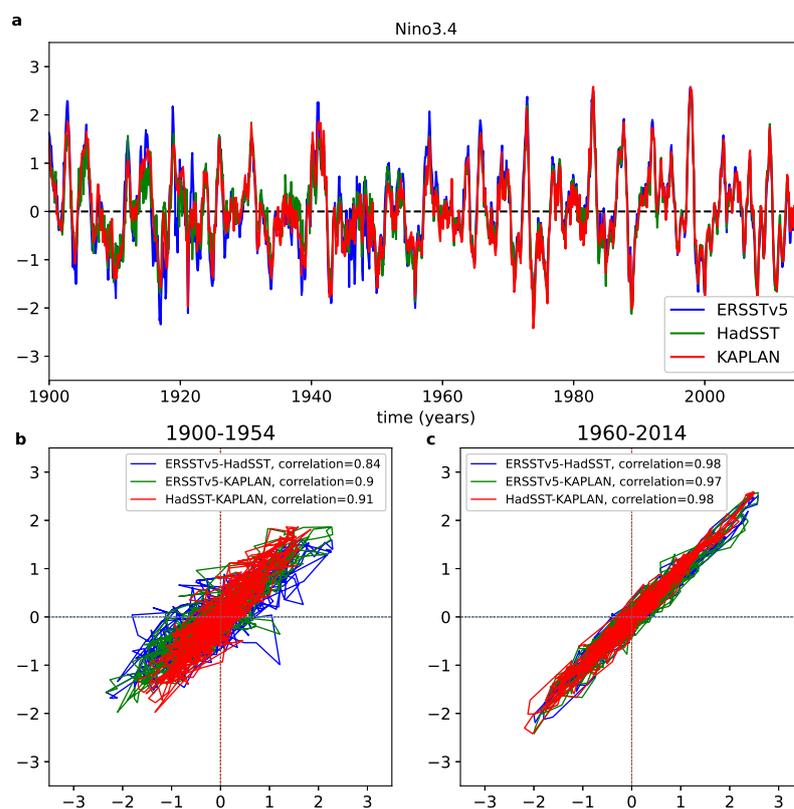


Рис. 3.4: Временные ряды индексов Nino3.4, рассчитанные по различным базам данных (а). Попарные зависимости между временными рядами индексов на интервале с 1900 по 1954 год (b) и с 1960 по 2014 год (c). В легендах на рисунках (b) и (c) указаны значения коэффициента корреляции Пирсона между временными рядами индексов.

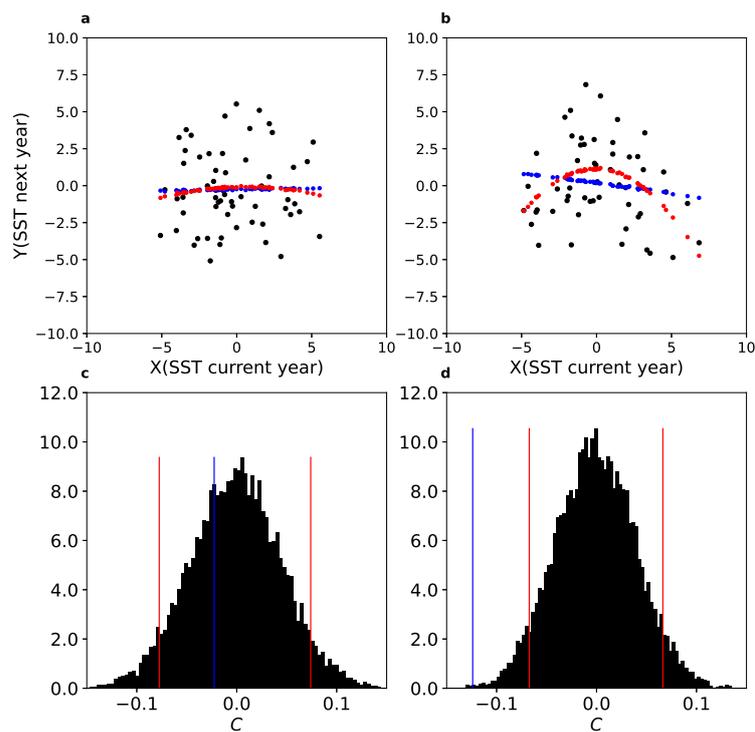


Рис. 3.5: Верхняя панель: зависимости между амплитудами годовых циклов аномалий ТПО на интервале с 1900 по 1954 год (а) и с 1960 по 2014 год (b). Нижняя панель: проверка статистической значимости квадратичной аппроксимации этих зависимостей (с)-(d) на основе теста, описанного в предыдущем разделе. Индекс аномалий ТПО Niño3.4 рассчитывался по данным ERSSTv5.

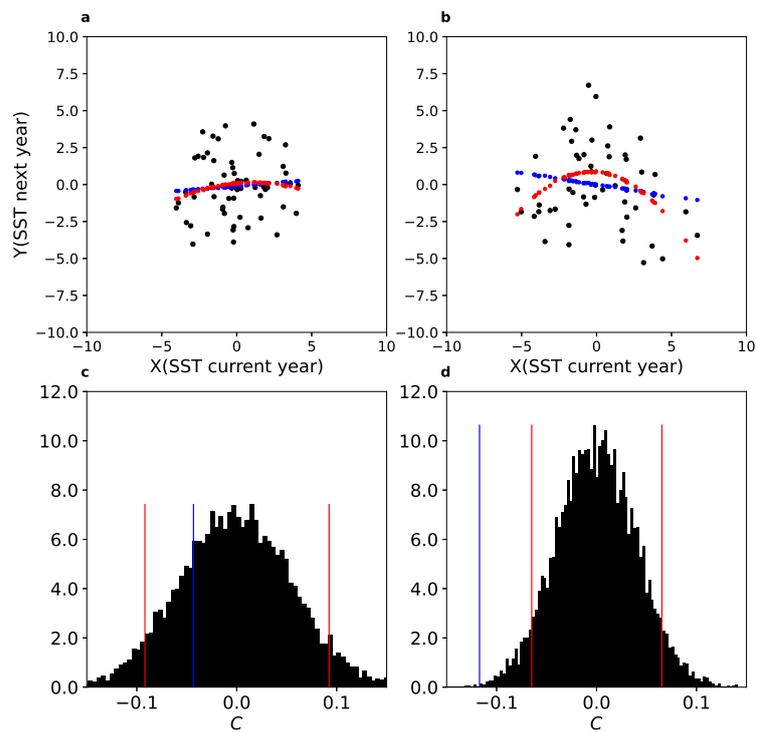


Рис. 3.6: То же, что на рисунке 3.5, но для индекса Niño3.4, рассчитанного по данным HadSST.

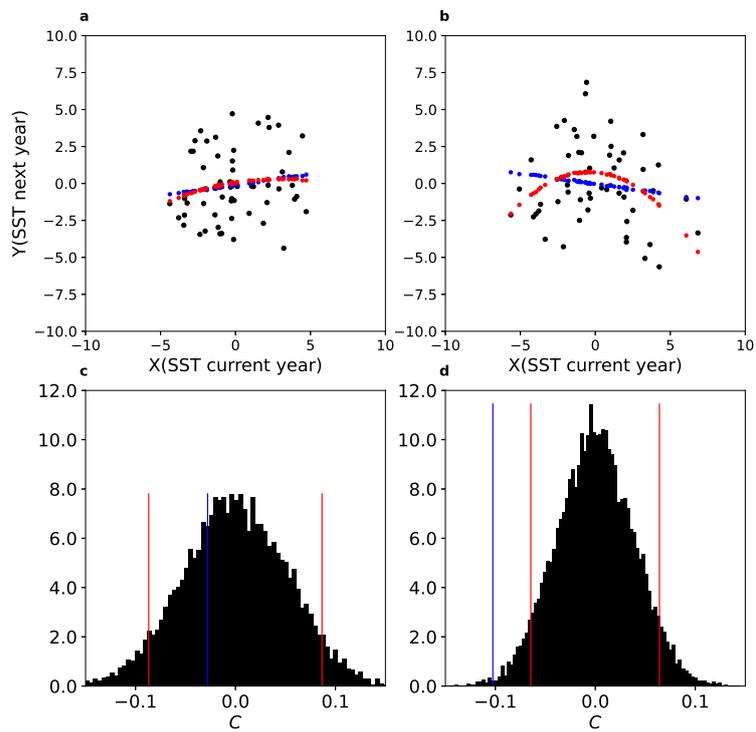


Рис. 3.7: То же, что на рисунках 3.5-3.6, но для индекса Niño3.4, рассчитанного по данным Карлан.

индексов Niño3.4 (см. рисунок 3.4 (а)) на интервале с 1900 по 2014 год с шагом по времени 1 месяц, рассчитанные по трем различным базам данных: ERSSTv5 (см. описание в разделе 2.2 второй главы диссертации), Hadley Centre SST dataset (HadISST v1.1) [146] и Kaplan v2 [139]. Как видно из рисунков 3.4 (b)-(c), в первой половине 20 века между временными рядами индексов присутствуют несущественные отличия, в то время как после 1960 года эти ряды, фактически, совпадают. С помощью методики, описанной в предыдущем разделе, мы проанализировали сезонные циклы ЭНЮК и межгодовые зависимости их амплитуд на двух неперекрывающихся временных интервалах с 1900 по 1954 и с 1960 по 2014 годы. Результаты анализа суммированы на рисунках 3.5-3.7. Мы получили, что на обоих рассматриваемых интервалах времени наибольшая часть вариации (80-82% на интервале с 1900 по 1954 год и 85-87% на интервале с 1960 по 2014 год в зависимости от используемых данных индекса Niño3.4) описывается первой временной ЭОФ в том случае, когда временные ряды индекса Niño3.4 разбиваются на 12-месячные сегменты, начинающиеся в мае либо в июне, что согласуется с результатами, полученными в предыдущем разделе. Таким образом, на протяжении 20 века характер сезонной изменчивости аномалий ТПО экваториального Тихого океана качественно не изменяется. В то же время, межгодовая зависимость амплитуд сезонных циклов аномалий ТПО на рассматриваемых интервалах выглядит совершенно по-разному. Как видно из рисунков 3.5-3.7, эта зависимость переходит от линейной в первой половине 20 века (рисунки 3.5-3.7 (а)), к существенно нелинейной во второй половине 20 и начале 21 века (рисунки 3.5-3.7 (b)). Данный результат согласуется с известным общезначимым принципом, который состоит в том, что нелинейные эффекты в колебательной системе обычно проявляются при увеличении амплитуды колебаний. В данном случае это увеличение, по-видимому, обусловлено упомянутым выше ростом числа экстремально сильных событий Эль-Ниньо во второй половине 20 века. Отметим, что полученные результаты полностью совпадают для трех различных баз данных ТПО.

3.3 Отклик ЭНЮК на вариации солнечной активности на столетних масштабах

3.3.1 Механизмы влияния солнечной активности на климат

В настоящее время отсутствует консенсус в отношении вопросов о степени влияния солнечной активности на климат и доминирующих физических механизмах этого влияния. Так, эмпирическим путем были выявлены [147] корреляции событий

Бонда (повышения интенсивности движения льда в северной Атлантике) с солнечной активностью в голоцене (современная эпоха, начавшаяся 12 тыс. лет назад). Однако крайне незначительные относительные изменения солнечной активности – не более 0.5% – говорят о том, что, если существенная связь солнечного форсинга с климатом действительно есть, то должен существовать механизм усиления этого форсинга [148]. Широко распространены концепции, задействующие стратосферу как основное звено передачи и усиления солнечного сигнала. Например, предполагается [149], что вариации разогрева стратосферы из-за поглощения озоном ультрафиолетовой части спектра солнечного излучения инициируют изменения меридионального градиента температуры в нижней стратосфере, вследствие чего меняется индекс рефракции планетарных волн, перераспределяются скорости в верхней тропосфере, что в результате приводит к климатическим изменениям в средних широтах. В рамках другой концепции [150] увеличение солнечной радиации разогревает океан в тропиках и субтропиках, увеличивает конвекцию и, тем самым, усиливает температурный градиент в тропопаузе, меняя рефракцию тропосферных волн так же, как и в механизме, упомянутом выше. Принципиально иной механизм влияния солнечного сигнала на климат посредством воздействия на подсистему, лежащую в основе ЭНЮК, описан в работе [151], где была выдвинута концепция так называемого океанского динамического термостата. На основе экспериментов с использованием модели ЭНЮК промежуточной сложности в данной работе было установлено, что даже слабые изменения солнечной активности приводят к заметному изменению градиента ТПО между западной и восточной частями экваториального Тихого океана. Данное обстоятельство обуславливается тем, что отклик океана на равномерное нагревание не является пространственно однородным [152]. В западной части Тихого океана ТПО в большей степени определяется тепловым балансом между океаном и атмосферой и, следовательно, чувствительна к изменениям приходящей солнечной энергии. ТПО в восточной части Тихого океана, напротив, в основном определяется процессами горизонтальной и вертикальной адвекции за счет пассатов и апвеллинга. Таким образом, увеличение солнечной активности приводит к увеличению градиента ТПО (за счет нагревания океана на западе) между востоком и западом Тихого океана, способствуя усилению пассатов и, как следствию, еще большему увеличению градиента ТПО (обратная связь Бьеркенса [60]), что приводит к преобладанию холодной фазы (Ла-Нинья) в динамике ЭНЮК. В противоположной ситуации, когда за счет уменьшения поступающей солнечной энергии ТПО в западной части океана снижается, происходит уменьшение градиента ТПО восток-запад, что обуславливает преобладание теплой фазы ЭНЮК – Эль-Ниньо. Такие возмущения тихоокеанской атмосферно-океанической системы, которая стремится скомпенсировать энергетический дисбаланс, возникающий из-за

вариаций солнечной активности, могут влиять на климат в других частях Земного шара через хорошо известные дальние связи ЭНЮК [153].

В работе [100] был проведен совместный анализ палеоданных за последние 850 лет и данных реанализа ТПО за последние 150 лет. В качестве палеоданных, отражающих ТПО, использовались различные климатические прокси – тропические кораллы, годовые кольца деревьев, ледяные керны, донные отложения. На интервале времени, охватывающем последние 150 лет, строились модели связи палеоданных с индексами Niño3.4, определяемыми по трем базам данных реанализа – ERSSTv3, HadSST2i и Kaplan SST. Далее, с помощью построенных моделей реконструировались индексы Niño3.4 на интервале с 1150 по 1995 год с шагом по времени 1 год, соответствующие трем используемым базам (рисунок 3.8, верхняя панель). Для представления сигнала солнечной активности в этой работе использовалась реконструкция ее индекса (рисунок 3.8, средняя панель), сделанная в работе [154] на основе сопоставления данных о космогенных изотопах ^{14}C и ^{10}Be и эмпирической связи между магнитной активностью Солнца и солнечным излучением. На рисунке 3.9 показаны скейлограммы (распределения энергии сигнала по временным масштабам), построенные по временным рядам индекса солнечной активности (рисунок 3.9 (a)) и индексов Niño3.4 (рисунки 3.9 (b)-(d)) с использованием процедуры вейвлет-преобразования. Видно, что в скейлограмме индекса солнечной активности ярко выражены хорошо известные пики на масштабах 70-100 лет (цикл Глейсберга [155]) и около 200 лет (цикл де Врие [156]). В скейлограммах индексов Niño3.4 ожидаемо прослеживаются пики на межгодовых масштабах (2-10 лет), которые становятся более регулярными начиная с конца 19 века. Кроме того, в этих скейлограммах значимо выражен пик на масштабах 200-250 лет. В работе [100] с использованием кросс-вейвлет преобразования и методов оценки фазовой вейвлет-когерентности исследовалась связь данного 200-летнего пика в рядах индекса Niño3.4 с влиянием солнечной активности. В результате был сделан вывод о значимой противофазной зависимости между индексами Niño3.4 и индексом солнечной активности в полосе частот, соответствующей характерному масштабу около 200 лет. Данный результат согласуется с упомянутой выше концепцией океанского динамического термостата о влиянии солнечной активности на ЭНЮК, предложенной в работе [151].

3.3.2 Построение эмпирической модели и анализ ее динамики

Мы предприняли попытку проверить достоверность сделанных в работе [100] выводов путем построения эмпирической модели вида (1.1) по временным рядам индексов Niño3.4 с добавлением форсингов, характеризующих вариации солнечной

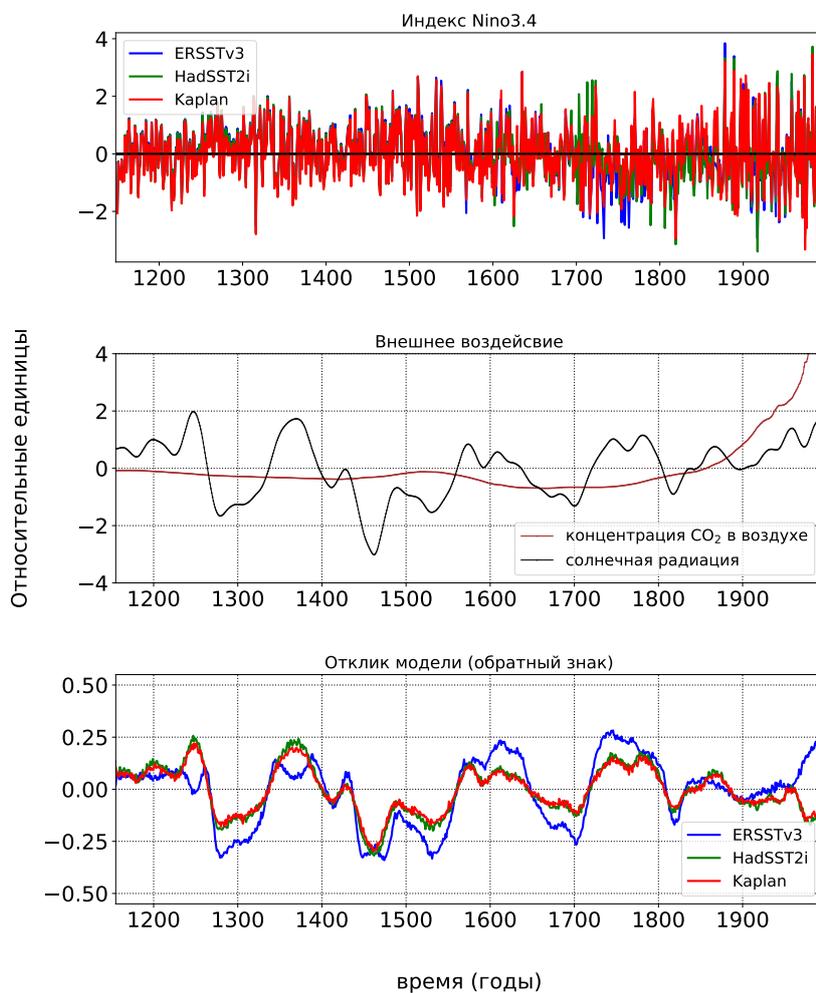


Рис. 3.8: Верхняя панель – временные ряды индексов Nino3.4, реконструированные в работе [100] по палеоклиматическим прокси с использованием трех различных баз данных реанализа ГПО. Средняя панель – временные ряды индекса солнечной активности и общего содержания CO_2 атмосфере, которые использовались в качестве форсингов при построении эмпирических моделей. Нижняя панель – средние по ансамблям из 10000 случайных реализаций эмпирических моделей, построенных по временным рядам индексов Nino3.4 с верхней панели.

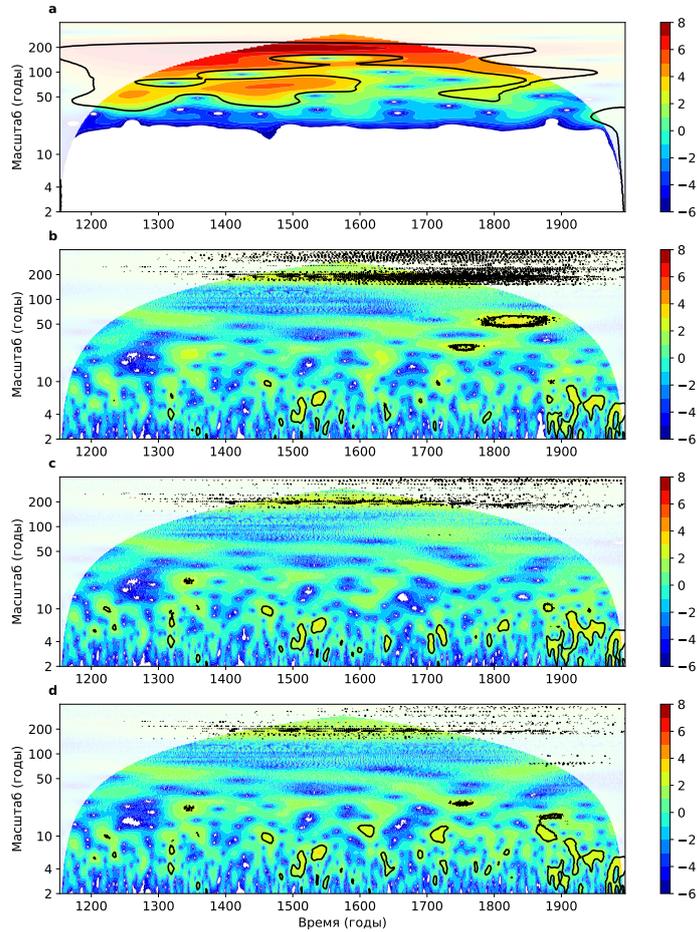


Рис. 3.9: Вейвлет-преобразование индекса солнечной активности (а) и индексов Niño3.4, основанных на различных базах данных – ERSSTv3 (b), HadSST2i (c) и Kaplan (d). В качестве материнской функции при вейвлет-преобразовании использовался комплексный вейвлет Морле с $\omega_0 = 6$ [157]. Черные контуры ограничивают области значимости на уровне 0.1, которые оценивались на основе AR(1)-суррогатного теста, описанного в работе [157]. Цветная шкала представлена в относительных единицах.

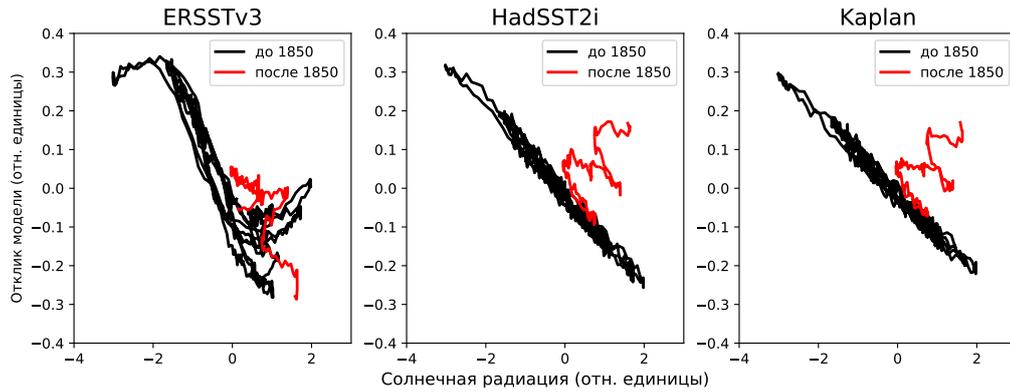


Рис. 3.10: Зависимости среднего по ансамблю случайных реализаций моделей (показанного на нижней панели рисунка 3.8) от индекса солнечной активности. Черным показаны зависимости на интервале времени до 1850 года (доиндустриальный период), красным – после 1850 года. Видно, что характер зависимостей существенно изменяется с началом роста антропогенных эмиссий CO_2 после 1850 года.

активности и общее содержание CO_2 в атмосфере. В качестве данных при построении эмпирической модели мы использовали те же временные ряды индексов Nino3.4 и индекса солнечной активности, что и в работе [100], а также временной ряд исторических значений концентрации CO_2 в атмосфере [158] (рисунок 3.8, средняя панель). Для параметризации детерминированной части модели использовалась нелинейная функция (1.3) в виде искусственной нейронной сети. По каждому из трех имеющихся временных рядов индекса Nino3.4 были построены модели в форме (1.1) с разными комбинациями форсингов. Анализ байесовой оптимальности (1.9) показал, что использование индекса солнечной активности в комбинации с сигналом содержания CO_2 в атмосфере в качестве двумерного форсинга модели является статистически обоснованным. Полученные оптимальные значения структурных параметров модели равны: $l = 3, m = 3$ для модели, обученной по ряду индекса Nino3.4, основанному на реанализе ERSSTv3, и $l = 2, m = 1$ для моделей, обученных по рядам индексов, основанным на двух других реанализах (HadSST2i и Kaplan). Результаты моделирования по каждому из трех рассматриваемых временных рядов индексов, качественно, совпадают. Мы получили, что в динамике построенных эмпирических моделей отклик на сигнал солнечной активности является, фактически, аддитивным по отношению к собственной динамике модели на мелких масштабах. Иными словами, собственная динамика модели не зависит от фазы солнечного цикла. На нижней панели рисунка 3.8 показаны (модельные) временные ряды, усредненные по ансамблям из 10000 случайных реализаций эмпирических моделей длиной 850 лет с исторически-

ми значениями форсингов. Фактически, они характеризуют вынужденные решения полученных моделей, являющиеся откликами на введенные форсинги. Видно, что для всех трех моделей этот отклик содержит сильную составляющую с характерным периодом около 200 лет. Следует отметить разброс средних откликов моделей в конце рассматриваемого интервала, примерно с 1950 года. Это связано с большой неопределенностью отклика моделей на аномальный рост CO_2 , случившийся на относительно коротком временном интервале в результате антропогенных выбросов. Причина такой неопределенности – незначительность вклада этого интервала в ценовую функцию при моделировании 850-летней динамики, а также единичность этого события. Отметим также, что отклик модели, основанной на реанализе ERSSTv3, на солнечную активность является более выраженным по амплитуде. Кроме того, в отличие от двух других баз данных, он является существенно нелинейным – как видно из рисунка 3.10, высокие значения солнечной радиации хотя и понижают значения Nino3.4, но приводят к их большему разбросу. Причины таких различий между реконструкциями на основе различных реанализов предстоит выяснить в дальнейших исследованиях. Тем не менее, рисунки 3.9-3.10 демонстрируют, что все три модели подтверждают вывод о противофазности между двухсотлетними колебаниями солнечной активности и индекса Nino3.4, сделанный в работе [100].

3.4 Заключение

В данной главе диссертации развитые нами методы эмпирического моделирования были применены для исследования динамики явления Эль-Ниньо на межгодовых и вековых временных масштабах. Предложена интерпретация сезонных барьеров предсказемости – весеннего в данных аномалий температуры поверхности и зимнего в данных аномалий теплосодержания верхнего слоя экваториального Тихого океана – на основе выделения сезонных циклов изменчивости этих ключевых характеристик явления Эль-Ниньо. Проанализированы связи между амплитудами этих циклов на межгодовых масштабах. Показано, что структура этих связей с течением времени переходит от линейной в начале 20 века к нелинейной в современном климате. Данный переход, по-видимому, связан с наблюдаемым увеличением числа экстремальных событий Эль-Ниньо во второй половине 20 века.

По временным рядам индексов Nino3.4 за последние 850 лет построена эмпирическая модель вековой изменчивости явления Эль-Ниньо. Показано, что использование сигнала, характеризующего вариации солнечной активности в комбинации с сигналом, отражающим содержание углекислого газа в атмосфере, в качестве двумерного форсинга модели является обоснованным с точки зрения байесова критерия опти-

мальности модели. С помощью построенной модели подтверждена противофазная зависимость индекса солнечной активности с индексом явления Эль-Ниньо Nino3.4 на масштабах порядка 200 лет.

В целом, стоит отметить, что новые результаты, касающиеся динамики явления Эль-Ниньо в широком диапазоне временных масштабов, полученные в диссертации напрямую по данным наблюдений, могут найти свое применение при решении широкого круга актуальных задач. В числе таких задач – сопоставление выходных данных глобальных климатических моделей с реальными данными и определение путей усовершенствования этих моделей для их лучшего соответствия реальному климату.

Заключение

В заключение приведем основные результаты диссертации и положения, выносимые на защиту, а также сформулируем научную новизну и практическую ценность работы.

Основные результаты диссертации

1. Разработан подход к реконструкции динамических систем (ДС) по многомерным данным наблюдений, который включает в себя два основных шага: редукцию исходных данных для определения набора фазовых переменных и построение в получаемом фазовом пространстве модели оператора эволюции в виде стохастической ДС. Рассмотрены различные типы параметризации модели оператора эволюции, допускающие учет внешних воздействий (форсингов) для анализа отклика исследуемой системы на них. Разработан подход к оценке оптимальной сложности эмпирической модели по имеющимся данным в рамках байесова формализма решения обратных задач и с использованием методов проверки статистических гипотез на основе суррогатных данных.
2. Эффективность разработанного подхода продемонстрирована на примере построения эмпирических моделей потребительской активности населения крупнейших городов Российской Федерации в условиях пандемии COVID-19. Получено, что учет в эмпирических моделях в качестве внешнего воздействия информации об интенсивности ограничительных мер, вводимых властями, является статистически обоснованным и позволяет значительно улучшить их предсказательные способности для отдельных категорий потребительской активности во всех рассмотренных городах.
3. На основе анализа данных давления на уровне моря (ДУМ) определен индекс, который отражает состояние субтропической атмосферной циркуляции в зимне-весенний сезон (февраль-март) и имеет статистически значимые длинные корреляции с колебанием Эль-Ниньо в предстоящий период с июня текущего года по май следующего года. Показано, что учет найденного предиктора в

качестве форсинга в линейной эмпирической модели индекса явления Эль-Ниньо Niño3.4 позволяет добиться статистически значимого улучшения ее предсказательной способности на масштабах от нескольких месяцев до одного года, в том числе при прогнозе через весенний барьер предсказуемости.

4. Проведен совместный анализ пространственно распределенных данных температуры поверхности (ТПО) и теплосодержания верхнего слоя (ТО) экваториального Тихого океана. По данным ТПО выявлен индекс, который описывает связанную с Эль-Ниньо изменчивость экваториального Тихого океана и имеет высокую корреляцию с традиционным индексом Niño3.4. По данным ТО определен индекс, который может характеризовать процессы накопления и разгрузки подповерхностных теплых вод в центральной и западной части Тихого океана, которые являются ключевой составляющей цикла колебания Эль-Ниньо. Установлено, что, в отличие от традиционных индексов Эль-Ниньо на основе ТПО найденный индекс ТО характеризуется зимним барьером предсказуемости.
5. Проанализированы предсказательные способности индивидуальных и совместных эмпирических моделей ТПО и ТО, как линейных так и нелинейных. Показано, что совместные модели, учитывающие найденные индексы как равноценные динамические переменные, имеют существенно лучшую предсказательную способность по сравнению с моделями основанными только на ТПО (ТО) и позволяют преодолеть сезонные (весенний в индексе ТПО и зимний в индексе ТО) барьеры предсказуемости. Установлено, что прогностические способности совместных линейной и нелинейной моделей практически не отличаются, что свидетельствует о том, что учет нелинейности в эмпирических прогностических моделях явления Эль-Ниньо не играет роли на внутригодовых масштабах.
6. На основе найденных индексов проанализированы связи между амплитудами сезонных циклов ТПО и ТО. Показано, что структура этих связей с течением времени переходит от линейной в начале 20 века к нелинейной в современном климате, что может быть связано с наблюдаемым увеличением числа экстремальных событий Эль-Ниньо во второй половине 20 века.
7. Построена эмпирическая модель вековой изменчивости явления Эль-Ниньо. Установлено, что использование сигнала, характеризующего вариации солнечной активности в комбинации с сигналом, отражающим содержание углекислого газа в атмосфере, в качестве двумерного форсинга модели является обоснованным с точки зрения байесова критерия оптимальности. С помощью построенной модели

подтверждена противофазная зависимость индекса солнечной активности с индексом явления Эль-Ниньо Niño3.4 на масштабах порядка 200 лет.

Положения, выносимые на защиту

1. Разработанные в диссертации методы позволяют восстанавливать эмпирические модели оператора эволюции распределенных систем, динамика которых протекает в присутствии внешних воздействий, по многомерным данным наблюдений.
2. Учет в линейной эмпирической модели, построенной по индексу явления Эль-Ниньо Niño3.4, предиктора, отражающего изменчивость крупномасштабной атмосферной циркуляции в субтропических широтах северного Тихого океана, позволяет добиться статистически значимого улучшения ее предсказательной способности на масштабах от нескольких месяцев до одного года, в том числе при прогнозе через весенний барьер предсказуемости.
3. Эмпирические модели явления Эль-Ниньо, учитывающие предикторы, отражающие изменчивость температуры поверхности (ТПО) и теплосодержания (ТО) верхнего слоя тропического Тихого океана, как равноценные динамические переменные, имеют существенно лучшую предсказательную способность по сравнению с моделями основанными только на ТПО (ТО) и позволяют преодолеть сезонные барьеры предсказуемости как ТПО, так и ТО.
4. Характер зависимостей между амплитудами сезонных циклов изменчивости температуры поверхности и теплосодержания верхнего слоя тропического Тихого океана переходит от линейного в начале 20 века к нелинейному в современном климате.
5. Эмпирическая модель вековой изменчивости явления Эль-Ниньо подтверждает гипотезу о противофазной зависимости индекса явления Эль-Ниньо Niño3.4 по отношению к индексу солнечной активности на масштабах порядка 200 лет.
6. Учет информации об интенсивности ограничительных мер, вводимых властями, в эмпирических моделях потребительской активности населения крупнейших городов Российской Федерации в условиях пандемии COVID-19 является обоснованным с точки зрения байесова критерия оптимальности и позволяет статистически значимо улучшить предсказательные способности этих моделей.

Научная новизна диссертационной работы состоит в следующем:

1. Предложен универсальный подход к реконструкции динамических систем по многомерным наблюдаемым данным, основанный на построении оптимальных редуцированных стохастических моделей оператора эволюции в рамках байесова формализма решения обратных задач и с использованием методов проверки статистических гипотез на основе суррогатных данных. Возможности подхода продемонстрированы на примере построения эмпирических прогностических моделей потребительской активности населения Российской Федерации в условиях пандемии COVID-19. На основе подхода построены эмпирические модели явления Эль-Ниньо в различных диапазонах временных масштабов.
2. Выявлены новые атмосферные и океанические предикторы, учет которых в эмпирических моделях позволяет преодолеть сезонные барьеры предсказуемости характеристик явления Эль-Ниньо и существенно улучшить качество их прогноза на внутригодовых масштабах.
3. Впервые продемонстрировано, что межгодовые зависимости между аномалиями температуры поверхности и теплосодержания верхнего слоя тропического Тихого океана эволюционируют с течением времени и переходят от линейных в начале 20 века к нелинейным в современном климате.

Практическая ценность работы обусловлена тем, что (1) при активном участии автора диссертации была создана прогностическая модель явления Эль-Ниньо, которая в июне 2019 года была включена (под названием «IAP NN») в список ведущих мировых моделей Эль-Ниньо, прогноз которых публикуется институтом физики Земли при Колумбийском университете (США). Планируется дальнейшее усовершенствование этой модели с учетом полученных в диссертации результатов; (2) построенная в первой главе диссертации прогностическая модель потребительской активности населения Российской Федерации в условиях пандемии COVID-19 в настоящий момент тестируется в качестве дополнительного блока для прогноза социальной активности в агентной модели развития эпидемий, разработанной во Всероссийском научно-исследовательском институте технической физики имени академика Е. И. Забабахина (ВНИИТФ) [119]. Ожидается, что такое дополнение модели ВНИИТФ позволит улучшить ее прогностические возможности.

Личный вклад автора. Все результаты, представленные в диссертации, получены либо лично автором, либо при его непосредственном участии. В частности, при построении эмпирической модели потребительской активности населения в условиях пандемии COVID-19, автором было предложено учитывать в ней в качестве внешнего воздействия информацию об интенсивности вводимых властями ограничительных

мерах и найдены индексы, отражающие эти ограничения. Исследования данных теплосодержания тропического Тихого океана с целью учета содержащейся в них информации в эмпирической модели явления Эль-Ниньо были инициированы и проведены автором. Автор самостоятельно выполнил большинство присутствующих в работе численных расчетов, а также построил соответствующие рисунки и графики.

Автор искренне благодарен своему научному руководителю Дмитрию Николаевичу Мухину за ценные советы, полезные обсуждения, поощрение инициативы при выборе и решении научных задач, заведующему отделом 240 ИПФ РАН, профессору Александру Марковичу Фейгину за постоянное внимание и всестороннюю поддержку, а также своим старшим коллегам – Андрею Гаврилову и Евгению Лоскутову – за активное и плодотворное сотрудничество.

Приложение

Функция правдоподобия эмпирической модели

С учетом уравнения (1.1) функция правдоподобия модели $P(\mathbf{Y}|\boldsymbol{\mu}_f, \boldsymbol{\mu}_{\widehat{\mathbf{g}}}, H_i)$ факторизуется следующим образом:

$$P(\mathbf{Y}|\boldsymbol{\mu}_f, \boldsymbol{\mu}_{\widehat{\mathbf{g}}}, H_i) = \prod_{n=1}^l P(\mathbf{y}_n|H_i) \times \prod_{n=l+1}^N P(\mathbf{y}_n|\mathbf{y}_{n-1}, \dots, \mathbf{y}_{n-l}, \boldsymbol{\mu}_f, \boldsymbol{\mu}_{\widehat{\mathbf{g}}}, H_i). \quad (1)$$

Первая группа сомножителей в (1) определяется априорными ограничениями на вектора $\mathbf{y}_1, \dots, \mathbf{y}_l$, задающие начальное состояние модели. Для того, чтобы определить вид каждого из однотипных сомножителей во второй группе заметим, что при заданной сложности модели H_i и фиксированных значениях параметров $\mathbf{y}_{n-1}, \dots, \mathbf{y}_{n-l}, \boldsymbol{\mu}_f, \boldsymbol{\mu}_{\widehat{\mathbf{g}}}$ выражение (1.1) определяет случайный вектор \mathbf{y}_n как функцию случайного вектора $\boldsymbol{\eta}_n = \widehat{\mathbf{g}} \cdot \boldsymbol{\xi}_n$, который имеет d -мерное нормальное распределение (см. раздел 1.2.1 первой главы диссертации) с плотностью вероятности (ПВ): $P_{\mathcal{N}}(\boldsymbol{\eta}_n, \widehat{\mathbf{g}}\widehat{\mathbf{g}}^T)$ где $P_{\mathcal{N}}(\mathbf{u}, \widehat{\Sigma}) := \frac{1}{\sqrt{(2\pi)^d |\widehat{\Sigma}|}} \exp\left(-\frac{1}{2}\mathbf{u}^T \widehat{\Sigma}^{-1} \mathbf{u}\right)$. С учетом этого, условная ПВ случайного вектора \mathbf{y}_n выражается как (см., например, формулу 6.5 в книге [159]):

$$P(\mathbf{y}_n|\mathbf{y}_{n-1}, \dots, \mathbf{y}_{n-l}, \boldsymbol{\mu}_f, \boldsymbol{\mu}_{\widehat{\mathbf{g}}}, H_i) = P_{\mathcal{N}}(\mathbf{y}_n - \mathbf{f}(\mathbf{y}_{n-1}, \dots, \mathbf{y}_{n-l}), \widehat{\mathbf{g}}\widehat{\mathbf{g}}^T). \quad (2)$$

Подстановка (2) в (1) с учетом гауссовых априорных ПВ на начальные вектора $\mathbf{y}_1, \dots, \mathbf{y}_l$ дает выражение (1.7) для правдоподобия эмпирической модели, приведенное в разделе 1.3 первой главы диссертации.

Как было отмечено в разделе 1.3, априорная ПВ $P(\boldsymbol{\mu}_f, \boldsymbol{\mu}_{\widehat{\mathbf{g}}}|H_i)$ представляет собой произведение гауссовых ПВ для каждого из параметров детерминированной части модели и константного распределения для параметров ее стохастической части:

$$P(\boldsymbol{\mu}_f, \boldsymbol{\mu}_{\widehat{\mathbf{g}}}|H_i) = const \cdot \prod_{i=1}^{M_f} P_{\mathcal{N}}(\mu_i, \sigma_{\mu_i}^2), \quad (3)$$

где μ_i – компоненты вектора $\boldsymbol{\mu}_f \in \mathbb{R}^{M_f}$, в котором собраны параметры детерминированной части модели, σ_{μ_i} – дисперсии гауссовых ПВ каждого параметра.

Подставив выражения (1) и (3) в (1.8) и взяв логарифм от левой и правой частей получим:

$$\begin{aligned}
\Psi_{H_i}(\boldsymbol{\mu}_f, \boldsymbol{\mu}_{\widehat{\mathbf{g}}}) &= \frac{1}{2} \sum_{n=1}^l \sum_{k=1}^d (\ln 2\pi + y_{k,n}^2) + \frac{N-l}{2} \ln |\widehat{\mathbf{g}}\widehat{\mathbf{g}}^T| + \\
&\frac{1}{2} \sum_{n=l+1}^N (\mathbf{y}_n - \mathbf{f}(\mathbf{y}_{n-1}, \dots, \mathbf{y}_{n-l}))^T (\widehat{\mathbf{g}}\widehat{\mathbf{g}}^T)^{-1} (\mathbf{y}_n - \mathbf{f}(\mathbf{y}_{n-1}, \dots, \mathbf{y}_{n-l})) + \\
&\frac{1}{2} \sum_{i=1}^{M_f} (\ln 2\pi \sigma_{\mu_i}^2 + \frac{\mu_i^2}{\sigma_{\mu_i}^2}) + \ln \text{const}.
\end{aligned} \tag{4}$$

Функция (4) представляет собой не что иное, как логарифмическую функцию правдоподобия [159] (взятую с обратным знаком) параметров эмпирической модели, дополненную L2-регуляризацией [160] за счет априорных ограничений, задаваемых ПВ (3).

Список публикаций по теме диссертации

- [A1] *Gavrilov A.S., Seleznev A.F., Mukhin D.N., Loskutov E.M., Feigin A.M., Kurths J.* Linear dynamical modes as new variables for data-driven ENSO forecast // *Climate Dynamics*. — 2019. — feb. — Vol. 52, no. 3-4. — Pp. 2199–2216.
- [A2] *Селезнев А.Ф., Гаврилов А.С., Мухин Д.Н., Лоскутов Е.М., Фейгин А.М.* Моделирование сложных систем по наблюдаемым данным на основе рекуррентных искусственных нейонных сетей // *Известия ВУЗов. Радиофизика*. — 2018. — Vol. 61, no. 12. — Pp. 1005–1021.
- [A3] *Seleznev Aleksei, Mukhin Dmitry, Gavrilov Andrey, Loskutov Evgeny, Feigin Alexander.* Bayesian framework for simulation of dynamical systems from multidimensional data using recurrent neural network // *Chaos: An Interdisciplinary Journal of Nonlinear Science*. — 2019. — Vol. 29, no. 12. — P. 123115.
- [A4] *Mukhin Dmitry, Gavrilov Andrey, Seleznev Aleksei, Buyanova Maria.* An Atmospheric Signal Lowering the Spring Predictability Barrier in Statistical ENSO Forecasts // *Geophysical Research Letters*. — 2021. — Vol. 48, no. 6. — Pp. 1–10.
- [A5] *Мухин Д. Н. Селезнев А. Ф. Гаврилов А. С. Фейгин А. М.* Оптимальные эмпирические модели динамических систем с внешними воздействиями: общий подход и примеры из климата // *Известия ВУЗов. Прикладная нелинейная динамика*. — 2021. — Vol. 29, no. 4. — Pp. 571–602.
- [A6] *Seleznev Aleksei, Mukhin Dmitry.* Improving statistical prediction and revealing nonlinearity of ENSO using observations of ocean heat content in the tropical Pacific // *Climate Dynamics*. — 2021. — no. 0123456789.
- [A7] *Селезнев А.Ф., Гаврилов А.С., Мухин Д.Н., Лоскутов Е.М.* Программа для расчета прогностической модели оператора эволюции в форме комплекснозначной

искусственной нейронной сети. Свидетельство о государственной регистрации программ для ЭВМ №2016661423. Зарегистрировано в реестре программ для ЭВМ 10 октября 2016 г.

- [A8] *Селезнев А.Ф., Гаврилов А.С., Мухин Д.Н., Лоскутов Е.М.* Программа для расчета Байесовой обоснованности модели оператора эволюции в форме комплекснозначной искусственной нейронной сети. Свидетельство о государственной регистрации программ для ЭВМ №2017660391. Зарегистрировано в реестре программ для ЭВМ 21 сентября 2017 г.
- [A9] *Гаврилов А.С., Мухин Д.Н., Лоскутов Е.М., Селезнев А.Ф.* Программа для расчета прогностической эмпирической модели оператора эволюции по многомерному временному ряду с учетом его гладкости. Свидетельство о государственной регистрации программ для ЭВМ №2017660485. Зарегистрировано в реестре программ для ЭВМ 22 сентября 2017 г.
- [A10] *Селезнев А.Ф., Гаврилов А.С.* Программа для построения байесовой линейной стохастической модели оператора эволюции с периодическими коэффициентами. Свидетельство о государственной регистрации программ для ЭВМ №2021610487. Зарегистрировано в реестре программ для ЭВМ 14 января 2021 г.
- [A11] *Seleznev A.F., Gavrilo A.S., Mukhin D.N.* Empirical complex-valued ANN-based model for ENSO forecast // Proceedings: Frontiers of nonlinear physics (FNP-2016). — 2016. — Pp. 184–185.
- [A12] *Фейгин А.М., Гаврилов А.С., Лоскутов Е.М., Мухин Д.Н., Селезнев А.Ф.* Нелинейные динамические моды: метод эмпирической реконструкции сложных систем // <<Нелинейные волны' - 2018>>. — Нижний Новгород: ИПФ РАН, 2019. — Pp. 193–224.
- [A13] *Селезнев А.Ф., Мухин Д.Н.* Оптимизация построения фазового пространства динамической системы по распределенным временным рядам. // Тезисы докладов <<Научная студенческая конференция Высшей школы общей и прикладной физики ННГУ >>, Нижний Новгород, 28-29 мая 2015. — 2015.
- [A14] *Seleznev A.F., Mukhin D.N.* Constructing the reduced phase space from climate variability: complex-valued spatio-temporal data decomposition / A.F. Seleznev and D.N. Mukhin // Abstract for International Workshop <<Analysis of dynamic networks and data driven modelling of the climate>>, 12-14 October 2015, Potsdam, Germany. — 2015.

- [A15] *Селезнев А.Ф.* Эмпирическое моделирование динамики Эль-Ниньо с использованием комплекснозначных искусственных нейронных сетей. // Тезисы научной школы <<Нелинейные волны – 2016>>, Нижний Новгород, 27 февраля – 4 марта 2016. — 2016.
- [A16] *Seleznev A.F., Gavrilov A.S., Mukhin D.N.* Empirical modeling ENSO dynamics with complex-valued artificial neural networks // European Geosciences Union General Assembly, 17-22 April 2016, Vienna, Austria. — 2016.
- [A17] *Seleznev A.S.* ANN-based empirical model for ENSO forecast // Abstract for Science of the future, Kazan, Russia, 20-23 September 2016. — 2016.
- [A18] *Селезнев А.Ф., Гаврилов А.С., Мухин Д.Н.* Эмпирическое моделирование динамики Эль-Ниньо // Тезисы 20-й школы-конференции молодых ученых <<Состав Атмосферы. Атмосферное электричество. Климатические процессы>>, Нижний Новгород, 24-26 мая 2016. — 2016.
- [A19] *Селезнев А.Ф., Гаврилов А.С., Мухин Д.Н., Фейгин А.М.* Построение динамических моделей по высокоразмерным временным рядам: новая методология и приложение к прогнозу климатической изменчивости // Тезисы докладов XXV сессии РАН по нелинейной динамике, Москва, 19-20 декабря 2016. — 2016.
- [A20] *Seleznev A.F., Gavrilov A.S., Mukhin D.N.* Construction of reduced dynamical models from high-dimensional time series // Book of abstracts. Dynamic Days, MD – Maryland, USA, 6-10 Jan 2017. — 2017.
- [A21] *Feigin A.M., Mukhin D.N., Loskutov E.M., Gavrilov A.S., Seleznev A.F.* Empirical approach to modeling & prognosis of climate systems // Abstract for Topical Problems of Nonlinear Wave Physics (NWP-2017), Moscow - St.-Petersburg, Russia, 22-28 July 2017. — 2017.
- [A22] *Gavrilov A.S., Seleznev A.F., Mukhin D.N., Loskutov E.M., Feigin A.M.* Reduced nonlinear data-driven prognostic climate model construction // Abstract: Topical Problems of Nonlinear Wave Physics (NWP-2017). — 2017.
- [A23] *Gavrilov A.S., Seleznev A.F., Mukhin D.N., Loskutov E.M., Feigin A.M.* Construction of reduced empirical prognostic model by climate data // Abstract for Fourth International Conference on Earth System Modelling, Hamburg, Germany, 28 August - 1 September 2017. — 2017.

- [A24] *Селезнев А.Ф., Гаврилов А.С., Мухин Д.Н., Лоскутов Е.М., Фейгин А.М.* Эмпирическое прогнозирование изменчивости Эль-Ниньо с использованием метода нелинейных динамических мод // Тезисы XXI Всероссийской школы-конференции молодых ученых <<Состав атмосферы. Атмосферное электричество. Климатические процессы>>, 6-10 июня 2017, Борок. — 2017.
- [A25] *Feigin A.M., Mukhin D.N., Gavrilov A.S., Seleznev A.F., Loskutov E.M.* Data-driven Nonlinear Dynamical Models for Forecast of Climate Variability // Abstract for JpGU-AGU Joint Meeting 2017, Makuhari Messe, Japan, 20-25 May 2017. — 2017.
- [A26] *Feigin A.M., Mukhin D.N., Gavrilov A.S., Seleznev A.F., Loskutov E.M.* Nonlinear Dynamical Modes as a Basis for Short-Term Forecast of Climate Variability // Abstract NG41A-0116, AGU Fall meeting, New Orleans, 11-15 December 2017. — 2017.
- [A27] *Feigin A.M., Gavrilov A.S., Seleznev A.F., Mukhin D.N., Loskutov E.M.* Linear dynamical modes: new variables for empirical prognostic model of ENSO variability // Abstract EGU2018-7611. European Geosciences Union General Assembly, 8-13 April 2018, Vienna, Austria. — 2018.
- [A28] *Gavrilov A.S., Seleznev A.F., Mukhin D.N., Loskutov E.M., Feigin A.M., Kurths J.* Empirical modeling of ENSO using linear dynamical mode decomposition // Abstract for 32nd IUGG Conference on Mathematical Geophysics. 23-28 June 2018, Nizhny Novgorod, Russia. — 2018.
- [A29] *Гаврилов А.С., Селезнев А.Ф., Мухин Д.Н., Лоскутов Е.М., Фейгин А.М.* Новые переменные для эмпирической прогностической модели Эль-Ниньо // Тезисы научной школы <<Нелинейные волны – 2018>>, Нижний Новгород, 26 февраля – 4 марта 2018. — 2018. — Pp. 27–29.
- [A30] *Seleznev Aleksei, Mukhin Dmitry, Gavrilov Andrey, Feigin Alexander.* Data-driven modeling decadal-to-centennial ENSO variability and its response to external forcing // EGU General Assembly Conference Abstracts. — EGU General Assembly Conference Abstracts. — 2021. — Pp. EGU21–8370.
- [A31] *Feigin Alexander, Mukhin Dmitry, Gavrilov Andrey, Seleznev Aleksei, Buyanova Maria.* An atmospheric forcing extending ENSO forecast horizon using statistical models // EGU General Assembly Conference Abstracts. — EGU General Assembly Conference Abstracts. — 2021. — Pp. EGU21–2813.

- [А32] *Мухин Д.Н., Селезнев А.Ф., Вдовин В.В., Лоскутов Е.М., Гаврилов А.С., Фейгин А.М.* ЭМПИРИЧЕСКОЕ МОДЕЛИРОВАНИЕ АКТИВНОСТИ ОБЩЕСТВА В УСЛОВИЯХ ПАНДЕМИИ COVID - 19: ВЫБОР ДИНАМИЧЕСКИХ ПЕРЕМЕННЫХ И ФОРСИНГОВ. // Забабахинские научные чтения, Снежинск, 27 сентября – 1 октября 2021. — 2021.
- [А33] *Лоскутов Е.М., Селезнев А.Ф., Вдовин В.В., Мухин Д.Н., Гаврилов А.С., Фейгин А.М.* ЭМПИРИЧЕСКОЕ МОДЕЛИРОВАНИЕ АКТИВНОСТИ ОБЩЕСТВА В УСЛОВИЯХ ПАНДЕМИИ COVID - 19: ПОСТРОЕНИЕ ПРОГНОСТИЧЕСКОЙ МОДЕЛИ. // Забабахинские научные чтения, Снежинск, 27 сентября – 1 октября 2021. — 2021.

Литература

- [1] *Анищенко В.С., Вадивасова Т.Е. Астахов В.В.* Нелинейная динамика хаотических и стохастических систем. — Саратов: Изд. Саратовского университета, 1999. — 368 С.
- [2] *Packard N H, Crutchfield J P, Farmer J D, Shaw R S.* Geometry from a Time Series // *Phys. Rev. Lett.* — 1980. — Vol. 45, no. 9. — Pp. 712–716.
- [3] *Takens Floris.* Detecting strange attractors in turbulence // *Dynamical Systems and Turbulence*, Warwick 1980. — Springer Berlin Heidelberg. — Pp. 366–381.
- [4] *Sauer Tim, Yorke James A, Casdagli Martin.* Embedology // *Journal of Statistical Physics.* — 1991. — Vol. 65, no. 3. — Pp. 579–616.
- [5] *Stark J, Broomhead D S, Davies M E, Huke J.* Takens embedding theorems for forced and stochastic systems // *Nonlinear Analysis: Theory, Methods and Applications.* — 1997. — Vol. 30, no. 8. — Pp. 5303–5314.
- [6] *Robinson James C.* A topological delay embedding theorem for infinite-dimensional dynamical systems // *Nonlinearity.* — 2005. — Vol. 18, no. 5. — Pp. 2135–2143.
- [7] *Casdagli Martin, Eubank Stephen, Farmer J. Doyme, Gibson John.* State space reconstruction in the presence of noise // *Physica D: Nonlinear Phenomena.* — 1991. — Vol. 51, no. 1-3. — Pp. 52–98.
- [8] *Deyle Ethan R, Sugihara George, Deyle E R, Gibson J, Ramanathan A.* Generalized Theorems for Nonlinear State Space Reconstruction // *PLoS ONE.* — 2011. — Vol. 6, no. 3. — P. e18295.
- [9] *Farmer J Doyme, Sidorowich John J.* Predicting chaotic time series // *Physical Review Letters.* — 1987. — Vol. 59, no. 8. — Pp. 845–848.

- [10] *Cremers J, Hubler A.* Construction of Differential Equations from Experimental Data // *Zeitschrift für Naturforschung A.* — 1987. — Vol. 42, no. 8. — Pp. 797–802.
- [11] *Crutchfield James P, McNamara Bruce S.* Equations of Motion from a Data Series // *Complex Systems.* — 1987. — Vol. 1.
- [12] *Gouesbet G, Letellier C.* Global vector-field reconstruction by using a multivariate polynomial L2 approximation on nets // *Physical Review E.* — 1994. — Vol. 49, no. 6. — Pp. 4955–4972.
- [13] *Anishchenko Vadim, Pavlov Alexey, Janson Natalia.* Global Reconstruction in the Presence of a priori Information // *Chaos Solitons and Fractals.* — 1998. — Vol. 9. — Pp. 1267–1278.
- [14] *Павлов А Н, Янсон Н Б, Анищенко В С.* Реконструкция динамических систем // *Радиотехника и электроника.* — 1999. — Т. 44, №. 9. — С. 1075-1092.
- [15] *Schelter Björn, Mader Malenka, Mader Wolfgang, Sommerlade Linda, Platt Bettina, Lai Ying Cheng, Grebogi Celso, Thiel Marco.* Overarching framework for data-based modelling // *Epl.* — 2014. — Vol. 105, no. 3.
- [16] *Shandilya Srinivas Gorur, Timme Marc.* Inferring network topology from complex dynamics // *New Journal of Physics.* — 2011. — Vol. 13.
- [17] *Wang Wen Xu, Yang Rui, Lai Ying Cheng, Kovanis Vassilios, Grebogi Celso.* Predicting catastrophes in nonlinear dynamical systems by compressive sensing // *Physical Review Letters.* — 2011. — Vol. 106, no. 15. — Pp. 1–4.
- [18] *Baake Ellen, Baake Michael, Bock Hans, M. Briggs K.* Fitting Ordinary Differential Equations to Chaotic Data // *Physical Review A.* — 2000. — Vol. 45.
- [19] *Bezruchko B P, Smirnov Dmitry, Sysoev Ilya.* Identification of chaotic systems with hidden variables (modified Bock’s algorithm) // *Chaos, Solitons and Fractals.* — 2006. — Vol. 29. — Pp. 82–90.
- [20] *Gorodetskyi V, Osadchuk M.* Analytic reconstruction of some dynamical systems // *Physics Letters, Section A: General, Atomic and Solid State Physics.* — 2013. — Vol. 377, no. 9. — Pp. 703–713.
- [21] *Mukhin D. N., Feigin A. M., Loskutov E. M., Molkov Ya I.* Modified Bayesian approach for the reconstruction of dynamical systems from time series // *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics.* — 2006. — Vol. 73, no. 3. — Pp. 1–7.

- [22] *Анищенко В С, Павлов А Н, Янсон Н Б.* Реконструкция динамических систем в приложении к решению задачи защиты информации // *ЖТФ.* — 1998. — Вып. 68, №. 12. — С. 1–8.
- [23] *Bezruchko Boris P, Smirnov Dmitry A.* Constructing nonautonomous differential equations from experimental time series // *Physical Review E - Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics.* — 2001. — Vol. 63, no. 1. — Pp. 1–7.
- [24] *Смирнов Д.А., Сысоев И.В., Селезнёв Е.П., Безручко Б.П.* Реконструкция моделей неавтономных систем с дискретным спектром воздействия // *Письма в ЖТФ.* — 2003. — Том 29. — Вып. 19. — С. 69-76.
- [25] *Molkov Y. I., Mukhin D. N., Loskutov E. M., Timushev R. I., Feigin A. M.* Prognosis of qualitative system behavior by noisy, nonstationary, chaotic time series // *Physical Review E.* — 2011. — Vol. 84, no. 3. — Pp. 1–7.
- [26] *Molkov Y I, Loskutov E M, Mukhin D N, Feigin A M.* Random dynamical models from time series // *Physical Review E.* — 2012. — mar. — Vol. 85, no. 3. — P. 36216.
- [27] *Ponomarenko V I, Prokhorov M D.* Extracting information masked by the chaotic signal of a time-delay system // *Phys. Rev. E.* — 2002. — Vol. 66, no. 2. — P. 26215.
- [28] *Sysoev I V, Prokhorov M D, Ponomarenko V I, Bezruchko B P.* Reconstruction of ensembles of coupled time-delay systems from time series // *Phys. Rev. E.* — 2014. — Vol. 89, no. 6. — P. 62911.
- [29] *Sysoev I V, Ponomarenko V I, Kulminskiy D D, Prokhorov M D.* Recovery of couplings and parameters of elements in networks of time-delay systems from time series // *Phys. Rev. E.* — 2016. — Vol. 94, no. 5. — P. 52207.
- [30] *Brunton Steven L., Proctor Joshua L., Kutz J. Nathan.* Discovering governing equations from data: Sparse identification of nonlinear dynamical systems. — 2015. — Vol. 113, no. 15. — Pp. 3932–3937.
- [31] *Mangan Niall M., Brunton Steven L., Proctor Joshua L., Kutz J. Nathan.* Inferring Biological Networks by Sparse Identification of Nonlinear Dynamics // *IEEE Transactions on Molecular, Biological, and Multi-Scale Communications.* — 2016. — Vol. 2, no. 1. — Pp. 52–63.
- [32] *Han Xiao, Shen Zhesi, Wang Wen-Xu, Di Zengru.* Robust Reconstruction of Complex Networks from Sparse Data // *Phys. Rev. Lett.* — 2015. — Vol. 114, no. 2. — P. 28701.

- [33] *Pikovsky A.* Reconstruction of a neural network from a time series of firing rates // *Physical Review E*. — 2016. — Vol. 93, no. 6. — Pp. 1–4.
- [34] *Sysoev Ilya V, Ponomarenko Vladimir I, Pikovsky Arkady.* Reconstruction of coupling architecture of neural field networks from vector time series // *Communications in Nonlinear Science and Numerical Simulation*. — 2018. — Pp. 342–351.
- [35] *Smirnov Dmitry A, Bezruchko Boris P, Seleznev Yevgeny P.* Choice of dynamical variables for global reconstruction of model equations from time series // *Physical Review E*. — 2002. — Vol. 65, no. 2. — P. 26205.
- [36] *Arnold Ludwig.* Random Dynamical Systems. Springer Monographs in Mathematics. — Berlin: Springer-Verlag, 1998.
- [37] *Berliner L. Mark, Wikle Christopher K., Cressie Noel.* Long-Lead Prediction of Pacific SSTs via Bayesian Dynamic Modeling // *Journal of Climate*. — 2000. — nov. — Vol. 13, no. 22. — Pp. 3953–3968.
- [38] *Penland Cécile, Magorian Theresa.* Prediction of Niño 3 Sea Surface Temperatures Using Linear Inverse Modeling // *Journal of Climate*. — 1993. — Vol. 6, no. 6. — Pp. 1067–1076.
- [39] *Penland C, Matrosova L.* Prediction of tropical Atlantic sea surface temperatures using linear inverse modeling. // *Journal of Climate*. — 1998. — no. 1992. — Pp. 483–496.
- [40] *Penland Cécile, Sardeshmukh Prashant D.* The Optimal Growth of Tropical Sea Surface Temperature Anomalies // *Journal of Climate*. — 1995. — Vol. 8, no. 8. — Pp. 1999–2024. — URL: https://journals.ametsoc.org/view/journals/clim/8/8/1520-0442_1995_008_1999_togots_2_0_co_2.xml.
- [41] *Фейгин А.М., Лоскутов Е М, Мольков Я И, Мухин Д Н.* Прогноз качественного поведения динамической системы по хаотическому временному ряду // *Известия ВУЗов. Радиофизика*. — 2001. — Вып. 44, №. 5-6. — С. 376 – 398.
- [42] *Feigin Alexander M, Molkov Yaroslav I, Mukhin Dmitrii N, Loskutov Eugeni M.* Investigation of nonlinear dynamical properties by the observed complex behaviour as a basis for construction of dynamical models of atmospheric photochemical systems // *Faraday Discussions*. — 2002. — Vol. 120. — Pp. 105–123.
- [43] *Kondrashov D, Kravtsov S, Robertson A W, Ghil M.* A Hierarchy of Data-Based ENSO Models // *Journal of Climate*. — 2005. — Vol. 18, no. 21. — Pp. 4425–4444.

- [44] *Kravtsov S, Kondrashov D, Ghil M.* Multilevel Regression Modeling of Nonlinear Processes: Derivation and Applications to Climatic Variability // *J. Climate.* — 2005. — Vol. 18. — Pp. 4404–4424.
- [45] *Kravtsov S, Kondrashov D, Ghil M.* Empirical model reduction and the modeling hierarchy in climate dynamics // *Stochastic Physics and Climate Modelling* / Ed. by T.N. Palmer, P Williams. — Cambridge Univ. Press, 2009. — Pp. 35–72.
- [46] *Grieger B, Latif M.* Reconstruction of the El Niño attractor with neural networks // *Climate Dyn.* — 1994. — Vol. 10. — Pp. 267–276.
- [47] *Mukhin Dmitry, Loskutov Evgeny, Mukhina Anna, Feigin Alexander, Zaliapin Ilya, Ghil Michael.* Predicting critical transitions in ENSO Models. Part I: Methodology and simple models with memory // *Journal of Climate.* — 2015. — Vol. 28, no. 5. — Pp. 1940–1961.
- [48] *Rossi V., Vila J.-P.* Bayesian Multioutput Feedforward Neural Networks Comparison: A Conjugate Prior Approach // *IEEE Transactions on Neural Networks.* — 2006. — jan. — Vol. 17, no. 1. — Pp. 35–47.
- [49] *Wu Aiming, Hsieh William W., Tang Benyang.* Neural network forecasts of the tropical Pacific sea surface temperatures // *Neural Networks.* — 2006. — mar. — Vol. 19, no. 2. — Pp. 145–154.
- [50] *Zhang G.P., Kline D.M.* Quarterly Time-Series Forecasting With Neural Networks // *IEEE Transactions on Neural Networks.* — 2007. — nov. — Vol. 18, no. 6. — Pp. 1800–1814.
- [51] *Gavrilov Andrey, Loskutov Evgeny, Mukhin Dmitry.* Bayesian optimization of empirical model with state-dependent stochastic forcing // *Chaos, Solitons and Fractals.* — 2017. — Vol. 104. — Pp. 327–337.
- [52] *Jolliffe I T.* Principal Component Analysis. Springer Series in Statistics. — 2nd edition. — New York, NY: Springer New York, 1986. — P. 271.
- [53] *Hannachi A, Jolliffe I T, Stephenson D B.* Empirical orthogonal functions and related techniques in atmospheric science: A review // *International Journal of Climatology.* — 2007. — jul. — Vol. 27, no. 9. — Pp. 1119–1152.
- [54] *Trenberth Kevin E.* The Definition of El Niño // *Bulletin of the American Meteorological Society.* — 1997. — Vol. 78, no. 12. — Pp. 2771–2777.

- [55] *Петросянец М.А., Семенов Е.К., Гущина Д.Ю., Соколкина Е.В., Соколкина Н.Н.* Циркуляция атмосферы в тропиках: Климат и изменчивость. — Москва: МАКС Пресс, 2005. — 670 С.
- [56] *Dijkstra H A.* Nonlinear Physical Oceanography: A Dynamical Systems Approach to the Large Scale Ocean Circulation and El Niño. Atmospheric and Oceanographic Sciences Library. — Springer Netherlands, 2005.
- [57] *Trenberth Kevin E.* El Niño southern oscillation (ENSO) // *Encyclopedia of Ocean Sciences.* — 2019. — no. March. — Pp. 420–432.
- [58] *Alexander Michael A, Bladé Ileana, Newman Matthew, Lanzante John R, Lau Ngarcheung, Scott James D.* The Atmospheric Bridge: The Influence of ENSO Teleconnections on Air–Sea Interaction over the Global Oceans // *Journal of Climate.* — 2002. — Vol. 15, no. 16. — Pp. 2205–2231.
- [59] *Осипов А.М., Гущина Д.Ю.* Эль-Ниньо 2015/2016 гг.: Эволюция, механизмы, сопутствующие удаленные аномалии // *Фундаментальная и прикладная климатология.* — 2018. — Вып. 3. — С. 54–81.
- [60] *Bjerknes J.* Monthly Weather Review Atmospheric Teleconnections From the Equatorial Pacific // *Monthly Weather Review.* — 1969. — Vol. 97, no. 3. — Pp. 163–172.
- [61] *Wyrski Klaus.* El Niño—The Dynamic Response of the Equatorial Pacific Ocean to Atmospheric Forcing // *Journal of Physical Oceanography.* — 1975. — Vol. 5, no. 4. — Pp. 572–584.
- [62] *Wyrski Klaus.* Water displacements in the Pacific and the genesis of El Niño cycles // *Journal of Geophysical Research: Oceans.* — 1985. — jul. — Vol. 90, no. C4. — Pp. 7129–7132.
- [63] *Железнова И В, Гущина Д Ю, Коленникова М А.* ОЦЕНКА ИЗМЕНЕНИЯ ВКЛАДА ОСНОВНЫХ ПРЕДИКТОРОВ В ПРОГНОЗ ЭЛЬ-НИНЬО В ПОСЛЕДНИЕ ДЕСЯТИЛЕТИЯ НА ОСНОВЕ ПРОСТОЙ СТАТИСТИЧЕСКОЙ МОДЕЛИ // *Фундаментальная и прикладная климатология.* — 2020. — №. 1. — С. 42–64.
- [64] *Cane Mark A, Zebiak Stephen E.* A Theory for El Niño and the Southern Oscillation // *Science.* — 1985. — Vol. 228, no. 4703. — Pp. 1085–1087.

- [65] *Jin Fei-Fei*. An Equatorial Ocean Recharge Paradigm for ENSO. Part I: Conceptual Model // *Journal of the Atmospheric Sciences*. — 1997. — Vol. 54, no. 7. — Pp. 811–829.
- [66] *Suarez Max J., Schopf Paul S.* A Delayed Action Oscillator for ENSO // *Journal of the Atmospheric Sciences*. — 1988. — 11. — Vol. 45, no. 21. — Pp. 3283–3287.
- [67] *Galanti Eli, Tziperman Eli*. ENSO’s Phase Locking to the Seasonal Cycle in the Fast-SST, Fast-Wave, and Mixed-Mode Regimes // *Journal of the Atmospheric Sciences*. — 2000. — 09. — Vol. 57, no. 17. — Pp. 2936–2950.
- [68] *Zebiak Stephen E, Cane Mark A.* A Model El Niño-Southern Oscillation // *Monthly Weather Review*. — 1987. — Vol. 115, no. 10. — Pp. 2262–2278.
- [69] *Anderson David L T, McCreary Julian P.* Slowly Propagating Disturbances in a Coupled Ocean-Atmosphere Model // *Journal of Atmospheric Sciences*. — 1985. — Vol. 42, no. 6. — Pp. 615–630.
- [70] *Jin Fei-Fei, Neelin J David*. Modes of Interannual Tropical Ocean–Atmosphere Interaction—a Unified View. Part I: Numerical Results // *Journal of Atmospheric Sciences*. — 1993. — Vol. 50, no. 21. — Pp. 3477–3503.
- [71] *Philander S George, Fedorov Alexey*. Is El Niño Sporadic or Cyclic? // *Annual Review of Earth and Planetary Sciences*. — 2003. — may. — Vol. 31, no. 1. — Pp. 579–594.
- [72] *Fedorov A V, Harper S L, Philander S G, Winter B, Wittenberg A*. How Predictable is El Niño? // *Bulletin of the American Meteorological Society*. — 2003. — Vol. 84, no. 7. — Pp. 911–920.
- [73] *Chen Shangfeng, Wu Renguang, Chen Wen, Yu Bin, Cao Xi*. Genesis of westerly wind bursts over the equatorial western Pacific during the onset of the strong 2015–2016 El Niño // *Atmospheric Science Letters*. — 2016. — jul. — Vol. 17, no. 7. — Pp. 384–391.
- [74] *Hu Shineng, Fedorov Alexey V*. The extreme El Niño of 2015–2016: the role of westerly and easterly wind bursts, and preconditioning by the failed 2014 event // *Climate Dynamics*. — 2019. — Vol. 52, no. 12. — Pp. 7339–7357.
- [75] *Martinez-Villalobos Cristian, Newman Matthew, Vimont Daniel J., Penland Cécile, David Neelin J*. Observed El Niño-La Niña Asymmetry in a Linear Model // *Geophysical Research Letters*. — 2019. — Vol. 46, no. 16. — Pp. 9909–9919.

- [76] *Levine Aaron F Z, Jin Fei Fei.* A simple approach to quantifying the noise–ENSO interaction. Part I: deducing the state-dependency of the windstress forcing using monthly mean data // *Climate Dynamics*. — 2017. — Vol. 48, no. 1. — Pp. 1–18.
- [77] *Zhang Chidong, Gottschalck Jonathan.* SST Anomalies of ENSO and the Madden–Julian Oscillation in the Equatorial Pacific // *Journal of Climate*. — 2002. — Vol. 15, no. 17. — Pp. 2429–2445.
- [78] *Chiodi Andrew M, Harrison D E, Vecchi Gabriel A.* Subseasonal Atmospheric Variability and El Niño Waveguide Warming: Observed Effects of the Madden–Julian Oscillation and Westerly Wind Events // *Journal of Climate*. — 2014. — Vol. 27, no. 10. — Pp. 3619–3642.
- [79] *Puy Martin, Vialard J, Lengaigne M, Guilyardi E.* Modulation of equatorial Pacific westerly/easterly wind events by the Madden–Julian oscillation and convectively-coupled Rossby waves // *Climate Dynamics*. — 2016. — Vol. 46, no. 7. — Pp. 2155–2178.
- [80] *Vimont Daniel J., Wallace John M., Battisti David S.* The seasonal footprinting mechanism in the Pacific: Implications for ENSO // *Journal of Climate*. — 2003. — Vol. 16, no. 16. — Pp. 2668–2675.
- [81] Generation of westerly wind bursts by forcing outside the tropics / Arnold Sullivan, Wenxiu Zhong, Gian Luca Eusebi Borzelli, Tao Geng, Chloe Mackallah, Benjamin Ng, Chi-Cherng Hong, Wenju Cai et al. // *Scientific Reports*. — 2021. — Vol. 11, no. 1. — P. 912.
- [82] *Barnston Anthony G., Tippett Michael K., L’Heureux Michelle L., Li Shuhua, Dewitt David G.* Skill of real-time seasonal ENSO model predictions during 2002–11: Is our capability increasing? // *Bulletin of the American Meteorological Society*. — 2012. — Vol. 93, no. 5. — Pp. 631–651.
- [83] *Wright Peter B.* Persistence of rainfall anomalies in the central Pacific // *Nature*. — 1979. — Vol. 277, no. 5695. — Pp. 371–374.
- [84] *Webster Peter J, Yang Song.* Monsoon and Enso: Selectively Interactive Systems // *Quarterly Journal of the Royal Meteorological Society*. — 1992. — Vol. 118, no. 507. — Pp. 877–926.
- [85] *McPhaden Michael J.* Tropical Pacific Ocean heat content variations and ENSO persistence barriers // *Geophysical Research Letters*. — 2003. — Vol. 30, no. 9.

- [86] Current status of ENSO prediction skill in coupled ocean-atmosphere models / Emilia K. Jin, James L. Kinter, B. Wang, C. K. Park, I. S. Kang, B. P. Kirtman, J. S. Kug, A. Kumar et al. // *Climate Dynamics*. — 2008. — Vol. 31, no. 6. — Pp. 647–664.
- [87] *Clarke Allan J, Van Gorder Stephen*. Improving El Niño prediction using a space-time integration of Indo-Pacific winds and equatorial Pacific upper ocean heat content // *Geophysical Research Letters*. — 2003. — Vol. 30, no. 7.
- [88] *Meinen Christopher S, McPhaden Michael J*. Observations of Warm Water Volume Changes in the Equatorial Pacific and Their Relationship to El Nino and La Nina // *Journal of Climate*. — 2000. — Vol. 13, no. 20. — Pp. 3551–3559.
- [89] *Chen Han Ching, Tseng Yu Heng, Hu Zeng Zhen, Ding Ruiqiang*. Enhancing the ENSO Predictability beyond the Spring Barrier // *Scientific Reports*. — 2020. — Vol. 10, no. 1. — Pp. 1–12.
- [90] *Бышев В. И., Нейман В. Г., Романов Ю А., Серых И. В., Сонечкин Д. М.* О СТАТИСТИЧЕСКОЙ ЗНАЧИМОСТИ И КЛИМАТИЧЕСКОЙ РОЛИ ГЛОБАЛЬНОЙ АТМОСФЕРНОЙ ОСЦИЛЛЯЦИИ // *Океанология*. — 2016. — Вып. 56, №. 2. — С. 165–171.
- [91] *McPhaden M J, Lee T, McClurg D*. El Niño and its relationship to changing background conditions in the tropical Pacific Ocean // *Geophysical Research Letters*. — 2011. — Vol. 38, no. 15.
- [92] *Yeh Sang-Wook, Kug Jong-Seong, Dewitte Boris, Kwon Min-Ho, Kirtman Ben P, Jin Fei-Fei*. El Niño in a changing climate // *Nature*. — 2009. — Vol. 461, no. 7263. — Pp. 511–514.
- [93] *Yeh Sang-Wook, Kirtman Ben P, Kug Jong-Seong, Park Wonsun, Latif Mojib*. Natural variability of the central Pacific El Niño event on multi-centennial timescales 2011. — Vol. 38, no. 2. — P. L02704.
- [94] *Mukhin Dmitry, Gavrillov Andrey, Feigin Alexander, Loskutov Evgeny, Kurths Juergen*. Principal nonlinear dynamical modes of climate variability // *Scientific Reports*. — 2015. — Vol. 5. — P. 15510.
- [95] *Mukhin Dmitry, Gavrillov Andrey, Loskutov Evgeny, Feigin Alexander, Kurths Juergen*. Nonlinear reconstruction of global climate leading modes on decadal scales // *Climate Dynamics*. — 2017. — Vol. 51, — Pp. 2301–2310.

- [96] *Serykh I.~V., Sonechkin D.~M., Byshev V.~I., Neiman V.~G., Romanov Yu. A.* Global Atmospheric Oscillation: An Integrity of ENSO and Extratropical Teleconnections // *Pure and Applied Geophysics*. — 2019. — Vol. 176, no. 8. — Pp. 3737–3755.
- [97] *И.И. Мохов, Д.А. Смирнов.* Исследование взаимного влияния процессов Эль-Ниньо - Южное колебание и Северо-Атлантического и Арктического колебаний // *Известия РАН. Физика атмосферы и океана*. — 2006. — Вып. 42, №. 5. — С. 650–667.
- [98] *Матвеева Т.А., Гущина Д.Ю.* Изменения характеристик центрально-тихоокеанского и восточно-тихоокеанского Эль-Ниньо в условиях потепления климата (по результатам экспериментов RCP 2.6 и RCP 8.5 CMIP5) // *Фундаментальная и прикладная климатология*. — 2017. — Вып. 2, №. 2. — С. 86–110.
- [99] *Гущина Д. Ю., Калиновская М. В., Матвеева Т. А.* Влияние тихоокеанского десятилетнего колебания на характеристики Эль-Ниньо двух типов при возможных изменениях климата // *Метеорология и гидрология*. — 2020. — №. 10. С. 14–28.
- [100] *Emile-Geay Julien, Cobb Kimberly M., Mann Michael E., Wittenberg Andrew T.* Estimating central equatorial pacific SST variability over the past millennium. part II: Reconstructions and implications // *Journal of Climate*. — 2013. — Vol. 26, no. 7. — Pp. 2329–2352.
- [101] *Серых И.В.* О роли Эль-Ниньо — Глобальной атмосферной осцилляции в межгодовой изменчивости гидрометеорологических процессов // *ГИДРОМЕТЕОРОЛОГИЯ И ЭКОЛОГИЯ*. — 2021. — №. 63. — С. 329–370.
- [102] *Мохов И И, Смирнов Д А, Наконечный П И, Козленко С С, Куртс Ю.* Взаимосвязь явлений Эль Ниньо / Южное Колебание и индийского муссона // *Известия Российской академии наук. Физика атмосферы и океана*. — 2012. — Вып. 48, №. 1. — С. 56–66.
- [103] *Кузнецов С.П.* Динамический хаос. — Москва: Физматлит, 2006. — 356 С.
- [104] *Schuster H G, Just Wolfram.* Deterministic Chaos: An Introduction: Fourth Edition. — 2005. — Pp. 1–287.
- [105] *Jaeger Lars, Kantz Holger.* Unbiased reconstruction of the dynamics underlying a noisy chaotic time series // *Chaos: An Interdisciplinary Journal of Nonlinear Science*. — 1996. — Vol. 6, no. 3. — Pp. 440–450.

- [106] GRASSBERGER PETER, SCHREIBER THOMAS, SCHAFFRATH CARSTEN. NONLINEAR TIME SEQUENCE ANALYSIS // *International Journal of Bifurcation and Chaos*. — 1991. — Vol. 01, no. 03. — Pp. 521–547.
- [107] Van Huffel Sabine, Zha Hongyuan. 10 The total least squares problem // *Computational Statistics*. — Elsevier, 1993. — Vol. 9 of *Handbook of Statistics*. — Pp. 377–408.
- [108] Kostelich Eric J. Problems in estimating dynamics from data // *Physica D: Nonlinear Phenomena*. — 1992. — Vol. 58, no. 1. — Pp. 138–152.
- [109] Бутковский О.Я., Кравцов Ю.А. Логунов М.Ю. Анализ погрешностей восстановления параметров нелинейного отображения по зашумленным хаотическим временным рядам. // *Известия ВУЗов. Радиофизика*. — 2002. — Вып. 45, №. 1. — С. 55 – 66.
- [110] Gorban Alexander N, Kgl Balzs, Wunsch Donald C, Zinovyev Andrei. *Principal Manifolds for Data Visualization and Dimension Reduction*. — 1st edition. — Springer Publishing Company, Incorporated, 2007.
- [111] Cybenko George. Approximations by superpositions of sigmoidal functions // *Approximation Theory and its Applications*. — 1989. — Vol. 9, no. 3. — Pp. 17–28.
- [112] Jeffreys Harold. *Theory of probability*. — Clarendon Press, 1998. — P. 459.
- [113] Karlik B., Vehbi A. Performance Analysis of Various Activation Functions in Generalized MLP Architectures of Neural Networks // *International Journal of Artificial Intelligence and Expert Systems*. — 2011. — Vol. 1, no. 4. — Pp. 111–122.
- [114] Loskutov E. M., Molkov Ya I., Mukhin D. N., Feigin A. M. Markov chain Monte Carlo method in Bayesian reconstruction of dynamical systems from noisy chaotic time series // *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*. — 2008. — jun. — Vol. 77, no. 6. — P. 066214.
- [115] Schwarz Gideon. Estimating the Dimension of a Model // *The Annals of Statistics*. — 1978. — Vol. 6, no. 2. — Pp. 461–464.
- [116] Fletcher Roger. *Practical Methods of Optimization*. — Second edition. — New York, NY, USA: John Wiley & Sons, 1987.
- [117] Theiler James, Eubank Stephen, Longtin André, Galdrikian Bryan, Doyme Farmer J. Testing for nonlinearity in time series: the method of surrogate data // *Physica D: Nonlinear Phenomena*. — 1992. — Vol. 58, no. 1-4. — Pp. 77–94.

- [118] A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker) / Thomas Hale, Noam Angrist, Rafael Goldszmidt, Beatriz Kira, Anna Petherick, Toby Phillips, Samuel Webster, Emily Cameron-Blake et al. // *Nature Human Behaviour*. — 2021. — Vol. 5, no. 4. — Pp. 529–538.
- [119] Техн. отчет: / В.А Адарченко, С.А Бабань, А.А Брагин, К.Ф Гребенкин, О.В Зацепин, А.С Козловских, В.В Легоньков, Е.Н Липилина и др. — Снежинск: Российский Федеральный Ядерный Центр - ВНИИ технической физики имени академика Е.И. Забабахина, 2020.
- [120] *Bamston Anthony G, Chelliah Muthuvel, Goldenberg Stanley B*. Documentation of a highly ENSO-related sst region in the equatorial pacific: Research note // *Atmosphere-Ocean*. — 1997. — sep. — Vol. 35, no. 3. — Pp. 367–383.
- [121] *Takahashi K, Montecinos A, Goubanova K, Dewitte B*. ENSO regimes: Reinterpreting the canonical and Modoki El Niño // *Geophysical Research Letters*. — 2011. — may. — Vol. 38, no. 10.
- [122] *Vimont Daniel J, Wallace John M, Battisti David S*. The Seasonal Footprinting Mechanism in the Pacific: Implications for ENSO* // *Journal of Climate*. — 2003. — Vol. 16, no. 16. — Pp. 2668–2675.
- [123] *Vimont Daniel J, Alexander Michael, Fontaine Abigail*. Midlatitude Excitation of Tropical Variability in the Pacific: The Role of Thermodynamic Coupling and Seasonality* // *Journal of Climate*. — 2009. — Vol. 22, no. 3. — Pp. 518–534.
- [124] *Fang Xiang Hui, Mu Mu*. Both air-sea components are crucial for El Niño forecast from boreal spring // *Scientific Reports*. — 2018. — Vol. 8, no. 1. — Pp. 1–8.
- [125] The NCEP/NCAR 40-Year Reanalysis Project / E Kalnay, M Kanamitsu, R Kistler, W Collins, D Deaven, L Gandin, M Iredell, S Saha at al. // *Bulletin of the American Meteorological Society*. — 1996. — Vol. 77, no. 3. — Pp. 437–472.
- [126] Extended Reconstructed Sea Surface Temperature, Version 5 (ERSSTv5): Upgrades, Validations, and Intercomparisons / Boyin Huang, Peter W. Thorne, Viva F. Banzon, Tim Boyer, Gennady Chepurin, Jay H. Lawrimore, Matthew J. Menne, Thomas M. Smith et al. // *Journal of Climate*. — 2017. — 09. — Vol. 30, no. 20. — Pp. 8179–8205.
- [127] *Tippett Michael K., L'Heureux Michelle L*. Low-dimensional representations of Niño 3.4 evolution and the spring persistence barrier // *npj Climate and Atmospheric Science*. — 2020. — Vol. 3, no. 1. — Pp. 1–11.

- [128] *Yu B, Tang Y M, Zhang X B, Nütsoo A*. An analysis on observed and simulated PNA associated atmospheric diabatic heating // *Climate Dynamics*. — 2009. — Vol. 33, no. 1. — Pp. 75–91.
- [129] El Niño–Southern Oscillation complexity / Axel Timmermann, Soon-Il An, Jong-Seong Kug, Fei-Fei Jin, Wenju Cai, Antonietta Capotondi, Kim M Cobb, Matthieu Lengaigne et al. // *Nature*. — 2018. — Vol. 559, no. 7715. — Pp. 535–545.
- [130] *Cheng Lijing, Trenberth Kevin E, Fasullo John, Boyer Tim, Abraham John, Zhu Jiang*. Improved estimates of ocean heat content from 1960 to 2015 // *Science Advances*. — 2017. — Vol. 3, no. 3.
- [131] *Deser Clara, Alexander Michael A, Xie Shang-Ping, Phillips Adam S*. Sea Surface Temperature Variability: Patterns and Mechanisms // *Annual Review of Marine Science*. — 2009. — dec. — Vol. 2, no. 1. — Pp. 115–143.
- [132] *Zebiak Stephen E*. Oceanic Heat Content Variability and El Niño Cycles // *Journal of Physical Oceanography*. — 1989. — Vol. 19, no. 4. — Pp. 475–486.
- [133] *Clarke Allan J, Gorder Stephen Van, Colantuono Giuseppe*. Wind Stress Curl and ENSO Discharge/Recharge in the Equatorial Pacific // *Journal of Physical Oceanography*. — 2007. — Vol. 37, no. 4. — Pp. 1077–1091.
- [134] *Cheng Lijing, Trenberth Kevin E, Fasullo John T, Mayer Michael, Balmaseda Magdalena, Zhu Jiang*. Evolution of Ocean Heat Content Related to ENSO // *Journal of Climate*. — 2019. — Vol. 32, no. 12. — Pp. 3529–3556.
- [135] *Burgers Gerrit, Jin Fei-Fei, van Oldenborgh Geert Jan*. The simplest ENSO recharge oscillator // *Geophysical Research Letters*. — 2005. — Vol. 32, no. 13.
- [136] *McPhaden Michael J*. A 21st century shift in the relationship between ENSO SST and warm water volume anomalies // *Geophysical Research Letters*. — 2012. — Vol. 39, no. 9.
- [137] *McPhaden Michael J*. Evolution of the 2002/03 El Niño* // *Bulletin of the American Meteorological Society*. — 2004. — Vol. 85, no. 5. — Pp. 677–696.
- [138] *Wang Chunzai, Picaut Joel*. Understanding Enso Physics—A Review // *Earth's Climate*. — 2004. — Vol. 147. — Pp. 21–48.
- [139] *Kaplan Alexey, Cane Mark A, Kushnir Yochanan, Clement Amy C, Blumenthal M Benno, Rajagopalan Balaji*. Analyses of global sea surface temperature 1856-1991 // 1998. — Vol. 103, no. C9. — Pp. 18,518–567,589.

- [140] *Liu Zhengyu, Jin Yishuai, Rong Xinyao*. A Theory for the Seasonal Predictability Barrier: Threshold, Timing, and Intensity // *Journal of Climate*. — 2019. — Vol. 32, no. 2. — Pp. 423–443.
- [141] *Chen Han-Ching, Jin Fei-Fei*. Fundamental Behavior of ENSO Phase Locking // *Journal of Climate*. — 2020. — Vol. 33, no. 5. — Pp. 1953–1968.
- [142] *Wang B, An S*. A mechanism for decadal changes of ENSO behavior: roles of background wind changes // *Climate Dynamics*. — 2002. — Vol. 18, no. 6. — Pp. 475–486.
- [143] *Zhang Yuan, Wallace John M, Battisti David S*. ENSO-like Interdecadal Variability: 1900–93 // *Journal of Climate*. — 1997. — Vol. 10, no. 5. — Pp. 1004–1020.
- [144] *Deser Clara, Phillips Adam S, Hurrell James W*. Pacific Interdecadal Climate Variability: Linkages between the Tropics and the North Pacific during Boreal Winter since 1900 // *Journal of Climate*. — 2004. — Vol. 17, no. 16. — Pp. 3109–3124.
- [145] *Ray Sulagna, Giese Benjamin S*. Historical changes in El Niño and La Niña characteristics in an ocean reanalysis // *Journal of Geophysical Research: Oceans*. — 2012. — Vol. 117, no. C11.
- [146] *Rayner N A, Parker D E, Horton E B, Folland C K, Alexander L V, Rowell D P, Kent E C, Kaplan A*. Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century // *J. Geophys. Res.* — 2003. — Vol. 108, no. D14. — P. 4407.
- [147] Persistent Solar Influence on North Atlantic Climate During the Holocene / Gerard Bond, Bernd Kromer, Juerg Beer, Raimund Muscheler, Michael N Evans, William Showers, Sharon Hoffmann, Rusty Lotti-Bond et al. // *Science*. — 2001. — Vol. 294, no. 5549. — Pp. 2130–2136.
- [148] *Emile-Geay Julien, Cane Mark, Seager Richard, Kaplan Alexey, Almasi Peter*. El Niño as a mediator of the solar influence on climate // *Paleoceanography*. — 2007. — Vol. 22, no. 3.
- [149] *Shindell Drew, Rind David, Balachandran Nambeth, Lean Judith, Lonergan Patrick*. Solar Cycle Variability, Ozone, and Climate // *Science*. — 1999. — Vol. 284, no. 5412. — Pp. 305–308.

- [150] *Shindell Drew T, Schmidt Gavin A, Mann Michael E, Rind David, Waple Anne.* Solar Forcing of Regional Climate Change During the Maunder Minimum // *Science*. — 2001. — Vol. 294, no. 5549. — Pp. 2149–2152.
- [151] *Clement Amy C, Seager Richard, Cane Mark A, Zebiak Stephen E.* An Ocean Dynamical Thermostat // *Journal of Climate*. — 1996. — Vol. 9, no. 9. — Pp. 2190–2196.
- [152] *Seager Richard, Zebiak Stephen E, Cane Mark A.* A model of the tropical Pacific sea surface temperature climatology // *Journal of Geophysical Research: Oceans*. — 1988. — Vol. 93, no. C2. — Pp. 1265–1280.
- [153] ENSO Atmospheric Teleconnections and Their Response to Greenhouse Gas Forcing / Sang-Wook Yeh, Wenju Cai, Seung-Ki Min, Michael J McPhaden, Dietmar Dommenges, Boris Dewitte, Matthew Collins, Karumuri Ashok et al. // *Reviews of Geophysics*. — 2018. — Vol. 56, no. 1. — Pp. 185–206.
- [154] *Steinhilber F, Beer J, Fröhlich C.* Total solar irradiance during the Holocene // *Geophysical Research Letters*. — 2009. — Vol. 36, no. 19.
- [155] *Peristykh Alexei N, Damon Paul E.* Persistence of the Gleissberg 88-year solar cycle over the last 12,000 years: Evidence from cosmogenic isotopes // *Journal of Geophysical Research: Space Physics*. — 2003. — Vol. 108, no. A1. — Pp. SSH 1–1–SSH 1–15.
- [156] Presence of the Solar de Vries Cycle (205 years) during the Last Ice Age / Gerhard Wagner, Jürg Beer, Jozef Masarik, Raimund Muscheler, Peter W Kubik, Werner Mende, Carlo Laj, Grant M Raisbeck, Françoise Yiou // *Geophysical Research Letters*. — 2001. — Vol. 28, no. 2. — Pp. 303–306.
- [157] *Torrence Christopher, Compo Gilbert P.* A Practical Guide to Wavelet Analysis // *Bulletin of the American Meteorological Society*. — 1998. — Vol. 79, no. 1. — Pp. 61–78.
- [158] Historical greenhouse gas concentrations for climate modelling (CMIP6) / M Meinshausen, E Vogel, A Nauels, K Lorbacher, N Meinshausen, D M Etheridge, P J Fraser, S A Montzka et al. // *Geoscientific Model Development*. — 2017. — Vol. 10, no. 5. — Pp. 2057–2116.
- [159] *Чистяков В.П.* Теория вероятностей. — 3 издание. — Москва: Наука, 1987. — 240 С.

- [160] *Hoerl Arthur E, Kennard Robert W.* Ridge Regression: Biased Estimation for Nonorthogonal Problems // *Technometrics*. — 1970. — feb. — Vol. 12, no. 1. — Pp. 55–67.