

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

на правах рукописи

Гаврилов Андрей Сергеевич

Методы эмпирической реконструкции
пространственно распределенных динамических
систем и их приложение к изучению
климатических процессов

01.04.03 – радиофизика

ДИССЕРТАЦИЯ

на соискание ученой степени
кандидата физико-математических наук

Научный руководитель
доктор физико-математических наук
Фейгин Александр Маркович

Нижний Новгород – 2019

Оглавление

| | |
|---|-----------|
| Введение | 5 |
| 1 Метод эмпирической реконструкции оператора эволюции по многомерным временным рядам | 16 |
| 1.1 Введение | 16 |
| 1.2 Стохастическая модель оператора эволюции на основе искусственных нейронных сетей | 17 |
| 1.3 Выбор оптимальной модели | 21 |
| 1.3.1 Байесов подход | 21 |
| 1.3.2 Пример: система Лоренца-63 с динамическим шумом | 29 |
| 1.4 Реконструкция оператора эволюции по пространственно распределенным временным рядам | 35 |
| 1.4.1 Описание и предварительная обработка используемых данных . | 35 |
| 1.4.2 Эмпирические ортогональные функции | 36 |
| 1.4.3 Модифицированная стохастическая модель оператора эволюции | 40 |
| 1.4.4 Реконструкция оператора эволюции по данным климатической модели INMCM4.0 | 43 |
| 1.5 Заключение | 48 |
| 2 Метод построения нелинейных динамических мод по пространственно распределенным временным рядам | 50 |
| 2.1 Введение | 50 |
| 2.2 Нелинейные динамические моды | 51 |
| 2.2.1 Параметризация НДМ | 53 |
| 2.2.2 Байесов подход к поиску НДМ | 54 |
| 2.2.3 Модельный пример | 61 |
| 2.2.4 Отличие линейных НДМ от ЭОФ | 64 |
| 2.3 Главные нелинейные моды климатической изменчивости по данным 1981-2014 гг. | 67 |

| | | |
|----------|--|------------|
| 2.3.1 | Наблюдаемые данные и результат их НДМ-разложения | 67 |
| 2.3.2 | Анализ главных мод | 70 |
| 2.3.3 | Климатический переход 1997-1998 годов | 74 |
| 2.3.4 | Возможность детектирования дальних связей | 75 |
| 2.4 | Главные моды доиндустриального климата модели INMCM4.0 | 76 |
| 2.4.1 | Параметры временного ряда и полученные НДМ | 76 |
| 2.4.2 | Структура главных многообразий | 81 |
| 2.5 | Заключение | 82 |
| 3 | Многомерные нелинейные динамические моды | 85 |
| 3.1 | Введение | 85 |
| 3.2 | Многомерное обобщение НДМ | 86 |
| 3.2.1 | Модификация модели НДМ | 86 |
| 3.2.2 | Алгоритм поиска МНДМ | 87 |
| 3.2.3 | Модельный пример | 88 |
| 3.3 | Двумерные моды ТПО на интервале с 1981 по 2014 год | 90 |
| 3.4 | Использование двумерных мод для анализа изменчивости климата на интервале с 1870 по 2014 год | 92 |
| 3.4.1 | Используемые данные и характеристики МНДМ | 92 |
| 3.4.2 | Реконструкция смены фаз тихоокеанского декадного колебания | 93 |
| 3.4.3 | Анализ дальних связей различных регионов в 1870-2014 годах . | 96 |
| 3.5 | Заключение | 100 |
| 4 | Эмпирический прогноз динамики явления Эль-Ниньо | 102 |
| 4.1 | Введение | 102 |
| 4.2 | Алгоритм построения модели и прогноза | 103 |
| 4.2.1 | Выбор переменных | 104 |
| 4.2.2 | Построение модели оператора эволюции | 105 |
| 4.2.3 | Схема построения прогноза | 106 |
| 4.3 | Описание используемых данных | 107 |
| 4.4 | Сравнение эффективности моделей МЛДМ-ИНС и ЭОФ-ИНС | 107 |
| 4.4.1 | Временные ряды главных компонент | 108 |
| 4.4.2 | Прогностическая способность моделей | 109 |
| 4.4.3 | Регионы с наилучшей предсказуемостью | 112 |
| 4.5 | Приложение к прогнозу динамики Эль-Ниньо | 114 |
| 4.5.1 | Прогноз индексов и событий Эль-Ниньо/Ла-Нинья | 115 |
| 4.5.2 | Прогностическая способность за пределами обучающей выборки | 118 |

| | |
|--|------------|
| 4.6 Заключение | 121 |
| Заключение | 123 |
| Список публикаций по теме диссертации | 127 |
| Литература | 136 |

Введение

Актуальность работы. Диссертация посвящена разработке и применению новых методов исследования сложных систем: пространственно распределенных систем, демонстрирующих мультимасштабную высокоразмерную динамику. В настоящее время все более и более актуальным становится эмпирическое исследование таких систем, основанное на построении их моделей путем прямого анализа наблюдаемых данных. Ввиду быстрого роста количества и качества методов и средств измерений, доступных вычислительных мощностей и сопутствующих методов машинного обучения, методы эмпирической реконструкции уже сегодня могут быть применены во многих активно развивающихся областях знаний (наиболее известными примерами являются нейросистемы и климатическая система Земли). По сравнению с традиционными моделями «из первых принципов», эмпирические модели в большей степени определяются имеющимися данными, чем конкретными параметризациями различных процессов, в том числе процессов, механизмы которых до сих пор до конца не изучены, и процессов подсеточного масштаба. В наиболее общей постановке, при эмпирическом моделировании конкретные уравнения, описывающие систему, заранее *не известны*.

К основным задачам, которые могут быть решены эмпирическими методами и которым посвящена настоящая работа, можно отнести выявление законов эволюции исследуемой системы и прогноз ее поведения; поиск низкоразмерных переменных для описания системы (эмпирическая редукция размерности) и, как следствие, выявление главных режимов (мод) ее изменчивости; поиск и анализ связей между компонентами исследуемой системы. Задача выявления законов эволюции системы имеет непосредственное отношение к нелинейной динамике: как правило, при ее решении система, породившая наблюдаемый временной ряд данных, рассматривается как динамическая система, т.е. предполагается наличие оператора эволюции и фазового пространства, в котором он действует [8, 22]. Обоснованием принципиальной возможности реконструкции оператора эволюции системы по наблюдаемому временному ряду конкретной величины служат доказанные Такенсом теоремы [116], из которых следует, что лишь по одному скалярному сигналу детерминированной динамической системы, «живущей» на конечномерном аттракторе (например, отвечающем хаотическому режиму

эволюции), можно топологически корректно восстановить этот аттрактор, используя в качестве переменных тот же самый сигнал, сдвинутый по времени достаточное число раз (переменные с задержками). Несмотря на то, что условия справедливости теоремы Такенса представляют собой достаточно идеализированный случай, идея реконструкции динамической системы с помощью переменных с задержками активно используется, и в последние десятилетия реконструкции динамических систем посвящено большое количество работ (см., например, [6, 22] и цитируемую там литературу). В частности, методы реконструкции детерминированного оператора эволюции можно разбить на две группы: локальные и глобальные. К первой группе относятся методы, в которых для каждой интересующей точки фазового пространства строится локальный оператор эволюции в малой окрестности этой точки. Для аппроксимации такого оператора эволюции могут использоваться как полиномиальные [38], так и более сложные функции, например, радиальные базисные функции [99]. Особенностью таких методов является требование высокой посещаемости каждой интересующей точки, т.е. большого размера выборки данных, ввиду большого количества независимых параметров в соответствующей модели. Кроме того, существенными ограничениями являются требование стационарности временного ряда и малого шума измерений для применения данных методов. Ко второй группе относятся методы, в которых оператор эволюции параметризуется определенной функцией, действующей глобально, т.е. во всей области фазового пространства, где происходит эволюция системы. В этом случае задача реконструкции состоит в поиске параметров данной функции, число которых, как правило, значительно меньше, чем в методах локальной реконструкции. Глобальные модели оператора эволюции способны не только реконструировать структуру фазового пространства, но и отслеживать изменения управляющих параметров системы и поэтому могут применяться в задачах реконструкции неавтономности системы [1, 40], восстановления бифуркационных диаграмм [12, 13, 119], передачи информации [2, 5, 10] и т.д. Фактически, на основе глобальных моделей к настоящему времени созданы и проверены на множестве модельных примеров методы реконструкции детерминированного оператора эволюции по хаотическим [39], нестационарным [20, 87, 112], зашумленным [24, 81, 86, 90] временным рядам. Кроме того, были предложены методы построения потоковых глобальных моделей в виде систем дифференциальных уравнений заранее выбранного вида [3, 21, 111]. Разрабатываемые в настоящей работе модели оператора эволюции также относятся к глобальным моделям.

Во многих практических приложениях, в частности, при исследовании климата, наблюдаемые данные представляют собой эволюционирующие во времени пространственные поля одной или нескольких переменных, измеренные с дискретным пространственным и временным шагом и имеющие ограниченную протяженность

по времени. Размерность таких пространственно распределенных временных рядов пропорциональна количеству узлов в пространственной сетке, т.е. огромна. С одной стороны, такие данные содержат больше информации о системе, чем каждый из составляющих их скалярных временных рядов, с другой стороны, конечная длина реально доступной временной выборки порождает существенное фундаментальное ограничение на максимально допустимую размерность фазового пространства, в котором возможно статистически обоснованное построение модели оператора эволюции (включая вышеупомянутые детерминированные модели на основе переменных с задержками), и, следовательно, исходные переменные на пространственно распределенной сетке напрямую для моделирования оператора эволюции не подходят. При этом ясно, что во многих случаях для описания системы в конкретном диапазоне пространственных и временных масштабов (например, сравнительно крупномасштабных процессов, хорошо представленных в конкретном наблюдаемом ряде данных) количество независимых переменных, необходимых для описания их динамики, может быть существенно меньше, чем количество узлов в полной пространственной сетке. Поэтому для таких систем активно разрабатываются методы эмпирической редукции размерности.

Эмпирическая редукция размерности является одним из наиболее важных этапов в задаче выбора фазовых переменных, поскольку она должна максимально возможным образом сохранять ключевые свойства системы, лежащие в основе наблюдаемой динамики. Низкоразмерные переменные, получаемые в результате эмпирической редукции размерности, могут открывать возможности к исследованию главных мод изменчивости системы и ее низкоразмерному моделированию. С формальной точки зрения, редукция размерности состоит в отображении наблюдаемых данных из высокоразмерного пространства исходных переменных в низкоразмерное подпространство (многообразие), задаваемое новыми определяющими его переменными. Одним из наиболее простых и при этом широко используемых способов редукции является метод главных компонент (principal component analysis – PCA), известный также в геофизических приложениях как разложение по базису эмпирических ортогональных функций (ЭОФ) [58, 91, 100]. Этот метод позволяет находить линейное отображение, максимизирующее долю вариации¹ исходных данных, описываемую новыми переменными – главными компонентами (ГК), – которые при этом являются линейно некоррелирующими друг с другом временными рядами; более подробно этот метод описан в одном из разделов диссертации. Существуют и более продвинутое линей-

¹Под вариацией скалярного временного ряда здесь и далее понимается средний квадрат отклонения значений временного ряда от их среднего. Под вариацией многомерного временного ряда понимается сумма вариаций его компонент.

ные методы и варианты ГК, такие как поворот «варимакс» [47, 124], дополнительно стремящийся разделить ГК в пространстве; анализ независимых подпространств, в котором требование линейной некоррелированности ГК заменено более общим требованием их независимости в смысле взаимной информации [97, 98]; главные динамические компоненты, учитывающие линейный оператор эволюции, действующий в пространстве искомым ГК [137]; главные паттерны взаимодействий [49, 73–75] – линейные ГК, учитывающие нелинейность оператора эволюции для ГК, но используемые обычно для редукции моделей из первых принципов; главные осцилляторные паттерны [138], представляющие собой нормальные моды линейного оператора эволюции, реконструируемого непосредственно по наблюдаемым данным; оптимально инерционные паттерны (optimally persistent patterns [28]) и разложение по среднему времени предсказуемости [29, 30], предоставляющие линейный поворот, после которого ГК оказываются упорядоченными по времени инерционности (persistence time) и среднему времени предсказуемости соответственно; и другие. Существует обобщение ЭОФ-разложения, называемое также многоканальным сингулярным спектральным анализом [43], которое основано на линейном преобразовании в расширенном пространстве, образованном исходными переменными, взятыми с временной задержкой в пределах некоторого временного окна, и учитывает, таким образом, запаздывающие корреляции между процессами в пространственно разнесенных точках (обусловленные, например, конечным временем распространения сигнала), что больше соответствует физике пространственно распределенных систем. Следует отметить, что, с одной стороны, линейность преобразования данных, лежащая в основе вышеперечисленных методов, приводит к простоте их численной реализации и сравнительно небольшой ресурсоемкости. С другой стороны, при этом не учитывается в явном виде нелинейность связей между различными наблюдаемыми переменными, которая, в случае общего положения, может иметь место в рассматриваемых системах (в частности, в климатической системе), ограничивая эффективность применения линейных методов в целом ряде задач.

Методы *нелинейной* эмпирической редукции размерности [77] пытаются тем или иным образом учесть и выявить, в том числе, нелинейные взаимосвязи. Многие из этих методов («kernel PCA» методы) [19, 27, 109] сводятся к использованию линейного метода главных компонент в специальном пространстве, которое можно представить, как результат определенного нелинейного преобразования исходных переменных, и уже имеют применение, в частности, в климатических приложениях [41, 48, 105]. Аналогично линейным методам редукции, такие методы обладают невысокой ресурсоемкостью (она определяется, главным образом, вычислительной сложностью закладываемого нелинейного преобразования), а также для них гаран-

тирована единственность решения; основным недостатком этих методов, который существенно ограничивает область их применения, является отсутствие в явном виде обратного преобразования из пространства найденных ГК в пространство исходных переменных. Другую группу методов нелинейной редукции размерности составляют методы, основанные на построении главного нелинейного многообразия напрямую [50], когда находятся одновременно скрытые переменные (нелинейные ГК) и соответствующее обратное преобразование. В частности, автоассоциативные нейронные сети [68] включают в себя как отображение исходных данных на низкоразмерное главное многообразие, так и обратное преобразование; нейронные сети с обучающимися входами [117] осуществляют лишь обратное отображение низкоразмерного главного многообразия в пространство исходных переменных, а нелинейные ГК при этом находятся в результате оптимизации сигнала на входе таких сетей. Как правило, при работе с высокоразмерными (пространственно распределенными) исходными данными в качестве предварительного шага для сжатия этих данных используются линейные методы редукции размерности [55]. Стоит иметь в виду, что задача поиска главного нелинейного многообразия, являясь наиболее общей по постановке, сама по себе, вообще говоря, не имеет единственного решения и поэтому требует наличия адекватной регуляризации (т.е. априорной информации) и применения методов оптимизации, что существенным образом определяет ее конкретное решение. В связи с этим ресурсоемкость соответствующих методов оказывается значительно выше, чем ресурсоемкость линейных методов редукции размерности. В диссертационной работе разработан новый метод разложения данных на основе поиска главного нелинейного многообразия (метод нелинейных динамических мод), учитывающий не только величину вариации данных, но и доминирующие временные масштабы эволюции системы в получаемых ГК. Метод реализован в рамках байесова подхода к эмпирическому моделированию [54], суть которого подробно изложена в соответствующем разделе диссертации.

Возвращаясь к задаче эмпирической реконструкции оператора эволюции пространственно распределенной системы, следует отметить, что необходимость редукции размерности данных неизбежно приводит к тому, что получаемые ГК описывают лишь низкоразмерную проекцию полного фазового пространства системы (точно так же такенсовы переменные с задержками описывают, вообще говоря, лишь проекцию полного фазового пространства, если допустимое количество переменных с задержкой недостаточно велико из-за ограниченной длины временного ряда). При этом оператор эволюции, действующий в полном высокоразмерном фазовом пространстве, даже если он является детерминированным, при таком низкоразмерном представлении становится, вообще говоря, неоднозначным, т.е. теряет детерминизм. Кроме

того, дискретность наблюдаемых данных делает невозможным корректное описание процессов с характерными масштабами, меньшими, чем шаг временного ряда. В данной ситуации одним из способов учесть возможное влияние плохо разрешенных процессов является их описание в виде стохастического (случайного) процесса, т.е. с помощью построения стохастической модели оператора эволюции [85]. В настоящее время существует множество работ, посвященных построению низкоразмерных эмпирических моделей (как правило, в пространстве ГК, получаемых с помощью ЭОФ-разложения). В некоторых из них оператор эволюции параметризуется в виде суммы линейной детерминированной компоненты и пространственно распределенного гауссова процесса в качестве случайной компоненты [18, 93, 95], в других моделях используются нелинейные параметризаторы, такие как полиномы [26, 67, 69–71, 114] и искусственные нейронные сети (ИНС) [44, 89, 106, 129, 136]. В работах [85, 89] была предложена и продемонстрирована на простых модельных примерах стохастическая модель на основе ИНС, в которой корреляционная матрица случайной компоненты нелинейно зависит от состояния системы в фазовом пространстве, в отличие от других упомянутых моделей, где случайная компонента моделировалась в форме простого гауссова шума. При этом ясно, что в общем случае заранее нельзя дать ответ на вопрос о том, насколько важен учет нелинейности детерминированной компоненты и нетривиальности случайной компоненты оператора эволюции, поскольку этот ответ зависит от информации, предоставленной наблюдаемым временным рядом данных. Другими словами, оптимальная для данного временного ряда структура модели заранее не известна. В диссертационной работе для такой модели разработан метод выбора ее оптимальной структуры и обучения в рамках байесова подхода и сделана модификация модели для случая ГК, полученных по пространственно распределенным климатическим данным.

Объектом применения вышеописанных методов в настоящей работе является климатическая система Земли. Из всего многообразия временных масштабов, имеющих в данной системе, внимание будет уделено диапазону масштабов от нескольких месяцев до десятилетий, представляемых пространственно распределенными временными рядами температуры поверхности океана (ТПО) в XIX–XXI столетиях. Одной из главных и наиболее известных составляющих изменчивости климата на этих временных масштабах является Эль-Ниньо – Южное Колебание [120] (ЭНЮК), которое регистрируется как крупномасштабная температурная аномалия в тропической зоне Тихого океана, повторяющаяся с нерегулярным временным интервалом от 2 до 7 лет². Данное явление существенно влияет на погодные условия (в том числе, экстре-

²Строго говоря, под Эль-Ниньо понимается фаза колебания, при котором наблюдается сильная положительная аномалия ТПО. Противоположная фаза называется «Ла-Нинья».

мальные) как непосредственно в тропиках Тихого океана, так и во многих других регионах земного шара [7, 17, 56, 96, 113]. Это делает задачу прогноза динамики ЭНЮК чрезвычайно актуальной, и рассмотрению этой задачи посвящена отдельная глава диссертации. В динамике ТПО присутствуют и более крупные временные масштабы: тренд глобального потепления, связанный как с естественной изменчивостью, так и с антропогенным ростом концентрации парниковых газов в атмосфере, представляющим собой внешнее воздействие на климатическую систему, декадная изменчивость ТПО в Тихом океане [82], северной Атлантике [63] и др. Относительная близость величин этих масштабов не позволяет рассматривать порождающие их процессы как полностью независимые, невзаимодействующие подсистемы. Так, в недавних работах [31, 122, 130] указывается, что естественная декадная изменчивость существенно модулирует рост глобальной средней температуры; в частности, по мнению авторов работы [122], замедление глобального потепления 1998-2015 гг. [35] тесно связано с наступлением отрицательной фазы Тихоокеанского декадного колебания (ТДК) [14, 84, 122], которое также сдвинуло изменчивость ЭНЮК в сторону доминирования условий Ла-Нинья. В настоящее время ясно, что корректное моделирование естественной изменчивости чрезвычайно важно для оценки будущего поведения глобального климата; при этом воспроизведение декадной изменчивости существующими климатическими моделями все еще не является достаточно надежным [36, 122]. В рамках диссертации альтернативный подход, основанный на разработанном эмпирическом методе редукции размерности, применен для выявления и анализа мод изменчивости климата на данных масштабах.

Основной целью диссертации является разработка новых методов эмпирической реконструкции пространственно распределенных систем, применимых, в том числе, к наблюдаемым данным об эволюции климатической системы Земли. Для достижения цели были поставлены и решены следующие конкретные задачи: (1) разработка и тестирование метода построения оптимальной эмпирической модели оператора эволюции, учитывающего такие особенности моделируемой системы, как ее пространственная распределенность, нелинейность, нестационарность, а также наличие различных внешних воздействий; (2) разработка и тестирование нового метода эмпирической редукции размерности пространственно распределенных данных, учитывающего оптимальным образом как нелинейность связей между процессами в пространственно разнесенных точках, так и динамическую природу этих процессов, проявляющуюся в существовании различных характерных временных масштабов эволюции системы; (3) применение разработанных методов редукции размерности к данным температуры поверхности океана и интерпретация результатов; (4) создание прогностической эмпирической модели Эль-Ниньо Южного Колебания – одной из

главных мод изменчивости климата на межгодовых масштабах – на основе разработанных методов.

Структура диссертации. Диссертация состоит из введения, четырех глав, заключения, списка публикаций по теме диссертации и списка цитируемой литературы. Общий объем диссертации 148 страниц, включая 44 рисунка, 3 таблицы и список литературы из 127 наименований.

В **первой главе** диссертации излагается метод построения оптимальной нелинейной стохастической модели оператора эволюции системы по наблюдаемому многомерному временному ряду данных: объясняется целесообразность выбора функциональной формы модели, подробно формулируются общий байесов подход к обучению модели и оптимизации ее структурных параметров и алгоритм, численно реализующий данный подход, приводится модельный пример, наглядно демонстрирующий способность модели воспроизводить динамические свойства моделируемой системы. Во второй части первой главы рассматриваются особенности применения подхода к временному ряду данных, порожденному пространственно распределенной системой. Для моделирования в этом случае берутся временные ряды главных компонент – переменных, полученных с помощью методов эмпирической редукции размерности пространственно распределенных данных (в качестве примера, в данной главе используется наиболее простой и употребимый из таких методов – ЭОФ-разложение) и описывающих главную часть динамики системы в низкоразмерной проекции фазового пространства. Проводится модификация формы стохастической модели оператора эволюции, дополнительно учитывающая (1) «медленную» нестационарность системы, связанную с неизбежным наличием в сложной пространственно распределенной системе процессов с временными масштабами, превышающими доступную длину наблюдаемого временного ряда и не описываемыми первоначально предложенной формой стохастической модели, (2) наличие возможных внешних воздействий, приводящих к явной неавтономности оператора эволюции системы, (3) детерминированные причинно-следственные связи, проявляющиеся в ненулевом временном масштабе автокорреляции («гладкости») главных компонент пространственно распределенной системы. Адекватность предложенной модифицированной стохастической модели и способность модели как давать количественный прогноз эволюции главных компонент, так и воспроизводить качественное поведение системы, включая медленную нестационарность, демонстрируются на примере данных ТПО, сгенерированных глобальной климатической моделью INMCM4.0 в ходе численного эксперимента по воспроизведению климата XX века в рамках международного проекта по сравнению климатических моделей «CMIP5».

Вторая глава посвящена разработке нового метода эмпирической редукции раз-

мерности пространственно распределенных данных, состоящего в поиске нелинейных динамических мод (НДМ). Главными свойствами НДМ, отличающими их от других методов редукции размерности, таких как ЭОФ-разложение, являются (1) учет в явном виде нелинейной связи между величинами, измеряемыми в пространственно разнесенных точках наблюдаемой системы, и (2) учет причинно-следственных связей в системе, проявляющихся в наличии характерных временных масштабов у получаемых временных рядов главных компонент. Оба данных свойства являются ключевыми свойствами, присущими всем пространственно распределенным динамическим системам, и, в частности, климатической системе Земли. Формулируется итерационный алгоритм поиска НДМ, при котором НДМ ищутся последовательно друг за другом на основе байесова подхода, подробно изложенного в первой главе диссертации. Фактически, каждая НДМ представляет собой нелинейный пространственно распределенный отклик на скалярную скрытую переменную – ГК – и соответствует проекции исходных данных на одномерное нелинейное многообразие, при этом временной масштаб ГК и степень нелинейности многообразия находятся с помощью байесова критерия оптимальности. Работа метода НДМ и его преимущества по отношению к ЭОФ-разложению демонстрируются на наглядных модельных примерах.

Во второй части главы разработанный метод НДМ-разложения применяется к выявлению и анализу главных мод изменчивости современного климата по спутниковым измерениям ТПО на интервале с 1981 по 2014 год и «доиндустриального» климата конца XIX века по данным, сгенерированным моделью INMCM4.0 в ходе соответствующего численного эксперимента SMIP5. По данным современного климата было найдено три существенно нелинейных НДМ, которые, помимо прочего, описывают такие явления, как годовой ход ТПО, ЭНЮК и тихоокеанское декадное колебание. В частности, анализ структуры этих НДМ подтверждает наличие климатического сдвига, приведшего к вышеупомянутой смене фазы ТДК в 1997-1998 годах, а также гипотезу о решающей роли ЭНЮК в наступлении этого сдвига. Кроме того, демонстрируется возможность использования НДМ для исследования структуры дальних связей в климате Земли.

НДМ, соответствующие доиндустриальному климату, также описывают годовой ход, ЭНЮК и ТДК, но при этом являются линейными (за исключением НДМ, описывающей годовой ход), свидетельствуя об ином режиме функционирования климатической системы и другой связи явлений в различных регионах земного шара. Таким образом, в данной главе демонстрируется свойство НДМ быть как линейными, так и нелинейными, в зависимости от свойств наблюдаемой системы.

Третья глава посвящена дальнейшему развитию НДМ-разложения. В ней указываются ограничения ранее сформулированного алгоритма поиска НДМ, связанные

с итерационностью этого алгоритма и, как следствие, одномерностью находимых главных многообразий и их неспособностью описать в общем виде произвольное неодномерное многообразие. В связи с этим предлагается многомерное обобщение НДМ, формулируется соответствующий алгоритм и на простом модельном примере показывается его лучшая эффективность, по сравнению с разложением на одномерные НДМ. Демонстрируется, что применение многомерных НДМ к данным ТПО современного климата на интервале с 1981 по 2014 год также приводит к большей адекватности полученных ГК, по сравнению с результатом итерационного поиска одномерных НДМ, и при этом подтверждает основные выводы о современном климате, сделанные во второй главе.

После этого с помощью метода многомерных НДМ анализируется пространственно распределенный временной ряд, основанный на доступных результатах измерений ТПО с 1870 по 2014 год, разбитый на 6 пересекающихся по времени фрагментов длиной 33 года (для уменьшения вклада медленных трендов в анализируемые ряды): на каждом из этих фрагментов находится двумерная НДМ. В результате такого анализа подтверждено наличие четырех климатических переходов в XX веке, связанных с изменением фазы ТДК. Кроме того, сделан вывод, согласующийся с результатами второй главы, о том, что структура детектируемых с помощью НДМ дальних связей сменяется с линейной (в доиндустриальном климате) на нелинейную (в современном климате).

В **четвертой главе** разработанные в предыдущих главах методы применяются к задаче прогноза ежемесячных временных рядов ТПО в тропическом поясе, где доминирующим режимом изменчивости на межгодовых масштабах является ЭНЮК. В частности, построение эмпирической прогностической модели проводится в два шага: сначала производится эмпирическая редукция размерности данных с помощью линейного варианта многомерных НДМ (линейный вариант выбран по причине приемлемой ресурсоемкости) – многомерных линейных динамических мод; затем в пространстве полученных ГК строится модифицированная стохастическая модель оператора эволюции, разработанная в первой главе. Показывается, что использование многомерных линейных динамических мод приводит к лучшей прогностической способности модели, чем использование переменных, полученных ЭОФ-разложением, на котором основано большинство существующих эмпирических моделей ЭНЮК. Проводится углубленное исследование эффективности построенной эмпирической модели применительно к задаче краткосрочного (в пределах одного года) прогноза индексов ЭНЮК и событий Эль-Ниньо/Ла-Нинья. Демонстрируется, что по точности и качеству прогноза индексов ЭНЮК созданная модель не уступает существующим мировым моделям ЭНЮК (как эмпирическим, так и моделям из первых принципов) и

может быть использована, наряду с ними, для ежемесячного прогноза этих индексов.

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

Апробация представленных в работе результатов. Основные результаты работы докладывались на семинарах ИПФ РАН, а также на международных и российских научных конференциях: XIV, XV, XVI, XVIII и XXI Всероссийских конференциях молодых ученых «Состав атмосферы. Атмосферное электричество. Климатические процессы» (2010, 2011, 2012, 2014, 2017), Всероссийских научных школах «Нелинейные волны» (Нижний Новгород, 2012, 2016, 2018), международных конференциях «Хаотические автоколебания и образование структур» (2010, 2013), «Topical problems of nonlinear wave physics» (Нижний Новгород, 2014, 2017), «Frontiers of nonlinear physics» (Нижний Новгород, 2016), «European Geosciences Union General Assembly» (2011, 2013, 2014, 2015, 2016, 2017, 2018), «American Geophysical Union Fall Meeting» (2011, 2012, 2013, 2014, 2015, 2016, 2017), «Science of the future» (2014, 2016), «Challenges in Data Science: a complex systems perspective» (2015), «CLIVAR Open Science Conference» (2016), «Fourth International Conference on Earth System Modelling» (2017), «International work-conference on Time Series» (2014), «JpGU-AGU Joint Meeting» (2017), «Davos Atmosphere and Cryosphere Assembly» (2013), «Dynamics Days Europe» (2012), 31 и 32 международных конференциях «IUGG Conference on mathematical geophysics» (2016, 2018).

Результаты, полученные в ходе выполнения работы, рекомендованы в доклад Президента РАН за 2015 год и в Отчеты РАН за 2017, 2018 годы. Результаты исследований также использовались при выполнении работ по гранту Правительства РФ «Новые подходы к исследованию климатических процессов и прогнозу экстремальных явлений» (14.Z50.31.0033), грантам РНФ (16-12-10198, 18-12-00231, 19-42-04121), РФФИ (19-02-00502 А). По теме диссертации опубликована 71 работа, в числе этих работ 6 статей в рецензируемых научных журналах из списка ВАК [A1–A6], 4 свидетельства о регистрации программы для ЭВМ [A7–A10], 6 работ в сборниках трудов конференций [A11–A16].

Личный вклад автора. Все приведенные в диссертации результаты получены либо лично автором, либо при его непосредственном участии.

Глава 1

Метод эмпирической реконструкции оператора эволюции по многомерным временным рядам

1.1 Введение

В данной главе рассматривается метод построения эмпирической модели (реконструкции) оператора эволюции динамической системы по многомерному временному ряду конечной длины на основе байесова подхода. Исходя из общих принципов, формулируется стохастическая модель, параметризуемая с помощью универсального нелинейного аппроксиматора – искусственных нейронных сетей (ИНС), и обосновывается адекватность ее выбора в данной задаче (раздел 1.2). В разделе 1.3 подробно излагается байесов подход к выбору конкретной структуры и параметров сформулированной модели, формулируется алгоритм ее обучения, а также на модельном примере системы Лоренца-63 с интерактивным шумом наглядно демонстрируются различные аспекты работы данного метода реконструкции. Более того, излагаемый в разделе 1.3.1 общий байесов подход будет являться методической основой и при построении нелинейных динамических мод в следующих главах диссертации.

Отдельное и главное внимание будет уделено важному частному случаю, когда наблюдаемый многомерный временной ряд представляет собой распределенный в пространстве сигнал (раздел 1.4), сгенерированный сложной мультимасштабной системой, что имеет место, в частности, при исследовании климатической системы. Будет предложена модификация формы стохастической модели оператора эволюции, учитывающая особенности его реконструкции в данной ситуации. В качестве нетривиального модельного примера при этом будет использован временной ряд,

сгенерированный глобальной климатической моделью INMCM4.0, разработанной в институте вычислительной математики РАН и участвовавшей в международном проекте по сравнению климатических моделей «CMIP5».

Материалы данной главы представлены в статьях [A1, A4, A6] и публикациях [A8, A10–A12, A14, A17–A34].

1.2 Стохастическая модель оператора эволюции на основе искусственных нейронных сетей

Пусть временной ряд $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_N)$, где $\mathbf{x}_n \in \mathbb{R}^D$, является результатом измерения характеристик некоторой неизвестной динамической системы в N последовательных эквидистантных моментах времени t_1, \dots, t_N (без ограничения общности здесь и всюду далее будем считать среднее по времени значение каждой компоненты ряда \mathbf{X} равным нулю). Под динамической системой (детерминированной [6] или случайной [11]) в данном случае будем понимать систему, состояние которой меняется во времени в соответствии с определенным законом эволюции (детерминированным или случайным). Математическая модель динамической системы включает фазовое пространство, каждый элемент которого однозначно описывает состояние системы, и оператор эволюции, действующий в этом пространстве. Заметим, что образующие фазовое пространство переменные (фазовые переменные) для одной и той же динамической системы могут быть выбраны разными способами. В то же время, наблюдаемые величины \mathbf{x} совершенно не обязательно представляют собой полный набор фазовых переменных.

Более того, анализ наблюдаемых данных, вообще говоря, не позволяет сделать вывода о размерности фазового пространства системы, породившей временной ряд \mathbf{X} . Применительно к реальным сложным системам, таким как климатическая система, это обстоятельство является следствием высокой размерности системы: число степеней свободы слишком велико для того, чтобы его можно было статистически обоснованно выявить по доступному для наблюдения временному ряду конечной длины. Поэтому при эмпирической реконструкции оператора эволюции модель неизбежно строится лишь в некотором подпространстве (проекции) размерности D_u , элементы которого $\mathbf{u}_n \in \mathbb{R}^{D_u}$ конструируются из наблюдаемых данных \mathbf{X} (конкретные способы построения переменных \mathbf{u} рассмотрены ниже в этой и следующих главах). Очевидно, что в общем случае не существует однозначного отображения (т.е. детерминированного оператора эволюции) $\mathbf{Q} : \mathbf{u}_n \rightarrow \mathbf{u}_{n+1}$ в таком подпространстве. Возникающая при этом неоднозначность отображения может быть описана стохастически посредством

определения \mathbf{Q} как оператора эволюции случайной динамической системы [11]. Для практического применения (см. детали в работе [85]) случайный оператор эволюции можно строить в следующем достаточно общем виде:

$$\mathbf{u}_{n+1} = \mathbf{f}(\mathbf{u}_n) + \hat{\mathbf{g}}(\mathbf{u}_n) \cdot \boldsymbol{\xi}_n, \quad (1.1)$$

где отображение $\mathbf{f} : \mathbb{R}^{D_u} \rightarrow \mathbb{R}^{D_u}$ представляет собой «детерминированную» часть модели; $\boldsymbol{\xi}_n \in \mathbb{R}^{D_u}$ – нормальный дельта-коррелированный случайный процесс, компоненты которого взаимно независимы и имеют нулевое среднее и единичную дисперсию; матричная функция $\hat{\mathbf{g}} : \mathbb{R}^{D_u} \rightarrow \mathcal{L}_{D_u}(\mathbb{R})$ зависит от состояния \mathbf{u}_n и вместе с $\boldsymbol{\xi}_n$ образует стохастическую часть модели – случайное возмущение, неоднородно распределенное в фазовом пространстве модели. Здесь $\mathcal{L}_{D_u}(\mathbb{R})$ – множество нижних треугольных матриц размерности d над действительным полем. Очевидно, множество $\mathcal{L}_{D_u}(\mathbb{R})$ является достаточным для задания произвольной положительно определенной ковариационной матрицы \mathbf{G} случайного возмущения в силу существования и единственности разложения Холецкого $\mathbf{G} = \hat{\mathbf{g}}\hat{\mathbf{g}}^T$ [51].

Физический смысл стохастической части в (1.1) состоит в том, что она моделирует действие процессов с малым временем автокорреляции, которые плохо разрешены в наблюдаемом временном ряде и не могут быть корректно описаны детерминированной частью модели. В отличие от традиционного описания стохастической части в виде Гауссова шума с постоянной ковариационной матрицей [86, 106], модель (1.1) является, вообще говоря, более эффективной: в областях фазового пространства, где наблюдаемая зависимость текущего состояния от предыдущего определена с меньшей неоднозначностью, оператор эволюции может быть реконструирован более точно за счет меньшей стохастической части модели, уменьшая, таким образом, систематическую погрешность реконструкции. В работах [85, 89] эффективность модели (1.1) продемонстрирована на примерах реконструкции систем различной степени сложности.

Ключевым утверждением, на основе которого будут строиться переменные \mathbf{u} для модели оператора эволюции (1.1) в данном разделе, является теорема Такенса о вложении [116], указывающая метод построения фазовых переменных в достаточно идеализированном случае. Согласно этой теореме, для динамической системы с гладким детерминированным оператором эволюции φ , действующим на компактном многообразии S размерности s , и для одномерной измеряемой величины x , являющейся гладкой функцией на M , в типичном случае отображение $\Phi_x : S \rightarrow \mathbb{R}^{2s+1}$ следующего вида

$$\Phi_x(\cdot) = (x(\varphi^{2s}(\cdot)), \dots, x(\varphi(\cdot)), x(\cdot)) \quad (1.2)$$

является гладким вложением многообразия S в пространство \mathbb{R}^{2s+1} (здесь φ^k обозначает k -кратное применение φ), или, говоря более простым языком, переменные

$$\mathbf{u}(t) := (x(t), x(t - \tau), \dots, x(t - (l - 1)\tau)) \quad (1.3)$$

задают топологически корректную копию фазового пространства системы при $l \geq 2s + 1$ и почти при любом значении задержки τ . Из этой теоремы, например, следует, что по одномерному дискретному временному ряду наблюдений за системой, находящейся на хаотическом аттракторе, принадлежащем многообразию размерности s , можно получить реконструкцию самого аттрактора в пространстве \mathbb{R}^{2s+1} . Подобный (1.3) метод производства переменных называется методом переменных с задержками [6]. Существует также обобщение теоремы Такенса [34], позволяющее строить переменные методом задержек по аналогии с (1.3) в случае векторной (многомерной) измеряемой величины \mathbf{x} :

$$\mathbf{u}(t) := (\mathbf{x}(t), \mathbf{x}(t - \tau), \dots, \mathbf{x}(t - (l - 1)\tau)) \quad (1.4)$$

Фактически, теорема Такенса и ее обобщение аналогичны общему утверждению, следующему из теоремы Уитни о вложении [128]: почти любое гладкое отображение многообразия S в пространство \mathbb{R}^{2s+1} является вложением. Несмотря на то, что, как уже было отмечено, при работе с наблюдаемыми временными рядами конечной длины размерность фазового пространства неизвестна (и, как правило, велика), метод переменных с задержками может давать удачную конечномерную проекцию этого пространства для построения стохастической модели (1.1). Качество такой модели можно оценить на основе анализа наблюдаемых данных с помощью байесова подхода, излагаемого в разделе 1.3.1.

Для дискретного временного ряда \mathbf{X} при величине задержки τ , равной одному отсчету, выражение (1.4) для переменной \mathbf{u} принимает вид:

$$\mathbf{u}_k := (\mathbf{x}_k, \mathbf{x}_{k-1}, \dots, \mathbf{x}_{k-(l-1)}) . \quad (1.5)$$

где l – количество задержанных переменных, $\mathbf{u}_k \in \mathbb{R}^{l \cdot D}$. Форма модели (1.1) в этом случае может быть значительно упрощена благодаря наличию $(l - 1) \cdot D$ тривиальных соотношений между переменными:

$$\begin{aligned} (u_1, \dots, u_D)_{n+1} &= \mathbf{f}(\mathbf{u}_n; \boldsymbol{\mu}_f) + \hat{\mathbf{g}}(\mathbf{u}_n; \boldsymbol{\mu}_g) \cdot \boldsymbol{\xi}_n, \\ (u_{D+1}, \dots, u_{l \cdot D})_{n+1} &= (u_1, \dots, u_{(l-1) \cdot D})_n \end{aligned} \quad (1.6)$$

Здесь отображения $\mathbf{f}(\cdot; \boldsymbol{\mu}_f) : \mathbb{R}^{l \cdot D} \rightarrow \mathbb{R}^D$ и $\hat{\mathbf{g}}(\cdot; \boldsymbol{\mu}_g) : \mathbb{R}^{l \cdot D} \rightarrow \mathcal{L}_D(\mathbb{R})$ имеют меньшую размерность выхода (D вместо $l \cdot D$) и снабжены дополнительно аргументами $\boldsymbol{\mu}_f$ и $\boldsymbol{\mu}_g$, показывающими, что они будут явно параметризованы определенным классом аппроксимирующих функций; $\boldsymbol{\xi}_n \in \mathbb{R}^D$ – нормальный дельта-коррелированный случайный

процесс. Второе уравнение в (1.6) выполняется по определению (1.5), т.е. переменные $\mathbf{x}_{k-1}, \dots, \mathbf{x}_{k-(l-1)}$ в момент времени t_{n+1} являются переменными $\mathbf{x}_k, \dots, \mathbf{x}_{k-(l-2)}$ в предыдущий момент времени t_n .

Наконец, выражение для модели (1.6) удобно записать, введя обозначение $\tilde{\mathbf{u}}_k = (u_1, \dots, u_D)_k$:

$$\begin{aligned} \tilde{\mathbf{u}}_{n+1} &= \mathbf{f}(\tilde{\mathbf{u}}_n, \tilde{\mathbf{u}}_{n-1}, \dots, \tilde{\mathbf{u}}_{n-l+1}; \boldsymbol{\mu}_f) + \hat{\mathbf{g}}(\tilde{\mathbf{u}}_n, \tilde{\mathbf{u}}_{n-1}, \dots, \tilde{\mathbf{u}}_{n-l+1}; \boldsymbol{\mu}_g) \cdot \boldsymbol{\xi}_n, \\ n &= l, \dots, N-1. \end{aligned} \quad (1.7)$$

Заметим, что величины $\tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_l$ не получаются в результате применения оператора эволюции и лишь играют роль начального состояния системы (т.е. параметров), что будет учтено при формулировании критерия выбора оптимальной модели в следующем разделе.

Соответствующее обратное отображение фазового пространства в пространство наблюдаемых переменных в нашем случае определяется простым соотношением:

$$\mathbf{x}_n = \tilde{\mathbf{u}}_n, \quad n = 1, \dots, N. \quad (1.8)$$

Уравнения для модели (1.7)-(1.8) можно записать в виде:

$$\begin{aligned} \mathbf{x}_{n+1} &= \mathbf{f}(\mathbf{x}_n, \mathbf{x}_{n-1}, \dots, \mathbf{x}_{n-l+1}; \boldsymbol{\mu}_f) + \hat{\mathbf{g}}(\mathbf{x}_n, \mathbf{x}_{n-1}, \dots, \mathbf{x}_{n-l+1}; \boldsymbol{\mu}_g) \cdot \boldsymbol{\xi}_n, \\ n &= l, \dots, N-1, \end{aligned} \quad (1.9)$$

имея в виду, что начальное состояние модели определяется равенством $\mathbf{x}_n = \tilde{\mathbf{u}}_n$ для $n = 1, \dots, l$.

В данной работе для параметризации функций \mathbf{f} и $\hat{\mathbf{g}}$ будет использоваться ИНС в форме персептрона с одним скрытым слоем [52]:

$$\mathbf{A}_{d_{in}, d_{out}}^m(\mathbf{z}) = \sum_{i=1}^m \boldsymbol{\alpha}_i \cdot \tanh(\boldsymbol{\omega}_i^T \mathbf{z} + \gamma_i). \quad (1.10)$$

Здесь d_{in} и d_{out} – размерности входа и выхода ИНС, соответственно, т.е. $\boldsymbol{\alpha}_i \in \mathbb{R}^{d_{out}}$, $\boldsymbol{\omega}_i \in \mathbb{R}^{d_{in}}$, $\mathbf{z} \in \mathbb{R}^{d_{in}}$; m – количество нейронов в скрытом слое, определяющее точность аппроксимации; $\boldsymbol{\omega}_i^T \mathbf{z}$ – скалярное произведение векторов-столбцов $\boldsymbol{\omega}_i$ и \mathbf{z} ; $\gamma_i \in \mathbb{R}$.

ИНС являются эффективными универсальными аппроксиматорами неизвестных зависимостей [52], их эффективность при построении прогностических моделей по временным рядам показана во множестве работ [40, 44, 85, 87, 129, 136]. Удобство использования ИНС заключается в том, что они позволяют существенно увеличить размерность и сложность модели, увеличив число параметров лишь немного. При этом, недостатком является вырождение в пространстве параметров ИНС. Этот недостаток можно преодолеть, используя методы регуляризации; в частности, при использовании

байесова подхода (раздел 1.3.1) регуляризация заключается во введении априорных ограничений на параметры ИНС.

Функции \mathbf{f} и $\hat{\mathbf{g}}$ в модели (1.9) представляются посредством ИНС (1.10) с разными количествами нейронов m_f и m_g :

$$\begin{aligned} \mathbf{f}(\mathbf{u}; \boldsymbol{\mu}_f) &= \mathbf{A}_{l,D,D}^{m_f}(\mathbf{u}), \\ \hat{\mathbf{g}}(\mathbf{u}; \boldsymbol{\mu}_g) &= \begin{cases} \mathit{low_triang}(\mathbf{A}_{l,D,D \cdot (D+1)/2}^{m_g}(\mathbf{u})) & \text{при } m_g \geq 1, \\ \mathit{low_triang}(\boldsymbol{\mu}_g) & \text{при } m_g = 0. \end{cases} \end{aligned} \quad (1.11)$$

Здесь под параметрами $\boldsymbol{\mu}_f$ и $\boldsymbol{\mu}_g$ из левой части понимаются наборы коэффициентов ИНС $\{\boldsymbol{\alpha}_i, \boldsymbol{\omega}_i, \gamma_i\}_{i=1}^{m_f}$ и $\{\boldsymbol{\alpha}_i, \boldsymbol{\omega}_i, \gamma_i\}_{i=1}^{m_g}$, параметризующие функции \mathbf{f} и $\hat{\mathbf{g}}$ соответственно. Функция $\mathit{low_triang} : \mathbb{R}^{D \cdot (D+1)/2} \rightarrow \mathcal{L}_D(\mathbb{R})$ обозначает процедуру записи компонент вектора размерности $D \cdot (D+1)/2$ в нижнюю треугольную матрицу размерности D , т.е. $(\mathit{low_triang}(\mathbf{z}))_{ij} = \mathbf{z}_{j+i(i-1)/2}$ для всех $1 \leq j \leq i \leq D$.

Кроме того, в (1.11) в виде обозначения $m_g = 0$ учтен случай «традиционной» стохастической части модели (1.9) с постоянной ковариационной матрицей, когда $\hat{\mathbf{g}} = \mathit{const}$. Одним из важных вопросов, возникающих при использовании модели (1.9) является следующий: насколько статистически обосновано увеличение числа параметров модели за счет введения дополнительной функции $\hat{\mathbf{g}}(\mathbf{u})$ при анализе конкретного временного ряда \mathbf{X} конечной длины? Этот вопрос является частным случаем более общего вопроса об определении оптимальной сложности модели, ответ на который в данной работе дается с помощью байесова критерия оптимальности, изложенного в следующем разделе.

1.3 Выбор оптимальной модели

1.3.1 Байесов подход

Определение байесовой обоснованности

Из (1.9) – (1.11) ясно, что сложность модели (которая, вообще говоря, характеризуется общим количеством параметров модели) для выбранной функциональной формы модели определяется тремя структурными параметрами (или гиперпараметрами): количеством нейронов m_f и m_g в детерминированной и стохастической частях модели и количеством задержанных переменных l , которое пропорционально размерности фазового пространства $D_u = l \cdot D$. Таким образом, проблема выбора оптимальной сложности модели может быть формализована: необходимо найти точку в пространстве структурных параметров, соответствующую наилучшей, в смысле определенного

критерия оптимальности, модели. В данной работе предлагается критерий оптимальности, основанный на байесовом подходе.

Пусть задано пространство моделей $\mathbf{H} = \{H_i\}$. В нашем случае каждая модель H_i задается структурными параметрами $(m_f, m_g, l)_i$. Согласно байесову подходу, мерой оптимальности модели H_i является вероятность того, что эта модель является источником наблюдаемого временного ряда \mathbf{X} . В соответствии с теоремой Байеса, эта вероятность равна:

$$\mathcal{P}(H_i|\mathbf{X}) = \frac{\mathcal{P}(\mathbf{X}|H_i)\mathcal{P}(H_i)}{\sum_i \mathcal{P}(\mathbf{X}|H_i)\mathcal{P}(H_i)}. \quad (1.12)$$

Здесь $\mathcal{P}(\mathbf{X}|H_i)$ – байесова обоснованность модели H_i [54], т.е. вероятность временного ряда \mathbf{X} в ансамбле всевозможных временных рядов, генерируемых моделью H_i (более точно – плотность вероятности (ПВ) в пространстве таких рядов, вычисленная на временном ряде \mathbf{X}). Множитель $\mathcal{P}(H_i)$ – априорная вероятность модели H_i , отражающая априорные знания о выборе данной модели по отношению к другим моделям из \mathbf{H} . Знаменатель в (1.12) играет роль нормировочного члена и не зависит от H_i .

Далее, считая (предполагая), что все модели из \mathbf{H} априори равновероятны, можно упростить (1.12) до $\mathcal{P}(H_i|\mathbf{X}) = \alpha \mathcal{P}(\mathbf{X}|H_i)$, где α не зависит от H_i . Таким образом, оптимальная модель соответствует максимуму байесовой обоснованности $\mathcal{P}(\mathbf{X}|H_i)$ по H_i .

Почему максимизация байесовой обоснованности позволяет выбрать модель оптимальной сложности? Качественно ответ на этот вопрос выглядит следующим образом. «Слишком простая» модель может с высокой вероятностью генерировать лишь малое подмножество всех временных рядов, в которое наблюдаемый ряд \mathbf{X} почти наверняка не входит, и поэтому величина $\mathcal{P}(\mathbf{X}|H_i)$ для нее мала. «Слишком сложная» модель, напротив, может с примерно одинаковой вероятностью генерировать слишком большое количество временных рядов, в том числе и ряд \mathbf{X} , и поэтому вероятность сгенерировать именно этот ряд, а не другие, для нее мала. Такую модель еще называют «переобученной», потому что она, за счет большого количества параметров, воспроизводит конкретную реализацию шума («быстрых процессов») и, вследствие этого, не способна воспроизводить динамику системы за пределами обучающей выборки. Отсюда следует, что существует не слишком простая и не слишком сложная модель – модель оптимальной сложности.

Проведем вывод аналитического выражения для байесовой обоснованности стохастической модели (1.9) – (1.11).

Аналитическое выражение для байесовой обоснованности стохастической модели

Индивидуальная модель H_i (уравнения (1.9) – (1.11)) полностью задается параметрами $\boldsymbol{\mu}_f, \boldsymbol{\mu}_g$ и вектором $\mathbf{u}_{init} = (\tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_l)$ начального состояния. Обозначим их через вектор параметров $\boldsymbol{\theta} = (\boldsymbol{\mu}_f, \boldsymbol{\mu}_g, \mathbf{u}_{init})$. Тогда байесова обоснованность $\mathcal{P}(\mathbf{X}|H_i)$ модели H_i может быть выражена через функцию правдоподобия $\mathcal{P}(\mathbf{X}|\boldsymbol{\theta}, H_i)$ и априорную вероятностную меру \mathcal{P}_{Θ_i} , введенную в пространстве Θ_i параметров $\boldsymbol{\theta}$ модели H_i :

$$\mathcal{P}(\mathbf{X}|H_i) = \int_{\Theta_i} \mathcal{P}(\mathbf{X}|\boldsymbol{\theta}, H_i) d\mathcal{P}_{\Theta_i}, \boldsymbol{\theta} \in \Theta_i. \quad (1.13)$$

С учетом уравнений (1.7)-(1.8), определяющих форму модели, функция правдоподобия $\mathcal{P}(\mathbf{X}|\boldsymbol{\theta}, H_i)$ может быть факторизована следующим образом:

$$\begin{aligned} \mathcal{P}(\mathbf{X}|\boldsymbol{\theta}, H_i) = & \int \left[\prod_{n=1}^N \delta(\mathbf{x}_n - \tilde{\mathbf{u}}_n) \times \right. \\ & \left. \times \prod_{n=l+1}^N \mathcal{P}(\tilde{\mathbf{u}}_n | \tilde{\mathbf{u}}_{n-1}, \dots, \tilde{\mathbf{u}}_{n-l}; \boldsymbol{\mu}_f, \boldsymbol{\mu}_g) \right] d\tilde{\mathbf{u}}_{l+1} \dots d\tilde{\mathbf{u}}_N. \end{aligned} \quad (1.14)$$

Далее, учитывая предположение о Гауссовости и дельта-коррелированности шума $\boldsymbol{\xi}_n$ в (1.9), имеем:

$$\begin{aligned} \mathcal{P}(\tilde{\mathbf{u}}_n | \tilde{\mathbf{u}}_{n-1}, \dots, \tilde{\mathbf{u}}_{n-l}, \boldsymbol{\mu}_f, \boldsymbol{\mu}_g) = & \det [2\pi \mathbf{G}_n(l, \boldsymbol{\mu}_g)]^{-1/2} \times \\ & \times \exp \left\{ -\frac{1}{2} (\tilde{\mathbf{u}}_n - \mathbf{f}_n(l, \boldsymbol{\mu}_f))^T \cdot \mathbf{G}_n^{-1}(l, \boldsymbol{\mu}_g) \cdot (\tilde{\mathbf{u}}_n - \mathbf{f}_n(l, \boldsymbol{\mu}_f)) \right\}, \end{aligned} \quad (1.15)$$

где

$$\begin{aligned} \mathbf{f}_n(l, \boldsymbol{\mu}_f) & := \mathbf{f}(\tilde{\mathbf{u}}_{n-1}, \dots, \tilde{\mathbf{u}}_{n-l}; \boldsymbol{\mu}_f) \\ \mathbf{G}_n^{-1}(l, \boldsymbol{\mu}_g) & := \hat{\mathbf{g}}(\tilde{\mathbf{u}}_{n-1}, \dots, \tilde{\mathbf{u}}_{n-l}; \boldsymbol{\mu}_g) \cdot \hat{\mathbf{g}}^T(\tilde{\mathbf{u}}_{n-1}, \dots, \tilde{\mathbf{u}}_{n-l}; \boldsymbol{\mu}_g) \end{aligned} \quad (1.16)$$

Вероятностную меру \mathcal{P}_{Θ_i} зададим, введя априорные ПВ параметров $\boldsymbol{\theta}$:

$$d\mathcal{P}_{\Theta_i} = \mathcal{P}_{pr}(\boldsymbol{\mu}_f, \boldsymbol{\mu}_g | H_i, \mathbf{q}) d\boldsymbol{\mu}_f d\boldsymbol{\mu}_g \cdot \mathcal{P}_{pr}(\mathbf{u}_{init} | H_i) d\mathbf{u}_{init}. \quad (1.17)$$

В выражении (1.17) $\mathcal{P}_{pr}(\boldsymbol{\mu}_f, \boldsymbol{\mu}_g | H_i, \mathbf{q})$ – априорная ПВ коэффициентов $\boldsymbol{\mu}_f, \boldsymbol{\mu}_g$ функций $\mathbf{f}, \hat{\mathbf{g}}$. В ней использован дополнительный гиперпараметр \mathbf{q} , смысл которого будет пояснен ниже при записи $\mathcal{P}_{pr}(\boldsymbol{\mu}_f, \boldsymbol{\mu}_g | H_i, \mathbf{q})$ в явном виде. На самом деле данный гиперпараметр может рассматриваться как идентификатор модели H_i наряду с $(m_f, m_g, l)_i$, но из соображений удобства изложения он отделен от H_i . С учетом этого, будем далее обозначать обоснованность как $\mathcal{P}(\mathbf{X}|H_i, \mathbf{q})$ вместо $\mathcal{P}(\mathbf{X}|H_i)$.

Для начального состояния \mathbf{u}_{init} естественно использовать нормальное априорное распределение с единичной дисперсией, полагая здесь и далее в разделе, что наблюдаемый временной ряд \mathbf{X} на время построения модели был приведен к единичной

вариации¹:

$$\mathcal{P}_{pr}(\mathbf{u}_{init}|H_i) = \frac{1}{(2\pi)^{l \cdot D/2}} \exp \left\{ -\frac{1}{2} \sum_{n=1}^l |\tilde{\mathbf{u}}_n|^2 \right\}. \quad (1.18)$$

Подставив (1.14)-(1.18) в (1.13) и проинтегрировав по переменным $\tilde{\mathbf{u}}_1, \dots, \tilde{\mathbf{u}}_N$, получим аналитическое выражение для байесовой обоснованности:

$$\begin{aligned} \mathcal{P}(\mathbf{X}|H_i, \mathbf{q}) &= (2\pi)^{-\frac{l \cdot D}{2}} \cdot \exp \left\{ -\frac{1}{2} \sum_{n=1}^l |\mathbf{x}_n|^2 \right\} \times \\ &\times \int \Lambda(\boldsymbol{\mu}_f, \boldsymbol{\mu}_g, \mathbf{X}; H_i) \mathcal{P}_{pr}(\boldsymbol{\mu}_f, \boldsymbol{\mu}_g; H_i, \mathbf{q}) d\boldsymbol{\mu}_f d\boldsymbol{\mu}_g, \end{aligned} \quad (1.19)$$

где, в рамках обозначений (1.16):

$$\begin{aligned} \Lambda(\boldsymbol{\mu}_f, \boldsymbol{\mu}_g, \mathbf{X}; H_i) &= \left[\prod_{n=l+1}^N \det [2\pi \mathbf{G}_n(l, \boldsymbol{\mu}_g)] \right]^{-1/2} \times \\ &\times \exp \left\{ -\frac{1}{2} \sum_{n=l+1}^N (\mathbf{x}_n - \mathbf{f}_n(l, \boldsymbol{\mu}_f))^T \cdot \mathbf{G}_n^{-1}(l, \boldsymbol{\mu}_g) \cdot (\mathbf{x}_n - \mathbf{f}_n(l, \boldsymbol{\mu}_f)) \right\}. \end{aligned} \quad (1.20)$$

Априорное распределение коэффициентов ИНС

Смысл априорной ПВ $\mathcal{P}_{pr}(\boldsymbol{\mu}_f, \boldsymbol{\mu}_g|H_i, \mathbf{q})$ состоит в том, что она ограничивает область определения модели в пространстве параметров. Введение данной ПВ помогает скомпенсировать вырождение модели в пространстве параметров, являющееся, как уже было сказано, недостатком ИНС. При этом область для поиска параметров (обучения) модели становится ограниченной, упрощая численный анализ. Для удобства вычислений, следуя [92], будем использовать Гауссову форму априорной ПВ, введя для каждой группы параметров ИНС (1.10) соответствующую дисперсию:

$$\begin{aligned} \mathcal{P}_{ANN}(\{\boldsymbol{\alpha}_i, \boldsymbol{\omega}_i, \gamma_i\}_{i=1}^m; \sigma_\alpha, \sigma_\omega, \sigma_\gamma) &= \\ &= \prod_{i=1}^m (2\pi\sigma_\alpha^2)^{-\frac{d_{out}}{2}} \exp \left(-\frac{|\boldsymbol{\alpha}_i|^2}{2\sigma_\alpha^2} \right) \times \\ &\times \prod_{i=1}^m (2\pi\sigma_\omega^2)^{-\frac{d_{in}}{2}} \exp \left(-\frac{|\boldsymbol{\omega}_i|^2}{2\sigma_\omega^2} \right) \times \\ &\times \prod_{i=1}^m (2\pi\sigma_\gamma^2)^{-\frac{1}{2}} \exp \left(-\frac{\gamma_i^2}{2\sigma_\gamma^2} \right). \end{aligned} \quad (1.21)$$

Под гиперпараметрами \mathbf{q} в $\mathcal{P}_{pr}(\boldsymbol{\mu}_f, \boldsymbol{\mu}_g|H_i, \mathbf{q})$ будем понимать набор априорных стандартных отклонений параметров ИНС, определяющих Гауссово распределение. Фактически, гиперпараметры \mathbf{q} определяют объем области определения модели в пространстве параметров и, таким образом, позволяют управлять сложностью (мощностью) модели, наряду с гиперпараметрами $(m_f, m_g, l)_i$.

¹Вариацией здесь и далее называется средний (по времени) квадрат отклонения величины от своего среднего (по времени) значения

В итоге для параметров $\boldsymbol{\mu}_f, \boldsymbol{\mu}_g$ имеем в случае $m_g \geq 1$:

$$\begin{aligned} \mathcal{P}_{pr}(\boldsymbol{\mu}_f, \boldsymbol{\mu}_g | H_i, \mathbf{q}) &= \mathcal{P}_{ANN}(\boldsymbol{\mu}_f; \sigma_\alpha^f, \sigma_\omega^f, \sigma_\gamma^f) \cdot \mathcal{P}_{ANN}(\boldsymbol{\mu}_g; \sigma_\alpha^g, \sigma_\omega^g, \sigma_\gamma^g), \\ \mathbf{q} &= (\sigma_\alpha^f, \sigma_\omega^f, \sigma_\gamma^f, \sigma_\alpha^g, \sigma_\omega^g, \sigma_\gamma^g), \end{aligned} \quad (1.22)$$

и в случае однородной стохастической части ($m_g = 0$):

$$\begin{aligned} \mathcal{P}_{pr}(\boldsymbol{\mu}_f, \boldsymbol{\mu}_g | H_i, \mathbf{q}) &= \mathcal{P}_{ANN}(\boldsymbol{\mu}_f; \sigma_\alpha^f, \sigma_\omega^f, \sigma_\gamma^f) \cdot \prod_{i=1}^{D(D+1)/2} \mathcal{U}_{[-\sigma_{\mu_g}, \sigma_{\mu_g}]}(\mu_{g_i}), \\ \mathbf{q} &= (\sigma_\alpha^f, \sigma_\omega^f, \sigma_\gamma^f, \sigma_{\mu_g}), \end{aligned} \quad (1.23)$$

Здесь $\mathcal{U}_{[a,b]}(z)$ обозначает ПВ однородного распределения величины z на отрезке $[a, b]$. В (1.23), таким образом, априорное распределение коэффициентов матрицы $\hat{\mathbf{g}}$ является однородным. Это связано с тем, что в случае однородной стохастической части задача выбора модели является хорошо обусловленной по отношению к $\hat{\mathbf{g}}$ и не требует существенной регуляризации, как показано ниже при формулировке алгоритма построения модели.

Ценовая функция для поиска оптимальной модели

Через байесову обоснованность $\mathcal{P}(X | H_i, \mathbf{q})$ удобно выразить длину описания данных [104] с помощью модели (H_i, \mathbf{q}) , т.е. количество информации, необходимое для кодирования данных \mathbf{X} с использованием модели (H_i, \mathbf{q}) :

$$I(\mathbf{X}) = -\ln \mathcal{P}(\mathbf{X} | H_i, \mathbf{q}). \quad (1.24)$$

Тогда минимум функции (1.24) по \mathbf{q} (максимум байесовой обоснованности) будет являться критерием оптимальности модели H_i , т.е. гиперпараметров $(m_f, m_g, l)_i$:

$$E(H_i) = \min_{\mathbf{q}} [-\ln \mathcal{P}(\mathbf{X} | H_i, \mathbf{q})]. \quad (1.25)$$

Наконец, модель H_i , имеющая наибольшую вероятность сгенерировать наблюдаемый временной ряд \mathbf{X} (байесову обоснованность), или, что то же самое, соответствующая наименьшему значению $E(H_i)$, т.е. обладающая наименьшей длиной описания данных (принцип минимальной длины описания), считается оптимальной. Заметим, что при этом произойдет самосогласованное определение характеристик \mathbf{q} априорного распределения $\mathcal{P}_{pr}(\boldsymbol{\mu}_f, \boldsymbol{\mu}_g | H_i, \mathbf{q})$ параметров модели H_i , т.е. оптимальной регуляризации.

Вычисление байесовой обоснованности

Точное вычисление интеграла (1.19) является чрезвычайно трудоемкой задачей ввиду большой размерности пространства подынтегральных аргументов и сложного

устройства подынтегрального выражения (наличием, вообще говоря, нескольких локальных максимумов), и поэтому в большинстве случаев невозможно. Существуют различные методы приближенного вычисления таких интегралов: метод Лапласа [37], метод сэмплирования Гиббса [42], EM-алгоритм [32]. В данной работе в качестве оценки данного интеграла мы будем использовать наибольшее из значений, полученных методом Лапласа, т.е. приближенным интегрированием в окрестности того или иного локального максимума подынтегрального выражения (по аналогии с тем, как это делалось в [81] для латентных переменных).

Для этого рассмотрим минус логарифм подынтегрального выражения в (1.19) как функцию переменных $(\boldsymbol{\mu}_f, \boldsymbol{\mu}_g)$:

$$\Psi_{\mathbf{X}, H_i, \mathbf{q}}(\boldsymbol{\mu}_f, \boldsymbol{\mu}_g) = -\ln [\Lambda(\boldsymbol{\mu}_f, \boldsymbol{\mu}_g, \mathbf{X}; H_i) \cdot \mathcal{P}_{pr}(\boldsymbol{\mu}_f, \boldsymbol{\mu}_g | H_i, \mathbf{q})], \quad (1.26)$$

Нетрудно видеть (см. выражение (1.20)), что значение данной функции в точке локального минимума по порядку величины сравнимо с величиной $ND/2$, т.е. оно много больше единицы. Поэтому разложение этой функции в ряд Тейлора до второго порядка в окрестности параметров $(\boldsymbol{\mu}_{f0}, \boldsymbol{\mu}_{g0})$, соответствующих ее локальному минимуму, позволяет приближенно вычислить минус логарифм интеграла (1.19), взятого в окрестности точки $(\boldsymbol{\mu}_{f0}, \boldsymbol{\mu}_{g0})$:

$$\begin{aligned} \Phi(\mathbf{X}, H_i, \mathbf{q}, \boldsymbol{\mu}_{f0}, \boldsymbol{\mu}_{g0}) &= \\ &= \Psi_{\mathbf{X}, H_i, \mathbf{q}}(\boldsymbol{\mu}_{f0}, \boldsymbol{\mu}_{g0}) + \frac{1}{2} \ln \det \frac{1}{2\pi} \nabla \nabla^T \Psi_{\mathbf{X}, H_i, \mathbf{q}} + \frac{1}{2} \sum_{n=1}^l (|\mathbf{x}_n|^2 + D \cdot \ln 2\pi). \end{aligned} \quad (1.27)$$

Здесь $\nabla \nabla^T \Psi_{\mathbf{X}, H_i, \mathbf{q}}$ – матрица вторых производных (гессиан) функции $\Psi_{\mathbf{X}, H_i, \mathbf{q}}$ по параметрам $(\boldsymbol{\mu}_f, \boldsymbol{\mu}_g)$ в точке минимума $(\boldsymbol{\mu}_{f0}, \boldsymbol{\mu}_{g0})$. Фактически, выражение (1.27) является оценкой ценовой функции для оптимальности модели (1.24) в случае, когда локальный минимум является единственным минимумом функции (1.26). Заметим, что выражение (1.27) также по порядку величины сравнимо с $ND/2 \gg 1$. Поэтому в случае, когда у функции (1.26) имеется несколько локальных минимумов, их вклады в интеграл (1.19), оцениваемые как $\exp(-\Phi(\mathbf{X}, H_i, \mathbf{q}, \boldsymbol{\mu}_{f0}, \boldsymbol{\mu}_{g0}))$, как правило, существенно (на порядки) отличаются друг от друга, и только один из них вносит определяющий вклад в интеграл (1.19). Исходя из этого предположения, при наличии нескольких локальных минимумов в качестве оценки ценовой функции для оптимальности (1.24) будем брать выражение (1.27) при том локальном минимуме, при котором оно является наименьшим (а вклад в интеграл – наибольшим).

Таким образом, выражение (1.27) позволяет оценить степень оптимальности модели, если нужная точка минимума $(\boldsymbol{\mu}_{f0}, \boldsymbol{\mu}_{g0})$ найдена. Само наличие оптимума основано на конкуренции первого и второго слагаемых в (1.27). Первое слагаемое

$\Psi_{\mathbf{X}, H_i, \mathbf{q}}(\boldsymbol{\mu}_{f0}, \boldsymbol{\mu}_{g0})$ характеризует ошибку описания конкретной выборки \mathbf{X} моделью и убывает с увеличением сложности модели, т.е. количества нейронов в ИНС и величины задержки l . Второе слагаемое $\frac{1}{2} \ln \det \frac{1}{2\pi} \nabla \nabla^T \Psi_{\mathbf{X}, H_i, \mathbf{q}}$ в (1.27), напротив, предотвращает «переобучение», свойственное сложной модели, поскольку это слагаемое растет с ростом общего количества M параметров $(\boldsymbol{\mu}_f, \boldsymbol{\mu}_g)$. Другими словами, первое слагаемое штрафует слишком простые модели, не способные захватить динамические свойства системы, а второе слагаемое штрафует слишком сложные модели, которым свойственно воспроизводить конкретную реализацию шума вместо динамических свойств системы. Последним слагаемым в окрестности оптимума, когда $l \ll N$, можно пренебречь, а при больших l оно действует аналогично второму слагаемому.

Чтобы пояснить поведение второго слагаемого, заметим, что в асимптотическом случае $N \gg M$ оно оценивается как $\frac{1}{2} \bar{M} \ln N$, где \bar{M} – эффективное количество «невыврожденных» параметров модели (1.9); это следует из того, что собственные значения гессиана $\nabla \nabla^T \Psi_{\mathbf{X}, H_i, \mathbf{q}}$, соответствующие сингулярным (вырожденным) направлениям в пространстве параметров, не зависят от N и определяются сугубо априорной ПВ, в то время как все остальные собственные значения гессиана по порядку величины сравнимы с N и дают главный вклад в определитель. Также в случае $N \gg M$ последним слагаемым заведомо можно пренебречь. Таким образом, критерий оптимальности в случае стремящейся к бесконечности длины временного ряда аналогичен известному критерию Шварца [110], в котором штрафной член оценивается как $\frac{1}{2} M \ln N$ и который слишком сильно штрафует модель при наличии вырождения, свойственного ИНС.

Алгоритм построения оптимальной модели

В рамках байесова подхода обучение модели (H_i, \mathbf{q}) состоит в максимизации по $(\boldsymbol{\mu}_f, \boldsymbol{\mu}_g)$ апостериорного распределения параметров ИНС, определяемого следующим выражением:

$$\mathcal{P}(\boldsymbol{\mu}_f, \boldsymbol{\mu}_g | \mathbf{X}, H_i, \mathbf{q}) = \int \mathcal{P}(\boldsymbol{\theta} | \mathbf{X}, H_i, \mathbf{q}) d\mathbf{u}_{init}, \quad (1.28)$$

которое, учитывая теорему Байеса для данного случая $\mathcal{P}(\boldsymbol{\theta} | \mathbf{X}, H_i, \mathbf{q}) = \mathcal{P}(\mathbf{X} | \boldsymbol{\theta}, H_i) \cdot \mathcal{P}_{pr}(\boldsymbol{\mu}_f, \boldsymbol{\mu}_g | H_i, \mathbf{q}) \cdot \mathcal{P}_{pr}(\mathbf{u}_{init} | H_i) / \mathcal{P}(\mathbf{X} | H_i, \mathbf{q})$, может быть записано в виде:

$$\mathcal{P}(\boldsymbol{\mu}_f, \boldsymbol{\mu}_g | \mathbf{X}, H_i, \mathbf{q}) \propto \int \mathcal{P}(\mathbf{X} | \boldsymbol{\theta}, H_i) \cdot \mathcal{P}_{pr}(\boldsymbol{\mu}_f, \boldsymbol{\mu}_g | H_i, \mathbf{q}) \cdot \mathcal{P}_{pr}(\mathbf{u}_{init} | H_i) d\mathbf{u}_{init}. \quad (1.29)$$

Несложно показать, путем подстановки (1.14)-(1.18) и интегрирования, что это апостериорное распределение как функция переменных $\boldsymbol{\mu}_f, \boldsymbol{\mu}_g$ пропорционально подынтегральному выражению в (1.19), т.е.:

$$\mathcal{P}(\boldsymbol{\mu}_f, \boldsymbol{\mu}_g | \mathbf{X}, H_i, \mathbf{q}) \propto \exp(-\Psi_{\mathbf{X}, H_i, \mathbf{q}}(\boldsymbol{\mu}_f, \boldsymbol{\mu}_g)). \quad (1.30)$$

Следовательно, поиск точки минимума $(\boldsymbol{\mu}_{f0}, \boldsymbol{\mu}_{g0})$ для вычисления ценовой функции для оптимальности (1.27) является одновременно обучением модели (H_i, \mathbf{q}) . Таким образом, предлагаемый алгоритм для построения оптимальной модели является достаточно простым. Он состоит в повторении итераций, каждая из которых включает следующие шаги:

1. Генерация значений структурных параметров m_f, m_g, l ;
2. Генерация случайного вектора априорных дисперсий \mathbf{q} ;
3. Обучение модели (H_i, \mathbf{q}) и получение точки $(\boldsymbol{\mu}_{f0}, \boldsymbol{\mu}_{g0})$ с помощью градиентного метода минимизации функции (1.26) со случайной стартовой точки;
4. Вычисление ценовой функции для оптимальности (1.27);
5. Многократное изменение гиперпараметров \mathbf{q} с обновлением («доучиванием») точки минимума $(\boldsymbol{\mu}_{f0}, \boldsymbol{\mu}_{g0})$ с помощью градиентного метода минимизации и пересчетом ценовой функции для оптимальности (1.27);
6. Вычисление не зависящей от \mathbf{q} ценовой функции для оптимальности (1.25) посредством выбора наиболее оптимального значения \mathbf{q} из ансамбля, полученного на шаге 5.

Отделение минимизации по \mathbf{q} от минимизации по остальным гиперпараметрам позволяет сэкономить значительные вычислительные ресурсы, потому что многократное доучивание $(\boldsymbol{\mu}_{f0}, \boldsymbol{\mu}_{g0})$, очевидно, быстрее, чем минимизация, стартующая со случайной начальной точки при каждом \mathbf{q} , особенно для близких значений \mathbf{q} .

Таким образом, в результате каждой итерации находятся одновременно оптимальная регуляризация \mathbf{q} , параметры $(\boldsymbol{\mu}_{f0}, \boldsymbol{\mu}_{g0})$ модели H_i при оптимальном \mathbf{q} и величина ценовой функции (1.25), характеризующая байесову обоснованность модели. Наконец, далее находятся оптимальные гиперпараметры m_f, m_g, l путем минимизации ценовой функции $E(H_i)$ по ансамблю всех итераций.

В заключение обратим внимание, что в случае однородной стохастической части ($m_g = 0$) максимизация выражения (1.20), входящего в ценовую функцию (1.26), по $\hat{\mathbf{g}}$ при любом фиксированном $\boldsymbol{\mu}_f$ приводит к ответу $\mathbf{g}\mathbf{g}^T = \frac{1}{N-l} \sum_{n=l+1}^N (\mathbf{x}_n - \mathbf{f}_n(l, \boldsymbol{\mu}_f)) \cdot (\mathbf{x}_n - \mathbf{f}_n(l, \boldsymbol{\mu}_f))^T$, означающему единственность матрицы $\hat{\mathbf{g}}$ в силу единственности разложения Холецкого [51]. Отсюда следует, что решение задачи поиска $\hat{\mathbf{g}}$ в этом случае однозначно определено параметрами $\boldsymbol{\mu}_f$ и не требует существенной регуляризации, в полном соответствии с априорным распределением (1.23).

1.3.2 Пример: система Лоренца-63 с динамическим шумом

Система Лоренца-63

Продemonстрируем предложенный подход на простом примере – реконструкции системы Лоренца-63 [80] с классическими значениями параметров $\tilde{b} = \frac{8}{3}$, $\tilde{\sigma} = 10$, соответствующими наличию хаотического аттрактора при третьем параметре $\tilde{r} = 25$, и дополнительным ланжевенновским источником шума:

$$\begin{aligned}\frac{d\tilde{x}}{dt} &= \tilde{\sigma} (\tilde{y} - \tilde{x}) \\ \frac{d\tilde{y}}{dt} &= \tilde{r}\tilde{x} - \tilde{y} - \tilde{x}\tilde{z} \\ \frac{d\tilde{z}}{dt} &= -\tilde{b}\tilde{z} + \tilde{x}\tilde{y} + \sigma\eta(t)\end{aligned}\tag{1.31}$$

Здесь $\eta(t)$ – белый гауссов шум, параметр σ определяет величину шума. Переменные и параметры классической системы Лоренца-63 помечены тильдой во избежание конфликта с введенными выше обозначениями.

В роли наблюдаемых данных \mathbf{X} рассмотрим скалярный ($D = 1$) временной ряд переменной \tilde{y} , взятый с шагом по времени $\tau = 0.17$:

$$x_n = \tilde{y}(t_0 + n\tau)\tag{1.32}$$

Выбранный шаг $\tau = 0.17$ соответствует первому хорошо выраженному минимуму взаимной информации между сигналами $\tilde{y}(t)$ и $\tilde{y}(t + \tau)$ по τ в случае невозмущенной системы Лоренца-63. Данный критерий выбора τ широко используется [8] в случае использования такенсовых координат. Он ориентирован на наилучшее разрешение топологии аттрактора в реконструируемом фазовом пространстве. Заметим, однако, что данный критерий является очень приближенным, и значение τ , определяющее «хорошую» проекцию аттрактора, может, вообще говоря, варьироваться в широком диапазоне и даже быть объектом оптимизации модели (1.9) вместе с другими структурными параметрами. Оптимизация временной задержки делалась, например, в работе [59] в приложении к оператору эволюции в более простой форме. В данном примере мы не будем варьировать τ .

Фрагменты временного ряда \mathbf{X} , соответствующего системе Лоренца-63 (1.31) без шума ($\sigma = 0$) и с шумом ($\sigma = 10$), представлены на рис. 1.1. Также на этом рисунке показаны проекции фазовой траектории системы на плоскость (x_n, x_{n+1}) для двух различных длин временного ряда, которые будут анализироваться в дальнейшем, и соответствующие инвариантные меры, посчитанные по достаточно длинным временным рядам, генерируемым системой при указанных значениях интенсивности шума. В отсутствие шума система аperiodически переключается между двумя регионами («крыльями») аттрактора Лоренца. Соответствующая структура аттрактора видна в

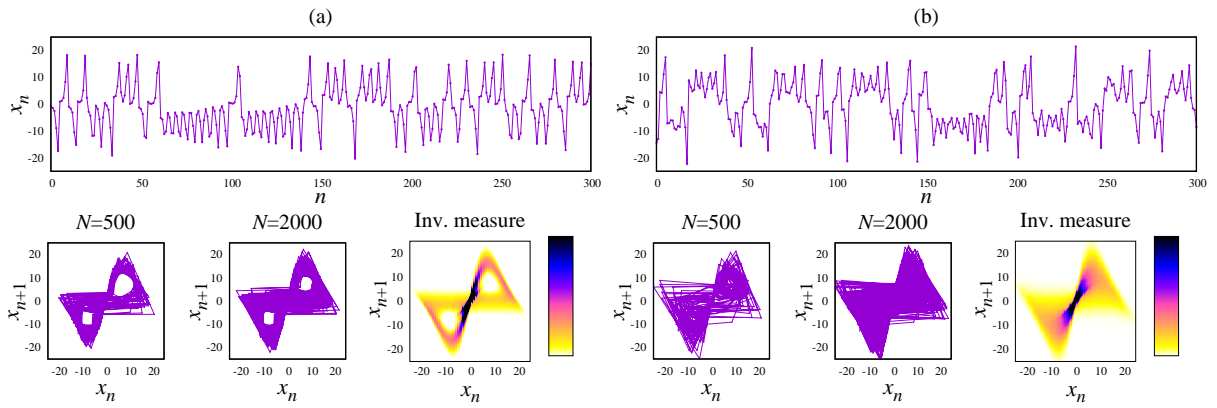


Рис. 1.1: Временные ряды, генерируемые системой (1.31) при $\sigma = 0$ (a) и $\sigma = 10$ (b). Вверху – фрагменты временных рядов x_n , внизу (слева направо) – проекции дискретной фазовой траектории на плоскость задержанных переменных (x_n, x_{n+1}) (для длин временного ряда $N = 500$ и $N = 2000$) и инвариантные меры системы в этих же переменных.

2-D проекциях и 2-D инвариантной мере на рис. 1.1a. Можно показать, например, с помощью широко известного метода «ложных ближайших соседей» [8, 61], что данный аттрактор может быть вложен в пространство размерности $D_u = l \cdot D = 4$, образуемое выбранными такенсовыми координатами с задержкой (см. рис. 1.2 и подпись к нему). Очевидно, эта низкоразмерная структура оператора эволюции разрушается с увеличением интенсивности шума σ (рис. 1.1b).

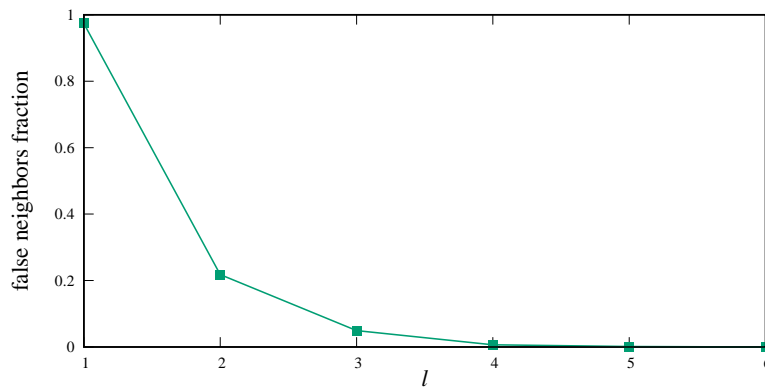


Рис. 1.2: Зависимость доли ложных соседей от количества задержек l , полученная по временному ряду (1.32) системы (1.31) при $\sigma = 0$ методом ложных ближайших соседей. Оценкой размерности вложения может быть размерность $D_u = l \cdot D = 4$, соответствующая точке $l = 4$, в которой доля ложных соседей перестает сильно убывать с ростом l .

При этом известно, что методы, аналогичные методу ложных ближайших соседей, т.е. основанные на детерминированном описании и анализе геометрии фазового пространства, плохо применимы для случайных или высокоразмерных динамических систем (для последних основной причиной является недостаточно большая длина доступных временных рядов): в этом случае такое описание становится менее адекватным. Поэтому поиск оптимальной размерности фазового пространства для стохастических моделей таких систем является актуальной проблемой. В ее решении может помочь, в частности, байесова обоснованность.

Реконструкция системы при различных уровнях шума и длинах временных рядов

Проанализируем результаты построения оптимальной модели по временным рядам системы (1.31) при трех значениях величины динамического шума: $\sigma = 5$, $\sigma = 10$ и $\sigma = 20$. Для всех значений σ рассмотрим случаи «короткого» ($N = 500$) и «длинного» ($N = 2000$) временного ряда. В каждом из этих случаев был применен алгоритм, описанный в разделе 1.3.1, и найдена зависимость ценовой функции для оптимальности (1.25) от гиперпараметров (m_f, m_g, l).

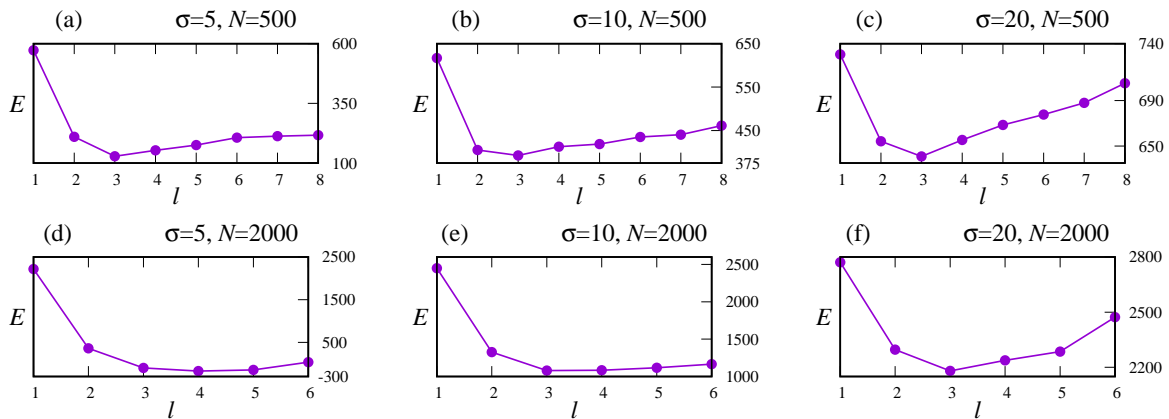


Рис. 1.3: Поведение ценовой функции для оптимальности (1.25) (в условных единицах) при уровнях шума в системе (1.31) $\sigma=5, 10$ и 20 и длинах ряда $N=500$ и 2000 . На каждом рисунке построена зависимость от количества задержек l (выражение (1.33)). Оптимальная размерность в каждом случае соответствует минимуму $E(l)$.

На рис. 1.3 показана зависимость этой ценовой функции от количества задержек l , которое в данном примере равно размерности фазового пространства $D_u = l \cdot D$,

при условии минимума по двум остальным параметрам m_f и m_g :

$$E(l) = \min_{m_f, m_g, \mathbf{q}} [-\ln P(\mathbf{X}|H_i, \mathbf{q})]. \quad (1.33)$$

Из рисунка можно видеть, что при увеличении σ и/или уменьшении длины ряда N оптимальное значение l уменьшается. Это объясняется ухудшением способности стохастической модели к фильтрации шума: при увеличении σ и уменьшении N детерминированная часть системы (1.31) становится все менее детально представлена детерминированной частью модели \mathbf{f} и все больше описывается за счет стохастической компоненты модели, что проиллюстрировано на рис. 1.4; в результате размерность фазового пространства, необходимая для параметризации \mathbf{f} , уменьшается. Так, при

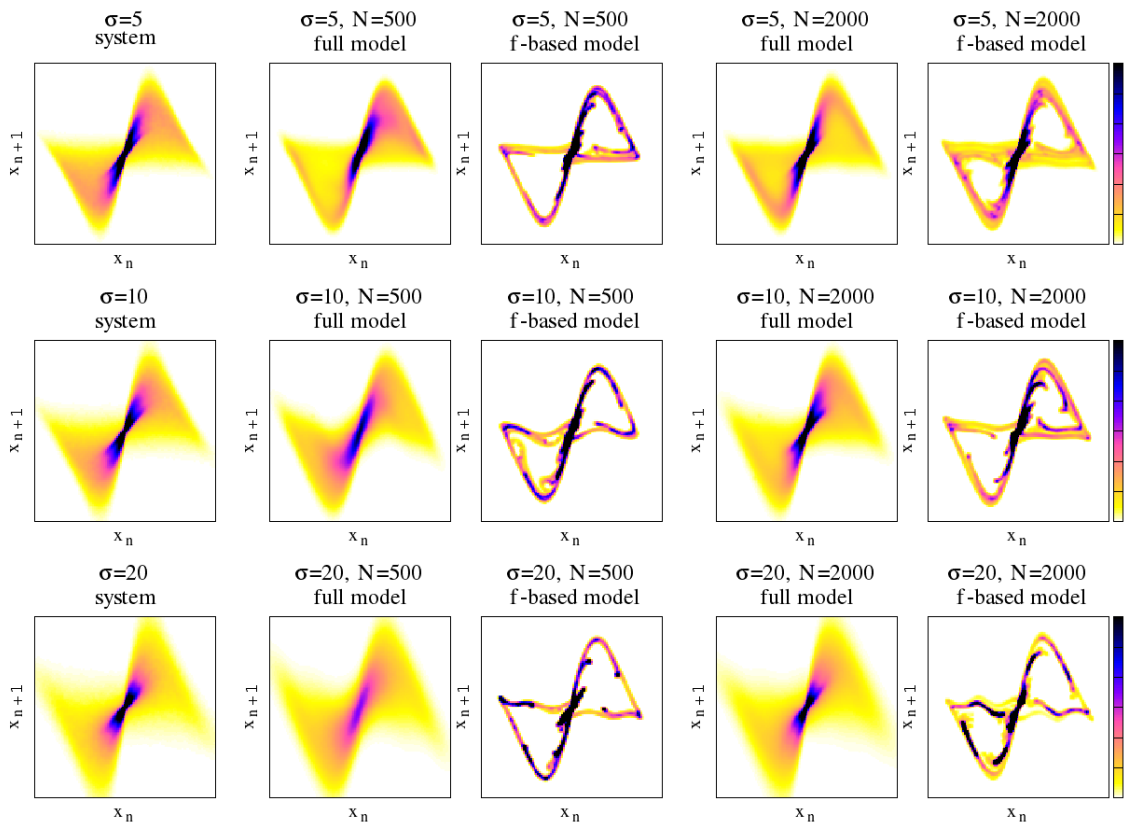


Рис. 1.4: Сравнение инвариантных мер системы, оптимальной стохастической модели (1.9) и детерминированной компоненты этой модели \mathbf{f} при различных уровнях шума в системе (по строкам). Модель была обучена по длинам временных рядов $N = 500$ и $N = 2000$ (по столбцам). По строкам: строка 1 – $\sigma = 5$ (a1-a5), строка 2 – $\sigma = 10$ (b1-b5), строка 3 – $\sigma = 20$ (c1-c5). По столбцам: столбец 1 – инвариантные меры системы (1.31), столбцы 2 и 4 – инвариантная мера полной модели (1.9) при $N = 500$ и $N = 2000$ соответственно, столбцы 3 и 5 – инвариантные меры детерминированной компоненты модели (1.9) при $N = 500$ и $N = 2000$ соответственно.

$\sigma = 5$, $N = 2000$ инвариантная мера, обусловленная детерминированной частью модели, наиболее близка к инвариантной мере невозмущенной системы, изображенной на рис. 1.1а, а оптимальная размерность фазового пространства равна 4, что соответствует размерности вложения невозмущенной системы. Напротив, при $\sigma = 20$ оптимальная размерность фазового пространства равна 3 и детерминированная часть модели устроена более просто (рис. 1.4).

На рис. 1.5 представлены зависимости ценовой функции для оптимальности (1.25) от количества нейронов в детерминированной части модели m_f при различных числах нейронов в стохастической части m_g для случая $\sigma = 10$ и соответствующего ему оптимального количества задержек $l = 3$. Во-первых, видно, что оптимум по m_f (а

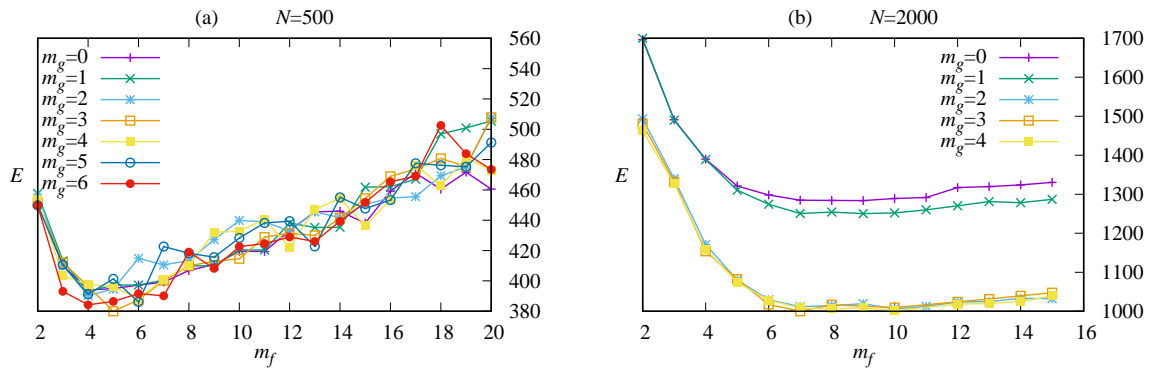


Рис. 1.5: Ценовая функция для оптимальности (1.25) (в условных единицах) модели с количеством задержек $l = 3$ как функция количества нейронов m_f в детерминированной части модели. Кривые разного цвета соответствуют разному количеству нейронов m_g в стохастической части модели. Модель была обучена при длинах ряда $N = 500$ (a) и $N = 2000$ (b) системы Лоренца-96 с уровнем шума $\sigma = 10$.

также по l , см. рис. 1.3) выражен более отчетливо в случае короткого временного ряда ($N = 500$), чем при $N = 2000$. Это связано с тем, что при меньшей длине выборки переобучение наступает, как правило, при меньшем количестве параметров модели. Во-вторых, при длине ряда $N = 2000$ модель с однородной стохастической частью ($m_g = 0$) или с одним нейроном в стохастической компоненте ($m_g = 1$) является намного менее обоснованной, чем модели с $m_g = 2, 3, 4$ (рис. 1.5b). В то же время, длина ряда $N = 500$ в данном случае является недостаточной для того, чтобы хорошо разрешить неоднородную структуру стохастической компоненты, и обоснованность модели с $m_g = 0$ слабо отличается от остальных случаев (рис. 1.5a). Заметим также, что для $\sigma = 5$ и $\sigma = 20$ оптимальные значения m_g полностью совпадают с рассмотренным случаем $\sigma = 10$. Таким образом, важность учета неоднородности стохастической

компоненты растут с ростом длины временного ряда.

В качестве дополнительной верификации критерия оптимальности на рис. 1.6 для случая $\sigma = 10$, $N = 2000$, когда модель с неоднородной стохастической частью является оптимальной, приведено сравнение этой модели с менее оптимальной моделью, имеющей однородную стохастическую часть, при тех же структурных параметрах m_f и l . Из инвариантных мер, показанных в верхней части рисунка, видно, что в

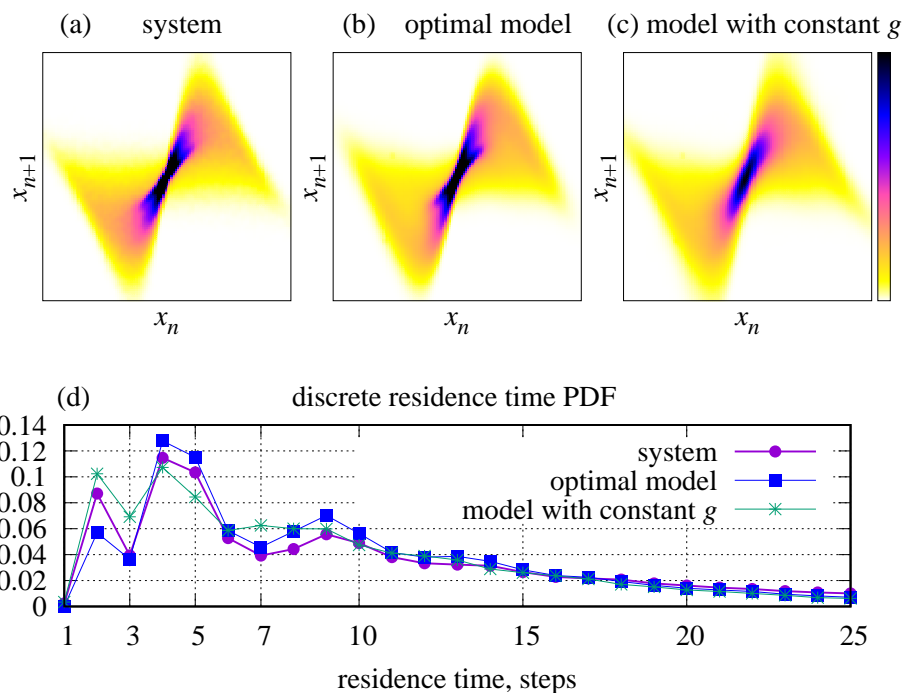


Рис. 1.6: Сравнение поведения моделей с однородной и неоднородной стохастической частью, обученных по временному ряду длиной $N = 2000$ системы (1.31) с уровнем шума $\sigma = 10$: (a) инвариантная мера системы ; (b) инвариантная мера оптимальной модели – $l = 3$, $m_f = 7$, $m_g = 2$; (c) инвариантная мера модели с однородной стохастической частью – $l = 3$, $m_f = 7$, $m_g = 0$; (d) – распределение времени жизни между переходами фазовой траектории с одного крыла системы Лоренца-96 на другое (для системы – фиолетовый цвет, для оптимальной модели – синий, для модели с однородной стохастической частью – зеленый).

целом обе модели воспроизводят очертания инвариантной меры системы, при этом на более детальном уровне модель с неоднородной стохастической частью лучше соответствует системе. В нижней части рис. 1.6 показана статистика времени жизни фазовой траектории на крыльях аттрактора Лоренца. Это время в данном случае измеряется в дискретном количестве отсчетов временного ряда. Видно, что обе модели воспроизводят два основных пика в дискретном распределении времени жизни, при этом данная бимодальность лучше выражена у модели с неоднородной стохастиче-

ской частью. Кроме того, эта оптимальная модель лучше выделяет вторичный пик, соответствующий времени жизни в 9 отсчетов.

В заключение отметим, что во всех рассмотренных случаях значения дисперсий априорных ПВ для коэффициентов ИНС (гиперпараметры \mathbf{q} , введенные в выражении (1.17)) в точке оптимума оказались достаточно хорошо удовлетворяющими следующим оценкам, сделанным из простых эвристических соображений в ряде предыдущих работ [85–87] (с учетом того, что ряд \mathbf{X} на время построения модели был приведен к единичной вариации):

$$\begin{aligned}\sigma_\alpha &= 1/\sqrt{m}, \\ \sigma_\omega &\sim 1, \\ \sigma_\gamma &= \sigma_\omega\sqrt{d_{in}},\end{aligned}\tag{1.34}$$

где m и d_{in} – число нейронов и размерность аргумента (входа) соответствующей ИНС. Данный результат, с одной стороны, подтверждает правильность сделанных оценок, и, с другой стороны, позволяет в реальных задачах использовать напрямую эти оценки, не проводя численную оптимизацию \mathbf{q} и уменьшая, тем самым, ресурсоемкость алгоритма обучения модели.

1.4 Реконструкция оператора эволюции по пространственно распределенным временным рядам

1.4.1 Описание и предварительная обработка используемых данных

Утверждения и выводы, сделанные в данном разделе, будут проиллюстрированы и проверены на пространственно распределенных временных рядах, сгенерированных глобальной климатической моделью института вычислительной математики РАН версии 4.0 (INMCM4.0) [125]. Данная модель содержит два основных блока: модель общей циркуляции атмосферы и модель общей циркуляции океана. В атмосферном блоке пространственное разрешение составляет 2×1.5 градусов по горизонтали и 21 уровень по вертикали (верхний уровень соответствует 10 гПа). В модели океана пространственное разрешение составляет 1×0.5 градусов по горизонтали и 40 уровней по вертикали. В качестве наблюдаемого временного ряда для анализа будет взят пространственно распределенный ряд ТПО на интервале с 1950 по 2000 год с шагом 1 месяц, сгенерированный данной моделью в ходе численного эксперимента по воспроизведению климата 1850-2005 годов (Historical run) в рамках международного проекта по сравнению климатических моделей «CMIP5» (Coupled Model Intercomparison Project,

phase 5) [126].

Выбранный временной ряд содержит ежемесячные значения ТПО одновременно в 85640 узлах эквидистантной по широте и долготе пространственной сетки. Ясно, что узлы такой сетки распределены по поверхности земного шара неравномерно: их плотность обратно пропорциональна косинусу широты, т.е. она выше у полюсов и ниже на экваторе. Учитывая этот факт, мы будем нормировать временные ряды так, чтобы среднее по времени значение для каждого ряда было равно нулю, а вариация ряда, деленная на площадь соответствующей элементарной ячейки сетки, равнялась единице:

$$x(t_n, \theta, \phi) = \frac{T(t_n, \theta, \phi) - \langle T(t_n, \theta, \phi) \rangle}{\sqrt{\langle (T(t_n, \theta, \phi) - \langle T(t_n, \theta, \phi) \rangle)^2 \rangle}} \cdot \sqrt{\cos \theta \Delta \theta \Delta \phi}. \quad (1.35)$$

В выражении (1.35) $T(t_n, \theta, \phi)$ – значение ТПО в момент времени t_n на широте θ и долготе ϕ , $\Delta \theta$ и $\Delta \phi$ – шаг сетки по широте и долготе соответственно, угловые скобки $\langle \dots \rangle$ обозначают операцию усреднения по времени, $x(t_n, \theta, \phi)$ – результирующее нормированное значение. Заметим, что данный способ нормировки несколько отличается от обычно используемого при анализе ТПО (когда знаменатель в выражении (1.35) берется равным единице) и, фактически, отражает предположение о том, что отношение сигнал/шум не зависит от координат. Вообще говоря, можно поставить задачу о выборе оптимальной нормировки, однако ее рассмотрение выходит за рамки диссертации. Всюду далее в диссертации при анализе ТПО в качестве нулевого шага будет предполагаться нормировка (1.35).

Кроме того, в данном разделе (1.4) перед дальнейшим анализом из временных рядов нормированной ТПО будет вычтен годовой ход, представляющий собой набор среднемесячных значений ТПО за наблюдаемый интервал времени. Таким образом, вместо значений ТПО мы будем иметь дело с отклонениями ТПО от среднегодового значения, которые называются также аномалиями ТПО (АТПО).

1.4.2 Эмпирические ортогональные функции

В отличие от примера, рассмотренного в конце предыдущего раздела, временной ряд пространственно распределенной системы, как правило, имеет огромную размерность D , пропорциональную количеству узлов в пространственной сетке (в рассматриваемом примере $D = 85640$). С одной стороны ясно, что такие данные содержат больше информации о системе, чем каждый из составляющих их скалярных временных рядов, с другой стороны, как было отмечено во Введении, конечная длина доступных в реальных приложениях временных рядов (в частности, рассматриваемых в данной работе) не позволяет строить статистически обоснованную модель в фазовом пространстве такой большой размерности. Ясно также, что для описания системы в

конкретном диапазоне пространственных и временных масштабов (например, крупномасштабных процессов, хорошо представленных в наблюдаемом ряде данных) эффект пространственной распределенности важен, но при этом количество независимых переменных, необходимых для описания их динамики, может быть существенно меньше D . Поэтому необходимым звеном в решении задачи реконструкции оператора эволюции пространственно распределенной системы является редукция размерности наблюдаемых данных с целью выявления наиболее адекватного набора переменных для описания динамики системы.

Одним из наиболее простых и часто применяемых способов редукции размерности является разложение данных по базису эмпирических ортогональных функций (ЭОФ), основанное на широко известном методе главных компонент [58, 100]. Суть его состоит в переходе к новым переменным \mathbf{y} (главным компонентам – ГК) с помощью проекции исходных данных $\mathbf{x}_1, \dots, \mathbf{x}_N$ на собственные вектора пространственной матрицы ковариаций $\mathbf{C} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}_n \mathbf{x}_n^T$ (где \mathbf{x}_n – вектора-столбцы, n – как и ранее, номер момента времени t_n):

$$\mathbf{y}_n = \mathbf{V}^T \cdot \mathbf{x}_n. \quad (1.36)$$

Здесь матрица \mathbf{V} имеет размер $D \times D$. Ее столбцы – собственные вектора $\mathbf{v}_1, \dots, \mathbf{v}_D$ матрицы \mathbf{C} , упорядоченные по убыванию собственных значений $\lambda_1, \dots, \lambda_D$, – представляют собой пространственные структуры, на которые проецируются исходные данные, и называются ЭОФ. Собственные значения при этом равны вариациям главных компонент \mathbf{y} .

Исходный временной ряд выражается через ЭОФ и временные ряды ГК следующим образом:

$$\mathbf{x}_n = \mathbf{V} \cdot \mathbf{y}_n = \sum_{i=1}^D y_{in} \cdot \mathbf{v}_i. \quad (1.37)$$

Таким образом, полное ЭОФ-разложение соответствует полному ортогональному повороту исходных переменных к собственному базису матрицы \mathbf{C} . При этом временные ряды ГК \mathbf{y} не коррелируют друг с другом по времени, а структуры ЭОФ ортогональны в пространстве.

Редукция размерности данных состоит в том, что в качестве переменных для дальнейшего моделирования берутся лишь первые d ГК ($d \ll D$):

$$\mathbf{p}_n = \begin{pmatrix} y_{1n} \\ \vdots \\ y_{dn} \end{pmatrix} = \mathbf{V}_d^T \cdot \mathbf{x}_n, \quad (1.38)$$

где матрица $\mathbf{V}_d := (\mathbf{v}_1, \dots, \mathbf{v}_d)$ имеет размер $D \times d$ и содержит d главных ЭОФ.

В итоге исходные данные с учетом редукции представляются следующим образом:

$$\mathbf{x}_n = \mathbf{V}_d \cdot \mathbf{p}_n + \mathbf{r}_n = \sum_{i=1}^d p_{in} \cdot \mathbf{v}_i + \mathbf{r}_n, \quad (1.39)$$

где первое слагаемое – главная часть разложения, описываемая d -мерной переменной \mathbf{p}_n , а \mathbf{r}_n – остаток разложения. Можно показать [58], что для любого d ЭОФ-разложение соответствует максимально возможной вариации переменных \mathbf{p} и минимально возможной вариации остатка \mathbf{r}_n . Именно это свойство определяет смысл редукции размерности с помощью ЭОФ-разложения: выбирается линейный поворот к таким d переменным, которые описывают наибольшую часть суммарной вариации наблюдаемого временного ряда.

На рис. 1.7 и 1.8 показаны пространственные структуры первых четырех ЭОФ $\mathbf{v}_1, \dots, \mathbf{v}_4$, отображенных на географическую сетку, и соответствующие им временные ряды ГК p_{1n}, \dots, p_{4n} , рассчитанные по данным АТПО модели INMCM4.0, описанным в разделе 1.4.1. В структуре первой ЭОФ (рис. 1.7) наиболее отчетливо представлен

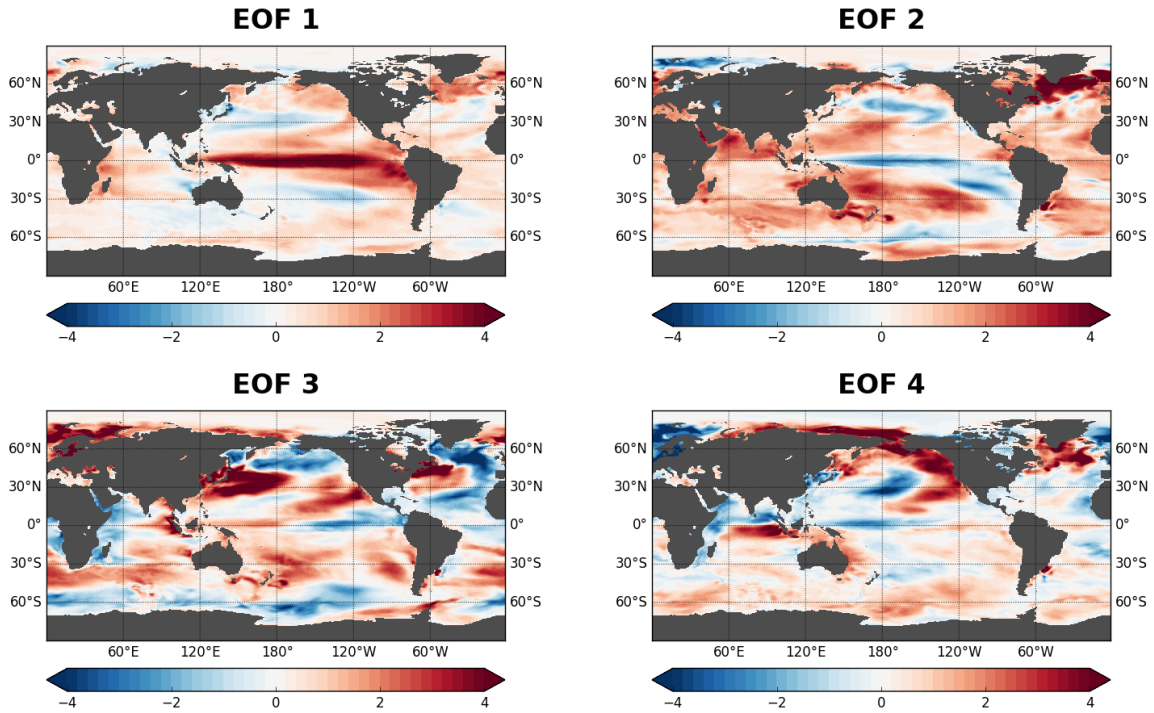


Рис. 1.7: Пространственные структуры первых четырех ЭОФ, рассчитанных в результате ЭОФ-разложения данных АТПО на интервале с 1950 по 2000 год, сгенерированных с помощью модели INMCM4.0 (раздел 1.4.1)

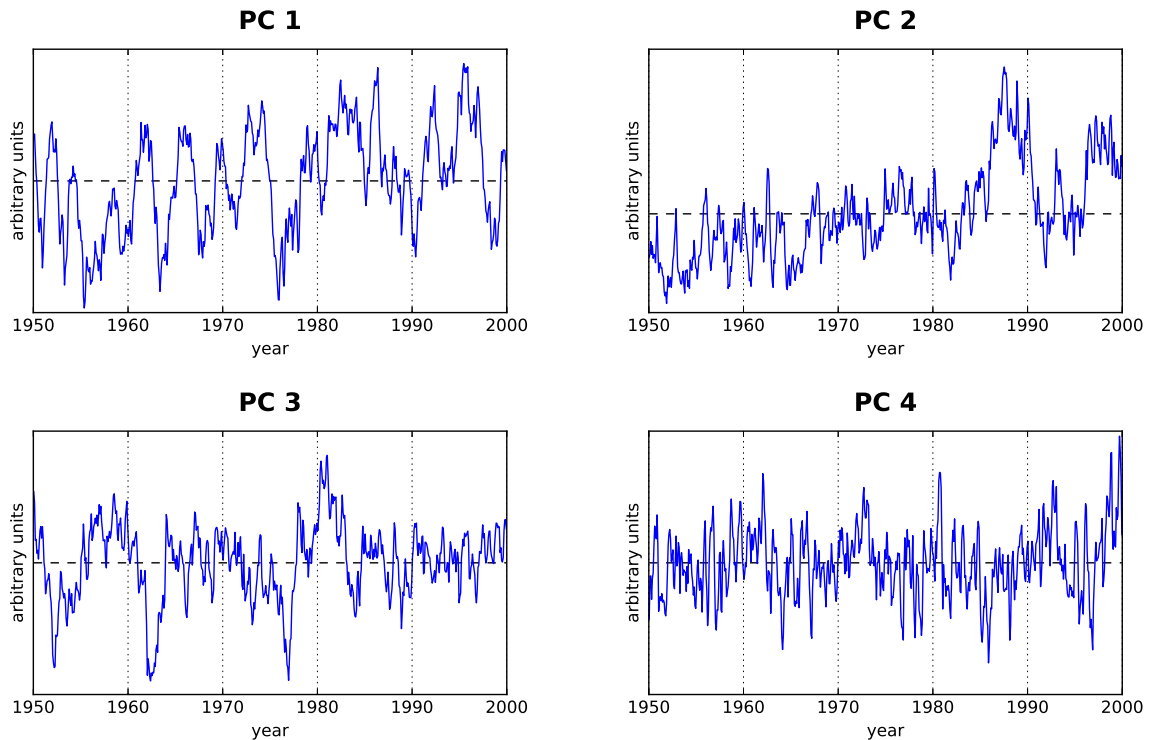


Рис. 1.8: Временные ряды первых четырех ГК, рассчитанных в результате ЭОФ-разложения данных АТПО на интервале с 1950 по 2000 год, сгенерированных с помощью модели INMCM4.0 (раздел 1.4.1)

характерный паттерн в тропической зоне Тихого океана, соответствующий ЭНЮК; при этом во временном ряде первой ГК (рис. 1.8) доминирует межгодовая изменчивость ЭНЮК: характерный период 2-7 лет, с которым осциллирует паттерн ЭНЮК в модели INMCM4.0. Во временных рядах ГК 1-3 (наиболее отчетливо – в ГК 2, рис. 1.8) виден самый медленный масштаб, заключающийся в монотонном изменении характерного среднего значения ГК, который можно ассоциировать с трендом глобального потепления в модели INMCM4.0, особенно заметным в районах Индийского океана, Индонезии и северной части Атлантического океана (рис. 1.7). Кроме того, во всех четырех ГК присутствует в той или иной степени декадная изменчивость (рис. 1.8). В частности, во всех пространственных структурах ЭОФ прослеживаются элементы паттерна ТДК (рис. 1.7) – характерной дипольной структуры в Тихом океане в северном полушарии [82]. Наконец, обратим внимание, что вклад в динамику ГК изменчивости с масштабами, не превышающими нескольких месяцев, увеличивается с ростом номера ГК.

1.4.3 Модифицированная стохастическая модель оператора эволюции

Как уже было сказано, основная идея реконструкции оператора эволюции по пространственно распределенным данным сводится к построению его модели в редуцированном пространстве ГК, имеющем сравнительно невысокую размерность. Таким образом, роль исходных данных для построения модели будут играть временные ряды ГК $\mathbf{p}_1, \dots, \mathbf{p}_N$.

Пример ГК по данным АТПО модели INMCM4.0, показанных в конце предыдущего раздела, демонстрирует важные свойства, присущие практически всем наблюдаемым пространственно распределенным временным рядам сложных систем. Во-первых, это наличие нескольких групп масштабов в наблюдаемой пространственно-временной динамике системы: (i) масштабы, хорошо разрешенные в наблюдаемом временном ряде и укладывающиеся на длине ряда несколько раз, которые, как правило, представлены в виде повторяющегося сигнала с ненулевым масштабом автокорреляции (т.е. «гладкого» сигнала) и наибольшей амплитудой изменчивости, такие как межгодовая изменчивость ЭНЮК (в большей степени) или декадная изменчивость (в меньшей степени) в рассматриваемом примере; (ii) медленные масштабы, сравнимые с длиной наблюдаемого временного ряда или превышающие ее и поэтому видимые как медленные тренды (например, тренд глобального потепления), детальной информации о законе эволюции которых в данных заведомо недостаточно из-за конечной длительности наблюдения; (iii) быстрые масштабы, плохо разрешаемые или не разрешаемые в наблюдаемом ряде из-за недостаточно малого шага по времени, которые видны как быстропеременный нерегулярный («случайный») сигнал. Видно, что в первые ЭОФ в рассматриваемом примере попали, главным образом, процессы с масштабами из группы (i), имеющие наибольшую по амплитуде изменчивость.

Во-вторых, являясь различными, вышеперечисленные временные масштабы, тем не менее, относительно близки друг к другу (отличаются лишь в несколько раз). Это означает, что части системы, соответствующие этим масштабам, вообще говоря, нельзя считать полностью независимыми ни в пространстве, ни во времени, поэтому при моделировании необходимо учитывать их совместно.

В-третьих, даже для масштабов из группы (i), которые могли бы достаточно хорошо моделироваться с помощью детерминированного оператора эволюции, из-за существенно ограниченной длины доступных наблюдаемых временных рядов зачастую не выполняется условие достаточно низкой размерности соответствующего им многообразия в фазовом пространстве системы (см. Введение).

Модель, сформулированная в разделе 1.2, решает перечисленные проблемы лишь

отчасти. В частности, она позволяет учесть влияние быстрых процессов с масштабами из группы (iii) в форме стохастического воздействия. Стохастическая компонента модели может также существенно смягчить условие на максимально допустимую размерность фазового пространства за счет возможности адекватного описания части детерминированной динамики системы, как показано в разделе 1.3.2.

Идея модификаций модели, предлагаемых в данном разделе, состоит в том, что в условиях ограниченной длины временного ряда улучшить качество реконструкции системы можно за счет явного введения в модель априорной информации о неавтономности оператора эволюции, связанной с наличием медленных трендов (ii) и известных внешних воздействий на систему, а также информации о ненулевом временном масштабе его детерминированной компоненты, отражающем присущие системе детерминированные причинно-следственные связи.

По аналогии с выведенным из более общей формы выражением (1.9), запишем математическую форму предлагаемой модифицированной стохастической модели оператора эволюции для ГК:

$$\begin{aligned} \mathbf{p}_{n+1} &= \mathbf{p}_n + \mathbf{f}(\mathbf{p}_n, \dots, \mathbf{p}_{n-l+1}; \boldsymbol{\kappa}_n; t_n; \boldsymbol{\mu}_f) + \hat{\mathbf{g}}(\mathbf{p}_n, \dots, \mathbf{p}_{n-l+1}; \boldsymbol{\mu}_g) \cdot \boldsymbol{\xi}_n, \\ n &= l, \dots, N - 1. \end{aligned} \quad (1.40)$$

Как и модель (1.9), она основана на использовании переменных с задержками (l – количество задержанных переменных, или «глубина памяти» модели) и состоит из детерминированной и случайной компонент, параметризованных с помощью параметров $\boldsymbol{\mu}_f$ и $\boldsymbol{\mu}_g$. Случайная компонента, описываемая последним слагаемым, полностью эквивалентна случайной компоненте модели (1.9): ее роль заключается в моделировании неоднозначности, возникающей при реконструкции детерминированной части в низкоразмерном фазовом пространстве ГК и включающей, главным образом, быстрые процессы из группы (iii). Детерминированная компонента модели представлена первыми двумя слагаемыми в выражении (1.40). Первое слагаемое отвечает за явный учет ненулевого масштаба автокорреляции, т.е. гладкости, временного ряда ГК $\mathbf{p}_1, \dots, \mathbf{p}_N$, подобно потоковым моделям (см. Введение). Второе слагаемое представляет собой детерминированный нелинейный оператор, применяемый к задержанным переменным, который, в отличие от модели (1.9), дополнительно включает явную медленную зависимость от времени t_n (нестационарность) и может зависеть от внешнего воздействия $\boldsymbol{\kappa}_n$ (если временной ряд $\boldsymbol{\kappa}_1, \dots, \boldsymbol{\kappa}_N$ этого воздействия известен, см. пример ниже), что соответствует явному учету вышеупомянутой неавтономности моделируемой системы.

Как и ранее, функции \mathbf{f} и $\hat{\mathbf{g}}$ параметризуются с помощью ИНС. Если функциональная форма $\hat{\mathbf{g}}$ полностью определяется выражением (1.11), то для детерминированной

компоненты необходимо конкретизировать зависимость \mathbf{f} от t_n и $\boldsymbol{\kappa}_n$:

$$\mathbf{f}(\mathbf{z}_n, \boldsymbol{\kappa}_n, t_n; \boldsymbol{\mu}_f) = \sum_{i=1}^{m_f} (\boldsymbol{\alpha}_i + t_n \boldsymbol{\beta}_i) \cdot \tanh(\boldsymbol{\omega}_i^T \mathbf{z}_n + \boldsymbol{\nu}_i^T \boldsymbol{\kappa}_n + \gamma_i). \quad (1.41)$$

Здесь вектор-столбец \mathbf{z}_n обозначает набор переменных с задержками $(\mathbf{p}_n, \dots, \mathbf{p}_{n-l+1})$; $\boldsymbol{\alpha}_i \in \mathbb{R}^d$, $\boldsymbol{\beta}_i \in \mathbb{R}^d$, $\boldsymbol{\omega}_i \in \mathbb{R}^{ld}$, $\boldsymbol{\nu}_i \in \mathbb{R}^{\dim \boldsymbol{\kappa}_n}$, $\gamma_i \in \mathbb{R}$ – коэффициенты ИНС, собранные в вектор параметров $\boldsymbol{\mu}_f$; время t_n полагается меняющимся от -0.5 до 0.5 на интервале наблюдения. Вообще говоря, для моделирования медленной нестационарности параметров оператора эволюции реальной системы, связанной с наличием трендов, следовало бы ввести слабую зависимость от времени во все коэффициенты ИНС (1.10). Разложение этой слабой зависимости в ряд Тейлора по времени до первого порядка эквивалентно линейной зависимости оператора \mathbf{f} от явного времени t_n . Для данного случая ИНС можно показать [87], что форма (1.41) представляет собой универсальный аппроксиматор такого разложения для любой нелинейной функции от $(\mathbf{z}_n, \boldsymbol{\kappa}_n, t_n)$.

Аналогично выводу, приведенному в разделе 1.3.1, можно получить ценовую функцию для обучения и оптимизации модели (1.40) в рамках байесова подхода, которая будет задаваться выражением (1.26), с той лишь разницей, что ряд $\mathbf{x}_1, \dots, \mathbf{x}_N$ заменится на ряд ГК $\mathbf{p}_1, \dots, \mathbf{p}_N$, а обозначения (1.16) примут вид:

$$\begin{aligned} \mathbf{f}_n(l, \boldsymbol{\mu}_f) &:= \tilde{\mathbf{u}}_n + \mathbf{f}(\tilde{\mathbf{u}}_{n-1}, \dots, \tilde{\mathbf{u}}_{n-l}, \boldsymbol{\kappa}_n, t_n; \boldsymbol{\mu}_f) \\ \mathbf{G}_n^{-1}(l, \boldsymbol{\mu}_g) &:= \hat{\mathbf{g}}(\tilde{\mathbf{u}}_{n-1}, \dots, \tilde{\mathbf{u}}_{n-l}; \boldsymbol{\mu}_g) \cdot \hat{\mathbf{g}}^T(\tilde{\mathbf{u}}_{n-1}, \dots, \tilde{\mathbf{u}}_{n-l}; \boldsymbol{\mu}_g). \end{aligned} \quad (1.42)$$

Кроме того, в целях уменьшения вычислительных затрат при построении данной модели мы воспользуемся результатом, приведенным в конце раздела 1.3.2, т.е. не будем проводить численную оптимизацию априорных значений дисперсий коэффициентов ИНС, а выберем их на основе оценок (1.34). При этом важно иметь в виду, что эти оценки делались для рассматриваемого в разделе 1.3 случая единичной нормировки временного ряда \mathbf{X} . В модифицированной модели для ГК мы не будем считать ГК приведенными к единичной вариации, поскольку соотношение их вариаций несет дополнительную информацию о динамике исходной (нередуцированной) системы. Учет этого фактора, а также наличия нестационарности и внешнего воздействия, приводит к следующей априорной ПВ для детерминированной компоненты, отличающейся от (1.21) добавлением соответствующих априорных множителей для коэффициентов на выходе ИНС и перемасштабированием коэффициентов при входах ИНС в соответствии с суммарными вариациями \mathbf{z}_n и $\boldsymbol{\kappa}_n$:

$$\begin{aligned}
\mathcal{P}_{pr}(\boldsymbol{\mu}_f) &= \prod_{i=1}^m \prod_{k=1}^d (2\pi\sigma_\alpha^2 s_k)^{-1} \exp\left(-\frac{|\boldsymbol{\alpha}_i|^2 + |\boldsymbol{\beta}_i|^2}{2\sigma_\alpha^2 s_k}\right) \times \\
&\times \prod_{i=1}^m \left(\frac{2\pi\sigma_\omega^2}{s_z}\right)^{-\frac{ld}{2}} \left(\frac{2\pi\sigma_\omega^2}{s_\kappa}\right)^{-\frac{\dim \kappa}{2}} \exp\left(-\frac{|\boldsymbol{\omega}_i|^2 s_z + |\boldsymbol{\nu}_i|^2 s_\kappa}{2\sigma_\omega^2}\right) \times \\
&\times \prod_{i=1}^m (2\pi\sigma_\gamma^2)^{-\frac{1}{2}} \exp\left(-\frac{\gamma_i^2}{2\sigma_\gamma^2}\right),
\end{aligned} \tag{1.43}$$

где $s_k = \frac{1}{N-1} \sum_{n=2}^N (p_{k,n} - p_{k,n-1})^2$, $s_z = \frac{1}{Nd} \sum_{n=1}^N |\mathbf{p}_n|^2$, $s_\kappa = \frac{1}{N \dim \kappa} \sum_{n=1}^N |\boldsymbol{\kappa}_n|^2$. Аналогичным образом, для коэффициентов стохастической компоненты при $m_g \geq 1$ имеем:

$$\begin{aligned}
\mathcal{P}_{pr}(\boldsymbol{\mu}_g) &= \prod_{i=1}^m \prod_{k=1}^d \prod_{j=1}^k (2\pi\sigma_\alpha^2 s_k)^{-\frac{1}{2}} \exp\left(-\frac{|\boldsymbol{\alpha}_{j+k(k-1)/2}|^2}{2\sigma_\alpha^2 s_k}\right) \times \\
&\times \prod_{i=1}^m \left(\frac{2\pi\sigma_\omega^2}{s_z}\right)^{-\frac{ld}{2}} \exp\left(-\frac{|\boldsymbol{\omega}_i|^2 s_z + |\boldsymbol{\nu}_i|^2 s_\kappa}{2\sigma_\omega^2}\right) \times \\
&\times \prod_{i=1}^m (2\pi\sigma_\gamma^2)^{-\frac{1}{2}} \exp\left(-\frac{\gamma_i^2}{2\sigma_\gamma^2}\right).
\end{aligned} \tag{1.44}$$

Конкретные значения $\sigma_\alpha, \sigma_\omega, \sigma_\gamma$, исходя из результата (1.34), для модифицированной стохастической модели в настоящей работе взяты равными $\sigma_\alpha = 1/\sqrt{m}$, $\sigma_\omega = 2$, $\sigma_\gamma = \sigma_\omega \sqrt{ld + \dim \kappa}$.

Таким образом, с учетом описанных модификаций, обучение и оптимизация модели (1.40) с помощью ценовой функции (1.26) проходят в соответствии с алгоритмом, изложенным в разделе 1.3.1, из которого исключен шаг оптимизации априорных дисперсий коэффициентов ИНС (шаг 5 алгоритма).

1.4.4 Реконструкция оператора эволюции по данным климатической модели INMCM4.0

Продemonстрируем работу метода построения модифицированной стохастической модели на примере реконструкции оператора эволюции в пространстве четырех ГК по данным АТПО модели INMCM4.0, найденных в разделе 1.4.2. С одной стороны, при формировании временных рядов АТПО из данных был вычтен годовой ход, представляющий собой сильно нелинейный режим, который при этом легко предсказывается без привлечения вышеописанных методов, что автоматически упрощает строящуюся модель. С другой стороны, такое линейное вычитание, вообще говоря, не исключает нелинейного влияния годового сигнала на другие процессы, представленные в получившихся ГК. Поэтому в модель (1.40) для ГК был введен в качестве заданного внешнего воздействия периодический сигнал с периодом 1 год (12 месяцев):

$$\boldsymbol{\kappa}_n = \left(\sin \frac{2\pi n}{12}, \cos \frac{2\pi n}{12} \right). \tag{1.45}$$

Очевидно, нелинейная функция от определенного таким образом внешнего воздействия способна моделировать нелинейный отклик на сигнал с периодом 1 год в любой фазе.

Как было отмечено, первое слагаемое в модели (1.40) отвечает за явный учет ненулевого временного масштаба автокорреляции (гладкости) временных рядов ГК. Дело в том, что в случае гладкого сигнала оператор эволюции близок к тождественному (\mathbf{p}_{n+1} близко к \mathbf{p}_n), и явный учет слагаемого \mathbf{p}_n позволяет сэкономить нейроны ИНС, нужные для аппроксимации данной тождественности. По сути, форма модели в этом случае представляет собой дискретизацию нелинейного дифференциального уравнения, когда оператор \mathbf{f} относительно мал, но при этом определяет динамику системы. На рис. 1.9 (слева) приведена зависимость ценовой функции для оптимальности (1.25) от количества нейронов m_f для модели (1.40), построенной по найденным четырем ГК, и для такой же модели, в которой отсутствовало слагаемое \mathbf{p}_n . Как видно из

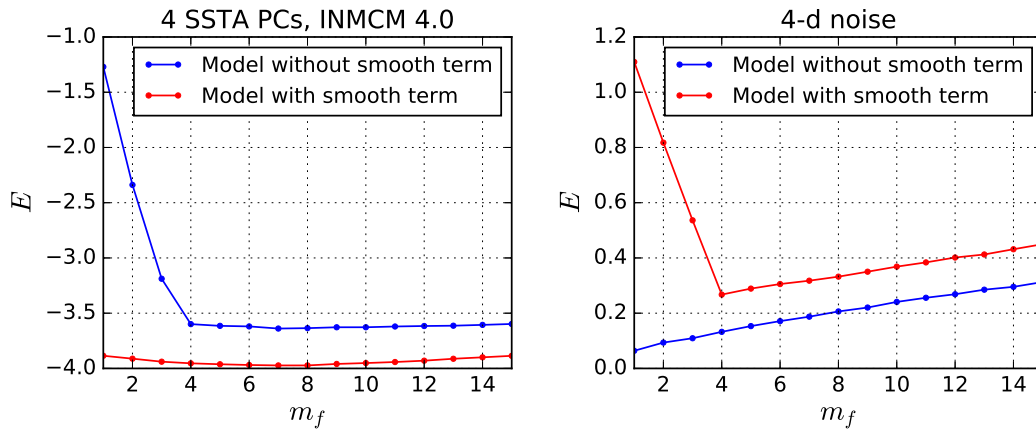


Рис. 1.9: Ценовая функция для оптимальности (1.25) в зависимости от количества нейронов m_f (при оптимальных прочих структурных параметрах) для модели (1.40) (красный цвет) и для такой же модели, в которой отсутствует слагаемое \mathbf{p}_n (синий цвет). Слева показан результат для моделей, построенных по четырем ГК АТПО климатической модели INMCM4.0, справа – для моделей, построенных по четырем временным реализациям белого шума, имеющих те же длительности и вариации, что и четыре ГК АТПО.

рисунка, модель с явным учетом гладкости ГК при любом количестве нейронов m_f соответствует меньшим значениям ценовой функции, т.е. является более оптимальной. Кроме того, модель, не учитывающая ответственную за гладкость компоненту \mathbf{p}_n , при количестве нейронов в детерминированной части, меньшем 4, имеет значительно худшую обоснованность, поскольку такого количества нейронов не хватает даже для

аппроксимации зависимости $\mathbf{p}_{n+1} \approx \mathbf{p}_n$, где размерность \mathbf{p}_n равна 4. Для сравнения, на рис. 1.9 справа показаны аналогичные графики для таких же моделей, построенных по четырем временным рядам, являющимся реализациями белого шума и имеющим ту же длину и те же значения вариаций, что и четыре ГК модели INMCM4.0. Поскольку временные ряды белого шума не являются гладкими, их модель с явным учетом слагаемого, ответственного за ненулевой временной масштаб автокорреляции ГК (1.40), является менее оптимальной, чем модель без учета данного слагаемого; при этом модель (1.40) в этом случае вынуждена затрачивать дополнительные 4 нейрона на компенсацию слагаемого \mathbf{p}_n , которое оказалось излишним, и ее оптимальность особенно плоха при $m_f < 4$.

Таким образом, явный учет временного масштаба автокорреляции временных рядов ГК АТПО в модели (1.40), наряду с неавтономностью, позволяет построить более оптимальную модель. На рис. 1.10 представлена зависимость ценовой функции для оптимальности этой модели от числа нейронов в детерминированной части m_f при различных значениях числа нейронов в стохастической части m_g . Видно, что

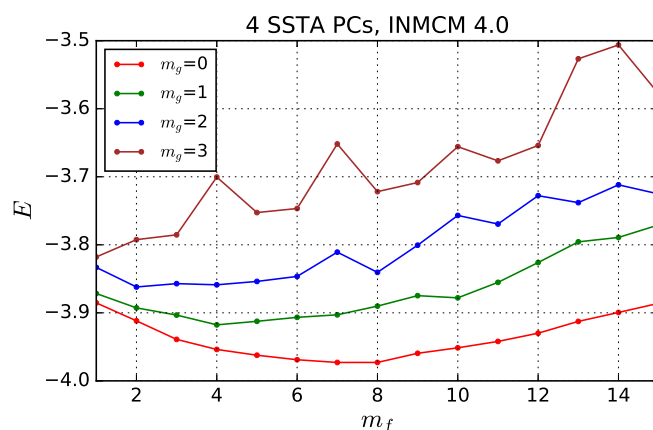


Рис. 1.10: Ценовая функция для оптимальности (1.25) в зависимости от количества нейронов m_f для модели (1.40), построенной по четырем ГК АТПО климатической модели INMCM4.0, при различных значениях количества нейронов в стохастической части m_g (при оптимальном значении задержки l).

оптимальной оказалась однородная стохастическая часть ($m_g = 0$). Аналогично случаю с моделью по временному ряду системы Лоренца длиной $N = 500$ (рис. 1.5), такой результат можно объяснить сравнительно короткой длиной рассмотренного временного ряда по отношению к моделируемым временным масштабам. Отметим, что, в силу большого количества временных масштабов в климатической системе и относительной близости этих масштабов друг другу, недостаточность длины практически любого временного ряда реальной климатической системы представляется

неизбежной: попытка рассмотреть более длинный интервал времени приводит к включению в рассмотрение новых временных масштабов, перестающих играть роль медленных трендов, для которых новый интервал снова оказывается коротким.

В итоге мы имеем стохастическую модель с оптимальными значениями структурных параметров $m_f = 7$, $l = 2$, $m_g = 0$. Рис. 1.11 демонстрирует способность построенной модели предсказывать траекторию системы, спроецированную в пространство четырех главных ГК, на некоторое время в будущее (количественный прогноз). На этом рисунке приведена среднеквадратичная ошибка (СКО) прогноза

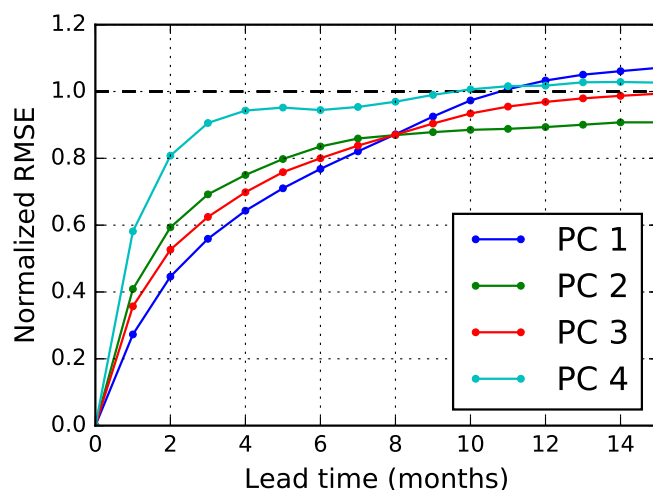


Рис. 1.11: Зависимость нормированной СКО прогноза (по оси ординат) каждой из четырех ГК от дальности прогноза, отложенной по оси абсцисс в месяцах. Нормированная СКО, равная единице, соответствует СКО, равной среднеквадратичному отклонению значений ГК от линейного тренда, содержащегося в этой ГК (т.е., фактически, амплитуде изменчивости ГК на наиболее разрешенных наблюдаемой выборкой временных масштабах).

каждой из четырех ГК как функция дальности прогноза, определяемой в данном случае количеством последовательных применений стохастического оператора эволюции системы; усреднение при этом проводилось по ансамблю начальных условий, взятому из наблюдаемого интервала времени. Фактически, данная СКО определяет реально даваемую моделью точность прогноза траектории системы. Процедура определения СКО в случае ансамбля реализаций прогноза стохастической моделью будет более подробно пояснена в главе 4, посвященной количественному прогнозу индексов ЭНЮК. Как и следовало ожидать, СКО увеличивается с ростом дальности прогноза, и прогноз траектории системы с помощью построенной модели в данном

случае становится практически полностью неинформативным при дальности более 8 месяцев. При этом видно, что при дальности прогноза, все еще имеющей смысл (менее 8 месяцев), точность прогноза тем выше, чем меньше номер ГК. Эту закономерность можно объяснить тем, что при ЭОФ-разложении степень «зашумленности» ГК, как правило, растет с номером ГК (см. рис. 1.8), увеличивая, тем самым, неопределенность прогноза (величину стохастической части модели) и приближая горизонт прогноза.

Кроме того, как показано в разделе 1.3.2, стохастическая модель оператора эволюции способна воспроизводить качественные характеристики наблюдаемого низкоразмерного режима поведения системы. В качестве иллюстрации данного утверждения на примере пространственно распределенных данных климатической модели INMCM4.0 на рис. 1.12 приведена одна случайная реализация стохастической модели, запущенной с начальных условий, соответствующих 1950 году. Заметим, что, в отличие от

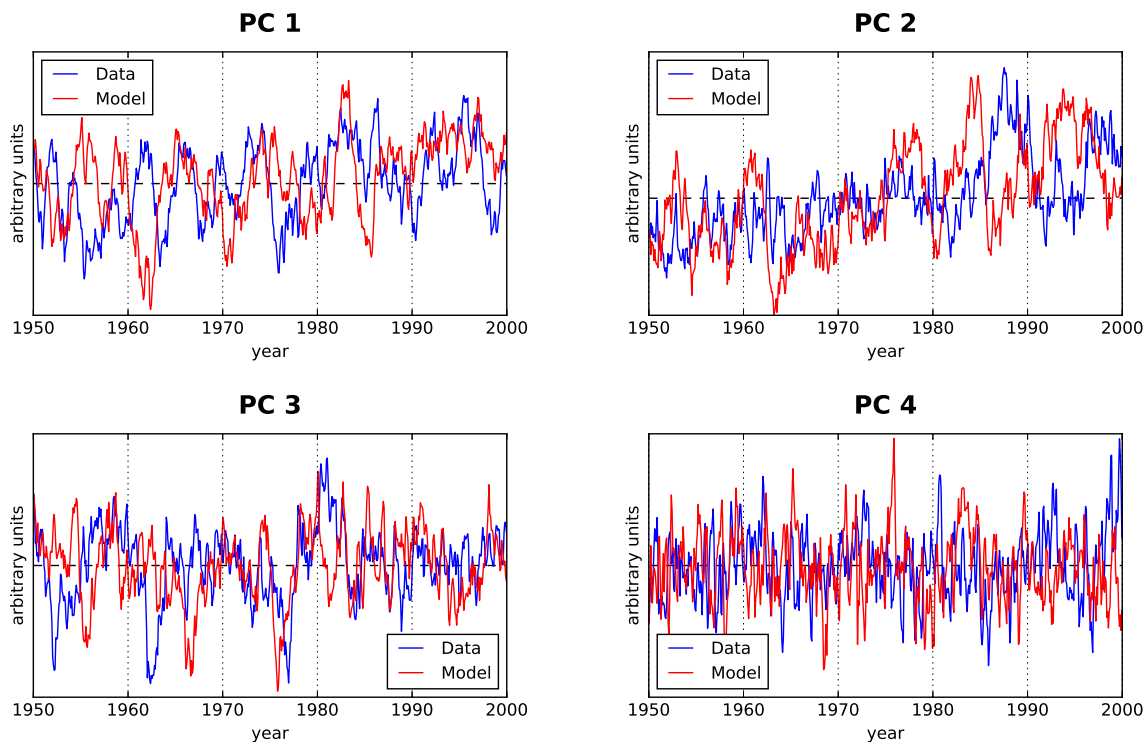


Рис. 1.12: Временные ряды главных четырех ГК АТПО климатической модели INMCM4.0 (синий цвет) и временные ряды одной случайной реализации построенной по данным ГК модифицированной стохастической модели (1.40), запущенной с начальных условий, соответствующих 1950 году (красный цвет).

примера, рассмотренного в разделе 1.3.2, в данном случае в системе присутствовал медленный тренд, который в модели (1.40) был учтен в виде слабой неавтономности

оператора эволюции. Как видно из рисунка, динамика на наблюдаемых временных масштабах (включая медленный тренд) для каждой ГК качественно воспроизводится построенной стохастической моделью.

Таким образом, предложенная модифицированная стохастическая модель оператора эволюции, действующая в редуцированном пространстве главных ГК, способна как к количественному прогнозу, так и к реконструкции качественного поведения этих ГК, а, значит, и описываемой ими главной части исходной пространственно распределенной системы.

1.5 Заключение

В данной главе был рассмотрен метод построения оптимальной стохастической модели оператора эволюции по наблюдаемому многомерному временному ряду неизвестной динамической системы. Практическая значимость такой модели объясняется тем, что наблюдаемые в природе временные ряды с эмпирической точки зрения зачастую являются плохо отличимыми от стохастических процессов, поэтому требуются статистически корректные алгоритмы извлечения информации об их законах эволюции. В частности, рассмотрена модель в форме дискретизованного по времени оператора эволюции (точечного отображения), который является суперпозицией детерминированной и случайной компонент, представляемых с помощью трехслойной ИНС. Важной частью такой модели является зависимость параметров случайной компоненты оператора эволюции от текущего состояния системы в фазовом пространстве; при этом сложность (нелинейность) как детерминированной, так и стохастической компонент оптимизируется в рамках байесова подхода. Алгоритм обучения эмпирической модели на основе общего байесова критерия оптимальности подробно изложен и будет использоваться также в следующих главах.

Эффективность стохастической модели проиллюстрирована на примере системы Лоренца-63. Показано, что оптимальная, с точки зрения байесова подхода, структура стохастической модели оператора эволюции зависит от имеющегося в распоряжении исследователя временного ряда данных. В частности, для достаточно длинного временного ряда и при наличии нелинейной детерминированной динамики (как в системе Лоренца-63) модель, учитывающая неоднородность стохастической компоненты в фазовом пространстве, становится более предпочтительной и точнее описывает поведение системы.

Рассмотрен вопрос эмпирической реконструкции оператора эволюции с помощью стохастической модели по пространственно распределенным временным рядам сложных систем, таких как климатическая система: в этом случае вместо того, чтобы

проводить реконструкцию напрямую в наблюдаемых переменных, имеющих высокую размерность, предложено строить стохастическую модель в низкоразмерном пространстве, полученном с помощью эмпирической редукции размерности исходных данных. В качестве наиболее простого примера редукции размерности, демонстрирующего работоспособность модели, в данной главе взято ЭОФ-разложение. Кроме того, исходя из общих свойств природных пространственно распределенных систем, предложена модификация формы стохастической модели, учитывающая возможные присутствие в системе медленной нестационарности и ее зависимость от внешних воздействий, а также ненулевой временной масштаб автокорреляции (гладкость) ГК, получаемых после редукции размерности. Сформулирован байесов алгоритм обучения модифицированной стохастической модели.

На примере временного ряда, сгенерированного глобальной климатической моделью INMCM4.0 в ходе численного эксперимента по воспроизведению климата XX века в рамках международного проекта по сравнению климатических моделей «CMIP 5», продемонстрирована эффективность и работоспособность модифицированной эмпирической стохастической модели, а именно ее способность как давать краткосрочный прогноз фазовой траектории, так и качественно воспроизводить режим поведения системы.

Глава 2

Метод построения нелинейных динамических мод по пространственно распределенным временным рядам

2.1 Введение

В разделе 1.4.2 предыдущей главы был рассмотрен один из наиболее известных методов эмпирической редукции размерности данных – ЭОФ-разложение. В данной главе предлагается новый подход к решению этой задачи, основанный на выделении нелинейных динамических мод. В отличие от традиционного ЭОФ-разложения, а также других методов, учитывающих связь различных компонент многомерного пространственно распределенного вектора наблюдаемых данных в линейной форме, предлагаемый метод способен реконструировать эту связь в нелинейной, физически более адекватной, форме, проецируя данные из высокоразмерного пространства наблюдаемых переменных на низкоразмерное нелинейное многообразие и находя, таким образом, новые «скрытые» низкоразмерные переменные (нелинейные ГК), описывающие главную часть динамики наблюдаемой системы.

Вторая особенность нелинейных динамических мод состоит в том, что они, в отличие от ЭОФ-разложения, предполагают априорную информацию о наличии простого оператора эволюции, обеспечивающего определенный временной масштаб нелинейных ГК, что является одним из ключевых свойств динамических систем, и учитывают, таким образом, причинно-следственную связь последовательных временных отсчетов наблюдаемого временного ряда. В результате нелинейные динамические моды позво-

ляют достичь лучшего выделения скрытых переменных в соответствии с временными масштабами различных компонент наблюдаемой системы. Данная особенность отличает предлагаемый метод от других упомянутых во Введении нелинейных методов редукции размерности.

Формулировка метода и реализующий его алгоритм излагаются в разделе 2.2. Здесь активно используется изложенный в разделе 1.3.1 байесов подход к эмпирическому моделированию. Также в данном разделе приводятся модельные примеры, демонстрирующие работу алгоритма, эффективность представления пространственно распределенных временных рядов в виде нелинейных динамических мод по отношению к линейным методам, а также отличие линейного варианта динамических мод – частного случая нелинейных динамических мод, в котором учтена априорная информация о временном масштабе получаемых ГК, – от обычного ЭОФ-разложения.

В разделах 2.3 и 2.4 новый метод применяется к выявлению и анализу главных мод современного климата по спутниковым измерениям ТПО на интервале с 1981 по 2014 год и доиндустриального климата конца XIX века на основе данных, сгенерированных глобальной климатической моделью INMCM4.0 в соответствующем режиме. Показывается возможность использования нелинейных динамических мод для проверки физических гипотез, детектирования дальних связей в климатической системе. Кроме того, полученные результаты демонстрируют свойство нелинейных динамических мод быть линейными или нелинейными в зависимости от свойств наблюдаемой системы на наблюдаемом интервале времени.

Материалы данной главы представлены в статьях [A2, A3] и публикациях [A7, A13, A16, A32, A35–A50].

2.2 Нелинейные динамические моды

Из выражения (1.39), типичного для линейных методов редукции размерности, ясно, что оно описывает d -мерное линейное многообразие в исходном пространстве, на которое проецируется по тому или иному правилу наблюдаемый временной ряд \mathbf{X} . При этом вектора $\mathbf{v}_1, \dots, \mathbf{v}_d$ являются параметрами, задающими это многообразие, а ГК представляют собой результат проекции на него данных в последовательные моменты времени. Заметим, что выбор размерности d , как правило, не входит в метод разложения и возлагается на пользователя.

Каждая компонента $p_{in} \cdot \mathbf{v}_i$ в выражении (1.39) представляет собой векторный временной ряд, линейно зависящий от скалярного ряда i -й ГК p_{i1}, \dots, p_{iN} , т.е. координаты этой компоненты являются линейно связанными друг с другом. Отсюда очевидно, что линейные методы по построению не могут эффективно учесть возможные нели-

нейные связи между различными компонентами ряда \mathbf{X} . Естественным обобщением выражения (1.39), способным преодолеть указанный недостаток, является следующий вид разложения:

$$\mathbf{x}_n = \sum_{i=1}^d \mathbf{h}_i(p_{in}) + \mathbf{r}_n. \quad (2.1)$$

В отличие от (1.39), здесь каждое слагаемое $\mathbf{h}_i(p_{in})$ обозначает нелинейную векторную функцию \mathbf{h}_i от скалярного аргумента p_{in} , т.е. нелинейное отображение некоторого скалярного временного ряда скрытой переменной p_{i1}, \dots, p_{iN} в пространство исходных переменных. Таким образом, компоненты вектора \mathbf{p}_n представляют собой соответствующие моменту времени t_n с номером n координаты на d -мерном нелинейном многообразии, определяемом отображениями $\mathbf{h}_1, \dots, \mathbf{h}_d$. Данное представление может хорошо описывать данные, формируемые аддитивными нелинейными откликами на скалярные сигналы, и является более общим, чем любое линейное представление¹.

Можно показать [58], что при традиционном ЭОФ-разложении компоненты $p_{in} \cdot \mathbf{v}_i$ могут быть найдены последовательно одна за другой с помощью пошаговой процедуры: на i -м шаге решается задача минимизации вариации i -го остатка разложения одновременно по аргументам \mathbf{v}_i и p_{i1}, \dots, p_{iN} (фактически, решается задача аппроксимации данных в предположении, что текущий остаток разложения – белый гауссов шум). Аналогичная схема применима к разложению (2.1):

- (i) Сначала находится первая компонента на основе аппроксимации данных \mathbf{x}_n с помощью неизвестных отображения \mathbf{h}_1 и скалярного временного ряда p_{11}, \dots, p_{1N} в предположении гауссовости и дельта-коррелированности по времени остатка \mathbf{r}_{1n} (предлагаемый алгоритм байесовой аппроксимации описан ниже):

$$\mathbf{x}_n = \mathbf{h}_1(p_{1n}) + \mathbf{r}_{1n}. \quad (2.2)$$

- (ii) Далее делается вторая итерация, в которой полученный на первой итерации остаток \mathbf{r}_{1n} играет роль данных \mathbf{x}_n и с ним проделывается та же процедура, что и в первой итерации, по поиску второй компоненты \mathbf{h}_2 и p_{21}, \dots, p_{2N} (аналогично выражению (2.2)).
- (iii) Далее берется остаток после второй итерации, с ним проделывается следующая итерация по поиску следующей компоненты и т. д.

Кроме того, традиционное ЭОФ-разложение сфокусировано на максимизации доли вариации исходных временных рядов, захваченной получаемыми линейными ГК. При этом не производится разделение наблюдаемой изменчивости системы по

¹Отметим, что при этом оно, вообще говоря, не описывает произвольное d -мерное многообразие.

временным масштабам: информация о динамике на различных временных масштабах оказывается перемешанной в различных ГК. В предлагаемом методе нелинейного разложения временной масштаб каждой нелинейной ГК будет явно учтен: в рамках байесова подхода будут находиться как оптимальная сложность нелинейной функции \mathbf{h}_i , так и оптимальный временной масштаб нелинейной ГК. В результате предлагаемое разложение должно существенно лучше разделять временные масштабы и соответствующие им подсистемы. Полученные на основе описанной итерационной схемы компоненты $\mathbf{h}_i(p_{in})$ будем называть *нелинейными динамическими модами* (НДМ) [A1].

Заметим, что обычное ЭОФ-решение можно считать частным случаем НДМ (см. раздел 2.2.4 ниже), т.е. оно может оказаться оптимальным для соответствующих входных данных. Кроме того, благодаря оптимизации сложности и временного масштаба НДМ, возможна ситуация, когда на некотором шаге нулевое решение $\mathbf{h}_i(p_{in}) = 0$ является оптимальным (в частности, более оптимальным, чем обычное ЭОФ-решение). Это означает, что при данных априорных предположениях статистически значимые НДМ отсутствуют, и НДМ-разложение закончено. В этом случае к остаткам может быть применено любое другое разложение (например, ЭОФ-разложение).

2.2.1 Параметризация НДМ

В этом и следующем разделах рассмотрим одну итерацию поиска НДМ. Для удобства выкладок, при формулировании алгоритма будем использовать выражение (2.2), в котором опустим у \mathbf{h} , p и \mathbf{r} первый индекс, обозначающий номер НДМ:

$$\mathbf{x}_n = \mathbf{h}(p_n) + \mathbf{r}_n. \quad (2.3)$$

Таким образом, здесь \mathbf{h} – искомое нелинейное отображение, p_n – значение искомой ГК, временной ряд которой обозначается теперь (p_1, \dots, p_N) , в момент дискретного времени t_n , а \mathbf{r}_n – остаток на данной итерации.

Итак, на этой итерации мы имеем остаток от предыдущей итерации в качестве входных данных \mathbf{x}_n . Первой возникающей проблемой при поиске НДМ является чрезвычайно большая размерность D вектора \mathbf{x}_n (в частности, в климатических приложениях, которые рассматриваются в диссертации), что делает прямую задачу аппроксимации практически не реализуемой численно. Поэтому начальным шагом является предварительный ЭОФ-поворот (1.36) входного ряда \mathbf{x}_n (для удобства изложения здесь и далее будем называть ГК, получаемые ЭОФ-разложением, «ЭОФ-ГК»). В результате данные представляются в виде $\mathbf{x}_n = \mathbf{V} \cdot \mathbf{y}_n$, где \mathbf{y} – полный набор ЭОФ-ГК.

После этого в пространстве нескольких главных ЭОФ-ГК будем строить НДМ, параметризованную с помощью полиномов Эрмита:

$$y_{kn} = \sum_{j=0}^m a_{jk} w_j(p_n) + \sigma_h \xi_{kn}, \quad k = 1, \dots, K. \quad (2.4)$$

Здесь $\{w_j(p_n)\}_{j=0}^{\infty}$ – базис нормированных полиномов Эрмита (как будет указано ниже, именно этот базис наиболее удобен в данном случае, хотя использование другого базиса также возможно); a_{jk} – набор $K \cdot (m + 1)$ коэффициентов для K полиномов степени m , который будет также обозначаться как матрица \mathbf{a} ; $\xi_{1n}, \dots, \xi_{Kn}$ – дельта-коррелированные гауссовы шумы с нулевым средним и единичной дисперсией, а σ_h – их амплитуда. В данной параметризации учтена возможность того, что только часть главных ЭОФ являются нелинейно связанными посредством искомого НДМ.

Для остальных ЭОФ-ГК строится тривиальная модель, в которую НДМ непосредственно не входит:

$$y_{kn} = \sigma_k \xi_{kn}, \quad k = K + 1, \dots, D, \quad (2.5)$$

где ξ_{kn} – дельта-коррелированные гауссовы шумы с нулевым средним и единичной дисперсией, σ_k – их амплитуды ($k = K + 1, \dots, D$). На основе полного обратного ЭОФ-поворота (1.37) легко видеть, что выражения (2.4)-(2.5) соответствуют модели в форме (2.3), в которой функция \mathbf{h} задает одномерное нелинейное многообразие (т.е. одномерную кривую), лежащее в подпространстве K главных ЭОФ:

$$\mathbf{h}(p_n) = \sum_{k=1}^K \left(\sum_{j=0}^m a_{jk} w_j(p_n) \right) \cdot \mathbf{v}_k \quad (2.6)$$

2.2.2 Байесов подход к поиску НДМ

В рамках байесова подхода, детально изложенного в разделе 1.3.1, выражение (2.3) рассматривается как модель, реконструирующая НДМ на основе данных \mathbf{X} . Ценовая функция для обучения выбранной модели НДМ, т.е. поиска коэффициентов полинома \mathbf{a} , временного ряда нелинейной ГК p_1, \dots, p_N , а также неизвестных параметров $\boldsymbol{\sigma} = (\sigma_h; \sigma_{K+1}, \dots, \sigma_D)$, представляет собой ПВ неизвестных переменных модели при условии наблюдения \mathbf{X} . Обучение состоит в поиске максимума данной ПВ по этим неизвестным переменным $(\mathbf{a}; p_1, \dots, p_N; \boldsymbol{\sigma})$. Согласно теореме Байеса, данная ПВ может быть представлена следующим образом:

$$\mathcal{P}(\mathbf{a}; p_1, \dots, p_N; \boldsymbol{\sigma} | \mathbf{X}) \propto \mathcal{P}(\mathbf{X} | \mathbf{a}, p_1, \dots, p_N; \boldsymbol{\sigma}) \cdot \mathcal{P}_{pr}(\mathbf{a}, p_1, \dots, p_N; \boldsymbol{\sigma}). \quad (2.7)$$

Здесь $\mathcal{P}(\mathbf{X} | \mathbf{a}, p_1, \dots, p_N; \boldsymbol{\sigma})$ – функция правдоподобия, т.е. вероятность того, что данные \mathbf{X} сгенерированы моделью (2.3) при конкретных значениях параметров

$(\mathbf{a}; p_1, \dots, p_N; \boldsymbol{\sigma})$. По аналогии с разделом 1.3.1, она может быть легко написана на основе уравнений модели НДМ (2.4)-(2.5) и предположений о дельта-коррелированности и гауссовости остатков в них:

$$\mathcal{P}(\mathbf{X}|\mathbf{a}; p_1, \dots, p_N; \boldsymbol{\sigma}) = \prod_{n=1}^N \prod_{k=1}^D \mathcal{P}(x_{kn}|\mathbf{a}, p_n, \boldsymbol{\sigma}) = \prod_{n=1}^N \prod_{k=1}^D \mathcal{P}(y_{kn}|\mathbf{a}, p_n, \boldsymbol{\sigma}), \quad (2.8)$$

где при $k = 1, \dots, K$:

$$\mathcal{P}(y_{kn}|\mathbf{a}, p_n, \boldsymbol{\sigma}) = (2\pi\sigma_h^2)^{-\frac{1}{2}} \cdot \exp\left(-\frac{\left|y_{kn} - \sum_{j=0}^m a_{jk}w_j(p_n)\right|^2}{2\sigma_h^2}\right), \quad (2.9)$$

и при $k = K + 1, \dots, D$:

$$\mathcal{P}(y_{kn}|\mathbf{a}, p_n, \boldsymbol{\sigma}) = (2\pi\sigma_k^2)^{-\frac{1}{2}} \cdot \exp\left(-\frac{|y_{kn}|^2}{2\sigma_k^2}\right). \quad (2.10)$$

Максимум (2.8) достигается при минимальном значении вариации остатка \mathbf{r}_n в (2.3) и, фактически, задает проекцию данных на главное многообразие, определяемое функцией \mathbf{h} , и результат проекции – координаты p_1, \dots, p_N на многообразии. При этом, вообще говоря, существует бесконечное количество способов выбора координат на одном и том же многообразии, среди которых есть более или менее адекватные решаемой практической задаче, т.е. в задаче максимизации (2.8) имеется вырождение по переменным p_1, \dots, p_N и параметрам отображения \mathbf{h} , особенно существенное в случае нелинейного многообразия (например, при наличии участков с высокой степенью кривизны и самопересечений). Поэтому в данном случае особенно важен выбор априорной ПВ $\mathcal{P}_{pr}(\mathbf{a}, p_1, \dots, p_N; \boldsymbol{\sigma})$ для неизвестных параметров, т.е. выбор требуемых свойств решения, за счет знания которых происходит регуляризация задачи проекции данных на многообразие. Для задачи поиска НДМ мы будем использовать априорную ПВ, факторизованную в форме произведения трех ПВ, которые пояснены ниже:

$$\mathcal{P}_{pr}(\mathbf{a}, p_1, \dots, p_N; \boldsymbol{\sigma}) = \mathcal{P}_{pr}(p_1, \dots, p_N) \cdot \mathcal{P}_{pr}(\mathbf{a}) \cdot \mathcal{P}_{pr}(\boldsymbol{\sigma}). \quad (2.11)$$

Априорная ПВ для нелинейных ГК p_1, \dots, p_N

Главное априорное предположение, которое будет сделано для НДМ, состоит в явном учете динамической природы получаемых нелинейных ГК p_1, \dots, p_N , а именно, в требовании наличия связи каждого значения p_n с предшествующим по времени значением p_{n-1} . Конкретно, мы будем использовать априорную ПВ для p_1, \dots, p_N в следующем виде:

$$\mathcal{P}_{pr}(p_1, \dots, p_N) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{p_1^2}{2}\right) \cdot \prod_{n=2}^N \frac{1}{\sqrt{2\pi\sigma_p^2}} \exp\left(-\frac{|p_n - b \cdot p_{n-1}|^2}{2\sigma_p^2}\right), \quad (2.12)$$

$$b = \exp(-1/\tau), \quad \sigma_p = \sqrt{1 - b^2}. \quad (2.13)$$

Можно показать (см. ниже), что ПВ (2.12) представляет собой ПВ для ансамбля реализаций длины N стационарного красного шума с единичной дисперсией и временем автокорреляции τ , имеющего оператор эволюции:

$$p_n = bp_{n-1} + \sigma_p \eta_n, \quad (2.14)$$

где η_n – белый гауссов шум с нулевым средним и единичной дисперсией.

Несмотря на общность такого априорного предположения, оно позволяет ограничить класс получаемых решений за счет исключения из рассмотрения коротких временных масштабов (меньше τ) в получаемой ГК. Другими словами, априорная ПВ (2.12) предполагает существование оператора эволюции для временного ряда ГК, действующего в определенном диапазоне временных масштабов, что является априорным базовым свойством реконструируемой динамической системы. Тем самым, введение данного априорного множителя осуществляет вышеупомянутую регуляризацию операции проецирования данных на соответствующее одномерное многообразие.

Заметим также, что априорное ограничение (2.12) приводит к тому, что формально каждое значение скрытой переменной p_n зависит сразу от всех значений исходного временного ряда $\mathbf{x}_1, \dots, \mathbf{x}_N$ через ценовую функцию (2.7), однако наибольший вклад в p_n дают элементы исходного временного ряда, ближайšie по времени к \mathbf{x}_n . Такая зависимость текущего состояния НДМ одновременно от прошлых и будущих значений наблюдаемых данных приводит к наличию «краевого эффекта» метода: p_n определяются менее точно (смещенно) на концах временного ряда.

Итак, покажем, что конечная реализация красного шума (2.14) удовлетворяет распределению (2.12). В самом деле, уравнение (2.14) определяет стохастический марковский процесс первого порядка с переходной вероятностью:

$$\mathcal{P}(p_n|p_{n-1}) = \frac{1}{\sqrt{2\pi\sigma_p^2}} \exp\left(-\frac{|p_n - b \cdot p_{n-1}|^2}{2\sigma_p^2}\right). \quad (2.15)$$

Далее, стационарный режим процесса (2.14) характеризуется инвариантной мерой [11] $\mu(p)$ для переменной p , удовлетворяющей следующему уравнению:

$$\mu(p_n) = \int \mathcal{P}(p_n|p_{n-1})\mu(p_{n-1})dp_{n-1} \quad (2.16)$$

Легко проверить, что решение уравнения (2.16) есть следующая гауссова ПВ:

$$\mu(p) = \frac{1}{\sqrt{2\pi\sigma_\mu^2}} \exp\left(-\frac{p^2}{2\sigma_\mu^2}\right) \quad (2.17)$$

$$\sigma_\mu^2 = \frac{\sigma_p^2}{1 - b^2} \quad (2.18)$$

Рассмотрим теперь конечную реализацию (p_1, \dots, p_N) процесса (2.14) в стационарном режиме. ПВ данной реализации полностью определяется ПВ начального состояния p_1 , т.е. инвариантной мерой (2.17), и переходной вероятностью марковского процесса (2.15):

$$\mathcal{P}(p_1, \dots, p_N) = \mathcal{P}(p_N | p_{N-1}) \cdot \dots \cdot \mathcal{P}(p_2 | p_1) \cdot \mu(p_1) \quad (2.19)$$

Подставляя (2.15), (2.17), (2.18), (2.13) в (2.19), получаем выражение (2.12) для априорной ПВ.

Заметим, что в условиях (2.13) дисперсия процесса (2.14) равна единице ($\sigma_\mu^2 = 1$). Кроме того, используя аналогичное (2.19) выражение для $\mathcal{P}(p_n, \dots, p_{n+t})$, несложно получить, что величина корреляции между p_n и p_{n+t} равна:

$$\langle p(n) \cdot p(n+t) \rangle_{\mathcal{P}(p_n, \dots, p_{n+t})} = b^t = \exp(-t/\tau) \quad (2.20)$$

Другими словами, величина τ , введенная в (2.13), действительно является временем автокорреляции красного шума (2.14).

Априорная ПВ для коэффициентов НДМ

Несмотря на выполнение условия единичной вариации $\sigma_p^2 + b^2 = 1$ (см. (2.13), (2.18)), ПВ (2.12) не исключает случая с бесконечно малой вариацией временного ряда p_1, \dots, p_N и связанного с этим вырождения в задаче поиска НДМ. Поэтому целью априорной ПВ $\mathcal{P}_{pr}(\mathbf{a})$ для коэффициентов полинома \mathbf{a} является снятие этого вырождения (путем исключения бесконечно больших значений коэффициентов полинома) при сохранении максимальной общности априорно допустимых форм искомого нелинейного многообразия. Идея выбора данной ПВ состоит в том, что она должна (i) обеспечивать модели НДМ возможность воспроизводить матрицу ковариаций исходных данных и (ii) быть инвариантной по отношению к ортогональному повороту базиса полиномов в (2.4), т.е. не зависеть от выбора конкретного ортонормированного базиса полиномов (при фиксированном определении скалярного произведения в пространстве полиномов, т.е. их ортогональности).

Условие (i) может быть выражено следующим равенством:

$$\frac{1}{N} \sum_{n=1}^N y_{kn} y_{k'n} = \left\langle \frac{1}{N} \sum_{n=1}^N \left[\sum_{j=0}^m a_{jk} w_j(p_n) \right] \cdot \left[\sum_{j'=0}^m a_{j'k'} w_{j'}(p_n) \right] \right\rangle_{\mathcal{P}_{pr}(p_1, \dots, p_N) \mathcal{P}_{pr}(\mathbf{a})} \quad (2.21)$$

В левой части выражения (2.21) стоит ковариация временных рядов данных, соответствующих компонентам с номерами k и k' , а в правой – ковариация компонент модели

НДМ (2.4) с теми же номерами, вычисленная на априорном распределении параметров НДМ $\mathcal{P}_{pr}(p_1, \dots, p_N) \mathcal{P}_{pr}(\mathbf{a})$. Нашей целью является выбор $\mathcal{P}_{pr}(\mathbf{a})$, удовлетворяющего равенству (2.21).

Условие (ii), фактически, состоит в том, что равенство (2.21) сохраняется при замене в нем $w_j(p)$ на $\tilde{w}_j(p) = \sum_{s=0}^m U_{js} w_s(p)$, где \mathbf{U} – произвольная ортогональная матрица размера $(m+1) \times (m+1)$, $j = 0, \dots, m$.

Оказывается, что при использовании нормированных полиномов Эрмита, которые ортогональны относительно меры $\mu(p)$ (см. (2.17)), оба условия выполняются для следующего гауссового распределения $\mathcal{P}_{pr}(\mathbf{a})$:

$$\mathcal{P}_{pr}(\mathbf{a}) = \prod_{k=1}^K \prod_{j=0}^m \sqrt{\frac{m+1}{2\pi\lambda_k}} \exp\left(-\frac{m+1}{2\lambda_k} |a_{jk}|^2\right), \quad (2.22)$$

где $\lambda_k = \frac{1}{N} \sum_{n=1}^N y_{kn}^2$ – вариации ЭОФ-ГК, в пространстве которых строится модель НДМ.

Действительно, поскольку ЭОФ-ГК не коррелируют друг с другом, левая часть выражения (2.21) равна $\delta_{kk'} \lambda_k$, где $\delta_{kk'}$ – символ Кронекера. Правая часть выражения (2.21), с учетом (2.22), может быть преобразована следующим образом:

$$\begin{aligned} & \left\langle \frac{1}{N} \sum_{n=1}^N \left[\sum_{j=0}^m a_{jk} w_j(p_n) \right] \cdot \left[\sum_{j'=0}^m a_{j'k'} w_{j'}(p_n) \right] \right\rangle_{\mathcal{P}_{pr}(p_1, \dots, p_N) \mathcal{P}_{pr}(\mathbf{a})} = \\ & = \left\langle \left[\sum_{j=0}^m a_{jk} w_j(p) \right] \cdot \left[\sum_{j'=0}^m a_{j'k'} w_{j'}(p) \right] \right\rangle_{\mu(p) \mathcal{P}_{pr}(\mathbf{a})} = \\ & = \sum_{j=0}^m \sum_{j'=0}^m \langle a_{jk} a_{j'k'} \rangle_{\mathcal{P}_{pr}(\mathbf{a})} \cdot \langle w_j(p) w_{j'}(p) \rangle_{\mu(p)} = \\ & = \sum_{j=0}^m \sum_{j'=0}^m \frac{\delta_{kk'} \delta_{jj'} \lambda_k}{m+1} \cdot \langle w_j(p) w_{j'}(p) \rangle_{\mu(p)} = \\ & = \frac{\delta_{kk'} \lambda_k}{m+1} \sum_{j=0}^m \langle w_j^2(p) \rangle_{\mu(p)} \end{aligned} \quad (2.23)$$

Далее, поскольку нормированные полиномы Эрмита являются ортонормированными относительно меры $\mu(p)$ при $\sigma_\mu^2 = 1$, для них справедливо равенство $\langle w_j(p) w_{j'}(p) \rangle_{\mu(p)} = \delta_{jj'}$. Подставляя это равенство в (2.23) и выполняя суммирование, получаем $\delta_{kk'} \lambda_k$, тем самым доказывая выполнение условия (i). Кроме того, любой ортогональный поворот нормированных полиномов Эрмита сохраняет ортонормированность нового базиса $\tilde{w}_j(p)$ относительно меры $\mu(p)$, что, учитывая (2.23), приводит к выполнению условия (ii).

Заметим, что выбор соответствующего априорного распределения коэффициентов \mathbf{a} , может быть произведен, вообще говоря, для любого базиса $\{w_j(p)\}_{j=0}^m$ (не обяза-

тельно даже полиномиального). Более детальное и общее решение данной задачи описано в приложении к статье [А3].

В результате использование априорной ПВ (2.22) ограничивает значения коэффициентов полинома и снимает вырождение, связанное с бесконечно малой вариацией p_1, \dots, p_N , присутствующее в ПВ (2.12).

Априорная ПВ для параметра σ

Что касается априорного распределения параметров σ , рассматриваемая задача аппроксимации по ним хорошо регуляризована: можно показать, что для каждого набора \mathbf{a} и p_1, \dots, p_N существует единственное решение σ . Поэтому априорное распределение $\mathcal{P}_{pr}(\sigma)$, как и в случае $m_g = 0$ в разделе 1.3.1, берется однородным в некотором достаточно широком интервале значений и на результат максимизации (2.7) не влияет:

$$\mathcal{P}_{pr}(\sigma) = const. \quad (2.24)$$

Оптимизация структурных параметров НДМ

Можно заметить, с учетом всех сделанных утверждений и определений, что в главное уравнение модели НДМ (2.4) и в выражение для ценовой функции (2.7) входят три структурных параметра, выбор которых до сих пор остался не определенным – временной масштаб τ нелинейных ГК, степень полинома m и параметр предварительной редукции K (количество ЭОФ-ГК, в пространстве которых строится НДМ). Данные параметры определяют допустимую сложность модели и ансамбль допустимых решений. Ясно, что при слишком простой модели (например, $m = 1$) точность аппроксимации данных с помощью НДМ будет слишком плохой, и модель может по построению не учесть важные свойства, содержащиеся в исходных данных; если же модель, наоборот, чересчур сложна (например, очень большая степень m), то модель опишет наблюдаемые данные с очень хорошей точностью, воспроизведя при этом и реализацию шума, присутствующего в данных. В результате нелинейное многообразие будет иметь сложную форму и не будет иметь физического смысла, т.к. оно полностью зависит от реализации шума в системе (переобученная модель). Поэтому необходимо найти оптимальные значения структурных параметров, при которых модель НДМ не слишком проста, но и не является переобученной.

В рамках байесова подхода эта задача решается с помощью байесовой обоснованности, определение и метод приближенного вычисления которой в данном случае полностью аналогичны изложенным в разделе 1.3.1. В соответствии с определением, байесова обоснованность характеризует вероятность модели с фиксированными

структурными параметрами (τ, m, K) и произвольными параметрами $(a; p_1, \dots, p_N; \sigma)$ сгенерировать наблюдаемый временной ряд \mathbf{X} и выражается усредненной по параметрам $(a; p_1, \dots, p_N; \sigma)$ правой частью выражения (2.7):

$$\mathcal{P}(\mathbf{X}|\tau, m, K) = \int_{\mathbf{a}; p_1, \dots, p_N; \sigma} \mathcal{P}(\mathbf{X}|\mathbf{a}; p_1, \dots, p_N; \sigma) \cdot \mathcal{P}_{pr}(\mathbf{a}; p_1, \dots, p_N; \sigma). \quad (2.25)$$

Здесь интеграл берется по всем параметрам $\mathbf{a}; p_1, \dots, p_N; \sigma$ в бесконечных пределах. Оптимальными будем считать значения (τ, m, K) , которые максимизируют вероятность (2.25).

Как и в разделе 1.3.1, для приближенного вычисления оптимальности удобно рассмотреть в качестве ценовой функции для обучения модели минус логарифм правой части выражения (2.7):

$$\Psi_{\mathbf{X}, \tau, m, K}(\mathbf{a}; p_1, \dots, p_N; \sigma) = -\ln [\mathcal{P}(\mathbf{X}|\mathbf{a}; p_1, \dots, p_N; \sigma) \cdot \mathcal{P}_{pr}(\mathbf{a}; p_1, \dots, p_N; \sigma)]. \quad (2.26)$$

Тогда обучение модели будет состоять в поиске минимума $(\bar{\mathbf{a}}; \bar{p}_1, \dots, \bar{p}_N; \bar{\sigma})$ этой функции, а приближенное вычисление интеграла (2.25) методом, используемым в разделе 1.3.1, приведет к следующему выражению для минус логарифма байесовой обоснованности:

$$-\ln \mathcal{P}(\mathbf{X}|\tau, m, K) \cong \Psi_{\mathbf{X}, \tau, m, K}(\bar{\mathbf{a}}; \bar{p}_1, \dots, \bar{p}_N; \bar{\sigma}) + \frac{1}{2} \ln \det \frac{1}{2\pi} \nabla \nabla^T \Psi_{\mathbf{X}, \tau, m, K}. \quad (2.27)$$

Здесь $\nabla \nabla^T \Psi_{\mathbf{X}, \tau, m, K}$ – матрица вторых производных (гессиан) функции $\Psi_{\mathbf{X}, \tau, m, K}$ по параметрам $(\mathbf{a}; p_1, \dots, p_N; \sigma)$ в точке минимума $(\bar{\mathbf{a}}; \bar{p}_1, \dots, \bar{p}_N; \bar{\sigma})$. Как и в разделе 1.3.1, при наличии нескольких локальных минимумов для оценки обоснованности используется тот, который соответствует наименьшему значению выражения (2.27).

Алгоритм поиска НДМ

На основе приведенных ценовых функций можно сформулировать следующий алгоритм для поиска НДМ:

1. Генерируется случайный набор структурных параметров (τ, m, K) ;
2. Генерируются случайные значения $(\mathbf{a}; p_1, \dots, p_N; \sigma)$ и, стартуя с этих значений, производится минимизация выражения (2.26) квазиградиентным методом. Таким образом, находится один из локальных минимумов ценовой функции $\Psi_{\mathbf{X}, \tau, m, K}$;
3. Вычисляется ценовая функция для оптимальности (2.27) для данного набора (τ, m, K) на основе найденных значений $(\bar{\mathbf{a}}; \bar{p}_1, \dots, \bar{p}_N; \bar{\sigma})$, соответствующих найденному минимуму (2.26);

4. Шаги 1-3 повторяются нужное количество раз;
5. Из полученного ансамбля решений в качестве оптимального выбирается решение $(\bar{\mathbf{a}}; \bar{p}_1, \dots, \bar{p}_N; \bar{\sigma})$, найденное при структурных параметрах (τ, m, K) , соответствующих минимальному значению ценовой функции для оптимальности (2.27). Это решение и есть НДМ.

Наиболее ресурсоемким в данном алгоритме является шаг минимизации 2. Как правило, для достижения ответа требуется достаточно большое число итераций, включающих такую минимизацию (шаг 4). При этом, достоинством предложенного алгоритма поиска оптимальной модели, как при расчете НДМ, так и при расчете оптимальной стохастической модели в разделе 1.3.1, является возможность его использования в режиме параллельных вычислений.

2.2.3 Модельный пример

Продемонстрируем на модельном примере метод поиска одной НДМ (один шаг полного НДМ-разложения). Для этой цели сгенерируем с помощью некоторой динамической системы скалярный временной ряд q_1, \dots, q_N и применим к нему зашумленное нелинейное преобразование $\phi : \mathbb{R} \rightarrow \mathbb{R}^2$:

$$\mathbf{x}_n = \phi(q_n) + \zeta_n, \quad (2.28)$$

где $\zeta_n \in \mathbb{R}^2$ – гауссов дельта-коррелированный случайный процесс. В этом случае задача НДМ-разложения состоит в том, чтобы по наблюдаемому двумерному ряду $\mathbf{x}_1, \dots, \mathbf{x}_N$ (см. рис. 2.1a, 2.1c) реконструировать как отображение ϕ , так и временной ряд скрытой переменной q_1, \dots, q_N настолько хорошо, насколько это позволяет сделать шум.

В качестве нелинейного преобразования ϕ возьмем следующую векторную функцию:

$$\phi(q_n) = \begin{pmatrix} 0.12q_n^3 + 0.89q_n^2 - 0.35q_n - 0.83 \\ 0.81q_n^3 - 0.26q_n^2 - 2.30q_n + 0.49 \end{pmatrix}. \quad (2.29)$$

Данная функция задает нелинейную кривую в \mathbb{R}^2 , которая показана на рис. 2.2a.

Для задания временного ряда скрытой переменной q_1, \dots, q_N используем переменную $\tilde{y}(t)$ системы Лоренца [80] с «классическими» значениями параметров ($\tilde{b} = \frac{8}{3}, \tilde{\sigma} = 10, \tilde{r} = 28, \sigma = 0$ в уравнении (1.31)), функционирующей в хаотическом режиме. В данном режиме траектория системы переключается между двумя крыльями аттрактора Лоренца, совершая на каждом крыле повторяющееся вращательное движение, при этом среднее время между переключениями примерно в 2.26 раз больше характерного периода вращательного движения. Сначала получим дискретный

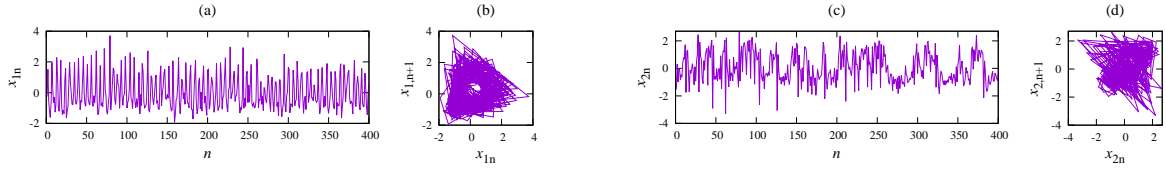


Рис. 2.1: Двумерный временной ряд, сгенерированный системой (2.28): временные ряды наблюдаемых переменных (a, c) и их представление в координатах с задержкой (b, d).

временной ряд \tilde{y}_n взятием $N = 400$ значений переменной $\tilde{y}(t)$ с шагом $\Delta t = 0.14$ в установившемся режиме системы, что соответствует в среднем 5.3 измерений за один характерный период вращательного движения и в среднем 33.2 переключений за интервал наблюдения, т.е. временной ряд \tilde{y}_n является достаточно гладким и при этом значительно покрывает аттрактор. Далее, с помощью нормировки ряда \tilde{y}_n на единичную вариацию получим временной ряд q_n . Данный временной ряд показан на рис. 2.3а. Также на рис. 2.3б показана проекция фазовой траектории системы Лоренца на плоскость задержанных переменных q_n, q_{n+1} (аналогично рис. 1.1).

Наконец, амплитуду обеих компонент процесса ζ_n возьмем равной 0.4. Наблюдаемый зашумленный временной ряд \mathbf{x}_n показан на плоскости переменных (x_1, x_2) на рис. 2.2б. Стандартное отклонение истинных (незашумленных) образов $\phi_1(q_n)$ и $\phi_2(q_n)$ скрытой переменной q_n примерно равно 1, поэтому уровень шума в наблюдаемых переменных можно оценить как σ_ϕ , т.е. 40%.

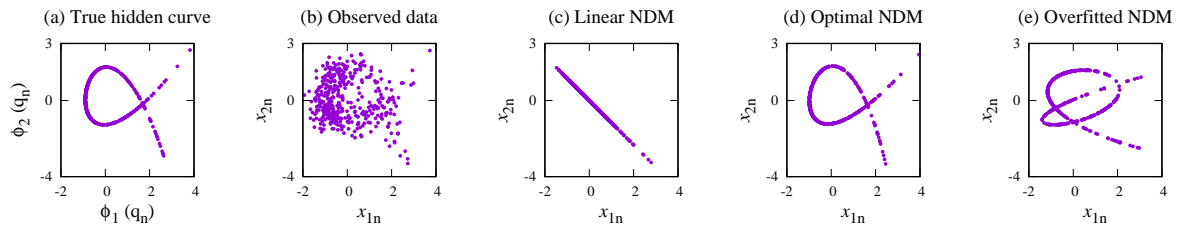


Рис. 2.2: Изображение в пространстве наблюдаемых переменных (x_1, x_2) : (a) истинное нелинейное отображение (2.29) системы (2.28); (b) наблюдаемый временной ряд системы (2.28); (c) слишком простая (линейная) НДМ со структурными параметрами $K = 2$, $m = 1$, $\tau = 10$; (d) НДМ с оптимальными структурными параметрами $K = 2$, $m = 3$, $\tau = 7.5$; (e) переобученная НДМ со структурными параметрами $K = 2$, $m = 7$, $\tau = 2.6$.

Из рис. 2.1b и 2.1d видно, что характерная структура аттрактора Лоренца в

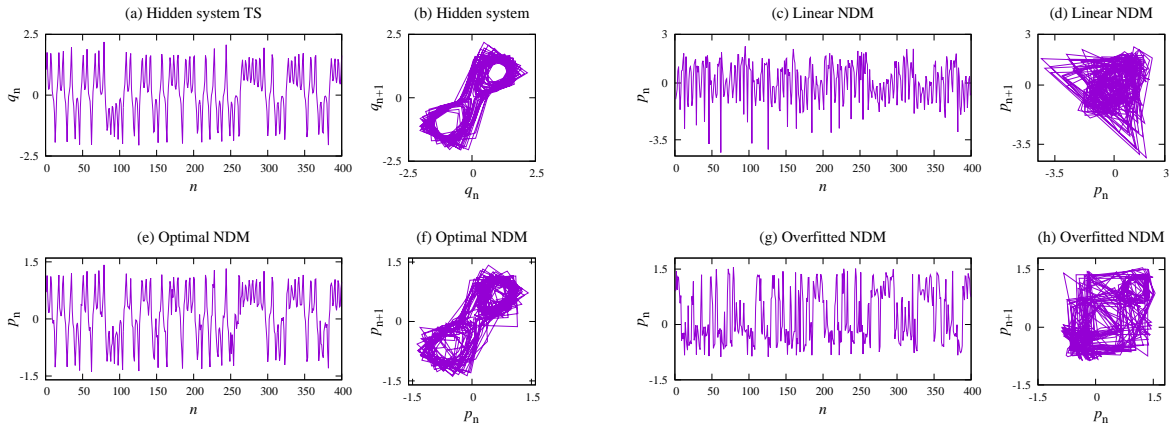


Рис. 2.3: Временные ряды скрытой переменной (а, с, е, г) и их представление в координатах с задержкой (b, d, f, h) для случаев: (а, b) система (2.28); (с, d) слишком простая (линейная) НДМ со структурными параметрами $K = 2$, $m = 1$, $\tau = 10$; (е, f) НДМ с оптимальными структурными параметрами $K = 2$, $m = 3$, $\tau = 7.5$; (г, h) переобученная НДМ со структурными параметрами $K = 2$, $m = 7$, $\tau = 2.6$.

наблюдаемых переменных (x_1, x_2) полностью разрушена нелинейностью и шумом. Фактически, в данной ситуации НДМ отвечает на важный вопрос: какая одномерная фазовая переменная наиболее оптимально учитывает динамические свойства наблюдаемой системы? В нашем модельном примере такая переменная представляет собой проекцию данных на истинную кривую, изображенную на рис. 2.2а, и задача состоит в отыскании как этой кривой, так и фазовой переменной.

Результат поиска НДМ с помощью разработанного метода показан на рис. 2.2 и 2.3. Нелинейная кривая, определяемая коэффициентами \mathbf{a} при оптимальных структурных параметрах НДМ, минимизирующих ценовую функцию для оптимальности (2.27), изображена на рис. 2.2d. Видно, что эта кривая очень близка к истинной кривой, а временной ряд p_1, \dots, p_N найденной нелинейной ГК (рис. 2.3е) при представлении в координатах с задержками демонстрирует структуру аттрактора Лоренца (рис. 2.3f). В полном соответствии с байесовым критерием оптимальности кривые для НДМ, полученных при неоптимальных структурных параметрах, менее похожи на истинную кривую: слишком простая (линейная) модель НДМ является недостаточно «гибкой» (в данном случае она получилась практически идентичной проекции на главную ЭОФ, рис. 2.2с), в то время как слишком сложная модель дает сложную «переобученную» кривую (рис. 2.2е), частично описывающую конкретную реализацию шума в (2.28). Ни линейная (рис. 2.3с, d), ни переобученная (рис. 2.3г, h) НДМ не позволяют выявить настолько хорошую структуру аттрактора в найденной ГК, насколько позволяет

оптимальная НДМ.

2.2.4 Отличие линейных НДМ от ЭОФ

Рассмотрим случай НДМ со структурным параметром $m = 1$, т.е. параметризованной с помощью полинома первой степени, и оптимизированной по другим структурным параметрам K, τ . В этом случае мы имеем дело с поиском одномерного линейного многообразия (прямой) в пространстве исходных переменных. Данный линейный вариант НДМ для краткости будем называть *линейной динамической модой* (ЛДМ).

С одной стороны, ЛДМ можно считать схожими с ЭОФ, потому что они удовлетворяют такой же линейной форме представления данных (1.36). С другой стороны, ЭОФ можно считать частным случаем ЛДМ (и, следовательно, НДМ), потому что ЭОФ-решение соответствует, с точностью до перенормировки, случаю бесконечно малого (нулевого) значения структурного параметра τ . В самом деле, в этом случае в априорной ПВ (2.12) b близко к 0 и σ_p близко к 1, т.е. данная ПВ ограничивает лишь вариацию сигнала p_1, \dots, p_N , но не его форму. В результате максимизация ПВ (2.7) приводит к линейному многообразию (прямой), ориентированному в соответствии с наименьшей вариацией остатка (при условии оставшихся априорных ограничений (2.12) и (2.22) на нормы p и \mathbf{a} , которые выбраны так, чтобы ЛДМ всегда имела возможность воспроизвести вариацию исходных ЭОФ-ГК).

Таким образом, в процессе байесовой оптимизации ЛДМ происходит выбор из различных структурных параметров, среди которых есть параметры, соответствующие и ЭОФ-решению. Фактически, мы ищем многообразие, которое стремится не только описать максимальную долю вариации исходных данных, но и найти временной ряд ГК с «наиболее вероятным» временным масштабом для рассматриваемой системы. При этом, поскольку априорная ПВ (2.12) учитывает наличие связи между последовательными временными отсчетами временного ряда скрытой переменной p_1, \dots, p_N , каждое найденное значение p_n определяется, вообще говоря, не одним текущим значением \mathbf{x}_n , а последовательностью данных. В результате даже в классе линейных преобразований полученные моды оказываются ориентированными согласно направлениям в фазовом пространстве, вдоль которых происходит динамика системы, в отличие от ЭОФ-разложения, при котором рассматриваются лишь корреляции временных рядов данных лишь в один момент времени, а связь последовательных временных отсчетов полностью игнорируется. И только в случае, когда временной масштаб τ найденного сигнала не превышает нескольких временных отсчетов, ЛДМ дают тот же результат, что и ЭОФ (как получилось в модельном примере, см. рис. 2.2с). Если же в имеющихся данных недостаточно информации для того, чтобы выделить какие-либо

линейные/нелинейные структуры (случай «белого шума»), то оптимальное значение τ получится бесконечно большим, что приведет к постоянному временному ряду $p_n = const$, т.е. к практически нулевой ЛДМ/НДМ.

Приведем простой пример, иллюстрирующий разницу между ЛДМ и ЭОФ. Пусть двумерный временной ряд \mathbf{x}_n содержит периодический сигнал, пропорциональный $y_n = \sin \frac{2\pi n}{T}$, возмущенный белым гауссовым шумом ξ_n с единичной вариацией компонент и их ковариацией $\langle \xi_{1n}\xi_{2n} \rangle_n = 0.5$:

$$\begin{aligned} x_{1n} &= 0.5y_n + \xi_{1n}, \\ x_{2n} &= -0.2y_n + \xi_{2n}. \end{aligned} \tag{2.30}$$

Будем рассматривать две реализации временного ряда $\mathbf{x}_1, \dots, \mathbf{x}_N$, каждая из которых имеет длину $N = 1000$ отсчетов: одна содержит «медленный» сигнал y с периодом $T = 500$, а для другой период $T = 20$. Обе реализации, сгенерированные в соответствии с уравнением (2.30), нормируем на единичную вариацию, т.е. $\langle x_{1n}^2 \rangle_n = \langle x_{2n}^2 \rangle_n = 1$. В левой части рис. 2.4 сверху и снизу на плоскости (x_1, x_2) изображены эти реализации (точками) и их невозмущенные компоненты, отвечающие сигналу y_n (синими линиями).

Сравним линейные преобразования, соответствующие ЛДМ и ЭОФ (рис. 2.4). Неудивительно, что в рассматриваемом случае ЭОФ ориентированы вдоль диагонали на плоскости (x_1, x_2) , независимо от направления, связанного с полезным сигналом y (синяя линия на рис. 2.4 слева), и величины T , поскольку ряды переменных x_1 и x_2 имеют единичную вариацию и положительную ковариацию (она определяется, в первую очередь, положительной ковариацией компонент шума в (2.30)). Можно легко показать аналитически, что главный собственный вектор соответствующей двумерной матрицы ковариаций (т.е. ЭОФ) равен $\begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}$ независимо от конкретного значения ковариации.

В случае ЛДМ такого не происходит: метод поиска ЛДМ стремится выявить автокорреляционные свойства скрытого процесса y за счет правильного выбора временного масштаба τ – структурного параметра ЛДМ, находимого в рамках байесова критерия оптимальности. В результате главная ЛДМ настраивается более точно на полезный сигнал, имеющий большой временной масштаб (по сравнению с временным масштабом шума): из рис. 2.4 видно, что при $T = 500$ полученная ЛДМ стремится ориентироваться в том же направлении, что и невозмущенная компонента, соответствующая сигналу y , а соответствующая ГК по форме воспроизводит временной ряд полезного сигнала. Более того, длина отрезка прямой, на которой сконцентрирован сигнал ЛДМ, примерно равна длине такого же отрезка, соответствующего истинному сигналу, т.е. ЛДМ эффективно отфильтровывает шумовую компоненту; в то же

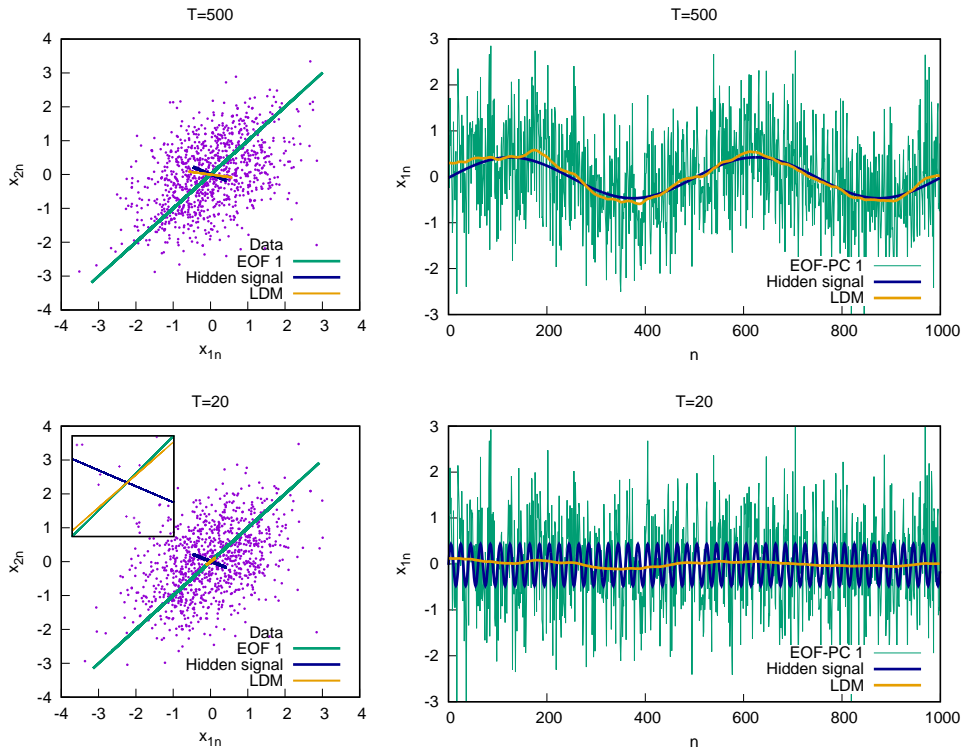


Рис. 2.4: Главная ЛДМ временных рядов данных, сгенерированных уравнением (2.30). Верхняя половина относится к временному ряду данных x , сгенерированному при $T = 500$, нижняя – при $T = 20$. В левой части показаны полученные моды на плоскости переменных (x_1, x_2) : ЛДМ изображена желтым цветом, главная ЭОФ изображена зеленой линией, фиолетовые точки соответствуют исходным данным, синяя линия показывает ассоциированную с сигналом y часть данных. В углу левого нижнего изображения ($T = 20$) показана увеличенная копия центральной части этого изображения. В правой части показаны временные ряды, соответствующие вкладу главной ЛДМ (желтым), главной ЭОФ (зеленым) и истинной периодической скрытой переменной y (синим) в наблюдаемую переменную x_1 .

время, соответствующий ЭОФ отрезок прямой по величине совпадает с диапазоном зашумленных данных.

Наконец, говоря о фильтрации шума с помощью ЛДМ, можно ожидать, что, при условии небольшой вариации полезного сигнала y большой временной масштаб этого сигнала лучше выделяется с помощью ЛДМ. Действительно, это полностью подтверждается правой частью рис. 2.4: сигнал y захватывается ЛДМ при коротком периоде $T = 20$ хуже, чем при $T = 500$. Дополнительно, как это видно из левой части рис. 2.4, в случае $T = 20$ соответствующий ЛДМ отрезок прямой имеет еще меньшую длину и ориентирован существенно ближе к ЭОФ. Причиной этому является то,

что, в силу принципиальной плохой отличимости моды с коротким масштабом и/или с низкой вариацией на фоне белого шума, метод ЛДМ стремится трактовать все данные как шум. Отсюда ясно, что в предельном случае очень большого уровня шума (очень плохо разрешенного полезного сигнала) найденное многообразие стягивается в точку, что означает отсутствие в наблюдаемых данных структур, распознаваемых методом ЛДМ/НДМ, и является критерием окончания разложения (см. вводную часть раздела 2.2).

2.3 Главные нелинейные моды климатической изменчивости по данным 1981-2014 гг.

В данном разделе разработанный метод поиска НДМ применяется к результатам современных спутниковых измерений ТПО, доступных начиная с 1980-ых годов. Сравнительно небольшая длина выбранного для анализа интервала (33 года) позволяет предположить, что влияние медленных внешних воздействий и трендов, таких как эмиссии парниковых газов, не очень велико, и, таким образом, ассоциировать искомые моды ТПО, главным образом, с проявлением естественной («внутренней») изменчивости климата на межгодовых и декадных масштабах (см. Введение).

2.3.1 Наблюдаемые данные и результат их НДМ-разложения

Для анализа были взяты временные ряды ежемесячного поля ТПО NOAA_OI.v2² [103], покрывающие поверхность Земли с разрешением 1×1 градус, на 33-летнем интервале с ноября 1981 года по октябрь 2014 года. В результате размерность наблюдаемого поля данных составила $D = 44219$ (число узлов пространственной сетки, приходящихся на поверхность океана). Предварительная нормировка данных ТПО описана в разделе 1.4.1, исключая шаг получения АТПО.

В результате НДМ-разложения этих данных было найдено три главных нелинейных моды (четвертая НДМ получилась равной нулю, см. вводную часть раздела 2.2). Соответствующие им временные ряды p_{i1}, \dots, p_{iN} ($i = 1, 2, 3$) представлены на рис. 2.5а в порядке убывания энергии, т.е. захваченной ими доли вариации ТПО (сверху вниз). Толщина линии на этом рисунке соответствует 95% байесову доверительному интервалу³, характеризующему неопределенность оценки скрытого состояния p_{in} .

²http://iridl.ldeo.columbia.edu/SOURCES/.IGOSS/.nmc/.Reyn_SmithOIv2/.monthly/.sst/

³Байесов доверительный интервал вычислялся с помощью генерации ансамбля параметров НДМ в соответствии с окрестностью максимума ПВ (2.7), аппроксимированной методом Лапласа

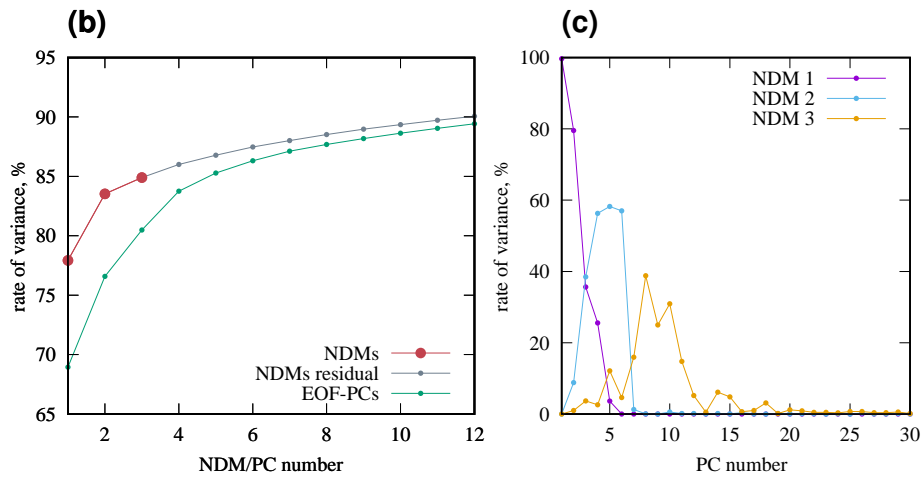
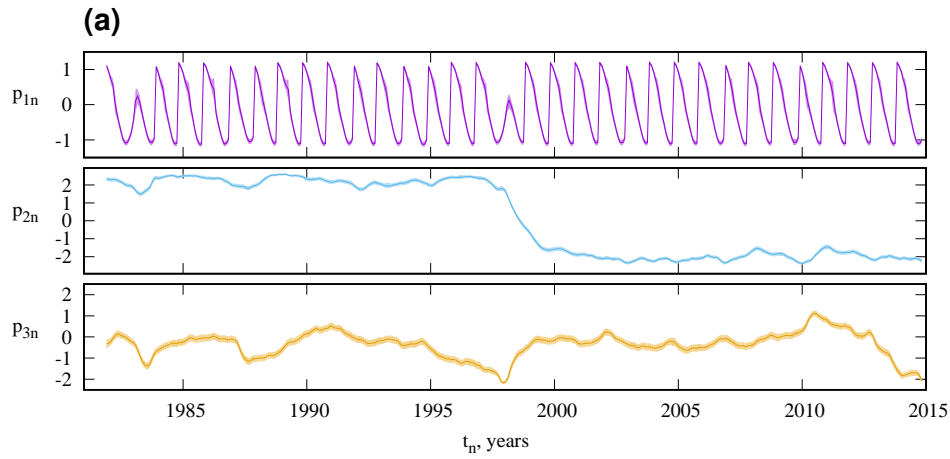


Рис. 2.5: (a) Временные ряды p_{i1}, \dots, p_{iN} ($i = 1, 2, 3$), соответствующие найденным НДМ, расположенные в порядке убывания захваченной вариации. (b) Доля кумулятивной вариации, захватываемой модами. Большие точки соответствуют НДМ, серые точки – ЭОФ-ГК, полученным в результате ЭОФ-разложения остатков, зеленые точки – ЭОФ-ГК, полученным в результате ЭОФ-разложения данных ТПО. (c) Вклад НДМ в различные ЭОФ-ГК (показана доля захватываемой вариации). Структурные параметры K и m первой НДМ – $K = 5, m = 3$, второй НДМ – $K = 4, m = 6$, третьей НДМ – $K = 8, m = 5$.

Сравнивая вариацию ТПО, захватываемую тремя НДМ, с вариацией, захватываемой обычными ЭОФ-ГК, полученными посредством проекции полных данных ТПО на главные ЭОФ, можно видеть, что три НДМ описывают примерно 85% полной вариации ТПО, что примерно соответствует вариации пяти ЭОФ-ГК; см. рис. 2.5b, демонстрирующий кумулятивную вариацию, захватываемую соответствующим количеством НДМ. Рис. 2.5c показывает вклад каждой НДМ в ЭОФ-ГК. Видно, что каждой НДМ соответствует определенная фракция ЭОФ-ГК. При этом суммарно

вовлеченными в нелинейное разложение оказываются примерно 20 ЭОФ-ГК: определяемое ими подпространство содержит всю нелинейность, которая может быть извлечена из данных ТПО с помощью НДМ-разложения.

Поскольку для каждой i -й НДМ найдена соответствующая функция \mathbf{h}_i , отображающая скалярный временной ряд p_{i1}, \dots, p_{iN} в пространство наблюдаемых переменных ТПО, мы имеем возможность визуализировать временные ряды НДМ в географических координатах. Такая визуализация полученных НДМ 1-3 представлена на рис. 2.6а-с для января 1992 года, а также в виде анимационного файла, демонстрирующего весь наблюдаемый временной интервал и доступного по интернет-ссылке⁴ (см. также статью [A2]). Кроме того, на рис. 2.6d и в вышеуказанном анимационном файле в том же формате изображен остаток от НДМ-разложения.

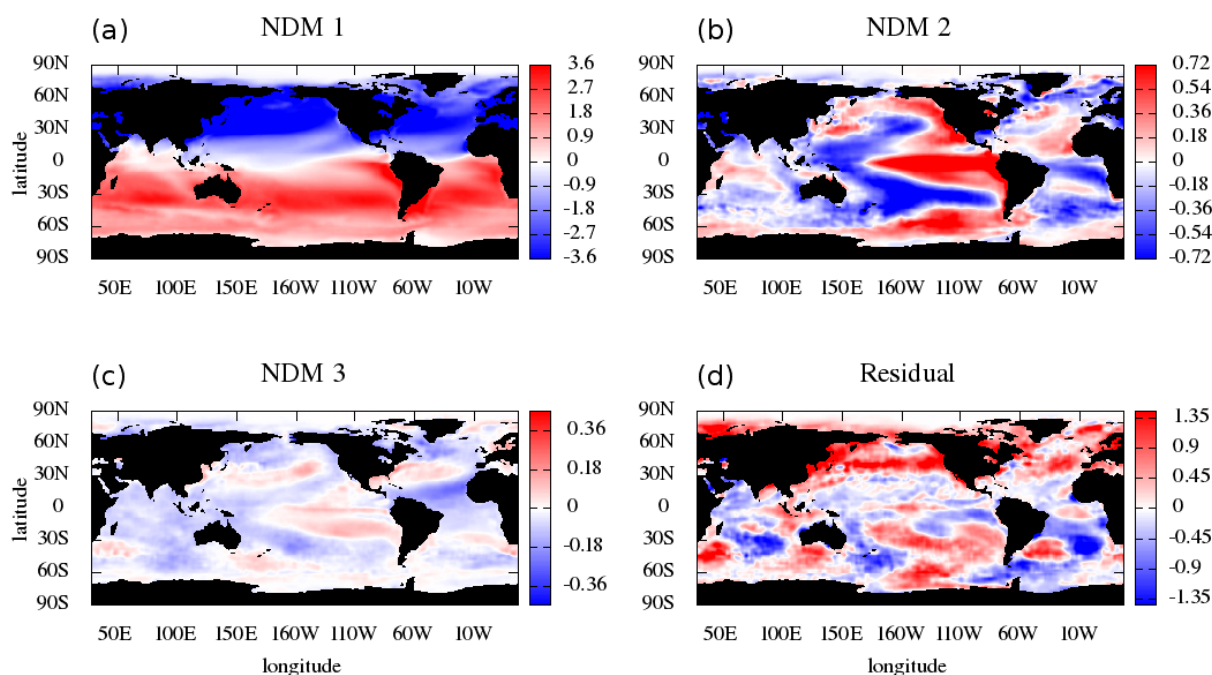


Рис. 2.6: Мгновенный “снимок” на географической сетке, соответствующий январю 1992 года. Цветовая шкала соответствует вкладу в градусах по Цельсию соответствующей НДМ (с учетом знака) в отклонение ТПО от своего среднего значения в каждой географической точке. НДМ 1-3 показаны на рис. 2.6(а-с). Остаток от НДМ-разложения показан на рис. 2.6d.

Первая мода осциллирует с периодом 1 год, отражая доминирующий вклад в нее годового хода, имеющегося в данных ТПО. Вторая и третья моды имеют более длинные временные масштабы, по порядку величины соответствующие масштабам

⁴media.nature.com/original/nature-assets/srep/2015/151022/srep15510/extref/srep15510-s2.gif

изменчивости ЭНЮК и ТДК. Вторая мода очевидным образом связана с ЭНЮК: она возбуждается с соответствующим знаком во время эпизодов Эль-Ниньо и Ла-Нинья (см., например, рис. 2.6b, на котором изображен момент времени, соответствующий эпизоду Эль-Ниньо 1992 года). В остатке разложения (см. рис. 2.6d и анимационный файл) присутствуют масштабы меньше 1 года; в рамках процедуры НДМ-разложения его оптимальным описанием является описание в виде шума.

2.3.2 Анализ главных мод

Проведем более детальный анализ вклада найденных НДМ в изменчивость ТПО в различных географических регионах, а также соответствующих этим НДМ климатических явлений. С этой целью рассмотрим различные индексы, основанные на аномалиях ТПО (АТПО, см. конец раздела 1.4.1). В Таблице 2.1 приведены индексы, соответствующие различным регионам в Северном полушарии и связанные с глобальным климатом: Nino 3.4 [120] – основанный на ТПО индекс ЭНЮК, PDO (Pacific Decadal Oscillation) – индекс ТДК, определяемый как проекция данных на главную ЭОФ АТПО в северной части Тихого океана [25], NTA (North Tropical Atlantic) – средняя температура в северной тропической Атлантики [94], IOD (Indian Ocean Dipole) – индийский океанический диполь [108]. Также в Таблицу 2.1 добавлен индекс NA (North Atlantic), определяемый средней ТПО в северном Атлантическом океане.

Таблица 2.1: Индексы, вычисленные по полю ТПО, реконструированному с помощью НДМ.

| Индекс | Регион | Описание | Корреляция с АТПО-индексом |
|----------|--|--|----------------------------|
| Nino 3.4 | [5°S – 5°N; 190°E – 120°W] | Среднее по региону | 0.88 |
| PDO | [20°N – 60°N; 140°E – 120°W] | Проекция на главную ЭОФ АТПО в регионе | 0.75 |
| NTA | 1. [6°N – 18°N; 60°W – 20°W] 2. [6°N – 10°N; 20°W – 10°W] | Среднее по объединению регионов 1 и 2 | 0.74 |
| NA | [45°N – 70°N; 60°W – 0°W] | Среднее по региону | 0.73 |
| IOD | 1. [10°S – 10°N; 50°E – 70°E] 2. [0°N – 10°N; 90°E – 110°E] | Разница между средними по регионам 1 и 2 | 0.38 |

Каждый из этих индексов был вычислен для каждой из трех НДМ, согласно приведенному в Таблице 2.1 определению. Наиболее интенсивная (первая) НДМ дает вклад в виде почти периодического, слабо модулированного сигнала в каждый из индексов. Он соответствует годовому ходу – аддитивной (линейной) части отклика климата на внешнее воздействие с периодом 1 год, обусловленное зависимостью инсоляции в каждой географической точке от времени года. Существенно, что процедура НДМ

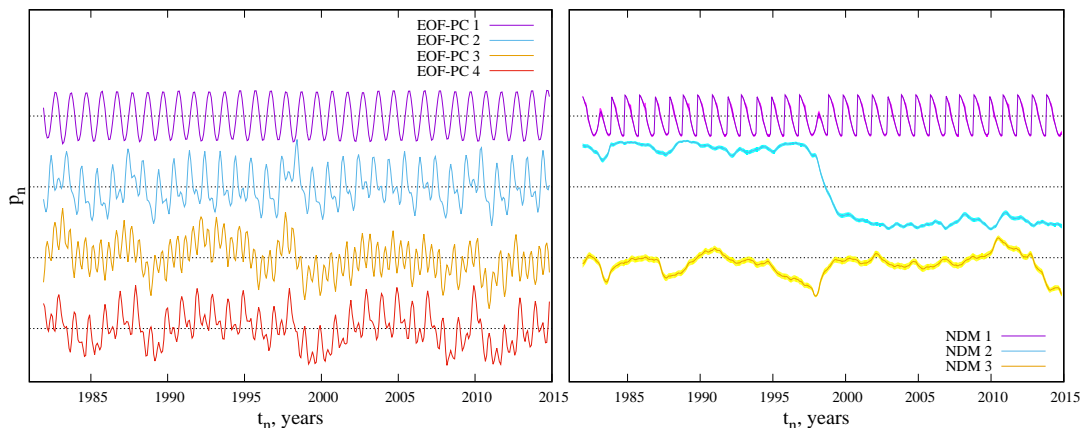


Рис. 2.7: Сравнение временных рядов ЭОФ-ГК (в условных единицах, слева) и НДМ (в условных единицах, справа). Временные масштабы гораздо лучше разделяются с помощью НДМ.

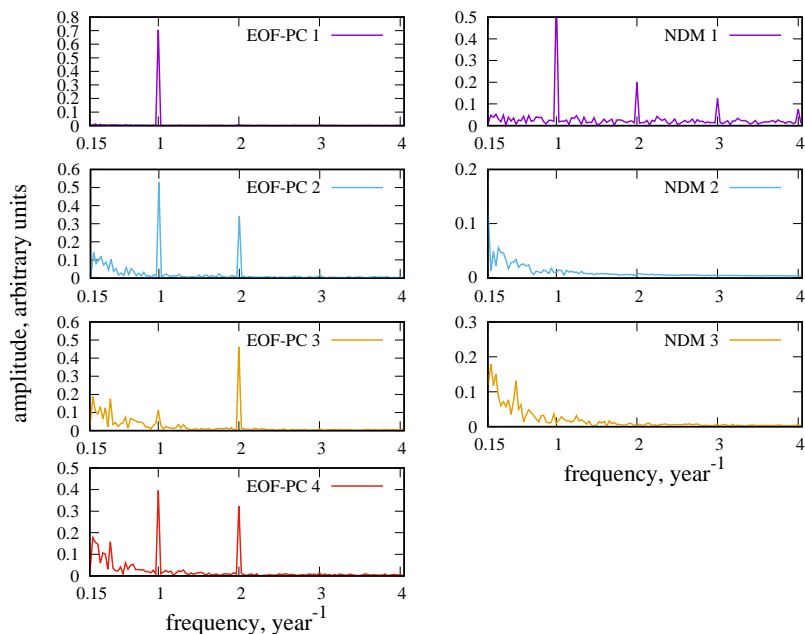


Рис. 2.8: Сравнение Фурье-спектров ЭОФ-ГК (слева) и нелинейных ГК, соответствующих НДМ (справа).

позволила полностью захватить годовой ход одной первой НДМ: гармоники сигнала с периодом 1 год отсутствуют во второй и третьей НДМ, в отличие от главных ЭОФ-ГК (см. рис. 2.7 с временными рядами ГК и рис. 2.8 с их Фурье-спектрами). Заметим, что это произошло естественным образом, т.е. никаких специальных периодических условий в процедуру разложения не закладывалось. В качестве подтверждения вышесказанного на рис. 2.9 показана разность между средним вкладом первой НДМ в ТПО в зимние месяцы и средним вкладом первой НДМ в ТПО в летние месяцы – получающийся паттерн, очевидно, представляет собой пространственный паттерн годового хода.

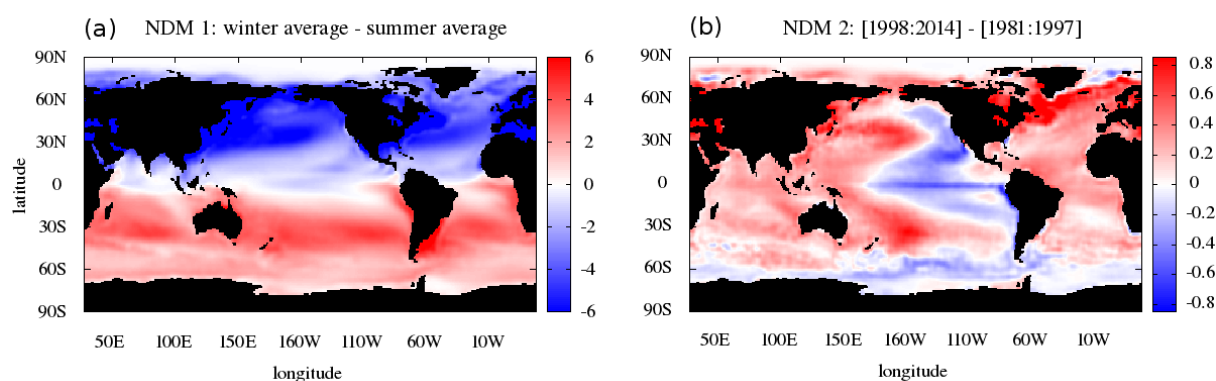


Рис. 2.9: (а) Паттерн годового хода, присутствующего в первой НДМ: разность между вкладами первой НДМ в ТПО, усредненными по зимним и летним месяцам на наблюдаемом временном интервале. (б) Паттерн отрицательной фазы ТДК, присутствующего во второй НДМ: разность между вкладами второй НДМ в ТПО, усредненными по интервалам [1998-2014] и [1981-1997].

В левой части рис. 2.10 представлены индексы, реконструированные на основе комбинаций НДМ, дающих наибольшие корреляции с соответствующими индексами, основанными на наблюдаемых данных АТПО (черный цвет). Соответствующие коэффициенты корреляции представлены в правом столбце Таблицы 2.1. Фиолетовые столбцы в правой части рис. 2.10 показывают корреляции индексов, реконструированных по разным НДМ (годовой периодический сигнал был удален из первой НДМ перед расчетом корреляции), с наблюдаемыми индексами. Для сравнения, на этих же рисунках показаны синими прозрачными столбцами корреляции различных ЭОФ-ГК (после удаления из них годового сигнала) с наблюдаемыми индексами. Из рис. 2.10 видно, что вторая НДМ является доминирующей в изменчивости ЭНЮК: она обеспечивает очень высокую корреляцию (0.88) с исходным индексом Niño 3.4.

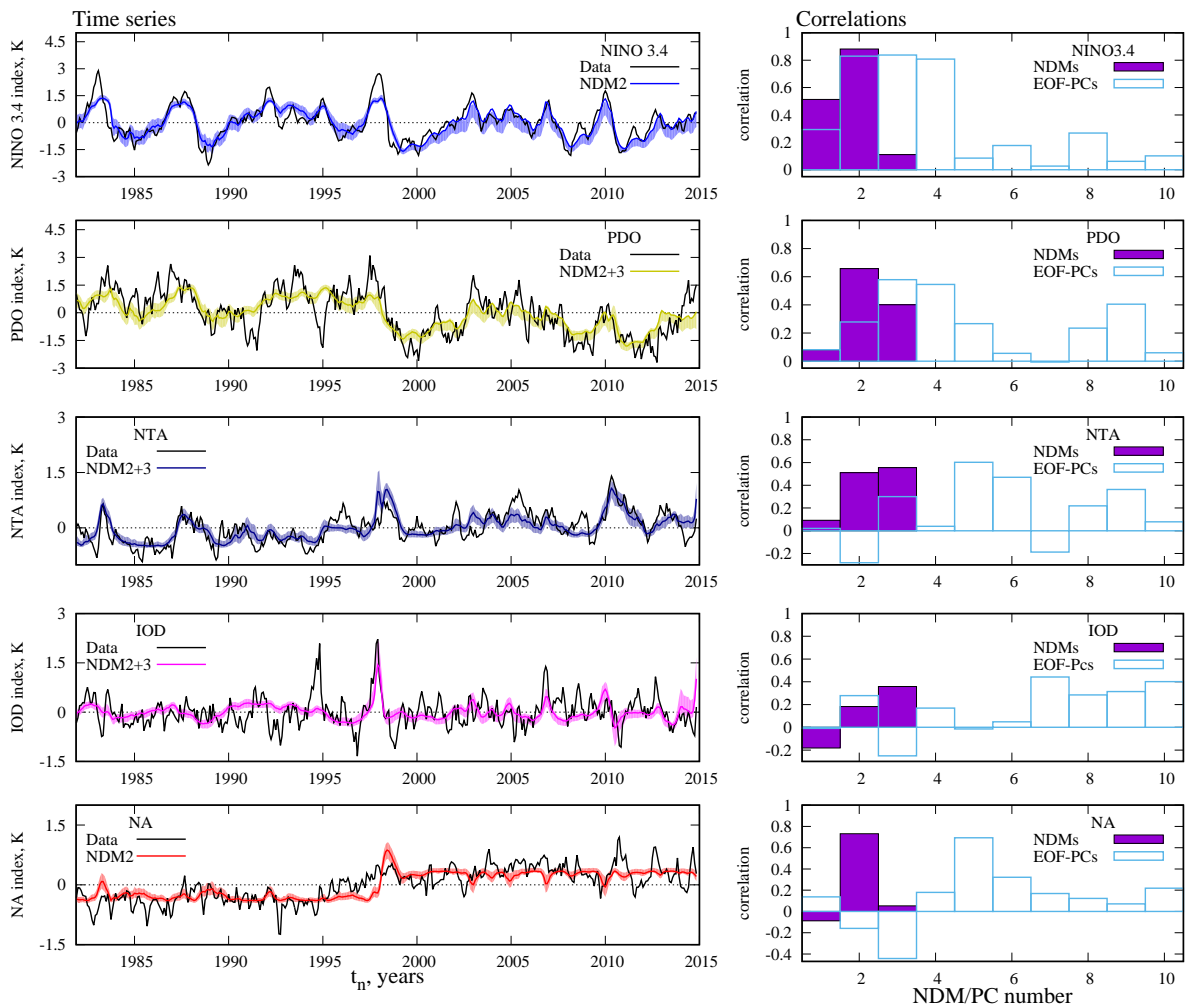


Рис. 2.10: Реконструкция климатических индексов, перечисленных в Таблице 2.1. Левая половина: сравнение индексов, вычисленных по полю АТПО (черный цвет), и индексов, реконструированных по различным комбинациям второй и третьей НДМ. Байесов доверительный интервал реконструкции показан тенью вокруг соответствующего индекса. Правая половина: фиолетовые столбцы показывают корреляции соответствующего индекса (см. подписи к вертикальной оси слева) с индексом, реконструированным по различным НДМ (номер НДМ отложен по оси абсцисс). Для сравнения прозрачными столбцами показаны корреляции этих же индексов с ЭОФ-ГК.

Фактически, не полностью описанными ею остались лишь два достаточно больших по амплитуде события Эль-Ниньо 1982-83 и 1997-98 годов: эти события оказались частично учтенными в первой НДМ, корреляция которой с индексом Nino 3.4 составила 0.5. Аналогичная ситуация имеет место в северо-Атлантическом регионе (индекс NA): вторая НДМ имеет гораздо больший вклад в этот регион, чем другие НДМ.

Изменчивость двух других рассмотренных индексов – PDO и NTA – наряду со второй НДМ существенно определяется третьей НДМ. Две эти моды вместе объясняют существенную часть изменчивости АТПО в обоих регионах. При этом динамика в Индийском океане отражена НДМ-разложением не так хорошо: реконструкция с помощью полученных НДМ дает лишь корреляцию 0.38 с наблюдаемым индексом IOB.

Заметим, что вклады ЭОФ-ГК в рассматриваемые регионы распределены по многим ГК, как это видно из правой половины рис. 2.10. Например, есть как минимум три ГК, одинаково значимых для динамики ЭНЮК, а также множество ГК, дающих вклад в индексы PDO и Атлантические регионы. Очевидно, это является результатом линейности ЭОФ-разложения, приводящей к смешиванию различных мод в полученных ГК в случае, когда нелинейность существенна. Рис. 2.7 наглядно демонстрирует преимущество разработанного нелинейного подхода по сравнению с традиционным ЭОФ-разложением в задаче разделения временных масштабов в реконструируемых ГК.

2.3.3 Климатический переход 1997-1998 годов

Наиболее интересным событием на анализируемом интервале является событие Эль-Ниньо 1997-1998 годов – одно из наиболее сильных за время наблюдения (см. поведение индекса Niño 3.4 на рис. 2.10). Существует интересная гипотеза [122] о том, что это событие сыграло решающую роль в переключении климата с положительной фазы ТДК в отрицательную и последующем изменении сценария глобального потепления: вскоре после этого события рост температуры атмосферы замедлился (см. также Введение), а перед этим событием происходило потепление атмосферы с середины 1970-х годов во время положительной фазы ТДК. Так, было показано [122], что во время отрицательной фазы ТДК значительная часть тепла уходит в глубину океана, предотвращая нагрев атмосферы. В то же время, ледники продолжают таять, а уровень океана – расти [121], указывая на то, что само глобальное потепление продолжается. Таким образом, изменился лишь режим потепления, что, по-видимому, есть результат собственной динамики климатической системы.

Данный климатический переход проявился во второй НДМ в виде резкого изменения («скачка») в ее временном ряде p_{21}, \dots, p_{2N} (рис. 2.5а), имеющего масштаб, сравнимый с одним годом, и происходящего как раз во время Эль-Ниньо 1997-1998 годов. В индексе PDO и двух Атлантических индексах, реконструированных по данной НДМ, это отражается в виде резкого изменения среднего значения индекса (см. рис. 2.10). Проверить, что данное событие ассоциируется со сменой фазы ТДК, можно,

убедившись, что вторая НДМ, в которой наблюдается переход, воспроизводит паттерн ТДК. Для этого были вычислены пространственно распределенные средние вклады второй НДМ в ТПО за интервалы времени до и после перехода, после чего построена разность двух полученных паттернов (рис. 2.9b). В результате, на рис. 2.9b отчетливо виден характерный паттерн отрицательной фазы ТДК в тихоокеанском регионе, в полном соответствии с рис. 9 из работы [122], на котором показана аналогичная разность для полного поля температуры атмосферы. Таким образом, можно заключить, что пространственно-временная структура второй НДМ, описываемой одним скалярным сигналом, подтверждает гипотезу о причине климатического перехода [122].

2.3.4 Возможность детектирования дальних связей

Нахождение главных одномерных многообразий (кривых) в пространстве наблюдаемых переменных позволяет исследовать дальние связи – связи между пространственно разнесенными регионами планеты. В самом деле, вторая НДМ описывает примерно 7% полной вариации изменчивости ТПО. Если исключить из рассмотрения первую

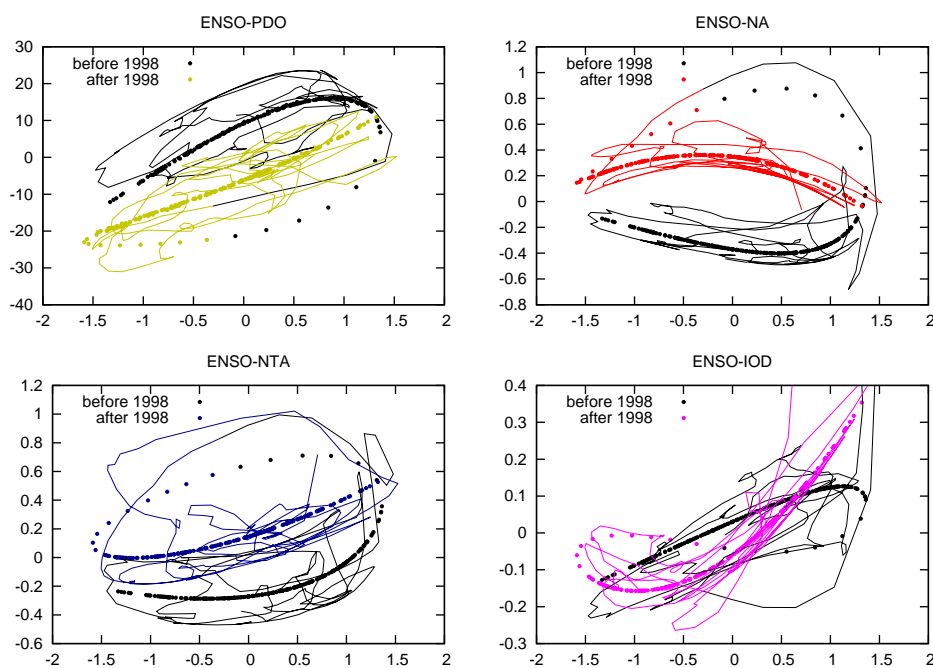


Рис. 2.11: Дальние связи между ЭНЮК (ENSO) и динамикой в других регионах, детектируемые с помощью НДМ. Зависимости, основанные на второй НДМ, показаны точками; основанные на сумме второй и третьей НДМ – тонкими линиями. Черный цвет соответствует интервалу времени до климатического перехода 1997-1998 годов, другие цвета – интервалу после перехода. Индексы, связь между которыми изображена, указаны над каждым из рисунков.

НДМ, имеющую вклад около 78% и содержащую, главным образом, годовой ход, то получится, что вторая НДМ описывает около 27% АТПО (в данном случае АТПО трактуются как результат вычитания первой НДМ из данных ТПО). При этом данная мода ставит в соответствие динамику во всех узлах географической сетки точкам на одномерной кривой, поэтому различные точки земного шара и, следовательно, регионы и климатические индексы, оказываются связанными посредством этой моды.

На рис. 2.11 точками показаны кривые, соответствующие второй НДМ: они определяют доминирующее совместное поведение региона Nino 3.4 и других рассмотренных регионов. Каждая из них состоит из двух ветвей: одна ветвь соответствует интервалу времени до смены фазы ТДК, другая – после этого перехода; видно, что переход существенно меняет данные связи. Добавление третьей НДМ приводит к небольшому «размытию» этих связей, однако их основная форма сохраняется: тонкие линии на рис. 2.11 соответствуют связям, полученным в результате реконструкции с помощью суммы второй и третьей НДМ, которые вместе описывают около 30% вариации АТПО.

2.4 Главные моды доиндустриального климата модели INMCM4.0

В данном разделе НДМ-разложение будет применено к ежемесячным данным ТПО, сгенерированным климатической моделью INMCM4.0 (см. раздел 1.4.1) в ходе «контрольного эксперимента» CMIP5. В отличие от современного периода, рассмотренного в разделах 1.4 и 2.3, контрольный эксперимент предусматривает неизменные внешние воздействия на климатическую систему (в частности, уровень эмиссий CO₂), соответствующие «доиндустриальному» климату конца XIX века, который модель стремится воспроизвести. В этом случае естественно ожидать, что НДМ-разложение приведет к иному результату.

2.4.1 Параметры временного ряда и полученные НДМ

Отсутствие во временном ряде контрольного эксперимента трендов, связанных с медленными изменениями внешних воздействий, позволяет взять для исследования собственных мод интервал значительно большей протяженности по времени. В частности, из полного 500-летнего временного ряда контрольного эксперимента, содержащего значения ТПО в 85640 узлах пространственной сетки, был отброшен начальный 250-летний переходный участок и взят для анализа оставшийся 250-летний интервал с 2100 по 2350 год модельного времени (если датировать начало полного временного ряда 1850 годом), соответствующий практически установившемуся собственному

модельному климату. Нормировка ряда ТПО производилась в соответствии с (1.35).
 Применение НДМ-разложения к данному временному ряду выявило пять отлич-

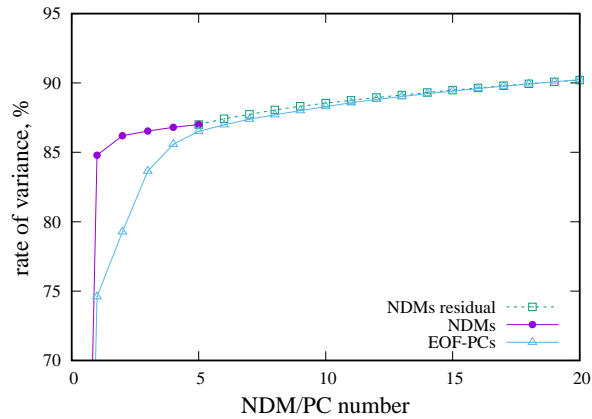


Рис. 2.12: Кумулятивная доля вариации, описываемая модами, в зависимости от количества мод. Фиолетовый цвет соответствует НДМ-разложению, синий – ЭОФ-ГК. Зеленым показан вклад НДМ с учетом разложения остатка на обычные ЭОФ.

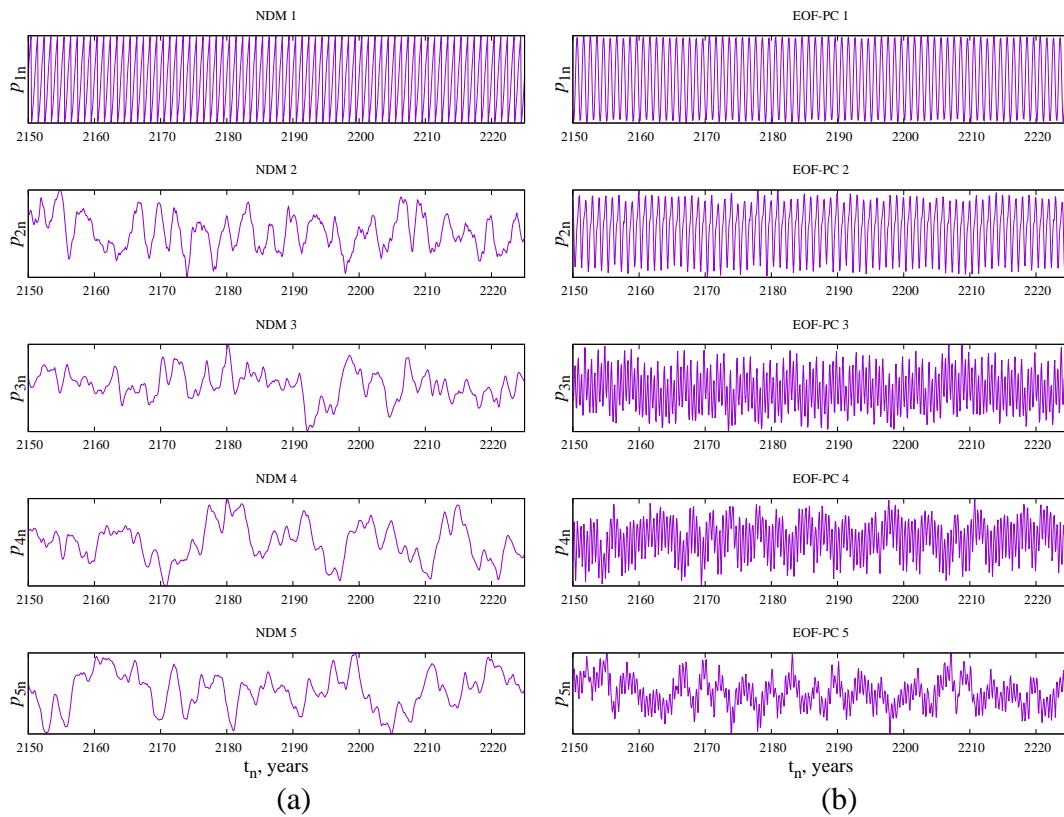


Рис. 2.13: Фрагменты пяти временных рядов НДМ (а) в сравнении с пятью главными ЭОФ-ГК (b).

ных от нуля мод. На рис. 2.12 показана кумулятивная доля вариации, описываемая полученными НДМ, продолженная с учетом дальнейшего ЭОФ-разложения остатка. Аналогичная зависимость приведена для ЭОФ-разложения исходных данных. Очевидно, что в данном случае кривая кумулятивной вариации для НДМ идет выше, чем для ЭОФ, главным образом, благодаря значительному вкладу первой НДМ в полную вариацию данных. Ниже будет показано, что, как и в случае с НДМ в разделе 2.3, это является результатом эффективного выделения годового хода первой НДМ, в то время как при ЭОФ-разложении годового хода распределен по многим ЭОФ-ГК. При этом суммарная доля вариации, захваченной пятью НДМ, очень близка к суммарной вариации пяти ЭОФ-ГК: на самом деле НДМ содержат информацию примерно из 13 ЭОФ-ГК, с существенным преобладанием пяти главных ЭОФ-ГК.

Тем не менее, важно помнить, что преимущество НДМ относительно ЭОФ-ГК, делающее их физически более адекватными, состоит еще в том, что НДМ обеспечивают лучшее выделение характерных временных масштабов системы. Это наглядно отражено во временных рядах ГК для НДМ и ЭОФ на рис. 2.13. В отличие от ЭОФ-разложения, где различные временные масштабы перемешаны в ЭОФ-ГК, полученные НДМ можно на основе рис. 2.13а разделить на 3 группы в соответствии с их временными масштабами:

i. Временной ряд первой НДМ p_{1n} представляет собой сигнал с периодом 1 год, отражая доминирующий вклад годового хода в эту моду. На рис. 2.14а показано изображение этой НДМ на географической сетке, т.е. компонента $\mathbf{h}_1(p_{1n})$ уравнения (2.1), в один из моментов времени (см. также мультимедийный файл, содержащий 20-летние фрагменты всех нижеследующих групп НДМ, изображенных на рис. 2.14, и доступный по интернет-ссылке⁵ из работы [A3]). Пространственная структура данной НДМ соответствует сезонным вариациям ТПО и аналогична рис. 2.6: колебания ТПО в северном полушарии находятся в противофазе с колебаниями ТПО в южном полушарии. При этом из рис. 2.13 видно, что гармоники годового сигнала отсутствуют в НДМ 2-5, но присутствуют во всех пяти ЭОФ-ГК.

ii. В ГК второй и третьей НДМ доминирует временной масштаб 2-5 лет. Можно сказать, что эти две НДМ образуют двумерное многообразие в пространстве исходных переменных. Их комбинация, очевидно, описывает динамику ЭНЮК: на географической сетке суперпозиция этих НДМ $\mathbf{h}_2(p_{2n}) + \mathbf{h}_3(p_{3n})$ содержит хорошо известные паттерны ЭНЮК, а также дальние связи ЭНЮК с Атлантическим, северным Тихим и Индийским океанами (см. рис. 2.14b, соответствующий положительной фазе ЭНЮК). При этом можно заметить, что положительные аномалии ТПО рис. 2.14b заходят

⁵ aip.scitation.org/doi/video_original/10.1063/1.4968852/1.4968852.mmm.original.v1.avi

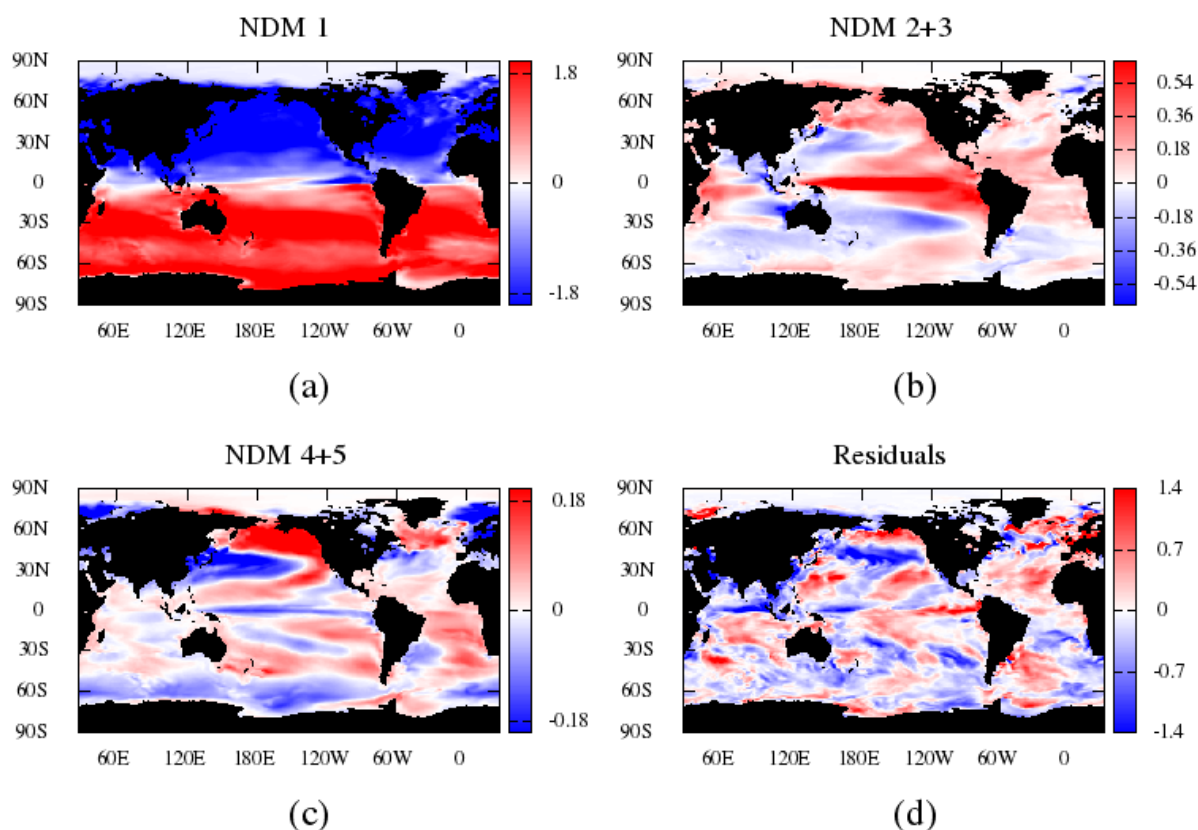


Рис. 2.14: Изображения комбинаций НДМ на географической сетке в момент времени, соответствующий положительной фазе ЭНЮК в зимнее время. Цветом показаны отклонения ТПО от своего среднего значения в каждой точке. (a) Первая НДМ, отражающая, главным образом, годовой ход; (b) комбинация НДМ 2 и 3 – двумерная мода, ассоциирующаяся с изменчивостью ЭНЮК в тропиках; (c) комбинация НДМ 4 и 5 – декадная двумерная мода; (d) остаток от НДМ-разложения.

слишком далеко в западную часть Тихого океана: на самом деле, этот эффект не связан с разложением НДМ, но является свойством модели INMCM4.0, присутствующим также во многих других глобальных климатических моделях. Установленная связь мод с ЭНЮК может быть подкреплена сравнением спектров индекса Niño 3.4 (см. Таблицу 2.1), реконструированного по комбинации НДМ 2 и 3 и по полным данным аномалий ТПО. Как видно из рис. 2.15а, комбинация второй и третьей НДМ описывает пик спектра индекса Niño 3.4 с масштабом 2-5 лет. Более того, данная комбинация мод также хорошо описывает декадный пик ЭНЮК с масштабом 15-20 лет, который может быть связан с проявлением ТДК в данном регионе. В дополнение к спектрам на рис. 2.16а показаны фрагменты индекса Niño 3.4 и его реконструкции с помощью НДМ 2 и 3, коэффициент корреляции между которыми (по полной длине

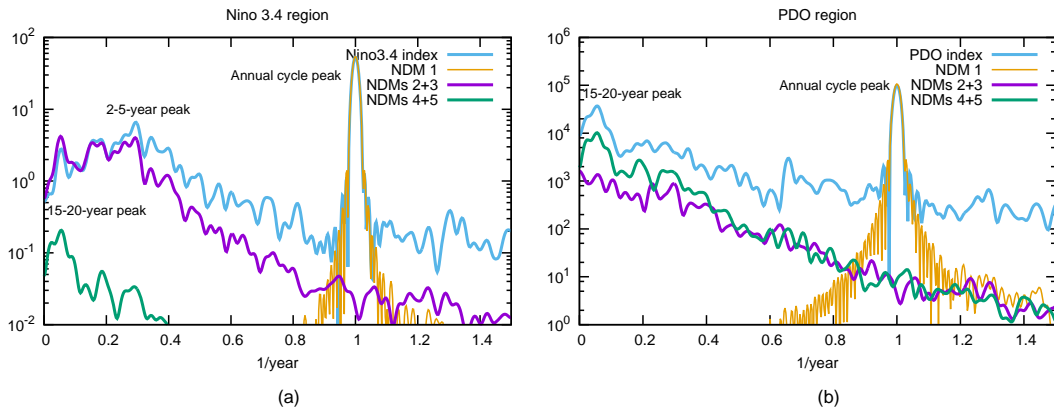


Рис. 2.15: Спектры индексов Nino3.4 (a) и PDO (b), реконструированных по полученным комбинациям НДМ (фиолетовый цвет – НДМ 2 и 3, зеленый цвет – НДМ 4 и 5) и по исходным данным (синий цвет). Межгодовой (2-5 лет) и декадный (15-20 лет) пики ЭНЮК в спектре индекса Nino 3.4 описываются НДМ 2 и 3, а декадный пик ТДК в спектре индекса PDO – НДМ 4 и 5. Пик, соответствующий годовому ходу, в обоих индексах полностью описывается первой НДМ (желтый цвет). Спектры вычислены с помощью применения коррелограмм Блэкмана-Тьюки [123] с окном 500 месяцев.

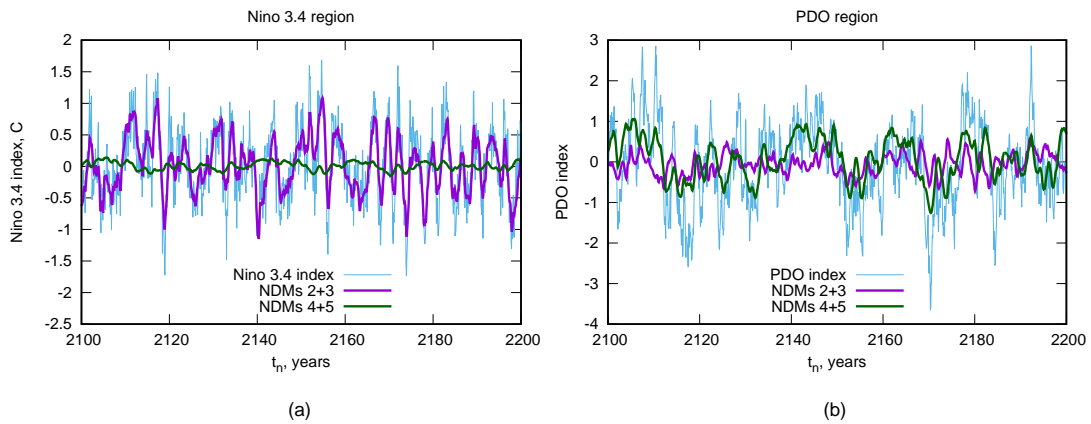


Рис. 2.16: Фрагменты временных рядов индексов Nino 3.4 (a) и PDO (b), реконструированных по полученным комбинациям НДМ, в сравнении с полными индексами, вычисленными по АТПО исходных данных. Цвета соответствуют цветам рис. 2.15. Корреляция комбинации НДМ 2 и 3 с индексами Nino 3.4 и PDO равна 0.82 и 0.38 соответственно. Корреляция комбинации НДМ 4 и 5 с индексами Nino 3.4 и PDO равна 0.29 и 0.69 соответственно.

ряда) составляет 0.82.

iii. Следующая группа – четвертая и пятая НДМ – образует другую двумерную моду, описывающую, главным образом, более длинные, декадные временные масштабы. Ее пространственная структура (см. рис. 2.14с) содержит, помимо прочего, характерный паттерн ТДК в северной части Тихого океана. Из рис. 2.15b видно, что эта двумерная мода воспроизводит пик с масштабом 15-20 лет в спектре индекса PDO (см. Таблицу 2.1), характеризующего динамику ТДК. Коэффициент корреляции реконструированного по НДМ 4 и 5 индекса PDO с полным индексом PDO (рис. 2.16b) также достаточно высок: 0.69.

Наконец, остаток от НДМ-разложения содержит значительно меньшие пространственные и временные масштабы (по отношению к НДМ 1-5). Пример соответствующего ему пространственного распределения изображен на рис. 2.14d, см. также вышеуказанный мультимедийный файл.

2.4.2 Структура главных многообразий

Исследуем структуру полученных НДМ в пространстве исходных переменных. Для этой цели на рис. 2.17 показаны проекции данных и мод на плоскость главных ЭОФ-ГК. Первая НДМ существенно нелинейна: она содержит периодические орбиты с периодом 1 год, почти идеально описывая годовой ход. Оптимизация данной НДМ в соответствии с ценовой функцией (2.27) привела к кривой, лежащей в пространстве семи ЭОФ-ГК и аппроксимируемой с помощью полинома степени 8.

Интересно, что для следующих НДМ 2-5 оптимальная степень полинома оказалась равной единице, т.е. они оказались равными ЛДМ (см. раздел 2.2.4) в данном случае. Вторая НДМ, хотя и лежит в пространстве шести ЭОФ-ГК, ориентирована практически полностью в направлении главной ЭОФ для остатка (т.е. данных после вычитания первой НДМ). Однако третья, четвертая и пятая НДМ существенно повернуты в пространстве ЭОФ-ГК, что отличает их от обычных ЭОФ. Раздел 2.2.4 подробно разъясняет причину этого отличия: оптимизация временных масштабов τ получаемых ГК приводит к лучшему выделению характерных масштабов, присутствующих в системе.

При поиске НДМ 6 наступает ситуация, описанная в конце раздела 2.2.4. Модель НДМ становится неспособной выделить одномерный полезный сигнал и поэтому трактует входные данные как шум, при этом НДМ получается равной нулю, означая окончание НДМ-разложения.

Таким образом, годовой ход представляет единственную по-настоящему нелинейную моду ТПО в доиндустриальном климате, генерируемой моделью INMCM4.0.

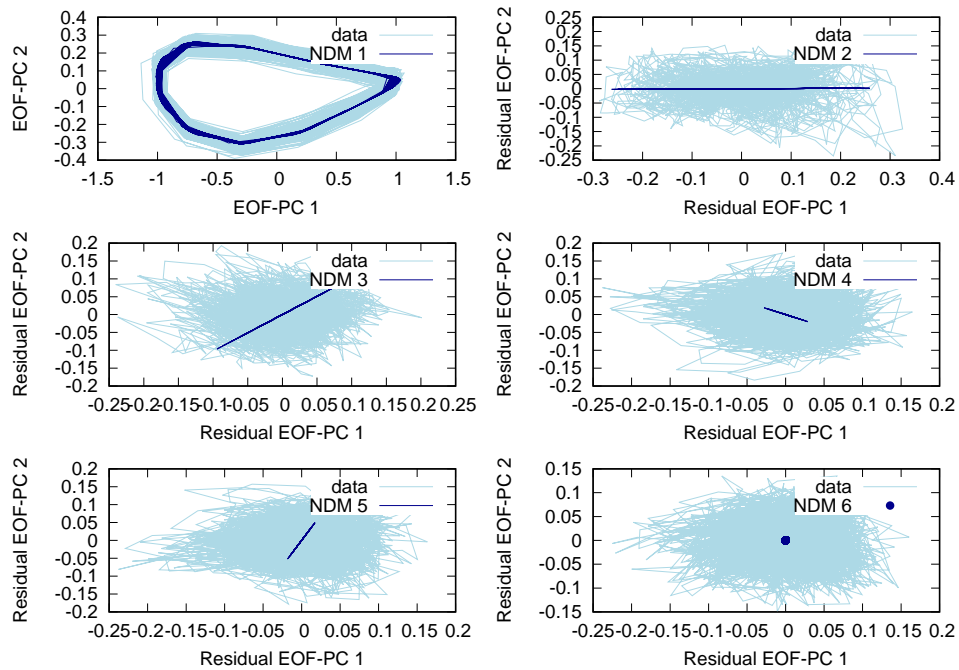


Рис. 2.17: Проекции данных (светлый цвет) и НДМ (темный цвет) на плоскость ЭОФ-ГК 1 и 2. Для каждой из НДМ 2-6 использовались ЭОФ-ГК для остатка после вычета предшествующих НДМ, т.е. компоненты y из выражения (2.4). НДМ 6 равна нулю, что означает конец НДМ-разложения. Оптимальное количество ЭОФ-ГК (ЭОФ-ГК остатков), в пространстве которых расположены НДМ 1-5, равны, соответственно: $K_1 = 7, K_2 = 6, K_3 = 5, K_4 = 5, K_5 = 5$.

Фактически, первая НДМ осуществляет вычитание годового хода и получение АТПО из данных ТПО, которое делается более простым способом в разделе 1.4.1. То, что следующие НДМ оказались линейными, позволяет предположить, что и некоторые другие линейные методы, учитывающие временную задержку и информацию о временных масштабах (см. Введение), примененные к АТПО, иногда могут давать адекватный результат.

Приведенный в данном разделе пример наглядно демонстрирует адаптивные свойства НДМ: они могут быть как существенно нелинейными, так и полностью линейными, в зависимости от свойств системы и имеющихся данных.

2.5 Заключение

В данной главе разработан новый метод эмпирической редукции данных, позволяющий представить наблюдаемые пространственно распределенные данные в виде суперпозиции небольшого числа нелинейных динамических мод. Каждая НДМ

представляет собой нелинейное отображение одномерного сигнала ГК в пространство наблюдаемых переменных и, таким образом, способна реконструировать *нелинейную* связь между переменными, наблюдаемыми в различных пространственно разнесенных точках, что отличает НДМ-разложение от ЭОФ-разложения и других линейных методов редукции. Кроме того, благодаря заложенному в НДМ априорному учету информации о наличии характерного временного масштаба у искомым временных рядов ГК, НДМ-разложение приводит к лучшему выделению временных масштабов, доминирующих в динамике системы на рассматриваемом интервале времени.

Важно подчеркнуть, что как временной масштаб ГК, так и степень нелинейности НДМ оптимизируются в рамках байесова критерия оптимальности. В результате алгоритм выбирает НДМ оптимальной сложности. В частности, в зависимости от длины наблюдаемого интервала, соотношения временных масштабов наблюдаемой динамики системы, сложности соответствующих пространственно-временных паттернов, получаемые НДМ могут оказаться как существенно нелинейными, так и абсолютно линейными, и даже равными нулю (отсутствие распознаваемых паттернов), что является критерием останковки НДМ-разложения. Адекватность работы критерия оптимальности продемонстрирована на наглядном модельном примере со скрытой переменной, сгенерированной системой Лоренца-63. Кроме того, на простом примере показано, что даже в линейном варианте НДМ осуществляют выделение доминирующих временных масштабов и соответствующего «полезного» сигнала лучше, чем ЭОФ-разложение.

Разработанный метод применен к пространственно распределенным данным ТПО, основанным на спутниковых измерениях с 1981 по 2014 год и характеризующим современный климат. В результате выделено три статистически значимых нелинейных моды климатической изменчивости. В первой моде доминирует годовая изменчивость климатической системы. Вторая мода описывает ЭНЮК; кроме того, она, совместно с третьей модой, реконструирует значительную часть тихоокеанской и атлантической динамики. Продемонстрирована связь найденных мод с декадной изменчивостью климата: подтверждено наличие климатического сдвига в конце XX века, приведшего к холодной (отрицательной) фазе ТДК, во многом определяющей наблюдаемое вплоть до 2015 года замедление потепления атмосферы. Показана возможность детектирования и описания дальних связей с помощью НДМ-разложения (на примере связи ЭНЮК с динамикой в других регионах).

Также метод был применен к анализу доиндустриального климата конца XIX века на основе данных ТПО, сгенерированных глобальной климатической моделью INMCM4.0 в ходе соответствующего эксперимента CMIP5. В данном случае в результате НДМ-разложения получилось пять мод с различной структурой. Первая мода,

как и в случае с НДМ современного климата, является существенно нелинейной и описывает годовую изменчивость климата. Однако остальные моды, в отличие от мод современного климата, оказались линейными. Они могут быть разделены на две группы, первая из которых отражает, главным образом, динамику в тропическом Тихом океане, связанную с ЭНЮК, вторая – декадную изменчивость, ассоциированную с ТДК. Показано, что эти моды, будучи линейными, все же отличаются от обычного ЭОФ-разложения, потому что их временные масштабы подстраиваются под временные масштабы, доминирующие в наблюдаемой динамике.

Отличие степени нелинейности ответственных за межгодовые и декадные масштабы мод современного (нелинейные НДМ) и доиндустриального (линейные НДМ) климата позволяет сделать предположение об отличии самих климатов и присутствующих в них дальних связей, обусловленном изменением внешних условий (таких, как глобальная концентрация парниковых газов в атмосфере), произошедшим за время от конца XIX до начала XXI века. Заметим, что другими возможными причинами этого отличия могут быть как различные источники данных (в первом случае это данные измерений реальной системы, во втором – собственный климат модели INMCM4.0), так и существенно разная длина анализируемого интервала времени (в частности, в первом случае за время наблюдения произошла лишь одна смена фазы ТДК, а во втором – несколько). Анализ, включающий различные «промежуточные» временные интервалы, будет являться частью следующей главы.

Глава 3

Многомерные нелинейные динамические моды

3.1 Введение

В данной главе предлагается дальнейшее обобщение разработанного в предыдущей главе метода нелинейных динамических мод, позволяющее вместо набора нескольких НДМ, каждая из которых соответствует проекции траектории системы на одномерное нелинейное многообразие, лежащее в пространстве наблюдаемых переменных, искать совместную многомерную НДМ (МНДМ), т.е. проекцию на многомерное главное многообразие, которое, вообще говоря, может принципиально не описываться суперпозицией одномерных НДМ. В разделе 3.2 формулируется суть предлагаемого обобщения и соответствующий алгоритм поиска МНДМ, а также рассматривается модельный пример, демонстрирующий лучшую физическую адекватность МНДМ, по сравнению с НДМ.

В разделе 3.3 рассматривается главная двумерная МНДМ для данных ТПО на интервале с 1981 по 2014 год. Проводится сравнение данной МНДМ с набором НДМ, полученных по тем же данным в разделе 2.3.

Последний раздел данной главы (раздел 3.4) посвящен применению метода МНДМ для анализа временных рядов ТПО на других временных интервалах, связанных с динамикой ТДК, начиная с конца XIX века и заканчивая настоящим временем. В частности, с помощью МНДМ исследуются паттерны дальних связей ЭНЮК, соответствующие различным циклам ТДК за последние 145 лет, находятся моменты смены фазы ТДК и реконструируются соответствующие глобальные паттерны ТДК. На основе адаптивных свойств МНДМ показывается, что дальние связи ЭНЮК с различными регионами становятся существенно нелинейными, начиная со второй

половины XX века.

Материалы данной главы представлены в статье [A5] и публикациях [A9, A15, A16, A50–A58].

3.2 Многомерное обобщение НДМ

3.2.1 Модификация модели НДМ

Несмотря на свою очевидную полезность, описанный в предыдущей главе метод НДМ-разложения имеет определенные ограничения. Во-первых, как было отмечено в разделе 2.2, форма разложения (2.1) не описывает произвольное нелинейное многообразие ввиду заложенной в нее аддитивности искомым одномерным НДМ. В частности, для используемой в предложенном алгоритме полиномиальной параметризации НДМ это означает, что в финальном разложении (2.1) неизбежно будут отсутствовать «перекрестные» нелинейные слагаемые, пропорциональные $p_{in}^{m_i} p_{kn}^{m_k}$. Понятно, что учет таких слагаемых, обеспечивающих более общую форму параметризации нелинейной связи ГК, описывающих состояние низкоразмерной подсистемы, с наблюдаемыми переменными, задающими состояние системы в целом, может более полно отражать свойства исследуемой системы и, тем самым, повысить точность ее реконструкции.

Во-вторых, даже для формы разложения (2.1) предложенная итерационная схема поиска НДМ (вводная часть раздела 2.2) не идеальна. Дело в том, при поиске очередной НДМ в процедуре никак не учитывается наличие следующих НДМ, потому что остаток на каждой итерации полагается гауссовым белым шумом. В результате наличие следующих НДМ, вообще говоря, может влиять на результат поиска текущей НДМ и, как следствие, на результат полного разложения, автоматически уменьшая описываемую суммарно всеми модами долю полной изменчивости системы. Это представляется особенно важным в тех случаях, когда механизмы, ответственные за динамику разных НДМ, не независимы, что проявляется в схожести разных НДМ по величине захватываемой вариации или по описываемым временным масштабам.

Указанные ограничения могут быть преодолены за счет следующего обобщения формы разложения (2.1):

$$\mathbf{x}_n = \mathbf{h}(\mathbf{p}_n) + \mathbf{r}_n. \quad (3.1)$$

Здесь, в отличие от разложения (2.1), различные нелинейные ГК ищутся совместно в виде векторного временного ряда скрытой переменной $\mathbf{p}_1, \dots, \mathbf{p}_N$, имеющего размерность d . Нелинейная векторная функция \mathbf{h} представляет собой нелинейное отображение сразу всех искомым нелинейных ГК в пространство исходных переменных. При этом, как и ранее, компоненты вектора \mathbf{p}_n представляют собой соответствующие

моменту t_n координаты на d -мерном нелинейном многообразии, определяемом теперь единым отображением \mathbf{h} , параметризуемым с помощью универсального аппроксиматора на основе многомерных полиномов Эрмита. Полное слагаемое $\mathbf{h}(\mathbf{p}_n)$ будем называть *многомерной нелинейной динамической модой* (МНДМ).

3.2.2 Алгоритм поиска МНДМ

Алгоритм поиска МНДМ аналогичен итерации поиска одной НДМ: проводится байесова оптимальная аппроксимация данных \mathbf{x}_n с помощью неизвестных отображения \mathbf{h} и временного ряда $\mathbf{p}_1, \dots, \mathbf{p}_N$ в предположении гауссовости и дельта-коррелированности по времени остатка \mathbf{r}_n . В частности, изложенные в разделах 2.2.1 и 2.2.2 шаги остаются прежними с точностью до следующих модификаций.

Во-первых, во всех выражениях скалярные величины p_n заменяются на вектора \mathbf{p}_n . Соответствующее априорное ограничение (2.12) на эти нелинейные ГК принимает вид:

$$\mathcal{P}_{pr}(\mathbf{p}_1, \dots, \mathbf{p}_N) = \prod_{i=1}^d \left[\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{p_{i1}^2}{2}\right) \cdot \prod_{n=2}^N \frac{1}{\sqrt{2\pi\sigma_{pi}^2}} \exp\left(-\frac{|p_{in} - b_i \cdot p_{i,n-1}|^2}{2\sigma_{pi}^2}\right) \right], \quad (3.2)$$

$$b_i = \exp(-1/\tau_i), \quad \sigma_{pi} = \sqrt{1 - b_i^2}.$$

Можно видеть, что в данную априорную ПВ для каждой нелинейной ГК заложен свой временной масштаб τ_i , который будет объектом оптимизации. Поэтому в выражениях (2.25)-(2.27) скалярная величина τ заменяется на вектор временных масштабов $\boldsymbol{\tau} = (\tau_1, \dots, \tau_d)$.

Во-вторых, величина m , характеризовавшая степень полинома, по аналогии с τ , заменяется на вектор $\mathbf{m} = (m_1, \dots, m_d)$, где m_i обозначает степень полинома \mathbf{h} по аргументу p_{in} – i -й компоненте вектора \mathbf{p}_n . Соответственно, число $(m+1)$ в априорной ПВ (2.22) меняется на зависящее от m количество коэффициентов многомерного полинома (учитывая все «перекрестные» коэффициенты).

Таким образом, сформулированный алгоритм поиска МНДМ является обобщением процедуры поиска одномерных НДМ. Стоит отметить, что данный алгоритм является еще более ресурсоемким, главным образом, ввиду резкого роста количества получающихся перекрестных коэффициентов многомерного полинома с ростом размерности нелинейного многообразия d и степеней полинома. Задача оптимизации вычислительного алгоритма, и, в том числе, вида аппроксимирующей функции, является одним из важных будущих шагов в данном направлении.

3.2.3 Модельный пример

Рассмотрим в качестве модельного примера трехмерный временной ряд, полученный в результате зашумленного нелинейного преобразования двумерной скрытой переменной \mathbf{q} .

По аналогии с разделом 2.2.3, для задания временного ряда скрытой переменной $\mathbf{q}_1, \dots, \mathbf{q}_N$ воспользуемся переменными \tilde{x} и \tilde{y} стохастической системы Лоренца-63 (система (1.31)):

$$\begin{aligned} q_{1n} &= 0.07(\tilde{y}_n - \tilde{x}_n) \\ q_{2n} &= 0.11(\tilde{y}_n + \tilde{x}_n) \end{aligned} \quad (3.3)$$

В данном случае будем использовать параметры системы (1.31) $\tilde{b} = \frac{8}{3}$, $\tilde{\sigma} = 10$, $\tilde{r} = 28$, $\sigma = 20$, шаг временного ряда $\Delta t = 0.125$ и длину временного ряда $N = 400$. Траектория двумерной переменной \mathbf{q} показана на рис. 3.1с.

Соответствующий трехмерный временной ряд $\mathbf{x}_1, \dots, \mathbf{x}_N$ сгенерируем с помощью следующего преобразования переменной \mathbf{q} :

$$\mathbf{x}_n = \phi(\mathbf{q}_n) + \sigma_\phi \cdot \zeta_n. \quad (3.4)$$

Здесь $\zeta_n \in \mathbb{R}^3$ – гауссов дельта-коррелированный случайный процесс, $\sigma_\phi = 0.05$ – его амплитуда, ϕ – следующая векторная функция:

$$\phi(\mathbf{q}_n) = \begin{pmatrix} \cos q_{1n} \cos q_{2n} \\ \cos q_{1n} \sin q_{2n} \\ \sin q_{1n} \end{pmatrix}. \quad (3.5)$$

Нетрудно видеть, что уравнение (3.5) задает двумерное многообразие в виде сферы в трехмерном пространстве, параметризованной параметром \mathbf{q}_n . Таким образом, модельный временной ряд $\mathbf{x}_1, \dots, \mathbf{x}_N$ представляет собой траекторию, лежащую на поверхности сферы, к которой добавлен шум (незашумленная и зашумленная траектории показаны на рис. 3.1а синим и черным цветами).

Применение одномерного НДМ-разложения приводит к одной НДМ, которая описывает 77% полной вариации временного ряда $\mathbf{x}_1, \dots, \mathbf{x}_N$. Соответствующее многообразие показано зеленым цветом на рис. 3.1b. Вторая НДМ в данном случае получается равной нулю, т.е. остаток от вычитания первой НДМ не содержит хорошо распознаваемой структуры, что неудивительно, поскольку траектория на сфере, вообще говоря, не может быть точно описана с помощью суперпозиции двух НДМ, а приближенной ее аппроксимации мешает шум.

В то же время, применение двумерной МНДМ позволяет хорошо описать нужный сегмент сферы двумерным многообразием (рис. 3.1b), а двумерный временной

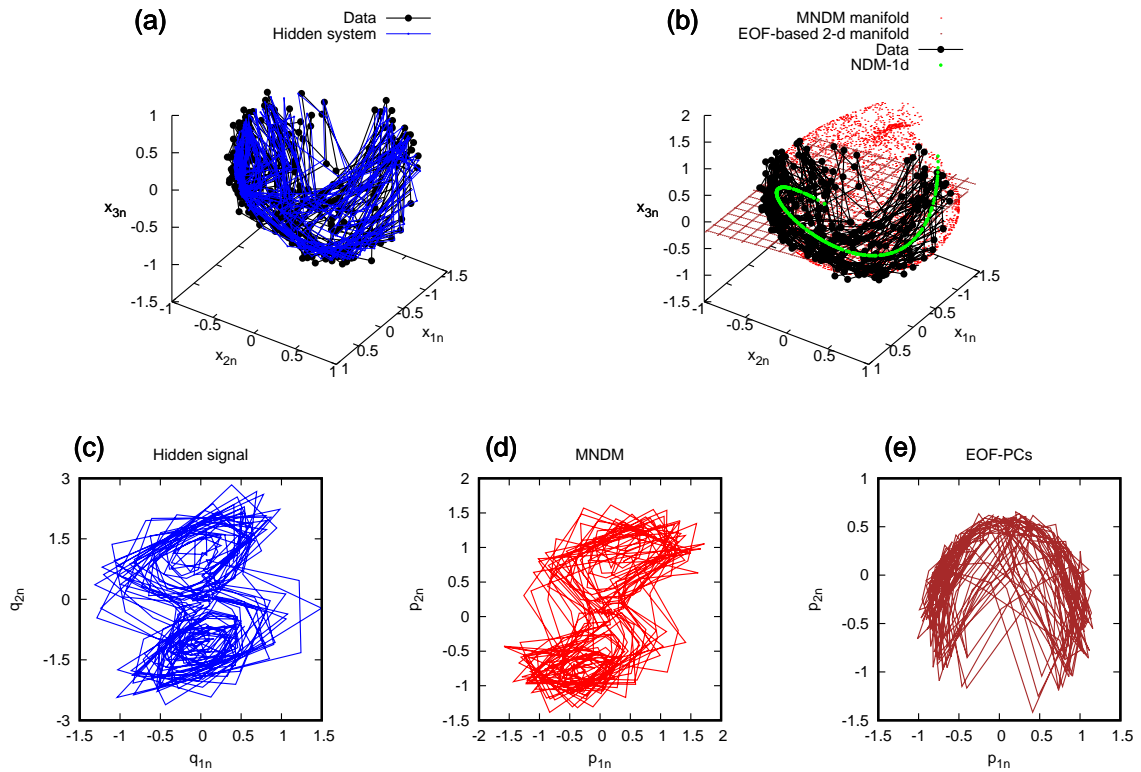


Рис. 3.1: Вверху в трехмерном пространстве «наблюдаемых» переменных показаны: (а) результат нелинейного преобразования (3.5) скрытой переменной \mathbf{q} (синий цвет) и полученный из него добавлением шума временной ряд данных (черный цвет); (б) временной ряд данных (черный цвет) и многообразия, соответствующие одномерной НДМ (зеленый цвет), двумерной МНДМ (красные точки) и двум ЭОФ (коричневая плоскость). Внизу на плоскости своих компонент изображены (с) двумерные временные ряды скрытой переменной \mathbf{q} , (d) нелинейные ГК двумерной МНДМ и (е) две главные ЭОФ-ГК.

ряд соответствующих нелинейных ГК (рис. 3.1d) топологически соответствует сигналу переменных \tilde{x} и \tilde{y} системы Лоренца-63. Таким образом, двумерная МНДМ описывает 99% вариации исходного временного ряда и справляется с задачей поиска «правильного» двумерного многообразия, в отличие от одномерных НДМ.

В дополнение, на рис. 3.1b приведена плоскость, соответствующая двум главным ЭОФ, которые, очевидно, не могут корректно описать сферу, и в пространстве соответствующих ЭОФ-ГК сигнал системы Лоренца-63 существенно искажен (рис. 3.1e).

3.3 Двумерные моды ТПО на интервале с 1981 по 2014 год

Напомним, что в разделе 2.3 был получен набор одномерных НДМ ТПО для интервала с 1981 по 2014 год, которые находились независимо друг от друга в рамках последовательной процедуры. При этом наиболее информативными оказались две главные моды, одна из которых была ассоциирована, главным образом, с годовым ходом, а вторая – с изменчивостью, связанной с ЭНЮК и ТДК. В контексте изложенных в разделе 3.2 ограничений последовательного метода разложения на одномерные НДМ обратим внимание на то, что изображенные на рис. 2.5а скрытые временные ряды первой и второй НДМ выглядят связанными друг с другом: имеются очевидные скачки как фазы, так и амплитуды в почти периодических временных рядах первой НДМ, которые происходят во время сильных событий Эль-Ниньо 1982-83 и 1997-98 годов, в результате чего корреляция вклада первой НДМ в индекс Niño3.4 с этим индексом оказывается около 0.5 (см. раздел 2.3.2). Этот факт можно интерпретировать как проявление связи ЭНЮК с годовым циклом (а именно, влияния годового цикла на динамику ЭНЮК) и, в то же время, указанных выше ограничений одномерного НДМ-разложения. Таким образом, можно ожидать, что эмпирическая реконструкция этих мод может быть более адекватной при построении общего многомерного нелинейного многообразия.

На рис. 3.2а изображены временные ряды скрытых переменных p_{1n} и p_{2n} двумерной МНДМ, построенной по последнему фрагменту рассматриваемого временного ряда. Аналогично результатам одномерного НДМ-разложения, годовой ход присутствует, главным образом, в первой компоненте моды p_{1n} , в то время как изменчивость ЭНЮК проявляется во второй компоненте p_{2n} , которая также отражает смену фазы ТДК в 1997-98 годах. Заметим, что полученная двумерная МНДМ описывает примерно 86% вариации данных (рис. 3.2b), что больше, чем суммарная вариация, описываемая тремя одномерными НДМ (около 85%): за счет одновременной реконструкции двух скрытых переменных вариация перераспределилась между временными рядами p_{1n} и p_{2n} более оптимально, что привело к лучшему захвату вариации наблюдаемых данных построенной двумерной МНДМ. Скорректировалось также и распределение вклада мод в ЭОФ-ГК: суммарно вовлеченными в двумерную МНДМ оказались первые 10 ЭОФ-ГК (рис. 3.2с). Кроме того, можно видеть, что в результате многомерного разложения исчезли сильные сдвиги фазы и амплитуды во временном ряде первой компоненты МНДМ (рис. 3.2а), и она стала практически идеально описывать годовой ход.

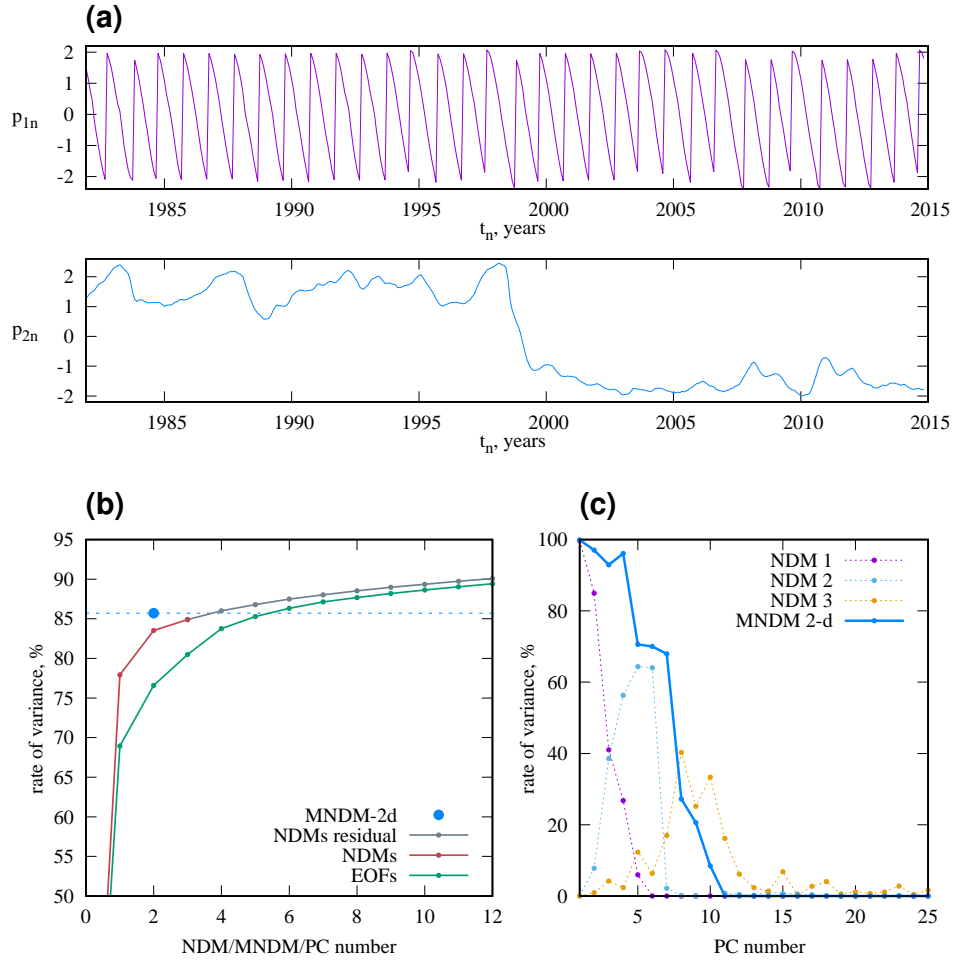


Рис. 3.2: (а) Временные ряды p_{i1}, \dots, p_{iN} ($i = 1, 2$), соответствующие найденной двумерной МНДМ. (б) Доля кумулятивной вариации, захватываемой модами, в процентах от вариации наблюдаемых данных. Синяя горизонтальная линия показывает вариацию двумерной МНДМ (примерно 86%), коричневые точки соответствуют одномерным НДМ, как на рис. 2.5b, серые точки – ЭОФ-ГК, полученным в результате ЭОФ-разложения остатков, зеленые точки – ЭОФ-ГК, полученным в результате ЭОФ-разложения данных ТПО. (с) Вклад одномерных НДМ (пунктирные линии) и двумерной МНДМ (сплошная линия) в различные ЭОФ-ГК (показана доля захватываемой вариации).

Важной особенностью полученной двумерной МНДМ является структура полиномиальной функции \mathbf{h} в (3.1): оказывается, часть этой функции, зависящая от перекрестных членов (см. раздел 3.2), близка к нулю. Иными словами, функция \mathbf{h} может быть представлена как сумма двух полиномиальных функций от одной переменной:

$$\mathbf{h}(\mathbf{p}_n) = \mathbf{h}_1(p_{1n}) + \mathbf{h}_2(p_{2n}). \quad (3.6)$$

Это позволяет отделить вклад компоненты p_{2n} от вклада компоненты p_{1n} (годового хода) в наблюдаемых данных. Таким образом, данные, полученные после вычитания вклада годового хода, $\mathbf{x}_n - \mathbf{h}_1(p_{1n})$, могут рассматриваться как полученные более корректно, чем при вычитании одномерной НДМ, временные ряды АТПО.

3.4 Использование двумерных мод для анализа изменчивости климата на интервале с 1870 по 2014 год

3.4.1 Используемые данные и характеристики МНДМ

Спутниковые измерения ТПО стали доступны лишь с конца XX века. Для того, чтобы проанализировать ТПО на более длинном интервале, необходимо привлечь данные других имеющихся измерений, например, данные корабельных наблюдений. Такие данные отличаются меньшей регулярностью и менее полным покрытием земного шара, особенно в более ранние годы. Поэтому для получения информации о ТПО в более регулярном виде (временные ряды на фиксированной пространственной сетке) применяются различные методы реанализа, задействующие как данные из различных источников, так и глобальные климатические модели. В данном разделе нами использовались данные реанализа ТПО из базы HadISST1 [102], представляющие собой временные ряды протяженностью 145 лет (с 1870 по 2014 год) с шагом 1 месяц на дискретной пространственной сетке с разрешением 1×1 градус.

В отличие от случая короткого временного ряда ТПО, рассмотренного в разделах 2.3 и 3.3, в этих данных в гораздо большей степени присутствуют сигналы с большими временными масштабами (тренды), связанные как с глобальным потеплением, так и собственной изменчивостью климата. Наличие этих трендов существенным образом влияет на построение нелинейных мод, описывающих динамику системы на декадных и межгодовых временных масштабах. Применение описанного выше алгоритма поиска МНДМ путем построения *независящего от «медленного» времени* многообразия по данному временному ряду становится затруднительным, поэтому в данном случае МНДМ находились по шести 33-летним перекрывающимся фрагментам полного 145-летнего временного ряда ТПО (точные границы этих фрагментов приведены в следующем разделе), причем последний фрагмент совпадал с временным рядом, рассмотренным в предыдущем разделе. Ясно, что таким образом можно уменьшить влияние трендов и лучше разрешить динамику системы на временных масштабах ТДК и ЭНЮК.

В результате применения алгоритма поиска двумерных МНДМ к каждому из шести фрагментов оказалось, что временной ряд первой компоненты p_{11}, \dots, p_{1N} у каждой МНДМ содержит периодический сигнал, соответствующий годовому ходу, а сами МНДМ также обладают свойством (3.6), позволяющим явно вычлест первую компоненту из данных и рассматривать остаток как АТПО.

Кроме того, заметим, что полученные по последнему фрагменту нелинейные ГК оказались практически идентичными найденным по другому массиву данных ТПО на том же интервале в предыдущем разделе. В частности, соответствующий годовому ходу временной ряд первой компоненты не содержит сильных сбоев фазы, как и ряд p_{1n} на рис. 3.2а, а во временном ряде второй компоненты (см. последний фрагмент на рис. 3.3 ниже) имеется переход, соответствующий смене фазы ТДК в 1998 году. Это дополнительно подтверждает грубость (в динамическом смысле, см. [4]) найденных МНДМ по отношению к конкретному источнику данных ТПО.

3.4.2 Реконструкция смены фаз тихоокеанского декадного колебания

В верхней части рис. 3.3 показаны временные ряды скрытых переменных p_{2n} , соответствующие МНДМ по шести взятым для анализа 33-летним фрагментам временного ряда ТПО: (1870-1902), (1887-1920), (1906-1939), (1930-1963), (1961-1994) и (1981-2014). Как видно, для четырех из этих фрагментов во второй компоненте МНДМ наблюдается переход, аналогичный переходу, ассоциирующемуся со сменой фазы ТДК и упоминаемому в разделе 2.3.3, что позволяет предположить связь найденных четырех переходов с ТДК.

В самом деле, связь между этими МНДМ-компонентами и ТДК очевидна из сравнения их временных рядов с временными рядами индекса PDO. В нижней части рис. 3.3 показаны индекс PDO, рассчитанный по наблюдаемым временным рядам АТПО, а также индексы PDO, реконструированные с помощью нелинейного преобразования временного ряда скрытой переменной p_{2n} . Такая реконструкция делается с помощью функции $\mathbf{h}_2(p_{2n})$ (см. раздел 3.3), которая отображает отсчеты временного ряда p_{2n} в пространство данных. После этого индекс PDO рассчитывается в соответствии с определением [25] путем проецирования данных на главную ЭОФ, рассчитанную по АТПО, как указано в Таблице 2.1. В реконструированном индексе прослеживается выраженная мультидекадная изменчивость ТДК и четко видны все переходы, проявляющиеся в виде резких скачков индекса между эпохами, характеризующимися либо отрицательными, либо положительными фазами ТДК. Таким образом, можно заключить, что выявленные с помощью двумерных МНДМ четыре климатических

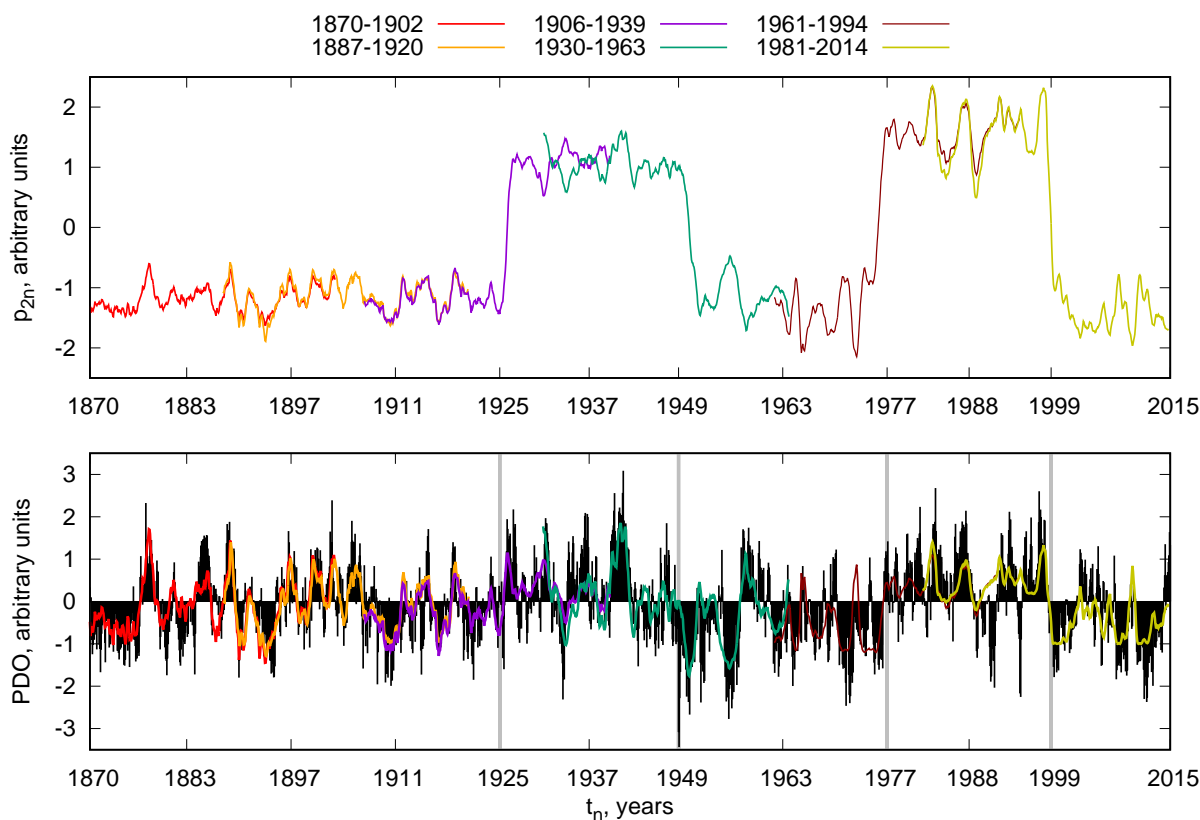


Рис. 3.3: МНДМ, построенные для различных фрагментов временного ряда ТПО (см. легенду). Вверху показаны временные ряды скрытой переменной p_{2n} (вторая компонента МНДМ). Внизу – индексы PDO, построенные на основе второй компоненты МНДМ (цветные линии) вместе с исходным индексом PDO, рассчитанным по наблюдаемым данным (черная линия).

перехода в XX веке, действительно, связаны с изменениями фазы ТДК: в 1925 и 1976 годах это переходы от отрицательной к положительной фазе ТДК, в 1949 и 1998 годах – обратные переходы. Эти результаты согласуются с работой [33], за исключением перехода в конце 40-х годов, который в данной работе был отнесен к 1946, а не к 1949 году. Кроме того, переходы в 1925 и 1976 годах согласуются с результатами [25].

Общепринятый способ визуализации глобального паттерна ТПО, ассоциирующегося с климатическим переходом, заключается в вычислении разницы между АТПО, усредненной за некоторый промежуток времени до и после перехода, либо наоборот (см. пример в разделе 2.3.3). Такие паттерны были построены для АТПО, реконструированных на основе временных рядов p_{2n} (рис. 3.4, слева): как видно из рисунка, они практически идентичны паттернам, вычисленным для АТПО по наблюдаемому временному ряду (рис. 3.4, справа). Характерная форма глобального ТДК-подобного

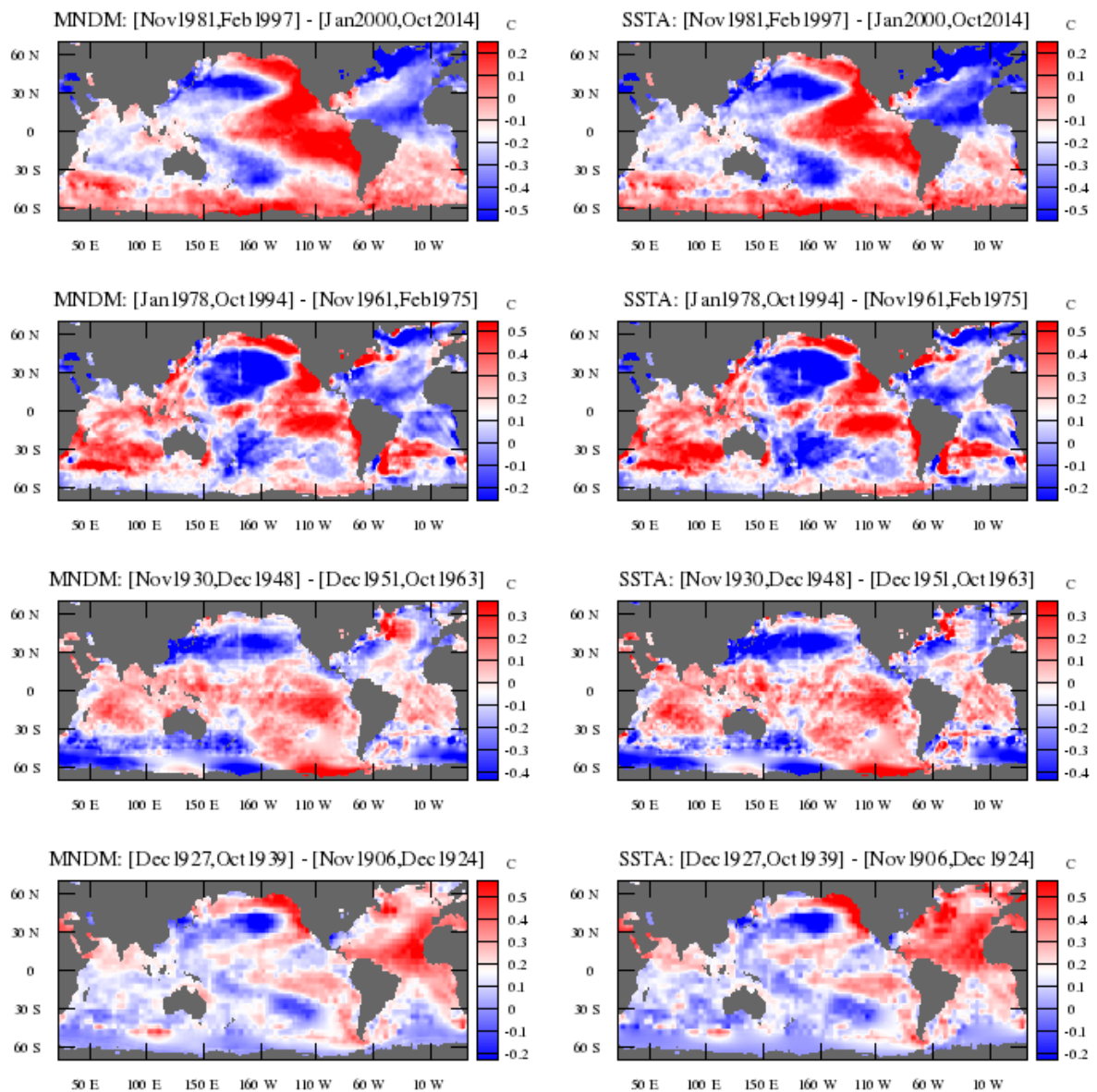


Рис. 3.4: Разность средних АТПО в положительной и отрицательной фазах ТДК для различных фрагментов временного ряда. Слева – паттерны, рассчитанные по временным рядам ТПО на основе второй компоненты МНДМ. Справа – те же паттерны, но рассчитанные по наблюдаемым рядам АТПО.

паттерна в Тихом океане имеет место для всех четырех рассмотренных переходов, тем самым подтверждая, что эти переходы в глобальном поведении ТПО действительно связаны с поведением ТДК. Тем не менее, паттерны для различных фрагментов ряда ТПО не полностью одинаковы, особенно для фрагмента (1930-1963), где теплые аномалии ТПО в разнице между положительной и отрицательной фазой ТДК имеют

большую протяженность в южной части Тихого океана, но необычно малы вдоль западного побережья Северной Америки. Такая картина, вероятно, связана с менее выраженным переходом ТДК в 1949 году, по сравнению с другими переходами, и, следовательно, большей неопределенностью во времени этого перехода, оцениваемом с помощью МНДМ.

3.4.3 Анализ дальних связей различных регионов в 1870-2014 годах

В Таблице 3.1 представлена доля суммарной вариации АТПО, захваченная МНДМ-компонентой p_{2n} для разных фрагментов временного ряда. Видно, что, в зависимости от фрагмента, от 30 до 42% вариации АТПО описывается одним скалярным временным рядом скрытой переменной p_{2n} . Это значит, что, по аналогии с разделом 2.3.4, данная компонента двумерной МНДМ связывает временные ряды аномалий ТПО в разных местах Земного шара посредством нелинейной функции $\mathbf{h}_2(p_{2n})$ от скалярного сигнала и, таким образом, описывает глобальную систему нелинейных дальних связей АТПО. Как видно из Таблицы 3.1, оптимальная степень полинома m_2 , задающего функцию $\mathbf{h}_2(p_{2n})$, меняется от 1 до 4 для разных фрагментов временного ряда. Заметим, что данный результат хорошо согласуется с результатами главы 2, в которой вторая НДМ получилась линейной для доиндустриального климата и нелинейной для современного климата.

| Фрагмент | 1870-1902 | 1887-1920 | 1906-1939 | 1930-1963 | 1961-1994 | 1981-2014 |
|--------------------|-----------|-----------|-----------|-----------|-----------|-----------|
| Доля вариации АТПО | 0.41 | 0.4 | 0.35 | 0.3 | 0.36 | 0.42 |
| $m_1; m_2$ | 7; 1 | 9; 2 | 9; 2 | 9; 4 | 9; 4 | 9; 4 |

Таблица 3.1: Характеристики МНДМ, найденных по различным фрагментам ряда ТПО: доля полной вариации АТПО, захваченная второй компонентой МНДМ, и оптимальные значения степеней полинома, соответствующих первой и второй компонентам МНДМ.

Для интерпретации картины дальних связей удобно рассмотреть карту значений коэффициентов корреляции между временными рядами компоненты $\mathbf{h}_2(p_{2n})$ в каждой точке пространственной сетки и соответствующими временными рядами исходного поля АТПО. Эти корреляции (рис. 3.5) указывают на вклад компоненты p_{2n} во временные ряды АТПО в различные регионы. Как видно из рисунка, они неравномерно распределены по земному шару.

Ясно, что для всех фрагментов временного ряда наибольший вклад второй ком-

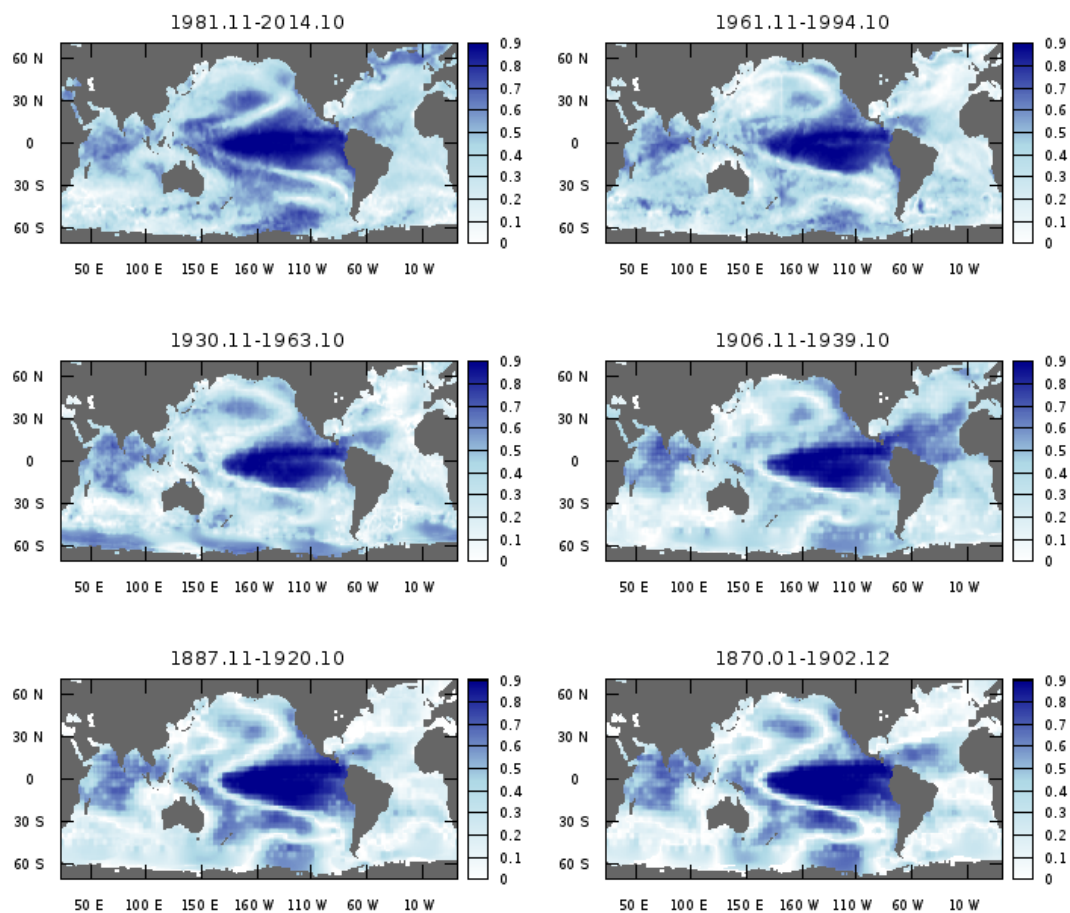


Рис. 3.5: Коэффициенты корреляции временных рядов второй компоненты МНДМ $h_2(p_{2n})$ и АТПО, рассчитанные в каждой точке пространственной сетки для различных фрагментов временного ряда.

| Индекс \ Фрагмент | Фрагмент | | | | | |
|-------------------|-----------|-----------|-----------|-----------|-----------|-----------|
| | 1870-1902 | 1887-1920 | 1906-1939 | 1930-1963 | 1961-1994 | 1981-2014 |
| Nino 3.4 | 0.96 | 0.96 | 0.96 | 0.95 | 0.94 | 0.96 |
| PDO | 0.71 | 0.78 | 0.7 | 0.75 | 0.66 | 0.71 |
| IO | 0.71 | 0.78 | 0.8 | 0.79 | 0.74 | 0.68 |
| SO | 0.69 | 0.72 | 0.72 | 0.63 | 0.71 | 0.8 |
| AMO | 0.4 | 0.25 | 0.83 | 0.3 | 0.28 | 0.74 |

Таблица 3.2: Коэффициенты корреляции между индексами, реконструированными на основе МНДМ, и индексами, вычисленными на основе данных АТПО. Индексы Nino 3.4 и PDO определены в Таблице 2.1, индексы IO, SP, AMO определяются как среднее значение в регионах $[30^{\circ}\text{S} - 30^{\circ}\text{N}; 40^{\circ}\text{E} - 100^{\circ}\text{E}]$, $[45^{\circ}\text{S} - 60^{\circ}\text{N}; 100^{\circ}\text{W} - 150^{\circ}\text{W}]$, $[0^{\circ}\text{N} - 75^{\circ}\text{N}; 0^{\circ}\text{W} - 80^{\circ}\text{W}]$ соответственно.

поненты МНДМ приходится на тропическую область Тихого океана: наибольшие значения корреляции наблюдаются в области, связанной с изменчивостью ЭНЮК (рис. 3.5). В частности, данные корреляции в регионе Niño 3.4 намного больше, чем в других местах земного шара. Тем не менее, МНДМ также вносит значительный вклад в регионы ТДК, Индийского океана (ИО) и южной части Тихого океана (SO) (см. рис. 3.5 и Таблицу 3.2). Из рис. 3.6, на котором изображены указанные в Таблице 3.2 индексы и их МНДМ-реконструкции на каждом фрагменте, видно, что высокие значения корреляции МНДМ/ТДК в регионе ЭНЮК (индекс Niño 3.4) означают почти идеальную реконструкцию межгодовой изменчивости Эль-Ниньо/Ла-Нинья по МНДМ. Исходя из этого, можно интерпретировать корреляционные распределения на рис. 3.5 как схему дальних связей ЭНЮК: чем выше корреляции в определенной

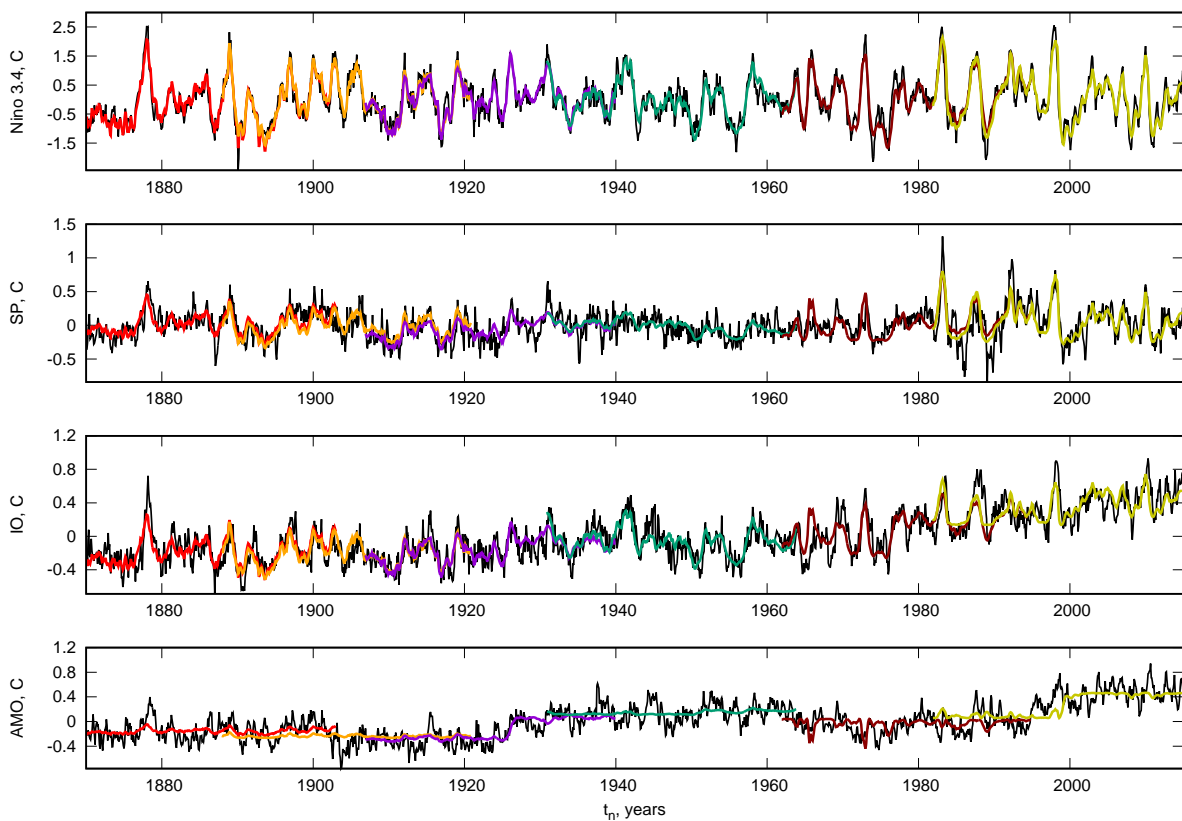


Рис. 3.6: Индексы ТПО Niño 3.4, SO, IO и AMO, рассчитанные на основе второй компоненты МНДМ (разные цвета соответствуют разным фрагментам временного ряда, как на рис. 3.3), в сравнении с индексами, рассчитанными по наблюдаемым данным (черный цвет). В данном случае индексы на различных фрагментах приведены к среднему значению всего временного ряда 1870-2014 гг. для того, чтобы обеспечить «сшиваемость» фрагментов.

области, тем сильнее связь динамики в этой области с изменчивостью ЭНЮК.

Из рис. 3.5, 3.6 и Таблицы 3.2 следует, что МНДМ указывает на тесную связь между регионами ЭНЮК, ТДК и южной частью Тихого океана. Это согласуется с известной гипотезой о наличии в Тихом океане связей между тропиками и субтропиками, обусловленных так называемым «атмосферным мостом», который передает сигнал ЭНЮК за пределы тропического пояса, а также посредством сигналов, распространяющихся в направлении экватора через океан (на более длинных временных масштабах) и низкоширотные атмосферные возмущения [79]. Из рис. 3.5 также видно, что в последние десятилетия сила этих связей возросла как для региона ТДК, так и для региона в южной части Тихого океана: для последнего фрагмента временного ряда корреляции данных и МНДМ-реконструкции в этих регионах выше, чем для всех предыдущих фрагментов.

Другим регионом, который тесно связан с динамикой тропического Тихого океана, является Индийский океан. Такая дальняя связь обусловлена вызываемыми ЭНЮК искажениями циркуляции Уолкера [76] и изменениями глубины термоклина в Индийском океане [132]; в свою очередь, Индийский океан может также влиять на динамику Эль-Ниньо/Ла-Нинья в результате изменения касательного напряжения зонального ветра в западной части Тихого океана [72]. Паттерн МНДМ на рис. 3.5 демонстрирует стабильные, почти неизменные связи между ЭНЮК и Индийским океаном на всех рассматриваемых фрагментах временного ряда ТПО.

Аналогично анализу, представленному в разделе 2.3.4, структура дальних связей, выявленных с помощью МНДМ, может быть изучена путем построения реконструиро-

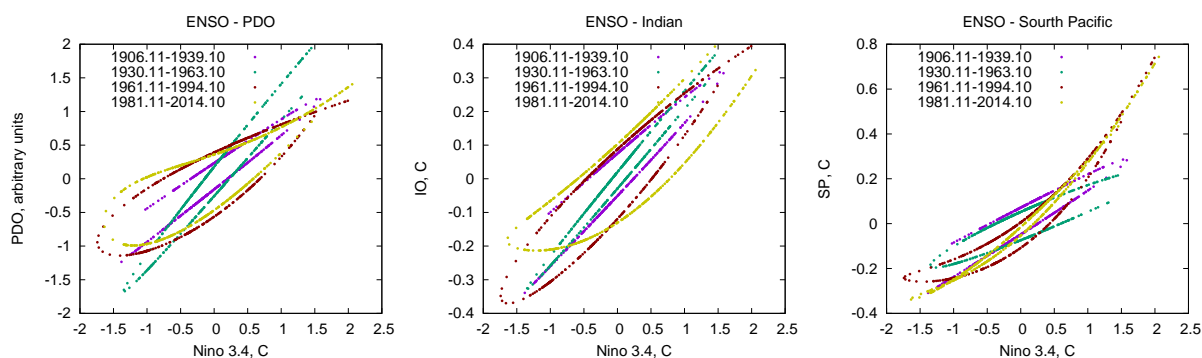


Рис. 3.7: Дальние связи между различными регионами, детектируемые с помощью МНДМ. Временные ряды индексов PDO, IO, SO и Nino 3.4 реконструированы на основе МНДМ и имеют нулевое среднее для каждого фрагмента временного ряда (цвета соответствуют рис. 3.3, 3.6). Показаны зависимости индексов PDO, IO, SO (слева направо) от Nino 3.4.

ванных индексов PDO, SO и IO в зависимости от индекса Nino 3.4 (рис. 3.7). В таком представлении различные фазы ТДК соответствуют различным ветвям одномерных кривых, визуализирующих вторую компоненту МНДМ. Видно, что до 1961 года ветви фаз ТДК близки к линейным функциям, однако затем, на двух последних фрагментах временного ряда, нелинейность кривых и расстояние между ветвями заметно увеличиваются. Интересно, что на двух последних фрагментах временного ряда структура выявленных зависимостей ЭНЮК-ТДК и ЭНЮК-IO не изменяется качественно. Связь между ЭНЮК и АТПО в южной части Тихого океана (индекс SO), согласно найденной МНДМ, перестает зависеть от фазы ТДК в последние десятилетия: желтая кривая, соответствующая последнему фрагменту 1981-2014 гг., выглядит почти как однозначная функция на графике ЭНЮК-SO.

Как видно из индекса Атлантической мультидекадной осцилляции (АМО), изображенного на рис. 3.6, найденная МНДМ крайне незначительно проявляется в Атлантическом океане. В основном, это связано с низкой корреляцией между МНДМ и изменчивостью ТПО в Атлантике на межгодовых временных масштабах. Тем не менее, как видно из рис. 3.4, 3.6, МНДМ отражает некоторые особенности декадной изменчивости АМО. Таким образом, аппарат МНДМ, вообще говоря, может быть полезен и для эмпирического исследования дальних связей с большими временными масштабами между ТПО в Атлантическом и Тихом океанах, с учетом ограничений вследствие сравнительно короткой для таких масштабов длины рассматриваемых в настоящей работе фрагментов (33 года).

3.5 Заключение

В данной главе предложен более общий способ построения нелинейных динамических мод, состоящий в поиске единого многомерного нелинейного многообразия. Соответствующие моды названы многомерными нелинейными динамическими модами (МНДМ). На модельном примере и на примере спутниковых измерений ТПО на интервале с 1981 по 2014 год показана лучшая адекватность МНДМ, по сравнению с набором одномерных НДМ, в задаче поиска скрытых переменных, описывающих наблюдаемую систему. В частности, двумерная МНДМ ТПО описывает большую часть вариации данных, чем три одномерных НДМ, а также лучше выделяет скрытые переменные, соответствующие годовому ходу и явлениям ЭНЮК и ТДК с межгодовыми и декадными масштабами.

Результаты применения данного метода к более длинному ряду реанализа ТПО (с 1870 по 2014 год), разбитому на несколько перекрывающихся по времени фрагментов, на основе прямого анализа данных измерений подтверждают ряд имеющихся гипотез

о причинах наблюдаемой динамики климатической системы. В частности, с помощью МНДМ получено подтверждение четырех климатических переходов в XX веке, связанных с изменением фазы ТДК. Кроме того, аппарат двумерных НДМ позволяет сделать выводы о дальних связях в климате Земли и их эволюции: показано, что структура дальних связей, определяемых Эль-Ниньо Южным Колебанием, стала существенно нелинейной, начиная с 60-х годов XX века.

Анализ изменения степени нелинейности второй компоненты МНДМ, ответственной за доминирующий режим изменчивости АТПО, при переходе от доиндустриального климата к современному подтверждает высказанную в конце предыдущей главы гипотезу о том, что смена линейной структуры мод на нелинейную вызвана скорее нестационарностью климатической системы, обусловленной различными медленными воздействиями (например, ростом концентрации парниковых газов в атмосфере), чем источником и особенностями выборки данных доиндустриального климата, использованных в главе 2.

Глава 4

Эмпирический прогноз динамики явления Эль-Ниньо

4.1 Введение

В этой главе методы эмпирической реконструкции, разработанные в предыдущих главах, будут использованы для построения прогностической модели для пространственно распределенных аномалий ТПО в тропическом поясе, где доминирующим режимом изменчивости на межгодовых масштабах является Эль-Ниньо Южное Коледание [120], которое, в свою очередь, оказывает влияние на режимы погоды во многих регионах земного шара (см. Введение). Поэтому модель, построенная по ежемесячным данным АТПО в тропиках, будет ориентирована, главным образом, на прогноз динамики индексов ЭНЮК.

Несмотря на то, что теоретическое объяснение Эль-Ниньо, основанное на гипотезе Бьеркнеса о положительной океано-атмосферной обратной связи [23] сейчас является практически общепринятым, конкретные механизмы, вовлеченные в динамику ЭНЮК, до сих пор являются предметом исследований [127]. Так, согласно одной точке зрения, Эль-Ниньо является одним из режимов эволюции нелинейного океано-атмосферного тропического осциллятора [101, 115, 131, 135]. Другое объяснение трактует Эль-Ниньо как результат стохастического (мелкомасштабного) внешнего воздействия на такой осциллятор [62, 95]. При этом крайне актуальная задача прогноза динамики Эль-Ниньо все еще далека от удовлетворительного решения [16, 118]. На данный момент в мире существует ансамбль моделей, предсказывающих индексы Эль-Ниньо в режиме «реального времени» (ежемесячно), называемый «ENSO Prediction Plume» [16], в который входят как современные климатические модели «из первых принципов» (например, [107, 133]), так и эмпирические модели (например, [70, 93]). Коэффици-

ент корреляции прогнозов наилучших моделей из этого ансамбля с наблюдениями составляет лишь 0.6 для дальности прогноза 6 месяцев, согласно [16].

Стоит также отметить, что эмпирические модели из этого ансамбля основаны, как правило, на использовании ЭОФ-разложения. В данной главе предлагается эмпирическая модель, основанная на многомерном обобщении ЛДМ-разложения, введенного в разделе 2.2.4, т.е. на линейном варианте МНДМ, и показывается, что она имеет достаточно хорошую прогностическую способность, в сравнении с моделями из упомянутого ансамбля. Грубо говоря, данная модель описывает ЭНЮК как результат эволюции низкоразмерной нелинейной динамической системы, находящейся под возможным периодическим воздействием с периодом 1 год (годовой цикл) и возможным стохастическим воздействием, т.е. способна оптимальным (в смысле байесова подхода) образом учесть обе существующих версии объяснения ЭНЮК, основываясь на прямом анализе данных. Предполагается, что созданная модель может быть использована для регулярного ежемесячного прогноза индексов ЭНЮК, а также включена в ансамбль «ENSO Prediction Plume».

Алгоритм построения предлагаемой модели и схемы прогноза представлен в разделе 4.2. Используемые данные ТПО описываются в разделе 4.3. В разделе 4.4 проводится сравнение модели, основанной на использовании многомерных ЛДМ, с такой же моделью, основанной на использовании обычного ЭОФ-разложения; показывается, что модель на основе нового разложения имеет лучшую прогностическую способность. Анализ прогностической способности предложенной модели для динамики ЭНЮК, а также сравнение ее с прогностической способностью моделей из ансамбля «ENSO Prediction Plume», представлены в разделе 4.5

Материалы данной главы представлены в статье [А6] и публикациях [А10, А21, А59–А71].

4.2 Алгоритм построения модели и прогноза

При построении прогностической модели мы будем придерживаться схемы реконструкции оператора эволюции пространственно распределенной системы в виде модифицированной стохастической модели (раздел 1.4 главы 1), а для выбора фазовых переменных будем использовать отличный от ЭОФ метод редукции размерности данных.

4.2.1 Выбор переменных

Исходя из результатов глав 2 и 3, в задаче прогноза АТПО представляется адекватным вместо ЭОФ-разложения использовать многомерные нелинейные динамические моды (МНДМ): при одной и той же размерности главного многообразия d (т.е. количества ГК) оно должно более эффективно описывать вариацию исходных данных (за счет нелинейности) и лучше выделять временные масштабы, доминирующие в динамике системы (за счет заложенной байесовой оптимизации временных масштабов ГК), приводя к лучшим фазовым переменным. Однако созданный алгоритм МНДМ является весьма ресурсоемким, что обусловлено, в первую очередь, быстрым ростом количества коэффициентов многомерного полинома, используемого для параметризации МНДМ \mathbf{h} в уравнении (3.1), при увеличении размерности аргумента d и степени полинома \mathbf{m} .

Фактически, в текущей реализации метода уже при размерности многообразия $d = 3$ возникают проблемы с ресурсоемкостью поиска оптимальной МНДМ для рассматриваемой задачи (особенно с учетом необходимости ежемесячного прогноза ЭНЮК, см. Введение 4.1). Как отмечено в главе 3, оптимизация метода поиска МНДМ за счет выбора других параметризующих функций или повышения эффективности вычислительного алгоритма может являться направлением будущего исследования. Поскольку в данной работе будут анализироваться, в том числе, случаи с размерностью главного многообразия $d > 2$, мы будем использовать для редукции размерности линейный вариант МНДМ – многомерные линейные динамические моды (МЛДМ). В этом случае количество коэффициентов линейного полинома оказывается приемлемым для численной реализации. При этом, несмотря на линейность, МЛДМ могут быть более адекватными с точки зрения выделения доминирующих режимов динамики, чем ЭОФ-разложение (см. примеры с одномерными ЛДМ в разделах 2.2.4 и 2.4). Заметим также, что эффекты, связанные с нелинейностью системы и нестационарностью временного ряда могут быть учтены при последующем построении оператора эволюции (см. раздел 4.2.2).

МЛДМ приводит к следующему представлению исходного временного ряда $\mathbf{x}_1, \dots, \mathbf{x}_N, \mathbf{x}_n \in \mathbb{R}^D$:

$$\mathbf{x}_n = \mathbf{L}(\mathbf{p}_n) + \mathbf{r}_n, \quad (4.1)$$

где \mathbf{L} – линейное отображение, \mathbf{p}_n – значение искомым ГК в момент времени t_n , \mathbf{r}_n – остаток разложения. Алгоритм поиска МЛДМ эквивалентен описанному в разделе 3.2 алгоритму поиска МНДМ, в котором степени полинома по каждому из аргументов p_{in} зафиксированы и равны $m_i = 1$, а также отсутствуют перекрестные слагаемые вида $p_{in}p_{kn}$.

Эффективность переменных, выбранных с помощью метода МЛДМ, применительно к задаче прогноза ЭНЮК, будет сравниваться с эффективностью переменных на основе ЭОФ-разложения, представляющего данные в форме (1.39).

4.2.2 Построение модели оператора эволюции

Как следует из предыдущего раздела, при обоих способах выбора фазовых переменных данные представляются через ГК $\mathbf{p}_n \in \mathbb{R}^d$ и временные ряды остатка $\mathbf{r}_n \in \mathbb{R}^D$. Мы будем реконструировать оператор эволюции для них по отдельности.

Оператор эволюции для ГК будем строить в форме модифицированной стохастической модели на основе ИНС (см. раздел 1.4.3):

$$\begin{aligned} \mathbf{p}_{n+1} &= \mathbf{p}_n + \mathbf{f}(\mathbf{p}_n, \dots, \mathbf{p}_{n-l+1}; \boldsymbol{\kappa}_n; t_n; \boldsymbol{\mu}_f) + \hat{\mathbf{g}}(\mathbf{p}_n, \dots, \mathbf{p}_{n-l+1}; \boldsymbol{\mu}_g) \cdot \boldsymbol{\xi}_n, \\ n &= l, \dots, N-1. \end{aligned} \quad (4.2)$$

Все обозначения здесь идентичны обозначениям выражения (1.40). Байесов алгоритм построения данной модели также полностью описан в разделе 1.4.3.

Переходя к модели для остатка \mathbf{r}_n , заметим, что как в случае использования МЛДМ, так и при ЭОФ-разложении этот остаток может оказаться автокоррелированным во времени и пространстве из-за упрощающих предположений (линейность многообразия, дельта-коррелированность и гауссовость остатка). Учет этой автокоррелированности может быть особенно важен в задаче краткосрочного прогноза траектории, который будет делаться для индексов ЭНЮК в данной главе. Поэтому мы будем использовать следующую простую модель для остатков.

Во-первых, сначала произведем полный ЭОФ-поворот остатков и получим их ГК \mathbf{q}_n :

$$\mathbf{r}_n = \mathbf{U} \cdot \mathbf{q}_n. \quad (4.3)$$

Здесь \mathbf{U} – соответствующая матрица поворота размера $D \times D$. Очевидно, в случае ЭОФ-разложения исходных данных данные ГК являются просто продолжением этого разложения.

Во-вторых, для каждой из полученных ГК построим линейную авторегрессионную модель первого порядка:

$$q_{kn} = b_k \cdot q_{k,n-1} + \sigma_k \cdot \eta_{kn}. \quad (4.4)$$

Здесь k – номер ГК остатка, $\eta_{k1}, \dots, \eta_{kN}$ – дельта-коррелированный гауссов процесс с единичной дисперсией, b_k и σ_k – обучаемые параметры модели. Фактически, уравнение (4.4) трактует каждую ГК остатка как красный шум (если в результате обучения $b_k \neq 0$).

Таким образом, необходимо вычислить набор параметров $\mathbf{b} = (b_1, \dots, b_D)$ и $\boldsymbol{\sigma} = (\sigma_1, \dots, \sigma_D)$, определяющий модель для остатка.

Заметим, что если бы остаток был просто отброшен, как делается обычно в моделях на основе ЭОФ, то это было бы эквивалентно модели в форме белого шума вместо (4.4). Для правомерности такого предположения требовалось бы включить в «основную» модель (4.2) большое количество ГК. Здесь же, наоборот, мы имеем более мощную модель (4.4) для остатка, что позволяет, вообще говоря, использовать меньшее количество ГК в основной модели без ухудшения общей прогностической способности.

4.2.3 Схема построения прогноза

Пусть имеется построенная эмпирическая модель, т.е. известны $\mathbf{f}, \hat{\mathbf{g}}, \mathbf{U}, \mathbf{b}, \boldsymbol{\sigma}$ и \mathbf{V} (или \mathbf{L}) при использовании ЭОФ (или МЛДМ, соответственно) в результате применения алгоритмов оптимизации и обучения, описанных в разделах 4.2.1, 4.2.2 и предыдущих главах. Тогда с помощью этой модели для любого начального момента времени $t_n \geq t_l$ возможен прогноз следующих за ним T состояний системы: сначала делается прогноз ГК и остатков на основе уравнений (4.2)-(4.4); затем с помощью применения уравнения (1.39) для ЭОФ-ГК (уравнения (4.1) для МЛДМ, соответственно) получается прогноз исходных наблюдаемых переменных \mathbf{x}_n .

Поскольку модель является стохастической, результатом прогноза является случайная последовательность $\mathbf{x}'_{n+1}(\boldsymbol{\xi}, \boldsymbol{\eta}), \dots, \mathbf{x}'_{n+T}(\boldsymbol{\xi}, \boldsymbol{\eta})$, которая может быть вычислена для любой реализации гауссовых случайных процессов $\boldsymbol{\xi}$ и $\boldsymbol{\eta}$.

Вообще говоря, дальнейшая обработка такого прогноза определяется пользователем. Например, для любого скалярного индекса $I(\mathbf{x}_n)$, определяемого через исходные переменные, можно вычислить его медианное значение по множеству реализаций процессов $(\boldsymbol{\xi}, \boldsymbol{\eta})$, которое можно считать «типичным» значением индекса I в момент времени $n + j$, предсказанным моделью с дальностью j временных отсчетов:

$$\tilde{I}'_{n,j} = \underset{\boldsymbol{\xi}, \boldsymbol{\eta}}{\text{Median}}[I(\mathbf{x}'_{n+j}(\boldsymbol{\xi}, \boldsymbol{\eta}))]. \quad (4.5)$$

Использование для этой цели медианы вместо среднего значения представляется более адекватным, потому что при $j \geq 2$ ПВ индекса становится негауссовой ввиду нелинейности оператора эволюции.

В настоящей работе медианное значение (4.5) будет считаться главным спрогнозированным значением при вычислении среднеквадратичной ошибки (СКО) прогноза – одной из стандартных мер качества прогноза при сравнении различных моделей.

4.3 Описание используемых данных

Алгоритм построения эмпирической модели будет применен к ежемесячным данным ТПО в тропическом поясе $[30^{\circ}\text{S}–30^{\circ}\text{N}]$ на интервале с января 1960 года по декабрь 2007 года, взятым из базы NOAA_ERSST_V4¹ [53, 78] и имеющим пространственное разрешение $2^{\circ} \times 2^{\circ}$.

Как видно, длина интервала обучения модели составляет 48 лет, что очень близко к рассмотренной в разделе 1.4 длине 50 лет для временного ряда ТПО модели INMCM4.0. Кроме того, прогностическая способность построенной модели будет исследована на независимом интервале с 2008 по 2017 год в режиме, когда прогнозируемые значения лежат за пределами обучения и оптимизации модели, см. подробнее раздел 4.5.2.

Как и в разделе 1.4, модель будет строиться по данным АТПО, а нелинейное влияние годового сигнала на систему будет учитываться в виде зависимости оператора эволюции от периодического внешнего воздействия (1.45) в модели (4.2).

Таким образом, в следующих разделах будут рассмотрены следующие два типа эмпирической модели АТПО:

1. Нестационарная нелинейная стохастическая модель на основе ИНС для ЭОФ-ГК (модель ЭОФ-ИНС);
2. Нестационарная нелинейная стохастическая модель на основе ИНС для новых ГК на основе метода МЛДМ (модель МЛДМ-ИНС).

Дополнительно, при рассмотрении различных климатических индексов мы будем сравнивать эффективность двух вышеупомянутых моделей с эффективностью стандартной авторегрессионной линейной модели первого порядка для каждого индекса (модель AR1).

4.4 Сравнение эффективности моделей МЛДМ-ИНС и ЭОФ-ИНС

В данном разделе проводится сравнение эффективности моделей, описанных выше, при размерностях главного многообразия d .

Для каждого d в диапазоне от 1 до 6 были построены модели ЭОФ-ИНС и МЛДМ-ИНС, т.е. найдены структурные параметры оптимальной МЛДМ и оптимальных операторов эволюции (набор временных масштабов τ ; параметр предварительной редукции K ; числа нейронов m_f, m_g ; величина задержки l ; см. выражения (3.2), (2.4), (4.2),

¹Данные предоставлены NOAA/OAR/ESRL PSD, Boulder, Colorado, USA и находятся в открытом доступе на сайте <http://www.esrl.noaa.gov/psd/>

а также все сопутствующие параметры $\mu_f, \mu_g, \mathbf{U}, \mathbf{b}, \sigma, \mathbf{V}, \mathbf{L}$ при этих структурных параметрах (см. выражения (4.1)-(4.4), (1.39)). Размерности d выше 6 не были рассмотрены, главным образом, из-за повышающейся вычислительной ресурсоемкости второго этапа эмпирической реконструкции – построения модифицированной стохастической модели оператора эволюции – с ростом d .

4.4.1 Временные ряды главных компонент

Согласно разделу 4.2.1, МЛДМ учитывает доминирующие в системе временные масштабы, в то время как ЭОФ-разложение ориентировано только на захват максимальной вариации исходных данных. Рис. 4.1 иллюстрирует данное отличие в случае размерности $d = 6$ для рассматриваемых данных АТПО. В левой колонке рис. 4.1

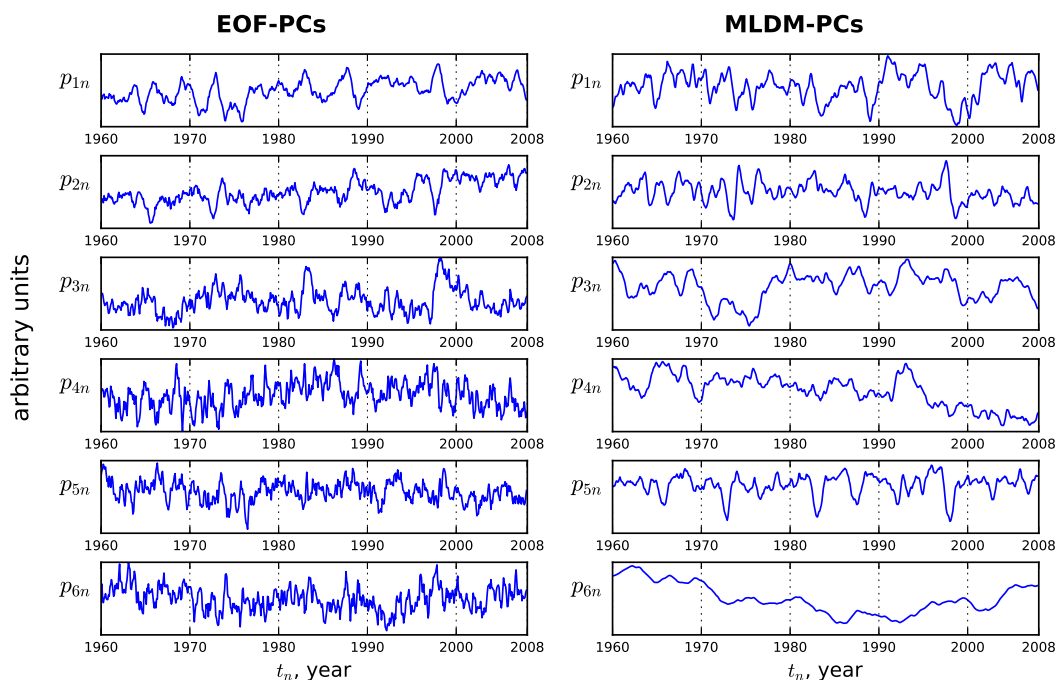


Рис. 4.1: Временные ряды ГК, соответствующие проекции 48-летних данных АТПО в тропиках на главное 6-мерное линейное многообразие, полученные с помощью ЭОФ-разложения (левая колонка) и поиска МЛДМ (правая колонка).

показаны 6 временных рядов ЭОФ-ГК, которые описывают примерно 71.3% полной вариации АТПО, а в правой колонке изображены временные ряды ГК 6-мерной МЛДМ, захватывающей 68.6% полной вариации данных. Но при этом видно, что МЛДМ лучше выделяет характерные временные масштабы, по сравнению с ЭОФ-разложением. В частности, МЛДМ лучше сфокусирована на полосе межгодовых и

декадных масштабов, а также на медленных трендах, которые доминируют в наблюдаемых временных рядах, в то время как «быстрая» шумоподобная изменчивость с масштабам, не превышающим нескольких временных отсчетов, практически исключена. В то же время, межгодовые, декадные и медленные масштабы отчасти перемешаны в различных ГК МЛДМ; это можно объяснить линейностью МЛДМ, которая естественным образом ограничивает гибкость разделения масштабов, особенно при не очень большой длине временного ряда.

Заметим, что ЭОФ-ГК на рис. 4.1 упорядочены по убыванию доли полной вариации, описываемой каждой из них. В то же время, компоненты МЛДМ, фактически, не упорядочены, потому что они, вообще говоря, коррелируют друг с другом, и поэтому вклад каждой из них в полную вариацию зависит от других. Также, несмотря на то, что компоненты МЛДМ описывают суммарно меньшую долю вариации, чем такое же количество ЭОФ-ГК, определяемое ими линейное многообразие не является подмножеством этих ЭОФ. В частности, для изображенного на рис. 4.1 случая в МЛДМ дают ненулевой вклад первые 13 ЭОФ, хотя вклад ЭОФ 1-6 в нее является доминирующим.

Наконец, временные ряды МЛДМ другой размерности d (от 1 до 5), которые здесь не показаны, демонстрируют аналогичные свойства. Заметим также, что, в отличие от ЭОФ-ГК, временные ряды МЛДМ другой размерности в общем случае не совпадают с изображенными на рис. 4.1.

4.4.2 Прогностическая способность моделей

Для анализа эффективности моделей в данном разделе будем использовать СКО прогноза, вычисленную для ансамбля прогнозов того или иного скалярного индекса I , сделанных с разных моментов времени:

$$RMSE_I(j) = \sqrt{\frac{\sum_{n=l}^{N-j} (I_{n+j} - \tilde{I}'_{n,j})^2}{N - j - l + 1}}. \quad (4.6)$$

Здесь $RMSE_I(j)$ – СКО прогноза индекса I при дальности j месяцев (от англ. root-mean-square error); I_{n+j} – истинное значение индекса I в момент t_{n+j} ; $\tilde{I}'_{n,j}$ – значение, предсказанное моделью, стартовавшей в момент t_n , которое получено на основе ансамбля из 500 стохастических реализаций по формуле (4.5); l – величина временной задержки в полученной оптимальной модели оператора эволюции (4.2). Меньшее значение СКО соответствует лучшей прогностической способности модели для рассматриваемого индекса; заметим, что возможный эффект влияния «переобучения» модели на СКО при этом практически исключен благодаря проведенной

байесовой оптимизации структурных параметров модели. Как правило, СКО прогноза любого индекса растёт с увеличением дальности прогноза j , насыщаясь при некотором ее значении, которое можно назвать горизонтом «краткосрочного» (см. главу 1) прогноза.

Для демонстрации работы моделей и сравнительного анализа их эффективности рассмотрим в качестве набора индексов временные ряды главных ЭОФ-ГК: каждый индекс есть линейная ортогональная проекция данных \mathbf{x}_n на соответствующую ЭОФ:

$$I(\mathbf{x}_n) = \mathbf{v}^T \mathbf{x}_n, \quad (4.7)$$

где \mathbf{v} – вектор-столбец одной из главных ЭОФ. Очевидно, СКО прогноза таких индексов связана с СКО прогноза главной части изменчивости полного пространственно распределенного поля АТПО.

На рисунках в этом разделе будет показана СКО, нормированная на среднеквадратичное отклонение индекса, из которого удален линейный тренд; таким образом, данная СКО измеряет долю неправильного предсказания не связанной с линейным трендом изменчивости индекса. Фактически, теперь вклад вариации линейного тренда в полную вариацию индекса не занижает СКО краткосрочного прогноза внутренней изменчивости, которая нас интересует.

На рис. 4.2 (слева вверху) показана СКО прогноза ЭОФ-ГК 1-4 с помощью моделей ЭОФ-ИНС и МЛДМ-ИНС при размерности главного многообразия $d = 1$, а также СКО прогноза линейной авторегрессионной моделью AR1, которая не зависит от величины d . Как видно, в случае $d = 1$ нелинейные модели МЛДМ-ИНС и ЭОФ-ИНС улучшают прогноз только для первого индекса. Кроме того, модель МЛДМ-ИНС демонстрирует немного лучшую прогностическую способность при дальностях до 9 месяцев, чем модель ЭОФ-ИНС; это подтверждает, что МЛДМ более точно выделяют информацию о динамике системы. Очевидно, прогноз других индексов (ЭОФ-ГК с номером 2 и больше) в случае $d = 1$ обеспечивается, главным образом, моделью для остатка (4.4), которая, фактически, является моделью AR1 для ЭОФ-ГК остатка; поэтому все СКО совпадают с СКО модели AR1 для этих индексов.

Аналогичным образом, в других случаях $d = 2 \div 6$ (рис. 4.2) мы имеем значительное улучшение СКО прогноза первых d индексов нелинейными моделями, по сравнению с моделью AR1; а также превосходство модели МЛДМ-ИНС над моделью ЭОФ-ИНС в краткосрочном прогнозе. Более того, начиная с размерности $d = 4$, модель МЛДМ-ИНС дает вклад в прогноз индексов с номерами выше d , например, СКО прогноза ЭОФ-ГК 6 при $d = 4$ и $d = 5$ на рис. 4.2 (ЭОФ-ГК с номерами 7 и выше здесь не приводятся). Заметим, что модель ЭОФ-ИНС на такое не способна по построению.

Наконец, на рис. 4.2 отчетливо видно, что улучшение общей прогностической

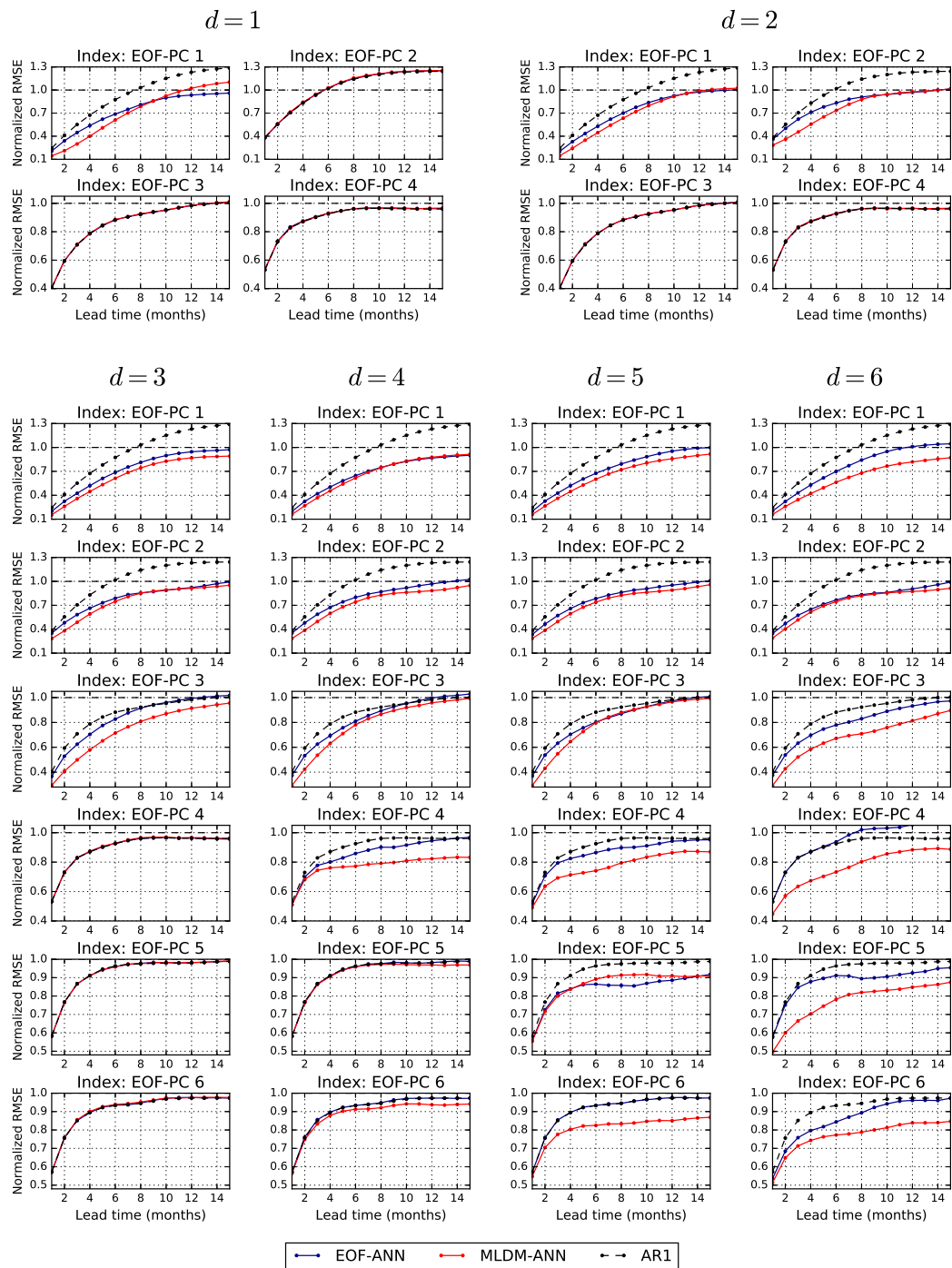


Рис. 4.2: Нормированная СКО прогноза различных ЭОФ-ГК с помощью моделей ЭОФ-ИНС (синий цвет), МЛДМ-ИНС (красный цвет), AR1 (черный пунктир), см. раздел 4.3. Модели ЭОФ-ИНС и МЛДМ-ИНС были построены для размерностей главного многообразия $d = 1$ (4 верхних левых рисунка для ЭОФ-ГК 1-4), $d = 2$ (4 верхних правых рисунка для ЭОФ-ГК 1-4), $d = 3, 4, 5, 6$ (нижние столбцы рисунков 1,2,3,4, соответственно; в каждом случае показана СКО прогноза ЭОФ-ГК 1-6).

способности модели МЛДМ-ИНС, по сравнению с моделью ЭОФ-ИНС, становится все более и более значимым с ростом d . Это полностью соответствует идее, используемой при построении модели МЛДМ-ИНС: поиску многообразия, наилучшим образом сохраняющего динамические свойства системы при минимальных потерях полной вариации исходных данных.

4.4.3 Регионы с наилучшей предсказуемостью

Лучшую СКО на рис. 4.2 имеет модель МЛДМ-ИНС при значении размерности главного многообразия $d = 6$ (в рамках рассмотренного диапазона значений d). Далее исследуем эту модель более детально. Оптимальные значения ее структурных параметров следующие: величина временной задержки $l = 3$, количество нейронов в детерминированной части оператора эволюции $m_f = 7$, количество нейронов в стохастической части оператора эволюции $m_g = 0$ (как и в разделе 1.4, оптимальная стохастическая компонента получилась однородной), предварительная размерность редукции для построения МЛДМ $K = 13$.

На рис. 4.3 показана СКО прогноза всех компонент ряда \mathbf{x}_n с помощью данной модели, т.е. пространственное распределение СКО в тропическом поясе. Она вычислена для временного ряда АТПО в каждом узле пространственной сетки согласно выражению (4.6) и затем нормирована, как описано в разделе 4.4.2. Видно, что АТПО

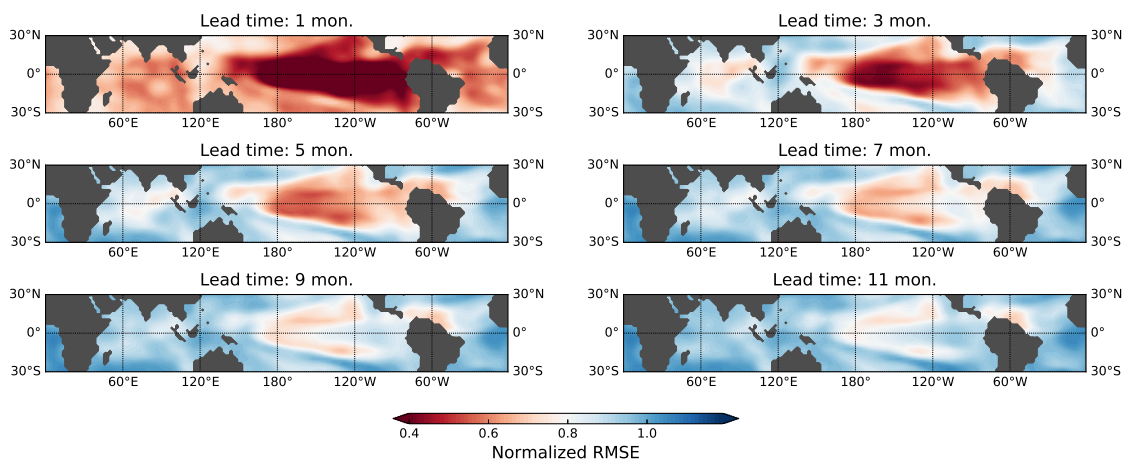


Рис. 4.3: Пространственное распределение нормированной СКО 6-мерной модели МЛДМ-ИНС для АТПО в тропическом поясе (СКО изображена цветом в каждой географической точке) при различных дальностях прогноза (см. заголовки рисунков).

в Индийском и тропическом Атлантическом океанах может быть предсказана не более чем на 3 месяца вперед, в то время как в Тихоокеанском регионе горизонт

прогноза значительно выше (вплоть до 11 месяцев в отдельных регионах). Причина такого отличия кроется в различии динамических свойств АТПО в данных регионах. Изменчивость АТПО в тропическом Тихом океане происходит в сравнительно узком диапазоне временных масштабов, соответствующем ЭНЮК (от 2 до 7 лет), при этом она дает основной вклад в полную вариацию АТПО в тропиках (рис. 4.4). Оставшаяся часть изменчивости АТПО в тропиках обеспечивается Индийским и тропическим Атлантическим океанами, диапазон временных масштабов динамики которых значительно шире – от сильного тренда глобального потепления до быстрых процессов, близких к шагу временного ряда (рис. 4.4). Ясно, что из перечисленных процессов

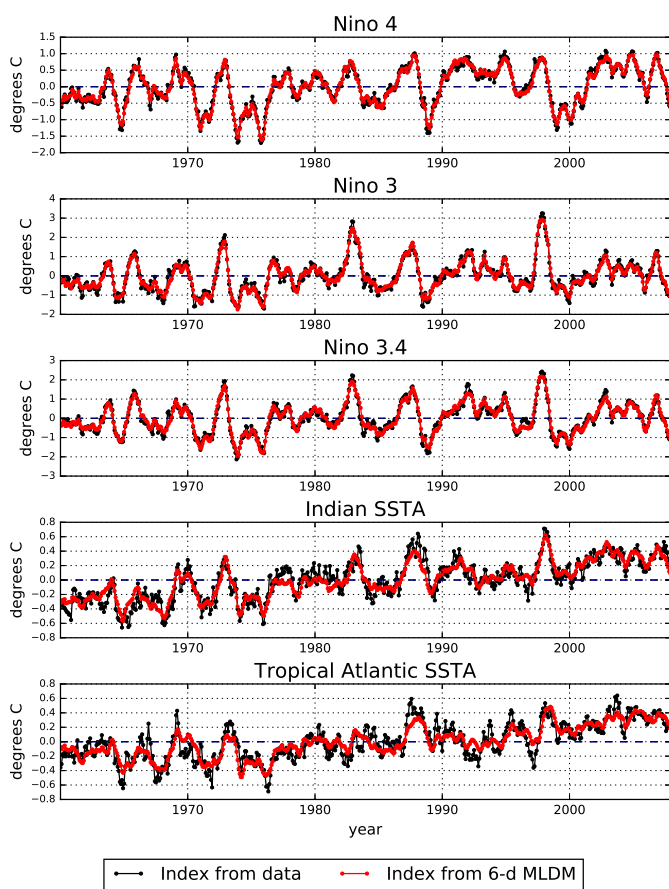


Рис. 4.4: Временные ряды (индексы) АТПО, усредненной по различным регионам: Nino 4 [$160^{\circ}\text{E} - 150^{\circ}\text{W}, 5^{\circ}\text{S} - 5^{\circ}\text{N}$], Nino 3 [$150 - 90^{\circ}\text{W}, 5^{\circ}\text{S} - 5^{\circ}\text{N}$], Nino 3.4 [$170 - 120^{\circ}\text{W}, 5^{\circ}\text{S} - 5^{\circ}\text{N}$], Индийский океан [$40 - 100^{\circ}\text{E}, 30^{\circ}\text{S} - 30^{\circ}\text{N}$], тропический Атлантический океан [$300^{\circ}\text{W} - 10^{\circ}\text{E}, 30^{\circ}\text{S} - 30^{\circ}\text{N}$]. Черные линии соответствуют наблюдаемым значениям индексов, красными линиями показана реконструкция индексов на основе 6-мерной МЛДМ.

низкоразмерная (6-мерная) МЛДМ описывает, в первую очередь, процессы с большей вариацией, т.е. межгодовую изменчивость АТПО, ассоциированную с тропическим

Тихим океаном, и тренд глобального потепления, а остальные процессы описываются ей хуже. Фактически, интенсивные быстрые процессы в Индийском и тропическом Атлантическом океанах уже на этапе выбора переменных исключаются из модели (т.е. моделируются как шум), что и ведет к более низкой прогностической способности модели в этих регионах.

Таким образом, из рис. 4.3 можно заключить, что наилучшую прогностическую способность (с дальностью вплоть до 9-11 месяцев) модель МЛДМ-ИНС имеет в центральном и восточном тропическом Тихоокеанском регионе, ассоциирующемся с явлением ЭНЮК. Заметим, что в северо-западном тропическом Атлантическом океане также имеется хорошо предсказуемая область, которая, возможно, связана с ЭНЮК через дальние связи в атмосфере (см., например, [9]).

4.5 Приложение к прогнозу динамики Эль-Ниньо

В данном разделе будет исследована эффективность построенной модели МЛДМ-ИНС (а также, для сравнения, модели ЭОФ-ИНС) в задаче краткосрочного прогноза динамики ЭНЮК. Для этой цели будут использованы две меры качества прогноза, в точности совпадающие с использованными в работе [16] при оценке моделей, предоставляющих прогноз индексов ЭНЮК, из ансамбля «ENSO prediction plume» (см. раздел 4.1). Данные меры качества прогноза применяются к прогнозируемым значениям индекса, сглаженным с временным окном 3 месяца; при этом дальность прогноза нумеруется таким образом, что среднее значение на трехмесячном интервале, идущем непосредственно за последним «наблюдаемым» месяцем при построении прогноза, соответствует дальности прогноза в 1 месяц.

Первая мера основана на выражении для СКО (4.6), вычисленном с иной нормировкой. В частности, каждое отклонение прогнозируемого значения индекса от истинного $I_{n+j} - \tilde{I}'_{n,j}$ нормируется на среднеквадратичное отклонение значения индекса в соответствующем календарном месяце, т.е. нормировочный коэффициент зависит от календарного месяца. СКО, вычисленная по нормированным таким образом отклонениям трехмесячного среднего значения прогнозируемого индекса от истинных трехмесячных средних, здесь и далее будет называться «стандартизованной СКО» (вслед за авторами работы [16]).

Вторая мера представляет собой обычный коэффициент корреляции истинных трехмесячных средних значений индекса с прогнозируемыми трехмесячными средни-

ми:

$$r_I(j) = \frac{\sum_{n=l}^{N-2-j} \Delta I_{n+j} \cdot \Delta \tilde{I}'_{n,j}}{\sqrt{\sum_{n=l}^{N-2-j} (\Delta I_{n+j})^2 \cdot \sum_{n=l}^{N-2-j} (\Delta \tilde{I}'_{n,j})^2}}. \quad (4.8)$$

Здесь $r_I(j)$ – значение данного коэффициента корреляции при дальности прогноза j месяцев; ΔI и $\Delta \tilde{I}'$ – отклонения трехмесячных средних значений индекса I и его прогноза \tilde{I}' от их средних за все время; остальные обозначения совпадают с введенными в (4.6).

В то время как СКО является количественной оценкой величины ошибки прогноза, мера, основанная на корреляции, отражает способность модели корректно предсказывать качественные характеристики индекса, такие как его знак, минимумы, максимумы, форма сигнала.

4.5.1 Прогноз индексов и событий Эль-Ниньо/Ла-Нинья

В настоящий момент известно, что ЭНЮК может наблюдаться как в центральном тихоокеанском (ЦТ) регионе, так и в восточном тихоокеанском (ВТ) регионе, демонстрируя, в зависимости от региона, различное поведение [46, 60, 83]. ЭНЮК в этих регионах можно характеризовать с помощью традиционных индексов Nino 4 (ЦТ регион) и Nino 3 (ВТ регион), определение которых дано в заголовке к рис. 4.4. В двух левых колонках рис. 4.5 показаны зависимости стандартизованной СКО и корреляций (см. вводную часть этого раздела) от дальности прогноза для моделей МЛДМ-ИНС и ЭОФ-ИНС. Видно, что ЭНЮК предсказывается обеими эмпирическими моделями немного лучше в ЦТ регионе (моделью МЛДМ-ИНС – вплоть до 11 месяцев), чем в ВТ регионе (моделью МЛДМ-ИНС – вплоть до 8 месяцев). Это может быть связано с более сильной зависимостью ТПО в ВТ регионе от параметров подповерхностных слоев воды и глубины термоклина, информация о которых плохо представлена в данных АТПО, используемых в данной главе, в то время как динамика ТПО в ЦТ регионе больше определяется параметрами атмосферы и воды на поверхности океана [60]. Данный результат также согласуется с пространственным распределением СКО в ЦТ и ВТ регионах на рис. 4.3.

В третьей колонке рис. 4.5 представлены аналогичные зависимости для индекса Nino 3.4 [120]. Данный индекс отражает поведение как в регионе Nino 3, так и в регионе Nino 4, и является сейчас общепринятым индикатором состояния ЭНЮК [15]. Естественным образом, прогностическая способность моделей для данного индекса чуть хуже, чем для индекса Nino 4, и чуть лучше, чем для индекса Nino 3 (для модели МЛДМ-ИНС горизонт прогноза составляет примерно 10 месяцев).

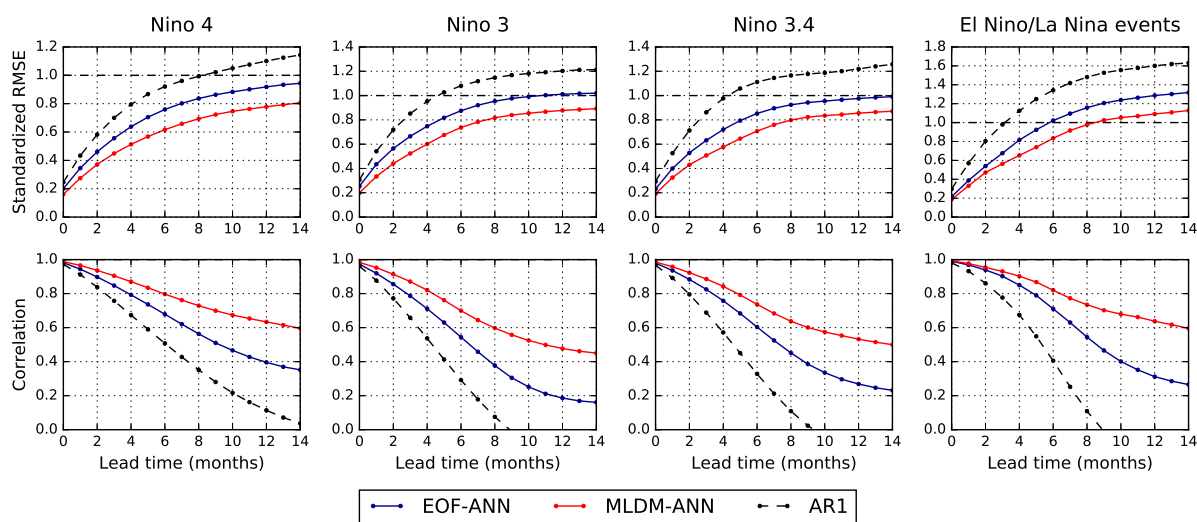


Рис. 4.5: Стандартизованная СКО (верхняя часть рисунка) и коэффициент корреляции прогноза и наблюдений (нижняя часть рисунка) в зависимости от дальности прогноза для моделей МЛДМ-ИНС, ЭОФ-ИНС и AR1, вычисленные для различных индексов на интервале обучения моделей: индекс Niño 4 (первый столбец), индекс Niño 3 (второй столбец), индекс Niño 3.4 (третий столбец), индекс Niño 3.4 во время событий Эль-Ниньо/Ла-Нинья (четвертый столбец, подробности см. в основном тексте).

Последняя колонка рис. 4.5 демонстрирует способность построенной модели предсказывать значения индекса Niño 3.4 в моменты времени, соответствующие событиям Эль-Ниньо и Ла-Нинья. Для определения этих моментов использовалось стандартное определение на основе индекса, представляющего собой сглаженный с окном 3 месяца индекс Niño 3.4, принятого Национальным управлением океанических и атмосферных исследований США (NOAA): Эль-Ниньо (или Ла-Нинья) регистрируется, если значение данного индекса в течение по крайней мере пяти последовательных месяцев было выше 0.5°C (или ниже -0.5°C). Таким образом, СКО и корреляции в этом случае подсчитывались по ансамблю из тех месяцев, когда регистрировалось либо событие Эль-Ниньо, либо событие Ла-Нинья. Видно, что количественный прогноз (СКО) ухудшился, по сравнению с прогнозом всех значений индекса Niño 3.4, в то время как качественный прогноз (корреляции), наоборот, немного улучшился, т.е. зависимость при больших значениях индекса предсказывается скорее качественно, чем количественно. Вывод о худшей количественной предсказуемости событий ЭНЮК (особенно, сильных) также подтверждается анализом концептуальных моделей [45].

Результаты, представленные на рис. 4.5, сложно сравнить напрямую с результатами

других моделей, поскольку различные исследователи используют различные базы данных и предварительную обработку. Однако грубое сравнение показывает, что эффективность моделей ЭОФ-ИНС и МЛДМ-ИНС близка к эффективности других эмпирических моделей, способных давать прогноз индексов Niño 3, Niño 3.4 и событий ЭНЮК [57, 67, 134] (см., например, аналогичные СКО и корреляции в этих работах с поправкой на отличие в определении дальности прогноза). Следовательно, поскольку в нашем случае модель МЛДМ-ИНС превосходит модель ЭОФ-ИНС, есть основания полагать, что и эффективность других эмпирических моделей в задаче прогноза индексов ЭНЮК также может повыситься при использовании переменных на основе метода МЛДМ вместо ЭОФ-разложения.

В заключение данного раздела, на рис. 4.6 показана зависимость эффективности модели МЛДМ-ИНС от календарного месяца при различных дальностях прогноза (на примере СКО и корреляции при прогнозе моделью индекса Niño 3.4). Из анализа

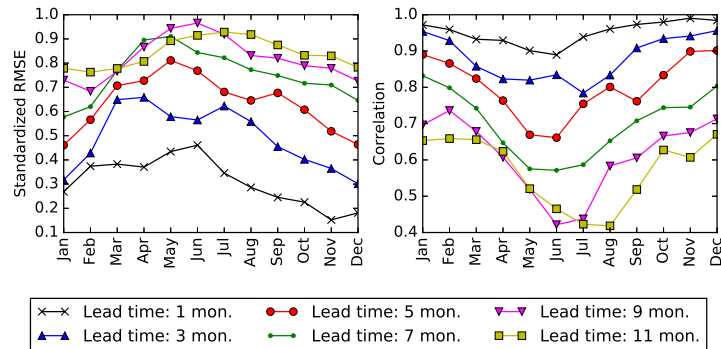


Рис. 4.6: Стандартизованная СКО (слева) и корреляция прогноза с наблюдениями (справа) при прогнозе индекса Niño 3.4 моделью МЛДМ-ИНС на обучающем интервале, построенные в зависимости от календарного месяца. Различные линии соответствуют различным значениям дальности прогноза (см. легенду).

обеих мер прогноза можно видеть, что прогноз значения индекса в весеннее и летнее время, когда событие ЭНЮК, как правило, начинает развиваться (либо не развиваться), является наиболее сложной задачей. Данное свойство хорошо известно и для других моделей ЭНЮК и называется «весенним браьером», потому что ухудшение прогностической способности наблюдается для прогнозов, стартовавших в первые месяцы календарного года. Например, значение коэффициента корреляции прогноза и наблюдений при дальности прогноза 5 месяцев (рис. 4.6) минимально в июне, что соответствует модели, стартовавшей в январе.

4.5.2 Прогностическая способность за пределами обучающей выборки

Несмотря на то, что модель МЛДМ-ИНС демонстрирует хорошую эффективность прогноза индексов на интервале обучения (1960-2007), не менее важным является анализ ее прогностической способности за пределами этого интервала, т.е. с 2008 по 2017 год. Во-первых, при таком анализе модель должна воспроизвести значения индексов, которые никак не задействовались в процедурах обучения и оптимизации. Во-вторых, обратим внимание, что временные ряды МЛДМ в конкретный момент времени определяются состояниями системы как в прошлом, так и в будущем относительно этого момента, и поэтому хуже определены на концах временного ряда (см. раздел 2.2.2). Данный «краевой эффект» не влияет на обучение модели, но может влиять на реальный прогноз, который стартует непосредственно с конца наблюдаемого интервала; этот эффект, очевидно, не учитывается при анализе СКО и корреляций прогнозов с наблюдениями на интервале обучения.

Для того, чтобы получить набор прогнозов для интервала с 2008 по 2017 год, было сделано следующее. Структурные параметры модели (размерность многообразия d , величина задержки l , количества нейронов m_f и m_g и др.) были взяты такими же, какие получились для интервала с 1960 по 2007 год длиной $N = 576$ месяцев. Далее, при этих структурных параметрах для каждого момента времени t_n ($n > N$) из проверочного интервала 2008-2017 была проделана полная процедура обучения модели, где в качестве наблюдаемых данных был взят временной ряд на сдвинутом интервале $[t_{n-N+1}, t_n]$, заканчивающемся моментом t_n и имеющем ту же длину N , т.е. были вычислены новые значения аномалий ТПО, получены обновленные ЭОФ/МЛДМ, обучена стохастическая модель оператора эволюции для них. Наконец, прогноз требуемого индекса I_{n+j} , соответствующий дальности прогноза j месяцев, получен с помощью запуска новой стохастической модели с момента времени t_n , как описано в разделе 4.2.3. Заметим, что сохранение одной и той же длины обучающего интервала N , во-первых, лучше соответствует предположению о сохраняющихся оптимальных структурных параметрах модели, и, во-вторых, сохраняет одну и ту же степень нестационарности в данных, которая *a priori* предполагается слабой в модели (раздел 1.4).

Данная схема, не задействующая прогнозируемые значения при обучении модели, была применена к проверочному интервалу с 2008 по 2017 год. Полученный в результате набор предсказанных значений индекса Nino 3.4 изображен на рис. 4.7 для различных дальностей прогноза. Можно видеть, что соответствие предсказанных и наблюдаемых значений ухудшается с ростом дальности прогноза. Важно отметить,

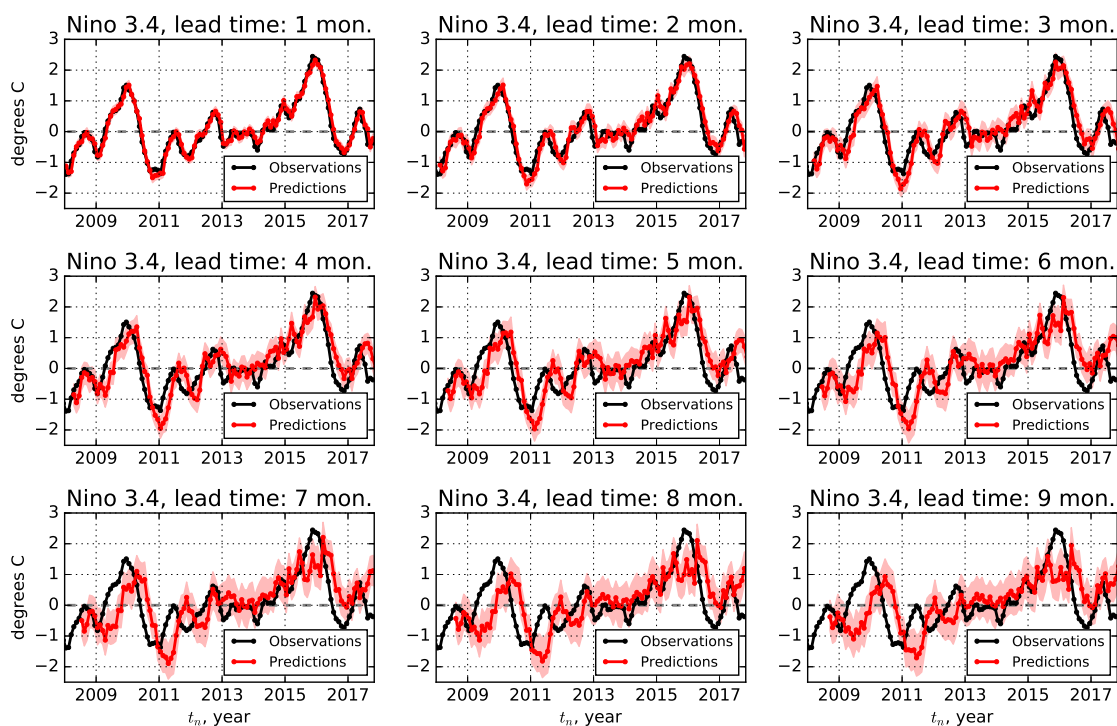


Рис. 4.7: Сравнение прогнозов индекса Nino 3.4 (красный цвет) с наблюдениями (черный цвет) на проверочном интервале с 2008 по 2017 год по описанной в тексте схеме, не задействующей прогнозируемые значения индексов при обучении модели. На разных рисунках показаны предсказанные значения при дальности прогноза от 1 до 8 месяцев (см. заголовки рисунков). Доверительный интервал величиной 65%, даваемый моделью, закрашен тенью красного цвета (он вычислен как диапазон значений между 0.175 и 0.825 квантилями ансамбля стохастических реализаций модели).

что, помимо медианных значений (4.5), построенная стохастическая модель способна давать доверительный интервал прогноза, который также представлен на рис. 4.7 (величина интервала составляет 65%). Приведенный рисунок подтверждает, что в построенной модели, как и в других моделях ЭНЮК, все еще присутствует весенний барьер, например, для Эль-Ниньо 2010 года он становится видимым при дальности 3-5 месяцев (отрезок кривой, соединяющей предсказанные значения, сдвинут «в будущее» по отношению к наблюдаемым значениям индекса, соответствующим этому событию Эль-Ниньо).

На рис. 4.8 красными квадратными маркерами показаны графики СКО и корреляции наблюдений с прогнозами модели МЛДМ-ИНС для проверочного интервала с 2008 по 2017 год. Для сравнения, те же самые меры приведены для прогнозов моделей ЭОФ-ИНС и AR1, сделанных по аналогичной схеме на том же проверочном

интервале (квадратные маркеры). Естественным образом, эффективность моделей на проверочном интервале ухудшилась, по сравнению с эффективностью на обучающем интервале (рис. 4.5); но все же и в данном случае модель МЛДМ-ИНС остается немного более эффективной, чем модель ЭОФ-ИНС, несмотря на принятое во внимание возможное влияние краевого эффекта МЛДМ.

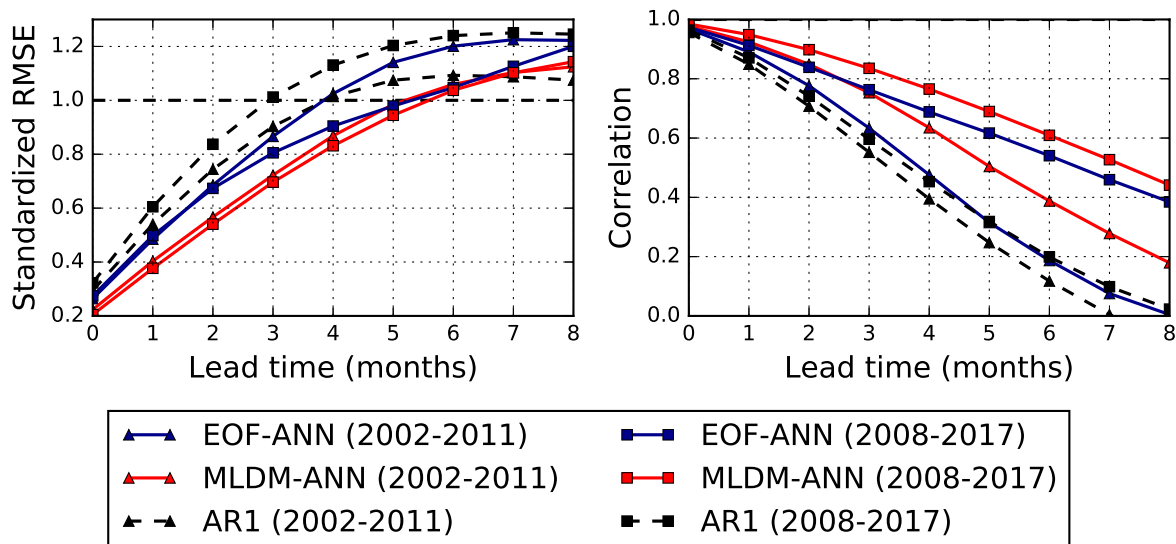


Рис. 4.8: Стандартизованная СКО (слева) и корреляция наблюдаемых и предсказанных значений индекса Nino 3.4 (справа) на проверочных интервалах 2008-2017 и 2002-2011 с помощью описанной в тексте схемы прогноза. Красным цветом показаны результаты для модели МЛДМ-ИНС, синим – для модели ЭОФ-ИНС, черным – для авторегрессионной модели AR1.

Для более прямого и однозначного сравнения с ансамблем моделей «ENSO Prediction Plume», анализируемым подробно в работе [16] на интервале с 2002 по 2011 год, на рис. 4.8 треугольными маркерами показаны также аналогичные СКО и коэффициенты корреляции для моделей МЛДМ-ИНС, ЭОФ-ИНС, AR1 на этом интервале (2002-2011). Чтобы сделать прогнозы на данном интервале без вовлечения прогнозируемых значений в процедуру обучения, были привлечены дополнительные данные ТПО из базы NOAA_ERSST_V4 с января 1954 по декабрь 1959. Визуальное сравнение полученных СКО и корреляций (рис. 4.8) с такими же СКО и корреляциями для ансамбля моделей ЭНЮК (рисунки 6, 9 в работе [16]) позволяет сказать, что результаты модели МЛДМ-ИНС находятся в пределах разброса результатов моделей из этого ансамбля с достаточно хорошей эффективностью. Также интересно отметить, что корреляции прогнозов с наблюдениями для моделей МЛДМ-ИНС и ЭОФ-ИНС

выше для интервала 2008-2017, чем для интервала 2002-2011. С одной стороны, это может означать, что интервал 2002-2011 сложнее для прогноза, чем интервал 2008-2017, но, с другой стороны, следует иметь в виду, что длина интервала 10 лет является не очень длинной для «честной» оценки прогностических способностей (за 10 лет, как правило, может произойти лишь 2-3 события ЭНЮК), и поэтому такая оценка имеет, вообще говоря, достаточно сомнительную статистическую значимость; наблюдаемое отличие графиков для интервалов 2002-2011 и 2008-2017 на рис. 4.8 может давать грубое представление об уровне этой значимости.

4.6 Заключение

В данной главе был предложен метод построения эмпирической прогностической модели для данных температуры поверхности океана, состоящий в комбинации двух отдельных шагов: эмпирической редукции размерности данных с помощью поиска главного многообразия и построении эмпирической стохастической модели оператора эволюции. Первый шаг основан на многомерном обобщении ЛДМ-разложения, представленного в главе 2. С одной стороны, многомерные ЛДМ представляют собой лишь частный случай многомерных нелинейных динамических мод, представленных в главе 3; с другой стороны, они тоже позволяют выявлять доминирующие в динамике системы временные масштабы и описывать их, но при гораздо меньшей ресурсоемкости, чем многомерные нелинейные динамические моды. Второй шаг метода основан на изложенном в главе 1 методе построения модифицированной стохастической модели оператора эволюции при анализе пространственно распределенных данных.

Метод применен к аномалиям ТПО в тропиках, где доминирующим режимом межгодовой изменчивости является ЭНЮК. При различных размерностях главного многообразия (от 1 до 6) показано, что МЛДМ предоставляют лучшие переменные для построения эмпирической модели оператора эволюции и краткосрочного прогноза траектории, чем традиционное ЭОФ-разложение, используемое многими прогностическими эмпирическими климатическими моделями. Во-первых, данный результат говорит о том, что есть потенциальная возможность улучшить другие эмпирические модели ЭНЮК с помощью использования в них МЛДМ. Во-вторых, можно ожидать, что другие методы, более ориентированные на выделение временных масштабов (например, многоканальный сингулярный спектральный анализ и др., см. Введение) также могут быть полезны для выбора переменных при построении модели.

Паттерны, наилучшим образом предсказываемые построенной моделью ТПО, соответствуют явлению ЭНЮК, что означает, что МЛДМ предоставляют хорошие переменные для прогноза динамики ЭНЮК. Действительно, модель демонстрирует

способность предсказывать различные индексы ЭНЮК, а также формально определенные события Эль-Ниньо и Ла-Нинья. По результатам анализа двух десятилетних проверочных интервалов можно заключить, что по прогностической способности модель не уступает существующим лучшим мировым моделям ЭНЮК (как эмпирическим моделям, так и моделям из первых принципов). Отметим, что данный результат явился основанием для включения в июне 2019 года разработанной модели (с названием «IAP-NN») в список ведущих моделей для прогноза Эль-Ниньо, прогноз которых ежемесячно публикуется институтом физики Земли при Колумбийском университете.

Стоит отметить, что в настоящем исследовании не была рассмотрена зависимость качества прогноза ЭНЮК от используемых данных. В частности, данные ТПО могли быть взяты в другом пространственном регионе, например, данные ТПО в Тихом океане, которые могли бы лучше учесть связь ЭНЮК с ТДК, или глобальные данные ТПО. Возможно также рассмотрение влияния других переменных на качество прогноза ЭНЮК, таких как температура подповерхностных слоев воды или глубина термоклина, которые могли бы улучшить прогноз ЭНЮК в ВТ регионе, или характеристики атмосферы, которые могли бы улучшить прогноз ЭНЮК в ЦТ регионе.

Другое возможное направление будущих исследований может быть связано с дальнейшим расширением метода. Во-первых, в данной работе исследование было ограничено рассмотрением линейных динамических мод вместо более общего нелинейного разложения ввиду ограниченности вычислительных ресурсов. Однако ограничения связаны, главным образом, с полиномиальной формой параметризации нелинейных динамических мод; использование другого типа параметризации, более удобного при больших размерностях многообразий, например, искусственных нейронных сетей, могло бы решить данную проблему. Кроме того, существующее разделение шагов эмпирической редукции размерности и построения оператора эволюции может приводить к все еще не оптимальным переменным для построения модели. Объединение этих двух шагов в рамках одной байесовой ценовой функции, несомненно, могло бы лучше решить задачу. Несмотря на то, что реализация такого объединения является более ресурсоемкой задачей, она может также в скором времени привести к успеху ввиду продолжающегося роста вычислительных мощностей.

Заключение

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

Основные результаты.

1. Разработан и реализован байесов подход к эмпирической реконструкции оператора эволюции в форме стохастической модели, применимый к пространственно распределенным данным и учитывающий такие особенности, как ненулевой временной масштаб автокорреляции временных рядов переменных, получаемых в результате применения методов редукции размерности к данным, а также возможные присутствие в системе медленной нестационарности и зависимость системы от внешних воздействий. Показано, что учет перечисленных особенностей улучшает байесову обоснованность модели. На примере временного ряда, сгенерированного глобальной климатической моделью ИВМ РАН, продемонстрирована способность разработанной эмпирической модели как давать краткосрочный прогноз фазовой траектории, так и качественно воспроизводить режим поведения системы.

2. Разработан и реализован новый метод эмпирической редукции размерности, позволяющий с помощью байесова подхода представить пространственно распределенные данные в виде суперпозиции нелинейных динамических мод, каждая из которых представляет собой результат нелинейного преобразования одномерной скрытой переменной и учитывает как наличие нелинейных связей между значениями, наблюдаемыми в различных пространственно разнесенных точках, так и наличие причинно-следственных связей в системе. Показано, что использование нелинейных динамических мод приводит к лучшему выделению главных пространственно-временных структур и временных масштабов, доминирующих в динамике системы, чем использование эмпирических ортогональных функций.

3. На основе ежемесячных данных температуры поверхности океана с 1981 по 2014 год выделено три статистически значимых нелинейных моды изменчивости современного климата, которые описывают годовую изменчивость климатической системы, а также значительную часть тихоокеанской и атлантической динамики, включая Эль-Ниньо Южное Колебание. Кроме того, с помощью нелинейных динамических мод

подтверждено наличие климатического сдвига, ассоциирующегося с наступлением холодной фазы Тихоокеанского Декадного Колебания в 1997-1998 годах.

4. На основе данных глобальной климатической модели ИВМ РАН выделено пять динамических мод доиндустриального климата, соответствующего концу XIX века, из которых первая является нелинейной и описывает годовую изменчивость системы, а остальные четыре являются линейными и описывают, главным образом, межгодовую и декадную изменчивость Эль-Ниньо Южного Колебания и Тихоокеанского Декадного Колебания.

5. Создан метод поиска многомерных нелинейных динамических мод, являющихся обобщением одномерных нелинейных динамических мод. Показана лучшая адекватность многомерных нелинейных динамических мод, по сравнению с набором одномерных нелинейных динамических мод, в задаче поиска скрытых переменных, описывающих наблюдаемую систему.

6. В результате применения двумерных нелинейных динамических мод к данным реанализа температуры поверхности океана с 1870 по 2014 год подтверждено наличие четырех климатических сдвигов, ассоциирующихся со сменами фазы Тихоокеанского Декадного Колебания, а также показано, что структура дальних связей Эль-Ниньо Южного Колебания, детектируемая этими модами, будучи линейной в конце XIX века, стала существенно нелинейной, начиная с 60-х годов XX века.

7. На основе эмпирической реконструкции динамики температуры поверхности океана в тропическом поясе Земли с помощью многомерных линейных динамических мод создана эмпирическая прогностическая модель индексов Эль-Ниньо Южного колебания. Показано, что по прогностической способности данная модель не уступает лучшим существующим в мире моделям Эль-Ниньо Южного Колебания. Это явилось основанием для включения разработанной модели в список ведущих моделей Эль-Ниньо, прогноз которых ежемесячно публикуется институтом Земли при Колумбийском университете.

Положения, выносимые на защиту.

1. Учет ненулевого временного масштаба автокорреляции переменных, получаемых в результате эмпирической редукции размерности пространственно распределенных данных, при построении по этим переменным стохастической модели оператора эволюции, улучшает байесову обоснованность данной модели. Модифицированная таким образом модель способна как давать краткосрочный прогноз фазовой траектории, так и качественно воспроизводить режим поведения исходной системы.

2. Использование разработанного метода разложения пространственно распределенных данных на нелинейные динамические моды, учитывающие как наличие

нелинейных связей между значениями, наблюдаемыми в различных пространственно разнесенных точках, так и наличие причинно-следственных связей в системе, приводит к эффективному выделению главных пространственно-временных структур и, в том числе, временных масштабов, доминирующих в динамике исследуемой системы.

3. Найденные три одномерные нелинейные моды изменчивости температуры поверхности океана в 1981-2014 гг. описывают годовую изменчивость климатической системы и значительную часть тихоокеанской и атлантической динамики, включая Эль-Ниньо Южное Колебание, а также подтверждают наличие климатического сдвига, ассоциирующегося с наступлением холодной фазы Тихоокеанского Декадного Колебания в 1997-1998 годах.

4. Из пяти найденных динамических мод доиндустриального климата, соответствующего концу XIX века, в модели ИВМ РАН первая является нелинейной и описывает годовую изменчивость системы, а остальные четыре являются линейными и описывают, главным образом, межгодовую и декадную изменчивость Эль-Ниньо Южного Колебания и Тихоокеанского Декадного Колебания.

5. Структура найденных двумерных нелинейных динамических мод температуры поверхности океана подтверждает наличие четырех климатических сдвигов, ассоциирующихся со сменами фазы Тихоокеанского Декадного Колебания. Кроме того, структура дальних связей Эль-Ниньо Южного Колебания, детектируемая этими модами, будучи линейной в конце XIX века, становится существенно нелинейной, начиная с 60-х годов XX века.

6. Прогностическая способность созданной по данным температуры поверхности океана в тропиках эмпирической модели в отношении индексов Эль-Ниньо Южного Колебания не уступает прогностической способности лучших существующих в мире моделей, предсказывающих эти индексы.

Научная новизна работы заключается в том, что были разработаны (1) новый метод эмпирической реконструкции оператора эволюции системы, применимый к пространственно распределенным данным и позволяющий учесть их такие особенности, как гладкость временных рядов переменных, получаемых в результате применения методов редукции размерности к данным, а также возможные присутствие в системе медленной нестационарности и зависимость системы от внешних воздействий (см. результат 1), и (2) новый метод эмпирической редукции размерности пространственно распределенных данных, заключающийся в поиске нелинейных динамических мод и учитывающий как наличие нелинейных связей между значениями, наблюдаемыми в различных пространственно разнесенных точках, так и наличие причинно-следственных связей в исследуемой системе (см. результаты 2 и 5). Кроме

того, указанные методы впервые были применены к реальным данным измерений.

Практическая ценность работы состоит в том, что разработанные методы были применены к исследованию климатической системы Земли на интервале с 1870 по 2014 год и дали объективную информацию о ее устройстве (см. результаты 3, 4, 6). Кроме того, разработанные методы позволили создать эмпирическую прогнозную модель явления Эль-Ниньо, существенно влияющего на режимы погоды (в том числе, экстремальные) в различных регионах земного шара. Разработанные методы также могут применяться как к исследованию климатической системы на других пространственных и временных масштабах, так и к анализу сложных систем в других областях науки (в качестве примера см. работы с участием автора [64–66,88]).

Автор выражает искреннюю благодарность своему научному руководителю, Александру Марковичу Фейгину, без чуткого и грамотного руководства которого данная работа бы не появилась на свет, своим ближайшим коллегам и соавторам – Дмитрию Мухину, Евгению Лоскутову, Алексею Селезневу – за плодотворное сотрудничество и активные дискуссии, Евгению Михайловичу Володину за предоставление доступа к данным модели ИВМ РАН и обсуждение полученных результатов, Юргену Куртцу, Сергею Кравцову, Дмитрию Кондрашову за поддержку и ценные советы, а также сотрудникам отдела 240 ИПФ РАН за доброе отношение и хорошую рабочую атмосферу.

Список публикаций по теме диссертации

- [A1] *Mukhin D.N., Kondrashov D.A., Loskutov E.M., Gavrilov A.S., Feigin A.M., Ghil M.* Predicting Critical Transitions in ENSO models. Part II: Spatially Dependent Models // *Journal of Climate*. — 2015. — mar. — Vol. 28, no. 5. — Pp. 1962–1976.
- [A2] *Mukhin D.N., Gavrilov A.S., Feigin A.M., Loskutov E.M., Kurths J.* Principal nonlinear dynamical modes of climate variability // *Scientific Reports*. — 2015. — oct. — Vol. 5. — P. 15510.
- [A3] *Gavrilov A.S., Mukhin D.N., Loskutov E.M., Volodin E.M., Feigin A.M., Kurths J.* Method for reconstructing nonlinear modes with adaptive structure from multidimensional data // *Chaos: An Interdisciplinary Journal of Nonlinear Science*. — 2016. — dec. — Vol. 26, no. 12. — P. 123101.
- [A4] *Gavrilov A.S., Loskutov E.M., Mukhin D.N.* Bayesian optimization of empirical model with state-dependent stochastic forcing // *Chaos, Solitons and Fractals*. — 2017. — Vol. 104.
- [A5] *Mukhin D.N., Gavrilov A.S., Loskutov E.M., Feigin A.M., Kurths J.* Nonlinear reconstruction of global climate leading modes on decadal scales // *Climate Dynamics*. — 2018. — sep. — Vol. 51, no. 5-6. — Pp. 2301–2310.
- [A6] *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.
- [A7] *Гаврилов А.С., Мухин Д.Н., Лоскутов Е.М., Беликлович М.В., Караитин Д.А.* Программа для расчета нелинейной динамической моды и ее обоснованности по многомерному временному ряду (версия 1). Свидетельство о государственной

регистрации программ для ЭВМ №2015614269. Зарегистрировано в реестре программ для ЭВМ 13 апреля 2015 г.

- [A8] *Гаврилов А.С., Мухин Д.Н., Лоскутов Е.М.* Программа для расчета стохастической модели оператора эволюции на базе искусственных нейронных сетей, ее обоснованности и прогноза поведения по скалярному временному ряду. Свидетельство о государственной регистрации программ для ЭВМ №2015661379. Зарегистрировано в реестре программ для ЭВМ 27 октября 2015 г.
- [A9] *Гаврилов А.С., Мухин Д.Н., Лоскутов Е.М.* Программа для расчета нелинейной динамической моды и ее обоснованности по многомерному временному ряду с оптимизацией по размерности. Свидетельство о государственной регистрации программ для ЭВМ №2015661380. Зарегистрировано в реестре программ для ЭВМ 27 октября 2015 г.
- [A10] *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.
- [A11] *Гаврилов А.С., Мухин Д.Н., Фейгин А.М.* Байесов подход к реконструкции климатических систем по временным рядам на примере модели явления Эль-Ниньо // Состав атмосферы. Атмосферное электричество. Климатические эффекты: труды XIV Всероссийской школы конференции молодых ученых. — 2010.
- [A12] *Гаврилов А.С., Мухин Д.Н., Лоскутов Е.М.* Применение стохастического моделирования для реконструкции по временным рядам модели явления Эль-Ниньо средней сложности // Состав атмосферы. Атмосферное электричество. Климатические эффекты : труды XVI Международной школы-конференции молодых ученых. — 2012. — Pp. 92–95.
- [A13] *Gavrilov A.S., Mukhin D.N., Loskutov E.M., Feigin A.M.* Decomposition of complex systems: nonlinear mode extraction // Proceedings: Topical problems of nonlinear wave physics (NWP-2014). — 2014. — Pp. 156–157.
- [A14] *Loskutov E.M., Mukhin D.N., Gavrilov A.S., Feigin A.M.* Constructing an embedding for reduced dynamical models of climate variability // Proceedings: Frontiers of nonlinear physics (FNP-2016). — 2016. — P. 179.

- [A15] *Gavrilov A.S., Mukhin D.N., Loskutov E.M., Feigin A.M.* Multidimensional nonlinear dynamical modes expansion of spatially distributed time series // Proceedings: Frontiers of nonlinear physics (FNP-2016). — 2016. — Pp. 213–214.
- [A16] *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.
- [A17] *Гаврилов А.С., Лоскутов Е.М., Мухин Д.Н., Фейгин А.М.* Использование стохастических моделей для реконструкции высокоразмерных систем и прогноза качественных изменений // Тезисы 9-й Международной школы <<Хаотические автоколебания и образование структур>> (ХАОС-2010). — 2010.
- [A18] *Mukhin D.N., Gavrilov A.S., Loskutov E.M., Feigin A.M.* Reconstruction of a dynamical system underlying an observed time series by optimal stochastic models // Abstract EGU2011-951, European Geosciences Union General Assembly, 3-8 April 2011, Vienna, Austria. — 2011.
- [A19] *Gavrilov A.S., Mukhin D.N., Loskutov E.M., Feigin A.M.* Stochastic approach to reconstruction of dynamical systems: optimal model selection criterion // Abstract NG51E-1678, AGU Fall Meeting, San-Francisco, California, 5-9 December 2011. — 2011.
- [A20] *Гаврилов А.С., Мухин Д.Н., Фейгин А.М.* Стохастический подход к прогнозу эволюции климатических систем по временным рядам: приложение к динамике явления Эль-Ниньо // Тезисы XV Всероссийской школы-конференции молодых ученых <<Состав атмосферы. Атмосферное электричество. Климатические процессы>> (САТЭП-2011), 30 мая - 4 июня 2011, Борок. — 2011.
- [A21] *Mukhin D.N., Mukhina A.Yu., Gavrilov A.S., Loskutov E.M.* ANN-based empirical modeling of SST variability in Equatorial Pacific region // Abstract NG43B-1579, AGU Fall Meeting, San-Francisco, California, 3-7 December 2012. — 2012.
- [A22] *Gavrilov A.S., Mukhin D.N., Loskutov E.M., Feigin A.M.* Bayesian evidence technique in application to stochastic modeling of dynamic systems // Abstract P8, The XXXII Dynamic Days Europe, Gothenburg, Sweden, 2012. — 2012.
- [A23] *Гаврилов А.С., Мухин Д.Н., Лоскутов Е.М., Фейгин А.М.* Стохастический подход к реконструкции динамических систем: критерий выбора оптимальной модели // Тезисы научной школы <<Нелинейные волны — 2012>>. — 2012.

- [A24] *Feigin A.M., Mukhin D.N., Mukhina A.Yu., Gavrilov A.S., Loskutov E.M.* Empirical modeling of climate dynamics // Abstract EGU2013-5966, European Geosciences Union General Assembly, 7-12 April 2013, Vienna, Austria. — 2013.
- [A25] *Mukhin D.N., Loskutov E.M., Gavrilov A.S., Feigin A.M.* Empirical prediction of climate dynamics: optimal models, derived from time series // Abstract NG33A-1575, AGU Fall meeting, San-Francisco, 9-13 December 2013. — 2013.
- [A26] *Мухин Д.Н., Гаврилов А.С., Лоскутов Е.М., Фейгин А.М.* Эмпирические прогностические модели климата // Тезисы X Международной школы-конференции <<Хаотические автоколебания и образование структур>> (ХАОС-2013), 7-12 октября 2013, Саратов. — 2013. — P. 114.
- [A27] *Gavrilov A.S., Mukhin D.N., Loskutov E.M., Feigin A.M.* Empirical ANN-based prognostic models for climate modes // Abstract EGU2014-202, European Geosciences Union General Assembly, 27 April - 2 May 2014, Vienna, Austria. — 2014.
- [A28] *Mukhin D.N., Gavrilov A.S., Loskutov E.M., Feigin A.M.* Optimal prognostic models from spatially distributed time series // Abstract for Topical problems of nonlinear wave physics (NWP-2014), 17-23 July 2014, Russia, Nizhny Novgorod. — 2014.
- [A29] *Loskutov E.M., Mukhin D.N., Gavrilov A.S., Feigin A.M.* Optimal Empirical Prognostic Models of Climate Dynamics // Abstract NG43A-3763, AGU Fall meeting, San-Francisco, 15-19 December 2014. — 2014.
- [A30] *Feigin A.M., Mukhin D.N., Loskutov E.M., Gavrilov A.S.* Empirical prognostic models from spatially distributed time series: shot-term forecast and long-term prognosis // Abstract: International work-conference on Time Series (ITISE-2014). — 2014.
- [A31] *Loskutov E.M., Mukhin D.N., Gavrilov A.S., Feigin A.M.* Empirical forecast of interannual climate variability. // Abstract for 31st IUGG Conference on Mathematical Geophysics, Paris, France, 6-10 June 2016. — 2016.
- [A32] *Feigin A.M., Gavrilov A.S., Loskutov E.M., Mukhin D.N.* Empirical reconstruction of complex systems: prognostic models of evolution operator, optimal pre-processing of highdimensional data, applications to climate // Abstract for 31st IUGG Conference on Mathematical Geophysics, Paris, France, 6-10 June 2016. — 2016.
- [A33] *Feigin A.M., Mukhin D.N., Gavrilov A.S., Loskutov E.M.* Nonlinear dynamical models of interannual to decadal climate variability from observed time series: data

pre-processing, model construction, forecast of climate indices // Abstract for CLIVAR Open Science Conference, Qingdao, China, 18-25 September 2016. — 2016.

- [A34] *Гаврилов А.С., Мухин Д.Н., Лоскутов Е.М., Фейгин А.М.* Построение вложения для эмпирических прогностических климатических моделей // Тезисы 20-й школы-конференции молодых ученых <<Состав Атмосферы. Атмосферное электричество. Климатические процессы>>, Нижний Новгород, 24-26 мая 2016. — 2016.
- [A35] *Mukhin D.N., Gavrilov A.S., Loskutov E.M., Feigin A.M.* Empirical prognostic models from spatially distributed time series: construction of nonlinear principal modes // Abstract: International work-conference on Time Series (ITISE-2014). — 2014.
- [A36] *Feigin A.M., Mukhin D.N., Loskutov E.M., Gavrilov A.S.* Global empirical reconstruction of complex systems: general approach and application to climate modeling // Abstract for Topical problems of nonlinear wave physics (NWP-2014), Russia, Nizhny Novgorod, 17-23 July 2014. — 2014.
- [A37] *Gavrilov A.S., Mukhin D.N., Loskutov E.M., Feigin A.M.* Construction of nonlinear dynamic modes by climate data // Abstract for Science of the future. 17-20 September 2014, Saint Petersburg, Russia. — 2014.
- [A38] *Гаврилов А.С., Мухин Д.Н., Лоскутов Е.М., Фейгин А.М.* Нелинейные климатические моды: постановка задачи и первые результаты // Тезисы XVIII Всероссийской школы-конференции молодых ученых <<Состав атмосферы. Атмосферное электричество. Климатические процессы>>, 29 сентября - 3 октября 2014, Борок. — 2014.
- [A39] *Gavrilov A.S.* Nonlinear modes decomposition of observed multivariate data // Abstract for workshop <<Network analysis and data driven modelling of the climate>>, 27-29 October 2014, Potsdam, Germany. — 2014.
- [A40] *Mukhin D.N., Gavrilov A.S., Loskutov E.M., Feigin A.M.* Nonlinear Decomposition of Climate Data: a New Method for Reconstruction of Dynamical Modes // Abstract NG43A-3752, AGU Fall meeting, San-Francisco, 15-19 December 2014. — 2014.
- [A41] *Gavrilov A.S., Mukhin D.N., Loskutov E.M., Feigin A.M., Kurths J.* Empirical decomposition of climate data into nonlinear dynamic modes // Abstract EGU2015-627, European Geosciences Union General Assembly, 12-17 April 2015, Vienna, Austria. — 2015.

- [A42] *Feigin A.M., Gavrilov A.S., Loskutov E.M., Mukhin D.N.* Decomposition of the complex system into nonlinear spatio-temporal modes: algorithm and application to climate data mining // Abstract EGU2015-6217 , European Geosciences Union General Assembly, 12-17 April 2015, Vienna, Austria. — 2015.
- [A43] *Mukhin D.N., Gavrilov A.S., Loskutov E.M., Feigin A.M., Kurths J.* Reconstruction of principal dynamical modes from climatic variability: nonlinear approach // Abstract EGU2015-5729, European Geosciences Union General Assembly, 12-17 April 2015, Vienna, Austria. — 2015.
- [A44] *Гаврилов А.С., Мухин Д.Н., Лоскутов Е.М., Фейгин А.М.* Анализ нелинейных мод поверхностной температуры океана // Тезисы 19-й Международной школы-конференции молодых ученых <<Состав атмосферы. Атмосферное электричество. Климатические процессы>>, 25 - 29 мая 2015, Шепси, Россия. — 2015.
- [A45] *Gavrilov A.S., Mukhin D.N., Loskutov E.M., Feigin A.M., Kurths J.* Construction of nonlinear dynamical modes: new algorithm // Abstract for International Workshop <<Analysis of dynamic networks and data driven modelling of the climate>>, 12-14 October 2015, Potsdam, Germany. — 2015.
- [A46] *Mukhin D.N., Gavrilov A.S., Loskutov E.M., Feigin A.M., Kurths J.* Nonlinear empirical reduction of climate system dimensionality // Abstract for Challenges in Data Science: a complex systems perspective, International Conference, Torino, Italy, October 14-17 2015. — 2015.
- [A47] *Loskutov E.M., Mukhin D.N., Gavrilov A.S., Feigin A.M.* Nonlinear Empirical Decomposition of Space-Time Distributed Data // Abstract NG23B-1798, AGU Fall meeting, San-Francisco, 14-18 December 2015. — 2015.
- [A48] *Mukhin D.N., Gavrilov A.S., Loskutov E.M., Feigin A.M., Kurths J.* Extraction of Nonlinear Dynamical Modes Underlying Climate Variability // Abstract NG23B-1803, AGU Fall meeting, San-Francisco, 14-18 December 2015. — 2015.
- [A49] *Mukhin D.N., Gavrilov A.S., Loskutov E.M., Feigin A.M.* Extraction of leading nonlinear dynamical modes of climate from data // Abstract: Topical Problems of Nonlinear Wave Physics (NWP-2017). — 2017.
- [A50] *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.

- [A51] *Гаврилов А.С., Мухин Д.Н., Лоскутов Е.М., Фейгин А.М.* Построение многомерных нелинейных динамических мод пространственно-распределенных систем // Тезисы научной школы <<Нелинейные волны – 2016>>, Нижний Новгород, 27 февраля – 4 марта 2016. — 2016.
- [A52] *Gavrilov A.S., Mukhin D.N., Loskutov E.M., Feigin A.M.* Nonlinear dynamical modes of climate variability: from curves to manifolds // Abstract EGU2016-651. European Geosciences Union General Assembly, 17-22 April 2016, Vienna, Austria. — 2016.
- [A53] *Gavrilov A.S.* Evolution of principal nonlinear patterns of global sea surface temperature in XX century // Abstract for Science of the future, Kazan, Russia, 20-23 September 2016. — 2016.
- [A54] *Mukhin D.N., Gavrilov A.S., Loskutov E.M., Feigin A.M., Kurths J.* Principal nonlinear dynamical modes of multidecadal climate variability // Abstract for CLIVAR Open Science Conference, Qingdao, China, 18-25 September 2016. — 2016.
- [A55] *Mukhin D.N., Gavrilov A.S., Loskutov E.M., Feigin A.M., Kurths J.* Leading Nonlinear Mode of Inter-Annual and Decadal Climate Behavior // Abstract NG24A-03, AGU Fall meeting, San-Francisco, 12-16 December 2016. — 2016.
- [A56] *Mukhin D.N., Gavrilov A.S., Loskutov E.M., Feigin A.M.* Detection of principal dynamical modes of changing climate // Abstract for JpGU-AGU Joint Meeting 2017, Makuhari Messe, Japan, 20-25 May 2017. — 2017.
- [A57] *Гаврилов А.С., Мухин Д.Н., Лоскутов Е.М., Фейгин А.М.* Анализ многолетней климатической изменчивости с помощью метода нелинейных мод // Тезисы XXI Всероссийской школы-конференции молодых ученых <<Состав атмосферы. Атмосферное электричество. Климатические процессы>>, 6-10 июня 2017, Борок. — 2017.
- [A58] *Mukhin D.N., Gavrilov A.S., Loskutov E.M., Feigin A.M., Kurths J.* Extracting Leading Nonlinear Modes of Changing Climate From Global SST Time Series // Abstract NG41A-0121, AGU Fall meeting, New Orleans, 11-15 December 2017. — 2017.

- [A59] *Mukhin D.N., Loskutov E.M., Gavrilov A.S., Mukhina A.Yu.* ENSO prediction by the analysis of observed variability of SST // Abstract MO-11_A3.1, Davos Atmosphere and Cryosphere Assembly DACA-13, 8-12 July, 2013, Davos, Switzerland. — 2013. — P. 85.
- [A60] *Seleznev A.F., Gavrilov A.S., Mukhin D.N.* ANN-based empirical model for ENSO forecast // Abstract for Science of the future, Kazan, Russia, 20-23 September 2016. — 2016.
- [A61] *Gavrilov A.S., Mukhin D.N., Loskutov E.M., Feigin A.M.* Construction of Optimally Reduced Empirical Model by Spatially Distributed Climate Data // Abstract NG31A-1824, AGU Fall meeting, San-Francisco, 12-16 December 2016. — 2016.
- [A62] *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.
- [A63] *Gavrilov A.S., Mukhin D.N., Loskutov E.M., Feigin A.M.* Reduced nonlinear prognostic model construction from high-dimensional data // Abstract EGU2017-1040. European Geosciences Union General Assembly, 23-28 April 2017, Vienna, Austria. — 2017.
- [A64] *Feigin A.M., Mukhin D.N., Gavrilov A.S., Seleznev A.F., Loskutov E.M.* Nonlinear Dynamical Modeling and Forecast of ENSO Variability // Abstract EGU2017-6689. European Geosciences Union General Assembly, 23-28 April 2017, Vienna, Austria. — 2017.
- [A65] *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.
- [A66] *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.
- [A67] *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.

- [A68] *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.
- [A69] *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.
- [A70] *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.
- [A71] *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. — Рр. 27–29.

Литература

- [1] *Лоскутов Е.М., Мольков Я.И., Мухин Д.Н., Фейгин А.М.* Прогноз качественного поведения динамической системы по хаотическому временному ряду // *Известия ВУЗов. Радиофизика.* — 2001. — Vol. 44, no. 5-6. — P. 376.
- [2] *Анищенко В.С., Павлов А.Н., Янсон Н.Б.* Реконструкция динамических систем в приложении к решению задачи защиты информации // *ЖТФ.* — 1998. — Vol. 68, no. 12. — P. 1.
- [3] *Павлов А.Н., Янсон Н.Б., Анищенко В.С.* Реконструкция динамических систем // *Радиотехника и электроника.* — 1999. — Vol. 44, no. 9. — P. 1075.
- [4] *Андронов А.А., Понтрягин Л.С.* Грубые системы // *Доклады Академии Наук СССР.* — 1937. — Vol. 14, no. 5. — Pp. 247–250.
- [5] *Дмитриев А.С., Панас А.И.* Динамический хаос: новые носители информации для систем связи. — М.: Физматлит, 2002. — P. 252.
- [6] *Анищенко В.С., Вадивасова Т.Е. Астахов В.В.* Нелинейная динамика хаотических и стохастических систем. — Саратов: Изд. Саратовского университета, 1999.
- [7] *Серых И.В.* Реакция Индийского океана на события Эль-Ниньо // *Молодой ученый.* — 2010. — Vol. 3, no. 14. — Pp. 83–89.
- [8] *Abarbanel Henry D. I.* Analysis of Observed Chaotic Data. Institute for Nonlinear Science. — New York, NY: Springer New York, 1996.
- [9] *Alexander Michael A., Bladé Ileana, Newman Matthew et al.* The Atmospheric Bridge: The Influence of ENSO Teleconnections on Air–Sea Interaction over the Global Oceans // *Journal of Climate.* — 2002. — aug. — Vol. 15, no. 16. — Pp. 2205–2231.

- [10] *Anishchenko Vadim S., Pavlov Alexey N.* Global reconstruction in application to multichannel communication // *Physical Review E*. — 1998. — feb. — Vol. 57, no. 2. — Pp. 2455–2457.
- [11] *Arnold Ludwig.* Random Dynamical Systems. Springer Monographs in Mathematics. — Berlin: Springer-Verlag, 1998.
- [12] *Bagarinao Epifanio, Nomura Taishin, Pakdaman K., Sato Shunsuke.* Generalized one-parameter bifurcation diagram reconstruction using time series // *Physica D: Nonlinear Phenomena*. — 1998. — dec. — Vol. 124, no. 1-3. — Pp. 258–270.
- [13] *Bagarinao Epifanio, Pakdaman K., Nomura Taishin, Sato Shunsuke.* Time series-based bifurcation diagram reconstruction // *Physica D: Nonlinear Phenomena*. — 1999. — jun. — Vol. 130, no. 3-4. — Pp. 211–231.
- [14] *Balmaseda Magdalena A., Trenberth Kevin E., Källén Erland.* Distinctive climate signals in reanalysis of global ocean heat content // *Geophysical Research Letters*. — 2013. — may. — Vol. 40, no. 9. — Pp. 1754–1759.
- [15] *Barnston 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.
- [16] *Barnston Anthony G., Tippett Michael K., L'Heureux Michelle L. et al.* Skill of real-time seasonal ENSO model predictions during 2002-11: Is our capability increasing? // *Bulletin of the American Meteorological Society*. — 2012. — may. — Vol. 93, no. 5. — Pp. 631–651.
- [17] *Barriopedro David, Calvo Natalia.* On the Relationship between ENSO, Stratospheric Sudden Warmings, and Blocking // *Journal of Climate*. — 2014. — jun. — Vol. 27, no. 12. — Pp. 4704–4720.
- [18] *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.
- [19] *Berry T., Cressman J. R., Gregurić-Ferenček Z., Sauer T.* Time-Scale Separation from Diffusion-Mapped Delay Coordinates // *SIAM Journal on Applied Dynamical Systems*. — 2013. — jan. — Vol. 12, no. 2. — Pp. 618–649.

- [20] *Bezruchko Boris P., Smirnov Dmitry A.* Constructing nonautonomous differential equations from experimental time series // *Physical Review E*. — 2000. — dec. — Vol. 63, no. 1. — P. 016207.
- [21] *Bezruchko Boris P., Smirnov Dmitry A.* Constructing nonautonomous differential equations from experimental time series // *Physical Review E*. — 2000. — dec. — Vol. 63, no. 1. — P. 016207.
- [22] *Bezruchko Boris P., Smirnov Dmitry A.* Extracting Knowledge From Time Series. Springer Series in Synergetics. — Berlin, Heidelberg: Springer Berlin Heidelberg, 2010.
- [23] *Bjercknes J.* Atmospheric Teleconnections from the Equatorial {Pacific} // *Mon. Wea. Rev.* — 1969. — Vol. 97, no. 3. — Pp. 163–172.
- [24] *Casdagli Martin, Eubank Stephen, Farmer J.Doyne, Gibson John.* State space reconstruction in the presence of noise // *Physica D: Nonlinear Phenomena*. — 1991. — aug. — Vol. 51, no. 1-3. — Pp. 52–98.
- [25] *Chao Yi, Ghil Michael, McWilliams James C.* Pacific interdecadal variability in this century's sea surface temperatures // *Geophysical Research Letters*. — 2000. — Vol. 27, no. 15. — Pp. 2261–2264.
- [26] *Chekroun Mickaël D., Kondrashov Dmitri.* Data-adaptive harmonic spectra and multilayer Stuart-Landau models // *Chaos: An Interdisciplinary Journal of Nonlinear Science*. — 2017. — sep. — Vol. 27, no. 9. — P. 093110.
- [27] *Coifman Ronald R., Lafon Stéphane.* Diffusion maps // *Applied and Computational Harmonic Analysis*. — 2006. — Vol. 21, no. 1. — Pp. 5–30.
- [28] *DelSole Timothy.* Optimally Persistent Patterns in Time-Varying Fields // *Journal of the Atmospheric Sciences*. — 2001. — jun. — Vol. 58, no. 11. — Pp. 1341–1356.
- [29] *DelSole Timothy, Tippett Michael K.* Average Predictability Time. Part I: Theory // *Journal of the Atmospheric Sciences*. — 2009. — may. — Vol. 66, no. 5. — Pp. 1172–1187.
- [30] *DelSole Timothy, Tippett Michael K.* Average Predictability Time. Part II: Seamless Diagnoses of Predictability on Multiple Time Scales // *Journal of the Atmospheric Sciences*. — 2009. — may. — Vol. 66, no. 5. — Pp. 1188–1204.
- [31] *DelSole Timothy, Tippett Michael K., Shukla Jagadish.* A Significant Component of Unforced Multidecadal Variability in the Recent Acceleration of Global Warming // *Journal of Climate*. — 2011. — feb. — Vol. 24, no. 3. — Pp. 909–926.

- [32] *Dempster A. P., Laird N. M., Rubin D. B.* Maximum Likelihood from Incomplete Data via the EM Algorithm // *Journal of the Royal Statistical Society. Series B (Methodological)*. — 1977. — Vol. 39. — Pp. 1–38.
- [33] *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. — aug. — Vol. 17, no. 16. — Pp. 3109–3124.
- [34] *Deyle Ethan R., Sugihara George, Deyle ER et al.* Generalized Theorems for Nonlinear State Space Reconstruction // *PLoS ONE*. — 2011. — mar. — Vol. 6, no. 3. — P. e18295.
- [35] *Easterling David R., Wehner Michael F.* Is the climate warming or cooling? // *Geophysical Research Letters*. — 2009. — apr. — Vol. 36, no. 8. — P. L08706.
- [36] *England Matthew H., McGregor Shayne, Spence Paul et al.* Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus // *Nature Climate Change*. — 2014. — mar. — Vol. 4, no. 3. — Pp. 222–227.
- [37] *Erdelyi Arthur.* Asymptotic expansions. — Dover Publications, 1956. — P. 108.
- [38] *Farmer J. Doyne, Sidorowich John J.* Predicting chaotic time series // *Physical Review Letters*. — 1987. — aug. — Vol. 59, no. 8. — Pp. 845–848.
- [39] *Feigin A. M., Molkov Ya. I., Mukhin D. N., Loskutov E. M.* Prognosis of Qualitative Behavior of a Dynamic System by the Observed Chaotic Time Series // *Radiophysics and Quantum Electronics*. — 2001. — Vol. 44, no. 5/6. — Pp. 348–367.
- [40] *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. — jan. — Vol. 120. — Pp. 105–123.
- [41] *Gámez A. J., Zhou C. S., Timmermann A., Kurths J.* Nonlinear dimensionality reduction in climate data // *Nonlinear Processes in Geophysics*. — 2004. — sep. — Vol. 11, no. 3. — Pp. 393–398.
- [42] *Gelfand Alan E., Smith Adrian F. M.* Sampling-Based Approaches to Calculating Marginal Densities // *Journal of the American Statistical Association*. — 1990. — jun. — Vol. 85, no. 410. — P. 398.
- [43] *Ghil M., Allen M. R., Dettinger M. D. et al.* Advanced spectral methods for climatic time series // *Reviews of Geophysics*. — 2002. — feb. — Vol. 40, no. 1. — P. 1003.

- [44] *Grieger B, Latif M.* Reconstruction of the El Niño attractor with neural networks // *Climate Dyn.* — 1994. — Vol. 10. — Pp. 267–276.
- [45] *Guckenheimer John, Timmermann Axel, Dijkstra Henk, Roberts Andrew.* (Un)predictability of strong El Niño events // *Dynamics and Statistics of the Climate System.* — 2017. — dec.
- [46] *Gushchina Daria, Dewitte Boris.* Intraseasonal Tropical Atmospheric Variability Associated with the Two Flavors of El Niño // *Monthly Weather Review.* — 2012. — nov. — Vol. 140, no. 11. — Pp. 3669–3681.
- [47] *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.
- [48] *Hannachi A., Turner A. G.* 20th century intraseasonal Asian monsoon dynamics viewed from Isomap // *Nonlinear Processes in Geophysics.* — 2013. — oct. — Vol. 20, no. 5. — Pp. 725–741.
- [49] *Hasselmann K.* PIPs and POPs: The reduction of complex dynamical systems using principal interaction and oscillation patterns // *Journal of Geophysical Research.* — 1988. — Vol. 93, no. D9. — P. 11015.
- [50] *Hastie T.* Principal Curves and Surfaces: Ph.D Dissertation: Ph.D. thesis / Stanford Linear Accelerator Center, Stanford University. — 1984. <http://www.slac.stanford.edu/cgi-wrap/getdoc/slac-r-276.pdf>.
- [51] *Horn Roger A., Johnson Charles R.* Matrix analysis. — Cambridge University Press, 1985. — P. 561.
- [52] *Hornik Kurt, Stinchcombe Maxwell, White Halbert.* Multilayer feedforward networks are universal approximators // *Neural Networks.* — 1989. — Vol. 2, no. 5. — Pp. 359–366.
- [53] *Huang Boyin, Banzon Viva F., Freeman Eric et al.* Extended Reconstructed Sea Surface Temperature Version 4 (ERSST.v4). Part I: Upgrades and Intercomparisons // *Journal of Climate.* — 2015. — feb. — Vol. 28, no. 3. — Pp. 911–930.
- [54] *Jeffreys Harold.* Theory of probability. — Clarendon Press, 1998. — P. 459.
- [55] *Jia F., Martin E.B., Morris A.J.* Non-linear principal components analysis for process fault detection // *Computers & Chemical Engineering.* — 1998. — mar. — Vol. 22. — Pp. S851–S854.

- [56] *Jien Jerry Y., Gough William A., Butler Ken.* The Influence of El Niño–Southern Oscillation on Tropical Cyclone Activity in the Eastern North Pacific Basin // *Journal of Climate.* — 2015. — mar. — Vol. 28, no. 6. — Pp. 2459–2474.
- [57] *Johnson Scot D., Battisti David S., Sarachik E. S.* Seasonality in an Empirically Derived Markov Model of Tropical Pacific Sea Surface Temperature Anomalies* // *Journal of Climate.* — 2000. — sep. — Vol. 13, no. 18. — Pp. 3327–3335.
- [58] *Jolliffe I T.* Principal Component Analysis. Springer Series in Statistics. — 2nd edition. — New York, NY: Springer New York, 1986. — P. 271.
- [59] *Judd Kevin, Mees Alistair.* Embedding as a modeling problem // *Physica D: Nonlinear Phenomena.* — 1998. — sep. — Vol. 120, no. 3-4. — Pp. 273–286.
- [60] *Kao Hsun-Ying, Yu Jin-Yi.* Contrasting Eastern-Pacific and Central-Pacific Types of ENSO // *Journal of Climate.* — 2009. — feb. — Vol. 22, no. 3. — Pp. 615–632.
- [61] *Kennel Matthew B., Brown Reggie, Abarbanel Henry D. I.* Determining embedding dimension for phase-space reconstruction using a geometrical construction // *Physical Review A.* — 1992. — mar. — Vol. 45, no. 6. — Pp. 3403–3411.
- [62] *Kessler William S.* Is ENSO a cycle or a series of events? // *Geophysical Research Letters.* — 2002. — dec. — Vol. 29, no. 23. — Pp. 40–1–40–4.
- [63] *Knight Jeff R., Folland Chris K., Scaife Adam A.* Climate impacts of the Atlantic Multidecadal Oscillation // *Geophysical Research Letters.* — 2006. — sep. — Vol. 33, no. 17. — P. L17706.
- [64] *Kocharovskaya E R, Gavrilov A S, Kocharovskiy V V et al.* Empirical mode with a variable spatial-temporal structure and the dynamics of superradiant lasers // *Journal of Physics: Conference Series.* — 2016. — Vol. 740, no. 1. — P. 12007.
- [65] *Kocharovskiy V., Feigin A., Gavrilov A. et al.* An approach of the space-time empirical modes to the nonlinear phenomena in lasers with Low-Q cavities // *Optics InfoBase Conference Papers.* — Vol. Part F81-E. — 2017.
- [66] *Kocharovskiy Vl V, Gavrilov A S, Kocharovskaya E R et al.* Comparative Analysis of the Dynamical Spectra of a Polarization of an Active Medium and an Electromagnetic Field in the Superradiant Heterolasers // *KnE Engineering.* — 2018. — oct. — Vol. 3, no. 6. — Pp. 160–173.

- [67] *Kondrashov D., Kravtsov S., Robertson A. W., Ghil M.* A Hierarchy of Data-Based ENSO Models // *Journal of Climate*. — 2005. — nov. — Vol. 18, no. 21. — Pp. 4425–4444.
- [68] *Kramer Mark A.* Nonlinear principal component analysis using autoassociative neural networks // *AIChE Journal*. — 1991. — feb. — Vol. 37, no. 2. — Pp. 233–243.
- [69] *Kravtsov S.* An empirical model of decadal ENSO variability // *Climate Dynamics*. — 2012. — nov. — Vol. 39, no. 9-10. — Pp. 2377–2391.
- [70] *Kravtsov S, Kondrashov D, Ghil M.* Multilevel Regression Modeling of Nonlinear Processes: {D}erivation and Applications to Climatic Variability // *J. Climate*. — 2005. — Vol. 18. — Pp. 4404–4424.
- [71] *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.
- [72] *Kug Jong-Seong, Li Tim, An Soon-Il et al.* Role of the ENSO–Indian Ocean coupling on ENSO variability in a coupled GCM // *Geophysical Research Letters*. — 2006. — may. — Vol. 33, no. 9. — P. L09710.
- [73] *Kwasniok F.* The reduction of complex dynamical systems using principal interaction patterns // *Physica D: Nonlinear Phenomena*. — 1996. — apr. — Vol. 92, no. 1-2. — Pp. 28–60.
- [74] *Kwasniok F.* Optimal Galerkin approximations of partial differential equations using principal interaction patterns // *Physical Review E*. — 1997. — Vol. 55, no. 5. — Pp. 5365–5375.
- [75] *Kwasniok Frank.* Reduced Atmospheric Models Using Dynamically Motivated Basis Functions // *Journal of the Atmospheric Sciences*. — 2007. — oct. — Vol. 64, no. 10. — Pp. 3452–3474.
- [76] *Lau Ngar-Cheung, Nath Mary Jo.* Atmosphere–Ocean Variations in the Indo-Pacific Sector during ENSO Episodes // *Journal of Climate*. — 2003. — jan. — Vol. 16, no. 1. — Pp. 3–20.
- [77] *Lee John A. (John Aldo), Verleysen Michel.* Nonlinear dimensionality reduction. — Springer, 2007.

- [78] *Liu Wei, Huang Boyin, Thorne Peter W. et al.* Extended Reconstructed Sea Surface Temperature Version 4 (ERSST.v4): Part II. Parametric and Structural Uncertainty Estimations // *Journal of Climate*. — 2015. — feb. — Vol. 28, no. 3. — Pp. 931–951.
- [79] *Liu Zhengyu, Alexander Mike.* Atmospheric bridge, oceanic tunnel, and global climatic teleconnections // *Reviews of Geophysics*. — 2007. — jun. — Vol. 45, no. 2. — P. RG2005.
- [80] *Lorenz Edward N.* Deterministic Nonperiodic Flow // *Journal of the Atmospheric Sciences*. — 1963. — mar. — Vol. 20, no. 2. — Pp. 130–141.
- [81] *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.
- [82] *Mantua Nathan J, Hare Steven R, Zhang Yuan et al.* A Pacific Interdecadal Climate Oscillation with Impacts on Salmon Production // *Bulletin of the American Meteorological Society*. — 1997. — Vol. 78, no. 6. — Pp. 1069–1079.
- [83] *Matveeva Tatiana, Gushchina Daria, Dewitte Boris.* The seasonal relationship between intraseasonal tropical variability and ENSO in CMIP5 // *Geoscientific Model Development*. — 2018. — jun. — Vol. 11, no. 6. — Pp. 2373–2392.
- [84] *Meehl Gerald A., Arblaster Julie M., Fasullo John T. et al.* Model-based evidence of deep-ocean heat uptake during surface-temperature hiatus periods // *Nature Climate Change*. — 2011. — oct. — Vol. 1, no. 7. — Pp. 360–364.
- [85] *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. 036216.
- [86] *Molkov Ya I., Mukhin D. N., Loskutov E. M. et al.* Using the minimum description length principle for global reconstruction of dynamic systems from noisy time series // *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*. — 2009. — oct. — Vol. 80, no. 4. — P. 046207.
- [87] *Molkov Y. I., Mukhin D. N., Loskutov E. M. et al.* Prognosis of qualitative system behavior by noisy, nonstationary, chaotic time series // *Physical Review E*. — 2011. — sep. — Vol. 84, no. 3. — P. 036215.

- [88] *Mukhin Dmitry, Gavrilov Andrey, Loskutov Evgeny et al.* Bayesian Data Analysis for Revealing Causes of the Middle Pleistocene Transition // *Scientific Reports 2019 9:1*. — 2019. — may. — Vol. 9, no. 1. — P. 7328.
- [89] *Mukhin Dmitry, Loskutov Evgeny, Mukhina Anna et al.* Predicting Critical Transitions in ENSO Models. Part I: Methodology and Simple Models with Memory // *Journal of Climate*. — 2015. — mar. — Vol. 28, no. 5. — Pp. 1940–1961.
- [90] *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. — mar. — Vol. 73, no. 3. — P. 036211.
- [91] *Navarra A, Simoncini V.* A Guide to Empirical Orthogonal Functions for Climate Data Analysis. — Springer-Verlag, 2010.
- [92] *Neal Radford M.* Tech.Rep.: Probabilistic inference using Markov chain Monte Carlo methods. — Department of Computer Science, University of Toronto, Toronto, Ontario, Canada, 1993.
- [93] *Penland Cécile, Magorian Theresa.* Prediction of Niño 3 Sea Surface Temperatures Using Linear Inverse Modeling // *Journal of Climate*. — 1993. — jun. — Vol. 6, no. 6. — Pp. 1067–1076.
- [94] *Penland C, Matrosova L.* Prediction of tropical Atlantic sea surface temperatures using linear inverse modeling. // *Journal of Climate*. — 1998. — no. 1992. — Pp. 483–496.
- [95] *Penland Cécile, Sardeshmukh Prashant D.* The Optimal Growth of Tropical Sea Surface Temperature Anomalies // *Journal of Climate*. — 1995. — aug. — Vol. 8, no. 8. — Pp. 1999–2024.
- [96] *Pielke Roger A., Landsea Christopher N.* La Niña, El Niño and Atlantic Hurricane Damages in the United States // *Bulletin of the American Meteorological Society*. — 1999. — oct. — Vol. 80, no. 10. — Pp. 2027–2033.
- [97] *Pires Carlos A. L., Hannachi Abdel.* Independent Subspace Analysis of the Sea Surface Temperature Variability: Non-Gaussian Sources and Sensitivity to Sampling and Dimensionality // *Complexity*. — 2017. — aug. — Vol. 2017. — Pp. 1–23.
- [98] *Pires Carlos A. L., Ribeiro Andreia F. S.* Separation of the atmospheric variability into non-Gaussian multidimensional sources by projection pursuit techniques // *Climate Dynamics*. — 2016. — apr. — Pp. 1–30.

- [99] *Powell M. J. D.* Approximation Theory and Methods. — Cambridge University Press, 1981. — mar.
- [100] *Preisendorfer R.W.* Principal Component Analysis in Meteorology and Oceanography / Ed. by C.D. Mobley. — Elsevier, 1988.
- [101] *Rasmusson Eugene M., Carpenter Thomas H.* Variations in Tropical Sea Surface Temperature and Surface Wind Fields Associated with the Southern Oscillation/El Niño // *Monthly Weather Review.* — 1982. — may. — Vol. 110, no. 5. — Pp. 354–384.
- [102] *Rayner N A, Parker D E, Horton E B et al.* 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.
- [103] *Reynolds Richard W., Rayner Nick A., Smith Thomas M. et al.* An Improved In Situ and Satellite SST Analysis for Climate // *Journal of Climate.* — 2002. — jul. — Vol. 15, no. 13. — Pp. 1609–1625.
- [104] *Rissanen J.* Modeling by shortest data description // *Automatica.* — 1978. — sep. — Vol. 14, no. 5. — Pp. 465–471.
- [105] *Ross I., Valdes P. J., Wiggins S.* ENSO dynamics in current climate models: an investigation using nonlinear dimensionality reduction // *Nonlinear Processes in Geophysics.* — 2008. — apr. — Vol. 15, no. 2. — Pp. 339–363.
- [106] *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.
- [107] *Saha Suranjana, Moorthi Shrinivas, Wu Xingren et al.* The NCEP Climate Forecast System Version 2 // *Journal of Climate.* — 2014. — mar. — Vol. 27, no. 6. — Pp. 2185–2208.
- [108] *Saji N H, Goswami B N, Vinayachandran P N, Yamagata T.* A dipole mode in the tropical Indian Ocean // *Nature.* — 1999. — Vol. 401, no. 6751. — Pp. 360–363.
- [109] *Scholkopf Bernhard, Smola Alexander, Muller Klaus R.* Nonlinear Component Analysis as a Kernel Eigenvalue Problem // *Neural Comp.* — 1998. — Vol. 10, no. 5. — Pp. 1299–1319.
- [110] *Schwarz Gideon.* Estimating the Dimension of a Model // *The Annals of Statistics.* — 1978. — mar. — Vol. 6, no. 2. — Pp. 461–464.

- [111] *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. — jan. — Vol. 65, no. 2. — P. 026205.
- [112] *Smirnov D. A., Sysoev I. V., Seleznev E. P., Bezruchko B. P.* Reconstructing nonautonomous system models with discrete spectrum of external action // *Technical Physics Letters*. — 2003. — oct. — Vol. 29, no. 10. — Pp. 824–827.
- [113] *Stolbova Veronika, Surovyatkina Elena, Bookhagen Bodo, Kurths Jürgen.* Tipping elements of the Indian monsoon: Prediction of onset and withdrawal // *Geophysical Research Letters*. — 2016. — apr. — Vol. 43, no. 8. — Pp. 3982–3990.
- [114] *Strounine K, Kravtsov S, Kondrashov D, Ghil M.* Reduced models of atmospheric low-frequency variability: Parameter estimation and comparative performance // *Physica D*. — 2010. — Vol. 239. — Pp. 145–166.
- [115] *Suarez Max J., Schopf Paul S.* A Delayed Action Oscillator for ENSO // *Journal of the Atmospheric Sciences*. — 1988. — nov. — Vol. 45, no. 21. — Pp. 3283–3287.
- [116] *Takens Floris.* Detecting strange attractors in turbulence // *Dynamical Systems and Turbulence*, Warwick 1980. — Springer Berlin Heidelberg. — Pp. 366–381.
- [117] *Tan Shufeng, Mayrovouniotis Michael L.* Reducing data dimensionality through optimizing neural network inputs // *AIChE Journal*. — 1995. — jun. — Vol. 41, no. 6. — Pp. 1471–1480.
- [118] *Tippett Michael K., Barnston Anthony G., Li Shuhua.* Performance of Recent Multimodel ENSO Forecasts // *Journal of Applied Meteorology and Climatology*. — 2012. — mar. — Vol. 51, no. 3. — Pp. 637–654.
- [119] *Tokunaga Ryuji, Kajiwara Shihoko, Matsumoto Takashi.* Reconstructing bifurcation diagrams only from time-waveforms // *Physica D: Nonlinear Phenomena*. — 1994. — dec. — Vol. 79, no. 2-4. — Pp. 348–360.
- [120] *Trenberth Kevin E.* The Definition of El Niño // *Bulletin of the American Meteorological Society*. — 1997. — dec. — Vol. 78, no. 12. — Pp. 2771–2777.
- [121] *Trenberth Kevin E, Fasullo John T.* Tracking Earth’s Energy // *Science*. — 2010. — Vol. 328, no. 5976. — Pp. 316–317.
- [122] *Trenberth Kevin E., Fasullo John T.* An apparent hiatus in global warming? // *Earth’s Future*. — 2013. — dec. — Vol. 1, no. 1. — Pp. 19–32.

- [123] *UCLA*. SSA-MTM toolkit for spectral analysis (cited 2019). Available online at research.atmos.ucla.edu/tcd/ssa/.
- [124] *Vejmělka Martin, Pokorná Lucie, Hlinka Jaroslav et al.* Non-random correlation structures and dimensionality reduction in multivariate climate data // *Climate Dynamics*. — 2015. — may. — Vol. 44, no. 9-10. — Pp. 2663–2682.
- [125] *Volodin E. M., Dianskii N. A., Gusev A. V.* Simulating present-day climate with the INMCM4.0 coupled model of the atmospheric and oceanic general circulations // *Izvestiya, Atmospheric and Oceanic Physics*. — 2010. — aug. — Vol. 46, no. 4. — Pp. 414–431.
- [126] *Volodin E. M., Diansky N. A., Gusev A. V.* Simulation and prediction of climate changes in the 19th to 21st centuries with the Institute of Numerical Mathematics, Russian Academy of Sciences, model of the Earth’s climate system // *Izvestiya, Atmospheric and Oceanic Physics*. — 2013. — jul. — Vol. 49, no. 4. — Pp. 347–366.
- [127] *Wang Chunzai, Deser Clara, Yu Jin-Yi et al.* El Niño and Southern Oscillation (ENSO): A Review // *Coral Reefs of the Eastern Tropical Pacific*. — Springer, Dordrecht, 2017. — Pp. 85–106.
- [128] *Whitney Hassler*. Differentiable Manifolds // *The Annals of Mathematics*. — 1936. — jul. — Vol. 37, no. 3. — P. 645.
- [129] *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.
- [130] *Wu Zhaohua, Huang Norden E., Wallace John M. et al.* On the time-varying trend in global-mean surface temperature // *Climate Dynamics*. — 2011. — aug. — Vol. 37, no. 3-4. — Pp. 759–773.
- [131] *Wyrtki Klaus*. El Niño—The Dynamic Response of the Equatorial Pacific Ocean to Atmospheric Forcing // *Journal of Physical Oceanography*. — 1975. — oct. — Vol. 5, no. 4. — Pp. 572–584.
- [132] *Xie Shang-Ping, Annamalai H., Schott Friedrich A., McCreary Julian P.* Structure and Mechanisms of South Indian Ocean Climate Variability* // *Journal of Climate*. — 2002. — apr. — Vol. 15, no. 8. — Pp. 864–878.
- [133] *Xin Xiaoge, Gao Feng, Wei Min et al.* Decadal prediction skill of BCC-CSM1.1 climate model in East Asia // *International Journal of Climatology*. — 2017. — aug.

- [134] *Xue Yan, Leetmaa Ants, Ji Ming.* ENSO Prediction with Markov Models: The Impact of Sea Level // *Journal of Climate*. — 2000. — feb. — Vol. 13, no. 4. — Pp. 849–871.
- [135] *Zebiak Stephen E., Cane Mark A.* A Model El Niño–Southern Oscillation // *Monthly Weather Review*. — 1987. — oct. — Vol. 115, no. 10. — Pp. 2262–2278.
- [136] *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.
- [137] *de la Iglesia Manuel D., Tabak Esteban G.* Principal Dynamical Components // *Communications on Pure and Applied Mathematics*. — 2013. — jan. — Vol. 66, no. 1. — Pp. 48–82.
- [138] *von Storch Hans, Bürger Gerd, Schnur Reiner, von Storch Jin-Song.* Principal Oscillation Patterns: A Review // *Journal of Climate*. — 1995. — mar. — Vol. 8, no. 3. — Pp. 377–400.