

"Утверждаю"
Директор ФГБУН Институт теории прогноза
землетрясений и математической геофизики РАН

Член-корреспондент РАН П.Н. Шебалин

13 апреля _____ 2021 г.



Отзыв ведущей организации о диссертации Кислицына Алексея Алексеевича

«Моделирование индикаторов разладки в нестационарных временных рядах электроэнцефалограмм», представленной на соискание ученой степени кандидата физико-математических наук по специальности 05.13.18 -математическое моделирование, численные методы и комплексы программ.

Диссертационная работа А.А. Кислицына посвящена численному моделированию непараметрического индикатора разладки для рядов, которые не являются стационарными. Исследование имеет как теоретический, так и прикладной характер. Соискателем предложена методика индикации разладки в нестационарных случайных процессах, что потребовало теоретического исследования специальных статистик, названных им стационарной точкой уровня значимости и согласованным уровнем (не)стационарности (автор использует термин «согласованный уровень стационарности», но по нашему мнению, возможно точнее было бы говорить о согласованном или фоновом уровне нестационарности). Также к теоретической части исследования относится разработка приближенного алгоритма вычисления этих статистик в скользящих окнах наблюдения за процессом; алгоритм основан на теоремах о конечнократных аппроксимациях в теории полугрупп. Практическая часть работы направлена на анализ потоков данных высокой интенсивности, что является в настоящее время актуальным направлением развития практически всех отраслей науки и промышленности (так называемая, методика “big data”). Соискатель применяет свою методику на примерах больших массивов данных, получаемых с многоканальных отведений электроэнцефалограмм у больных эпилепсией. По-видимому, эта методика впервые применена к детальному анализу состояния пациентов за длительный период наблюдения, что является весьма важным шагом в указанной области. Соискателем предложена методика классификации состояний временного ряда по уровню нестационарности. Оказалось, что применительно к рассмотренным случаям заболеваний распределение индикатора разладки для каждого из выделенных состояний стационарно. Это позволило построить предиктор приступа эпилепсии как индикатор изменения распределения уровня нестационарности данных по сравнению с паттерном спокойного состояния и корректно оценить вероятность ложного срабатывания.

Актуальность темы исследования. Традиционный подход к анализу нестационарных временных рядов состоит в том, что рассматриваются только такие ряды, которые с помощью линейных преобразований сводятся к стационарным. Эти ряды изучаются в рамках авторегрессионных интегрированных моделей скользящего среднего. Указанные модели оперируют не с выборочными функциями распределения, а непосредственно с элементами временного ряда. Ряды, не укладывающиеся в рамки регрессионного анализа, изучаются разными эвристическими методами, называемыми адаптивными, не имеющими

четкого математического обоснования. Следовательно, разработка методов, позволяющих анализировать более широкий класс нестационарных временных рядов, является важным шагом в развитии математической статистики.

Повторимся, диссертантом рассматриваются принципиально нестационарные ряды данных, и ищется критерий перехода системы к иной степени нестационарности, в данном конкретном случае, означающей вероятность развития приступа эпилепсии. В качестве стационарного фона рассматриваются не собственно ряды данных, а построенные на их основе распределения величин нестационарности этих рядов. Подход этот представляется достаточно общим и, предположительно, может быть использован и в других случаях эволюции сильно неравновесных и сильно изменчивых систем, в частности может быть применен и к случаю прогноза землетрясений, где задача, по сути, аналогична – найти признаки возможного приближения сильного землетрясения на фоне сильно изменчивого нестационарного фонового режима сейсмичности. Необходимым требованием при этом является наличие больших (плотных по времени) массивов данных, столь плотными потоками данных сейсмология на данный момент не располагает, но можно надеяться, будет иметь в некоторых районах в обозримом будущем.

Одной из актуальных практических задач анализа временных рядов в контексте больших данных является построение модели разрядки с учетом нестационарных свойств ряда. Это биржевые ряды цен сделок на финансовые инструменты, 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 можно трактовать как разрядку «нормы». Модель разрядки 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. Обычно решения в таких случаях принимаются на основе классических критериев, только уровень значимости постфактум оказывается хуже, чем утверждается критерием. Следовательно, доказательная корректировка вероятности наступления определенных событий является актуальной задачей. Для ее решения в данной работе использовались так называемые кинетические методы анализа временных рядов, основанные на аналитических и численных решениях уравнения Лиувилля для выборочных распределений, а также аппарат теории конечнократных аппроксимаций по формулам Чернова. Соискатель исследует асимптотические по количеству наблюдений свойства 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 соседних пар выборок из временного ряда.

Содержание диссертации. Диссертация «Моделирование индикаторов разрядки в нестационарных временных рядах электроэнцефалограмм» состоит из введения, четырех

глав и заключения общим объемом 94 страницы. Список использованных источников состоит из 75 наименований.

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 соискателем согласованным уровнем стационарности. Для стационарных рядов этот индикатор представляет собой стационарную точку - уровень значимости в критерии Колмогорова-Смирнова. Для задачи распознавания разладки с указанием вероятности ошибки первого рода формулируется численный алгоритм расчета этого индикатора по различным фрагментам исходного ряда как статистики, построенной для соседних выборок произвольной длины из фрагмента ряда также произвольной длины.

Во второй главе проводится теоретический анализ предложенной модели. Соискателем кратко представлены элементы теории конечнократных аппроксимаций полугрупп на основе теоремы Чернова и даны понятия эквивалентности по Чернову и средней полугруппы. На этой теоретической базе им построена методика аппроксимации стационарной точки уровня значимости. Результат аппроксимации сходится к точному значению при увеличении длины фрагмента ряда, причем получаемая оценка достигается существенно меньшими вычислительными затратами.

В третьей главе приведены вычислительные алгоритмы, применяемые для нахождения индикаторов разладки. Была проведена программная оптимизация времени вычисления с использованием OpenMP. Описана структура программного комплекса, реализующего разработанные алгоритмы, а также приведена структура пользовательского интерфейса.

В четвертой главе приводятся результаты численного эксперимента по индикации разладки в виде отклонения согласованного уровня стационарности от своего ожидаемого значения для прогноза наступления приступа эпилепсии в многомерных временных рядах, полученных из записей электроэнцефалограмм по совокупности отведений от мозга больного. Идея идентификации разладки как предвестника приступа эпилепсии в нестационарном потоке данных основывается на том, что распределение предложенного индикатора оказалось стационарным. В работе выяснено, что для данной задачи оптимальная длина встык-выборки составляет 5000 данных. Идентифицировано 8 отведений, по которым наблюдается существенное превышение значений индикатора в окрестности приступа. Выяснилось, что приступ эпилепсии всегда проходит на фоне индикации разладки, каковой считается превышение в двухминутном окне сканирования определенного критического уровня. 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 разладки.

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

Научная новизна. Все результаты диссертации, вынесенные на защиту, являются новыми. Новизна самой работы заключается в том, что впервые теоретический метод построения эквивалентной по Чернову полугруппы был применен в математической статистике для обоснования алгоритма приближенного вычисления индикатора разладки. 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 разладки, в качестве которого предложено использовать статистику, называемую согласованным уровнем стационарности ряда.

Практическая значимость работы. Развитые в диссертации методы позволяют исследовать разладку в нестационарных временных рядах и использовать предложенный автором алгоритм для решения задач выделения фрагментов временных рядов и их последующей кластеризации, что позволит распознавать состояние системы в скользящем окне наблюдения в реальном времени. Разработанная математическая модель и программный продукт могут быть использованы в научной работе в ИТПЗ РАН, МГУ, МИ РАН, МФТИ, ВШЭ и других научных организациях, а также в практической работе специалистами в различных прикладных областях, таких как финансовая аналитика, медицина, математическая лингвистика. Представляется весьма перспективным использовать разработанный инструментарий в области математической геофизики, поскольку с его помощью можно попытаться повысить точность предсказания сильных землетрясений.

Обоснованность и достоверность результатов диссертации. Все полученные в диссертации результаты сформулированы в виде теорем и снабжены строгими и подробными доказательствами, что гарантирует их достоверность. Корректность вычислительных операций основана на использовании апробированных в научной практике методов численного анализа.

Публикация результатов в печати. По материалам диссертации опубликовано 13 работ. Из них 9 статей в рецензируемых журналах, входящих в базы данных WoS и SCOPUS, и 4 препринта ИПМ им. М.В. Келдыша РАН. Материал диссертации достаточно полно представлен в опубликованных работах. Положения и выводы диссертации прошли серьезную научную апробацию.

Отмеченные недостатки.

1. В диссертации рассматривается индикатор, построенный по анализу статистики расстояний между соседними выборками. Однако не рассматривались выборки со сдвигом на произвольное окно наблюдения, хотя такой анализ напрашивается и соответствует развитой методике. Было бы полезно продолжить работу в этом направлении.
2. Недостаточно убедительно выглядит аналогия приближенного вычисления статистики расстояний между выборками и эквивалентной по Чернову полугруппой. Следовало бы сделать акцент не на полугруппе, а на доказанной сходимости итерационного процесса к стационарной точке распределения, что как раз и соответствует области применения этой теоремы.
3. При описании программного комплекса большое внимание уделено блок-схемам алгоритмов, но недостаточно полно описан интерфейс пользователя и возможности управления решаемыми задачами.

Заключение о диссертации. Данная диссертация представляет собой целостное научное исследование на актуальную тему, в котором получен ряд новых результатов по теории аппроксимации выборочных статистик нестационарных временных рядов электроэнцефалограмм и построены предикторы разладки, имеющие более высокую точность срабатывания, чем критерии, традиционно используемые в этой области. Результаты, полученные в диссертации Кислицына Алексея Алексеевича, несомненно,

являются результатами высокого научного уровня и имеют значительную научную и практическую ценность. Автореферат правильно и полно отражает содержание диссертации. На основании вышеизложенного считаем, что диссертационная работа Кислицына Алексея Алексеевича «Моделирование индикаторов разрядки в нестационарных временных рядах электроэнцефалограмм» удовлетворяет всем требованиям Положения о порядке присуждения ученых степеней, предъявляемым к кандидатским диссертациям, а ее автор заслуживает присуждения ему ученой степени кандидата физико-математических наук по специальности 05.13.18 -математическое моделирование, численные методы и комплексы программ.

Отзыв заслушан и утвержден на заседании семинара лаборатории №2 («Лаборатория регистрации и интерпретации волновых полей») института теории прогноза землетрясений и математической геофизики РАН (протокол № 2 от 13 апреля 2021 г.).

Отзыв составил

Зав лабораторией №2 ИТПЗ РАН,
дфмн М.В. Родкин
13 апреля 2021

