

На правах рукописи



Запорожцев Иван Федорович

**КРАТКОСРОЧНОЕ ПРОГНОЗИРОВАНИЕ
ПРОСТРАНСТВЕННО-ВРЕМЕННОЙ ИЗМЕНЧИВОСТИ
ОКЕАНОГРАФИЧЕСКИХ ХАРАКТЕРИСТИК
МЕТОДАМИ АНАЛИЗА МНОГОМЕРНЫХ ВРЕМЕННЫХ РЯДОВ**

Специальность: 05.13.18 – Математическое моделирование,
численные методы и комплексы программ

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата технических наук

Санкт-Петербург – 2016

Работа выполнена в ФГБОУ ВПО «Мурманский государственный технический университет» на кафедре математики, информационных систем и программного обеспечения

Научный руководитель: Доктор технических наук, профессор
Середа Альгирдас-Владимир Игнатьевич

Официальные оппоненты: Доктор технических наук, профессор
Нечаев Юрий Иванович,
заслуженный деятель науки РФ, «Санкт-Петербургский государственный морской технический университет», профессор кафедры вычислительной техники и информационных технологий

Кандидат физико-математических наук, доцент
Фролькис Виктор Абрамович,
«Петербургский государственный университет путей сообщения Императора Александра I», профессор кафедры математики и моделирования

Ведущая организация: Федеральное государственное бюджетное учреждение науки Мурманский морской биологический институт Кольского научного центра Российской академии наук (ММБИ КНЦ РАН, г. Мурманск)

Защита диссертации состоится «02» марта 2016 года в 15:30 на заседании диссертационного совета Д 212.238.01 Санкт-Петербургского государственного электротехнического университета «ЛЭТИ» им. В.И. Ульянова (Ленина) по адресу: 197376, Санкт-Петербург, ул. Проф. Попова, 5.

С диссертацией можно ознакомиться в библиотеке Санкт-Петербургского государственного электротехнического университета «ЛЭТИ» им. В.И. Ульянова (Ленина) и на сайте СПбГЭТУ «ЛЭТИ»: www.eltech.ru.

Автореферат разослан «24» декабря 2015 г.

Ученый секретарь
диссертационного совета Д 212.238.01, к.т.н.



Щеголева Н.Л.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы. Комплексное изучение временной изменчивости характеристик поверхности океана является неотъемлемой частью многолетнего анализа геофизических процессов в атмосфере и океане. Компьютерное моделирование и разработка математически обоснованных технологий численного анализа характеристик океанической поверхности представляют несомненный интерес для специалистов различных областей научных исследований (океанологов, биологов, экологов и др.) и практической деятельности (рыболовства, нефтегазодобывающей, транспортно-логистической и т.п.).

В последние десятилетия российскими учёными И.М. Ашиком, С.К. Поповым, А.Л. Лобовым, В.В. Елисовым, О.Н. Вербицкой и др. разработаны гидродинамические модели уровня поверхности для Баренцева, Белого, Карского и Каспийского морей. Эти теоретические модели могут быть адаптированы для исследования различных характеристик поверхности, однако они строятся исходя из анализа данных прибрежных станций и других источников, исключая спутниковый мониторинг, который имеет наибольший охват по территории. Известно, что в ряде случаев поля характеристик поверхности, являющиеся результатами применения методов численной гидродинамики и спутниковой океанографии, плохо согласованы. Несмотря на существенную погрешность данных дистанционного зондирования Земли спутниками, именно они являются основным источником для построения массивов гридированных (восстановленных в узлах равномерной планарной сетки) значений температуры и аномалий уровня поверхности моря. Анализ и краткосрочный прогноз временных рядов этих двух характеристик в узлах сетки вместе с учётом метеорологической информации является основой оперативного информационного обеспечения поддержки принятия управленческих решений на море. Эмпирические модели, устанавливающие зависимости в значениях временных рядов одной или нескольких целевых характеристик, появились задолго до гидродинамических в рамках физико-статистического подхода и существуют до сих пор. В основе нового направления лежат так называемые структурные модели, которые стали результатом усложнения по сравнению с традиционными статистическими аналитическими зависимостями: вводятся различные способы расчёта коэффициентов линейных комбинаций, стала возможна работа с нелинейными и нестационарными процессами, порождающими временные ряды. Примерами таких формализмов являются искусственные нейронные сети, модели, построенные по результатам многомерных вариантов сингулярного спектрального анализа (МССА), декомпозиции на эмпирические моды (МДЭМ) и т.д. Значительный вклад в становление и развитие двух последних групп методов внесли: Н.Э. Голяндина, А. Zhigljavsky, М. Ghil, Д.В. Степанов (МССА), N.E. Huang, В.А. Давыдов, Р. Yang, R.T. Rato, J. Fleureau, N. Rehman, D.P. Mandic (МДЭМ), результаты работ которых были использованы при написании диссертации. Общей чертой всех структурных моделей является алгоритмичность: необходимость выполнения ряда шагов, не все из которых хорошо формализованы; также в них присутствует значительное

число различных модельных параметров. Именно в рамках этих моделей в настоящее время ведётся работа над созданием системного адаптивного подхода к прогнозированию значений рядов без учёта специфики анализируемых данных, обусловленной их природой. При построении прогноза гридированных значений важен анализ не только системы рядов, но и взаимного расположения соответствующих им узлов, в то время как известные структурные модели ориентированы исключительно на работу с рядами. Таким образом, построение новой методики прогнозирования на основе структурных моделей, но с учётом пространственно-временной изменчивости является целесообразным.

Цель работы – создание методики краткосрочного прогнозирования как набора методов, использующих структурные и статистические модели нестационарных временных рядов, и реализующей их компьютерной технологии, позволяющей получать результаты в темпе поступления исходных экспериментальных данных (на примере температуры поверхности и аномалий уровня моря).

В работе поставлены и решены следующие **задачи исследования**:

1. Выполнение обзора методов и моделей прогнозирования, которые используются в настоящее время или имеют потенциал для применения в рамках морских гидрологических прогнозов; выявление их достоинств и недостатков.

2. Выбор и обоснование *модели пространственно-временной изменчивости* целевой физической характеристики как модели многомерного временного ряда её гридированных значений в заданной акватории.

3. Разработка методики краткосрочного прогнозирования как набора методов идентификации модели пространственно-временной изменчивости:

3.1. Разработка *метода кластеризации* узлов планарной равномерной сетки и соответствующих им одномерных рядов (выделение *кластеров узлов* и группировка по ним одномерных рядов; группы называются *кластерными рядами*).

3.2. Разработка *метода построения начального внутрикластерного прогноза (метода НВП)* для каждого кластерного ряда в отдельности.

3.3. Разработка *метода уточнения начального внутрикластерного прогноза (метода УВП)* для получения итогового краткосрочного прогноза.

4. Создание комплекса программ, реализующих методику прогнозирования, для построения прогнозов по сгенерированным («искусственным») и реальным данным, их статистической обработки и анализа их качества, а также для визуализации необходимой аналитической информации.

Объектом исследования диссертационной работы является океаническая поверхность в некоторой акватории. **Предметом изучения** являются модели и алгоритмы, позволяющие составить краткосрочный прогноз пространственно-временной изменчивости целевой характеристики поверхности океана.

Методы исследования. Для решения поставленных задач использовались: математическое моделирование, методы анализа многомерных рядов (МССА и МДЭМ), математической статистики, кластерного анализа, линейной алгебры; в экспериментальной части – методы теории графов, структурного и объектно-ориентированного программирования, интерактивные компьютерные техноло-

гии визуализации аналитической информации, вычислительный эксперимент.

Научная новизна работы заключается в следующем:

1. Предложена и экспериментально обоснована модель пространственно-временной изменчивости целевой физической характеристики, представляющая исходный многомерный временной ряд как систему многомерных кластерных рядов, аддитивные составляющие которых выражены структурными и статистическими моделями.

2. Разработана методика краткосрочного прогнозирования, включающая:

2.1. Метод разбиения множества узлов равномерной сетки на подмножества, названный автором «метод кластеризации», обеспечивающий группировку сильно взаимно коррелированных одномерных временных рядов, соответствующих соседним узлам этой сетки.

2.2. Метод построения начального внутрикластерного прогноза на основе предложенного автором и экспериментально подтверждённого способа совместного использования двух аддитивных моделей временного ряда (в рамках подходов МДЭМ и МССА).

2.3. Метод уточнения начального внутрикластерного прогноза.

3. Разработан комплекс программ (библиотека ядра, консольные утилиты и графическая оболочка), реализующих предложенную автором методику краткосрочного прогнозирования значений целевой характеристики.

Достоверность научных результатов и выводов обусловлена корректностью применяемого математического аппарата и подтверждается результатами вычислительных экспериментов для сгенерированных и реальных данных.

Практическая значимость результатов работы заключается в том, что они обеспечивают решение задачи краткосрочного прогнозирования пространственно-временной изменчивости океанографических характеристик, представляющей практический интерес для специалистов в области промысловой океанографии, экологии, морской биологии и т.п., а также других изменчивых во времени пространственно распределённых характеристик, независимо от их природы.

Положения, выносимые на защиту:

1. Модель пространственно-временной изменчивости целевой физической характеристики, структурная и параметрическая идентификация которой выполняется путём применения авторских методов кластеризации и прогнозирования.

2. Методика краткосрочного прогнозирования, которая включает в себя:

2.1. Метод кластеризации узлов планарной равномерной сетки и соответствующих им одномерных временных рядов.

2.2. Метод построения начального внутрикластерного прогноза многомерного временного ряда на базе МССА и МДЭМ.

2.3. Метод уточнения начального внутрикластерного прогноза, использующий введённые автором кластерные характеристики и модель ошибки начального внутрикластерного прогноза.

3. Комплекс программ, реализующий авторскую методику.

4. Результаты анализа и краткосрочного прогнозирования многомерных временных рядов по массивам сгенерированных и реальных данных.

Внедрение результатов работы. Разработанные методика и программные средства были успешно использованы автором при выполнении НИР в рамках государственного задания Федерального агентства по Рыболовству по теме «Компьютерное моделирование, численный анализ и прогнозирование уровня поверхности океана в акваториях Арктики» (2014 г, № ГР НИОКР 01201459892), а также специалистами научно-производственной компании «Морская информатика», осуществляющей информационное обеспечение рыболовского флота (2015 г., получен акт о внедрении). Основной компонент созданного комплекса программ (библиотека ядра) зарегистрирован в Реестре программ для ЭВМ.

Апробация результатов работы. Результаты, полученные в диссертации, докладывались и обсуждались на 7 конференциях: международных научно-технических конференциях «Наука и образование – 2013» и «Наука и образование – 2014» (Мурманск, 4–11 марта 2013 г., 24–28 марта 2014 г.); XIII и XIV международных научных конференциях студентов и аспирантов «Проблемы Арктического региона» (Мурманск, 14 мая 2013 г., 14 мая 2014 г.); международной научно-технической конференции «Computer Modeling and Simulation» (Санкт-Петербург, 2-4 июля 2014 г.); международной конференции «2014 International Conference on Computer Technologies in Physical and Engineering Applications» (Санкт-Петербург, 30 июня – 4 июля 2014 г.); Первой международной конференции молодых учёных «Проблемы и перспективы развития рыбохозяйственного комплекса на современном этапе» (Мурманск, 22-24 октября 2014 г.).

Исследования поддержаны стипендиями губернатора Мурманской области (2010-2011 уч. год, 2011-2012 уч. год) и стипендией Президента РФ для аспирантов, обучающихся по приоритетным направлениям модернизации и технологического развития российской экономики (2013-2014 уч. год).

Предложенные автором методы, алгоритмы и программные средства показали свою эффективность при обработке реальных данных, относящихся к акваториям Баренцева моря и моря Ирмингера.

Личный вклад автора в совместных работах. Разработка общей схемы и детализация этапов предложенной в диссертации методики краткосрочного прогнозирования, выбор и обоснование модели пространственно-временной изменчивости целевой физической характеристики, разработка необходимых методов и алгоритмов, реализующих эту методику, создание программных средств для экспериментальной оценки её эффективности, осуществление и анализ результатов вычислительных экспериментов проводились автором самостоятельно.

Публикации. По теме диссертации опубликовано 12 научных работ, в том числе: 3 статьи в ведущих рецензируемых научных журналах и изданиях, рекомендованных ВАК РФ; 1 свидетельство о государственной регистрации программы для ЭВМ; 1 работа, зарегистрированная в ЦИТИС; 7 статей в материалах международных научно-технических конференций.

Структура и объём работы. Диссертационная работа состоит из введения, четырёх глав, заключения, библиографического списка и приложений. Работа изложена на 183 страницах машинописного текста, включает 70 рисунков и 16 таблиц. Список использованных источников содержит 112 наименований.

СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность темы диссертационной работы и её практическая значимость, рассмотрены существующие подходы к решению задач предметной области, сформулированы цель и задачи диссертационного исследования, определены объект, предмет и методы исследования, указаны научная новизна и положения, выносимые на защиту. Отмечена практическая ценность полученных результатов, их внедрение и апробация. Представлена информация о структуре и объёме диссертации.

В первой главе исследована специфика формирования многомерных временных рядов значений физических характеристик океанической поверхности и обоснована целесообразность использования одномерных и многомерных рядов для решения ряда задач исследования пространственно-временной изменчивости этих характеристик. В главе приведены сведения об организации процесса сбора и обработки данных об изменчивости температуры поверхности и аномалий уровня моря, представлен обзор методов краткосрочного прогнозирования: гидродинамических, статистических и структурных. Их выбор обусловлен как практикой морских гидрологических прогнозов, так и приложениями в других предметных областях при прогнозировании многомерных временных рядов.

Решаемая в рамках диссертационного исследования задача может быть сформулирована следующим образом. В прямоугольной пространственной области Ω задана планарная равномерная сетка UG . Для каждого узла сетки UG известен временной ряд, отражающий временную изменчивость целевой характеристики, причём временные ряды для каждого из узлов сформированы единообразно: совпадают стартовый временной отсчёт, шаг по времени между измерениями и анализируемая длина временного ряда. Такой подход к формированию временных рядов позволяет говорить, что в области Ω задан многомерный временной ряд $F_N^{UG}(t) = (F_N^1(t), F_N^2(t), \dots, F_N^s(t))$ длины N , где s – число одномерных рядов. Требуется построить прогноз, то есть соответствующий исходному многомерный временной ряд спрогнозированных значений $\tilde{F}_\tau^{UG}(t)$, $t = N+1, N+2, \dots, N+\tau$, в предположении, что задан горизонт прогноза τ . Поскольку имеется в виду краткосрочный прогноз, выполняется условие $\tau \ll N$.

В преддверии решения поставленной задачи необходимо задать или оценить относительную (или нормированную) погрешность в значениях исходного многомерного временного ряда ε . Кроме того, одним из важных элементов решения является проблема определения по величинам τ и ε длины фрагмента T исходного многомерного временного ряда, продолжением которого будет спро-

гнозируемые значения целевой характеристики, при условии, что этот фрагмент имеет наименьшую возможную длину для обеспечения ошибки прогноза того же порядка, что и величина ε . В работе принято, что значения τ и ε одни и те же для всех одномерных рядов, составляющих многомерный ряд.

Вторая глава посвящена построению метода кластеризации двух типов объектов задачи краткосрочного прогнозирования: узлов сетки UG и соответствующих им одномерных временных рядов. Метод разбиения множества временных рядов с географической привязкой на подмножества назван «методом кластеризации» для упрощения ссылок, хотя в узком смысле термина таковым не является: для объектов (узлов или рядов) невозможно указать «центры сгущения», анализ «близости» этих объектов происходит алгоритмически без использования формальной меры (метрики) ввиду наличия как пространственных, так и временных особенностей объектов. Авторский метод использует элементы кластерного анализа, поэтому в главе рассмотрены как известные подходы к решению классической задачи кластеризации, так и новый метод кластеризации, адаптированный к сформулированной в работе задаче прогнозирования и являющийся способом построения системы кластеров узлов UG и кластерных рядов.

Сформулированы требования к методу кластеризации ряда $F_N^{UG}(t)$:

R1: адаптированность к сетке UG : кластеры должны быть прямоугольными или являться объединением прямоугольных областей;

R2: низкая чувствительность к амплитуде и частотному набору шумовой составляющей значений ряда;

R3: формирование кластеров в виде односвязных областей;

R4: обеспечение высокой степени согласованности временной изменчивости значений одномерных рядов одного кластера;

R5: запрет отставания или опережения в динамике временных рядов кластера относительно друг друга (то есть запрет наличия незначительно отличающихся фрагментов двух рядов, сдвинутых относительно друг друга по времени).

В основу метода положены следующие правила:

R1: на начальном этапе следует группировать узлы, находящиеся внутри прямоугольных областей, введя равномерное разбиение сетки UG , так как она является прямоугольной (выполнение R1);

R2: коэффициент кросс-корреляции, вычисленный для любой пары узлов (то есть для соответствующей пары временных рядов) одного кластера, не должен быть меньше наперед заданного порогового значения (выполнение R2 и R4);

R3: для любой пары узлов одного кластера наибольшее значение коэффициента корреляции для заданных фрагментов соответствующих им рядов должно приходиться на нулевой лаг (выполнение R5);

R4: кластеры необходимо укрупнять, объединяя прямоугольные кластеры в соответствии с жадным алгоритмом (то есть алгоритмом, на каждом шаге которого принимаются локально оптимальные решения): объединяются пары имеющих общую пространственную границу кластеров с наибольшим значением наименьшего коэффициента кросс-корреляции, который вычисляется для пар

узлов (то есть для соответствующих пар временных рядов), взятых из разных кластеров (выполнение R3 и R4).

В соответствии с указанными правилами разработан алгоритм (рисунок 1), который использует в качестве меры отличия кластеризуемых объектов выражение $(1 - r_{xy})$, где r_{xy} – коэффициент кросс-корреляции, вычисленный для пары временных рядов x и y . Необходимость дивизивного шага можно обосновать тем фактом, что кластеризация интерпретируется как этап предобработки в задаче краткосрочного прогнозирования. Исходя из требований объединения в определенном смысле наиболее согласованных в динамике изменения значений одномерных рядов, важно принимать во внимание не только значения коэффициента кросс-корреляции, но и лаг, на котором достигается его наибольшее значение. Так как далее система кластеров используется для взаимной коррекции, то есть ставится задача выбора корректирующих кластеров, опережение или отставание одного ряда относительно другого (случай, когда наибольшая корреляция наблюдается не для нулевого лага) следует рассматривать как неоднородность кластера, в котором они содержатся, поэтому кластеры с такими элементами запрещены. Из тех же соображений правила кластеризации не предусматривают работу с искусственными данными в виде «центров кластеров»: происходит анализ только реальных рядов. Выполнение указанных требований позволяет ожидать повышение эффективности авторских методов НВП и УВП.

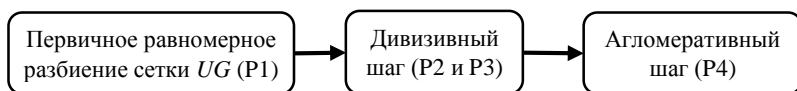


Рисунок 1. Основные этапы метода кластеризации

Во второй главе также вводятся кластерные характеристики:

коэффициент кросс-корреляции для пар «корректирующий-потенциальный корректируемый кластер», определяемый как наименьшее значение коэффициента кросс-корреляции фрагментов рядов одинаковой длины, вычисленного для всех возможных пар одномерных временных рядов, выбранных для разных кластеров при нулевом лаге;

отношения пар ближайших сингулярных чисел в разложении траекторных матриц (в смысле МССА) одномерных временных рядов кластера;

дисперсия кластера (дисперсия пространственной изменчивости приращений характеристики) в момент t за временной промежуток длины τ (определяется как среднее значение выборочных дисперсий приращений целевой характеристики в различных узлах кластера в моменты $t-1, t-2, \dots, t-\tau$ (осреднение по моментам времени, а выборки состоят из одновременных значений)).

направление градиента кластера в момент времени t : термин «градиент» используется не в строгом математическом смысле (определяется как сумма градиентов целевой характеристики в каждом узле кластера, а градиент в узле – по пяти значениям кластерного ряда: в самом узле и в 4-х его ближайших соседях);

его направление – результат аппроксимации путём сведения к одному из четырёх вариантов (по аналогии с направлениями к ближайшим соседям для узла).

Эти характеристики рассматриваются как меры однородности кластера, и согласованности в динамике временных рядов из разных кластеров локального временного масштаба в несколько отсчётов (дисперсия и градиент) и глобального временного масштаба (корреляция и отношения сингулярных чисел).

Разработанный метод кластеризации обеспечивает группировку сильно коррелированных одномерных рядов, если соответствующие узлы являются соседними в сетке UG , и формирует главным образом прямоугольные кластеры узлов, упрощая дальнейший анализ. Результатом выполнения метода кластеризации является построение системы кластеров узлов сетки UG и кластерных рядов; расчёт кластерных характеристик позволяет для каждого кластера C_i определить множество корректирующих кластеров C_j (в методе УВП из главы 3).

Третья глава содержит описание модели пространственно-временной изменчивости целевой физической характеристики и методов НВП и УВП авторской методики краткосрочного прогнозирования, которые служат для идентификации слагаемых в модели изменчивости.

Для всех кластерных рядов в диссертации принят единый вид модели $F^{C_i} \Big|_{t_0}(t)$, но параметрическая идентификация каждой её реализации выполняется отдельно для каждого кластера C_i и носит локальный характер. Она строится по фрагменту многомерного ряда длины T (и аппроксимирует этот фрагмент), оканчивающемуся моментом (t_0-1) , и используется для прогнозирования следующего фрагмента длиной τ , начинающегося с выбранного t_0 :

$$F^{C_i} \Big|_{t_0}(t) = AR_1^{C_i} \Big|_{t_0, T}(F^{C_i}) + AR_2^{C_i} \Big|_{t_0, \tau, u_i \cup C_j}(F^{C_j}, AR_1^{C_j}), \quad (1)$$

Модель ряда (1) содержит два авторегрессионных слагаемых: первое является линейной комбинацией значений кластерного ряда в предыдущие моменты времени, а второе – ошибок прогноза на один момент вперёд. Модель ошибки для кластера C_i строится по массиву ошибок как этого кластера, так и ошибок других кластерах по результатам НВП для u подряд идущих временных отсчётов. Идентификация первого слагаемого модели (1) выполняется по методу НВП, а второго – по методу УВП. Серию таких моделей для различных кластеров C_i сетки UG и моментов времени t_0 можно считать *моделью пространственно-временной изменчивости* целевой физической характеристики.

Метод НВП разработан на основе предложенного автором и экспериментально подтверждённого способа совместного использования двух аддитивных моделей нестационарного ряда (на основе МДЭМ и МССА). В МССА затруднена идентификация квазипериодических колебаний и есть неопределённость в выборе количества значимых компонент разложения, в то время как МДЭМ позволяет в некоторой степени решить эти проблемы. Длину фрагмента временного ряда при прогнозировании по МССА рекомендуется выбирать из априорных

знаний о природе ряда, которые могут отсутствовать, поэтому в данной работе выбор осуществляется из условия квазистационарности одномерного ряда.

В состав метода НВП входит итерационный метод K -продолжения многомерного временного ряда, введённый Н.Э. Голяндиной и Д.В. Степановым и использующий результаты МССА. Базовый вариант НВП, названный D -МССА («дополненный МССА»), строящий прогноз независимо для каждого кластерного ряда, начиная с номера отсчёта $N+1$, содержит следующие шаги.

Шаг 1. Задаются горизонт τ и погрешность входных данных ε , по которым определяется длина фрагмента исходного ряда $T = T_{opt}$ (из условия τ - ε -стационарности, сформулированного в работах Ю.Н. Орлова и К.П. Осминина, усредняется по значениям для всех одномерных рядов кластера).

Шаг 2. Формируются два фрагмента кластерного ряда длины T : обучающий $LF_T = (F_N(N-\tau-T+1), \dots, F_N(N-\tau))$ и тестовый $TF_T = (F_N(N-T+1), \dots, F_N(N))$, для каждого фрагмента выбирается число строк траекторной матрицы в МССА $L = [T/2]$.

Шаг 3. Для фрагмента LF_T строится серия прогнозов по методу K -продолжения на горизонт τ , в которых размерность линейной рекуррентной формулы МССА (количество значимых сингулярных чисел n) равна $1, 2, \dots, L$. По результатам этой серии выбирается оптимальный вариант значения n_{opt} , который представляет минимум ошибки $\delta_{t,T,\tau}^C$ (здесь $t = N - \tau + 1$), выражаемой в %, если все $\delta_{t,T,\tau}^C > \varepsilon$, или первый при упорядоченном переборе, если $\delta_{t,T,\tau}^C \leq \varepsilon$.

$$\delta_{t,T,\tau}^C = \frac{100}{d(t)} \cdot \max_{i=1, \dots, s_C} \sqrt{\frac{\sum_{j=0}^{\tau-1} (F_N^i(t+j) - \tilde{F}_\tau^i(t+j))^2}{\tau}}, \quad (2)$$

где $d(t) = \max_{k=1, \dots, s} \left\{ \max_{t' \leq t} F_N^k(t') - \min_{t' \leq t} F_N^k(t') \right\}$, s_C и s – число узлов кластера и сетки UG .

Шаг 4. Для фрагмента TF_T выполняется единственный эксперимент методом K -продолжения со значением $n = n_{opt}$. Значение $\delta_{N+1,T,\tau}^C$ для этого фрагмента будем считать итоговым значением ошибки НВП.

Важно отметить, что в расчётах погрешность ε считается равной отношению средней абсолютной погрешности к нормирующему выражению $d(N)$.

Были разработаны и экспериментально проверены модифицированные версии этого метода, реализующие МДЭМ: необходимо выполнить декомпозицию ряда длины $T+\tau$, полученного объединением LF_T и TF_T , и применить D -МССА (шаги 3 и 4) к каждой моде отдельно с последующим суммированием результатов или исключить первую (шумовую) моду и воспользоваться D -МССА. Было установлено экспериментально, что второй вариант, требующий меньшее количество операций, генерирует прогноз с меньшей ошибкой, поэтому в дальнейшем использовался именно этот вариант как метод НВП.

Метод уточнения начального внутрикластерного прогноза (метод УВП) использует линейную регрессионную модель ARX, предложенную Н. Tong, которая была успешно применена в метеорологических исследованиях J. Tastu, P.

Pinson и др. Она адаптирована автором диссертации к задачам данного исследования и предлагается в модифицированном виде как *модель ошибки НВП*:

$$e_t(C) = \beta_0 + \sum_{j=1}^{\tau} \beta_j e_{t-j}(C) + \sum_{i=1}^p \sum_{j=1}^{\tau} \beta_{ij} e_{t-j}(C_i), \quad (3)$$

где C – *корректируемый кластер*, то есть кластер, значения ряда \tilde{F}_t^C (то есть прогноза НВП) которого необходимо корректировать (уточнить);
 $e_t(C)$ – средняя ошибка НВП (для момента t и $\tau = 1$) для кластера C (здесь ошибка есть разность реального значения и прогноза);
 C_i – i -тый *корректирующий кластер*, то есть кластер, ошибка прогноза в котором выступает внешней переменной ARX;
 p – число корректирующих кластеров.

Метод уточнения заключается в структурной и параметрической идентификации модели (3), которая прогнозирует ошибку НВП для одного временного отсчёта (а не τ отсчётов). Прибавив эту ошибку к прогнозу и продолжив выполнять НВП (повторяя итерации K -продолжения вплоть до требуемого значения τ), получим итоговый прогноз. Структурная идентификация модели (3) выполняется заданием правил выбора корректирующих кластеров, а параметрическая – методом МНК. Кластер K не рассматривается как потенциальный корректирующий кластер для корректируемого C в трёх случаях: при непопадании кластера K в заданную окрестность кластера C в рамках сетки UG ; при отставании временных рядов кластера K , то есть при наличии пары одномерных рядов (F^i, F^j) , $F^i \in F^K, F^j \in F^C$, в которой первый отстаёт (в смысле комментария к требованию R5 из главы 2) от второго; при его значительной неоднородности, что определяется путём анализа отношений пар ближайших сингулярных чисел в разложении ССА одномерных рядов кластера.

Отбор p наилучших кластеров среди потенциальных корректирующих осуществляется в результате решения серии дискретных задач оптимизации (по убыванию приоритета) с введёнными порогами на кластерные характеристики:

- 1) задача максимизации: число совпадений направлений градиентов кластеров K и C для промежутка времени длины τ , непосредственно предшествующего первой прогнозной точке t ;
- 2) задача минимизации: значение дисперсии кластера;
- 3) задача максимизации: значение коэффициента корреляции для пары кластеров (K, C) , определяемое как минимум по всем возможным парам одномерных временных рядов из разных кластеров.

Уточнение при $p = 0$ названо *автокоррекцией*. Для различения метода с коррекцией и автокоррекцией используются аббревиатуры $UBII$ -ARX и $UBII$ -AR соответственно; далее полагается, что оба варианта включают НВП.

Прогноз для всей исследуемой области строится в результате выполнения трёх этапов (переход на следующий этап осуществляется, когда выполнен предыдущий для всех кластерных рядов): НВП, расчёт кластерных характеристик и УВП. Сделанная оценка временной сложности позволяет утверждать, что

метод кластеризации реализуется за $O(s^2 \log_2 s)$ операций, а этапы прогнозирования – за $O(T^2 s^2)$, поэтому методика может применяться в реальном времени (прогноз строится за время, существенно меньшее 24 ч – периода обновления данных в международных центрах обработки океанографической информации).

Четвёртая глава посвящена описанию разработанного комплекса программ и представлению результатов экспериментальных исследований, выполненных для сгенерированных и реальных данных.

Комплекс программ, реализующий авторскую методику, состоит из библиотеки ядра, консольных утилит и графической оболочки. Оболочка предоставляет функции как автоматизации оперативного извлечения информации – определения значений и визуализации полей физической характеристики в виде плоских прямоугольных карт или трёхмерных поверхностей с возможностью вывода значения характеристики для выбранного узла карты, так и проведения прогностических экспериментов. Стремление к высокой степени автоматизации обработки временных рядов с целью прогноза не исключает существенной роли экспертов в процессе использования данной методики, поэтому доступен ряд параметров, для которых имеются априорно заданные (рекомендуемые) значения, а также представлена возможность задания других конкретных значений в результате ручной настройки. Таким образом, методика реализована в виде интерактивной (на каждом шаге обработки данных) и удобной в использовании компьютерной технологии. Программное обеспечение создано в среде *Eclipse* на языке *Java*, отдельные модули, использующие картографию, написаны на языке *Python*.

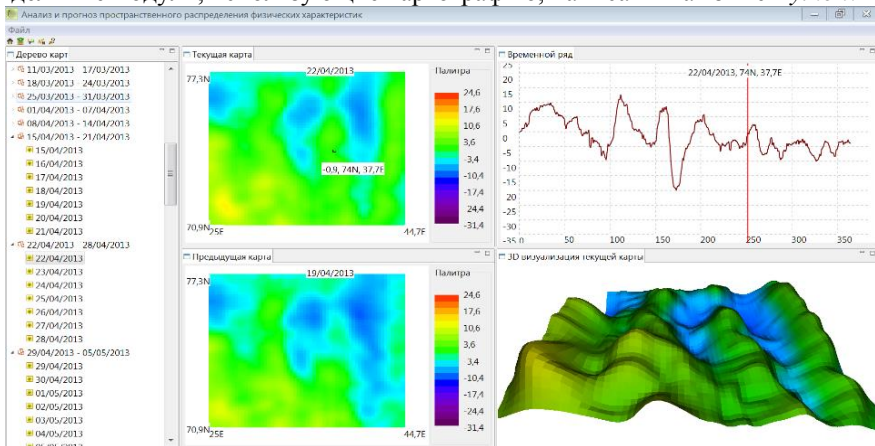


Рисунок 2. Скриншот окна графической оболочки

Вычислительные эксперименты со сгенерированными данными. Массив данных построен для сетки *UG* размером (30×20) , временной ряд имеет длину $N = 350$ отсчётов. Модель полезного сигнала одномерного ряда имеет вид:

$$y(t) = P_4(\vec{t}) + \sin \omega \vec{t}, \quad (4)$$

где для определённости

$$P_4(\tilde{t}) = 2,7356\tilde{t}^4 - 0,0714\tilde{t}^3 - 10,106\tilde{t}^2 - 4,0196\tilde{t} + 19,0154,$$

$$\tilde{t} = t / 90 - 2, \quad t = 0, \dots, 349, \quad \omega = 2,5\pi.$$

Одномерные ряды – это либо полезный сигнал (4), либо его линейная аппроксимация $f(t)$, либо сигнал с шумом $x(t)$:

$$x(t) = y(t) + z(t), \quad z(t) \sim N(0, \sigma^2), \quad \sigma = 0,7 \Rightarrow M(|z(t) / y(t)|) = 0,05. \quad (5)$$

На рисунке 3 приведены априорная (а) и апостериорная (б) системы кластеров: модельный кластер №7 разделился на пять (№№ 7-11) независимых кластеров; этот факт не противоречит исходным данным, так как ряды этого модельного кластера характеризовались наибольшим уровнем шума (до 10%).



Рисунок 3. Сгенерированные данные: а) априорная система кластеров, б) апостериорная система кластеров, в) примеры профилей рядов, г) осреднённая по кластерам ошибка прогноза (тонкая линия – НВП, толстая – УВП-ARX)

Прогнозирование выполнялся независимо для целевого временного интервала I , содержащего 20 последовательных временных отсчётов (с 220 по 239, рисунок 3в) на горизонт $\tau = 5$. Для данного примера оценкой оптимальной длины фрагмента временных рядов для прогноза с ошибкой ε не более 10% по условию τ - ε -стационарности (Ю.Н. Орлов, К.П. Осминин) стало значение, близкое к 20 ($T = T_{opt} = 20$). Результаты экспериментов для НВП и УВП-ARX представлены на рисунке 3г в виде профилей средней ошибки по формуле (2) для каждого момента интервала I . Коррекция УВП-ARX наиболее эффективна в случае продолжения монотонного профиля временного ряда (полезного сигнала) и сохранения знака второй производной, а в окрестности точки перегиба оптимален НВП.

Вычислительные эксперименты с реальными данными (рисунок 4)

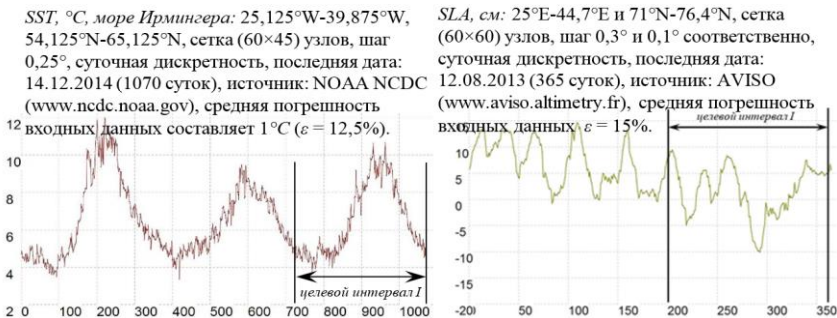


Рисунок 4. Параметры массивов входных данных, целевые интервал I и примеры профилей рядов для SST и SLA.

Для анализа методики выполнены два прикладных прогнозных исследования: для массивов SST (температуры поверхности) в акваториях моря Ирмингера и Баренцева моря, а также для массивов SLA (аномалий уровня моря) в акватории Баренцева моря. В каждом случае для тестирования выбран интервал I (рисунок 4), стартовые отсчёты устанавливались с пятидневным шагом. Таким образом, для SST (Баренцево море и море Ирмингера) количество независимых экспериментов по каждому методу прогнозирования составило 65, для SLA – 32.

Для массива SST в акватории моря Ирмингера была поставлена задача анализа методики в формате независимых экспериментов для отсчётов с 731 по 1070 (целевой интервал I) и горизонта $\tau = 5$. Построение системы кластеров выполнено по двухгодичный фрагменту исходного многомерного ряда, предшествующему по времени интервалу I ; результат (57 кластеров, рисунок 5) согласуется с физическими особенностями акватории: в северо-западной области имеет место повышенная дифференциация кластеров ввиду неоднородностей в динамике SST , вызванных перемешиванием вод тёплого и холодного течений.

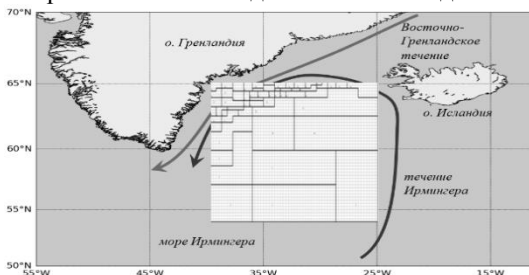


Рисунок 5. Система кластеров и географические особенности акватории

В результате исследования τ - ε -стационарности было установлено, что для большинства кластеров T_{opt} близко к 60. В дополнение к трём методам прогнозирования, а именно: МССА и авторским УВП-ARX и УВП-AR, рассмотрен вариант обновления прогностических результатов при оперативном поступлении данных для одного временного отсчёта (первого прогнозного). Средние и наибольшие ошибки прогноза SST по всем кластерам и всем точкам целевого интервала I без учёта (δ_1 и δ_{m1}) и с учётом (δ_2 и δ_{m2}) значений для одного дополнительного (первого прогнозного) временного отсчёта, а также среднеквадратичные отклонения σ каждой величины указаны в таблице ниже (для кластеров по отдельности вычисляются средние ошибки на всём целевом интервале по формуле (2), затем по ним определяется среднее и наибольшее значения (а также σ) как скалярный результат для целевой области – акватории моря Ирмингера).

№	Ошибка, % Метод	δ_1	σ_1	δ_2	σ_2	δ_{m1}	σ_{m1}	δ_{m2}	σ_{m2}
1	МССА	14,5	2,2	13,1	1,9	17,2	2,7	16,8	2,5
2	УВП-AR	11,5	1,8	11,2	1,8	16,7	2,5	16,1	2,3
3	УВП-ARX	10,3	1,9	9,8	1,8	15,9	2,5	15,3	2,4

Представленные в таблице значения позволяют сделать вывод о том, что предложенная коррекция результатов МССА одного кластера путём учёта ошибок прогноза в других кластерах целесообразна и эффективна. Также экспериментально подтверждено, что разработанный метод НВП с МДЭМ для исключения первой (шумовой) компоненты разложения показывает лучшие результаты, чем известный метод МССА и Д-МССА, то есть предложенный в данной работе вариант его модификации. Аналогичные выводы могут быть сделаны на основании полученных в ходе вычислительных экспериментов результатов прогнозирования температуры и аномалий уровня в акватории Баренцева моря. Также отметим, что ошибки прогноза согласованы с точностью исходных данных.

В заключении сформулированы полученные в диссертации научные и практические результаты.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

В ходе исследований проведён содержательный анализ существующих методов прогнозирования многомерных временных рядов, обоснована актуальность использования современного аппарата теории анализа временных рядов для решения поставленной задачи, разработана и экспериментально исследована оригинальная методика краткосрочного прогнозирования значений физической характеристики в узлах планарной равномерной сетки. Достоинствами методики являются: широкая область применения безотносительно природы данных, возможность построения прогноза значений нестационарного многомерного ряда в режиме реального времени, наличие программной реализации, что даёт возможность использования методики на практике. В качестве реальных данных для анализа были выбраны массивы аномалий уровня и температуры поверхности в акваториях Баренцева моря и моря Ирмингера.

К основным результатам относятся следующие:

1. Предложена модель пространственно-временной изменчивости целевой характеристики как основа авторской методики построения краткосрочного прогноза. Её актуальность обусловлена отсутствием известных гидродинамических моделей (в случае моря Ирмингера) или их недостаточной практической результативностью (для удалённых от берега акваторий Баренцева моря).

2. Разработана методика прогнозирования как совокупность методов, осуществляющих идентификацию модели и построение прогноза:

2.1. Метод кластеризации узлов равномерной сетки и формирования кластерных временных рядов. Классический кластерный анализ не применим в решаемой задаче, а кросскорреляционный анализ не учитывает пространственные особенности формирования многомерного ряда, поэтому авторский метод синтезирован на базе этих двух подходов.

2.2. Метод построения начального внутрикластерного прогноза (НВП) на базе структурных моделей МССА и МДЭМ. Метод прогнозирования в рамках МССА впервые дополнен автором практическими процедурами предобработки

по МДЭМ и выбора количества значимых аддитивных компонент в разложении многомерного ряда, критерием определения длины его фрагмента, по которому строится прогноз. Вычислительные эксперименты показали, что НВП позволяет получить лучшие результаты в смысле приведённой относительной ошибки, чем метод прогнозирования в рамках МССА.

2.3. Метод уточнения результатов начального внутрикластерного прогноза (УВП) для каждого кластера путём прогнозирования ошибки на основании известных ошибок для предыдущих временных отсчётов. В качестве модели ошибки используется модификация статистической авторегрессионной модели с внешними переменными, адаптированная автором для решения поставленной задачи; при её структурной идентификации используются введённые автором числовые характеристики кластеров.

3. Методика реализована в виде комплекса программ для проведения прогностических экспериментов и её применения для решения реальных практических задач. Для основных алгоритмов проведён анализ временной сложности: получена оценка количества операций как функция длины фрагмента временного ряда и числа узлов планарной равномерной сетки. Анализ показал, что методика может применяться в режиме реального времени (прогноз может быть построен за промежуток времени, существенно меньший периода обновления исходных данных, равного 24 часам, во внешних реально доступных источниках).

4. В результате вычислительных экспериментов подтверждена практическая эффективность применения методики как для сгенерированных, так и для реальных данных. Сравнительный анализ результатов прогнозирования, полученных по авторской методике и по МССА, показал, что авторский вариант имеет наименьшую приведённую относительную ошибку и она согласована с относительной погрешностью исходных данных.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Научные статьи, опубликованные в изданиях, рекомендованных ВАК:

1. Запорожцев, И.Ф. Использование искусственных нейронных сетей для прогнозирования временной изменчивости пространственного распределения характеристик океанической поверхности / И.Ф. Запорожцев, А.-В.И. Серда // Вестник МГТУ. Труды Мурманского государственного технического университета. – Мурманск, 2013. – Т. 16, № 4. – С. 708–714.

2. Запорожцев, И.Ф. Декомпозиция на эмпирические моды в задаче краткосрочного прогнозирования многомерных временных рядов геофизической природы / И.Ф. Запорожцев, А.-В.И. Серда // Цифровая обработка сигналов. – 2014. – № 2. – С. 34–40.

3. Запорожцев, И.Ф. Методика краткосрочного прогноза изменения значений характеристик динамической системы на основе численного анализа многомерных временных рядов / И.Ф. Запорожцев, А.-В.И. Серда // Цифровая обработка сигналов. – 2015. – № 1. – С. 17–24.

Работы, опубликованные в других изданиях:

4. Запорожцев, И.Ф. Экспериментальное исследование результатов прогнозирования многомерных временных рядов с использованием метода MSSA с предобработкой по EMD / И.Ф. Запорожцев, А.-В.И. Серeda // Наука и образование – 2013: материалы междунар. НТК, Мурманск, 4–11 марта 2013 г.– Мурманск: МГТУ, 2013. – С. 441-445.

5. Запорожцев, И.Ф. Прогноз временной изменчивости поля аномалий уровня поверхности моря в районе желоба Святой Анны с помощью нейросетевых методов // Проблемы Арктического региона: материалы XIII междунар. науч. конф., Мурманск, 14 мая 2013 г. Том II – Мурманск: ММБИ КНЦ РАН, 2013. – С. 42-44.

6. Запорожцев, И.Ф. Кластеризация узлов планарной регулярной сетки в задаче прогноза временной изменчивости пространственного распределения физических характеристик / И.Ф. Запорожцев, А.-В.И. Серeda // Наука и образование – 2014: материалы междунар. НТК, Мурманск, 24–28 марта 2014 г.– Мурманск : МГТУ, 2014. – С. 382-385.

7. Запорожцев, И.Ф. Методика краткосрочного прогноза поля аномалий уровня моря в акваториях Арктики // Проблемы Арктического региона: материалы XIV междунар. НТК, Мурманск, 14 мая 2014 г. Том II – Мурманск: ММБИ КНЦ РАН, 2014. – С. 65-66.

8. Запорожцев, И.Ф. Краткосрочный прогноз значений характеристик океанической поверхности в контексте решения задач рыбного промысла / И.Ф. Запорожцев, А.-В.И. Серeda // Проблемы и перспективы развития рыбохозяйственного комплекса на современном этапе: Доклады I междунар. конф. молодых учёных, г. Мурманск, 22-24 октября 2014г. – Мурманск : ФГБНУ «ПИНРО», 2014. – С. 63-64.

9. Zaporozhtsev, I.F. Analysis and Forecast of Sea Level Anomalies Spatio-Temporal Variability in the Barents Sea (Анализ и прогнозирование пространственно-временной изменчивости аномалий уровня в Баренцевом море) / I.F. Zaporozhtsev, A.-V.I. Sereda // Computer Modeling and Simulation : труды междунар. науч.-техн. конф., Санкт-Петербург, 2-4 июля 2014 года. – СПб. : Изд-во Политехн. ун-та, 2014. – С. 81-85.

10. Zaporozhtsev, I.F. Computer Technology for Short-Term Prediction of Sea Level Anomalies Spatial Distribution in the Arctic (Компьютерная технология краткосрочного прогнозирования пространственного распределения аномалий уровня моря в акваториях Арктики) / I.F. Zaporozhtsev, A.-V.I. Sereda // 2014 International Conference on Computer Technologies in Physical and Engineering Applications (ICCTPEA) : Proceedings, Saint-Petersburg, June 30 – July 4 2014 year. – P. 215-216.

11. Компьютерное моделирование, численный анализ и прогнозирование уровня поверхности океана в акваториях Арктики : отчёт о НИР (заключ.) / ФГБОУ ВПО «Мурманский государственный технический университет»; рук. А.-В.И. Серeda ; исполн. И.Ф. Запорожцев. – Мурманск, 2014. – 42 с. – Библиогр. : с. 41-42. – № ГР НИОКР 01201459892. – № ГР ИКРБС 215040170026. – Зарегистрирован в ЦИТиС 1 апреля 2015 г.

Свидетельство о государственной регистрации программы для ЭВМ:

12. Запорожцев, И.Ф. Библиотека для выполнения краткосрочного прогноза временной изменчивости пространственного распределения характеристик поверхности океана: Свидетельство о государственной регистрации программы для ЭВМ № 2015610857 / И.Ф. Запорожцев ; правообладатель ФГБОУ ВПО «МГТУ». – Зарегистрировано в Реестре программ для ЭВМ 20 января 2015 г.