

На правах рукописи

Кислицын Алексей Алексеевич

Моделирование индикаторов разрядки в нестационарных временных рядах
электроэнцефалограмм

Специальность 05.13.18 — математическое моделирование, численные методы
и комплексы программ

Автореферат диссертации на соискание ученой степени
кандидата физико-математических наук

Москва — 2021

Работа выполнена в ФГУ «ФИЦ ИПМ им. М.В. Келдыша РАН»

Научный руководитель: доктор физико-математических наук,
заведующий отделом вычислительной физики
и кинетических уравнений ИПМ
им. М.В. Келдыша РАН, Орлов Юрий
Николаевич

Официальные оппоненты: Шамаров Николай Николаевич, доктор
физико-математических наук, доцент кафедры
высшей математики Московского физико-
технического института (г. Москва)

Кочеткова Ирина Андреевна, кандидат
физико-математических наук, доцент кафедры
прикладной информатики и теории
вероятностей Российского университета
дружбы народов (г. Москва)

Ведущая организация: Институт теории прогноза землетрясений и
математической геофизики РАН (г. Москва)

Защита состоится «13» мая 2021 г. в « 11.00 » часов на заседании
диссертационного совета Д 002.024.03, при ИПМ им. М.В. Келдыша РАН,
125047, Москва, Миусская пл., д.4

С диссертацией можно ознакомиться в библиотеке и на сайте ИПМ
им. М.В. Келдыша РАН <http://keldysh.ru>

Автореферат разослан « ____ » _____ 2021 г.

Учёный секретарь диссертационного совета
к.ф.-м.н.

М.А. Корнилина



ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования.

Задача разработки высокопроизводительных численных алгоритмов и программных комплексов для обработки больших массивов нестационарных данных, возникающих во многих областях практической деятельности, в настоящее время приобрела большое практическое значение в связи с развившимися возможностями вычислительной техники и значительно увеличившейся статистической детализацией описания самих процессов. Одной из актуальных задач является проблема статистического распознавания разладки в нестационарных рядах. Для стационарных случайных процессов эта задача имеет решение в виде классических критериев – либо параметрических, когда оцениваются доверительные интервалы принадлежности параметра, характеризующего класс распределений наблюдаемой величины, либо непараметрических, основанных на близости выборочных функций распределения в различных нормах.

Анализ нестационарных процессов может быть проведен только при использовании высокопроизводительных вычислений, что связано с развитием так называемых систем с интенсивным программным обеспечением. Согласно Стратегии развития отрасли информационных технологий в Российской Федерации на 2014–2020 годы и на перспективу до 2025 года такими системами являются аппаратно-программные комплексы с большим удельным весом программной части. Тема настоящего исследования напрямую связана с обработкой данных, получаемых такими комплексами.

Существует много примеров временных рядов, в которых одним из инструментов анализа является модель разладки с учетом нестационарных свойств ряда. Это биржевые ряды цен сделок на финансовые инструменты, а также ряды данных о доходности соответствующих торговых операций, совершенных в рамках определенных стратегий (алгоритмов), где потоки событий составляют до нескольких десятков в секунду по отдельному инструменту. Другим примером являются потоки событий, связанных с анализом трафика беспроводной связи, достигают десятков тысяч в секунду. Менее мощные, но тоже достаточно большие потоки данных требуется обрабатывать при анализе различной биометрической информации о состоянии организма, а также телеметрической информации о состоянии технических систем. Анализ данных электроэнцефалограмм в медицине, статистика показаний сейсмограмм и различных счетчиков радиоактивности, последовательности символов в текстах различной природы представляют типичный набор задач, в которых результат наблюдения требуется отнести к определенным классам «норма» или «авария». Практически важным является определение промежуточного состояния, которое, собственно, и представляет собой разладку «нормы», чтобы успеть принять меры по недопущению состояния «авария». Модель разладки должна при этом давать вероятностную

картину перехода из одного состояния в другое, причем ошибка второго рода (пропуск цели) должна быть минимальна при заданном уровне ошибки первого рода (ложная тревога).

Для стационарных рядов функции распределения которых известны, вероятности ошибок при принятии или отклонении статистических гипотез могут быть вычислены теоретически. Однако если ряд нестационарный и его выборочные распределения не принадлежат определенному параметрическому классу, оценить ошибку затруднительно. Тем не менее, решения в таких случаях принимаются на основе классических критериев, только уровень значимости постфактум оказывается хуже, чем утверждается критерием.

В настоящей работе предложена модель непараметрического индикатора разладки для рядов, которые не являются стационарными, причем во многих практических случаях распределение этого индикатора стационарно. Это позволило корректно указать вероятность ошибки первого рода. Разработанный алгоритм и программный комплекс для вычисления этого индикатора направлены на анализ потоков данных высокой интенсивности, что является в настоящее время актуальным направлением развития практически всех отраслей науки и промышленности.

Степень разработанности научной проблемы.

Для нестационарных временных рядов общего подхода к решению задачи о разладке в настоящее время не существует. Есть результаты, полученные для частных случаев, когда ряды оказываются коинтегрированными. Тогда разладку можно искать с помощью классических стационарных критериев.

Для рядов, которые не сводятся с помощью регрессионного анализа к стационарным, на практике применяются недостаточно обоснованные адаптивные методы. В них предполагается, что ряды могут быть описаны некоторой стационарной моделью, но с течением времени параметры модели заново пересчитываются в скользящем временном окне с учетом сравнения предсказанного значения с фактом. Недостатком такого подхода для распознавания разладки является то, что ее можно спутать с изменением параметров модели, тогда как собственно разладкой в этом случае следует считать изменение вида самой модели.

Альтернативным подходом является применение кинетического метода, с помощью которого анализируется эволюция выборочных плотностей функции распределения. Этот подход к анализу нестационарных временных рядов развивается в настоящее время в ИПМ им. М.В. Келдыша РАН. Были предложены новые индикаторы нестационарности, такие как согласованный уровень стационарности и индекс нестационарности. Эти индикаторы использовались главным образом для классификации типов временных рядов, но для анализа разладки они ранее не применялись в силу увеличения размерности такой задачи. Таким образом, ни на уровне математической модели, ни в виде программного обеспечения индикаторы нестационарности ранее не рассматривались как предикторы разладки.

Цель и задачи диссертационного исследования.

Цель работы заключается в создании программного комплекса для обнаружения разладки в нестационарных временных рядах электроэнцефалограмм в реальном времени.

Научно-теоретическая задача состоит в разработке математической модели непараметрического индикатора – предиктора разладки в нестационарном временном ряде (п. 1 паспорта специальности 05.13.18).

В силу большой вычислительной сложности статистической процедуры следует также разработать приближенный метод анализа изучаемых статистик (п. 2 паспорта специальности 05.13.18).

Практическая задача заключается в создании численного алгоритма вычисления и тестирования индикатора в условиях многомерности потока событий (п. 4 паспорта специальности 05.13.18).

Объект и предмет исследования.

Объектом исследования является специальная статистика, получаемая по непересекающимся выборкам из нестационарных временных рядов, называемая согласованным уровнем стационарности.

Предмет исследования – свойства выборочных распределений этой статистики с целью выяснения возможности ее использования в качестве предиктора разладки в нестационарных временных рядах.

В качестве практического примера ряда, на котором тестируется соответствующий численный алгоритм, взят многомерный ряд записей электроэнцефалограмм по совокупности отведений для пациентов с приступами эпилепсии.

Теоретическая и методическая основа исследования.

В работе использовались кинетические методы анализа временных рядов, аппарат теории конечнократных аппроксимаций по формулам Фейнмана-Чернова, а также аналитические и численные методы решения уравнения Лиувилля.

Основу теоретического анализа составляют методы исследования асимптотических свойств стационарной точки уровня значимости, на котором принимается гипотеза об однородности двух выборочных распределений. Исходной статистикой для этого анализа является выборочная функция распределения расстояний между выборочными функциями распределения встык-выборки изучаемого временного ряда.

Информационная база исследования.

Информационная база исследования состоит из учебной литературы по теории вероятностей и математической статистике, а также статей в научных журналах и трудов международных конференций, относящихся к задаче моделирования разладки.

Обоснованность и достоверность результатов исследования.

Достоверность и обоснованность результатов подтверждается использованием строгих математических доказательств и рассуждений и апробированных в научной практике методов численного анализа.

Соответствие диссертации паспорту научной специальности.

Согласно вышеописанным основным целям и задачам данной диссертационной работы, она направлена на разработку математической модели явления разрядки, аппроксимирующего ее численного алгоритма и создания соответствующего программного комплекса. Тем самым диссертация соответствует пунктам 1, 2 и 4 паспорта специальности 05.13.18 – математическое моделирование, численные методы и комплексы программ.

Научная новизна результатов исследования.

Новизна работы заключается в том, что впервые теоретический метод построения эквивалентной по Чернову полугруппы был применен в математической статистике для обоснования алгоритма приближенного вычисления индикатора разрядки. Также впервые был проведен статистический анализ выборочных функций распределения этого индикатора, что являлось в вычислительном плане серьезным техническим затруднением, связанным с увеличением размерности задачи. Практический результат, полученный при тестировании предложенного метода на примерах электроэнцефалограмм, состоит в улучшении распознавания приближения приступа эпилепсии. Этот результат получен с использованием нового критерия разрядки, в качестве которого предложено использовать статистику, называемую согласованным уровнем стационарности ряда.

Теоретическая ценность и практическая значимость работы.

Теоретическая и практическая ценность работы состоит в том, что в ней построен алгоритм вычисления многомерного индикатора разрядки в нестационарном временном ряде, доказаны теоремы о свойствах его аппроксимации, позволяющие значительно сократить время счета, и построен программный комплекс с интерфейсом, реализующий предложенную методику.

Практическая значимость работы подтверждается также применением полученных результатов в проведении научно-исследовательских работ по тематике, утвержденной в ИПМ им. М.В. Келдыша РАН. Результаты исследований были использованы при выполнении проектов ИПМ им. М.В. Келдыша РАН, поддержанных грантами РФФИ (№ 14-21-00025, № 19-71-30004), в совместном проекте ИПМ им. М.В. Келдыша РАН и НМИЦ нейрохирургии им. ак. Н.Н. Бурденко (соглашение № 73 от 05.12.2018), а также в работе по проекту с ННГУ им. Н.И. Лобачевского «Надежный и логически прозрачный искусственный интеллект: технология, верификация и применение при социально-значимых и инфекционных заболеваниях» (договор № 111-20-ЕП от 16.11.2020). Практическая задача относилась к сфере анализа больших данных и была связана с построением предиктора разрядки применительно к записям электроэнцефалограмм для пациентов с приступами эпилепсии.

Апробация работы.

Материалы диссертации докладывались на научных семинарах отдела вычислительной физики и кинетических уравнений ИПМ им. М.В. Келдыша РАН, на кафедре прикладной информатики и теории вероятностей в Институте прикладной математики и телекоммуникаций РУДН, на рабочих совещаниях в НМИЦ нейрохирургии им. акад. Н.Н. Бурденко, а также на конференциях:

1. Gennady Bocharov, Alexey Kislitsyn, Rostislav Savinkov, Mario Novkovic, Lucas Onder. Modelling the FRC network of lymph node // International Congress on Systems Immunology, Immunoinformatics & Immune-computation, Taormina, Italy, July 17-18, 2015.

2. Kislitsyn A.A, Orlov Yu.N. On the Distribution of the Stationary Point of Significance Level for Empirical Distribution Function // International Congress on Ultra Modern Telecommunications and Control Systems ICUMT, Moscow, Russia, November 5-8, 2018.

3. Kislitsyn A.A, Orlov Yu.N. Nonstationary stochastic motion modeling by dynamical systems // ECMS, Caserta, Italy, June 11-14, 2019.

Публикации.

По материалам диссертации опубликовано 13 работ в рецензируемых научных изданиях, входящих в перечень ВАК, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени кандидата физико-математических наук по специальности 05.13.18 – математическое моделирование, численные методы и комплексы программ. Эти работы включают 4 препринта ИПМ им. М.В. Келдыша РАН, а также 9 статей, входящих в базы данных WoS и SCOPUS, из которых 4 статьи опубликованы в отечественных, а 5 – в иностранных журналах. Имеются также 1 патент на изобретение и 1 свидетельство государственной регистрации на программу для ЭВМ.

Личный вклад автора.

Автор диссертации, работая в коллективе соавторов, самостоятельно сформулировал теоретические утверждения, относящиеся к модели разрядки, разработал численный алгоритм и провел серию вычислительных экспериментов. Математическая постановка задач принадлежит научному руководителю. Физическая постановка задачи о предсказании приступа эпилепсии и экспериментальные временные ряды ЭЭГ принадлежат сотрудникам НМИЦ им. акад. Н.Н. Бурденко Козловой А.Б., Корсаковой М.Б. и Машерову Е.Л. В совместных публикациях лично соискателем получены следующие результаты.

В работах [1, 2, 3] соискателем проведены численные расчеты согласованного уровня стационарности и дана оценка точности предиктора.

В работах [5, 6, 7] соискателем проведены работы по созданию трехмерной визуализации результатов расчетов статистических характеристик применительно к задачам математического моделирования в медицине.

В работах [8, 9, 12, 13] соискателю принадлежит модель разрядки и доказательства теорем, относящихся к аппроксимации решения уравнения Лиувилля и к аппроксимации стационарной точки эмпирического уровня значимости распределений изучаемых статистик.

В работах [10, 11] соискателем разработан численный алгоритм и построена модель аппроксимации решения уравнения Лиувилля.

Вклад соискателя в патент [14] состоит в проведении верификационных статистических исследований

Вклад соискателя в программный комплекс [15] состоит в разработке архитектуры комплекса, программировании вычислительного алгоритма и интерфейса пользователя.

Основные положения, выносимые на защиту.

1. Разработана математическая модель индикатора разрядки нестационарного временного ряда на основе статистики согласованного уровня стационарности (п. 1 паспорта специальности 05.13.18).
2. Разработан метод приближенного вычисления значений индикатора разрядки на основе использования формул Фейнмана-Чернова конечнократной аппроксимации полугрупп (п. 2 паспорта специальности 05.13.18).
3. Построен численный алгоритм для определения многомерного индикатора разрядки в реальном времени и программный комплекс с пользовательским интерфейсом (п. 4 паспорта специальности 05.13.18).
4. Проведены численные эксперименты по индикации разрядки применительно к рядам электроэнцефалограмм для пациентов с приступами эпилепсии и показано, что построенный индикатор имеет стационарное распределение, а его применение характеризуется ошибкой первого рода, равной значению индикатора разрядки, и нулевой эмпирической ошибкой второго рода (п. 4 паспорта специальности 05.13.18).

Структура и объем диссертации.

Диссертация «Моделирование индикаторов разрядки в нестационарных временных рядах электроэнцефалограмм» состоит из введения, четырех глав и заключения. Каждая глава разбита на параграфы, имеющие двойную нумерацию, первая цифра которой указывает на соответствующую главу. Формулы внутри каждого параграфа имеют тройную нумерацию, с указанием на главу и параграф. Рисунки и таблицы имеют двойную нумерацию с указанием на номер главы.

Результаты исследования изложены на 94 страницах и содержат 40 рисунков и 1 таблицу. Библиографический список состоит из 75 наименований, расположенных в алфавитном порядке. В автореферате сохранена нумерация формул и рисунков в соответствии с диссертацией.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении проводится обзор современных направлений исследований в области построения статистических индикаторов разладки для нестационарных временных рядов. В результате обзора обосновываются и формулируются задачи, решению которых посвящена диссертация.

В первой главе даются основные определения и свойства индикатора, называемого согласованным уровнем стационарности (далее СУС). Для стационарных рядов этот индикатор представляет собой стационарную точку уровня значимости ε в критерии Колмогорова-Смирнова, которая определяется формулой

$$1 - K\left(\sqrt{\frac{N}{2}}\varepsilon\right) = \varepsilon, \quad (1.1.6)$$

где $K(z)$ есть функция Колмогорова. Решение этого уравнения $\varepsilon = \varepsilon^*(N)$ обладает тем свойством, что вероятность превышения значения $\varepsilon^*(N)$ как значимость соответствующего критерия равна расстоянию между выборками в норме C . Далее анализируется статистика расстояний в норме C между так называемыми встык-выборками изучаемого ряда, т.е. между выборочными функциями распределения (далее ВФР) $F_N(x, t)$ и $F_N(x, t + N)$, сдвинутыми одна относительно другой на величину окна выборки:

$$\rho(N; t) = \|F_N(x, t) - F_N(x, t + N)\| \quad (1.1.7)$$

и строится функция распределения $G_N(\rho)$ расстояний (1.1.7), которая представляет эмпирическую вероятность того, что расстояние между распределениями не больше ρ . Индикатор СУС $\rho^*(N)$ определяется так, что соответствующее расстояние равно значимости критерия, т.е. СУС является решением относительно ρ уравнения

$$G_N(\rho) = 1 - \rho. \quad (1.1.8)$$

Вводится также индекс нестационарности временного ряда, определяемый как отношение решения $\rho^*(N)$ уравнения (1.1.8) к стационарной точке $\varepsilon^*(N)$ уровня значимости стационарного критерия (1.1.6).

$$J(N) = \frac{\rho^*(N)}{\varepsilon^*(N)}. \quad (1.1.9)$$

В современных программах для идентификации разладки в основном используются параметрические методы оценивания, т.е. определяются доверительные интервалы, которым принадлежит оцениваемый параметр распределения изучаемой случайной величины, функциональный вид распределения которой считается известным.

Из непараметрических методов оценивания близости распределений как таковых, а не функционалов от них (близость средних, дисперсий и т.п.),

программно реализованы лишь классические стационарные методы, использующие критерии Колмогорова-Смирнова (норма C), Джини (норма L_1), Крамера – фон Мизеса (норма L_2) и разные взвешенные нормы, применяемые в критериях Андерсона-Дарлинга, Ватсона, Фроцини.

Для анализа нестационарных и не коинтегрированных рядов имеется только один программный комплекс, разработанный в ИПМ им. М.В. Келдыша РАН Ю.Н. Орловым и С.Л. Федоровым, в котором реализован расчет СУС в зависимости от длины выборки для анализа пошаговой эволюции выборок внутри заданного одномерного фрагмента данных. Для заданной длины встык-выборки этот алгоритм выдает ровно одно число. После вариации длины выборки результатом работы алгоритма является график СУС в виде $\rho^*(N)$, где N меняется в заданных пределах в соответствии с длиной L_{tot} полного набора анализируемых данных.

Для задачи распознавания разладки с указанием вероятности ошибки первого рода (ложной тревоги) необходимо дополнить вышеуказанный численный алгоритм расчета СУС возможностью проводить анализ по различным фрагментам исходного ряда, т.е. в качестве выходных данных он должен выдавать функцию распределения случайной величины СУС на основе статистики значений последовательности СУС $\rho_n^*(N, L)$, построенной для встык-выборок длины N в окнах длины L . Таким образом, Задача 1 при разработке программного комплекса формулируется как увеличение размерности блока статистического анализа СУС.

Кроме того, надо предусмотреть возможность многоканального анализа, для чего требуется провести распараллеливание алгоритма. Это важно в задачах анализа высокочастотных биометрических данных, которые записываются одновременно по нескольким отведениям (вплоть до нескольких тысяч различных рядов). Следовательно, возникает Задача 2 об использовании высокопроизводительных вычислений и проведении оценки их эффективности на основе вычислительных экспериментов для конкретных прикладных задач.

Задача 3 состоит в разработке метода сокращенных вычислений для построения статистики СУС, поскольку в условиях больших данных алгоритмы, основанные на полном переборе, неэффективны, а часто и не реализуемы в силу естественно ограниченных возможностей суперкомпьютеров.

Решение этих трех задач реализовано в программном комплексе NSSAT, который является практическим результатом данной диссертационной работы.

Результаты первой главы представлены в работах [3, 5, 6, 7, 14].

Во второй главе рассмотрены элементы теории конечнократных аппроксимаций полугрупп на основе теоремы Чернова. Вводятся следующие понятия эквивалентности по Чернову и средней полугруппы.

Операторнозначная функция $\hat{f}(t)$, действующая из некоторой правой полукрестности нуля на числовой оси в банахово пространство $B(X)$

линейных ограниченных операторов, действующих в банаховом пространстве X , называется эквивалентной по Чернову полугруппе $\hat{g}(t)$, если $\forall T > 0$ и $\forall u \in X$ выполняется условие:

$$\lim_{n \rightarrow \infty} \sup_{t \in [0; T]} \left\| \left(\hat{f}(t/n) \right)^n - \hat{g}(t) \right\|_X = 0. \quad (2.1.3)$$

Сильно непрерывная однопараметрическая полугруппа U ограниченных линейных преобразований банахова пространства X называется обобщенным средним значением случайной полугруппы ξ , если полугруппа U эквивалентна по Чернову математическому ожиданию $M\xi$.

Далее анализируется эквивалентная по Чернову стационарная точка уровня значимости для совокупности выборок. Пусть по длине L_{tot} фактически известного фрагмента временного ряда построен СУС как решение уравнения (1.1.8) для встык-выборок длины N . Следует учесть, что $L_{tot} \gg 2N$, так что на длине L_{tot} укладывается достаточно много длин N , чтобы можно было набрать статистику для построения распределения $G_N(\rho)$. На этой длине L_{tot} в доле случаев $\rho^*(N)$ расстояние между встык-выборками больше, чем $\rho^*(N)$. Идея идентификации разладки состоит в том, что если теперь в определенном скользящем окне длины L доля расстояний между встык-выборками длины N оказалась больше, чем вычисленная ранее величина $\rho^*(N)$, то в этом окне фиксируется разладка. Для такого вывода требуется, чтобы СУС нестационарного ряда был бы случайной величиной, имеющей стационарное распределение.

Исследуя локальное значение СУС – то есть значение, полученное по промежутку времени L , существенно меньшему, чем первоначальный объем данных L_{tot} , естественно возникает вопрос о флуктуациях этого локального СУС относительно значения, отвечающего множеству L_{tot} . Важно понимать, что СУС всего множества $\rho_{tot}^*(N)$ не является средним значением последовательности СУС $\rho_n^*(N, L)$, построенной для встык-выборок длины N в окнах длины L . Рассмотрим функцию

$$\Psi_N(\rho) = 1 - G_N(\rho). \quad (2.3.2)$$

Предположим, что ВФР $G_N(\rho)$ аппроксимирует дифференцируемую функцию распределения $G(\rho)$ соответствующей генеральной совокупности. Поскольку $G_N(\rho)$ монотонно возрастает от 0 до 1, то функция $\Psi_N(\rho)$ обладает свойствами:

$$\Psi_N(0) = 1, \Psi'_N(0) = -a_N \leq 0.$$

В таком случае существует предел

$$\lim_{n \rightarrow \infty} \left(\Psi_N \left(\frac{\rho}{n} \right) \right)^n = e^{-\rho a_N} \equiv \Phi_N(\rho). \quad (2.3.3)$$

Следуя определениям, данным выше для операторных функций, будем называть предельную функцию $\Phi_N(\rho)$ эквивалентной по Чернову функции $\Psi_N(\rho)$ (то есть уровню значимости распределения $G_N(\rho)$). Будем обозначать эту эквивалентность как $\Phi_N(\rho) \stackrel{Ch}{\sim} \Psi_N(\rho)$.

Согласно теореме об эквивалентности по Чернову средней полугруппы, если имеется (конечный или бесконечный) набор функций $\Psi_N^k(\rho)$ вида (2.3.2), каждая из которых эквивалентна в смысле (2.3.3) функции $\Phi_N^k(\rho)$ с коэффициентом $-a_N^k$ в показателе экспоненты, и если задан соответствующий набор неотрицательных коэффициентов p_k таких, что $\sum_k p_k = 1$, то средняя функция $\bar{\Psi}_N(\rho) = \sum_k p_k \Psi_N^k(\rho)$ эквивалентна в смысле (2.3.3) функции $\bar{\Phi}_N(\rho) = e^{-\rho \bar{a}_N}$, где

$$\bar{a}_N = \sum_k p_k a_N^k. \quad (2.3.4)$$

Рассмотрим далее последовательность непересекающихся промежутков длины L . Для каждого k -го промежутка построим эмпирическое распределение $G_N^k(\rho, L)$ расстояний между встык-выборками длины N . Пусть таких промежутков n , так что $nL = L_{tot}$ есть полная длина изучаемого фрагмента ряда. Тогда распределение расстояний, построенное по всему объему данных, есть среднее распределение, полученное усреднением по отдельным выборкам:

$$G_N(\rho, L_{tot}) = \frac{1}{n} \sum_{k=1}^n G_N^k(\rho, L), \quad (2.3.5)$$

$$\bar{\Psi}_N(\rho) = \sum_k p_k \Psi_N^k(\rho) = 1 - G_N(\rho, L_{tot}), \quad p_k = \frac{1}{n}.$$

Пусть $\rho_k^*(N)$ есть стационарная точка для функции $\Psi_N^k(\rho)$, а $\tilde{\rho}_k(N)$ - для эквивалентной ей функции $\Phi_N^k(\rho)$. В силу того, что $a_N^k \geq 0$, эта стационарная точка единственная. Обозначим через $\tilde{\rho}(N)$ стационарную точку эквивалентной средней функции $\bar{\Phi}_N(\rho) = e^{-\rho \bar{a}_N}$. С точностью до $o(\rho)$ имеем из (2.3.3), (2.3.4):

$$\tilde{\rho}_k(N) = \frac{1}{1 + a_N^k}, \quad \tilde{\rho}(N) = \frac{1}{1 + \bar{a}_N} = \frac{1}{\frac{1}{n} \sum_{k=1}^n \frac{1}{\tilde{\rho}_k(N)}}. \quad (2.3.6)$$

Поскольку эквивалентные функции отличаются от оригиналов на величину $o(\rho)$, то на такую же величину отличаются и соответствующие стационарные точки.

Таким образом, в работе доказана следующая теорема о стационарных точках распределений с непрерывными плотностями.

Теорема 2.4. Пусть распределения случайных величин имеют непрерывные плотности и пусть на множестве этих случайных величин задана некоторая неотрицательная мера. Тогда стационарная точка функции, эквивалентной по Чернову среднему уровню значимости данных распределений, с точностью до бесконечно малой второго порядка совпадает с обратной величиной к среднему значению обратных величин стационарных точек функций, эквивалентных по Чернову уровням значимости данных распределений:

$$\frac{1}{\tilde{\rho}(N)} = \sum_{k=1}^n \frac{p_k}{\tilde{\rho}_k(N)} \equiv \frac{1}{\bar{\rho}(N)}. \quad (2.3.7)$$

В силу сказанного ранее соотношение (2.3.7) с точностью $o(\rho)$ может быть распространено и на стационарные точки распределений $\Psi_N^k(\rho)$. Формула (2.3.7) принципиальна, поскольку она позволяет значительно уменьшить количество вычислительных процедур, поскольку для каждой длины N встык-выборок СУС вычисляется всего один раз в наименьших доступных для этой цели непересекающихся окнах длины L , покрывающих все исследуемое множество данных L_{tot} .

Такой подход существенно сокращает время расчета в задачах анализа Больших Данных. Он позволяет оценивать важную статистическую характеристику нестационарного временного ряда по аппроксимационной процедуре, сходящейся к точному значению с гарантированной точностью.

Результаты второй главы опубликованы в работах [2, 8, 9, 10, 11, 12, 13].

В третьей главе приведены вычислительные алгоритмы, применяемые для вычисления индикатора разладки. Блок-схема алгоритма для вычисления СУС как предиктора разладки приведена на Рис. 3.2.

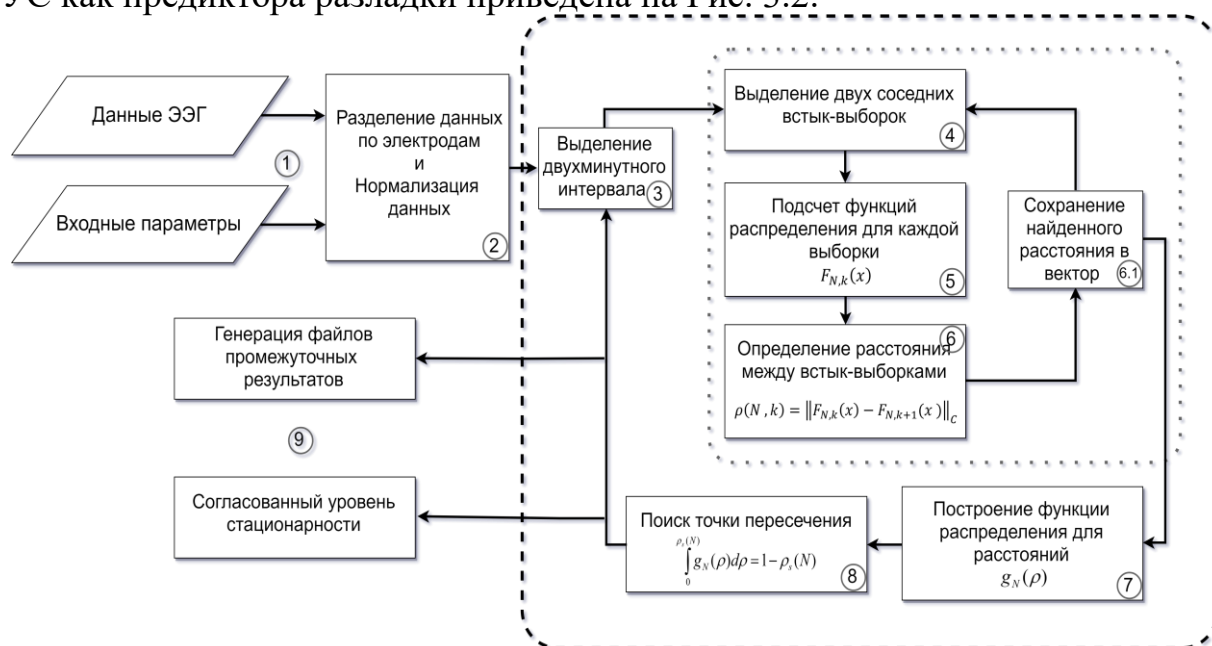


Рис. 3.2 – Блок-схема вычислительного алгоритма СУС

При обработке большого количества векторов данных применялось распараллеливание вычислений. Была проведена программная оптимизация времени вычисления описанного выше алгоритма с использованием интерфейса OpenMP. Расчеты производились на 2, 4, 6 и 8 параллельных потоках для двух вариантов входных данных, которые представляли собой 64 временных ряда длиной 690000 и 2340000 данных (стандартные размеры файлов данных, используемых в исследовании).

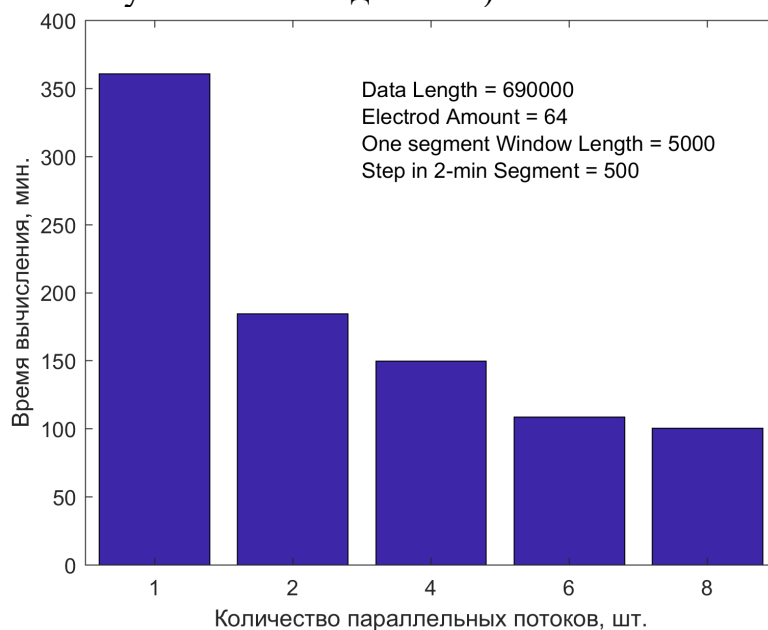


Рис. 3.6 – Время работы алгоритма при разном числе параллельных потоков

Программный комплекс имеет возможность работы в двух режимах: во-первых, подбор оптимальной ширины окна для дальнейшего анализа нестационарного случайного процесса; во-вторых, реализация расчета СУС по экспериментальным данным, с задаваемыми параметрами разбиения выборок данных.

В целях оптимизации процесса расчета, а также для целей практического использования, был разработан интерфейс для программного комплекса.

Рабочее поле пользовательского интерфейса для программного комплекса делится на несколько областей (Рис. 3.8): область меню, информационная область, область задания инициализирующих параметров, графическая область и область информирования пользователя.

В пункте меню «Файл» находятся элементы управления, относящиеся к работе с текущей сессией расчетов: пункт «Открыть файл данных» позволяет выбрать файл исходных данных определенного формата; пункт «Очистить форму» производит очистка памяти приложения, сброс всех настроек и очистка графической области; пункт «Сменить режим расчета» меняет алгоритм расчета и поля в области задания инициализирующих параметров.

В пункте меню «График» находятся элементы управления, относящиеся к работе с графической областью, где отображаются изображения, построенные по полученным результирующим данным; выбрав подпункт «Сохранить график», можно сохранить построенный график из результирующих данных.

Пункт меню «Справка» позволяет получить детальную информацию о каждом элементе управления программного комплекса.

Пункт меню «Отладочная консоль» позволяет вызвать командную консоль программного комплекса для детального анализа отладочной информации по происходящим расчетам.

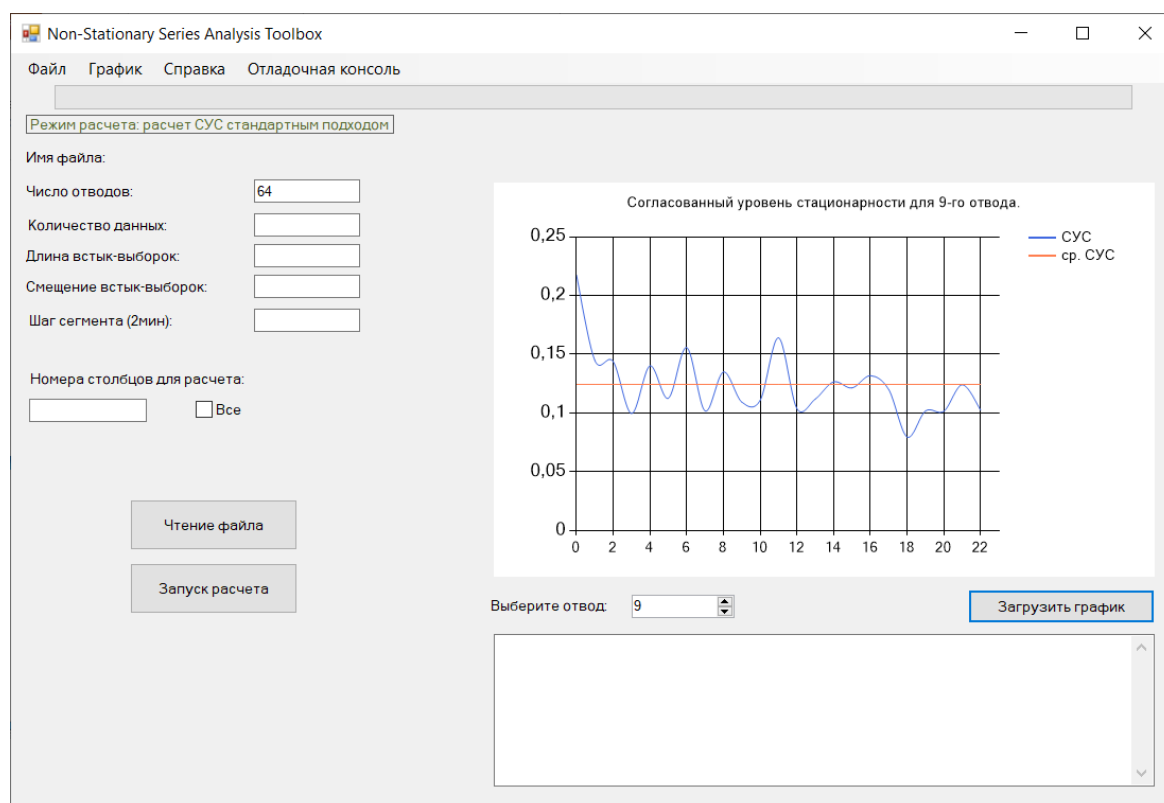


Рис. 3.8 – Скрин-шот рабочего окна программы

Далее приводится список полей в области задания инициализирующих параметров при работе программного комплекса в режиме расчета СУС с задаваемыми параметрами разбиения выборок данных. Число отводов – это количество столбцов данных в исходном файле данных. Количество данных – это количество точек в столбце данных; все столбцы должны быть одинаковой длины в одном файле. Длина встык-выборки – это ширина окна, в котором считается функция распределения. Смещение встык-выборок – выборки могут наслаиваться друг на друга для повышения качества вычисления СУС. Шаг сегмента (2мин) – размер выборки в котором вычисляется 1 точка СУС.

Кнопка «Чтение файла» используется для того, чтобы загрузить файл с исходными данными в память программного комплекса для дальнейших расчетов.

Кнопка «Запуск расчета» позволяет после заполнения всех необходимых полей в области задания инициализирующих параметров запустить расчет.

Кнопка «Загрузить график» выполняет построение графика по результирующему файлу, номер которого отвечает номеру столбца обработанных данных и может быть выбран в поле «Введите отвод».

Вычислительный алгоритм и пользовательский интерфейс написаны на языке C++/CLI в среде разработки MS Visual Studio 2017.

Результаты третьей главы опубликованы в [1, 15].

В четвертой главе приводятся результаты численного эксперимента по индикации разрядки в виде отклонения СУС от своего ожидаемого значения для прогноза наступления приступа эпилепсии в многомерных временных рядах, полученных из записей электроэнцефалограмм (далее ЭЭГ) по совокупности отведений.

Идея идентификации разрядки как предвестника приступа эпилепсии в нестационарном потоке данных ЭЭГ основывается на специфическом «законе сохранения» в виде функционала, построенного на траектории изучаемого случайного процесса и такой, что ее выборочное распределение, в свою очередь, стационарно. Таким функционалом и является СУС. В работе показано, что оптимальная длина встык-выборки составляет 5000 данных. Идентифицировано 8 отведений, по которым наблюдается существенное превышение СУС в окрестности приступа, причем распределение СУС оказалось стационарным. Выяснилось, что приступ эпилепсии всегда проходит на фоне индикации разрядки, каковой считается превышение в двухминутном окне сканирования уровня СУС промежутка в целом. Описанная методика была протестирована на ЭЭГ десяти пациентов. В общей сложности были проанализированы 20 достаточно длинных фрагментов, оканчивающихся приступом. Критический уровень СУС для разных пациентов оказывался различным, причем после приступа он существенно менялся и для одного пациента. Однако выборки длины 5000 каждый раз оказывались оптимальны для обнаружения разрядки.

Результаты четвертой главы опубликованы в работах [1, 2, 8, 10].

В заключении подытоживаются основные результаты диссертации и обсуждаются возможные области их применения, указываются ограничения построенной модели и возможности ее совершенствования.

ВЫВОДЫ И РЕКОМЕНДАЦИИ

В работе построен программный комплекс для вычисления многомерного индикатора разрядки, который может служить предиктором наступления приступа эпилепсии. Для его построения использовалась нестационарная статистка распределения расстояний между выборочными функциями распределения ЭЭГ-сигнала. Было выяснено, что ряды данных по каждому отведению нестационарны. На наборах данных ЭЭГ реальных пациентов была подтверждена гипотеза о том, что существуют участки головного мозга,

нестационарная активность которых перед приступом возрастает. Этот факт позволил применить модель согласованного уровня стационарности для индикации разладки нестационарного случайного процесса. Были определены оптимальная длина встык-выборки для вычисления СУС и длина окна для фиксации разладки, и показано, что выборочная функция распределения СУС стационарна. Показана также эффективность аппроксимации эквивалентной стационарной точки уровня значимости по Чернову. Полученный результат может быть использован как составная часть программного обеспечения медицинской техники при стационарном наблюдении за пациентами.

ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в изданиях, входящих в перечень ВАК:

1. Kislitsyn A.A., Kozlova A.B., Masharov E.L., Orlov Yu.N. Numerical Algorithm for Self-consistent Stationary Level for Multidimensional Non-stationary Time-series // Keldysh Institute Preprints. 2017. № 124. 14 p.
2. Кислицын А.А., Козлова А.Б., Корсакова М.Б., Машеров Е.Л., Орлов Ю.Н. Стационарная точка уровня значимости для нестационарных функций распределения // Препринты ИПМ им. М.В. Келдыша. 2018. № 113. 20 с.
3. Ивченко А.Ю., Кислицын А.А., Орлов Ю.Н. Распределение SIR для траекторий детерминированного движения на примере метрополитена // Препринты ИПМ им. М.В. Келдыша. 2018. № 155. 21 с.
4. Кислицын А.А. Программный комплекс для анализа статистики согласованного уровня стационарности временных рядов // Препринты ИПМ им. М.В. Келдыша. 2020. № 26. 22 с.

Публикации в журналах, входящих в базы данных WoS и SCOPUS:

5. Kislitsyn A., Savinkov R., Novkovic M., Onder L., Bocharov G. Computational Approach to 3D Modeling of the Lymph Node Geometry // Communication. 2015. 3(2): 222 – 234.
6. Топтыгина А.П., Азиатцева В.В., Савкин И.А., Кислицын А.А., Семикина Е.Л., Гребенников Д.С., Алешкин А.В., Сулимов А.В., Сулимов В.Б., Бочаров Г.А. Прогнозирование специфического гуморального иммунного ответа на основании исходных параметров иммунного статуса детей, привитых против кори, краснухи и эпидемического паротита // Иммунология, 2015. Т. 36. № 1. С. 22 – 30.
7. Savinkov R., Kislitsyn A., Watson D., Loon R., Novkovic M., Onder L., Bocharov G. Data-driven modelling of the FRC network for studying the fluid flow in the conduit system // Engineering Applications of Artificial Intelligence. 2016, V. 62. DOI: 10.1016/j.engappai.2016.10.007
8. Кислицын А.А., Козлова А.Б., Корсакова М.Б., Орлов Ю.Н. Индикатор разладки для нестационарных случайных процессов // Доклады РАН, сер. математическая, 2019. Т. 484. № 4. С. 393-396.

9. Kislitsyn A.A, Orlov Yu.N. Chernoff approximation for non-stationary random walk modeling // Lobachevsky Journal of Mathematics, 2019. Vol. 40. No 12. P. 2095-2102.
10. Kislitsyn A.A, Orlov Yu.N. On the Distribution of the Stationary Point of Significance Level for Empirical Distribution Function / IEEE Transactions, 2019. DOI:10.1109/ICUMT.2018.8631234.
11. Kislitsyn A.A, Orlov Yu.N. Nonstationary stochastic motion modeling by dynamical systems // ECMS Conference Proc., 2019. P. 466-472. DOI: 10.7148/2019-0466.
12. Кислицын А.А., Орлов Ю.Н. Моделирование эволюции выборочных распределений случайных величин с помощью уравнения Лиувилля // Математическое моделирование, 2020. Т. 32. №1. С. 111-128.
13. Kislitsyn A.A, Orlov Yu.N. Dynamical System Model with the use of Liouville Equation for Empirical Distribution Function Densities // Discontinuity, Nonlinearity and Complexity, 2020. Vol. 9. No 4. P. 529-540. DOI: 10.5890/DNC.2020.12.006

Патенты и свидетельства:

14. Патент № 2599506, 10.10.2016. Способ прогнозирования первичных и вторичных вакцинальных неудач при вакцинации против вирусов кори, краснухи и эпидемического паротита у детей с помощью вакцины Приорикс и способ персонифицированного подхода к коррекции вакцинальных неудач. Авторы: Топтыгина А.П., Азиатцева В.В., Кислицын А.А., Бочаров Г.А.
15. Свидетельство о регистрации программы для ЭВМ «Программный комплекс NSSAT (Non-Stationary Series Analysis Toolbox) для определения и визуализации характеристик нестационарности временных рядов». Правообладатель: ИПМ им. М.В. Келдыша РАН. Авторы: Кислицын А.А., Орлов Ю.Н. Свидетельство о государственной регистрации № 2019660374 от 05.08.2019.

