

Федеральное государственное бюджетное учреждение науки
Институт проблем передачи информации имени А. А. Харкевича
Российской академии наук

На правах рукописи

Артёмов Алексей Валерьевич

**МАТЕМАТИЧЕСКИЕ МОДЕЛИ
ВРЕМЕННЫХ РЯДОВ С ТРЕНДОМ
В ЗАДАЧАХ ОБНАРУЖЕНИЯ РАЗЛАДКИ**

05.13.18 — математическое моделирование,
численные методы и комплексы программ

Диссертация на соискание учёной степени
кандидата физико-математических наук

Научный руководитель:
доцент, к. ф.-м. н.
Бурнаев Е. В.

МОСКВА — 2016

Оглавление

Введение	5
1 Оценивание параметров сигнала, наблюдаемого во фрактальном гауссовском шуме	16
1.1 Введение	16
1.1.1 Фрактальное броуновское движение	17
1.1.2 Некоторые известные из литературы результаты по фильтрации для фрактального броуновского движения	19
1.2 Постановка задачи	20
1.2.1 Теорема Гирсанова для фрактального броуновского движения	22
1.3 Оценка максимального правдоподобия параметра сноса	24
1.4 Байесовская оценка параметра сноса. Случай нормального априорного распределения	27
1.5 Байесовская оценка. Случай равномерного априорного распределения	31
1.6 Выводы	33
2 Ансамбли «слабых» детекторов в задачах обнаружения разладки	34
2.1 Введение	34
2.2 Известные модели и процедуры обнаружения в задачах о разладке стационарной случайной последовательности	35
2.2.1 Модели разладки стационарной случайной последовательности	36
2.2.2 Некоторые широко используемые статистические процедуры обнаружения разладки стационарной случайной последовательности	39
2.3 Нарушение стандартных предположений о модели разладки. Ансамбли	41

2.3.1	Выполнимость широко используемых предположений о модели разладки	41
2.3.2	Ансамбли «слабых» детекторов	42
2.4	Критерии качества обнаружения разладки	45
2.4.1	Вычислительный алгоритм настройки параметров ансамбля	47
2.5	Сравнительный анализ эффективности ансамблей и классических процедур обнаружения разладки	48
2.6	Выводы	52
3	Математические модели сигналов с квазипериодическим трендом и обнаружение их разладок	56
3.1	Введение	56
3.2	Задача оценивания параметров сигнала с квазипериодическим трендом	57
3.2.1	Известные в литературе модели сигналов с периодической составляющей	57
3.2.2	Постановка задачи оценки гладкого тренда	59
3.3	Алгоритм оценивания параметров сигнала на основе фильтра для наблюдений с длинной памятью	60
3.3.1	Описание алгоритма	60
3.4	Алгоритм оценивания параметров сигнала на основе непараметрической регрессии	62
3.4.1	Модели наблюдений с явным учетом сезонности	62
3.4.2	Итеративные алгоритмы оценивания параметров моделей квазипериодических сигналов	64
3.4.3	Практическая реализация алгоритмов оценивания на основе непараметрической регрессии	66
3.5	Обнаружение моментов изменения свойств сигналов с квазипериодическим трендом	68
3.5.1	Модели разладки квазипериодических сигналов и их адекватность задачам обнаружения разладки	68
3.5.2	Процедуры обнаружения разладки характеристик квазипериодических сигналов	71

3.6	Эффективность обнаружения разладки квазипериодического временного ряда	72
3.6.1	Вычислительный эксперимент и наборы данных	72
3.6.2	Исследуемые процедуры	73
3.6.3	Точность аппроксимации тренда	75
3.6.4	Результаты	77
3.7	Выводы	78
4	Комплекс программ	79
4.1	Предпосылки и архитектура	79
4.2	Структура комплекса	81
4.3	Дополнительные функциональные возможности	84
4.3.1	Алгоритм «по умолчанию»	84
4.3.2	Возможности масштабирования системы обнаружения разладок	85
4.4	Выводы	86
5	Результаты решения прикладных задач	87
5.1	Введение	87
5.2	Задача прогнозирования значений финансовых показателей	87
5.3	Задача оценки параметров нагрузки сетей передачи данных	89
5.4	Задача обнаружения разладок и аномалий поисковой системы Яндекса	90
5.5	Задача исследования возможности детектирования изменения режима турбулентного течения	93
5.5.1	Задача обнаружения изменения дисперсии случайного процесса	94
5.5.2	Задача обнаружения изменения параметров процесса авторегрессии	95
5.5.3	Исследование оперативных характеристик решения задачи детектирования изменения режима турбулентного течения	98
5.6	Выводы	99
	Заключение	101
	Список литературы	103
	Условные обозначения	116

Список рисунков	118
Список таблиц	122

Введение

Актуальность и степень научной разработанности проблемы

1. В последние десять лет при активном развитии информационно-коммуникационных технологий возник новый тип высокотехнологичных систем: системы с *интенсивным программным обеспечением* (software-intensive systems¹). К нему относятся такие высокотехнологичные инженерно-технические системы, как

- цифровые системы широкополосной связи,
- интернет-системы (устройства, сети передачи данных и интернет-сервисы),
- информатизированные центры обслуживания и колл-центры,
- автоматизированные энергосети и системы распределения энергии,
- интеллектуальные транспортные системы (авиадиспетчерские службы, системы дорожного транспорта и др.),
- платформы электронной торговли,
- автоматизированные системы в здравоохранении

и другие. Для всех этих систем характерны: большой объем накапливаемых и обрабатываемых данных, сложная взаимозависимость программных компонент и огромное количество аппаратных элементов, а также чрезвычайно большое число людей, использующих систему для различных целей. Как показывает изучение существующих аппаратно-программных комплексов, в настоящее время одной из их центральных проблем является низкая надежность их эксплуатации: неизбежные при крупном масштабе программные, аппаратные и антропогенные отказы являются на практике нормой, а не исключением [5]. Согласно исследованию [86], доминирующей причиной системных отказов является именно возникновение сбоев программного обеспечения. Таким образом, обеспечение

¹Согласно стандарту ISO/IEC/IEEE 42010:2011(E) — системы, функциональность которых определяется главным образом их программными средствами [33]; согласно Стратегии развития отрасли информационных технологий в Российской Федерации на 2014–2020 годы и на перспективу до 2025 года — аппаратно-программные комплексы с большим удельным весом программной части [102]

бесперебойной и эффективной эксплуатации высокотехнологичных систем с интенсивным ПО представляет собой крупную проблему, для устранения которой необходимо прежде всего предотвращение отказов их программной составляющей, в частности, их быстрое и точное обнаружение.

2. В ряде работ предложены подходы к обнаружению проблемного поведения ПО (дефектов, вредоносного вмешательства и т. д.) на основании данных, собираемых при эксплуатации системы — измерений количества обработанных запросов и средней длительности ожидания в единицу времени, измерения объема переданного сетевого трафика и т. д.:

- в [63] измерения объема передаваемого сетевого трафика используются для обнаружения перегрузок сетевого ядра;
- в [82] измерения профиля сетевого трафика применяются для детектирования внедрений в компьютерные сети;
- в [3] измерения характеристик потоков данных в распределенной сети видеонаблюдения используются для обнаружения перегрузки автомагистралей;
- в [8] измерения погрешности определения псевдодальности используются для выявления неисправных навигационных спутников;
- в [21] измерения сетевого трафика используются для обнаружения аномалий компьютерных сетей (вредоносного или содержащего ошибки трафика), приводящих к сбою сети или сетевого сервиса;
- в [32] рассматривается подход к обнаружению системных отказов распределенных компьютерных систем на основании анализа данных в их узлах;
- работы [6; 45–47] исследуют подходы к обнаружению системных сетевых сбоев, используя измерения объема переданного сетевого трафика;
- в [20; 60] рассматриваются подходы к эффективному обнаружению изменений с помощью распределенных систем сенсоров.

Рассмотренные в этих работах задачи сводятся к выявлению момента резкого изменения некоторых характеристик рассматриваемой системы на основе наблюдаемых статистических данных о других характеристиках этой системы [95]. Задачи такого типа (*задачи о разладке*) были рассмотрены А. Н. Колмогоровым, А. Н. Ширяевым и рядом других авторов [9; 96; 99; 104; 106]. Для решения задач типа задачи о разладке, к которым относится и задача обнаружения отказов реальных информационных систем, рядом исследователей были предложены процедуры обнаружения разладки, такие как:

- метод кумулятивных сумм [49; 54; 64; 69; 105];
- метод контрольных карт [57; 75];
- процедура Ширяева-Робертса [70; 107];
- Байесовские подходы Гиршика, Рубина и Ширяева [2; 24; 108];
- процедуры на основе деревьев контекста [10; 11];
- процедуры на основе методов разложения многомерных данных, таких как анализ главных компонент [6; 45–47; 63] и анализ сингулярного спектра [27; 28; 87; 88].

и др. Однако прямое использование названных процедур для обнаружения разладок реальных систем с интенсивным ПО неэффективно по следующим причинам.

3. Системы с интенсивным ПО являются системами массового обслуживания и испытывают антропогенные циклы нагрузки на ряде масштабов времени (день, неделя, год). В силу изменчивости на большом масштабе времени основной цикл будет стохастическим. Стандартные процедуры обнаружения разладки, напротив, предполагают стационарность наблюдений на рассматриваемом промежутке времени, за исключением момента появления разладки. Поэтому для успешного решения задачи обнаружения разладок сложных систем необходим эффективный аппарат математического моделирования и оценивания квазипериодических сигналов.

4. Типичным свойством потоков данных в информационных системах является также длинная память (long-range dependence). Длинная память является основной причиной возникновения всплесков нагрузки и присутствует на чрезвычайно большом диапазоне масштабов времени; известно ее значительное влияние на эффективность систем массового обслуживания [18]. Таким образом, для идентификации и оценивания реальных сигналов, порожденных системами с интенсивным ПО, необходимо использование специальных стохастических моделей, позволяющих моделировать длинную память.

5. Применение процедур обнаружения разладки, упомянутых выше, требует определить математическую модель возникающего отказа в терминах распределений наблюдаемых характеристик. На практике сделать это часто невозможно, так как типы возникающих отказов и сопутствующие им изменения статистических характеристик априори произвольны; как следствие, в этих задачах могут

быть неэффективны даже теоретически оптимальные методы обнаружения разладки [16; 44; 53; 71; 83].

В области машинного обучения широко известен подход на основе алгоритмической композиции или ансамбля, который заключается в совместном использовании множества «слабых» алгоритмов для получения лучшей предсказательной силы [72]. Согласно композиционному подходу, процедуры обнаружения разладки, для которых сигналы тревоги слабо (однако больше, чем просто случайно) коррелируют с истинным разладками, естественно рассматривать как «слабые» детекторы. В этих условиях для эффективного обнаружения разладки достаточно использовать ее стандартную математическую модель² и для каждого класса наблюдений, представленного обучающей выборкой, выбрать наиболее эффективную композицию.

6. В последние десять лет возникли существенно новые практические условия, в которых беспрецедентные объемы данных обостряют проблему высокоэффективного автоматизированного обнаружения разладок современных больших систем [5]. В этих условиях возникают и новые усиленные требования к методологии и алгоритмике решения описываемых задач. До сих пор не было предложено единой архитектуры, пригодной для обнаружения разладок сложных естественных и инженерных систем крупного размера.

Таким образом, для обнаружения отказов крупных систем с интенсивным ПО актуально исследование методов моделирования сигналов с квазипериодическим трендом и с шумовой компонентой, обладающей длинной памятью, исследование методов обнаружения разладки в случае нарушения стандартных предположений о ее модели, а также разработка единой масштабируемой программной архитектуры для обнаружения разладок и аномалий в условиях больших объемов данных.

Цели работы

Целями настоящей диссертационной работы являются разработка и исследование математических методов, алгоритмов и комплексов программ обнаружения

²В литературе, как правило, стандартная модель разладки заключается в изменении среднего значения стационарной гауссовской случайной последовательности. В этом случае наблюдаемый процесс $\xi = (\xi_t)_{t \geq 0}$ имеет вид $\xi_t = \mu \mathbb{1}_{\{t \geq \theta\}}(t) + \nu_t$, где $\mu \in \mathbb{R}$ — магнитуа разладки, $\theta \geq 0$ — момент появления разладки, и $\nu = (\nu_t)_{t \geq 0}$ — последовательность независимых стандартно нормально распределенных случайных величин.

разладок и аномалий больших динамических систем при наличии квазипериодических трендов, шумовой компоненты с длинной памятью, в случае нарушения стандартных предположений о модели разладки.

Задачи работы

Для достижения поставленных целей в настоящей диссертационной работе рассматриваются следующие задачи:

- разработка и исследование математических методов оценки параметров сигнала по данным измерений, выполненных во фрактальном шуме;
- разработка и исследование алгоритма обнаружения разладки на основе ансамбля «слабых» детекторов для повышения эффективности обнаружения разладки в случае нарушения стандартных предположений о ее модели;
- разработка математических моделей и алгоритмов оценивания сигналов с трендом (в частности, квазипериодического сигнала) и обнаружения разладок и аномалий на фоне тренда;
- создание комплекса программ, реализующих разработанные методы для решения модельных и реальных задач обнаружения разладки.

Научная новизна работы

Научная новизна результатов, полученных в диссертационной работе, состоит в том, что в ней

1. Впервые поставлены и решены задачи фильтрации сигнала, представляемого в виде разложения по заданной системе функций, по данным его регистрации во фрактальном шуме и при различных типах дополнительной информации о сигнале.
2. Впервые разработан и исследован алгоритм обнаружения разладки временного ряда, основанный на совместном использовании множества процедур обнаружения разладки.
3. Предложены и исследованы математические модели временных рядов с трендом (в частности, квазипериодического временного ряда) и обнаружения разладок и аномалий на фоне тренда.

4. Создано и внедрено в производство в компании «Яндекс» новое программное обеспечение, реализующее методы оценки параметров и процедуры обнаружения разладок реальных сигналов.

Теоретическая и практическая значимость работы

Результаты диссертационной работы носят теоретический и практический характер. Теоретические результаты относятся к теории оптимальной фильтрации фрактальных динамических систем, позволяют теоретически исследовать фильтры, основанные на конкретных системах функций, и могут применяться при построении и оценке эффективности компонент автоматизированных информационных систем, используемых для решения задач прогнозирования сигналов. Практические результаты диссертационной работы заключаются в том, что разработанные в ней подходы моделирования сигналов с трендом и обнаружения разладок, были успешно применены для решения следующих прикладных задач:

1. Задача оценки параметров наблюдаемых сигналов больших информационных систем компании «Яндекс» в режиме реального времени.
2. Задача обнаружения отказов программного обеспечения больших информационных систем компании «Яндекс» в режиме реального времени.
3. Задача оценки нагрузки больших сети передачи данных Абилин на основе измерений объема передаваемого между узлами сети трафика.

Методология и методы исследования

В диссертационной работе используются:

- подходы стохастического анализа,
- подходы теории непараметрического оценивания сигналов,
- методы численной оптимизации выпуклых функций,
- методы Монте-Карло эмпирического исследования созданных алгоритмов,
- подходы объектно-ориентированного программирования с использованием языка программирования `python`.

Основные результаты, выносимые на защиту

Основными результатами настоящей диссертационной работы являются:

1. Разработаны новые математические методы оценки параметров сигнала по данным измерений, выполненным во фрактальном шуме, в том числе:
 - получена оценка максимального правдоподобия параметра сигнала;
 - получены оптимальные Байесовские оценки для случаев нормального и равномерного априорных распределений параметра сигнала;
 - охарактеризован оптимальный момент остановки измерений сигнала для случая нормального априорного распределения параметра сигнала.
2. Разработан и исследован алгоритм обнаружения разладки временного ряда на основе ансамбля процедур обнаружения разладки, предложен метод настройки параметров ансамбля.
3. Предложена и исследована методология моделирования квазипериодических сигналов и обнаружения их разладок, в том числе:
 - предложена математическая модель квазипериодического временного ряда на основе разложения по заданной системе функций и вычислительный алгоритм оценки ее параметров на основе оптимального фильтра п. 1;
 - предложена многокомпонентная математическая модель квазипериодического временного ряда и вычислительный алгоритм оценки ее параметров на основе непараметрической регрессии;
 - предложена математическая модель краткосрочной разладки квазипериодического временного ряда и процедура обнаружения этой разладки на основе ансамблей «слабых» детекторов.
4. Создан комплекс программ, реализующий предложенные в диссертационной работе вычислительные алгоритмы фильтрации тренда фрактального случайного сигнала, оценивания квазипериодического сигнала, настройки параметров ансамбля и обнаружения разладки временного ряда на основе ансамбля.

Степень достоверности и апробация работы

Достоверность полученных результатов гарантируется использованием строгих доказательств, основанных на хорошо изученных методах стохастического анализа; совпадением полученных оценок с известными результатами в частных случаях линейных задач; описаниями проведенных экспериментов, допускающих

ми их воспроизводимость; успешным применением результатов исследования в реальных задачах обнаружения программных отказов систем с интенсивным ПО.

Результаты работы докладывались и обсуждались на следующих российских и международных научно-технических конференциях и семинарах:

1. Научный семинар кафедры математического моделирования и информатики физического факультета МГУ им. М. В. Ломоносова под руководством профессора Ю. П. Пытьева (05.03.2015).
2. Научный семинар «Математические методы в естественных науках» физического факультета МГУ им. М. В. Ломоносова под руководством профессора А. Н. Боголюбова (26.03.2015).
3. XXII международная научная конференция студентов, аспирантов и молодых учёных «Ломоносов-2015», 13–17 апреля 2015 г., Москва, Россия.
4. Научный семинар «Practical Machine Learning» компании «Яндекс» под руководством к. ф.-м. н. М. А. Ройзнера (04.06.2015).
5. Научный семинар «Математические модели информационных технологий» департамента анализа данных и искусственного интеллекта Высшей школы экономики под руководством профессора С. О. Кузнецова (18.06.2015).
6. Научный семинар отдела Интеллектуальных систем ВЦ РАН под руководством члена-корреспондента РАН К. В. Рудакова (24.06.2015).
7. Научный семинар «Случайные процессы и стохастический анализ» кафедры теории вероятностей механико-математического факультета МГУ им. М. В. Ломоносова под руководством академика РАН А. Н. Ширяева (23.09.2015)
8. Научный семинар Yandex Data Factory под руководством к. ф.-м. н. Е. А. Рябенко (09.10.2015).
9. Научный семинар лаборатории математического моделирования сложных естественных и инженерных систем МГУ им. М. В. Ломоносова под руководством доцента Е. А. Грачева (06.11.2015).
10. The 8th International Conference on Machine Vision, 19–21 November 2015, Barcelona, Spain.
11. 58-я научная конференция МФТИ, 23–28 ноября 2015 г., г. Долгопрудный, Россия.

12. Общественный постоянный научный семинар «Теория автоматического управления и оптимизации» ИПУ РАН им. В. А. Трапезникова под руководством профессора Б. Т. Поляка (11.12.2015).
13. Deep Machine Intelligence Workshop, Skolkovo Institute of Science and Technology, 4–5 June 2016, Moscow, Russia.
14. Международная конференция по стохастическим методам, 27 мая–03 июня 2016 г., пос. Абрау-Дюрсо, г. Новороссийск, Россия.
15. Международная конференция по алгебре, анализу и геометрии, 26 июня–2 июля 2016 г., г. Казань, Россия.
16. 9th European Summer School in Financial Mathematics, 29 August–2 September 2016, Pushkin, St. Petersburg, Russia.

Личный вклад автора

Автор внес следующий личный вклад в работах, выполненных с соавторами:

1. В работах [1,4] предложены модели квазипериодических сигналов и алгоритмы оценивания их параметров, проведены вычислительные эксперименты для оценки качества предложенной методологии обнаружения разладок.
2. В работе [2] проведен теоретический подсчет структуры оптимальных фильтров во всех случаях, а также численное исследование функции штрафа для случая нормального априорного распределения.
3. В работе [3] предложен критерий качества процедур обнаружения разладки и алгоритм оптимизации этого критерия для ансамблей «слабых» детекторов, проведены вычислительные эксперименты для оценки качества ансамблей.

Публикации

По теме диссертационной работы опубликовано 7 печатных работ, в том числе 1 работа в журнале из списка ВАК и 3 работы в журналах из списка Scopus. Наиболее значимые работы:

- [1] Artemov A. V. Effective signal extraction via local polynomial approximation under long-range dependency conditions // Accepted for publication in Lobachevskii Journal of Mathematics. — 2016. — Vol. 37. — Issue 1.

- [2] Артёмов А. В., Бурнаев Е. В. Оптимальное оценивание сигнала, наблюдаемого во фрактальном гауссовском шуме // Теория вероятностей и ее применения. — 2015. — Т. 60. — №. 1. — С. 163–171.
- [3] Artemov A., Burnaev E. Ensembles of detectors for online detection of transient changes // Eighth International Conference on Machine Vision. — International Society for Optics and Photonics, 2015. — pp. 98751Z–98751Z-5.
- [4] Artemov A., Burnaev E., Lokot A. Nonparametric decomposition of quasi-periodic time series for change-point detection // Eighth International Conference on Machine Vision. — International Society for Optics and Photonics, 2015. — pp. 987520–987520-5.

Структура и объем диссертационной работы

Диссертация состоит из титульного листа, оглавления, введения, пяти глав, заключения и списка литературы, включающего 108 наименований. Работа изложена на 122 страницах и содержит 27 рисунков.

Введение. Охарактеризована проблематика исследования; приведен анализ ее разработанности; обоснована актуальность диссертационной работы; сформулированы цели и задачи диссертационной работы; охарактеризованы научная новизна и практическая значимость исследования, приведено краткое содержание диссертации; приведен список опубликованных по теме диссертации работ.

Первая глава. Теоретическое исследование задачи оценки параметров сигнала по данным его измерений, выполненным во фрактальном шуме (шуме с длинной памятью). Вычисление оценки максимального правдоподобия и Байесовских оценок, учитывающих дополнительную информацию о параметрах сигнала.

Вторая глава. Разработка математических методов обнаружения разладки динамической системы на основе ансамблей «слабых» детекторов при нарушении стандартных предположений о модели разладки. Описания и результаты вычислительных экспериментов для исследования эффективности разработанного подхода.

Третья глава. Разработка методологии моделирования и обнаружения разладок сигналов с квазипериодическим трендом. Модели сигналов с квазипериодическим трендом на основе локально полиномиального приближения и

на основе явного учета сезонности. Вычислительные алгоритмы оценивания параметров моделей квазипериодических сигналов. Алгоритмы обнаружения разладок и аномалий на основе ансамблей «слабых» детекторов.

Четвертая глава. Описание структуры и функционала разработанного комплекса программ.

Пятая глава. Применение разработанных математических методов и алгоритмов в задачах анализа реальных сигналов.

Заключение. Кратко перечислены основные результаты диссертационной работы.

Глава 1

Оценивание параметров сигнала, наблюдаемого во фрактальном гауссовском шуме

1.1 Введение

Проблема оценивания параметров сигнала, наблюдаемого в шуме с длинной памятью, является задачей статистики случайных процессов. В литературе рассматривается три основных класса задач такого типа, а именно

- задачи прогнозирования [25],
- задачи оценки параметра [14; 37; 61; 67; 92],
- задачи фильтрации [38–41].

История исследования явления длинной памяти и понятия фрактального броуновского движения, её моделирующего, приведена в разделе 1.1.1 Обзор некоторых работ по фильтрации для управляемых фрактальным броуновским движением процессов дан в разделе 1.1.2. Для некоторых из названных задач существуют эффективные процедуры оценивания, однако нельзя сказать, что теория оптимальной фильтрации процессов, управляемых фрактальным броуновским движением, завершена [38]. В частности, за исключением случая линейного тренда отсутствуют теоретические результаты о структуре оптимальной оценки параметра тренда фрактального броуновского движения. Кроме того, даже для случая линейного тренда в рамках Байесовского подхода отсутствуют теорети-

ческие результаты о структуре оптимальной оценки, за исключением случая нормального априорного распределения параметра тренда [14].

В данной главе рассматривается задача оценки векторного параметра тренда случайного сигнала, моделируемого фрактальным диффузионным процессом. Подсчитывается оценка максимального правдоподобия для этого параметра, а также последовательная Байесовская оценка для нормального априорного распределения в предположении нелинейного тренда и для случая равномерного априорного распределения в предположении линейного тренда. Также описывается структура оптимального момента остановки наблюдений в случае нормального априорного распределения.

1.1.1 Фрактальное броуновское движение

Броуновское движение является широко используемой математической моделью многих физических и биологических процессов, а его математическая теория — это мощный инструмент анализа широкого круга естественных явлений. Универсальная применимость модели броуновского движения основана на предположении о структуре рассматриваемой системы, которая должна состоять из большого числа независимых или слабо зависимых компонент. Однако существуют естественные структуры и процессы, проявляющие долгосрочные временные и пространственные корреляции. Например, в ряде задач статистической механики появляются распределения вероятностей, в которых коэффициенты корреляции довольно медленно стремятся к нулю. Это означает, что случайные величины в таких задачах нельзя считать слабо зависимыми [103]. Затухание корреляций определяется степенным характером убывания спектральной плотности мощности, обратно пропорциональной частоте. Такая зависимость спектральной плотности мощности от частоты характерна для весьма различных по своей природе процессов:

- флуктуаций нормального периода сердцебиений человека [42];
- флуктуаций тока и магнитного поля в системах заряженных частиц [34; 97];
- вариаций объема трафика, передаваемого по сети Ethernet [1; 62];
- вариаций объема информации, кодирующего видеосигнал [22];
- флуктуаций уровня воды в водоемах [31],

- и даже флуктуаций мощности и частоты аудиосигнала в музыке [89];
- исследования в области макроэкономики, финансов, моделей ценообразования и доходности также указывают на наличие «длинной памяти» для процессов, соответствующих динамике цен, эволюции экономических показателей, изменениям курсов обмена валют, динамике процентных ставок и т. п. [7].

Все эти и другие эмпирические данные говорят о необходимости изучения фрактальных стохастических процессов для моделирования, идентификации и оценивания реальных сигналов.

Теория фрактального броуновского движения — это математическое обобщение классической теории случайного блуждания и броуновского движения. Термин «фрактальное» соответствует фрактальному интегрированию и дифференцированию [79; 80]. Процесс фрактального броуновского движения был впервые рассмотрен Колмогоровым [98] и позднее был задан Мандельбротом и ван Нессом посредством стохастического интеграла по обыкновенному броуновскому движению [51]. Приведем обозначения на основе принятых в работе Клепценой [38].

Стандартное фрактальное броуновское движение $B^H = (B_t^H)_{0 \leq t \leq T}$ на $[0, T]$ с параметром Хёрста $H \in (0, 1)$ — это гауссовский процесс с непрерывными траекториями такой, что

$$B_0^H = 0, \quad \mathbb{E} B_t^H = 0, \quad \mathbb{E}[B_t^H B_s^H] = \frac{1}{2}(|t|^{2H} + |s|^{2H} - |t - s|^{2H}).$$

В случае, когда $H = 1/2$, фрактальное броуновское движение является обыкновенным броуновским движением, в случае же $H \neq 1/2$ процесс B^H не является мартингалом [101].

В противоположность классическому броуновскому движению, имеющему независимые приращения, основными свойствами фрактального броуновского движения являются длительность и интенсивность его пространственных и временных корреляций. Диапазон времен, для которых приращения фрактального броуновского движения являются зависимыми, можно назвать бесконечным [51]. С эмпирической точки зрения этот эффект (наличие «длинной памяти») может быть определен как устойчивость наблюдаемых автокорреляций [7]. Понятие фрактального броуновского движения тесно связано с рядом понятий статисти-

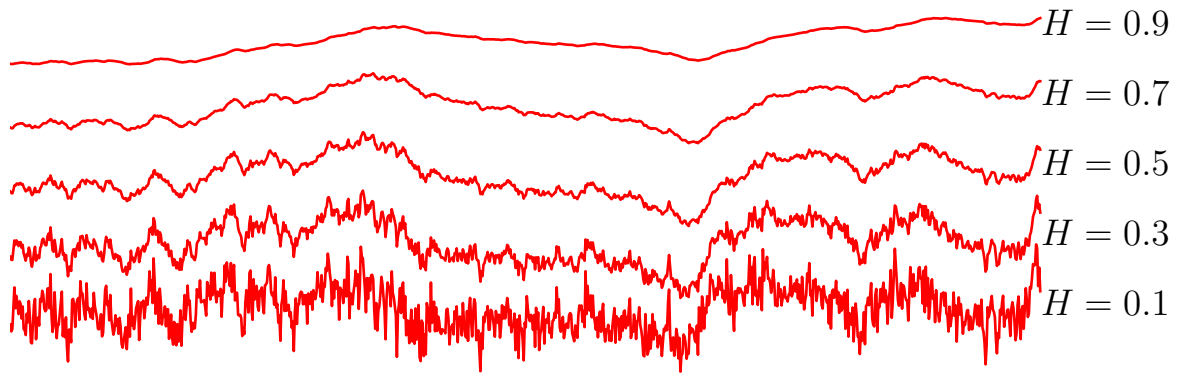


Рисунок 1.1: Примеры реализаций фрактального броуновского движения с различными значениями показателя H

ческой физики, включая понятия масштабной инвариантности, фрактальной размерности и показателя Херста, а также с описанными выше эмпирическими явлениями, такими как фликкер-шум или шум $1/f^a$. Последний является процессом фрактального броуновского движения с показателем Херста $H = \frac{a-1}{2}$.

1.1.2 Некоторые известные из литературы результаты по фильтрации для фрактального броуновского движения

Литература содержит ряд работ, посвященных оценке параметров фрактальных динамических систем. Отметим, что подходы стандартной теории стохастического интегрирования в случае, когда управляющим процессом является фрактальное броуновское движение, неприменимы. По этой причине значительный объем литературы посвящен обобщению результатов стандартной теории на случай фрактального броуновского движения: были предложены подходы к расширению теории стохастического интегрирования на системы, управляемые фрактальным броуновским движением и, в частности, предложен фрактальный аналог теоремы Гирсанова [38; 61; 80]. Для целей настоящей главы центральными являются работа Клепцовой [38], в которой описана структура процесса правдоподобия для общего вида процесса диффузионного типа, а также работа Четина [14], в которой рассматривается Байесовская задача оценивания параметра линейного сноса в системе, управляемой фрактальным броуновским движением.

Задачи оценивания параметров динамических систем, управляемых фрактальным броуновским движением, рассматриваются с 1990-х гг., когда Норрос доказал теорему Гирсанова для фрактального броуновского движения в предположении линейного сноса, получил явный вид оценки максимального правдоподобия коэффициента сноса и исследовал его асимптотические свойства [61]. Аналогичный результат получен Ху, который рассматривал задачу оценивания параметра линейного сноса fBm методом максимального правдоподобия в дискретном времени, используя теорему о нормальной корреляции [92]. Лишь недавно Четин, Новиков и Ширяев рассмотрели Байесовскую задачу последовательного оценивания параметра линейного сноса fBm, в которой последний априори нормально распределен, и аналитически получили оптимальную Байесовскую оценку и оптимальный момент остановки для квадратичного и дираковского штрафов [14]. Отметим, что в обоих этих задачах рассматривались модели, в которых тренд (снос) являлся линейной функцией, и лишь в работе Четина делались априорные предположения о параметре тренда. Также задачи оценки параметра сноса рассматривались для фрактальных аналогов процесса Орнштейна-Уленбека, где были получены оценки максимального правдоподобия этого параметра, исследованы их асимптотические свойства и доказана состоятельность [37; 67]. Наконец, рядом авторов рассматривалась задача оценки параметра сноса в частично наблюдаемой фрактальной диффузионной системе [12].

Среди задач фильтрации центральное место занимает цикл работ Клепцовой и ле Бретона, которые предложили алгоритмы фильтрации для однородных линейных систем, управляемых фрактальным броуновским движением [36; 39–41; 48] и расширили фильтр Калмана-Бьюси на фрактальные линейные системы [19].

1.2 Постановка задачи

Пусть на вероятностном пространстве (Ω, \mathcal{F}, P) выделена фильтрация $\mathcal{F} = (\mathcal{F}_t)_{t \geq 0}$ и задан случайный процесс $\xi = (\xi_t)_{0 \leq t \leq T}$, имеющий представление

$$\xi_t = f(t) + \nu_t, \quad (1.1)$$

где снос $f(t)$ удовлетворяет условию абсолютной интегрируемости на $[0, T]$, $\int_0^T |f(t)| dt < \infty$, а шум ν_t задается соотношением

$$\nu_t = \sigma(t) B_t^H, \quad (1.2)$$

в котором коэффициент диффузии $\sigma(t)$ — квадратично интегрируемая на $[0, T]$ функция, $\int_0^T |\sigma(t)|^2 dt < \infty$, а процесс $B^H = (B_t^H)_{0 \leq t \leq T}$ — фрактальное броуновское движение с показателем Хёрста $H \in (0, 1)$. В настоящей главе функция $\sigma(t)$ предполагается известной. Пусть коэффициент сноса $f(t)$ представим в виде разложения

$$f(t) = \sum_{i=0}^{n_\theta} \theta_i g_i(t) \quad (1.3)$$

по заданной системе $\{g_i(t)\}_{i=0}^{n_\theta}$ абсолютно интегрируемых функций, а параметры $\theta_i, i = 0, \dots, n_\theta$ — неизвестны. Для краткости примем векторные обозначения $\boldsymbol{\theta} = (\theta_0, \dots, \theta_{n_\theta})^\top$, $\mathbf{g}(t) = (g_0(t), \dots, g_{n_\theta}(t))^\top$, тогда

$$f(t) = \boldsymbol{\theta}^\top \mathbf{g}(t). \quad (1.4)$$

Рассмотрим задачу нахождения оценки значения $\boldsymbol{\theta}$ по реализации $\{\xi_s, 0 \leq s \leq t\}$, доступной до момента времени t . Для ее решения рассмотрим два широко используемых подхода: подход на основе метода максимального правдоподобия и Байесовский подход. В первом подходе $\boldsymbol{\theta}$ считается вектором неизвестных параметров и требуется отыскать оценку $\hat{\boldsymbol{\theta}}_{\text{ML}} = \hat{\boldsymbol{\theta}}_{\text{ML}}(t)$, максимизирующую правдоподобие соответствующего наблюдаемого процесса. Байесовский подход основан на предположении, что $\boldsymbol{\theta}$ — случайный вектор из $\mathbb{R}^{n_\theta+1}$, имеющий известное априорное распределение $p^\theta(\mathbf{x}), \mathbf{x} \in \mathbb{R}^{n_\theta+1}$. В таких предположениях рассматривается задача нахождения последовательного правила оценивания $\delta_{\text{BAYES}} = (\varrho_{\text{BAYES}}, \hat{\boldsymbol{\theta}}_{\text{BAYES}})$, такого, что

$$\inf_{\delta \in \mathbb{D}} \mathbb{E}[c\varrho + \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2] = \mathbb{E}[c\varrho_{\text{BAYES}} + \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{\text{BAYES}}\|^2], \quad (1.5)$$

где $\mathbb{D} = \{\delta \mid \delta = (\varrho, \hat{\boldsymbol{\theta}})\}$ — класс правил оценивания с конечными моментами остановки $\varrho \leq T < \infty$ относительно фильтрации $\mathcal{F}_t = \sigma(\{\xi_s, 0 \leq s \leq t\})$, а $c > 0$ — заданная постоянная, интерпретируемая как плата за длительность на-

блюдений. Байесовская стратегия последовательного оценивания $\boldsymbol{\theta}$ заключается в том, что наблюдения останавливаются в момент ϱ_{BAYES} , и $\hat{\boldsymbol{\theta}}_{\text{BAYES}}$ принимается оптимальной оценкой значения $\boldsymbol{\theta}$. Заметим, что если для случайной величины $\hat{\boldsymbol{\theta}}$ имеет место соотношение $\mathbb{E}[\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2] < \infty$, то выполнено равенство

$$\mathbb{E}[c\varrho + \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2] = \mathbb{E}[c\varrho + \mathbb{E}[\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2 | \mathcal{F}_t^\xi]].$$

Величину $\mathbb{E}[\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2 | \mathcal{F}_t^\xi]$ тогда можно подсчитать как среднее по условному распределению $p^\theta(\mathbf{x} | \mathcal{F}_t^\xi)$:

$$\mathbb{E}[\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2 | \mathcal{F}_t^\xi] = \int_{\mathbb{R}^{n_\theta+1}} \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}\|^2 p^\theta(\mathbf{x} | \mathcal{F}_t^\xi) d\mathbf{x}.$$

1.2.1 Теорема Гирсанова для фрактального броуновского движения

Приведем ряд результатов, относящихся к структуре процесса правдоподобия для фрактального броуновского движения со сносом [38]. Для этого вначале введем необходимые обозначения.

Известно следующее интегральное преобразование процесса B^H , результатом которого является мартингал. Обозначим для $0 \leq s < t \leq T$

$$\begin{aligned} \kappa_H &= 2H\Gamma\left(\frac{3}{2} - H\right)\Gamma\left(\frac{1}{2} + H\right), \\ k_H(t, s) &= \kappa_H^{-1} s^{1/2-H} (t-s)^{1/2-H}, \\ \lambda_H &= \frac{2H\Gamma(3-2H)\Gamma(\frac{1}{2}+H)}{\Gamma(\frac{3}{2}-H)}, \\ w_H(t) &= \lambda_H^{-1} t^{2-2H}, \quad dw_t^H = d(w_H(t)) \end{aligned}$$

и определим процесс $M^H = (M_t^H)_{0 \leq t \leq T}$ соотношением

$$M_t^H \equiv \int_0^t k_H(t, s) dB_s^H. \quad (1.6)$$

Так заданный процесс M_t^H является гауссовским мартингалом, причем его квадратическая вариация $\langle M_t^H \rangle$ есть функция $w_H(t)$ [38; 61]. Для дальнейшего

удобства зададим также процесс $m^H = (m_t^H)_{0 \leq t \leq T}$ соотношением

$$m_t^H = M_t^H / w_H(t).$$

Пусть на фильтрованном вероятностном пространстве $(\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})$ задан процесс $Y = (Y_t)_{0 \leq t \leq T}$, стохастический дифференциал dY_t которого удовлетворяет соотношению

$$dY_t = C(t) dt + B(t) dB_t^H,$$

где $B^H = (B_t^H)_{0 \leq t \leq T}$ — фрактальное броуновское движение с параметром $H \in (0, 1)$, а неслучайная функция $C(t)$ таковы, что функция $Q_H(t)$, задаваемая соотношением

$$Q_H(t) = \frac{d}{dw_t^H} \int_0^t k_H(t, s) C(s) ds, \quad (1.7)$$

корректно определена. В последнем соотношении дифференцирование по величине dw_t^H понимается в смысле следующей операции

$$\frac{df(t)}{dw_t^H} \equiv \frac{\lambda_H}{2 - 2H} t^{2H-1} \frac{df(t)}{dt}.$$

Задание такой функции позволяет сформулировать аналог теоремы Гирсанова для процесса Y .

Теорема (Доказательство см. в [38]). Пусть $Q_H(t)$ принадлежит $L^2([0, T], dw_t^H)$, где величина dw_t^H определена в (1.6). Зададим случайный процесс $\Lambda^H = (\Lambda_t^H)_{0 \leq t \leq T}$ соотношением

$$\Lambda_t^H = \exp \left\{ \int_0^t Q_H(s) dM_s^H - \frac{1}{2} \int_0^t (Q_H(s))^2 dw_s^H \right\}. \quad (1.8)$$

Тогда $\mathbb{E} \Lambda_t^H = 1$ и распределение процесса Y по мере $\mathbb{P}^Y = \Lambda_t^H \mathbb{P}$ совпадает с распределением процесса $\int_0^t B(s) dB_s^H$ по мере \mathbb{P} .

Случайный процесс Λ^H называется процессом правдоподобия или производной Радона-Никодима $\frac{d\mathbb{P}^Y}{d\mathbb{P}}$ меры \mathbb{P}^Y по мере \mathbb{P} .

1.3 Оценка максимального правдоподобия параметра сноса

Рассмотрим задачу отыскания оценки максимального правдоподобия для параметра $\boldsymbol{\theta}$ коэффициента сноса $f(t)$ в (1.1). Согласно (1.1)–(1.3) процесс ξ_t удовлетворяет уравнению

$$\xi_t = \sum_{i=0}^{n_\theta} \theta_i g_i(t) + \sigma(t) B_t^H, \quad (1.9)$$

а его стохастический дифференциал — уравнению

$$d\xi_t = \sum_{i=0}^{n_\theta} \theta_i g'_i(t) dt + \sigma(t) dB_t^H.$$

Структуру процесса правдоподобия и соответствующей оценки описывает следующая

Теорема 1. Пусть коэффициент сноса $f(\cdot)$ фрактального броуновского движения имеет вид (1.3)–(1.4). Тогда оценка $\hat{\boldsymbol{\theta}}_{\text{ML}}$ максимального правдоподобия параметра $\boldsymbol{\theta}$ имеет вид

$$\hat{\boldsymbol{\theta}}_{\text{ML}} = \mathbf{R}_H^{-1}(t) \boldsymbol{\psi}_t^H, \quad (1.10)$$

где компоненты $n_\theta + 1$ -мерного случайного процесса $\boldsymbol{\psi}^H = (\boldsymbol{\psi}_t^H)_{0 \leq t \leq T}$ и элементы детерминированной матрицы $\mathbf{R}_H(t)$ равны

$$(\boldsymbol{\psi}_t^H)_i = \int_0^t \psi_i(s) dM_s^H \quad \text{и} \quad (\mathbf{R}_H(t))_{ij} = \int_0^t \psi_i(s) \psi_j(s) dw_s^H, \quad (1.11)$$

$i, j = 0, \dots, n_\theta$, соответственно, а функции $\psi_i(t)$, $i = 0, \dots, n_\theta$, задаются соотношениями

$$\psi_i(t) = \frac{d}{dw_t^H} \int_0^t k_H(t, s) g'_i(s) ds, \quad i = 0, \dots, n_\theta. \quad (1.12)$$

Доказательство. Для функции $Q_H^\theta(t)$, общий вид которой определен в (1.7), получим

$$Q_H^\theta(t) = \sum_{i=0}^{n_\theta} \theta_i \frac{d}{dw_t^H} \int_0^t k_H(t,s) g'_i(s) ds = \sum_{i=0}^{n_\theta} \theta_i \psi_i(t),$$

где приняты обозначения (1.12) для функций $\psi_i(t)$, $i = 0, \dots, n_\theta$. Для процесса правдоподобия Λ^H , общий вид которого записан в (1.8), имеем

$$\Lambda_t^H(\boldsymbol{\theta}) = \exp \left\{ \sum_{i=0}^{n_\theta} \theta_i \int_0^t \psi_i(s) dM_s^H - \frac{1}{2} \int_0^t \left(\sum_{i=0}^{n_\theta} \theta_i \psi_i(s) \right)^2 dw_s^H \right\}. \quad (1.13)$$

Процесс Λ^H определяет производную Радона-Никодима меры, порожденной процессом ξ_s из (1.9), относительно меры процесса $\xi_s = B_s^H$, $s \leq t$. Формулу (1.13) можно записать более компактно, используя векторные обозначения:

$$\Lambda_t^H(\boldsymbol{\theta}) = \exp \left\{ \boldsymbol{\theta}^\top \boldsymbol{\psi}_t^H - \frac{1}{2} \boldsymbol{\theta}^\top \mathbf{R}_H(t) \boldsymbol{\theta} \right\}, \quad (1.14)$$

где компоненты $n_\theta + 1$ -мерного процесса $\boldsymbol{\psi}^H$ и элементы матрицы $\mathbf{R}_H(t)$ заданы соотношениями (1.11). Используя условие равенства нулю градиента $\nabla_{\boldsymbol{\theta}} \log \Lambda_t^H(\boldsymbol{\theta})$, получим оценку максимального правдоподобия $\hat{\boldsymbol{\theta}}_{\text{ML}} = \arg \max_{\boldsymbol{\theta}} \Lambda_t^H(\boldsymbol{\theta})$ как решение системы алгебраических уравнений

$$\int_0^t \psi_i(s) dM_s^H - \sum_{j=0}^{n_\theta} \theta_j \int_0^t \psi_i(s) \psi_j(s) dw_s^H = 0, \quad i = 0, \dots, n_\theta,$$

или в векторном виде

$$\boldsymbol{\psi}_t^H - \mathbf{R}_H(t) \boldsymbol{\theta} = 0.$$

Если матрица $\mathbf{R}_H(t)$ обратима для каждого t , то решением этой системы является оценка $\hat{\boldsymbol{\theta}}_{\text{ML}}$ из (1.10). Теорема доказана. \square

Следствие 1 (Случай полиномиального сноса). Пусть $g_i(t) = t^i$, $i = 0, \dots, n_\theta$, а коэффициент диффузии предполагается постоянным: $\sigma(t) = \sigma$. Тогда наблюдаемый процесс имеет структуру $\xi_t = \sum_{i=0}^{n_\theta} \theta_i t^i + \sigma B_t^H$, для функций $\psi_i(t)$ имеем $\psi_i(t) = \beta_H(i) t^{i-1}$, $i = 0, \dots, n_\theta$, а компоненты векторнозначного процес-

са $\boldsymbol{\psi}^H$ и элементы матрицы $\mathbf{R}_H(t)$ заданных в (1.11), равны

$$(\boldsymbol{\psi}_t^H)_i = \beta_H(i) \int_0^t s^{i-1} dM_s^H, \quad (\mathbf{R}_H(t))_{ij} = \alpha_H(i, j) t^{i+j-2H},$$

соответственно¹, где

$$\begin{aligned} \alpha_H(i, j) &= \lambda_H^{-1} \beta_H(i) \beta_H(j) \frac{2-2H}{i+j-2H}, \\ \beta_H(i) &= i \frac{2-2H+i-1}{2-2H} \frac{\Gamma(3-2H)}{\Gamma(3-2H+i-1)} \frac{\Gamma(3/2-H+i-1)}{\Gamma(3/2-H)}, \end{aligned}$$

где $i, j = 0, \dots, n_\theta$. Оценка $\hat{\boldsymbol{\theta}}_{\text{ML}}$ получается решением системы $\boldsymbol{\psi}_t^H - \mathbf{R}_H(t)\boldsymbol{\theta} = 0$.

Отметим, что в случае $n_\theta = 1$ наблюдаемый процесс удовлетворяет стохастическому дифференциальному уравнению $d\xi_t = \theta_1 dt + \sigma dB_t^H$, а выражение для процесса правдоподобия имеет вид $\Lambda_t^H(\boldsymbol{\theta}) = \exp\{\theta_1 M_t^H - \theta_1^2 \lambda_H^{-1} t^{2-2H}/2\}$, поэтому оценка максимального правдоподобия $(\hat{\theta}_1)_{\text{ML}}$ для θ_1 имеет вид

$$(\hat{\theta}_1)_{\text{ML}} = \frac{M_t^H}{w_H(t)}. \quad (1.15)$$

Этот результат получен Норросом [61].

¹ Для подсчета величин вида $\int_0^t s^i dM_s^H$ воспользуемся формулой интегрирования по частям

$$\int_0^t s^i dM_s^H = t^i M_t^H - i \int_0^t s^{i-1} M_s^H ds.$$

Пользуясь той же формулой далее, для второго слагаемого получим

$$\int_0^t s^{i-1} M_s^H ds = t^{i-1} N_t^{(i)} - (i-1) \int_0^t s^{i-2} N_s^{(i)} ds,$$

где $N_t^{(i)} = \int_0^t N_s^{(i-1)} ds, i = 1, 2, \dots, N_t^{(0)} = M_t^H$. Применяя формулу интегрирования по частям далее, получим и все остальные слагаемые.

1.4 Байесовская оценка параметра сноса. Случай нормального априорного распределения

Рассмотрим задачу отыскания Байесовской оценки параметра $\boldsymbol{\theta} \in \mathbb{R}^{n_\theta}$, имеющего априорное распределение с плотностью $p^\theta(\mathbf{x})$, $\mathbf{x} = (x_0, \dots, x_{n_\theta}) \in \mathbb{R}^{n_\theta}$.

Согласно обобщенной формуле Байеса (см. [14; 100]) условная плотность распределения $\boldsymbol{\theta}$ при условии наблюдений $\mathcal{F}_t^\xi = \sigma(\{\xi_s, 0 \leq s \leq t\})$ за процессом ξ выражается в виде

$$p^\theta(\mathbf{x} | \mathcal{F}_t^\xi) = \frac{dP(\theta_0 \leq x_0, \dots, \theta_n \leq x_n | \mathcal{F}_t^\xi)}{dx_0 \cdots dx_n} = \frac{p^\theta(\mathbf{x}) \Lambda_t^H(\mathbf{x})}{\int_{\mathbb{R}^{n_\theta}} p^\theta(\mathbf{z}) \Lambda_t^H(\mathbf{z}) d\mathbf{z}}, \quad \mathbf{x} \in \mathbb{R}^{n_\theta}, \quad (1.16)$$

где $\Lambda_t^H(\mathbf{x})$ — процесс правдоподобия, структура которого описана в разделе 1.3. Далее рассмотрим частные случаи, в которых априорное распределение является нормальным и равномерным.

Рассмотрим случай нормального априорного распределения. Основной результат настоящего раздела изложен в следующей теореме.

Теорема 2. Пусть $\boldsymbol{\theta}$ — нормальный случайный вектор с математическим ожиданием \mathbf{m} и ковариационной матрицей Σ . Тогда оптимальной в среднем квадратичном Байесовской оценкой $\hat{\boldsymbol{\theta}}_{\text{BAYES}}$ значения $\boldsymbol{\theta}$ является апостериорное среднее

$$\hat{\boldsymbol{\theta}}_{\text{BAYES}} = E[\boldsymbol{\theta} | \mathcal{F}_t^\xi] = (\mathbf{R}_H(t) + \Sigma^{-1})^{-1} (\boldsymbol{\psi}_t^H + \Sigma^{-1} \mathbf{m}). \quad (1.17)$$

Величина условной среднеквадратичной ошибки оценивания $E[\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{\text{BAYES}}\|^2 | \mathcal{F}_t^\xi]$ определяется следом условной ковариационной матрицей

$$\text{cov}[\boldsymbol{\theta} | \mathcal{F}_t^\xi] = (\mathbf{R}_H(t) + \Sigma^{-1})^{-1}. \quad (1.18)$$

Доказательство. Хорошо известно, что наилучшей в среднеквадратичном оценке значения вектора $\boldsymbol{\theta}$ при условии наблюдений $\{\xi_s, 0 \leq s \leq t\}$ является условное математическое ожидание $E[\boldsymbol{\theta} | \mathcal{F}_t^\xi]$, при этом величина среднеквадратичной погрешности оценивания равна следу условной ковариационной матрицы

$\text{cov}[\boldsymbol{\theta}|\mathcal{F}_t^\xi]$. Как будет показано ниже, для случая нормального априорного распределения эти величины легко подсчитать.

Используя формулы (1.14) и (1.16) и учитывая, что плотность нормального случайного вектора

$$p^\theta(\mathbf{x}) = \frac{1}{((2\pi)^{n_\theta} \det \boldsymbol{\Sigma})^{1/2}} \exp \left\{ -\frac{1}{2}(\mathbf{x} - \mathbf{m})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \mathbf{m}) \right\},$$

для условного распределения $\boldsymbol{\theta}$ при условии наблюдений $\mathcal{F}_t^\xi = \sigma(\{\xi_s, 0 \leq s \leq t\})$ получаем

$$p^\theta(\mathbf{x}|\mathcal{F}_t^\xi) = g(\mathbf{x}) / \int_{\mathbb{R}^{n_\theta}} g(\mathbf{z}) d^{n_\theta} \mathbf{z},$$

$$g(\mathbf{x}) = \exp \left\{ \mathbf{x}^\top (\boldsymbol{\psi}_t^H + \boldsymbol{\Sigma}^{-1} \mathbf{m}) - \frac{1}{2} \mathbf{x}^\top (\mathbf{R}_H(t) + \boldsymbol{\Sigma}^{-1}) \mathbf{x} \right\}.$$

Пользуясь известной формулой $\int_{\mathbb{R}^n} \exp \left\{ -\frac{1}{2} \mathbf{x}^\top \mathbf{A} \mathbf{x} + \mathbf{x}^\top \mathbf{b} \right\} d^n \mathbf{x} = \sqrt{\frac{(2\pi)^n}{\det \mathbf{A}}} \exp \left\{ \frac{1}{2} \mathbf{b}^\top \mathbf{A}^{-1} \mathbf{b} \right\}$, последнее выражение преобразуем к виду

$$p^\theta(\mathbf{x}|\mathcal{F}_t^\xi) = \sqrt{\frac{\det \mathbf{A}}{(2\pi)^{n_\theta}}} \exp \left\{ -\frac{1}{2} (\mathbf{x}^\top \mathbf{A} \mathbf{x} - 2 \mathbf{x}^\top \mathbf{b} + \mathbf{b}^\top \mathbf{A}^{-1} \mathbf{b}) \right\},$$

где детерминированная матрица $\mathbf{A} = \mathbf{A}_H(t)$ и n_θ -мерный случайный процесс $\mathbf{b}^H = (\mathbf{b}_t^H)_{0 \leq t \leq T}$ задаются соотношениями $\mathbf{A}_H(t) = \mathbf{R}_H(t) + \boldsymbol{\Sigma}^{-1}$ и $\mathbf{b} = \mathbf{b}_t^H = \boldsymbol{\psi}_t^H + \boldsymbol{\Sigma}^{-1} \mathbf{m}$ соответственно. Таким образом, условная плотность

$$p^\theta(\mathbf{x}|\mathcal{F}_t^\xi) = \sqrt{\frac{\det \mathbf{A}}{(2\pi)^{n_\theta}}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \mathbf{A}^{-1} \mathbf{b})^\top \mathbf{A} (\mathbf{x} - \mathbf{A}^{-1} \mathbf{b}) \right\}$$

является многомерной нормальной со средним и ковариационной матрицей

$$\begin{aligned} \mathbb{E}[\boldsymbol{\theta}|\mathcal{F}_t^\xi] &= \mathbf{A}^{-1} \mathbf{b} = (\mathbf{R}_H(t) + \boldsymbol{\Sigma}^{-1})^{-1} (\boldsymbol{\psi}_t^H + \boldsymbol{\Sigma}^{-1} \mathbf{m}) \quad \text{и} \\ \text{cov}[\boldsymbol{\theta}|\mathcal{F}_t^\xi] &= \mathbf{A}^{-1} = (\mathbf{R}_H(t) + \boldsymbol{\Sigma}^{-1})^{-1}, \end{aligned}$$

соответственно. Величина $E[\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{\text{BAYES}}\|^2 | \mathcal{F}_t^\xi]$ условной среднеквадратичной ошибки оценивания при этом равна

$$E[\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{\text{BAYES}}\|^2 | \mathcal{F}_t^\xi] = E\left[\text{tr}\left((\mathbf{R}_H(t) + \boldsymbol{\Sigma}^{-1})^{-1}\right)\right] = \text{tr}\left((\mathbf{R}_H(t) + \boldsymbol{\Sigma}^{-1})^{-1}\right).$$

Теорема доказана. \square

Следствие 2. Пусть выполнены условия теоремы 2, тогда оптимальный момент остановки в задаче (1.5) детерминирован.

Доказательство. Действительно, для определения оптимального момента остановки ϱ_{BAYES} необходимо решить следующую задачу оптимальной остановки

$$\varrho_{\text{BAYES}} = \arg \inf_{\varrho \in \mathbb{D}} E\left[c\varrho + E[\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{\text{BAYES}}\|^2 | \mathcal{F}_t^\xi]\right] = \arg \inf_{t \in [0, T]} F_H(t),$$

где функция

$$F_H(t) = ct + E[\|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{\text{BAYES}}\|^2 | \mathcal{F}_t^\xi] = ct + \text{tr}\left((\mathbf{R}_H(t) + \boldsymbol{\Sigma}^{-1})^{-1}\right), t \in [0, T] \quad (1.19)$$

детерминирована. \square

Следствие 3 (Случай полиномиального сноса). Пусть $g_i(t) = t^i, i = 0, \dots, n_\theta$, коэффициент диффузии $\sigma(t) = \sigma$ постоянен, а матрица $\boldsymbol{\Sigma}$ — диагональная (координаты вектора $\boldsymbol{\theta}$ независимы). Тогда функция $F_H(t)$ из (1.19) имеет единственный минимум при $t \in [0, T]$.

Доказательство. Действительно, пусть $\boldsymbol{\Sigma} = \text{diag}(\gamma_0^2, \dots, \gamma_n^2)$, где γ_i^2 — дисперсия координаты $\theta_i, i = 0, \dots, n_\theta$. Тогда след условной ковариационной матрицы равен

$$\text{tr}\left((\mathbf{R}_H(t) + \boldsymbol{\Sigma}^{-1})^{-1}\right) = \prod_{i=0}^n (\alpha_H(i, i) t^{2i-2H} / \sigma^4 + \gamma^{-2})^{-1}$$

и является строго убывающей функцией при $t > 0$. Поэтому $F_H(t)$ в (1.19) является суммой строго возрастающей и строго убывающей функций и имеет при $t \in [0, T]$ минимум, и притом единственный. \square

Аналогичный результат известен для случая линейного сноса $a(t) = \mu t$ [14]. Для случая кубического сноса и значений $H = 0.2, c = 0.02$ график функции $F_H(t)$ представлен на рис. 1.3.

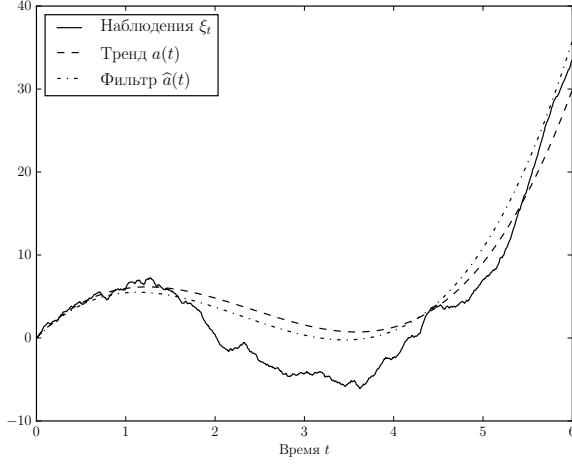


Рисунок 1.2: Траектории результата наблюдения ξ_t , тренда $f(=) \sum_{k=0}^3 \theta_k t^k$ и фильтра $\hat{f}(t) = \sum_{k=0}^3 (\hat{\theta}_{\text{BAYES}})_k t^k$, $0 \leq t \leq T$, в модельной задаче выделения кубического тренда при значении параметра $H = 0.8$

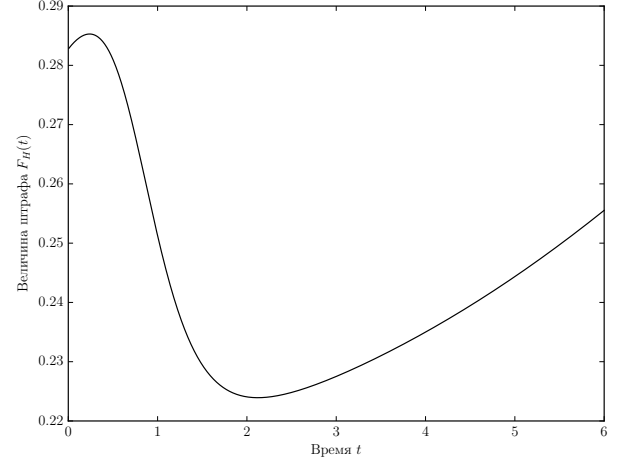


Рисунок 1.3: Значения функции штрафа $F_H(t)$ в модельной задаче выделения полиномиального (кубического) тренда при значениях параметров $H = 0.8, c = 0.02$

Заметим, что если наблюдаемый процесс удовлетворяет стохастическому дифференциальному уравнению с линейным сносом $d\xi_t = \theta_1 dt + \sigma dB_t^H$, в котором θ_1 — нормально распределенная случайная величина с математическим ожиданием m и дисперсией γ^2 , то условная плотность распределения θ_1 является нормальной:

$$p^{\theta_1}(x|\mathcal{F}_t^\xi) = \sqrt{\frac{w_H(t)/\sigma^2 + 1/\gamma^2}{2\pi}} \times \\ \times \exp \left\{ - \left(x - \frac{M_t^H/\sigma + m/\gamma^2}{w_H(t)/\sigma^2 + 1/\gamma^2} \right)^2 \frac{w_H(t)/\sigma^2 + 1/\gamma^2}{2} \right\},$$

а Байесовская оценка $(\hat{\theta}_1)_{\text{BAYES}} = \mathbb{E}[\theta_1|\mathcal{F}_t^\xi]$ параметра θ_1 и условная средне-квадратичная погрешность оценивания $(\hat{\gamma}_1^2)_{\text{BAYES}} = \mathbb{E}[(\theta_1 - (\hat{\theta}_1)_{\text{BAYES}})^2|\mathcal{F}_t^\xi]$ задаются соотношениями

$$(\hat{\theta}_1)_{\text{BAYES}} = \frac{M_t^H/\sigma + m/\gamma^2}{w_H(t)/\sigma^2 + 1/\gamma^2} \quad \text{и} \\ (\hat{\gamma}_1^2)_{\text{BAYES}} = \frac{1}{w_H(t)/\sigma^2 + 1/\gamma^2},$$

соответственно. Этот результат (при $\sigma = 1$) получен Норросом [61].

1.5 Байесовская оценка. Случай равномерного априорного распределения

Рассмотрим задачу отыскания Байесовской оценки параметра $\boldsymbol{\theta} \in \mathbb{R}^{n_\theta}$ в случае равномерного априорного распределения на n_θ -мерном параллелепипеде $\mathbf{r} = \prod_{i=0}^{n_\theta} [a_i, b_i]$.

Плотность распределения параметра $\boldsymbol{\theta}$ задается соотношением

$$p^\theta(\mathbf{x}) = \prod_{i=0}^{n_\theta} \frac{1}{b_i - a_i} \mathbb{1}_{[a_i, b_i]}(x_i) = \frac{1}{|\mathbf{r}|} \mathbb{1}_{\mathbf{r}}(\mathbf{x}),$$

где $\mathbb{1}_{\mathbf{r}}(\mathbf{x}) = \prod_{i=0}^{n_\theta} \mathbb{1}_{[a_i, b_i]}(x_i)$, $|\mathbf{r}| = \prod_{i=0}^{n_\theta} (b_i - a_i)$, а плотность условного распределения выражается формулой

$$p^\theta(\mathbf{x} | \mathcal{F}_t^\xi) = \frac{1}{Z_t^H} \mathbb{1}_{\mathbf{r}}(\mathbf{x}) \exp \left\{ \mathbf{x}^\top \boldsymbol{\psi}_t^H - \frac{1}{2} \mathbf{x}^\top \mathbf{R}_H(t) \mathbf{x} \right\} = \frac{1}{Z_t^H} \mathbb{1}_{\mathbf{r}}(\mathbf{x}) \Lambda_t^H(\mathbf{x}),$$

где процесс $Z^H = (Z_t^H)_{0 \leq t \leq T}$, задаваемый равенством $Z_t^H = \int_{\mathbf{r}} \Lambda_t^H(\mathbf{x}) d^{n_\theta} \mathbf{x}$, играет роль нормировочной постоянной, а процесс правдоподобия $\Lambda_t^H(\mathbf{x})$ задан согласно (1.13).

Аналитический расчет нормировочного множителя Z_t^H , условных среднего $E[\boldsymbol{\theta} | \mathcal{F}_t^\xi]$ и ковариационной матрицы $\text{cov}[\boldsymbol{\theta} | \mathcal{F}_t^\xi]$ в случае произвольного n_θ , вообще говоря, труден (соответствующий подсчет может быть выполнен численно, например, с использованием алгоритма, предложенного в [23]). Остановимся на расчете оценки для важного частного случая линейного сноса ($n_\theta = 1$), в котором наблюдаемый процесс ξ определяется стохастическим дифференциальным уравнением

$$d\xi_t = \theta_1 dt + \sigma dB_t^H, \quad (1.20)$$

где $\theta_1 \sim U(a, b)$. Результат подсчета в этой задаче составляет следующая

Теорема 3. Пусть в (1.20) θ_1 — равномерно распределенная на $[a, b]$ случайная величина, не зависящая от B_t^H . Тогда оптимальная в среднеквадратичном Байесовская оценка параметра θ_1 имеет вид

$$(\hat{\theta}_1)_{\text{BAYES}} = m_t^H + [Z_t^H w_H(t)]^{-1} [\Lambda_t^H(a) - \Lambda_t^H(b)], \quad (1.21)$$

а условная среднеквадратичная погрешность оценивания равна

$$\begin{aligned} (\hat{\gamma}_1^2)_{\text{BAYES}} = & \mathbb{E} \left[(\theta_1 - (\hat{\theta}_1)_{\text{BAYES}})^2 | \mathcal{F}_t^\xi \right] = [w_H(t)]^{-1} + \\ & + [Z_H(t)w_H(t)]^{-1} [\Lambda_t^H(a)(a - m_t^H) - \Lambda_t^H(b)(b - m_t^H)] - \\ & - [Z_H(t)w_H(t)]^{-2} [\Lambda_t^H(a) - \Lambda_t^H(b)]^2, \quad (1.22) \end{aligned}$$

где

$$\begin{aligned} Z_t^H &= \sqrt{\frac{2\pi}{w_H(t)}} \exp \left\{ \frac{1}{2} (m_t^H)^2 w_H(t) \right\} C_t^H, \\ C_t^H &= \Phi \left((b - m_t^H) \sqrt{w_H(t)} \right) - \Phi \left((a - m_t^H) \sqrt{w_H(t)} \right). \end{aligned} \quad (1.23)$$

Доказательство. Условное распределение $p^{\theta_1}(x | \mathcal{F}_t)$ легко подсчитать прямым вычислением, оно выражается формулой

$$p^{\theta_1}(x | \mathcal{F}_t^\xi) = \frac{1}{Z_t^H} \mathbb{1}_{[a,b]}(x) \exp \left\{ w_H(t) (xm_t^H - x^2/2) \right\},$$

где процесс Z^H определен согласно (1.23). Условные среднее и дисперсия получаются прямым вычислением соответствующих интегралов. \square

Рассмотрим несколько асимптотических свойств полученного байесовского фильтра (1.21).

При $a \rightarrow -\infty, b \rightarrow +\infty$ (т. е. в случае априори произвольного θ_1) байесовская оценка в (1.21) совпадает с оценкой максимального правдоподобия. Действительно, так как $\Lambda_t^H(x) \rightarrow 0$ при $x \rightarrow \pm\infty$, то второе слагаемое в (1.21) стремится к нулю при $x \rightarrow \pm\infty$, и $(\hat{\theta}_1)_{\text{BAYES}} \rightarrow m_t^H$.

При $t \rightarrow \infty$ байесовская оценка в (1.21) также совпадает с оценкой максимального правдоподобия. Действительно, так как $w_H(t) \rightarrow \infty$ при $t \rightarrow \infty$, то второй член в (1.21) стремится к нулю при $t \rightarrow \infty$, и $(\hat{\theta}_1)_{\text{BAYES}} \rightarrow m_t^H$.

Рассмотрим задачу отыскания оптимального момента остановки в (1.5). Функция штрафа в этой задаче задается выражением

$$\mathbb{E} \left[c\varrho + \mathbb{E} \left[(\theta_1 - (\hat{\theta}_1)_{\text{BAYES}})^2 | \mathcal{F}_\varrho^\xi \right] \right] = \mathbb{E} [c\varrho + \gamma_\varrho^H],$$

где случайный процесс $\gamma^H = (\gamma_t^H)_{0 \leq t \leq T}$ задается соотношением (1.22). Заметим, что при $t \rightarrow \infty$ выполняется $\gamma_t^H \rightarrow 0$. Для определения оптимального момента остановки необходимо найти

$$\varrho_{\text{BAYES}} = \arg \inf_{\varrho} E[c\varrho + \gamma_{\varrho}^H]. \quad (1.24)$$

В силу громоздкости формул (1.21), (1.22) и (1.23) не представляется возможным найти для ϱ_{BAYES} из (1.24) аналитическое описание (см. подходы в [65]), возможно лишь только численное оценивание этого момента остановки.

1.6 Выводы

В настоящей главе рассмотрена задача оценки параметров тренда случайного сигнала, моделируемого фрактальным диффузионным процессом. Вычислены оценки максимального правдоподобия для этих параметров, а также последовательные Байесовские оценки для нормального априорного распределения в предположении нелинейного тренда и для случая равномерного априорного распределения в предположении линейного тренда. Описана структура оптимального момента остановки наблюдений в случае нормального априорного распределения, а также сделаны замечания, относящиеся к случаю равномерного априорного распределения параметра линейного тренда. В случае, когда система функций, параметризующая тренд, является системой полиномов, оптимальный момент остановки является неслучайной величиной.

Глава 2

Ансамбли «слабых» детекторов в задачах обнаружения разладки

2.1 Введение

Математические задачи «о разладке» составляют часть статистики случайных процессов, связанную с вынесением (в режиме реального времени) суждений об их характеристиках на основе анализа их реализаций. В данной главе всюду принимается, что время дискретно: $t = 1, 2, \dots$. Пусть наблюдаемый случайный процесс $\xi = (\xi_t)_{t \geq 0}$ имеет структуру

$$\xi_t = \begin{cases} \xi_t^\infty, & \text{если } 0 \leq t < \theta, \\ \xi_t^0, & \text{если } t \geq \theta, \end{cases} \quad (2.1)$$

где $\xi^\infty = (\xi_t^\infty)_{t \geq 0}$ и $\xi^0 = (\xi_t^0)_{t \geq 0}$ суть два различных по своей структуре случайных процесса. Конкретный смысл процессов ξ^∞ и ξ^0 определяется приложением, в теории же принято говорить, что процесс ξ^∞ соответствует состоянию без разладки (*нормальному*), а ξ^0 — состоянию с разладкой (*аномальному*); момент θ принято называть моментом появления *разладки*. Задача «о разладке» состоит в следующем: пока наблюдения ξ согласуются с *нормальным* состоянием, требуется продолжать наблюдения. Если состояние изменяется на *аномальное*, требуется обнаружить изменение как можно скорее, избегая ложных тревог.

Класс задач «о разладке» охватывает самые разнообразные реальные ситуации: разладка — это нарушение однородности, это нарушение нормального хода

производственного процесса, это изменение вибрационных режимов, это изменение критических режимов в электроэнергетических системах, это появление «цели» в радиолокации, это выявление «разломов» в геологических данных, это возникновение землетрясений, цунами, это появление фронта ударных волн и т. д. [99]. Первые задачи типа задачи «о разладке» были рассмотрены еще в 40-х годах А. Вальдом в рамках теории последовательного различения статистических гипотез [96]. К настоящему времени теория оптимальных правил остановки получила значительное развитие, обзор которого см. в работах А. Н. Ширяева, M. Basseville и др. [9; 104; 106].

Отметим, что существует ряд общепринятых постановок задачи о разладке случайной последовательности, а именно ряд стандартных допущений о распределениях процессов ξ^∞ и ξ^0 и о моменте θ . Обзор этих допущений (стандартных моделей разладки) и некоторых оптимальных процедур обнаружения разладки приведен в разделе 2.2. Ситуация нарушения этих предположений и ее значение для эффективности обнаружения разладки обсуждается в разделе 2.3.1. В разделе 2.3.2 рассматривается проблема обнаружения разладки временного ряда с целью эффективного практического применения при нарушении стандартных предположений о модели возникающей разладки и определяется процедура обнаружения разладки на основе ансамбля. Критерии эффективности такого подхода и метод настройки параметров ансамблей приводятся в разделе 2.4. В конце главы в разделе 2.5 описаны результаты вычислительных экспериментов, направленные на численное сравнение эффективности ансамблей и классических процедур обнаружения разладки.

2.2 Известные модели и процедуры обнаружения в задачах о разладке стационарной случайной последовательности

Не задаваясь целью привести исчерпывающий обзор (если он вообще возможен) литературы по проблеме обнаружения изменения статистических свойств случайных процессов, приведем в разделе 2.2.1 ряд постановок задач о разладке,

для которых известны эффективные процедуры. (Для более полного обзора см., например, монографию [104].) Сами эти процедуры описываются в разделе 2.2.2.

2.2.1 Модели разладки стационарной случайной последовательности

В изложении ниже всюду предполагается, что наблюдаемый процесс $\xi = (\xi_t)_{t \geq 0}$ имеет конкретную структуру

$$\xi_t = \begin{cases} \sigma Z_t, & 0 \leq t < \theta, \\ \mu t + \sigma Z_t, & t \geq \theta, \end{cases} \quad (2.2)$$

где $\sigma > 0$, $\mu \neq 0$, θ – некоторый момент со значениями в $[0, \infty]$, и $Z = (Z_t)_{t \geq 0}$ – процесс стандартного нормального белого шума. Случай $\theta = 0$ соответствует появлению разладки в момент $t = 0$. Случай $\theta = \infty$ соответствует тому, что разладка не появляется и вовсе, следовательно, все время наблюдается процесс $\xi_t = \sigma Z_t$.

Обозначим $P_\theta = \text{Law}(\xi \mid \theta)$ – распределение вероятностей процесса $\xi = (\xi_t)_{t \geq 0}$ из (2.2), когда разладка происходит в момент времени θ . В частности, P_∞ есть распределение вероятностей процесса ξ , когда разладка не происходит. Математическое ожидание по мере P_θ обозначим E_θ .

Процедура обнаружения разладки – это *момент остановки*¹ $\tau = \tau(\omega)$ со значениями в $[0, \infty]$, интерпретируемый как момент подачи сигнала тревоги о появлении разладки. Рассмотрим два события:

$$\{\tau < \theta\} \text{ и } \{\tau \geq \theta\}.$$

Первое событие – это подача ложной тревоги ($\tau < \theta$). Когда же происходит второе событие, то естественно интересоваться, насколько велико время за-

¹ Случайная величина $\tau = \tau(\omega)$ со значениями в $[0, \infty]$ является моментом остановки, если при каждом $t \geq 0$ событие $\{\omega : \tau(\omega) \leq t\} \in \mathcal{F}_t$, где $\mathcal{F}_t = \mathcal{F}_t^\xi (= \sigma(\{\xi_s, 0 \leq s \leq t\}))$ есть σ -алгебра событий, порождаемых значениями $\{\xi_s, 0 \leq s \leq t\}$. Наглядный смысл этого условия состоит в том, что для каждого $t \geq 0$ решение вопроса о том, чтобы прекратить наблюдения или их продолжать зависит лишь от информации о процессе ξ , полученной на интервале времени $[0, t]$ и не зависит от «будущего».

паздывания $\tau - \theta$ при «правильной» подаче сигнала о появлении разладки. Сформулируем несколько вариантов оптимизационных задач о разладке.

Вариант А. Пусть $\theta = \theta(\omega)$ есть случайная величина со значениями в $[0, \infty]$, не зависящими от Z и такая, что

$$P(\theta = 0) = \pi, \quad P(\theta > t | \theta > 0) = e^{-\lambda t}, \quad (2.3)$$

где $\pi \in [0, 1)$ и λ являются известными величинами. Зафиксируем некоторую константу $\alpha \in (0, 1]$ и обозначим $\mathcal{M}_{(\alpha)} = \{\tau : P(\tau < \theta) \leq \alpha\}$ — класс тех моментов остановки τ (относительно $(\mathcal{F}_t)_{t \geq 0}$), для которых вероятность ложной тревоги $P(\tau < \theta)$ меньше или равна α . Требуется найти момент $\tau_{(\alpha)}^*$ (если он существует), являющийся решением задачи

$$E(\tau - \theta | \tau \geq \theta) \rightarrow \inf_{\tau \in \mathcal{M}_{(\alpha)}}. \quad (2.4)$$

Для решения этой *условно-вариационной* задачи рассматривают байесовскую постановку задачи о разладке — задачу на безусловный минимум

$$P(\tau < \theta) + c E(\tau - \theta)^+ \rightarrow \inf_{\tau \in \mathcal{M}}, \quad (2.5)$$

где \mathcal{M} — класс моментов остановки, $c > 0$ — константа и $E(\tau - \theta)^+ = E(\tau - \theta | \tau \geq \theta)P(\tau \geq \theta)$. Критерий (2.4) был предложен А. Н. Ширяевым [107].

В следующих вариантах (**В**, **С**, **Д** и **Е**) величина θ является просто параметром, принимающим числовые значения в $[0, \infty]$. При формулировке вариантов **В**, **С**, и **Д** фиксируется некоторое число $T > 0$ и рассматривается класс $\mathcal{M}_T = \{\tau : E_\infty \tau \geq T\}$, являющийся классом тех моментов остановки τ , для которых среднее время $E_\infty \tau$ до ложной тревоги (т. е. когда $\theta = \infty$) равно T .

Вариант В. Требуется найти момент $\tau_T^* \in \mathcal{M}_T$, являющийся решением задачи на минимум

$$\sup_{\theta \geq 0} E_\theta(\tau - \theta | \tau \geq \theta) \rightarrow \inf_{\tau \in \mathcal{M}_T}. \quad (2.6)$$

Этот критерий введен Поллаком [66].

Вариант С. Требуется найти момент $\tau_T^* \in \mathcal{M}_T$, являющийся решением задачи на минимум

$$\sup_{\theta \geq 0} \operatorname{ess\,sup}_{\omega} E_{\theta}((\tau - \theta)^+ | \mathcal{F}_{\theta})(\omega) \rightarrow \inf_{\tau \in \mathcal{M}_T}, \quad (2.7)$$

где $\mathcal{F}_{\theta} = \sigma(\{\xi_s, 0 \leq s \leq \theta\})$ и $\operatorname{ess\,sup}_{\omega}$ есть операция взятия существенного супремума². Описанный критерий называют критерием Лордена [49].

Вариант D. До сих пор рассматривались критерии, зависящие от времени запаздывания в подаче сигнала о появлении разладки; при этом ограничения на величину времени запаздывания не накладывались. Существуют приложения, в которых *необходимо* обнаружить разладку в течение заранее заданного интервала времени: для момента подачи тревоги необходимо выполнение условий $\theta < \tau \leq \theta + m$, где $m \geq 1$ [26]. Для решения этой задачи Дж. Мустакидесом [57] был предложен критерий, максимизирующий вероятность обнаружения разладки, оптимальным моментом остановки $\tau_T^* \in \mathcal{M}_T$ в котором является решение задачи на максимум

$$\inf_{\theta \geq 0} P_{\theta}(\theta < \tau \leq \theta + m | \tau > \theta) \rightarrow \sup_{\tau \in \mathcal{M}_T}. \quad (2.8)$$

Вариант E, в отличие от вариантов **A–D**, не вводит специальных критериев, а относится к задачам проверки статистических гипотез о равенстве средних в последовательности независимых наблюдений. По заданной выборке X_1, \dots, X_N требуется проверять гипотезу

$$H : \quad \mu_1 = \mu_2 = \dots = \mu_N = \mu \quad (2.9)$$

против альтернативы

$$K : \quad \mu = \mu_1 = \dots = \mu_{\theta} \neq \mu_{\theta+1} = \dots = \mu_N, \quad (2.10)$$

²Пусть $f = f(\omega)$ – неотрицательная случайная величина на вероятностном пространстве (Ω, \mathcal{F}, P) . Ее *существенным супремумом* (или *существенной верхней гранью*) называют нижнюю грань тех C , для которых $P(f(\omega) > C) = 0$. Это значение обозначают $\|f\|_{\infty}$, $\operatorname{ess\,sup}_{\omega} f(\omega)$ или $\sup \operatorname{vrai}_{\omega} f(\omega)$. Так что

$$\|f\|_{\infty} = \inf\{0 \leq C \leq \infty : P(f > C) = 0\}.$$

где $\mu_i = E X_i, i = 1, \dots, N$. Задачи проверки гипотез такого рода рассматривались в работах [73; 74].

2.2.2 Некоторые широко используемые статистические процедуры обнаружения разладки стационарной случайной последовательности

Приведем ряд процедур обнаружения разладки с указанием соответствующих критериев, для которых они являются оптимальными. Каждая процедура Π предписывает подавать сигнал тревоги в момент времени τ первого достижения некоторым процессом $S = (S_t)_{t \geq 0}$ заданного уровня $h > 0$: $\tau = \inf\{t \geq 0 : S_t \geq h\}$. Таким образом, различие между рассматриваемыми процедурами заключается в выборе процесса S (уровень h можем без ограничения общности выбрать равным единице).

Пусть в (2.1) (одномерные) плотности распределения процессов $\xi^\infty = (\xi_t^\infty)_{t \geq 0}$ и $\xi^0 = (\xi_t^0)_{t \geq 0}$ суть $p_\infty(\cdot)$ и $p_0(\cdot)$ соответственно. Обозначим

$$l_t = \frac{p_0(X_t)}{p_\infty(X_t)}, \quad \zeta_t = \log l_t, \quad t = 1, 2, \dots \quad (2.11)$$

отношение правдоподобия и его логарифм, соответственно.

Когда момент появления разладки θ является случайной величиной (вариант **A**), рассматривается следующий Байесовский алгоритм обнаружения разладки. Пусть θ принимает значения $0, 1, \dots$ с вероятностями

$$P(\theta = 0) = \pi, \quad P(\theta = t | \theta > 0) = p(1 - p)^{t-1}, \quad t = 1, 2, \dots,$$

где π и $p \in [0, 1]$ — параметры модели. Введем процесс **условной вероятности** $\pi = (\pi_t)_{t \geq 0}$, задаваемый рекуррентными соотношениями

$$\pi_t = \frac{\varphi_t}{1 + \varphi_t}, \quad \varphi_t = l_t \frac{p + \varphi_{t-1}}{1 - p}, \quad t = 1, 2, \dots \quad (2.12)$$

Величина $\pi_t = P(\theta < t | \mathcal{F}_t)$ интерпретируется как условная вероятность появления разладки до момента времени t . Тревога поднимается при значениях π_t , близких к единице. Процедура обнаружения разладки на основе статистики

условной вероятности была предложена А. Н. Ширяевым и является оптимальной для варианта **A** [107].

Процедура Ширяева-Робертса заключается в вычислении статистики $\psi = (\psi_t)_{t \geq 0}$, задаваемой рекуррентным соотношением

$$\psi_t = \sum_{\theta=1}^t \prod_{k=\theta}^t l_k = (1 + \psi_{t-1})l_t, \quad \psi_0 = 0, \quad t = 1, 2, \dots \quad (2.13)$$

Этот метод обнаружения разладки был независимо предложен А. Н. Ширяевым и С. Робертсом и является оптимальным для критерия Поллака (вариант **B**) [70; 107].

Процедура **кумулятивных сумм** (CUSUM) основана на рекуррентном подсчете процесса $T = (T_t)_{t \geq 0}$ согласно соотношениям

$$T_t = \max \left\{ 0, \max_{1 \leq \theta \leq t} \sum_{k=\theta}^t \zeta_k \right\} = \max(0, T_{t-1} + \zeta_t), \quad T_0 = 0, \quad t = 1, 2, \dots \quad (2.14)$$

Данная процедура была предложена Е. Пейджем, ее оптимальность для минимаксного критерия Лордена (вариант **C**) была показана Дж. Мустакидесом и Я. Ритовым для случая дискретного времени и А. Н. Ширяевым для случая непрерывного времени [54; 64; 69; 105].

Процесс **контрольных карт Шухарта** $S = (S_t)_{t \geq 0}$ [75] задается соотношением

$$S_t = \sum_{k=t-K+1}^t \zeta_k, \quad t = 1, 2, \dots, \quad (2.15)$$

где единственный параметр — величина K — некоторая заранее заданная переменная (размер скользящего окна). Дж. Мустакидесом [57] было показано, что статистика контрольных карт Шухарта при $K = 1$ максимизирует вероятность обнаружения разладки в варианте **E**.

Процедура, называемая в диссертационной работе **changepoint**, изначально была предложена для обнаружения изменения среднего значения гауссовской последовательности [74]. Эта процедура строится на основе статистики $S =$

$(S_t)_{t \geq 0}$, задаваемой соотношениями

$$S_t = \max_{t-K \leq \theta \leq t} \frac{\bar{X}_{\theta+1}^t - \bar{X}_{t-K}^\theta}{\sqrt{(t-\theta)^{-1} + (\theta-t+K)^{-1}} \sqrt{W}}, \quad t = 1, 2, \dots, \quad (2.16)$$

где K — размер скользящего окна, и

$$\begin{aligned} \bar{X}_i^j &= \frac{1}{j-i} \sum_{k=i}^j X_k, \\ W &= \frac{1}{K-2} \left[\sum_{k=t-K}^{\theta} (X_k - \bar{X}_{t-K}^\theta)^2 + \sum_{k=\theta+1}^t (X_k - \bar{X}_{\theta+1}^t)^2 \right]. \end{aligned}$$

2.3 Нарушение стандартных предположений о модели разладки. Ансамбли

2.3.1 Выполнимость широко используемых предположений о модели разладки

Большим ограничением стандартных моделей разладки (в том числе и моделей **A – E**) на практике является то, что при их использовании принимается, что распределения $p_\infty(\cdot)$ и $p_0(\cdot)$ — нормальные, с известными математическими ожиданиями и дисперсиями. А. Н. Ширяев пишет в своей работе [104]: «... значительный материал будет посвящен моделям, основанным на броуновском движении, что объясняется и тем, что такие модели представляют практический интерес, и тем, что для них во многих случаях удастся получить прозрачные и точные результаты. (... Напомним также, что в других случаях для соответствующих характеристик были получены лишь приближенные результаты.)» На практике это предположение нарушается по целому ряду пунктов, а именно:

1. Распределения $p_\infty(\cdot)$ и $p_0(\cdot)$ являются ненормальными.
2. Параметры распределений $p_\infty(\cdot)$ и $p_0(\cdot)$ неизвестны точно ни до появления разладки, ни тем более после появления разладки.
3. Присутствует временная корреляция между наблюдениями ξ_t и $\xi_{t+\Delta t}$ (например, процесс ξ обладает длинной памятью, см. главу 1).
4. Присутствуют тренды или циклы, см. главу 3.

5. Разладка не продолжается бесконечное время, возможен *пропуск разладки*, т.е. ситуация, в которой выдается ошибочный сигнал об отсутствии разладки.

В ряде работ исследуется вопрос о нарушении стандартных предположений о модели разладки, ведущий к снижению эффективности ее обнаружения [9; 44; 53; 71; 83]. В диссертационной работе рассматривается следующая общая ситуация. Пусть наблюдаемый случайный процесс $\xi = (\xi_t)_{t \geq 0}$ имеет структуру

$$\xi_t = \begin{cases} \xi_t^\infty, & \text{если } t \in \mathcal{T}_\infty, \\ \xi_t^0, & \text{если } t \in \mathcal{T}_0, \end{cases}$$

где случайные процессы $\xi^\infty = (\xi_t^\infty)_{t \geq 0}$ и $\xi^0 = (\xi_t^0)_{t \geq 0}$ имеют (вообще говоря, неизвестные) плотности $p_\infty(\cdot)$ и $p_0(\cdot)$ соответственно, а множества \mathcal{T}_∞ и \mathcal{T}_0 соответствуют состоянию без «разладки» (*нормальному*) и с «разладкой» (*аномальному*), соответственно. В ряде задач, возникающих на практике, разладка имеет конечную длительность либо должна быть обнаружена в течение заданного времени [26; 81; 85]. Поэтому мы рассматриваем ситуацию «кратковременного изменения», в которой

$$\mathcal{T}_\infty = [0, \theta) \cup [\theta + \Delta, \infty) \quad \text{и} \quad \mathcal{T}_0 = [\theta, \theta + \Delta),$$

предполагающую конечную длительность аномального состояния. Такая ситуация допускает возникновение ошибок обоих родов (как ложной тревоги, так и пропуска цели) и является более реалистичной при описании эффективности процедур обнаружения разладки. Пока наблюдения за процессом ξ согласуются с нормальным состоянием, требуется продолжать наблюдения. Если состояние изменяется, требуется обнаружить изменение как можно скорее, избегая ложных тревог. При возврате к нормальному состоянию, однако, требуется как можно скорее обнаружить последнее, «выключив» сигнал тревоги о наличии разладки.

2.3.2 Ансамбли «слабых» детекторов

Пусть $\Pi_1, \dots, \Pi_{n_\Pi}$ обозначают n процедур обнаружения разладки, причем каждая процедура Π_k предписывает подавать сигнал тревоги в момент τ_k первого

выхода некоторого процесса $S^k = (S_t^k)_{t \geq 0}$ на некоторый уровень $h_k > 0$: $\tau_k = \inf\{t \geq 0 : S_t^k \geq h_k\}$. Рассмотрим далее множество *сигналов* $s^k = (s_t^k)_{t \geq 0}$, $k = 1, \dots, n_{\Pi}$, задаваемых соотношениями $s_t^k = S_t^k/h_k$, $t = 1, 2, \dots$

Определение 1. Процедура А обнаружения разладки называется *ансамблем* детекторов $\Pi_1, \dots, \Pi_{n_{\Pi}}$, если она предписывает подавать тревогу в момент τ_A выхода процесса $a = (a_t)_{t \geq 0}$ на заданный уровень $h_A > 0$: $\tau_A = \inf\{t \geq 0 : a_t \geq h_A\}$:

$$a_t = \psi(\boldsymbol{\lambda}; \mathbf{S}_t^1, \dots, \mathbf{S}_t^{n_{\Pi}}), \quad (2.17)$$

где $\boldsymbol{\lambda} \in \mathbb{R}^d$ ($d \geq n_{\Pi}$) и $\mathbf{S}_t^k = \{s_s^k, 0 \leq s \leq t\}$ — история сигнала $s^k = (s_t^k)_{t \geq 0}$ до момента времени t , $k = 1, \dots, n_{\Pi}$.

Заметим, что конкретный ансамбль полностью определен выбором «агрегирующей функции» $\psi(\cdot)$. Ее параметры $\boldsymbol{\lambda} \in \mathbb{R}^d$ могут быть выбраны по размеченной выборке с помощью оптимизации некоторой меры эффективности (она будет введена ниже в п. 2.4. Несколько примеров конкретный ансамблей представлено ниже.

Ансамбль на основе голосования по большинству задается следующим выбором агрегирующей функции

$$\psi_{\text{MAJ}}(\boldsymbol{\lambda}; \mathbf{S}_t^1, \dots, \mathbf{S}_t^{n_{\Pi}}) = \frac{2}{n_{\Pi}} \sum_{k=1}^{n_{\Pi}} \mathbb{1}_{\{s_t^k \geq 1\}}(t). \quad (2.18)$$

Задавая $h_{\text{MAJ}} = 1$, получим правило остановки, предписывающее подавать сигнал тревоги о появлении разладки в первый момент времени τ_{MAJ} , когда число $n^+ = \sum_{k=1}^{n_{\Pi}} \mathbb{1}_{\{s_t^k \geq h_{\text{MAJ}}\}}(t)$ «голосов», поданных за появление разладки, превысит число $n^- = n_{\Pi} - n^+$ «голосов», поданных против разладки.

Другим вариантом ансамбля является *взвешенное голосование*, агрегирующая функция для которого задается в виде

$$\psi_{\text{WEIGHT}}(\boldsymbol{\lambda}; \mathbf{S}_t^1, \dots, \mathbf{S}_t^{n_{\Pi}}) = \sum_{k=1}^{n_{\Pi}} \lambda_k s_t^k. \quad (2.19)$$

Как и выше, зададим порог $h_{\text{WEIGHT}} = 1$ и выберем $\lambda_1, \dots, \lambda_{n_{\Pi}}$ таким образом, чтобы получить оптимальное значение некоторого показателя эффективности обнаружения разладки. Получим правило остановки, задаваемое моментом

$\tau_{\text{WEIGHT}} = \inf\{t \geq 0 : a_t \geq h_{\text{WEIGHT}}\}$. При появлении разладки детекторы $s^1, \dots, s^{n_{\text{п}}}$ накапливают информацию из наблюдений $\{\xi_1, \dots, \xi_t\}$, а агрегирующая функция усиливает сигнал детекторов, используемый для подачи тревоги.

При появлении разладки, информация, предварительно фильтруемая детекторами, накапливается со временем и формирует сигнал, на основании которого подается тревога о появлении разладки. Однако, на практике накопление статистики о появлении разладки может происходить медленно ввиду неточного задания модели разладки или ее небольшой «магнитуды» (т.е. различия в распределениях до и после разладки). Поэтому для более эффективного обнаружения разладки полезно использовать всю историю $\mathbf{S}_t^k = \{s_u^k, 1 \leq u \leq t\}$ сигнала $s^k = (s_t^k)_{t \geq 0}$ (а не только его текущее значение s_t^k). Для этого рассмотрим два класса ансамблей, использующих значения сигналов с запаздыванием величиной вплоть до p .

Ансамбль на основе *взвешенного голосования с историей* p задается выбором агрегирующей функции

$$\psi_{\text{WEIGHT-}p}(\boldsymbol{\lambda}; \mathbf{S}_t^1, \dots, \mathbf{S}_t^n) = \sum_{j=0}^p \sum_{k=1}^n \lambda_{kj} s_{t-j}^k. \quad (2.20)$$

Естественно, как и выше, задать $h_{\text{WEIGHT-}p} = 1$, поскольку такая «нормировка» может быть достигнута масштабированием параметров $\boldsymbol{\lambda} = (\lambda_{kj}) \in \mathbb{R}^{n \times p}$. Получим правило остановки, задаваемое моментом $\tau_{\text{WEIGHT-}p} = \inf\{t \geq 0 : a_t \geq h_{\text{WEIGHT-}p}\}$.

Еще один класс ансамблей, которые мы рассмотрим в диссертационной работе, основан на использовании линейного классификатора в пространстве признаков, образованном значениями сигналов каждого из детекторов. В качестве такого классификатора рассмотрим *логистическую регрессию*, агрегирующую функцию для которой можно записать в виде

$$\psi_{\text{LOG-}p}(\boldsymbol{\lambda}; \mathbf{S}_t^1, \dots, \mathbf{S}_t^n) = \sigma\left(\sum_{j=0}^p \sum_{k=1}^n \lambda_{kj} s_{t-j}^k - \lambda_0\right). \quad (2.21)$$

Значение $a_t = \psi_{\text{LOG-}p}(\boldsymbol{\lambda}; \mathbf{S}_t^1, \dots, \mathbf{S}_t^n)$ статистики ансамбля можно интерпретировать как условную вероятность события $t \in \mathcal{T}_0$ (т.е. действия «аномального состояния») при заданной истории $\Xi_t = \{\xi_u, 0 \leq u \leq t\}$ наблюдений до момента

времени t . Заметим, что для такого ансамбля пороговое значение $h_{\text{Log-}p} \in (0, 1)$, а момент останковки $\tau_{\text{Log-}p} = \inf\{t \geq 0 : a_t \geq h_{\text{Log-}p}\}$.

2.4 Критерии качества обнаружения разладки

Для того, чтобы охарактеризовать эффективность обнаружения разладки на реальных данных, необходимо иметь:

- специальные функционалы, описывающие, какие эффекты следует ожидать на практике при применении полученных методов;
- множество размеченных данных, т. е. таких реализаций $\xi = (\xi_t)_{t \geq 0}$, для которых множества \mathcal{T}_0 и \mathcal{T}_∞ известны (например, размечены пользователем или получены в ходе моделирования).

В диссертационной работе используется подход, принятый в области машинного обучения, состоящий в использовании размеченной выборки $\mathbf{X}^\ell = \{(X^i, Y^i)\}_{i=1}^\ell$. Каждая точка $(X^i, Y^i) \in \mathbf{X}^\ell$ в нашей схеме — это пара временных рядов $X^i = (X_t^i)_{t \geq 0}$ и $Y^i = (Y_t^i)_{t \geq 0}$, причем X^i соответствует отрезку наблюдений фиксированной длины T . Временной ряд Y^i — это разметка: в каждый момент времени $t \in [0, T]$ значение $Y_t^i \in \{0, 1\}$, где 1 обозначает разладку, а 0 — ее отсутствие. Таким образом, Y^i можно интерпретировать как индикатор аномального состояния: $Y_t^i = \mathbb{1}_{\mathcal{T}_0}(t)$. Поскольку мы рассматриваем ситуации, в которых возможно появление и исчезновение разладки, то значения $Y_t^i, t \in [0, T]$ формируются следующим образом: изначально в отсутствие разладки $Y_t^i = 0$; при появлении разладки происходит переключение $Y_t^i = 1$; как только разладка закончилась, Y_t^i снова становится равным 0.

Рассмотрим ряд общепринятых функционалов качества для процедур обнаружения разладки. Пусть, как и ранее, процедура Π заключается в подаче сигнала тревоги в момент $\tau = \inf\{t \geq 0 : s_t \geq h\}$ первого выхода статистики $s = (s_t)_{t \geq 0}$ на заданный уровень h .

1. Зависимость средней задержки в обнаружении разладки $\text{ADD}(\tau) = E_\theta(\tau - \theta | \tau > \theta)$ (average detection delay) от среднего времени до ложной тревоги $\text{ARL}(\tau) = E_\infty \tau$ (average run length, ARL) является наиболее часто используемой оперативной характеристикой процедуры детектирования разладки [50; 55; 56]. Ее подсчет заключается в оценке средней длины

серии, требуемой статистике S для выхода на уровень h , для случаев распределений $p_0(\cdot)$ и $p_\infty(\cdot)$, соответственно. Средняя длина серии должна быть как можно больше, когда действует распределение $p_\infty(\cdot)$, и, насколько возможно, меньше, когда действует распределение $p_0(\cdot)$.

2. Другая оперативная характеристика основана на оценивании все той же средней задержки в обнаружении разладки $\text{ADD}(\tau)$ на каждом уровне вероятности ложной тревоги $\text{PFA}(\tau) = P_\infty(\tau < \theta)$ (probability of false alarm, PFA), причем последняя понимается как ожидаемая доля таких точек заданной выборки, в которых при заданной задержке в обнаружении $\text{ADD}(\tau)$ имела место ложная тревога, а ее эмпирическая оценка $\frac{1}{\ell} \sum_{i=1}^{\ell} \mathbf{1}_{\{\tau_i < \theta_i\}}(i)$, где $\tau_i = \inf\{t \geq 0 : s_t^i \geq h\}$ — момент подачи сигнала тревоги в точке $(X^i, Y^i) \in \mathbf{X}^\ell$, а θ_i — момент появления разладки в этой точке.
3. Как нетрудно понять, характеристики из пп. 1–2 не описывают ситуацию, в которой возможен пропуск цели (ввиду конечной длительности разладки), поскольку они не включают соответствующий функционал. Для описания этой ситуации рассмотрим две следующие характеристики. Кривые типа «точность — полнота» используются в области машинного обучения для описания эффективности алгоритмов классификации и могут быть применены и при характеристике алгоритмов обнаружения разладок. Естественно рассматривать разладку как некоторое событие, подлежащее «выделению» или «фильтрации» из входящего потока наблюдений. С другой стороны, для каждого значения порога h любая процедура обнаружения разладки «выделяет» из входящего потока некоторое множество событий, интерпретируемых этой процедурой как разладки. Тогда «полнотой» назовем отношение количества всех происшедших за время наблюдений разладок, которые были выделены процедурой, к полному числу происшедших разладок. «Точностью» же назовем отношение количества всех «выделенных» процедурой разладок, которые в самом деле ими являются, к полному числу выделенных разладок.
4. В диссертационной работе предлагается следующий новый функционал для описания эффективности процедуры обнаружения разладки.

Определение 2. Математическое ожидание потерь, свойственных процедуре обнаружения разладки Π ,

$$\mathbf{F}(\Pi) = c_\infty \mathbb{E}_\infty \left[\frac{\sum \mathbf{1}_{\{s_t \geq h\}}(t) \mathbf{1}_{\mathcal{T}_\infty}(t)}{\sum \mathbf{1}_{\mathcal{T}_\infty}(t)} \right] + c_0 \mathbb{E}_0 \left[\frac{\sum \mathbf{1}_{\{s_t < h\}}(t) \mathbf{1}_{\mathcal{T}_0}(t)}{\sum \mathbf{1}_{\mathcal{T}_0}(t)} \right], \quad (2.22)$$

где c_0 и c_∞ суть потери за единицу времени, сопутствующие ошибочным решениям о наличии и отсутствии разладки, соответственно. Согласно этому определению, процедура обнаружения разладки Π тем лучше, чем меньше сопутствующие ей ожидаемые потери; оптимальная процедура обнаружения разладки определяется как решение задачи $\mathbf{F}(\Pi) \rightarrow \min_s$. Когда $c_\infty = c_0 = 1$, мера $\mathbf{F}(\Pi)$ имеет интерпретацию *средней относительной точности идентификации состояния наблюдений*, т. е. доли времени, в течение которого процедура обнаружения Π подает верный сигнал о наличии либо отсутствии разладки.

2.4.1 Вычислительный алгоритм настройки параметров ансамбля

Оптимизация штрафа, заданного в (2.22), позволяет осуществить выбор параметров $\boldsymbol{\lambda} \in \mathbb{R}^d$ ансамбля A в (2.17) и получить ансамбль A^* , для которого $\mathbf{F}(A^*) = \inf_{\boldsymbol{\lambda} \in \mathbb{R}^d} \mathbf{F}(A)$. Поскольку прямое вычисление математических ожиданий в (2.22) в общем случае невозможно, рассмотрим аппроксимацию $\mathbf{F}_{\text{ЕМР}}(\Pi)$ функции потерь $\mathbf{F}(\Pi)$, называемую эмпирическим риском:

$$\begin{aligned} \mathbf{F}_{\text{ЕМР}}(\Pi) &= c_\infty \frac{1}{\ell} \sum_{i=1}^{\ell} \left[\frac{\sum \mathbf{1}_{\{s_t^i \geq h\}}(t) \mathbf{1}_{\mathcal{T}_\infty^i}(t)}{\sum \mathbf{1}_{\mathcal{T}_\infty^i}(t)} \right] + c_0 \frac{1}{\ell} \sum_{i=1}^{\ell} \left[\frac{\sum \mathbf{1}_{\{s_t^i < h\}}(t) \mathbf{1}_{\mathcal{T}_0^i}(t)}{\sum \mathbf{1}_{\mathcal{T}_0^i}(t)} \right] \\ &= \frac{1}{\ell} \sum_{i=1}^{\ell} \left\{ \frac{c_\infty}{T_\infty^i} \sum_{t \in \mathcal{T}_\infty^i} \mathbf{1}_{\{s_t^i \geq h\}}(t) + \frac{c_0}{T_0^i} \sum_{t \in \mathcal{T}_0^i} \mathbf{1}_{\{s_t^i < h\}}(t) \right\}, \end{aligned} \quad (2.23)$$

где $s^i = (s_t^i)_{t \geq 0}$ — траектория процесса s , подсчитанная по наблюдениям X^i , а \mathcal{T}_∞^i и \mathcal{T}_0^i суть длительности нормального и аномального состояний в точке (X^i, Y^i) . Согласно классическому подходу статистической теории обучения, минимизация эмпирического риска $\mathbf{F}_{\text{ЕМР}}(\Pi)$ дает процедуру обнаружения разладки Π^* , для которой ожидаемые потери $\mathbf{F}(\Pi^*)$ близки к своему минимуму. Ввиду разрывности градиента прямая оптимизация эмпирического риска $\mathbf{F}_{\text{ЕМР}}(\Pi)$ трудна (если во-

обще возможна); по этой причине рассмотрим сглаженную версию эмирического риска $\mathbf{F}_{\text{EMP}}(\Pi)$, задаваемую соотношением

$$\mathbf{F}_{\text{DIFF}}(\Pi) = \frac{1}{\ell} \sum_{i=1}^{\ell} \left\{ \frac{c_{\infty}}{T_{\infty}^i} \sum_{t \in T_{\infty}^i} \sigma(s_t^i - h) + \frac{c_0}{T_0^i} \sum_{t \in T_0^i} \sigma(h - s_t^i) \right\}, \quad (2.24)$$

где $\sigma(x) = 1/(1 + e^{-x})$ — логистическая функция. Так определенная функция риска является дифференцируемой по параметрам ансамбля $\boldsymbol{\lambda} \in \mathbb{R}^d$, и может быть оптимизирована градиентными методами.

2.5 Сравнительный анализ эффективности ансамблей и классических процедур обнаружения разладки

Для всестороннего сравнения эффективности разработанных процедур обнаружения разладки была проведена серия вычислительных экспериментов с использованием искусственных наборов данных, различных по распределениям вероятностей их значений и изменяющимся при разладке параметрам. Свойства использованных данных приведены в таблице 2.1. В этих экспериментах для каждого типа данных были смоделированы 1024 независимые реализации длины $T = 1000$, использовавшиеся для подстройки параметров ансамблей, и другие 1024 реализации такой же длины для оценки качества и вычисления оперативных характеристик. Для 5 «слабых» детекторов и 7 вариантов ансамблей, описанных в таблице 2.2, были вычислены оперативные характеристики, описанные в разделе 2.4. Таблица 2.3 представляет результаты сравнения описанных процедур обнаружения разладки в терминах стандартной для области машинного обучения характеристики «площадь под кривой “точность–полнота”».

Набор	Тип данных	Изменяемые параметры	Момент разладки θ и длительность Δ	Магнитуда разладки
WhiteNoise	Белый гауссовский шум	математическое ожидание μ	случайные, $\theta \sim U(200, 800)$, $\Delta \sim U(5, 100)$	случайная, $\mu \sim U(0.1, 2)$
Fractal	Фрактальный гауссовский шум			
Cauchy	Белый шум Коши			
ARMA-AR	Процесс ARMA(10, 3)	AR-члены φ_i		случайная
ARMA-MA	Процесс ARMA(10, 3)	MA-члены θ_j		случайная
GARCH1-ARMA	GARCH(1, 1) + ARMA(10, 3)	$\alpha_1, \beta_1, \varphi_i, \theta_j$		случайная
GARCH1	Процесс GARCH(1, 1)	α_1, β_1		случайная, $\alpha_1 \sim U(.4, .8)$, $\beta_1 \sim U(.1, .2)$,

Таблица 2.1: Характеристики использованных при сравнительном анализе искусственных наборов данных

Процедура	Статистика процедуры	Параметры процедуры	Параметры распределения
Posterior	Статистика апостериорной вероятности (2.12)	$p = 0.05$	
SR	Статистика Ширяева-Робертса (2.13)		$p_\infty = \mathcal{N}(0, 1),$
CUSUM	Статистика кумулятивных сумм (2.14)		$p_0 = \mathcal{N}(1, 1)$
Shewhart	Статистика контрольных карт (2.15)	$K = 20$	
Changepoint	Статистика changepoint (2.16)	$K = 24$	
Maj	Ансамбль голосования большинством (2.18)		$p_\infty = \mathcal{N}(0, 1),$
Weight- p	Ансамбль взвешенного голосования (2.20)	$p \in \{0, 1, 2, 3\}$	$p_0 = \mathcal{N}(1, 1)$
Log- p	Ансамбль-логистическая регрессия (2.21)	$p \in \{0, 1, 2, 3\}$	

Таблица 2.2: Характеристики использованных при сравнительном анализе процедур обнаружения разладки

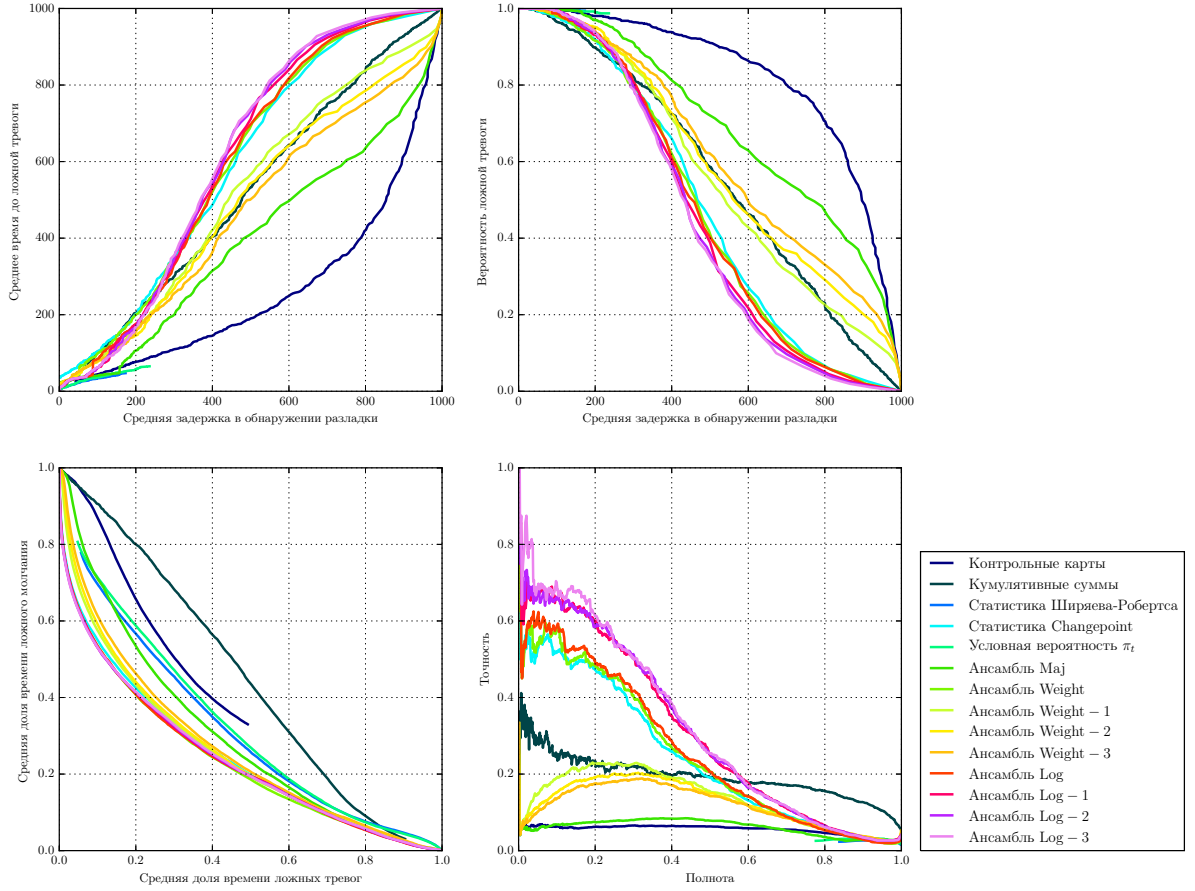


Рисунок 2.1: Сравнительный анализ оперативных характеристик «слабых» детекторов и ансамблей для набора данных Cauchy

Для всех наборов данных, исключая набор GARCH1-ARMA, предложенные процедуры показывают более высокую эффективность обнаружения разладки в терминах кривой «точность–полнота» при близких параметрах времени задержки в обнаружении разладки. Сказанное означает, что при одинаковой задержке в обнаружении разладки (скорости реакции на изменение распределения данных) предложенные процедуры выделяют больше истинных разладок из потока наблюдений, чем «слабые» детекторы. Ансамбли также показывают и значительно более высокую точность сегментации (согласно кривой зависимости средней доли времени ложного молчания от средней доли времени ложной тревоги). Последнее означает, что предложенные процедуры обеспечивают существенно лучшую среднюю относительную точность идентификации состояния наблюдений.

Для более детального анализа эффективности работы ансамблей была проведена визуализация их траекторий на выбранных точках $(X^i, Y^i) \in \mathbf{X}^\ell$. Сравнение траекторий, представленное на рис. 2.3, показывает, что сигнал о наличии разладки, генерируемый ансамблем, существенно «сильнее» коррелирует с ис-

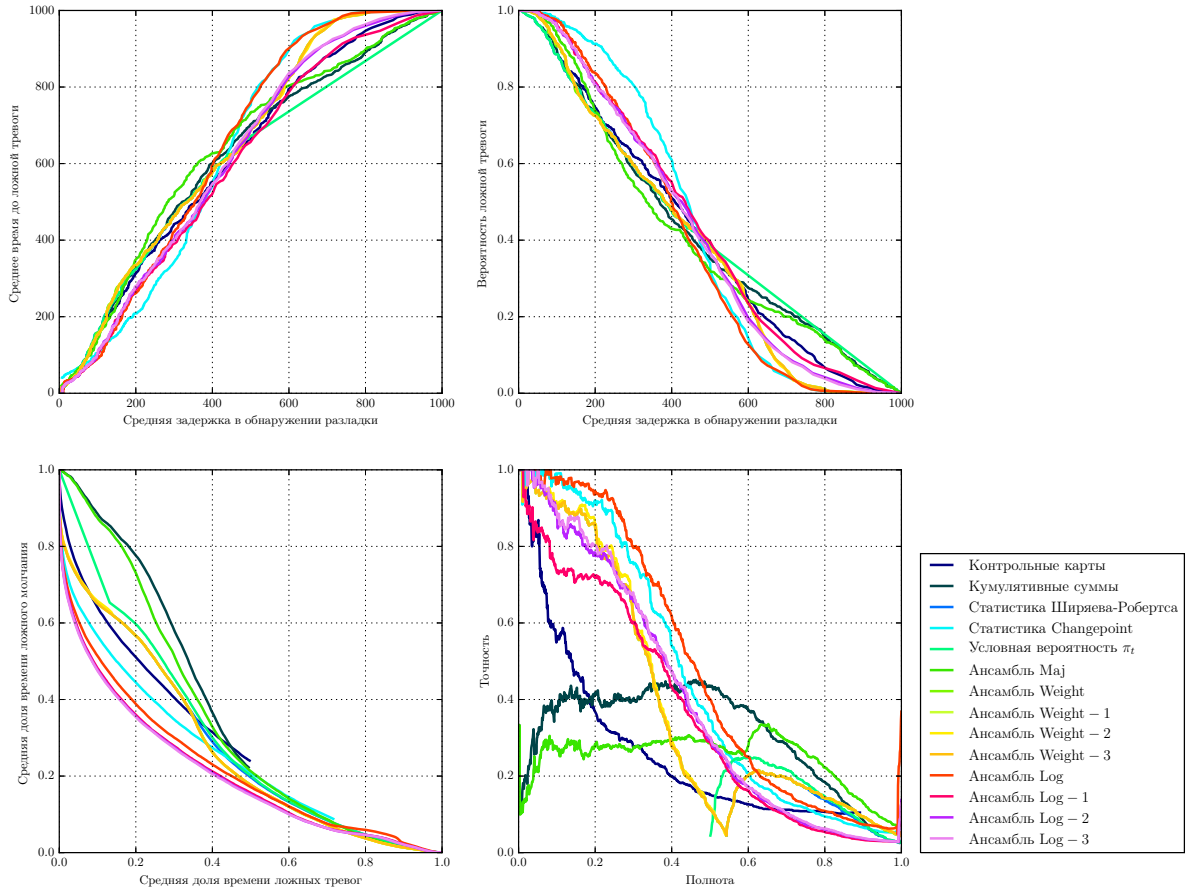


Рисунок 2.2: Сравнительный анализ оперативных характеристик «слабых» детекторов и ансамблей для набора данных Fractal

тинным индикатором разладки, чем сигнал любого из рассмотренных «слабых» детекторов.

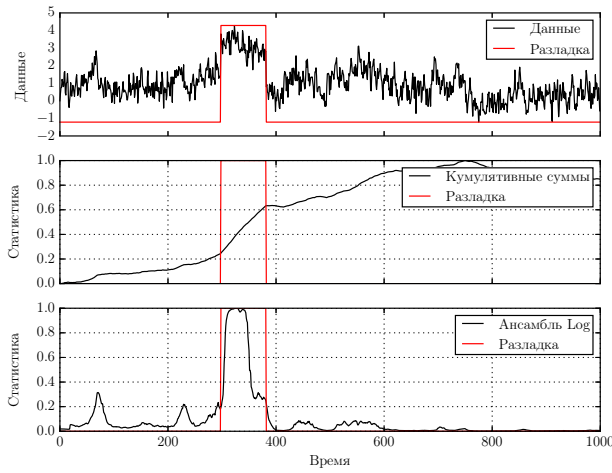
2.6 Выводы

В настоящей главе рассмотрен алгоритм обнаружения разладки случайной последовательности, модель разладки которой неизвестна. Этот алгоритм основан на одновременном использовании ансамбля статистик, лежащих в основе классических процедур обнаружения («слабых» детекторов), для достижения лучшей предсказательной силы. Рассмотрен ряд конкретных реализаций процедур обнаружения разладки на основе ансамблей «слабых» детекторов, отличающихся выбором конкретной агрегирующей функции. Предложен новый функционал качества процедур обнаружения разладки, физической смысл которого заключается в вероятности безошибочной сегментации траектории наблюдений на промежутки без разладки и с разладкой. Предложен градиентный алгоритм оптимизации дифференцируемой аппроксимации этого функционала по обучаю-

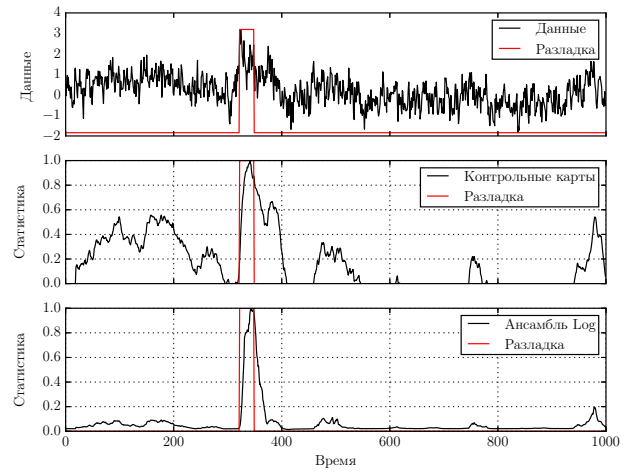
щей выборке. В ходе вычислительного эксперимента проведено эмпирическое сравнение эффективности ансамблей и классических процедур обнаружения разладки, показавшее эффективность первых в модельных задачах обнаружения разладки.

Процедура	WhiteNoise	Fractal	Cauchy	GARCH1	ARMA-AR	ARMA-MA	GARCH1-ARMA
Shewhart	77.52	24.44	05.45	32.08	19.80	76.37	40.00
CUSUM	61.11	30.70	19.24	59.88	28.74	89.90	75.44
SR	22.22	6.11	.40	50.06	24.17	7.15	72.72
Changepoint	60.62	45.42	24.18	21.94	13.03	57.15	22.98
Posterior π_t	27.38	7.76	.66	53.60	29.00	35.69	74.58
MAJ	62.13	24.62	6.11	47.80	28.74	92.71	67.01
WEIGHT – 0	71.73	38.94	25.08	55.62	24.94	79.48	67.23
WEIGHT – 1	71.58	38.97	13.46	57.65	29.60	91.92	71.28
WEIGHT – 2	73.89	39.63	12.29	56.57	30.45	91.13	69.28
WEIGHT – 3	73.25	38.98	11.61	57.83	26.24	90.70	72.11
LOG – 0	77.25	48.64	25.90	51.35	23.29	87.72	68.35
LOG – 1	76.27	36.20	31.03	50.03	23.49	88.47	65.97
LOG – 2	78.01	39.74	31.30	49.35	27.99	88.93	66.43
LOG – 3	78.85	40.31	32.24	49.08	27.99	88.88	66.77

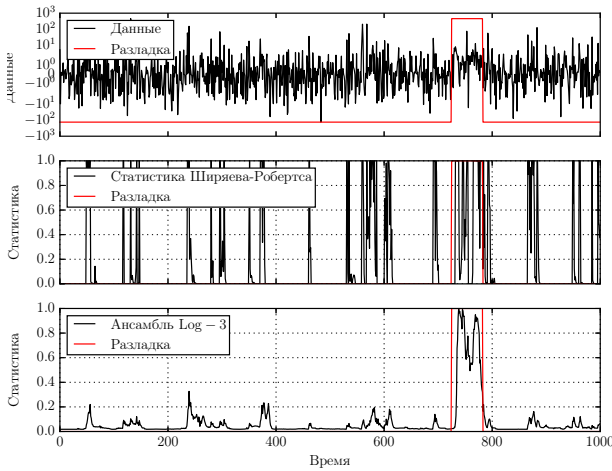
Таблица 2.3: Сравнительный анализ эффективности «слабых» детекторов и ансамблей для всех наборов данных в терминах площади под кривой «точность–полнота». Полужирным шрифтом выделен лучший результат для каждого набора данных.



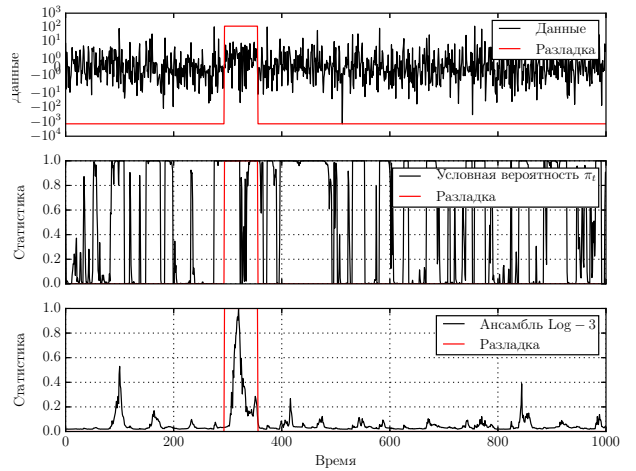
(а) Пример данных (верхний график), траектории статистики кумулятивных сумм (средний график) и статистики ансамбля LOG-0 (нижний график).



(б) Пример данных (верхний график), траектории статистики контрольных карт (средний график) и статистики ансамбля LOG-0 (нижний график).



(с) Пример данных (верхний график), траектории статистики Ширяева-Робертса (средний график) и статистики ансамбля LOG-3 (нижний график).



(д) Пример данных (верхний график), траектории статистик апостериорной вероятности (средний график) и ансамбля LOG-3 (нижний график).

Рисунок 2.3: (а)–(б): Сравнительный анализ эффективности «слабых» детекторов и ансамблей для набора данных Fractal.

(с)–(д): Сравнительный анализ эффективности «слабых» детекторов и ансамблей для набора данных Cauchy.

Глава 3

Математические модели сигналов с квазипериодическим трендом и обнаружение их разладок

3.1 Введение

Как было подчеркнуто в главе 2, выполнимость широко используемых предположений о модели разладки на практике всегда лишь частичная; это обстоятельство усугубляется при работе с циклическими временными сигналами. А поскольку многие сигналы, относящиеся к системам с интенсивным программным обеспечением, являются периодическими по своей природе, для обнаружения их разладок необходимо иметь математическую модель, позволяющую адекватно отображать их свойства. В настоящей главе рассматривается методология моделирования и оценивания квазипериодических сигналов и обнаружения их разладок для эффективного мониторинга сигналов больших систем с интенсивным ПО. В разделе 3.2 проведен краткий обзор литературы, связанной с моделированием и оцениванием параметров периодических временных рядов. Разделы 3.3 и 3.4 представляют математические модели и вычислительные алгоритмы оценивания параметров квазипериодических сигналов систем с интенсивным ПО в предположениях наличия длинной памяти и стохастических циклов, соответственно. Раздел 3.5 представляет модель разладки и алгоритм ее обнаружения на основе оценок, выводимых в двух предыдущих разделах и главе 2. Наконец, в разделе 3.6 содержатся результаты вычислительных экспери-

ментов, направленных на численное исследование эффективности предложенной методологии обнаружения разладок и аномалий с использованием искусственных данных.

3.2 Задача оценивания параметров сигнала с квазипериодическим трендом

Проблеме моделирования и оценивания периодических временных сигналов посвящен весьма большой объем литературы. В диссертационной работе проводится лишь краткий ее обзор и детально описывается два хорошо известных подхода к выделению периодического тренда, использованных ниже для сравнительного анализа с разработанными в работе подходами.

3.2.1 Известные в литературе модели сигналов с периодической составляющей

Задачи оценивания параметров сезонных сигналов и их математические модели рассматриваются в литературе с 1960-х гг. Одним из первых методов, предложенных для прогнозирования ожидаемого значения временного ряда, является алгоритм на основе экспоненциально взвешенного скользящего среднего (exponentially weighted moving average, EWMA), подробно описанный Питером Винтерсом [91]. В этом методе предлагается прогнозировать среднее значение $E \xi_t$ временного ряда $\xi = (\xi_t)_{t \geq 0}$ величиной

$$\hat{\mu}_t = (1 - \alpha)\hat{\mu}_{t-1} + \alpha\xi_t, \quad t \geq 0, \quad (3.1)$$

где постоянная $\alpha \in (0, 1)$ — параметр алгоритма. В случае, когда сигнал не имеет четкой сезонности или долгосрочного тренда, т. е. когда среднее значение $\mu = E \xi_t$ не зависит от времени, математическое ожидание оценки $E \hat{\mu}_t = E \xi_t = \mu$. В случае же, когда процесс наблюдений ξ нестационарный, статистические свойства оценки $\hat{\mu}_t$ не будут столь очевидными; для улучшения эффективности оценивания в таких ситуациях прибегают к методам экспоненциального сглаживания более высоких порядков. В настоящей работе они опускаются и для

базового сравнения с разработанными методами рассматривается оригинальный подход из (3.1).

Еще одним алгоритмом, эффективность которого была исследована в задачах обнаружения разладки, является алгоритм на основе метода главных компонент (principal component analysis, PCA) [6; 45–47; 63]. Он заключается в классификации поступающих наблюдений на «нормальные» и «аномальные» с помощью разделения некоторого пространства наблюдений на главное и остаточное подпространства. Соответствующий алгоритм получил название «гусеница» или анализ сингулярного спектра (singular spectrum analysis, SSA) [88]. Пусть процесс наблюдений $\xi = (\xi_t)_{t \geq 0}$ имеет некоторый известный период T и пусть доступна выборка $\mathbf{X}^\ell = \{(X_k, t_k)\}_{k=1}^\ell$, $X_k = \xi_{t_k}$, $k = 1, \dots, \ell$, — реализация сигнала до момента времени t_ℓ , $\ell \geq T$. Выберем некоторую размерность $d \geq T$ пространства наблюдений и построим векторы $\mathbf{x}_1, \dots, \mathbf{x}_n$ как отрезки наблюдений со сдвигом, так что для каждого $\mathbf{x}_m \in \mathbb{R}^d$ имеет место $\mathbf{x}_m = (X_{1+m}, \dots, X_{d+m})$, $m = 1, \dots, n$. Предполагая существование в \mathbb{R}^d системы векторов $\{\mathbf{e}_1, \dots, \mathbf{e}_d\}$ такой, что каждый $\mathbf{x}_m = \sum_{k=1}^d \alpha_k \mathbf{e}_k$, найдем векторы этой системы как собственные векторы матрицы Σ , строками которой являются наблюдения $\mathbf{x}_1, \dots, \mathbf{x}_n$. Нахождение векторов $\{\mathbf{e}_1, \dots, \mathbf{e}_d\}$ легко осуществить с помощью сингулярного разложения матрицы Σ .

Пространство наблюдений затем делят на d_1 -мерное главное пространство $\mathcal{L}(\mathbf{e}_1, \dots, \mathbf{e}_{d_1})$, описывающее «главную часть» информации выборки \mathbf{X}^ℓ , и остаточное подпространство $\mathcal{L}(\mathbf{e}_{d_1+1}, \dots, \mathbf{e}_d)$, описывающее «шумовую часть» информации в этой выборке. Наблюдения, имеющие высокую энергию (евклидову норму) в остаточном подпространстве, классифицируются затем как аномальные.

Среди других подходов к моделированию сигналов с циклами, отметим нашедшие применение в различных приложениях модели на основе процесса авторегрессии [59; 76], разложения временного ряда на компоненты [29; 78], и параметрической регрессии [4; 13]. Отметим, что применительно к задаче моделирования сигналов больших систем с интенсивным ПО такие подходы неприменимы или лишь ограниченно применимы в силу следующих причин:

- невозможности изменения жестко заданной длины периода (как правило, равной 12 месяцам в задачах моделирования экономических временных рядов);

- невозможности обработки пропусков в сигнале (что является типичной ситуацией при большом количестве сигналов и высокой частоте регистрации каждого сигнала);
- предположения о регулярном поступлении данных;
- предположения о фиксированной модели наблюдений (т. е. невозможности детектирования разладки);
- высокого числа параметров (в случае регрессионных моделей) и сложности модели.

Сверх упомянутого, ни один из описанных выше подходов не включает явной модели длинной памяти; следовательно, используя эти подходы, нельзя добиться оптимального оценивания тренда. С другой стороны, первый подход, разработанный в диссертационной работе, основан на теоретически оптимальном фильтре, включающем явную модель длинной памяти (фрактального броуновского движения), исследованную в главе 1. Во-вторых, второй предлагаемый подход основан на методе непараметрической регрессии, позволяющем получить явное разложение временного ряда на тренд, сезонную и шумовую компоненты. Этот подход основан на хорошо исследованной ядерной оценке Надарая-Ватсона (см. [58; 90]), причем в качестве используется ядро бесконечного порядка, что улучшает свойства сходимости оценки для гладких функций [52]. Оценка на основе непараметрической регрессии широко используется в приложениях (см., например, [84]) и хорошо зарекомендовала себя в задачах, рассмотренных в диссертационной работе, в частности для случая большой длины периода и гладких целевых функций.

3.2.2 Постановка задачи оценки гладкого тренда

Предполагается, что наблюдения $\xi = (\xi_t)_{t \geq 0}$ выполнены согласно общей модели

$$\xi_t = f(t) + \nu_t, \quad t \geq 0, \quad (3.2)$$

где $f(t)$ — неслучайная гладкая функция (тренд), наблюдаемая в шуме ν_t , $E \eta(t) = 0$. По данным зашумленных измерений $\mathbf{X}^\ell = \{(X_k, t_k)\}_{k=1}^\ell$, $X_k = \xi_{t_k}$, $k = 1, \dots, \ell$, выполненным согласно (3.2), требуется оценить значение $f(t) = E \xi_t$ для каждого $t \geq 0$. Для разработки алгоритмов решения этой задачи в следующих двух разделах приняты более детальные предположения о тренде и шуме.

3.3 Алгоритм оценивания параметров сигнала на основе фильтра для наблюдений с длинной памятью

3.3.1 Описание алгоритма

Рассмотрим алгоритм оценивания параметров тренда на основе фильтра, разработанного в главе 1. Этот алгоритм предполагает, что наблюдения $\xi = (\xi_t)_{t \geq 0}$ выполнены согласно модели (ср. (3.2)):

$$\xi_t = f(t) + \nu_t^H, \quad t \geq 0, \quad (3.3)$$

где локально гладкий тренд $f(t)$ наблюдается на фоне шума ν_t^H , обладающего свойством длинной памяти. Помеха ν_t^H моделируется с помощью фрактального гауссовского шума $Z^H = (Z_t^H)_{t \geq 0}$ (fractional Gaussian noise, fGn) с некоторой (неизвестной) дисперсией: $\nu_t^H = \sigma Z_t^H$.

Теоретические результаты раздела 1.3 могут применяться для оценивания тренда $f(t)$ при выполнении следующих условий:

- функция $f(t)$ является достаточно гладкой, что позволяет рассмотреть ее аппроксимацию многочленом конечной степени $\sum_{i=0}^{n_\theta} \theta_i (t - t_0)^i$ в окрестности любого $t_0 > 0$;
- значение параметра Херста H известно (на практике для его оценивания можно использовать подходы, предложенные в работах [15; 17; 31; 35]).

Кроме того, привлекательным свойством указанного подхода является тот факт, что разработанная в нем оценка не зависит от значения дисперсии σ .

Рассмотрим следующий алгоритм оценивания тренда $f(t)$.

1. Рассмотрим шагающее окно, позиция W которого принимает последовательно значения W_1, W_2, \dots , где $W_i = [(i-1)s, (i-1)s + h]$, $i = 1, 2, \dots$, $h > 0$ — ширина окна, $s < h$ — шаг, на который сдвигается положение окна.
2. Вычислим оценку $\hat{f}_{W_i}(t)$ тренда $f(t)$ для $t \in W_i$:
 - (а) Предполагая, что в окрестности точки $t_0 = (i-1)s + h/2$ модель тренда задается кубическим полиномом

$$X_k = \sum_{j=0}^3 \theta_j (t_k - t_0)^j + \sigma Z_k^H \quad (3.4)$$

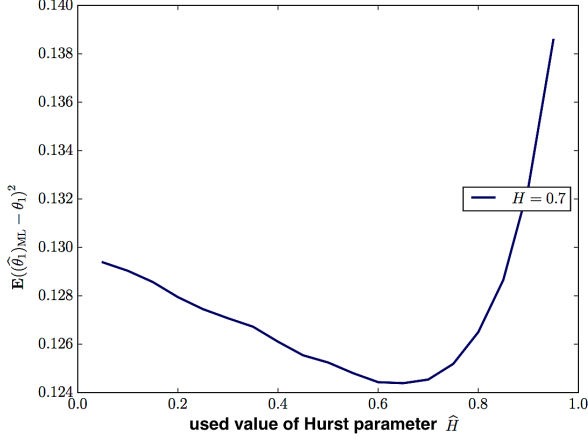


Рисунок 3.1: Зависимость среднеквадратичной погрешности оценивания коэффициента линейного сноса фрактального броуновского движения от значения \hat{H} параметра Херста, предполагаемого в оценке 1.3

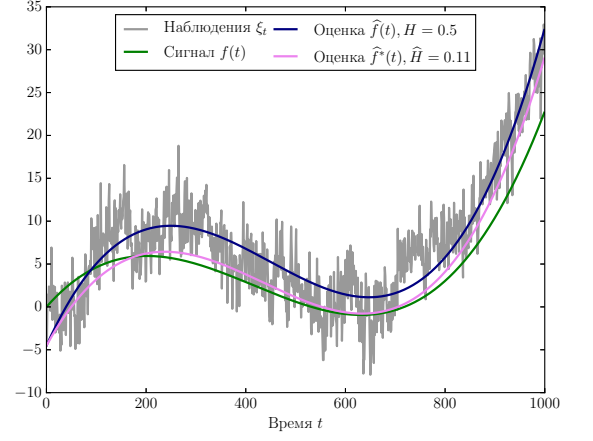


Рисунок 3.2: Результат применения алгоритма оценивания параметра тренда с поправкой на длинную память. Серым цветом показана траектория результата наблюдений ξ в (3.4), синим — истинная траектория тренда $f(t)$, зеленым — результат оценивания $\hat{f}(t)$ без поправки на длинную память ($H = 1/2$), розовым — результат оценивания $\hat{f}(t)$ с поправки на длинную память ($H = 0.11$). Истинное значение $H = 0.1$.

где $t_k \in W_i$, дисперсия σ постоянна, и принято $H = 1/2$, вычислим оценку максимального правдоподобия $\hat{\theta}_{\text{ML}}$ параметра $\theta = (\theta_0, \dots, \theta_3)$, используя соотношения из раздела 1.3.

- (b) Вычислим оценку тренда для каждого $t \in W_i$, используя соотношение $\hat{f}_{W_i}(t) = \sum_{i=0}^3 (\hat{\theta}_{\text{ML}})_i (t - t_0)^i$.
 - (c) Вычислим оценку дисперсии помехи $\hat{\sigma}$ как выборочную дисперсию величин $Y_k = X_k - \hat{f}_{W_i}(t_k)$, $t_k \in W_i$.
 - (d) Вычислим оценку показателя Херста \hat{H} , используя подход из работы [15] и стандартизованные остатки $R_k = Y_k / \hat{\sigma}$, $t_k \in W_i$.
 - (e) Используя полученную оценку показателя Херста \hat{H} , вычислим скорректированные оценки тренда и дисперсии в (a)–(c).
3. Вычислим оценку $\hat{f}(t)$ тренда $f(t)$, усредняя локальные оценки $\hat{f}_{W_1}(t), \hat{f}_{W_2}(t), \dots$, полученные для каждого положения W_1, W_2, \dots шагающего окна:

$$\hat{f}(t) = \frac{1}{n(t)} \sum_{i: t \in W_i} \hat{f}_{W_i}(t),$$

где $n(t)$ — число окон, пересекающихся в точке t (при равномерном поступлении наблюдений $n(t) = n = h/s + 1$).

Двухшаговая процедура вычисления оценки $\hat{f}(t)$ необходима на практике ввиду того, что показатель Херста H является неизвестной, но важной постоянной, влияющей на точность оценивания. Эта зависимость проиллюстрирована на рис. 3.1, где показано, что погрешность в задании постоянной Херста влияет на точность оценивания параметра тренда. Применяя поправку согласно схеме, описанной в шагах 2а– 2с алгоритма, возможно уменьшить неопределенность в значении H и достичь лучшей точности оценивания тренда по сравнению с подходом «по умолчанию», в котором $H = \frac{1}{2}$, см. рис. 3.2.

Кратко опишем алгоритм прогнозирования будущих значений временного ряда. Пусть $\mathbf{X}^\ell = \{(X_k, t_k)\}_{k=1}^\ell$ — данные, использовавшиеся для вычисления оценки $\hat{f}(t)$ тренда. Задача прогнозирования будущего значения временного ряда на один шаг вперед заключается в вычислении оценки $\hat{f}(t_{\ell+1})$ ожидаемого значения $E X_{t_{\ell+1}}$ измерения $X_{t_{\ell+1}}, t_{\ell+1} > t_\ell$. Вычислим $h/(2s)$ оценок $\hat{f}_{W_{-h}^F}(t), \dots, \hat{f}_{W_{-h/(2s)}^F}(t)$, используя описанный выше алгоритм (1)–(3) и позиции $W_{-h}^F = [t_{\ell-h}, t_\ell], \dots, W_{-h/(2s)}^F = [t_{\ell-h/(2s)}, t_\ell]$ шагающего окна. Оценка $\hat{f}(t_{\ell+1})$ значения $E X_{t_{\ell+1}} = f(t_{\ell+1})$, вычисляется согласно соотношению

$$\hat{f}(t_{\ell+1}) = \frac{2s}{h} \sum_{k=-h}^{-h/2} \hat{f}_{W_k^F}(t_{\ell+1}).$$

3.4 Алгоритм оценивания параметров сигнала на основе непараметрической регрессии

3.4.1 Модели наблюдений с явным учетом сезонности

Рассмотрим многокомпонентную математическую модель квазипериодического сигнала и алгоритм оценивания его параметров на основе непараметрической регрессии. Пусть равенство

$$\xi_t = f(t) + \nu_t, \quad t \geq 0, \quad (3.5)$$

задает модель сигнала $\xi = (\xi_t)_{t \geq 0}$, в которой моделью тренда служит соотношение $f(t) = Q_t S_t$, где процессы $Q = (Q_t)_{t \geq 0}$, $S = (S_t)_{t \geq 0}$ и $\nu = (\nu_t)_{t \geq 0}$ суть ненаблюдаемые амплитуда и сезонная составляющая; модель случайной помехи ν_t задается равенством $\nu_t = \sigma_t \varepsilon_t$, где $\sigma_t = \sigma_{\varphi(t)}$ — неслучайная функция, а $\varepsilon = (\varepsilon_t)_{t \geq 0}$ — процесс стандартного гауссовского белого шума. В модели (3.5) Q — случайный процесс, $S_t = S(\varphi(t))$ — детерминированная (неслучайная) функция. Величина $\varphi(t) = 2\pi\{t/T\}$ имеет смысл фазы (известного) периода T , соответствующего моменту времени t (где $\{x\} = x - [x]$ — дробная часть x). Назовем такое представление временного ряда моделью $[Q, S, \nu]$.

Факторизация представлений тренда и помехи в (3.5) позволяет выразить важные свойства реальных сигналов систем с интенсивным ПО, такие как медленный рост числа обрабатываемых запросов и его флуктуации в течение суток. В модели тренда $Q = (Q_t)_{t \geq 0}$ и $S = (S_t)_{t \geq 0}$ интерпретируются как ненаблюдаемые амплитуда и сезонная составляющая, соответственно.

Рассматривая различные модели амплитуды Q , можно получать и модели временного ряда ξ с различными свойствами. При $Q_t \equiv 1, t \geq 0$, соотношение

$$\xi_t = S_t + \nu_t \quad (3.6)$$

задает строго периодическую модель $[S, \nu]$ наблюдений, где S и ν имеют тот же смысл, что и выше. С другой стороны, при $Q_t = L_t A_t$, где $L = (L_t)_{t \geq 0}$ и $A = (A_t)_{t \geq 0}$ — случайные процессы, моделирующие ненаблюдаемые долгосрочный тренд и локальную амплитуду, соответственно, соотношение

$$\xi_t = L_t A_t S_t + \nu_t \quad (3.7)$$

задает модель $[L, A, S, \nu]$, учитывающую тренд на двух масштабах времени. А именно, тренд L предполагается случайным процессом, характерное время изменения которого порядка $\Delta t_L \gg T$, в то время как характерное время изменения процесса A порядка $\Delta t_A \sim T$. Моделирование тренда и амплитуды случайными процессами означает интерпретацию цикла $L_t A_t S_t$ в (3.7) как стохастического, а не детерминированного и позволяет называть модель в (3.7) *квазипериодической*. Рассматривая модели $[S, \nu]$, $[Q, S, \nu]$ или $[L, A, S, \nu]$, можно выбрать удобную детализацию модели при моделировании наблюдений.

3.4.2 Итеративные алгоритмы оценивания параметров моделей квазипериодических сигналов

Рассмотрим алгоритм оценивания значений \hat{S}_{ψ_j} и $\hat{\sigma}_{\psi_j}^2$ для каждой фазы $\psi_j = \varphi(t_j)$, где $t_j = j\Delta$, $\Delta = T/p$, $j = 1, \dots, p$, по данным зашумленных измерений $\mathbf{X}^\ell = \{(X_k, t_k)\}_{k=1}^\ell$.

Инициализация. Положим величину \hat{Q}_k равной 1 для каждого $k = 1, \dots, \ell$, а величину $\hat{\sigma}_{\psi_j}^2$ для каждого $j = 1, \dots, p$ положим равной дисперсии наблюдений X_1, \dots, X_ℓ .

Итерации. Повторим следующие шаги:

1. С использованием оценки Надарая-Ватсона переоценим \hat{S}_{ψ_j} :

$$\hat{S}_{\psi_j} = \frac{\sum_{k=1}^{\ell} w_k X_k / \hat{Q}_k K_h(\varphi_k, \psi_j)}{\sum_{k=1}^{\ell} w_k K_h(\varphi_k, \psi_j)} \quad (3.8)$$

где $\varphi_k = \varphi(t_k)$ суть фазы в моменты времени t_k , $k = 1, \dots, \ell$, $K_h(\varphi, \psi)$ — ядро ширины $h > 0$, и w_k — вес k -го измерения.

2. С использованием оценки Надарая-Ватсона переоценим $\hat{\sigma}_{\psi_j}^2$:

$$\hat{\sigma}_{\psi_j}^2 = \frac{\sum_{k=1}^{\ell} (X_k - \hat{X}_k)^2 K_h(\varphi_k, \psi_j)}{\sum_{k=1}^{\ell} K_h(\varphi_k, \psi_j)}. \quad (3.9)$$

3. Переоценим $\hat{X}_k = \hat{X}(t_k)$ и $\hat{Q}_k = \hat{Q}(t_k)$, $k = 1, \dots, \ell$. Для вычисления прогноза \hat{X}_k значения X_k выберем некоторое $H > 0$ и рассмотрим моменты времени t_{k-p}, \dots, t_k , где $t_k - H \leq t_{k-p} < \dots < t_k$. Согласно (3.5) в предположении локально постоянной амплитуды

$$X_i = Q_k \hat{S}_{\varphi(t_i)} + \nu_i, \quad i = k-p, \dots, k, \quad (3.10)$$

где значение $\hat{S}_{\varphi(t_i)}$ получим кубической интерполяцией значений \hat{S}_{ψ_j} по четырем ближайшим к $\varphi(t_i)$ точкам сетки ψ_1, \dots, ψ_n . В предположении $\nu_i \sim \mathcal{N}(0, \hat{\sigma}_{\psi_i}^2)$ амплитуду Q_k в (3.10) оценим методом взвешенной линейной

Алгоритм 1. Оценивание параметров квазипериодической модели 3.5

- 1: **Вход:** данные $\mathbf{X}^\ell = \{(X_k, t_k)\}_{k=1}^\ell$.
 - 2: Положить $\hat{Q}_k = 1, k = 1, \dots, \ell$.
 - 3: Положить $\hat{\sigma}_{\psi_j}^2 = \hat{\sigma}^2 = \text{var}(X_1, \dots, X_\ell), j = 1, \dots, p$.
 - 4: **Повторять**
 - 5: Положить $\text{Diff}_S = 1/p \sum_{j=1}^p (\hat{S}_{\psi_j} - \hat{S}_{\psi_j}^{\text{PREV}})^2$.
 - 6: Положить $\text{Diff}_{\sigma^2} = 1/p \sum_{j=1}^p (\hat{\sigma}_{\psi_j}^2 - (\hat{\sigma}_{\psi_j}^2)^{\text{PREV}})^2$.
 - 7: **Для всех** $j = 1, \dots, p$ **выполнять**
 - 8: $\hat{S}_{\psi_j}^{\text{PREV}} \leftarrow \hat{S}_{\psi_j}, (\hat{\sigma}_{\psi_j}^2)^{\text{PREV}} \leftarrow \hat{\sigma}_{\psi_j}^2$.
 - 9: Обновить \hat{S}_{ψ_j} согласно (3.8).
 - 10: Обновить $\hat{\sigma}_{\psi_j}^2$ согласно (3.9).
 - 11: **Конец цикла**
 - 12: **Для всех** $k = 1, \dots, \ell$ **выполнять**
 - 13: Вычислить \hat{Q}_k и \hat{X}_k согласно (3.10)–(3.11).
 - 14: **Конец цикла**
 - 15: **Пока выполняется** $\text{Diff}_S < 10^{-6}$ **и** $\text{Diff}_{\sigma^2} < 10^{-6}$
-

регрессии с весами $\lambda_i = 1/\hat{\sigma}_{\psi_i}^2, i = k - p, \dots, k$. Прогноз \hat{X}_k значения X_k вычислим согласно равенству

$$\hat{X}_k = \hat{Q}_k \hat{S}_{\varphi(t_k)}. \quad (3.11)$$

Итерации останавливаются, когда среднеквадратичное изменение оценок $\hat{\sigma}_{\psi_j}^2$ и \hat{S}_{ψ_j} становится меньше некоторого порогового значения (на практике, после 3–5 итераций алгоритма отличие в этих оценках становится меньше величины 10^{-6}).

Алгоритмы оценивания для моделей $[S, \nu]$ and $[L, A, S, \nu]$. Процедура оценивания параметров модели $[L, A, S, \nu]$ следует в точности тем же шагам 1–3, как и в случае модели $[Q, S, \nu]$ с замечой Q_t произведением $L_t A_t$. В этом случае прямая оценка локальной амплитуды становится недоступна.

Для модели $[S, \nu]$ не требуется оценивание амплитуды Q , и вся процедура обучения становится проще: шаг 3, относящийся к регрессии (3.10), опустим, и вычислим только \hat{S}_{ψ_j} и $\hat{\sigma}_{\psi_j}^2$ (использая кубическую интерполяцию, описанную выше) в (3.8)–(3.9) для $j = 1, \dots, p$. Оценка X_n вычисляется согласно соотношению $\hat{X}_n = \hat{S}_{\varphi_n}$.

3.4.3 Практическая реализация алгоритмов оценивания на основе непараметрической регрессии

Алгоритм 1 включает ряд параметров, выбор которых при программной имплементации неясен. Обсудим некоторые подробности выбора параметров, принятого в диссертационной работе.

Разность фаз. В диссертационной работе ядерная функция $K_h(\varphi, \psi)$ задана зависящей от разности фаз посредством величины

$$\delta(\varphi, \psi) = \min(|\varphi - \psi|, |\varphi - \psi - 2\pi|, |\varphi - \psi + 2\pi|),$$

которая имеет смысл кратчайшего расстояния между фазами φ и ψ с учетом цикла.

Сглаживание и веса. Веса w_k , назначаемые измерениям X_k , $k = 1, \dots, \ell$, в (3.8), задаются соотношениями:

$$w_k = (1 - \alpha)^{-(t-t_k)/\Delta}, \quad (3.12)$$

где $\alpha \in (0, 1)$ контролирует скорость затухания веса, назначаемого наблюдению, а $\Delta > 0$ является некоторой постоянной, имеющей размерность времени, так что степень $(t - t_k)/\Delta$ безразмерна (в диссертации для простоты принято, что эта величина равна периоду наблюдений: $\Delta = T$). Такой выбор весов в (3.8)–(3.9) позволяет объединить усреднение внутри периода (с помощью ядерного сглаживания) и усреднение по периодам (посредством экспоненциального сглаживания с весами w_k). Сглаживание внутри периода (ядерное) позволяет использовать временную локальность и вычислять оценку усреднением соседних точек (какие именно точки являются соседями, определяется шириной ядра h), а сглаживание по периодам (экспоненциальное) — фазовую локальность, вычисляя оценку усреднением точек с близкими фазами в предшествующих периодах наблюдений. Заметим, что при $h \rightarrow 0$ ядерная функция сколь угодно близко приближается к дельта-функции Дирака, а формулы (3.8)–(3.9) сводятся к векторному случаю экспоненциально взвешенного скользящего среднего по точкам с близкими фазами.

Выбор ядерной функции. В диссертационной работе рассматривались несколько типов ядерных функций: лапласовское, гауссовское и косинусное ядра, зада-

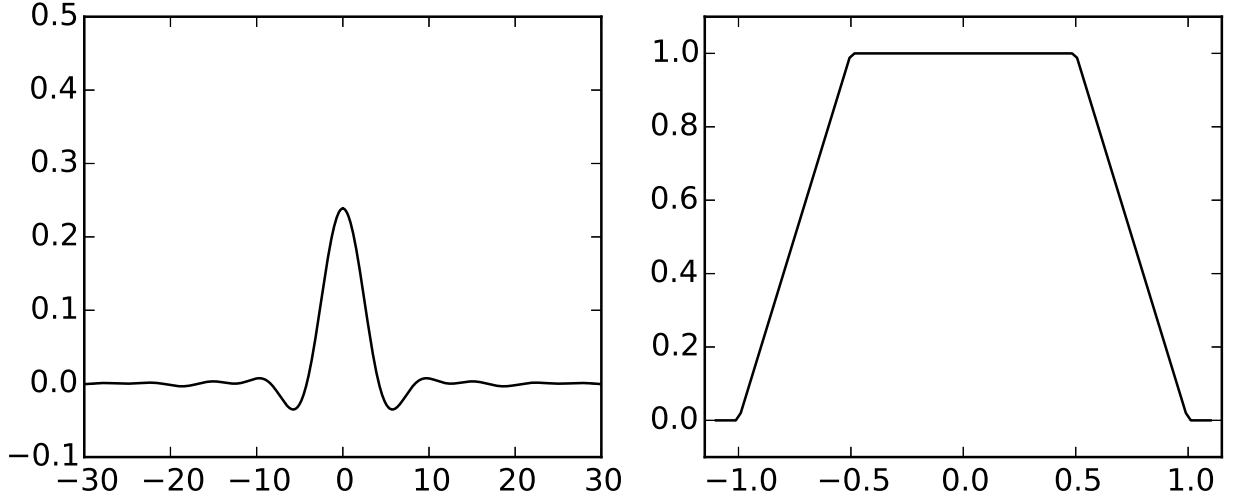


Рисунок 3.3: Косинусное ядро и его Фурье-образ

ваемые соотношениями

$$\begin{aligned}
 K_{\text{GAUSS}}(\delta) &= \exp \left\{ -\delta^2 / (2h^2) \right\}, \\
 K_{\text{LAPLACE}}(\delta) &= \exp \left\{ -|\delta|/h \right\}, \quad \text{и} \\
 K_{\text{COS}}(\delta) &= \frac{2}{\pi} \frac{\cos(\delta/(2h)) - \cos(\delta/h)}{\delta^2/h^2},
 \end{aligned} \tag{3.13}$$

соответственно. Лапласовское и гауссовское ядра наиболее широко применяются в непараметрической регрессии; эти функции являются ядрами второго порядка, что ограничивает скорость сходимости оценки величиной $O(h^2)$, где h — ширина ядра. Косинусное ядро, показанное на рис. 3.3, имеет бесконечный порядок, что означает, что скорость сходимости непараметрической оценки ограничена только степенью дифференцируемости оцениваемой функции [52].

Эффективная оценка дисперсии. Для малых выборок дисперсия в (3.9) часто оказывается переоцененной. Чтобы улучшить свойства оценки, в диссертационной работе используется эффективная оценка дисперсии согласно соотношению, справедливому для малых выборок

$$(\hat{\sigma}_{\psi_j})_{\text{eff}} = \frac{1}{\sqrt{n}} \hat{\sigma}_{\psi_j} t_{n-1, 1-\alpha/2}, \tag{3.14}$$

где $t_{n-1,1-\alpha/2}$ является α -квантилем распределения Стьюдента, а величина n имеет смысл эффективного объема выборки: $n = \sum_{k=1}^{\ell} w_k$ [43]. Такая поправка улучшает точность оценки дисперсии при малом объеме данных, например, в случае появления нового потока наблюдений $\xi = (\xi_t)_{t \geq 0}$ или при появлении пропуска в наблюдениях.

3.5 Обнаружение моментов изменения свойств сигналов с квазипериодическим трендом

Рассмотрим задачу обнаружения момента изменения вероятностно-статистических характеристик квазипериодического сигнала, модель измерения которого задана в (3.2). Методология, развиваемая в диссертационной работе, предполагает осуществление фильтрации тренда в непрерывном времени с генерированием ряда «текущих» характеристик наблюдаемого сигнала — бегущей оценки $\hat{f}(t)$ тренда $f(t)$, прогноза $\hat{\xi}_t$ значения ξ_t , оценки $\hat{\sigma}_t^2$ дисперсии σ_t^2 , а в случае алгоритмов фильтрации на основе явной модели сезонности (3.5) — и оценку \hat{Q}_t амплитуды Q_t наблюдений. Значительное увеличение ошибки прогнозирования (а значит, и погрешности оценивания тренда), сопутствующей фильтрации, указывает на возникновение структурных изменений модели (3.2). Обнаружение этих изменений возможно с помощью анализа траекторий упомянутых компонент сигнала, генерируемых фильтрацией. В настоящем разделе рассмотрим две важные для практики задачи обнаружения разладки для квазипериодических моделей.

3.5.1 Модели разладки квазипериодических сигналов и их адекватность задачам обнаружения разладки

Рассмотрим первый тип изменений, интересный в практике — это краткосрочные изменения с характерной длительностью много меньше периода наблюдений: $\Delta_S \ll T$. Они представляют собой локальное отклонение в значениях наблюдаемого сигнала. Пример такого отклонения показан на верхней половине рис. 3.4. Заметим, что квазипериодический тренд $f(t)$ при разладке этого типа сохраняется неизменным; учитывая это, рассмотрим модель краткосрочной разладки

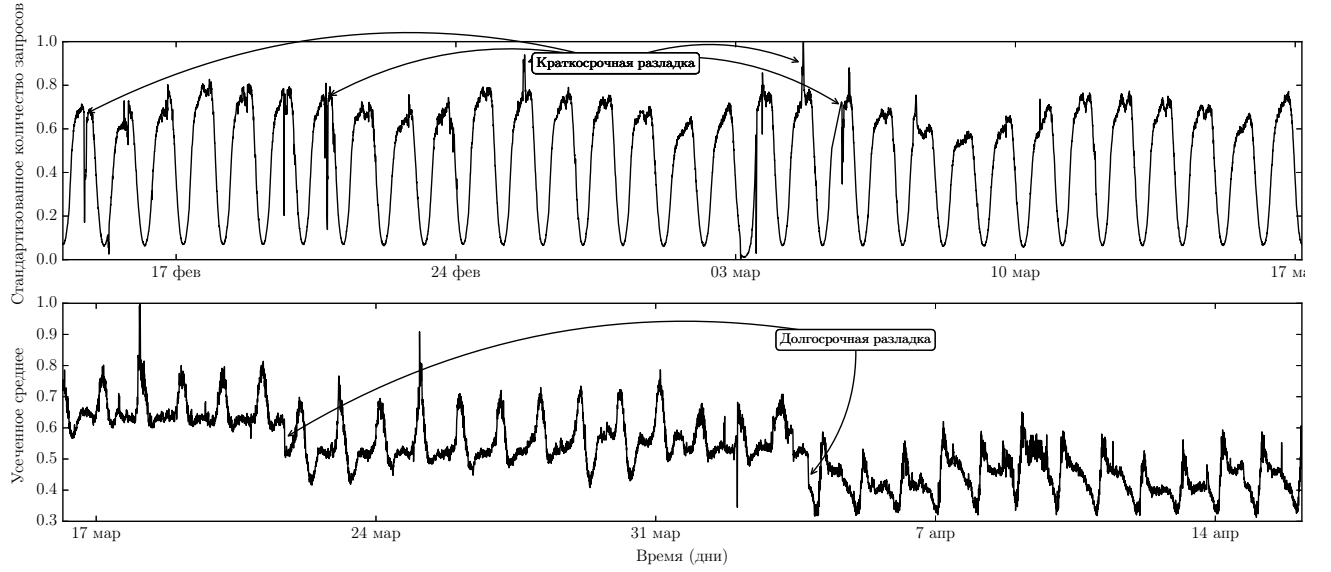


Рисунок 3.4: *Верхний график:* месячная траектория числа запросов, заданных некоторому интернет-сервису Яндекса, с отмеченными несколькими краткосрочными разладками. *Нижний график:* месячная траектория усредненного времени ответа, этого интернет-сервиса, с отмеченными двумя долгосрочными разладками. Показаны нормализованные на единицу значения.

в (3.2):

$$\nu_t = \mu 1_{[\theta, \theta + \Delta_S]}(t) + Z_t, \quad t \geq 0, \quad (3.15)$$

где θ — неизвестный момент разладки, μ — неизвестная величина разладки, Δ_S — неизвестная длительность разладки, и $Z = (Z_t)_{t \geq 0}$ — гауссовский процесс белого шума. Согласно (3.15) разладка заключается в кратковременном скачкообразном изменении среднего значения помехи ν_t с нуля на μ , причем естественно предполагать, что «магнитуда» изменения μ является неизвестным параметром (вопрос определения ее значения при практической реализации обсуждается ниже в п. 3.5.2). Отметим также, что, хотя описанное изменение и имеет «короткую» продолжительность Δ_S , она, тем не менее, может оказаться ощутимой с точки зрения пользователя (так, например, при получасовом интервале измерения $T = 48$, а продолжительность разладки $\Delta_S = 6$ выливается в 3 часа недоступности сервиса).

Второй тип разладки, рассматриваемый в диссертационной работе, связан с долгосрочными изменениями в модели $[Q, S, \nu]$. Рассмотрим модель долгосроч-

ной разладки, в которой квазипериодическая компонента Q

$$Q_t = \begin{cases} 1 + Z_t, & \text{если } t \notin [\theta, \theta + \Delta_L], \\ \mu + Z_t, & \text{если } t \in [\theta, \theta + \Delta_L], \end{cases} \quad (3.16)$$

где θ — неизвестный момент разладки, μ — неизвестная величина разладки, Δ_L — неизвестная длительность разладки, и $Z = (Z_t)_{t \geq 0}$ — стандартный гауссовский процесс белого шума. Согласно (3.16) разладка заключается в изменении масштаба наблюдаемого процесса ξ и имеет характерную длительность порядка $\Delta_L \gg T$, как показано на нижней части рис. 3.4. Изменения такого типа соответствуют, как правило, структурным изменениям в рассматриваемой системе (обновлению программного обеспечения, замене аппаратного обеспечения, инициации нового режима работы и т. д.).

Обсудим вопрос об адекватности рассматриваемых моделей разладки задачам их обнаружения. Очевидно, что эти модели являются лишь приближением некоторой неизвестной истинной модели разладки в силу ошибок различного типа, таких как

- ошибки аппроксимации, вызванной заменой в (3.4) истинного тренда его кубическим приближением;
- ошибки оценивания параметров истинного тренда алгоритмами разделов 3.3–3.4 по результатам зашумленных измерений;
- ошибки моделирования, связанной с интерпретацией изменения в (3.15)–(3.16) как скачкообразного изменения среднего значения того или иного случайного процесса;
- ошибки моделирования, связанной с интерпретацией помехи в (3.2) как гауссовского или фрактального гауссовского шума.

Таким образом, стандартные предположения о модели разладки в рассматриваемой методологии оказываются нарушены. Как обсуждалось в главе 2, раздел 2.3.1, нарушение стандартных предположений о модели разладки, ведет, естественно, к снижению эффективности ее обнаружения. Однако, как показано там же, раздел 2.5, в этих условиях эффективность обнаружения разладки может быть повышена использованием ансамблей «слабых» детекторов. Как будет показано в п. 3.6, модели (3.15)–(3.16) обладают достаточной для обнару-

жения разладок описательной силой в случае, если используются алгоритмы обнаружения на основе ансамблей «слабых» детекторов.

3.5.2 Процедуры обнаружения разладки характеристик квазипериодических сигналов

Алгоритмы детектирования разладки, рассматриваемые в настоящем разделе, основаны на преобразованиях «текущих» характеристик фильтруемого по наблюдениям $\xi = (\xi_t)_{t \geq 0}$ сигнала $f(t)$ и применении ансамблей «слабых» детекторов.

Обнаружение краткосрочных изменений. Рассмотрим случайный процесс $R = (R_t)_{t \geq 0}$, задаваемый соотношением

$$R_t = \frac{\xi_t - \hat{f}(t)}{\hat{\sigma}_t}, \quad t \geq 0, \quad (3.17)$$

и имеющий смысл нормализованной ошибки прогнозирования наблюдаемой в (3.2) случайной величины ξ_t значением $\hat{\xi}_t$. Так как согласно алгоритмам разделов 3.3–3.4 математическое ожидание прогноза $E \hat{\xi}_t = E \xi_t = f(t)$, то при отсутствии разладки разность $\xi_t - \hat{\xi}_t$ является оценкой значения ν_t , а значение R_t является оценкой значения Z_t в (3.15). При появлении разладки равенство $E \hat{\xi}_t = E \xi_t$ нарушается, поскольку $E \hat{\xi}_t = f(t) \neq f(t) + \mu \sigma_t^2 = E \xi_t$. Таким образом, для математического ожидания $E R_t$ имеем¹ $E R_t \approx \mu 1_{[\theta, \theta + \Delta_S]}(t)$. Для обнаружения разладки процесса R используем ансамбль «слабых» детекторов, параметры которого подберем по множеству размеченных траекторий процесса ξ .

Обнаружение долгосрочных изменений. Рассмотрим два вспомогательных процесса $Z = (Z_t)_{t \geq 0}$ и $Z^S = (Z_t^S)_{t \geq 0}$, задаваемых соотношениями

$$\begin{aligned} Z_t &= \ln \hat{Q}_t, \\ Z_t^S &= (1 - \lambda) Z_{t-1}^S + \lambda Z_t, \quad Z_0^S = Z_0, \quad t \geq 0, \end{aligned} \quad (3.18)$$

¹Знак “ \approx ” обусловлен приближенным характером прогноза $\hat{\xi}_t$, см. раздел 3.5.1.

соответственно. Согласно (3.16)–(3.18) процесс Z имеет смысл логарифма² локального масштаба наблюдений A , причем приближенно $E Z_t \approx \ln \mu 1_{[\theta, \theta + \Delta_L]}(t)$. В случае, когда медленно меняющийся тренд L отсутствует (используется модель $[Q, S, \nu]$), достаточно использовать процесс Z для обнаружения разладки процесса локальной амплитуды $Q \equiv A$. В случае же, когда в модели учитывается и медленно меняющийся тренд (рассматривается модель $[L, A, S, \nu]$, $Q_t = L_t A_t$), рассмотрим сглаженную версию локальной амплитуды Z^S в (3.18), в которой постоянная экспоненциального сглаживания λ выбрана таким образом, чтобы выполнялось соотношение $1/\lambda \gg T$. Тогда математическое ожидание $E[Z_t - Z_t^S] \approx \ln \mu 1_{[\theta, \theta + \Delta_L]}(t)$. Для обнаружения разладки процесса разности $Z_t - Z_t^S$ используем ансамбль «слабых» детекторов, параметры которого подберем по множеству размеченных траекторий процесса ξ .

3.6 Эффективность обнаружения разладки квазипериодического временного ряда

3.6.1 Вычислительный эксперимент и наборы данных

Эффективность алгоритмов фильтрации и обнаружения разладки была исследована в ходе вычислительного эксперимента на двух искусственных наборах данных, обозначаемых в диссертационной работе ARTIFICIAL-EASY и ARTIFICIAL-HARD. Результаты применения разработанной методологии в задачах анализа реальных сигналов представлены в главе 5.

Искусственные наборы данных состоят из отрезков измерений $\{(X_k, t_k)\}_{k=1}^{\ell}$, $\ell = 2016$, интервал $t_{k+1} - t_k, k = 1, \dots, \ell - 1$, между которыми фиксирован 5 минутами, выполненных согласно модели

$$X_k = f(t_k) + \nu_{t_k}^H, \quad k = 1, \dots, \ell,$$

в которой тренд моделируется строго периодической функцией

$$f(t_k) = A \sin\left(\frac{2\pi t_k}{T}\right), \quad k = 1, \dots, \ell,$$

²Можно использовать и процесс локальной амплитуды A без преобразований. Здесь взят натуральный логарифм для единообразия статистических свойств сигнала при обнаружении разладки.

Таблица 3.1: Точность выделения тренда для искусственного набора данных в терминах относительной среднеквадратичной погрешности для процедуры EWMA, вариантов процедур на основе анализа главных компонент (PCA и PCA-PRETRAINED) и рассматриваемого в диссертационной работе подхода.

Постановка	EWMA	PCA	PCA-PRETRAINED	Наш подход
Аппроксимация тренда	7.84	8.96	5.58	5.72
Прогноз на одну точку вперед	7.34	5.65	3.80	3.06

с параметрами $A = 1.5, T = 288$, а процесс ν^H является процессом с длинной памятью. Для моделирования разладки в искусственных данных в каждой репликации измерений шум с длинной памятью ν^H формируется согласно модели (3.15), в которой дисперсия помехи $\sigma_t \equiv 1, t \geq 1$, случайный момент появления разладки $\theta \sim U(T, 6T)$ и случайная длительность разладки $\Delta_S \sim U(5, 100)$, а процесс $Z^H = (Z_t^H)_{t \geq 0}$ формируется как дискретная аппроксимация процесса фрактального гауссовского шума с параметром Херста $H = 0.95$. Для набора данных ARTIFICIAL-EASY магнитуда разладки в (3.15) принята равной $\mu = 5$, а для набора данных ARTIFICIAL-HARD, магнитуда разладки $\mu = 3$. Несмотря на такую, казалось бы, высокую величину магнитуды разладки, как показано ниже, сгенерированные разладки весьма трудно обнаруживать ввиду наличия сезонного тренда и шума с длинной памятью. Было сформировано 1000 независимых репликаций выборки для обучения ансамбля и еще 1000 независимых репликация для оценивания его качества.

3.6.2 Исследуемые процедуры

В вычислительных экспериментах в качестве «слабых» рассматриваются детекторы на основе статистики кумулятивных сумм (2.14), статистики Ширяева-Робертса (2.13), контрольных карт Шухарта (2.15), статистики changepoint (2.16), а также статистики апостериорной вероятности (2.12) (детали см. в разделе 2.2.2). На вход каждому детектору подадим процесс «остатков» R , заданный согласно (3.17). Используя траектории статистик этих детекторов, обучим ансамбль на основе логистической регрессии, агрегирующая функция которого задана согласно (2.21). Глубина истории, используемой ансамблем, варьировалась в небольших пределах, $p \in \{0, 1, 2, 3, 4\}$; наилучшие результаты согласно кривой «точность–полнота» показывает алгоритм с глубиной истории $p = 4$.

Было проведено эмпирическое сравнение качества обнаружения разладки, получаемого с помощью ансамблей и ряда других подходов: процедуры на основе порогового фильтра, процедуры кумулятивных сумм, и методов на основе анализа сингулярного спектра. Кратко опишем рассматриваемые подходы и их применение в рамках сравнительного анализа.

Процедура на основе порогового фильтра EWMA-THRESHOLD использует экспоненциально взвешенное скользящее среднее для оценивания текущего среднего значения $\hat{\mu}_t$ и дисперсии $\hat{\sigma}_t^2$ временного ряда ξ , вычисляет «остатки» $R_t = (\xi_t - \hat{\mu}_t)/\hat{\sigma}_t$, и подсчитывает долю точек скользящего окна $[t - \Delta, t]$, которые превышают заданное пороговое значение h . Момент подачи тревоги в этой процедуре определяется согласно равенству $\tau_{\text{THR}} = \inf\{k \geq 1 : S_k \geq h_{\text{THR}}\}$, где $S_k = \Delta^{-1} \sum_{i=k-\Delta}^k \mathbb{1}_{\{R_i \geq h\}}(i)$. Пороговая доля точек h_{THR} , поточечный порог h и длина скользящего окна Δ являются параметрами алгоритма; представленные результаты вычислительного эксперимента для процедуры на основе порогового фильтра получены при подобранных по равномерной сетке значениях параметров, максимизирующих качество этой процедуры на тестовой выборке.

Процедура кумулятивных сумм EWMA-CUSUM заменяет статистику $S = (S_t)_{t \geq 0}$, заданную выше, статистикой кумулятивных сумм $T = (T_t)_{t \geq 0}$, заданной в (2.14). Плотности $p_\infty(\cdot)$ и $p_0(\cdot)$ в этой процедуре предполагаются нормальными с единичными дисперсиями и математическими ожиданиями $\mu_\infty = 0$ и $\mu_0 = \mu$, соответственно; в последнем равенстве параметр μ подбирается для максимизации качества на тестовой выборке.

Процедура PCA тесно связана с подходом на основе анализа сингулярного спектра (SSA) и методами разделения на подпространства, представленными в литературе [47; 63; 88]. Процедуре PCA основана на получении разложения временного ряда $\xi = (\xi_t)_{t \geq 0}$ с помощью процедуры SSA и анализе компоненты ξ_t^{RES} , принадлежащей остаточному подпространству. Статистика $P = (P_t)_{t \geq 0}$ процедуры задается нормой компоненты ξ_t^{RES} : $P_t = \|\xi_t^{\text{RES}}\|, t \geq 0$. Заметим, что эффективность процедуры PCA определяется прежде всего объемом данных, используемых для получения разложения пространства наблюдений в нормальное и аномальное подпространства. Для использования этой ее особенности подадим на вход процедуре дополнительную неделю исторических данных для получения более точного разложения. Полученную таким увеличением данных процедуру

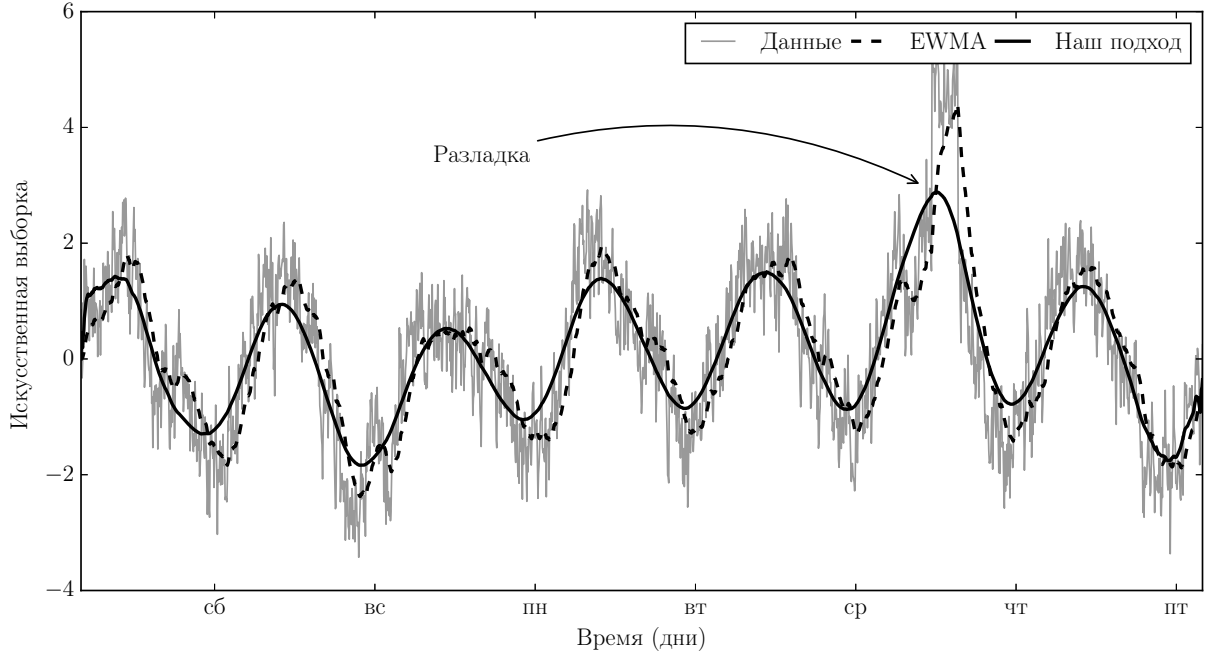


Рисунок 3.5: Пример данных из набора ARTIFICIAL-EASY и результат выделения тренда, полученный методом EWMA и предложенным в диссертационной работе подходом. Отмечены сгенерированные разладки.

назовем PCA-PRETRAINED. Заметим, что никакая другая процедура, включая ансамбль, не использовала никаких дополнительных данных для дообучения.

3.6.3 Точность аппроксимации тренда

Проведем сравнение точности оценивания тренда по выборке ARTIFICIAL-EASY для EWMA, PCA, PCA-PRETRAINED и предлагаемого подхода. Точность оценивания тренда характеризуется относительной среднеквадратичной погрешностью оценивания (relative root mean squared forecast error, RRMSE), задаваемой соотношением

$$\text{RRMSE}(\xi_t, \hat{\xi}_t) = \sqrt{\frac{1}{\ell} \sum_{t=1}^{\ell} \frac{(\xi_t - \hat{\xi}_t)^2}{\xi_t^2}}.$$

Табл. 3.1 представляет точность оценивания тренда в двух несколько отличающихся постановках: аппроксимации тренда и прогнозе на одну точку вперед. Точность оценивания тренда $\text{RRMSE}(f(t), \hat{f}(t))$ характеризует близость оценки $\hat{f}(t)$ и истинного тренда $f(t)$. Точность прогноза на одну точку вперед характеризует качество прогнозирования новых данных $\xi_{t_{\ell+1}}$ при фиксированной

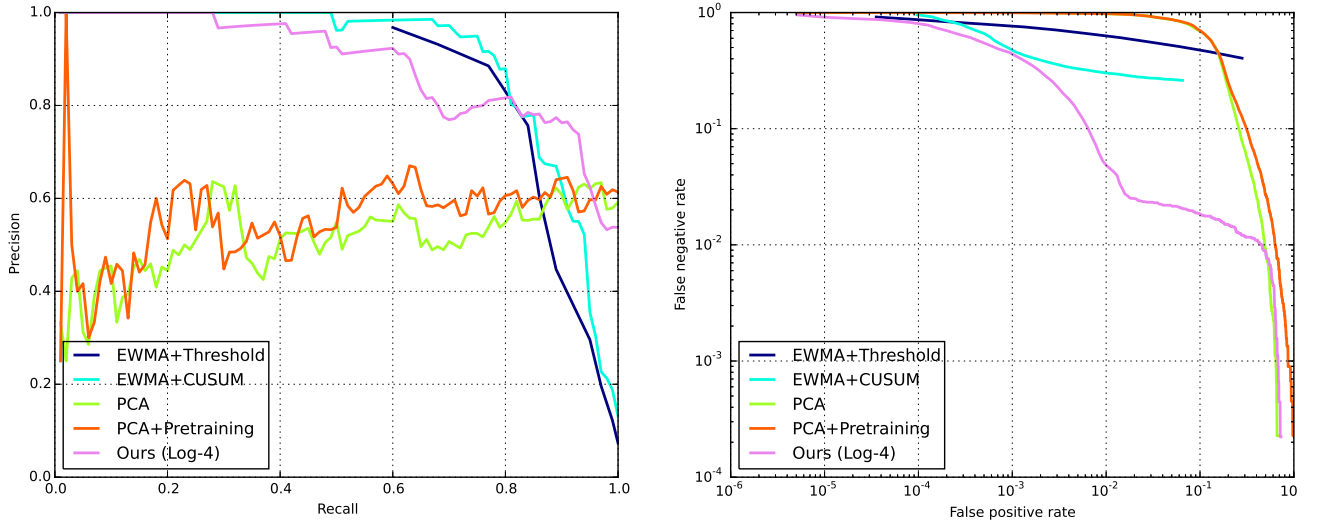


Рисунок 3.6: Эмпирическое сравнение эффективности обнаружения разладки в данных ARTIFICIAL-EASY для процедур на основе EWMA, процедур на основе PCA, и алгоритма раздела 3.3. Слева: кривые «точность–полнота». Справа: кривые средней относительной точности сегментации, см. раздел 2.4.

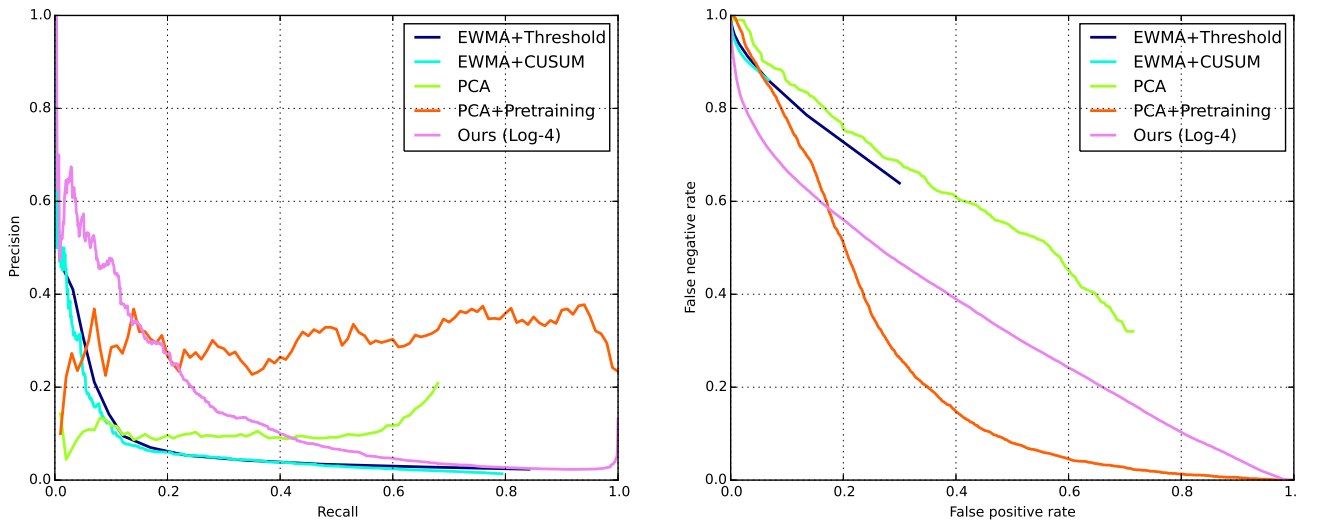


Рисунок 3.7: Эмпирическое сравнение эффективности обнаружения разладки в данных ARTIFICIAL-HARD для процедур на основе EWMA, процедур на основе PCA, и алгоритма раздела 3.3. Слева: кривые «точность–полнота». Справа: кривые средней относительной точности сегментации, см. раздел 2.4.

истории наблюдений $\{\xi_k, k = 1, \dots, t_\ell\}$. Исследование точности алгоритма на выборке ARTIFICIAL-EASY показывает, что предлагаемый в диссертационной работе подход работает существенно точнее, чем EWMA и сравним по точности с процедурами на основе анализа главных компонент. Пример решения задачи оценивания тренда представлен на рис. 3.5; можно заметить, что предложенный подход строит более робастную оценку тренда по сравнению с подходом EWMA.

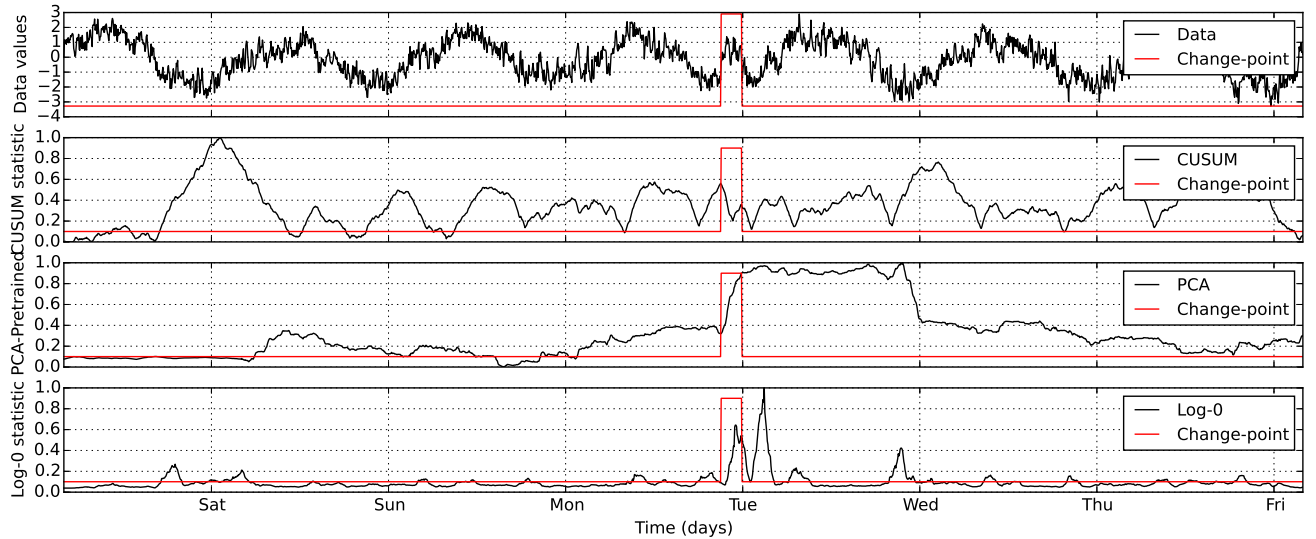


Рисунок 3.8: Сравнение траекторий статистик, используемых в процедурах CUSUM, PCA-Pretrained и ансамбле LOG-0 для данных ARTIFICIAL-HARD. Верхний рис.: траектория искусственных данных и индикатор разладки. Второй сверху рис.: траектория статистики процедуры CUSUM и индикатор разладки. Заметим отсутствие корреляции с индикатором разладки. Третий сверху рис.: траектория статистики процедуры PCA-Pretrained и индикатор разладки. Заметим слабую корреляцию с индикатором разладки. Нижний рис.: траектория статистики процедуры LOG-0 и индикатор разладки. Заметим высокую корреляцию с индикатором разладки. Все статистики были отмасштабированы для удобства просмотра.

3.6.4 Результаты

На данных ARTIFICIAL-EASY предлагаемый в диссертационной работе подход немного проигрывает лишь оптимальной процедуре кумулятивных сумм в терминах площади под кривой «точность–полнота», см. левый рис. 3.6. На данных ARTIFICIAL-HARD предлагаемый подход работает лучше, чем все другие процедуры, в равных условиях. Добавление дополнительной недели данных для улучшения точности разложения временного ряда в методе PCA позволяет этому подходу показать наилучшие результаты обнаружения разладки, см. левый рис. 3.7. Предлагаемый в диссертационной работе подход осуществляет наиболее точную сегментацию временного ряда, как можно заметить из правых частей рис. 3.6 и рис. 3.7. На практике это означает, что с использованием предлагаемого подхода можно добиться меньшей средней продолжительности как ложной тревоги, так и ложного молчания о разладке.

Разработанный подход к обнаружению разладки значительно превосходит другие процедуры в терминах кривой «точность–полнота». Причина повышения

эффективности обнаружения разладки заключается в высокой корреляции сигнала тревоги и индикатора истинной разладки, см. рис. 3.8. Заметим, однако, что ввиду сложной природы рассматриваемых данных многие разладки трудно обнаружить.

3.7 Выводы

Разработаны и исследованы математические модели временных рядов с трендом, отталкивающиеся от различных предположений о структуре наблюдений:

- математическая модель на основе локально полиномиального разложения тренда временного ряда, наблюдаемого в шуме с длинной памятью;
- математическая модель временного ряда, включающая явную модель сезонности.

Предложены алгоритмы оценивания параметров этих моделей по обучающей выборке, их эффективность исследована эмпирически в ходе вычислительного эксперимента. Сравнение точности оценивания тренда для указанных алгоритмов с рядом известных процедур показывает, что разработанные модели и алгоритмы эффективны в задачах оценивания параметров квазипериодических сигналов. Кроме того, эмпирическое исследование качества обнаружения разладки квазипериодических сигналов указывает на то, что эффективность обнаружения разладки с помощью предложенных подходов выше по сравнению с подходами, известными из литературы.

Глава 4

Комплекс программ

В настоящей главе обсуждаются идеологические и архитектурные предпосылки комплекса программ, реализованного в ходе диссертационной работы; его структуру и взаимосвязь его частей; краткое описание дополнительного функционала, обеспечивающего гибкость комплекса как целого. Настоящий комплекс, при грамотном использовании, позволяет реализовать высокоэффективную схему обнаружения отказов программно-аппаратных комплексов очень высокой сложности, и эта возможность будет продемонстрирована ниже в главе 5 на примере организации мониторинга поискового сервиса Яндекса.

В каком-то смысле и вся диссертационная работа выросла из его программной части, изучая и развивая которую, автор и его коллеги по Яндексу получили ценный опыт разработки интеллектуальных алгоритмов анализа данных.

4.1 Предпосылки и архитектура

Программные модули, реализующие алгоритмы интеллектуального анализа данных, как правило, удобно использовать в качестве «черных ящиков», т. е. задействуя лишь внешний интерфейс и абстрагируясь от деталей реализации. «Черный ящик» не только снижает сложность целой программной системы (как и любой вызываемый модуль), но и избавляет от необходимости отладки реализованного программного кода, часто недокументированного, запутанного и обладающего нетривиальной логикой. Согласно этой идеологии, на вход алгоритма обнаружения разладки, реализуемого таким модулем, поступает поток

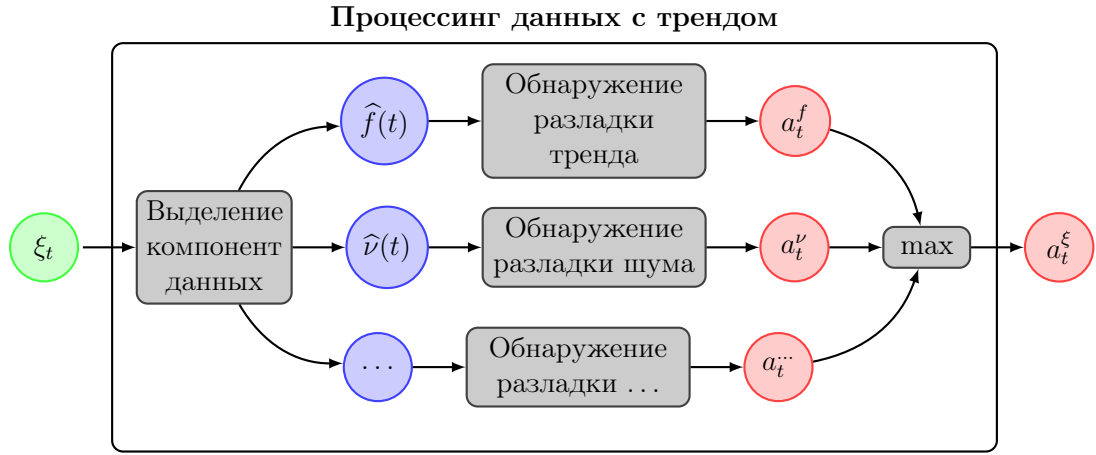


Рисунок 4.1: Принципиальная схема программного модуля процессинга данных с трендом

данных ξ_t ; на его выходе формируется сигнал a_t^ξ , который вспомогательный код использует для передачи сообщения о появлении отказа¹.

Пользуясь этой удобной логикой и применяя изложенные в главах 1–3 подходы, легко изобразить принципиальную схему компоненты, реализующей алгоритм обнаружения разладки в потоке данных — она приведена на рис. 4.1. Согласно терминологии программного комплекса, на вход автоматической компоненте обнаружения разладки — профилю мониторинга (Profile) — в режиме реального времени поступают данные — пары $(X_1, t_1), (X_2, t_2), \dots$ «измерение—момент времени». Профиль мониторинга осуществляет, если это требуется, выделение независимых компонент этого входного потока, таких как текущий профиль цикла и его амплитуду, и передает эти компоненты на обработку статистическим подпрограммам — статистикам (Statistic) и их ансамблям (Aggregation). Последние реализуют алгоритмы обнаружения разладки временного ряда, предполагая *стандартную модель разладки*², и подают сигналы тревоги (Alarm) $a_t^i, i = 1, \dots, K$, — временные ряды, множества значений каждого из которых ограничены интервалом $[0, 2]$, где K — число выделяемых независимых компонент входного потока, причем *аномальному* состоянию соответствуют значения $a_t^i \geq 1, i = 1, \dots, K$. Объединяя выходы этих подпрограмм (например, взяв их максимум), получаем сигнал тревоги профиля мониторинга.

¹например, сообщения электронной почты или смс-сообщения

²В литературе, как правило, стандартная модель разладки заключается в изменении среднего значения стационарной гауссовской случайной последовательности. В этом случае наблюдаемый процесс $\xi = (\xi_t)_{t \geq 0}$ имеет вид $\xi_t = \mu \mathbb{1}_{\{t \geq \theta\}} + \nu_t$, где $\mu \in \mathbb{R}$ — магнитуда разладки, $\theta \geq 0$ — момент появления разладки, и $\nu = (\nu_t)_{t \geq 0}$ — последовательность независимых стандартно нормально распределенных случайных величин.

4.2 Структура комплекса

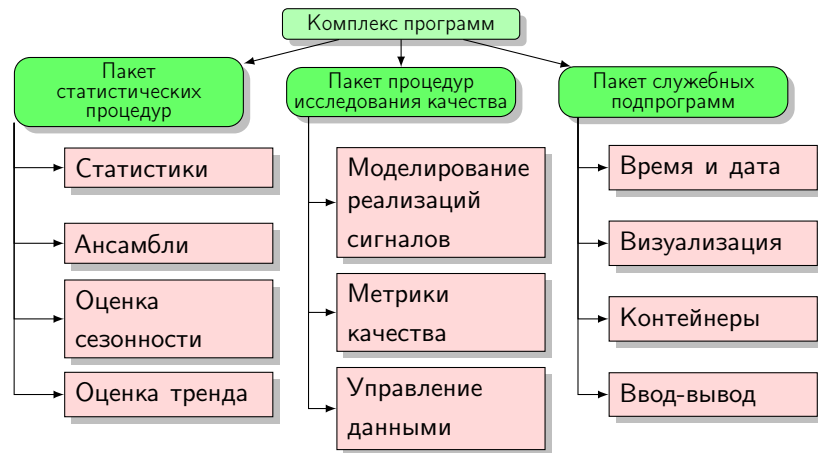


Рисунок 4.2: Структура основных составляющих реализованного программного комплекса

Разработанный программный комплекс организован в структуру из пяти высокомодульных пакетов. Каждый из пакетов реализует одну из частей функционала, описанного в диссертационной работе.

1. Пакет, реализующий алгоритмы оптимального оценивания параметров тренда сигнала, наблюдаемого во шуме с длинной памятью.
2. Пакет, реализующий вычислительный алгоритм оценивания компонент квазипериодической модели (3.5).
3. Пакет для работы с ансамблями «слабых» детекторов, в том числе:
 - модуль численной оптимизации сглаженного эмпирического риска (2.24) по заданной обучающей выборке;
 - модуль обнаружения кратковременной разладки случайного сигнала на основе ансамбля (2.17) (в том числе в режиме реального времени).
4. Пакет моделирования реализаций случайных сигналов с заданными статистическими параметрами, такими как кратковременные разладки, квазипериодические тренды и длинная память.
5. Пакет оценивания эффективности исследуемых алгоритмов и визуализации данных.

В качестве платформы для реализации разработанных математических методов и алгоритмов используется язык программирования **python**, библиотека математических функций **numpy** и библиотека научных расчетов **scipy**. Диаграммы всех классов, относящихся к описанным пакетам, приведены на рис. 4.3–4.4.

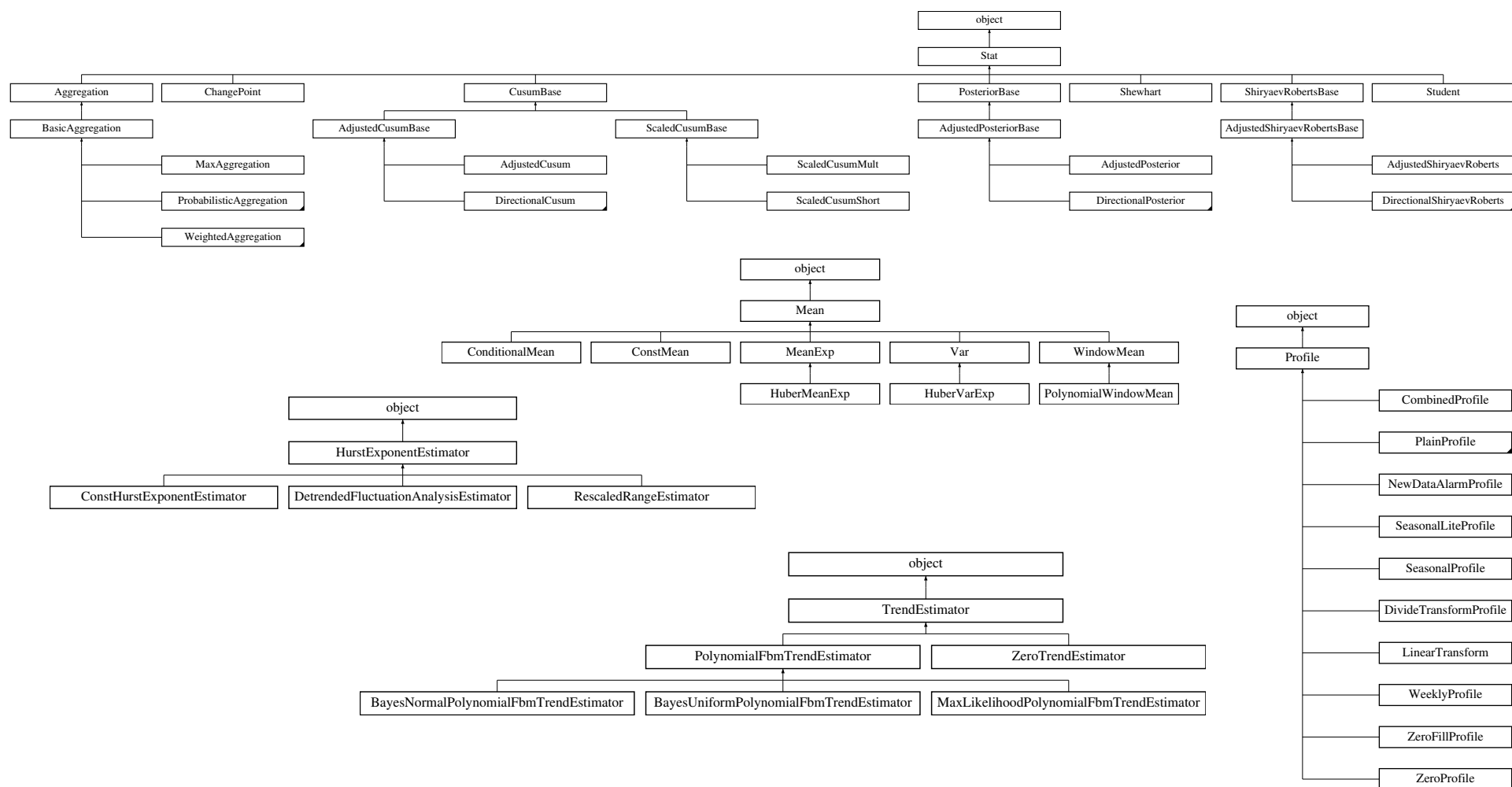


Рисунок 4.3: Иерархия классов пакета статистических процедур

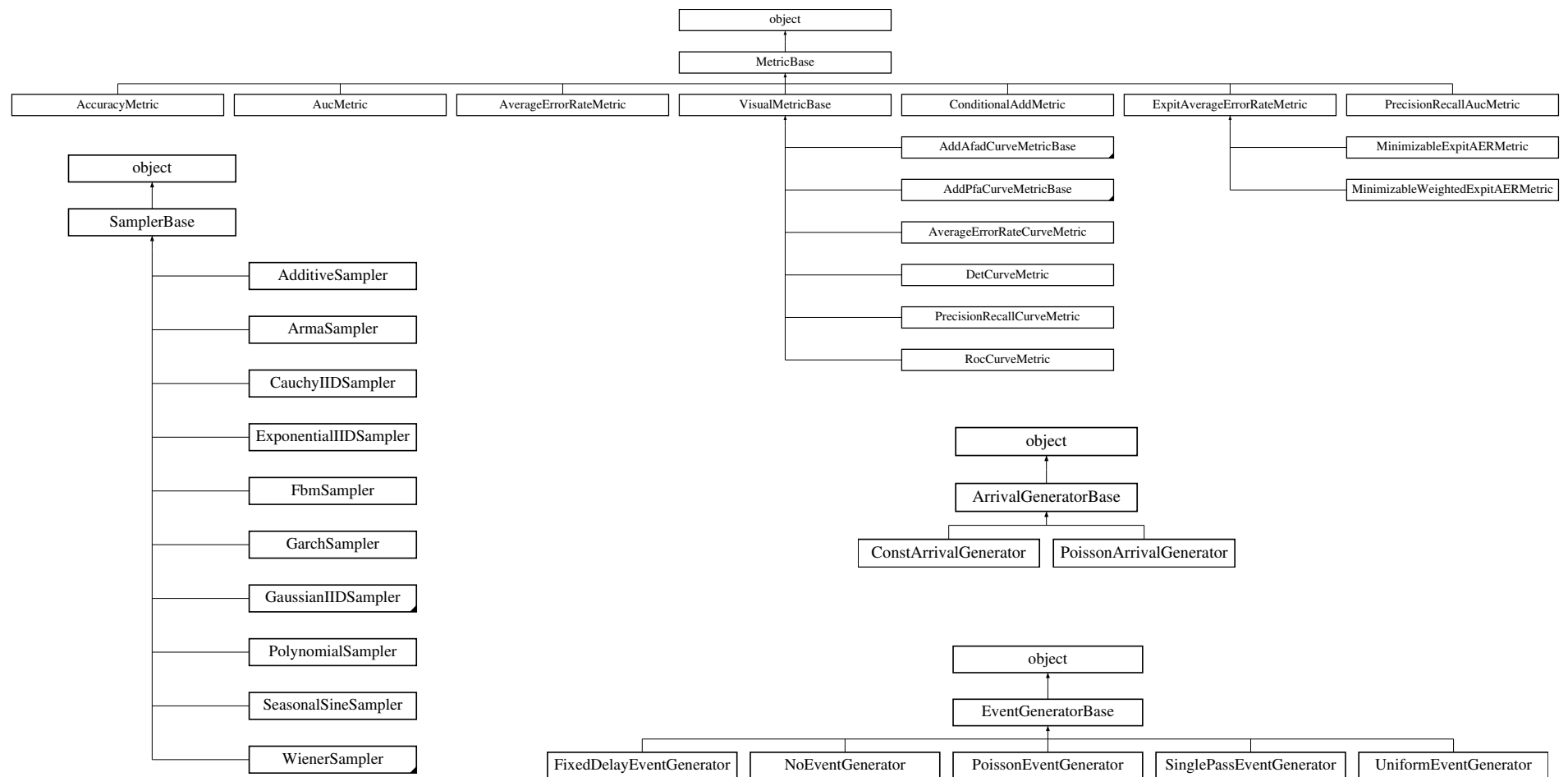


Рисунок 4.4: Иерархия классов пакета моделирования реализаций случайных сигналов и пакет оценивания эффективности исследуемых алгоритмов и визуализации данных

4.3 Дополнительные функциональные возможности

Помимо базовых возможностей разработанного программного комплекса отметим еще две его особенности, превращающие его в настоящую промышленную машину, способную оперативно разворачиваться для мониторинга крупных систем с высоким числом наблюдаемых характеристик. Этими особенностями являются возможность настройки алгоритма по умолчанию и возможность эффективного масштабирования системы мониторинга. Таким образом, при использовании программного комплекса можно осуществить «холодный старт» мониторинга и получать разумные результаты при отсутствии больших обучающих выборок.

4.3.1 Алгоритм «по умолчанию»

На практике для получения устойчивых оценок в алгоритмах из разделов 3.3–3.4 необходимо накопить не менее месяца наблюдений (что соответствует 1344 точкам для почасовых данных или 8064 точкам для пятиминутных данных). Однако при запуске мониторинга нового сервиса эти данные, как правило, недоступны.

Целью создания алгоритма «по умолчанию» являются автоматизация обработки новых наблюдений без необходимости оценивания параметров полноценных моделей тренда; сокращение времени, требуемого для оценивания параметров моделей тренда после передачи данных; создание настройки параметров разладок по умолчанию для заказчиков. При этом естественным требованием, предъявляемыми к алгоритму, является не слишком большое снижение эффективности обнаружения разладки.

В диссертационной работе предлагается простой подход, при котором оценка тренда и дисперсии помехи для общей модели (3.2) выполняются с помощью робастных версий скользящего среднего [30]. А именно, тренд $f(t)$ оценивается величиной $\hat{\mu} = (\hat{\mu}_t)_{t \geq 0}$

$$\hat{\mu}_t = (1 - \alpha_\mu)\hat{\mu}_{t-1} + \alpha_\mu\eta_t,$$

$$\eta_t = \xi_t + h(\xi_t - \hat{\mu}_{t-1}; \beta_\sigma \hat{\sigma}_t),$$

где α_μ — вес нового наблюдения, β_σ — фактор а функция $h(x; m_x)$ задается соотношением

$$h(x; m_x) = \begin{cases} x, & \text{если } |x| \leq m_x, \\ \text{sign}(x)m_x, & \text{если } |x| > m_x. \end{cases}$$

Таким образом, если величина $\xi_t - \hat{\mu}_{t-1}$ разности измеренного и ожидаемого значения временного ряда превышает величину $\beta_\sigma \hat{\sigma}_t$, оценка $\hat{\mu}_t$ изменяется не более, чем на величину $\alpha_\mu \beta_\sigma \hat{\sigma}_t$. Такая схема изменения значения статистики гарантирует защиту от сильного влияния выбросов в значениях наблюдений на результат оценивания. Оценка дисперсии временного ряда обновляется согласно схожему соотношению

$$\begin{aligned} \hat{\sigma}_t^2 &= (1 - \alpha_\sigma) \hat{\sigma}_{t-1}^2 + \alpha_\sigma \eta_t, \\ \eta_t &= \hat{\sigma}_{t-1}^2 h\left(\frac{\xi_t - \hat{\mu}_{t-1}}{\hat{\sigma}_{t-1}}; 1\right), \end{aligned}$$

где α_σ — вес нового наблюдения. Алгоритмы обнаружения разладки используют модель (3.17), в которой $\hat{f}(t) \equiv \hat{\mu}_t$.

4.3.2 Возможности масштабирования системы обнаружения разладок

Предложенный подход оценивания параметров модели является существенно непараметрическим. Большинство параметров, используемых в нем, имеют или фиксированные значения, например, тип ядерной функции или параметр Δ в (3.12), равный периоду сезонности, или оцениваются по данным, например ширина ядра h , периодическая компонента $S_{\varphi(t)}$ и ее дисперсия $\sigma_{\varphi(t)}$. Поэтому модели зависят лишь от параметров, имеющих ясный физический смысл: вес нового наблюдения α в (3.12), требуемое соотношение среднего времени до ложной тревоги $E_\infty \tau$ и задержки в обнаружении разладки $E_0 \tau$ (см. и другие критерии раздела 2.4) и минимальные значения детектируемых величин разладки μ в (2.11) (в предположении нормальных наблюдений). Задавая эти параметры, можем тонко настроить схему обнаружения для работы с потенциально очень большой группой измеряемых характеристик, обладающих схожими свойствами. Данный подход обеспечивает дешевое масштабирование системы детектирования

разладок, поскольку он, во-первых, избавляет от необходимости выбора модели для каждого регистрируемого временного ряда, и, во-вторых, делает возможным автоматическое добавление новых измеряемых характеристик, если они имеют известные системе свойства. Благодаря этим качествам система обнаружения разладок легко масштабируется с увеличением числа регистрируемых данных посредством добавления вычислительных узлов. Отметим, что система обнаружения разладок, развернутая в Яндексе, обрабатывает 3,5 млн точек в день (порядка 150 тыс. точек в час), при этом обработка одного нового наблюдения составляет около 0,1–0,4 с.

4.4 Выводы

Предложенные алгоритмы реализованы в виде комплекса программ с модульной структурой, который интегрирован в библиотеку решения задач обнаружения разладок и аномалий сигналов с трендом. Помимо основной функциональности у комплекса есть ряд дополнительных возможностей, востребованных в прикладных задачах. Среди таких функций отметим возможности одновременной обработки нескольких временных рядов, разметки полученных сообщений об ошибках, обработки данных с пропусками и плавающим интервалом поступления, а также возможность обнаружения разладки алгоритмом «по умолчанию», позволяющая осуществлять крупномасштабное развертывание системы в промышленном окружении.

Глава 5

Результаты решения прикладных задач

5.1 Введение

В данной главе описаны результаты применения классических и разработанных подходов в реальных задачах прогнозирования значений финансовых показателей, оценки параметров нагрузки сетей передачи данных, обнаружения разладок и аномалий системы Яндекс.Поиск и исследования возможности детектирования изменения режима турбулентного течения. Использовались как модельные (но полученных физически точным алгоритмом моделирования), так и реальные данные. В каждом из следующих разделов 5.2–5.5 подробно описываются: исследуемая задача; природа рассматриваемых данных; их источник или параметры порождающего алгоритма; характеристики всего набора данных, такие как число отсчетов и количество имеющихся траекторий; предпосылки и идеология использованного подхода анализа данных; его точные параметры и установки; иллюстрации результатов его работы.

5.2 Задача прогнозирования значений финансовых показателей

Рассмотрим задачу прогнозирования цены закрытия финансового показателя S&P 500¹, исторические данные значений которого доступны, например, на сервере Yahoo! Finance². Задача прогнозирования заключается в предсказании

¹https://en.wikipedia.org/wiki/S%26P_500_Index

²<https://finance.yahoo.com>

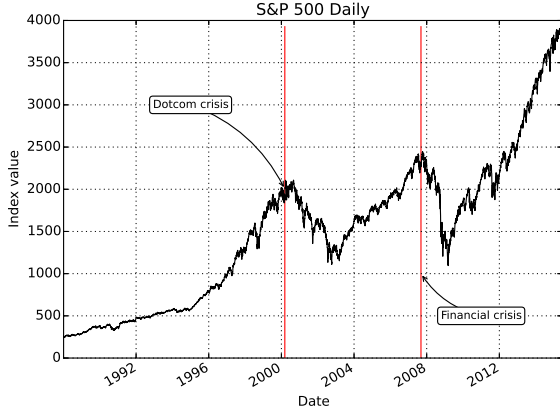


Рисунок 5.1: Значения индекса S&P 500, использованные при решении задачи прогнозирования значения этого индекса на один день вперед

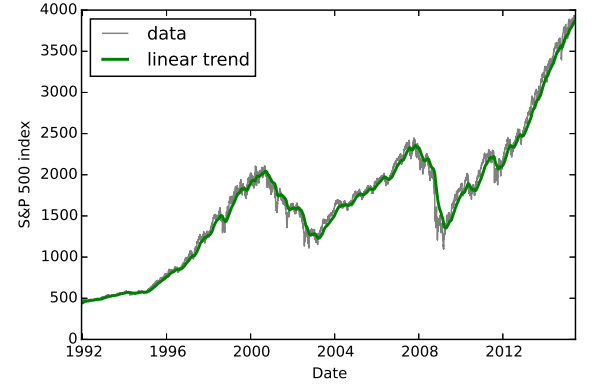


Рисунок 5.2: Значения индекса S&P 500 (серая линия) и значения прогноза этого индекса на один день вперед (зеленая линия)

значения ξ_{t+1} цены закрытия в день $t + 1$ по известному отрезку $\{\xi_s, 0 \leq s \leq t\}$ значений наблюдаемого временного ряда $\xi = (\xi_t)_{t \geq 0}$. В качестве критерия точности рассматривалась относительная погрешность оценивания, даваемая соотношением (3.6.3). В анализе использовалось 7073 точки, записанных за период с 04.01.1988 по 31.12.2015, значения которых были загружены с помощью библиотеки `pandas-datareader`³. Их значения показаны на рис. 5.1 вместе с двумя важными событиями, происшедшими в указанный период — началом так называемого кризиса доткомов 2001 г. и началом ипотечного финансового кризиса 2008 г.

Для решения задачи прогнозирования рассматривалась модель (1.1) в которой предполагался локально линейный тренд $f(t) = \theta_0 + \theta_1 t$, а модель помехи задавалась процессом фрактального броуновского движения $\nu_t = \sigma B_t^H$, причем значение показателя Херста H было неизвестным. Для оценивания параметров θ_0, θ_1 тренда рассматривалось скользящее окно $X_i, \dots, X_{i+\Delta}$ последовательных наблюдений, вычислялась оценка \hat{H} значения показателя Херста H и затем оценки $(\hat{\theta}_0)_{\text{ML}}, (\hat{\theta}_1)_{\text{ML}}$ согласно Теореме 1 раздела 1.3. Значение $\xi_{t+\Delta+1}$ оценивалось величиной $\hat{f}(t + \Delta + 1) = (\hat{\theta}_0)_{\text{ML}} + (\hat{\theta}_1)_{\text{ML}}(t + \Delta + 1)$.

Использовались значения $\Delta = 1000$. Для оценивания значения показателя Херста использовалась `python`-реализация⁴ алгоритма анализа бестрендовых корреляций [15].

³<https://github.com/pydata/pandas-datareader>

⁴<https://github.com/dokato/dfa>

Траектория результата прогноза $\hat{f}(t)$ показана на рис. 5.2. Значение относительной погрешности оценивания составило 1.16% против 1.86% при использовании $H = 1/2$, что представляет снижение погрешности оценивания на 62%.

5.3 Задача оценки параметров нагрузки сетей передачи данных

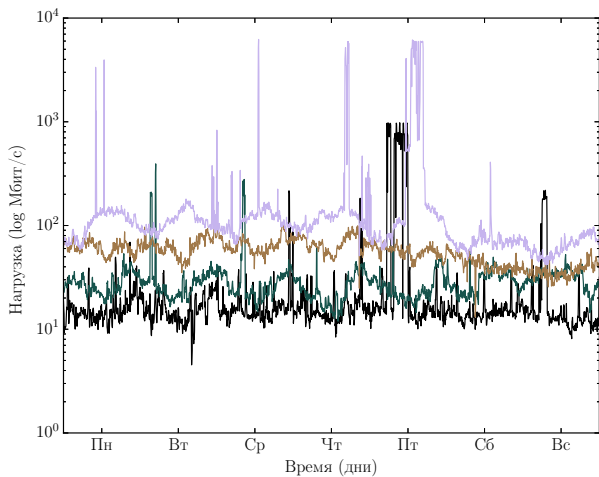


Рисунок 5.3: Примеры недельных временных рядов, отвечающих нагрузкам соединений Атланта–Лос Анджелес, Лос Анджелес–Чикаго, Сиэтл–Лос Анджелес, Чикаго–Лос Анджелес за неделю 14–21 июня 2004 г. в сети Абилин (2016 измерений)

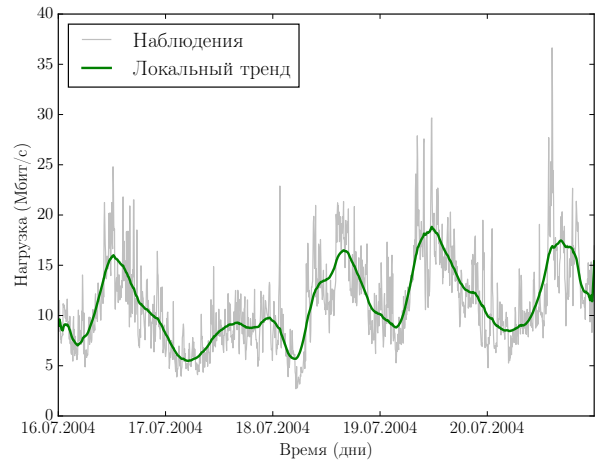


Рисунок 5.4: Оценивание квазипериодического тренда нагрузки соединения Хьюстон–Чикаго

Рассмотрим задачу оценивания сезонного профиля нагрузки сети передачи данных, используя публично доступные данные, описывающие нагрузку в Интернет 2-сети Абилин [77; 93]. Эта база данных часто используется для проверки эффективности тех или иных алгоритмов выделения сезонной составляющей и обнаружения разладок ввиду естественного наличия сезонности, а также аномалий, связанных с теми или иными аспектами работы сети [47; 63]. База данных Абилин описывает сетевую нагрузку в сети Абилин⁵, предоставляя агрегированную за пятиминутные интервалы информацию об объеме трафика, прошедшего между двумя узлами (endpoints) сети, в полугодовой период с 1 марта по 10 сентября 2004 года. В базе данных Абилин присутствует 132 временных ряда,

⁵https://en.wikipedia.org/wiki/Abilene_Network

соответствующих передаче данных между 12 узлами сети. Несколько отрезков этих временных рядов показано на рис. 5.3.

Задача оценивания сезонного профиля нагрузки заключалась в построении оценки $\hat{f}(t)$ гладкого тренда нагрузки $f(t)$ для каждого $t \geq 0$ по отрезку наблюдений $\{\xi_s, 0 \leq s \leq t\}$. Для решения этой задачи рассматривалась модель (3.3), в которой предполагался локально кубический тренд $f(t) = \sum_{i=0}^3 \theta_i t^i$, а модель помехи задавалась процессом фрактального броуновского движения $\nu_t^H = \sigma B_t^H$, причем значение показателя Херста H было неизвестным. Для вычисления оценки $\hat{f}(t)$ применялся алгоритм оценивания тренда с поправкой на длинную память, разработанный в разделе 3.3.

В качестве параметров использованного алгоритма были выбраны значения ширины окна $\Delta = 128$ точек, что соответствует примерно 10 часам наблюдений, использовался сдвиг окна на одну точку, число повторений с переоценкой параметра Херста равнялось двум. Для оценивания значения показателя Херста использовался алгоритм анализа бестрендовых корреляций [15].

Результат оценивания нагрузки представлен на рис. 5.4. Общее время работы алгоритма оценивания составило 12 секунд на одном вычислительном ядре.

5.4 Задача обнаружения разладок и аномалий поисковой системы Яндекс

Поисковая система Яндекс является высоконагруженной программной системой, обрабатывающей около 5 млрд поисковых запросов ежемесячно⁶, вследствие чего ее архитектура является существенно распределенной. Рассмотрим задачу обнаружения отказов поисковой системы по данным измерения ее характеристик, выполняемым в реальном времени. Основные типы измеряемых характеристик включают так называемые счетчики, характеризующие интенсивность нагрузки системы (число запросов, число срабатываний поискового алгоритма переранжирований, количество обращений к поисковому индексу), и так называемые метрики, т. е. усредненные по выборке поисковых запросов показатели, отражаю-

⁶что составляет около 165 млн запросов в день или около 2000 запросов в секунду, данные согласно источнику http://www.vedomosti.ru/technology/articles/2013/02/07/yandeks_obogal_microsoft

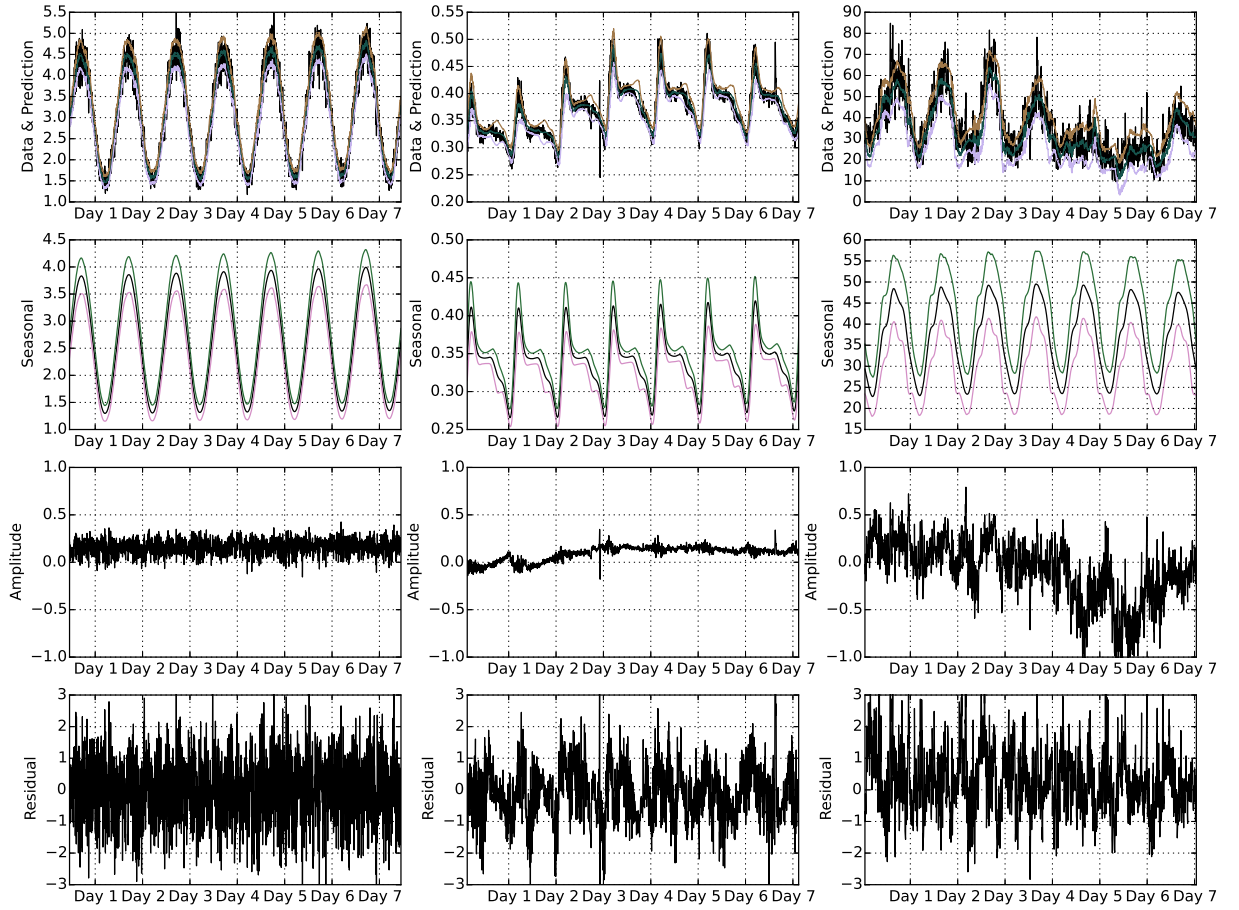


Рисунок 5.5: Результаты разложения временного ряда с использованием алгоритма раздела 3.4.2 для трех различных наборов данных: описанного в разделе 3.4 искусственного набора (левый столбец), описываемого в настоящем разделе набора данных Яндекса (центральный столбец), и набора данных Абилин (правый столбец). На графиках представлены: исходный временной ряд ξ , и результат оценивания тренда $\hat{f}(t)$ с доверительными интервалами $\hat{f}(t) \pm \hat{\sigma}_{\varphi(t)}$ (верхний рисунок), оценка $\hat{S}_{\varphi(t)}$ сезонной компоненты $S_{\varphi(t)}$ с доверительными интервалами $\hat{S}_{\varphi(t)} \pm \hat{\sigma}_{\varphi(t)}$ (второй сверху рисунок), оценка \hat{Q}_t амплитуды Q_t сезонного профиля (третий сверху рисунок), оценка $\hat{\nu}_t$ случайной помехи ν_t (нижний рисунок).

щие эффективность их обработки (среднее время ответа в секундах, показатели кликабельности ссылок, доли запросов без кликов по результатам поиска и т. п.).

База данных соответствующих показателей содержит 8147 одновременно наблюдаемых характеристик с временным разрешением 5 мин, 30 мин или 24 ч (пятиминутные, получасовые или дневные варианты данных, соответственно). В пятиминутных и получасовых данных присутствует суточная цикличность (длина периода $T = 288$ или $T = 48$ точек в день, соответственно). Для небольшого подмножества временных рядов (порядка 100) имеется экспертная разметка наличия отказов. Пример временных рядов обоих упомянутых типов представ-

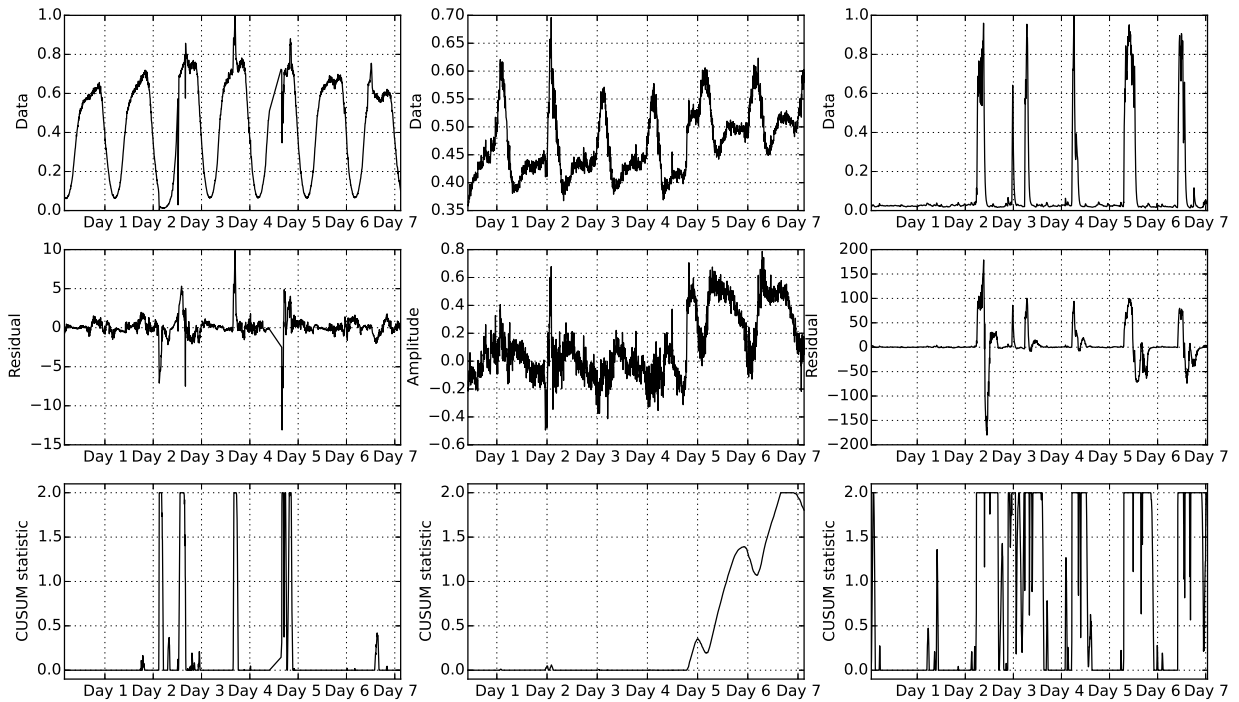


Рисунок 5.6: Результаты обнаружения разладки поисковой системы Яндекс для случаев краткосрочных изменений с характерной длительностью много меньше периода (левый столбец) и долгосрочных изменений масштаба (центральный столбец). Для сравнения типов рассматриваемых данных также представлены результаты обнаружения разладки по данным Абилин (правый столбец, ср. раздел 5.3). На графиках представлены: исходный временной ряд ξ (верхний рисунок), выделяемая компонента, анализируемая на предмет разладки (второй сверху рисунок): остатки R , вычисляемые согласно (3.17), для случая краткосрочных разладок в данных Яндекса и в данных Абилин, и амплитуда A , вычисляемая согласно алгоритму раздела 3.4.2, для случая долгосрочных разладок в данных Яндекса, и траектория статистики кумулятивных сумм (нижний рисунок).

лен на рис. 3.4 вместе с метками, указывающими на долгосрочные изменения в амплитуде и краткосрочные изменения в нерегулярной компоненте. Заметим, что в данных присутствует медленно меняющийся тренд, соответствующий изменениям масштаба наблюдаемого процесса.

Стандартный подход к обнаружению разладки распределенной поисковой системы включает тщательную ручную калибровку параметров алгоритмов обнаружения разладки для достижения оптимального компромисса между запазыванием в обнаружении разладки и вероятностью ложного срабатывания (или средним временем между ложными срабатываниями). Чрезвычайно высокий объем рассматриваемой системы (более 8000 различных одновременно наблюдаемых характеристик) делает эту задачу чрезмерно трудозатратной. Подход,

рассматриваемый в диссертационной работе, напротив, исключает необходимость ручного подбора параметров алгоритмов обнаружения разладки за счет использования множества размеченных траекторий и автоматического подбора параметров ансамбля. Поэтому для обнаружения разладок указанной системы использовался подход, состоящий в использовании алгоритма оценивания тренда, предложенного в разделе 3.4.2, классических и предложенных в главе 2 алгоритмов детектирования разладки временного ряда.

Результаты применения указанного подхода представлены на рис. 5.5–5.6. Для обучения параметров модели использовался непрерывный 4-недельный отрезок временного ряда, следующий за ним недельный сегмент — для оценки качества модели. Прогноз значения временного ряда использует как поступающие, так и исторические данные, т. е. алгоритм может обновлять оценки, адаптируясь ко входящим данным. Отрезки всех наборов данных имеют 8064 измерений (28 дней, состоящих из 288 5-минутных интервалов каждый) для обучения и 2016 точек (7 дней, состоящих из 288 5-минутных интервалов) для оценки, соответственно. Пример результирующего прогноза вместе с разложением временного ряда показан на рис. 5.5. Рис. 5.6 содержит результаты обнаружения разладки данных Яндекса и Абилин.

5.5 Задача исследования возможности детектирования изменения режима турбулентного течения

В этом разделе исследуется задача обнаружения изменения режима турбулентного течения гидродинамической системы, физическая модель которой разработана в [94]. Физический смысл рассматриваемой задачи заключается в поиске оптимального момента остановки гидродинамического эксперимента при возникновении поломки оборудования с использованием данных измерения относительного давления в различных точках объема модели. Предложены и исследованы простейшие модели данных, представляющие практический интерес в описанной задаче.

5.5.1 Задача обнаружения изменения дисперсии случайного процесса

Рассмотрим стандартную модель разладки в (2.1), в которой распределения $p_\infty(\cdot)$ и $p_0(\cdot)$ моделей наблюдений $\xi^\infty = (\xi_t^\infty)_{t \geq 0}$ и $\xi^0 = (\xi_t^0)_{t \geq 0}$, отвечающих состояниям до и после поломки, соответственно, суть

$$p_\infty = \mathcal{N}(0, \sigma_\infty^2), \quad p_0 = \mathcal{N}(0, \sigma_0^2),$$

что соответствует скачкообразному изменению дисперсии случайного процесса ξ в момент θ . Легко записать выражения для статистики кумулятивных сумм для рассматриваемой модели; она задается соотношениями (2.11), (2.14), в которых статистика ζ_t принимает вид

$$\zeta_t = \log \frac{p_0(X_t)}{p_\infty(X_t)} = \log \frac{\sigma_\infty}{\sigma_0} + \frac{\xi_t^2}{2} \frac{\sigma_0^2 - \sigma_\infty^2}{\sigma_0^2 \sigma_\infty^2}.$$

Принимая (без ограничения общности), что $\sigma_\infty < \sigma_0$, для величин $E_\infty \zeta_t$ и $E_0 \zeta_t$ получим

$$E_\infty \zeta_t = \log r + \frac{1}{2} (1 - r^2), \quad E_0 \zeta_t = \log r + \frac{1}{2} \left(\frac{1}{r^2} - 1 \right),$$

где $r = \sigma_\infty / \sigma_0 < 1$, откуда следует $E_\infty \zeta_t < 0, E_0 \zeta_t > 0$. Сигнал тревоги об обнаружении поломки следует поднимать в первый момент времени τ_B^* такой, что статистика T_n (2.14) превысит некоторую величину B_τ :

$$\tau_B^* = \inf \{t \geq 1, T_n \geq B_\tau\}.$$

Описанная модель сигнала требует, очевидно, оценивания дисперсии временного ряда X_t до и после разладки, в случае, если дисперсия неизвестна. Решить эту проблему можно стандартным способом, используя вместо величин σ_∞^2 и σ_0^2 их выборочные аналоги — статистики S_{0n}^2 и $S_{\infty n}^2$, — подсчитанные по соответствующим «отрезкам» данных, на протяжении которых заведомо не наступает разладка.

Результаты обнаружения разладки с использованием описанного подхода представлены на рис. 5.7. Легко заметить, что обнаружение разладки в этом случае не представляет трудностей.

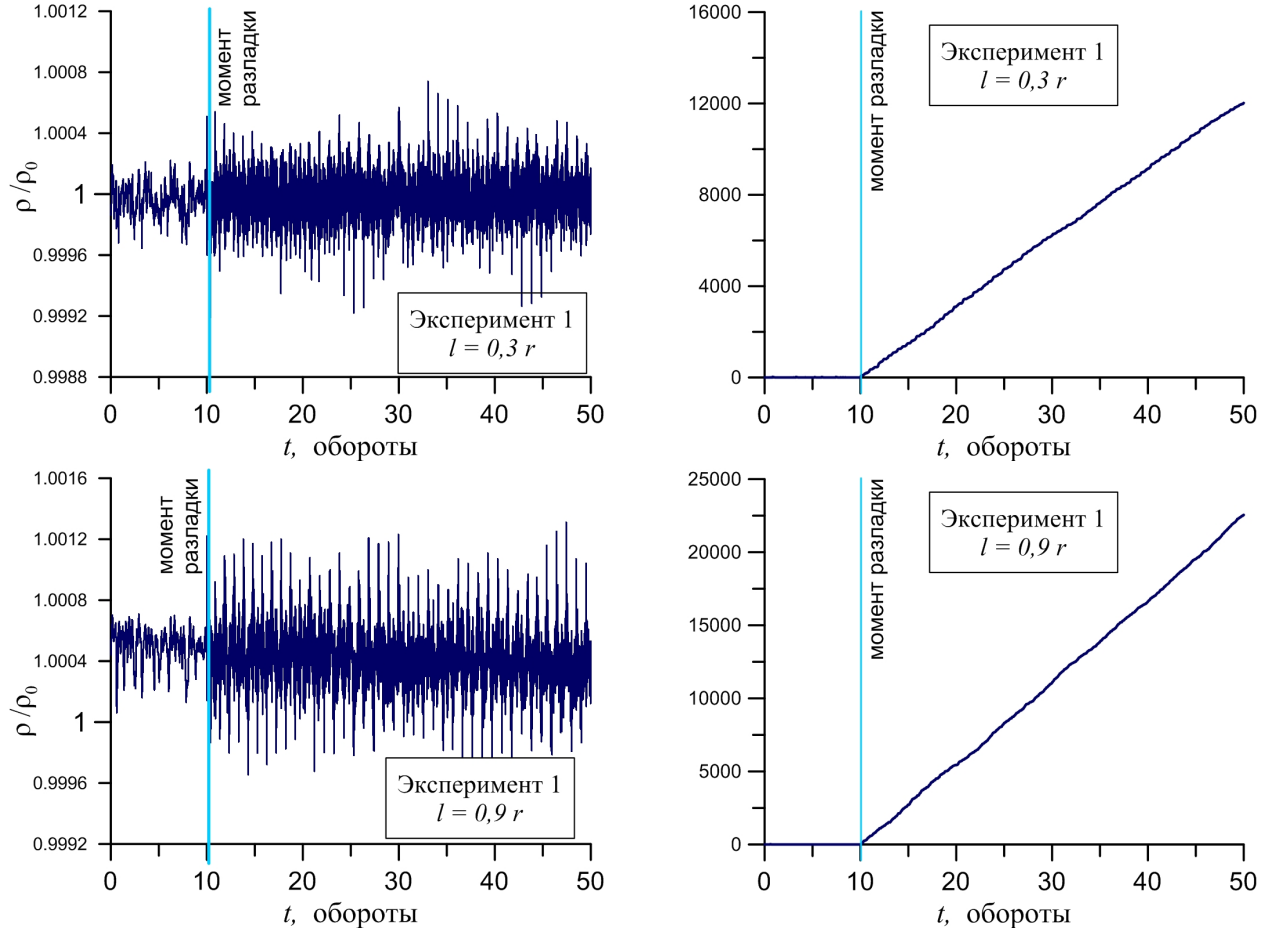


Рисунок 5.7: Результаты решения задачи обнаружения изменения режима турбулентного течения, соответствующего изменению дисперсии процесса относительного давления. Слева представлены реализации процесса относительного давления, измеренного в двух различных точках объема физической модели. Справа представлены реализации процесса кумулятивных сумм.

5.5.2 Задача обнаружения изменения параметров процесса авторегрессии

Рассмотрим модель разладки, которая заключается в изменении значений авторегрессионных коэффициентов в момент θ . Одномерная авторегрессионная модель порядка p может быть записана в виде [68]

$$\xi_t = \sum_{i=1}^p a_i \xi_{t-i} + \sigma Z_t,$$

где a_1, \dots, a_p — набор коэффициентов авторегрессии, определяющих конкретный вид процесса; $Z_t \sim \mathcal{N}(0,1)$ — последовательность независимых стандартных нормальных величин (случайный процесс «белого шума»); σ — дисперсия процесса «белого шума». Разладка в рассматриваемом случае состоит в изменении значений коэффициентов авторегрессии a_1, \dots, a_p . Пусть до момента разладки авторегрессионная модель определяется набором коэффициентов $(a_1^\infty, \dots, a_p^\infty)$, а после разладки — коэффициентами (a_1^0, \dots, a_p^0) , т. е.

$$\begin{aligned}\xi_t^\infty &= \sum_{i=1}^p a_i^\infty \xi_{t-i} + \sigma Z_t, & k = 1, \dots, \theta - 1, \\ \xi_t^0 &= \sum_{i=1}^p a_i^0 \xi_{t-i} + \sigma Z_t, & k = \theta, \theta + 1, \dots\end{aligned}$$

Легко показать, что при известных значениях коэффициентов $(a_1^\infty, \dots, a_p^\infty)$ и (a_1^0, \dots, a_p^0) задача о разладке для процесса авторегрессии сводится к задаче о разладке для среднего значения случайного процесса. В самом деле, распределения случайной величины ξ_t является условно-гауссовским с параметрами $E(\xi_t | \xi_1, \dots, \xi_{t-1}) = \sum_{i=1}^p a_i \xi_{t-i}$ и $V(\xi_t | \xi_1, \dots, \xi_{t-1}) = \sigma^2$. Поэтому явный вид величины ζ_t выражается формулой

$$\zeta_t = \left(\xi_t - \sum_{i=1}^p a_i^\infty \xi_{t-i} \right)^2 - \left(\xi_t - \sum_{i=1}^p a_i^0 \xi_{t-i} \right)^2.$$

Сигнал тревоги об обнаружении поломки следует поднимать в первый момент времени τ_B^* такой, что статистика T_n (2.14) превысит некоторую величину B_τ :

$$\tau_B^* = \inf \{ t \geq 1, T_n \geq B_\tau \}.$$

В заключение остановимся на вопросе получения оценок $\hat{a}_i^\theta, i = 1, \dots, p, \theta = 0, \infty$, величин $a_i^\theta, i = 1, \dots, p, \theta = 0, \infty$. Рассмотрим «отрезки» X_1, \dots, X_K и X_{n-K}, \dots, X_n данных, выбрав K таким, что распределение процесса X_t при $t = 1, \dots, K$ и при $t = n - K, \dots, n$ неизменно и равно P_∞ и P_0 соответственно. Предполагая реализацию X_1, \dots, X_n заданной, запишем уравнения авторегрес-

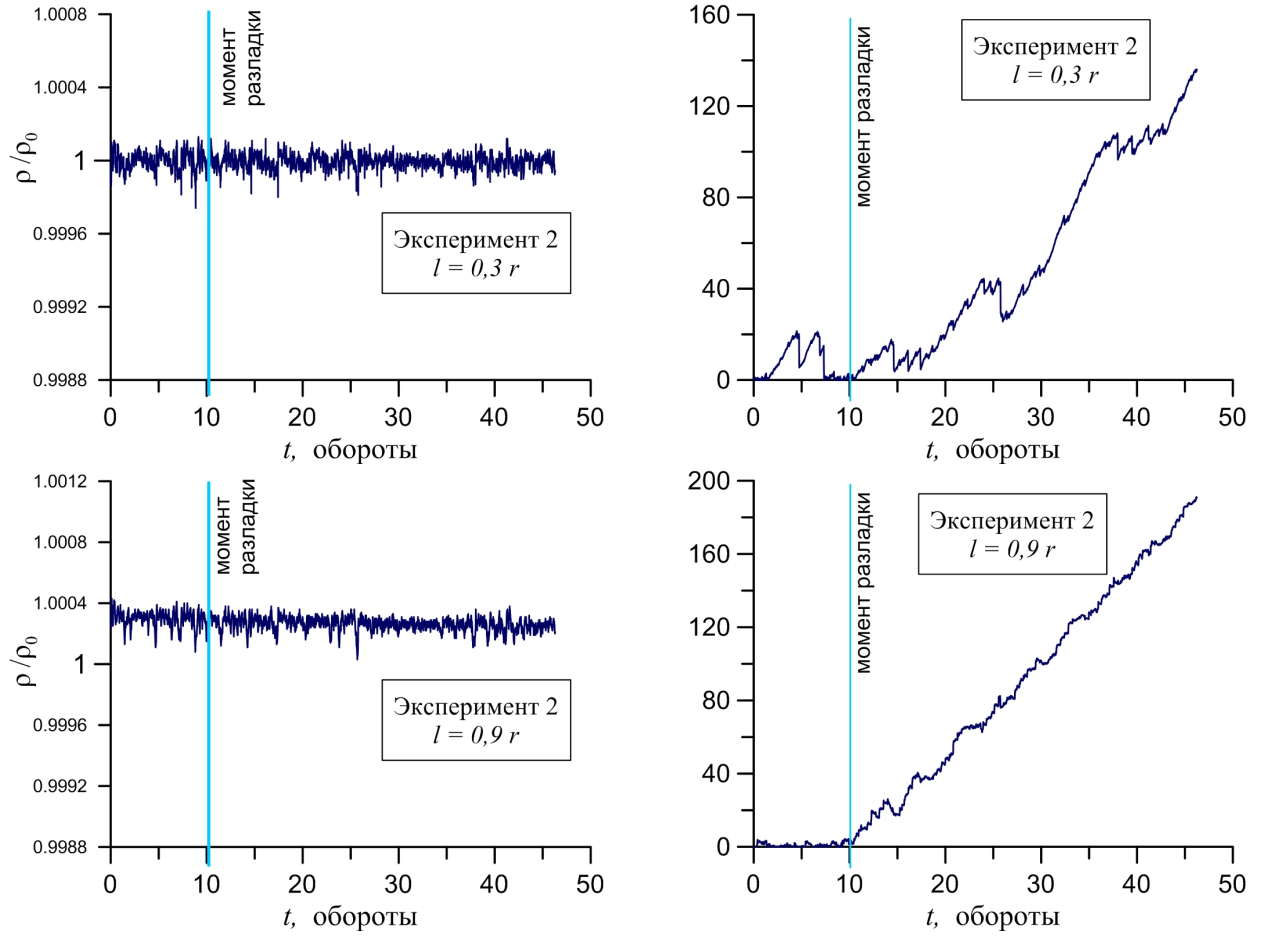


Рисунок 5.8: Результаты решения задачи обнаружения изменения режима турбулентного течения, соответствующего изменению спектра процесса относительного давления. Слева представлены реализации процесса относительного давления, измеренного в двух различных точках объема физической модели. Справа представлены реализации процесса кумулятивных сумм.

сии для отрезка X_1, \dots, X_K (систему уравнений Юла-Уокера [68])

$$\begin{aligned}
 X_{p+1} &= a_1^\infty X_p + \dots + a_p^\infty X_1 + \varepsilon_{p+1}, \\
 X_{p+2} &= a_1^\infty X_{p+1} + \dots + a_p^\infty X_2 + \varepsilon_{p+2}, \\
 &\dots \\
 X_K &= a_1^\infty X_{K-1} + \dots + a_p^\infty X_{K-p} + \varepsilon_K,
 \end{aligned}$$

или в матрично-векторном виде

$$\begin{bmatrix} X_{p+1} \\ X_{p+2} \\ \vdots \\ X_K \end{bmatrix} = \begin{bmatrix} X_p & X_{p-1} & \dots & X_1 \\ X_{p+1} & X_p & \dots & X_2 \\ \vdots & \vdots & \ddots & \vdots \\ X_{K-1} & X_{K-2} & \dots & X_{K-p} \end{bmatrix} \begin{bmatrix} a_1^\infty \\ a_2^\infty \\ \vdots \\ a_p^\infty \end{bmatrix} + \begin{bmatrix} \varepsilon_{p+1} \\ \varepsilon_{p+2} \\ \vdots \\ \varepsilon_K \end{bmatrix}. \quad (5.1)$$

Легко понять, что оценка $\hat{a}_1^\infty, \dots, \hat{a}_p^\infty$ параметров $a_1^\infty, \dots, a_p^\infty$ авторегрессии может быть получена в виде оценки наименьших квадратов. А именно, обозначая $\mathbf{a}^\infty = (a_1^\infty, \dots, a_p^\infty)$, $\mathbf{x} = (X_{p+1}, X_{p+2}, \dots, X_{p+K})$ и матрицу в (5.1) \mathbf{A} , получим оценку в виде

$$\hat{\mathbf{a}}^\infty = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{x}.$$

Оценку дисперсии σ «шума» Z_t можно получить также в стандартном виде

$$\hat{\sigma}^2 = \frac{\|\mathbf{x} - \mathbf{A}(\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{x}\|}{K - p}.$$

5.5.3 Исследование оперативных характеристик решения задачи детектирования изменения режима турбулентного течения

Видно, что статистика кумулятивных сумм позволяет с определенной точностью детектировать момент разладки. Для характеристики точности этого определения рассмотрим две основные метрики: частоту ложных срабатываний p_{fa} и среднее время запаздывания в обнаружении разладки $E(t - \theta)$. Качество каждого конкретного алгоритма, как и в случае любой многокритериальной задачи оптимизации, выражается в виде зависимости $E(t - \theta)$ от p_{fa} . В данной работе указанные зависимости, для различных экспериментов и положений датчиков, представлены на рис. 5.9.

Из полученных графиков можно сделать два важных вывода:

1. Если взять критерием адекватности полученного метода регистрации разладок $E(t - \theta)p_{fa} < 1$, т. е. факт того, что среднее запаздывание метода в определении разладки меньше среднего времени между двумя соседними

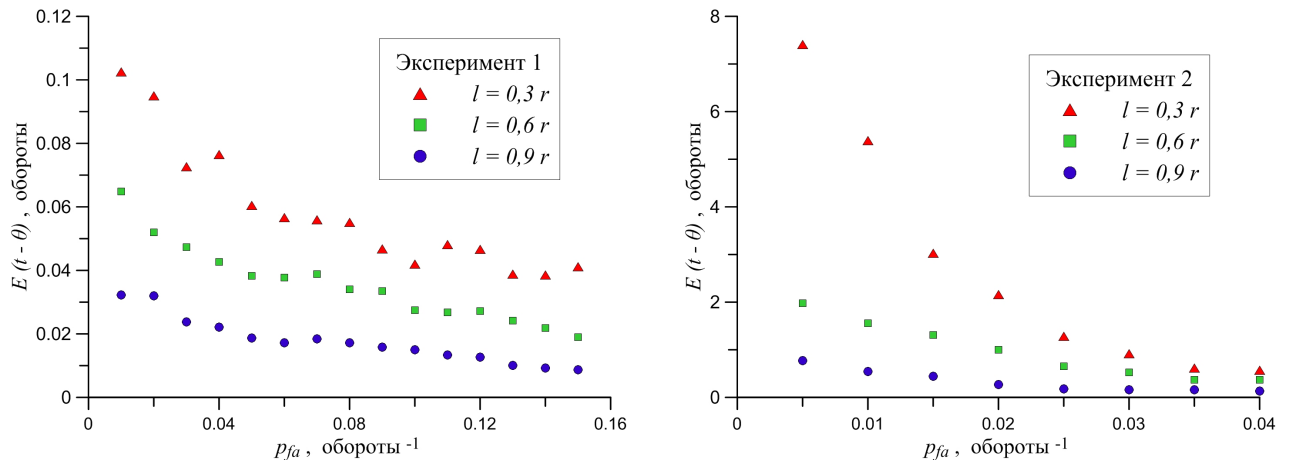


Рисунок 5.9: Оперативные характеристики решения задачи обнаружения изменения режима турбулентного течения: зависимость среднего времени запаздывания в обнаружении разладки от вероятности ложной тревоги для различных положений датчиков давления. l — расстояние от датчика до центра симметрии системы, r — характерный радиус системы. Левый рис.: изменение режима течения соответствует изменению дисперсии процесса относительного давления. Правый рис.: изменение режима течения соответствует изменению спектра процесса относительного давления.

ложными срабатываниями, то выбранные модели шума для использования аппарата разладок оказались адекватными цели исследования.

2. Лучшее качество работы алгоритма достигается при использовании датчиков давления, более удаленных от оси симметрии системы.

5.6 Выводы

Разработанные модели, алгоритмы и комплексы программ были применены для решения реальных задач: прогнозирования значений финансовых показателей, оценки параметров нагрузки сетей передачи данных, обнаружения разладок и аномалий системы Яндекс.Поиск и исследования возможности детектирования изменения режима турбулентного течения.

Алгоритмы оценки параметров временных рядов, использующие разработанные в диссертационной работе подходы, основаны на теоретически эффективных оценках. В частности, алгоритм оценки тренда, наблюдаемого в шуме с длинной памятью, включает явную аналитическую модель длинной памяти, а алгоритм оценивания сезонного тренда использует хорошо изученную оценку Надарая-Ватсона; именно поэтому в практических ситуациях эти математические методы гарантируют эффективность решения задач оценивания сигналов по их зашумленным измерениям.

Что касается алгоритма обнаружения разладки, использующего ансамбль «слабых» детекторов, ввиду сложности оптимизируемого функционала получить теоретические гарантии его эффективности чрезвычайно трудно. Однако на практике, как в ходе вычислительных экспериментов с большим числом независимых повторений (см. главу 2), так и в применении к реальным данным, этот метод обнаружения разладки неизменно дает лучшие результаты по сравнению с классическими процедурами детектирования разладок. Кроме того, важным преимуществом ансамблей по сравнению с классическими процедурами является отсутствие необходимости ручной настройки алгоритма, поскольку существует метод оптимизации, позволяющий автоматически выбирать его параметры.

Заключение

Комплексное рассмотрение проблем обнаружения отказов больших систем с интенсивным программным обеспечением дало возможность создать алгоритмический задел и программную реализацию, которые обладают повышенной эффективностью решения этих задач по сравнению с известными в литературе подходами. Разработанные методология, алгоритмика и программный комплекс гарантируют обнаружение разладок при малом числе ложных тревог; являются гибкими при адаптации для новых источников данных, обладают малым числом задаваемых вручную параметров; являются вычислительно легкими, линейно масштабируются с ростом числа обрабатываемых временных рядов.

Гарантии эффективности алгоритмов оценки параметров временных рядов проистекают из теоретически доказанных свойств оптимальности используемых статистических оценок. В частности, алгоритм оценки тренда, наблюдаемого в шуме с длинной памятью, включает явную аналитическую модель длинной памяти, а алгоритм оценивания сезонного тренда использует хорошо изученную оценку Надарая-Ватсона. Верификация свойств разработанных методов оценивания трендов на модельных задачах и эмпирическое сравнение их качества с известными подходами подтверждают высокую эффективность оценивания параметров сигналов.

Математический метод обнаружения разладки на основе ансамбля «слабых» детекторов всесторонне исследован в ходе интенсивных вычислительных экспериментов и доказал свою эффективность по сравнению с классическими процедурами детектирования разладок в ряде модельных условий по всем метрикам качества процедур этого класса. Кроме того, важным преимуществом ансамблей по сравнению с классическими процедурами является отсутствие необходимости ручной настройки алгоритма, поскольку существует метод оптимизации, позволяющий автоматически выбирать его параметры.

Программная реализация разработанных подходов, выполненная в виде модульного комплекса пакетов программ на языке `python`, обеспечивает гибкость как при вычислительных экспериментах, так и при промышленном развертывании. Решения реальных задач, представленные в диссертационной работе, полностью выполнены с использованием разработанного программного комплекса.

Основные результаты настоящей диссертационной работы заключаются в следующем.

1. Проведено теоретическое исследование модели случайного процесса, управляемого фрактальным броуновским движением, в том числе получены оценка максимального правдоподобия и Байесовская оценка параметра тренда этого процесса, охарактеризован оптимальный момент остановки измерений этого процесса.
2. Разработан и исследован математический метод обнаружения разладки временного ряда на основе ансамбля «слабых» детекторов (т. е. классических процедур обнаружения разладки, оперирующих в условиях нарушения предположений о модели разладки).
3. Разработана и исследована математическая модель временного ряда с гладким (в частности, квазипериодическим) трендом, наблюдаемым в шуме с длинной памятью, и алгоритм оценивания параметров этого тренда.
4. Разработана и исследована математическая модель квазипериодического временного ряда с явным учетом сезонности и алгоритм оценивания параметров профиля сезонности на основе непараметрической регрессии.
5. Предложена методология моделирования временных рядов с трендом и обнаружения их разладок и аномалий.
6. Создан программный комплекс, предоставляющий эффективное решение задач обнаружения разладок и аномалий реальных сигналов.
7. С помощью разработанного программного комплекса решен ряд задач оценивания параметров и обнаружения разладок реальных сигналов.

Список литературы

1. *Abry P., Veitch D.* Wavelet analysis of long-range-dependent traffic // IEEE Transactions on Information Theory. — 1998. — Vol. 44, no. 1.
2. *Adams R. P., MacKay D. J. C.* Bayesian Online Changepoint Detection // arXiv preprint arXiv:0710.3742. — 2007. — P. 7. — arXiv: 0710.3742. — URL: <http://arxiv.org/abs/0710.3742>.
3. *Ahmed T., Oreshkin B., Coates M.* Machine learning approaches to network anomaly detection // Proceedings of the 2nd USENIX workshop on Tackling computer systems problems with machine learning techniques. — USENIX Association. 2007. — Pp. 1–6.
4. An hourly periodic state space model for modelling French national electricity load / V. Dordonnat, S. J. Koopman, M. Ooms, A. Dessertaine, J. Collet // International Journal of Forecasting. — 2008. — Vol. 24, no. 4. — Pp. 566–587.
5. Analysis and modeling of time-correlated failures in large-scale distributed systems / N. Yigitbasi, M. Gallet, D. Kondo, A. Iosup, D. Epema // Proceedings — IEEE/ACM International Workshop on Grid Computing. — 2010. — Pp. 65–72.
6. Anomaly detection in large-scale data stream networks / D.-S. Pham, S. Venkatesh, M. Lazarescu, S. Budhaditya // Data Mining and Knowledge Discovery. — 2014. — Vol. 28, no. 1. — Pp. 145–189.

7. *Baillie R. T.* Long memory processes and fractional integration in econometrics // *Journal of Econometrics*. — 1996. — Vol. 73, no. 1. — Pp. 5–59.
8. *Basseville M., Nikiforov I.* Fault isolation for diagnosis: nuisance rejection and multiple hypothesis testing // *Annual Reviews in Control*. — 2002. — Vol. 26. — Pp. 189–202.
9. *Basseville M., Nikiforov I. V.* Detection of abrupt changes: theory and application. — N.J. : Prentice Hall Englewood Cliffs, 1993.
10. *Ben-Gal I., Morag G., Shmilovici A.* Context-Based Statistical Process Control // *Technometrics*. — 2003. — Vol. 45, no. 4. — Pp. 293–311.
11. *Ben-Gal I., Singer G.* Statistical process control via context modeling of finite-state processes: an application to production monitoring // *IIE Transactions*. — 2004. — Vol. 36, no. 5. — Pp. 401–415.
12. *Brouste A., Kleptsyna M., Popier A.* Fractional Diffusion with Partial Observations // *Communications in Statistics - Theory and Methods*. — 2011. — Vol. 40, no. 19–20. — Pp. 3479–3491.
13. *Bruhns A., Deurveilher G., Roy J.-S.* A non-linear regression model for mid-term load forecasting and improvements in seasonality // *Proceedings of the 15th Power Systems Computation Conference*. — 2005. — Pp. 22–26.
14. *Çetin U., Novikov A., Shiryaev A. N.* Bayesian Sequential Estimation of a Drift of Fractional Brownian Motion // *Sequential Analysis*. — 2013. — Vol. 32, no. 3. — Pp. 288–296.
15. Detrended fluctuation analysis: a scale-free view on neuronal oscillations / R. Hardstone, S.-S. Poil, G. Schiavone, R. Jansen, V. V. Nikulin, H. D. Mansvelder, K. Linkenkaer-Hansen // *Scale-free Dynamics and Critical Phenomena in Cortical Activity*. — 2012. — P. 75.

16. *Du W., Polunchenko A. S., Sokolov G.* On Robustness of the Shiryaev-Roberts Procedure for Quickest Change-Point Detection under Parameter Misspecification in the Post-Change Distribution // arXiv preprint arXiv:1504.04722. — 2015. — arXiv: `arXiv:1504.04722v1`.
17. *Dubovikov M. M., Starchenko N. V., Dubovikov M. S.* Dimension of the minimal cover and fractal analysis of time series // Physica A: Statistical Mechanics and its Applications. — 2004. — Vol. 339, no. 3–4. — Pp. 591–608.
18. *Erramilli A., Narayan O., Willinger W.* Experimental queueing analysis with long-range dependent packet traffic // IEEE/ACM Transactions on Networking (TON). — 1996. — Vol. 4, no. 2. — Pp. 209–223.
19. Extension of the Kalman–Bucy Filter to Elementary Linear Systems with Fractional Brownian Noises / M. L. Kleptsyna, M. L. Kleptsyna, a. L. E. Breton, a. L. E. Breton // Statistical Inference for Stochastic Processes. — 2002. — Pp. 249–271.
20. *Fellouris G., Moustakides G. V.* Bandwidth and Energy Efficient Decentralized Sequential Change Detection // arXiv preprint arXiv:1210.2029. — 2013. — arXiv: `arXiv:1210.2029v2`.
21. *Futamura K., Liu D.* ANOMALY DETECTION METHODS FOR A COMPUTER NETWORK. — 2007.
22. *Garrett M. W., Willinger W.* Analysis, modeling and generation of self-similar VBR video traffic // ACM SIGCOMM Computer Communication Review. — 1994. — Vol. 24, no. 4. — Pp. 269–280.
23. *Genz A.* Numerical computation of multivariate normal probabilities // Journal of computational and graphical statistics. — 1992. — Vol. 1, no. 2. — Pp. 141–149.

24. *Girshick M. A., Rubin H.* A Bayes approach to a quality control model // The Annals of Mathematical Statistics. — 1952. — Pp. 114–125.
25. *Gripenberg G., Norros I.* On the prediction of fractional Brownian motion // Journal of Applied Probability. — 1996. — Pp. 400–410.
26. *Guépié B. K., Fillatre L., Nikiforov I.* Sequential Detection of Transient Changes // Sequential Analysis. — 2012. — Vol. 31, no. 4. — Pp. 528–547.
27. *Hassani H.* A brief introduction to singular spectrum analysis. — 2010. — URL: http://ssa.cf.ac.uk/a_brief_introduction_to_ssa.pdf (visited on 06/14/2016).
28. *Hassani H.* Singular spectrum analysis: methodology and comparison // Journal of Data Science. — 2007. — Vol. 5, no. 2. — Pp. 239–257.
29. *Hodrick R. J., Prescott E. C.* Postwar US business cycles: an empirical investigation // Journal of Money, credit, and Banking. — 1997. — Pp. 1–16.
30. *Huber P. J.* Robust statistics. — Springer, 2011.
31. *Hurst H. E.* Long-term storage capacity of reservoirs // Trans. Amer. Soc. Civil Eng. — 1951. — Vol. 116. — Pp. 770–808.
32. In-Network PCA and Anomaly Detection / L. Huang, X. Nguyen, M. Garofalakis, M. I. Jordan, A. Joseph, N. Taft // Advances in Neural Information Processing Systems 19. — 2007. — Vol. 19. — Pp. 617–624.
33. *ISO/IEC/IEEE* Systems and software engineering — Architecture description // ISO/IEC/IEEE 42010:2011(E) (Revision of ISO/IEC 42010:2007 and IEEE Std 1471-2000). — 2011. — Jan. — Pp. 1–46.
34. *Johnson J. B.* The Schottky effect in low frequency circuits // Physical review. — 1925. — Vol. 26, no. 1. — P. 71.

35. *Kirichenko L., Radivilova T., Deineko Z.* Comparative analysis for estimating of the Hurst exponent for stationary and nonstationary time series // Information Technologies & Knowledge. — 2011. — Vol. 5, no. 1. — Pp. 371–388.
36. *Kleptsyna M. L., Breton A. L., Viot M.* Asymptotically optimal filtering in linear systems with fractional Brownian noises // SORT. 2004, Vol. 28, Nüm. 2 [July-December]. — 2004. — Vol. 28, December. — Pp. 177–190.
37. *Kleptsyna M. L., Le Breton a.* Statistical Analysis of the Fractional Ornstein-Uhlenbeck Type Process // Statistical Inference for Stochastic Processes. — 2002. — Vol. 5, no. 3. — Pp. 229–248.
38. *Kleptsyna M. L., Le Breton A., Roubaud M.-C.* Parameter estimation and optimal filtering for fractional type stochastic systems // Statistical Inference for Stochastic Processes. — 2000. — Vol. 3, no. 1–2. — Pp. 173–182.
39. *Kleptsyna M., Breton a. L., Roubaud M. C.* General approach to filtering with fractional brownian noises - application to linear systems // Stochastics An International Journal of Probability and Stochastic Processes. — 2000. — Vol. 71, no. 1. — Pp. 119–140.
40. *Kleptsyna M. L., Kloeden P. E., Anh V.* Nonlinear filtering with fractional Brownian motion // Problemy Peredachi Informatsii. — 1998. — Vol. 34, no. 2. — Pp. 65–76.
41. *Kleptsyna M., Kloeden P., Anh V.* Linear filtering with fractional Brownian motion // Stochastic Analysis and Applications. — 1998. — Vol. 16, no. 5. — Pp. 907–914.
42. *Kobayashi M., Musha T.* 1/f fluctuation of heartbeat period // IEEE transactions on Biomedical Engineering. — 1982. — Vol. 6, BME-29. — Pp. 456–457.

43. *Krishnamoorthy K.* Handbook of statistical distributions with applications. — CRC Press, 2006.
44. *Lai T. L., Xing H.* Sequential Change-Point Detection When the Pre- and Post-Change Parameters are Unknown // Sequential Analysis. — 2010. — Vol. 29, no. 2. — Pp. 162–175.
45. *Lakhina A., Crovella M., Diot C.* Characterization of network-wide anomalies in traffic flows // Proceedings of the 4th ACM SIGCOMM conference on Internet measurement - IMC '04. — 2004. — Vol. 6. — P. 201.
46. *Lakhina A., Crovella M., Diot C.* Detecting distributed attacks using network-wide flow traffic // Proceedings of FloCon 2005 Analysis Workshop. — 2005.
47. *Lakhina A., Crovella M., Diot C.* Diagnosing network-wide traffic anomalies // ACM SIGCOMM Computer Communication Review. — 2004. — Vol. 34, no. 4. — P. 219.
48. *Le Breton A.* Filtering and parameter estimation in a simple linear system driven by a fractional Brownian motion // Statistics & probability letters. — 1998. — Vol. 38, no. 3. — Pp. 263–274.
49. *Lorden G.* Procedures for Reacting to a Change in Distribution. — 1971.
50. *Lucas J. M.* Counted Data CUSUM's // Technometrics. — 1985. — May. — Vol. 27, no. 2. — Pp. 129–144. — (Visited on 03/20/2015).
51. *Mandelbrot B. B., Van Ness J. W.* Fractional Brownian motions, fractional noises and applications // SIAM review. — 1968. — Vol. 10, no. 4. — Pp. 422–437.
52. *McMurry T. L., Politis D. N.* Minimally biased nonparametric regression and autoregression // REVSTAT–Statistical Journal. — 2008. — Vol. 6, no. 2. — Pp. 123–150.

53. *Mei Y.* Sequential change-point detection when unknown parameters are present in the pre-change distribution // *The Annals of Statistics*. — 2006. — Pp. 92–122.
54. *Moustakides G. V.* Optimal Stopping Times for Detecting Changes in Distributions // *The Annals of Statistics*. — 1986. — Vol. 14, no. 4. — Pp. 1379–1387.
55. *Moustakides G. V., Polunchenko A. S., Tartakovsky A. G.* A Numerical Approach to Performance Analysis of Quickest Change-Point Detection Procedures. — 2009. — arXiv: 0907.3521. — URL: <http://arxiv.org/abs/0907.3521>.
56. *Moustakides G. V., Polunchenko A. S., Tartakovsky A. G.* Numerical comparison of CUSUM and Shiryaev–Roberts procedures for detecting changes in distributions // *Communications in Statistics—Theory and Methods*. — 2009. — Vol. 38, no. 16–17. — Pp. 3225–3239. — arXiv: arXiv:0908.4119v1.
57. *Moustakides G.* Multiple Optimality Properties of the Shewhart Test // arXiv preprint arXiv:1401.3408. — 2014. — arXiv: arXiv:1401.3408v1. — URL: <http://arxiv.org/abs/1401.3408>.
58. *Nadaraya E. A.* On estimating regression // *Theory of Probability & Its Applications*. — 1964. — Vol. 9, no. 1. — Pp. 141–142.
59. New capabilities and methods of the X-12-ARIMA seasonal-adjustment program / D. F. Findley, B. C. Monsell, W. R. Bell, M. C. Otto, B.-C. Chen // *Journal of Business & Economic Statistics*. — 1998. — Vol. 16, no. 2. — Pp. 127–152.

60. *Nguyen X., Wainwright M. J., Jordan M. I.* Nonparametric decentralized detection using kernel methods // Signal Processing, IEEE Transactions on. — 2005. — Vol. 53, no. 11. — Pp. 4053–4066.
61. *Norros I., Valkeila E., Virtamo J.* An Elementary Approach to a Girsanov Formula and Other Analytical Results on Fractional Brownian Motions // Bernoulli. — 1999. — Aug. — Vol. 5, no. 4. — P. 571.
62. On the self-similar nature of Ethernet traffic (extended version) / W. E. Leland, M. S. Taqqu, W. Willinger, D. V. Wilson. — 1994.
63. Optimal volume anomaly detection and isolation in large-scale IP networks using coarse-grained measurements / P. Casas, S. Vaton, L. Fillatre, I. Nikiforov // Computer Networks. — 2010. — Vol. 54, no. 11. — Pp. 1750–1766.
64. *Page E.* Continuous inspection schemes // Biometrika. — 1954. — Vol. 41, no. 1. — Pp. 100–115.
65. *Peskir G., Shiryaev A.* Optimal Stopping and Free-Boundary Problems. — ETH Zürich : Springer, 2006.
66. *Pollak M.* Optimal detection of a change in distribution // The Annals of Statistics. — 1985. — Pp. 206–227.
67. *Prakasa Rao B. L. S.* Parametric estimation for linear stochastic differential equations driven by fractional Brownian motion // Random Operators and Stochastic Equations. — 2003. — Vol. 11, no. 3. — Pp. 229–242.
68. *Rasmussen C. E.* Gaussian processes for machine learning. — 2006.
69. *Ritov Y.* Decision theoretic optimality of the CUSUM procedure // The Annals of Statistics. — 1990. — Pp. 1464–1469.
70. *Roberts S.* A comparison of some control chart procedures // Technometrics. — 1966. — Vol. 8, no. 3. — Pp. 411–430.

71. *Ryu J.-H., Wan H., Kim S.* Optimal design of a CUSUM chart for a mean shift of unknown size // *Journal of Quality Technology*. — 2010. — Vol. 42, no. 3. — P. 311.
72. *Schapire R. E., Freund Y.* Boosting: Foundations and algorithms. — MIT press, 2012.
73. *Sen A., Srivastava M. S.* On tests for detecting change in mean // *The Annals of statistics*. — 1975. — Pp. 98–108.
74. *Sen A., Srivastava M. S.* Some one-sided tests for change in level // *Technometrics*. — 1975. — Vol. 17, no. 1. — Pp. 61–64.
75. *Shewhart W. A.* Economic control of quality of manufactured product. — 1931.
76. *Shiskin J., Young A. H., Musgrave J. C.* The X-11 variant of the census method II seasonal adjustment program. — US Department of Commerce, Bureau of the Census, 1965.
77. *SNDlib* Networks with multiple demand matrices. — URL: <http://sndlib.zib.de/dynamicmatrices.overview.action>.
78. STL: A seasonal-trend decomposition procedure based on loess / R. B. Cleveland, W. S. Cleveland, J. E. McRae, I. Terpenning // *Journal of Official Statistics*. — 1990. — Vol. 6, no. 1. — Pp. 3–73.
79. Stochastic analysis of the fractional Brownian motion / L. Decreusefond [et al.] // *Potential analysis*. — 1999. — Vol. 10, no. 2. — Pp. 177–214.
80. Stochastic calculus for fractional Brownian motion and applications / F. Biagini, Y. Hu, B. Øksendal, T. Zhang. — Springer Science & Business Media, 2008.
81. *Streit R. L., Willett P. K.* Detection of random transient signals via hyperparameter estimation // *IEEE Transactions on Signal Processing*. — 1999. — Vol. 47, no. 7. — Pp. 1823–1834.

82. *Tartakovsky A.* Efficient computer network anomaly detection by changepoint detection methods // IEEE Journal of Selected Topics in Signal Processing. — 2013. — Vol. 7, no. 1. — Pp. 7–11. — arXiv: [arXiv:1212.1829](https://arxiv.org/abs/1212.1829).
83. *Tartakovsky A. G., Polunchenko A. S.* Quickest changepoint detection in distributed multisensor systems under unknown parameters // Information Fusion, 2008 11th International Conference on. — IEEE. 2008. — Pp. 1–8.
84. The elements of statistical learning. Vol. 2 / T. Hastie, R. Tibshirani, J. Friedman, T. Hastie, J. Friedman, R. Tibshirani. — Springer, 2009.
85. Time variant power spectrum analysis for the detection of transient episodes in HRV signal / Cerutti S., Cerutti S., a. Bianchi, a. Bianchi, Minardi L., Minardi L., S. M., S. M., M. M., M. M. // IEEE Transactions on biomedical engineering. — 1993. — Vol. 40, no. 2. — Pp. 136–144.
86. Ultra-Large-Scale Systems: The Software Challenge of the Future — SEI Digital Library / L. Northrop, R. P. Gabriel, M. Klein, D. Schmidt. — URL: <http://resources.sei.cmu.edu/library/asset-view.cfm?assetID=30519>.
87. *Vautard R., Ghil M.* Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series // Physica D: Nonlinear Phenomena. — 1989. — Vol. 35, no. 3. — Pp. 395–424.
88. *Vautard R., Yiou P., Ghil M.* Singular-spectrum analysis: A toolkit for short, noisy chaotic signals // Physica D: Nonlinear Phenomena. — 1992. — Vol. 58, no. 1. — Pp. 95–126.
89. *Voss R. F., Clarke J.* "1/f noise" in music: Music from 1/f noise // The Journal of the Acoustical Society of America. — 1978. — Vol. 63, no. 1. — Pp. 258–263.

90. *Watson G. S.* Smooth regression analysis // *Sankhya: The Indian Journal of Statistics, Series A.* — 1964. — Pp. 359–372.
91. *Winters P. R.* Forecasting Sales by Exponentially Weighted Moving Averages // *Management Science.* — 1960. — Vol. 6, no. 3. — Pp. 324–342.
92. *Yaozhong H., Weilin X., Weiguo Z.* Exact maximum likelihood estimators for drift fractional Brownian motions // *Acta Mathematica Scientia.* — 2011. — Vol. 31, no. 5. — Pp. 1851–1859. — arXiv: 0904.4186. — URL: <http://arxiv.org/abs/0904.4186>.
93. *Zhang Y.* Abilene Dataset. — URL: <http://www.cs.utexas.edu/~yzhang/research/AbileneTM>.
94. *Байдасов М. И.* Анализ шума в задаче об оптимальной остановке гидродинамического эксперимента: дис. ... маг. / Байдасов М. И. — М., 2013. — 34 с.
95. *Бурнаев Е. В.* О минимаксной и обобщенной байесовской задачах скорейшего обнаружения разладки для пуассоновского процесса [Текст] : дис. ... канд. физ.-мат. наук : 05.13.17 : защищена 22.01.02 : утв. 15.07.02 / Бурнаев Евгений Владимирович. — М., 2008. — 127 с.
96. *Вальд А.* Последовательный анализ. — 1960.
97. *Климонтович Ю. Л.* Естественный фликер шум ("шум $1/f$ ") и сверхпроводимость // *Письма в ЖЭТФ.* — 1990. — Т. 51, № 1. — С. 43–45.
98. *Колмогоров А. Н.* Спираль Винера и некоторые другие интересные кривые в гильбертовом пространстве // *ДАН СССР.* — 1940. — Т. 26. — С. 115–118.

99. *Колмогоров А. Н., Прохоров Ю. В., Ширяев А. Н.* Вероятностно-статистические методы обнаружения спонтанно возникающих эффектов // Труды Математического института им. В. А. Стеклова. — 1988. — Т. 182. — С. 4—23.
100. *Липцер Р., Ширяев А.* Статистика случайных процессов: Нелинейная фильтрация и смежные вопросы. — М. : Наука. Главная редакция физико-математической литературы, 1974.
101. *Липцер Р., Ширяев А.* Теория мартингалов. — М. : Наука. Главная редакция физико-математической литературы, 1986.
102. Об утверждении Стратегии развития отрасли информационных технологий в Российской Федерации на 2014–2020 годы и на перспективу до 2025 года. — 2013. — URL: <http://government.ru/docs/8024/> (дата обр. 06.05.2016) ; Проверен: 07.05.2016.
103. *Синай Я. Г.* Автомодельные распределения вероятностей // Теория вероятностей и ее применения. — 1976. — Т. 21, № 1. — С. 63—80.
104. *Ширяев А. Н.* Вероятностно-статистические методы в теории принятия решений. — Москва : МЦНМО, 2011.
105. *Ширяев А. Н.* Минимаксная оптимальность метода кумулятивных сумм (CUSUM) в случае непрерывного времени // Успехи математических наук. — 1996. — Т. 51, № 4. — С. 173—174.
106. *Ширяев А. Н.* Статистический последовательный анализ: оптимальные правила остановки. — 2-е, перер. — Москва : Наука, Физматлит, 1976.
107. *Ширяев А. Н.* Об оптимальных методах в задачах скорейшего обнаружения // Теория вероятностей и ее применения. — 1963. — Т. 8, № 1. — С. 26—51.

108. *Ширяев А.* Задача скорейшего обнаружения нарушения стационарного режима // Докл. АН СССР. Т. 138. — 1961. — С. 1039—1042.

Условные обозначения

Обозначения, принятые в диссертационной работе

Ниже представлены условные обозначения, общие для всей диссертационной работы.

Символ	Значение	Размерность
t, s, u	различные моменты времени	скаляр
ξ	сигнал-результат измерения	скаляр
X_t	значение результата измерения в момент t	скаляр
$f(t)$	тренд исследуемого сигнала	скаляр
ν_t	шум	скаляр
θ	оцениваемый параметр	скаляр
$\boldsymbol{\theta}$	вектор оцениваемых параметров	вектор, n_θ
$a(t)$	неслучайная функция t (вместо a может быть любая латинская буква)	скаляр
$\varphi_t = \varphi(t)$	случайный процесс (вместо φ может быть любая греческая буква)	скаляр
B^H	фрактальное броуновское движение	скаляр
H	параметр Хёрста фрактального броуновского движения	скаляр
\mathbf{X}^ℓ	обучающее множество пар «данные–разметка»	ℓ
Π	процедура обнаружения разладки	
\mathcal{A}	ансамбль процедур обнаружения разладки	
τ	момент подачи тревоги при обнаружении разладки	скаляр

Таблица 5.1: Общие для всей диссертационной работы условные обозначения

Общематематические обозначения

Символ	Значение	Размерность, тип
$P, P(\cdot)$	вероятность	
E	символ математического ожидания	
\mathcal{F}	σ -алгебра	
\mathbb{R}	множество действительных чисел	
\mathbb{R}^n	n -мерное евклидово пространство	

Таблица 5.2: Общематематические условные обозначения

Список рисунков

1.1	Примеры реализаций фрактального броуновского движения с различными значениями показателя H	19
1.2	Траектории результата наблюдения ξ_t , тренда $f(=) \sum_{k=0}^3 \theta_i t^i$ и фильтра $\hat{f}(t) = \sum_{k=0}^3 (\hat{\theta}_{\text{BAYES}})_i t^i$, $0 \leq t \leq T$, в модельной задаче выделения кубического тренда при значении параметра $H = 0.8$	30
1.3	Значения функции штрафа $F_H(t)$ в модельной задаче выделения полиномиального (кубического) тренда при значениях параметров $H = 0.8, c = 0.02$	30
2.1	Сравнительный анализ оперативных характеристик «слабых» детекторов и ансамблей для набора данных Cauchy	51
2.2	Сравнительный анализ оперативных характеристик «слабых» детекторов и ансамблей для набора данных Fractal	52
2.3	Сравнительный анализ эффективности «слабых» детекторов и ансамблей для набора данных Fractal.	55
3.1	Зависимость среднеквадратичной погрешности оценивания коэффициента линейного сноса фрактального броуновского движения от значения \hat{H} параметра Херста, предполагаемого в оценке 1.3	61
3.2	Результат применения алгоритма оценивания параметра тренда с поправкой на длинную память. Серым цветом показана траектория результата наблюдений ξ в (3.4), синим — истинная траектория тренда $f(t)$, зеленым — результат оценивания $\hat{f}(t)$ без поправки на длинную память ($H = 1/2$), розовым — результат оценивания $\hat{f}(t)$ с поправки на длинную память ($H = 0.11$). Истинное значение $H = 0.1$	61
3.3	Косинусное ядро и его Фурье-образ	67

3.4	<i>Верхний график:</i> месячная траектория числа запросов, заданных некоторому интернет-сервису Яндекса, с отмеченными несколькими краткосрочными разладками. <i>Нижний график:</i> месячная траектория усредненного времени ответа, этого интернет-сервиса, с отмеченными двумя долгосрочными разладками. Показаны нормализованные на единицу значения.	69
3.5	Пример данных из набора ARTIFICIAL-EASY и результат выделения тренда, полученный методом EWMA и предложенным в диссертационной работе подходом. Отмечены сгенерированные разладки. .	75
3.6	Эмпирическое сравнение эффективности обнаружения разладки в данных ARTIFICIAL-EASY для процедур на основе EWMA, процедур на основе PCA, и алгоритма раздела 3.3. Слева: кривые «точность–полнота». Справа: кривые средней относительной точности сегментации, см. раздел 2.4.	76
3.7	Эмпирическое сравнение эффективности обнаружения разладки в данных ARTIFICIAL-HARD для процедур на основе EWMA, процедур на основе PCA, и алгоритма раздела 3.3. Слева: кривые «точность–полнота». Справа: кривые средней относительной точности сегментации, см. раздел 2.4.	76
3.8	Сравнение траекторий статистик, используемых в процедурах CUSUM, PCA-PRETRAINED и ансамбле LOG-0 для данных ARTIFICIAL-HARD. Верхний рис.: траектория искусственных данных и индикатор разладки. Второй сверху рис.: траектория статистики процедуры CUSUM и индикатор разладки. Заметим отсутствие корреляции с индикатором разладки. Третий сверху рис.: траектория статистики процедуры PCA-PRETRAINED и индикатор разладки. Заметим слабую корреляцию с индикатором разладки. Нижний рис.: траектория статистики процедуры LOG-0 и индикатор разладки. Заметим высокую корреляцию с индикатором разладки. Все статистики были отмасштабированы для удобства просмотра.	77
4.1	Принципиальная схема программного модуля процессинга данных с трендом	80

4.2	Структура основных составляющих реализованного программного комплекса	81
4.3	Иерархия классов пакета статистических процедур	82
4.4	Иерархия классов пакета моделирования реализаций случайных сигналов и пакет оценивания эффективности исследуемых алгоритмов и визуализации данных	83
5.1	Значения индекса S&P 500, использованные при решении задачи прогнозирования значения этого индекса на один день вперед . . .	88
5.2	Значения индекса S&P 500 (серая линия) и значения прогноза этого индекса на один день вперед (зеленая линия)	88
5.3	Примеры недельных временных рядов, отвечающих нагрузкам соединений Атланта–Лос Анджелес, Лос Анджелес–Чикаго, Сиэтл–Лос Анджелес, Чикаго–Лос Анджелес за неделю 14–21 июня 2004 г. в сети Абилин (2016 измерений)	89
5.4	Оценивание квазипериодического тренда нагрузки соединения Хьюстон–Чикаго	89
5.5	Результаты разложения временного ряда с использованием алгоритма раздела 3.4.2 для трех различных наборов данных: описанного в разделе 3.4 искусственного набора (левый столбец), описываемого в настоящем разделе набора данных Яндекс (центральный столбец), и набора данных Абилин (правый столбец). На графиках представлены: исходный временной ряд ξ , и результат оценивания тренда $\hat{f}(t)$ с доверительными интервалами $\hat{f}(t) \pm \hat{\sigma}_{\varphi(t)}$ (верхний рисунок), оценка $\hat{S}_{\varphi(t)}$ сезонной компоненты $S_{\varphi(t)}$ с доверительными интервалами $\hat{S}_{\varphi(t)} \pm \hat{\sigma}_{\varphi(t)}$ (второй сверху рисунок), оценка \hat{Q}_t амплитуды Q_t сезонного профиля (третий сверху рисунок), оценка $\hat{\nu}_t$ случайной помехи ν_t (нижний рисунок).	91

- 5.6 Результаты обнаружения разладки поисковой системы Яндекс для случаев краткосрочных изменений с характерной длительностью много меньше периода (левый столбец) и долгосрочных изменений масштаба (центральный столбец). Для сравнения типов рассматриваемых данных также представлены результаты обнаружения разладки по данным Абилин (правый столбец, ср. раздел 5.3). На графиках представлены: исходный временной ряд ξ (верхний рисунок), выделяемая компонента, анализируемая на предмет разладки (второй сверху рисунок): остатки R , вычисляемые согласно (3.17), для случая краткосрочных разладок в данных Яндекса и в данных Абилин, и амплитуда A , вычисляемая согласно алгоритму раздела 3.4.2, для случая долгосрочных разладок в данных Яндекса, и траектория статистики кумулятивных сумм (нижний рисунок). 92
- 5.7 Результаты решения задачи обнаружения изменения режима турбулентного течения, соответствующего изменению дисперсии процесса относительного давления. Слева представлены реализации процесса относительного давления, измеренного в двух различных точках объема физической модели. Справа представлены реализации процесса кумулятивных сумм. 95
- 5.8 Результаты решения задачи обнаружения изменения режима турбулентного течения, соответствующего изменению спектра процесса относительного давления. Слева представлены реализации процесса относительного давления, измеренного в двух различных точках объема физической модели. Справа представлены реализации процесса кумулятивных сумм. 97
- 5.9 Оперативные характеристики решения задачи обнаружения изменения режима турбулентного течения: зависимость среднего времени запаздывания в обнаружении разладки от вероятности ложной тревоги для различных положений датчиков давления. l — расстояние от датчика до центра симметрии системы, r — характерный радиус системы. Левый рис.: изменение режима течения соответствует изменению дисперсии процесса относительного давления. Правый рис.: изменение режима течения соответствует изменению спектра процесса относительного давления. 99

Список таблиц

2.1	Характеристики использованных при сравнительном анализе искусственных наборов данных	49
2.2	Характеристики использованных при сравнительном анализе процедур обнаружения разладки	50
2.3	Сравнительный анализ эффективности «слабых» детекторов и ансамблей для всех наборов данных в терминах площади под кривой «точность–полнота». Полужирным шрифтом выделен лучший результат для каждого набора данных.	54
3.1	Точность выделения тренда для искусственного набора данных в терминах относительной среднеквадратичной погрешности для процедуры EWMA, вариантов процедур на основе анализа главных компонент (PCA и PCA-PRETRAINED) и рассматриваемого в диссертационной работе подхода.	73
5.1	Общие для всей диссертационной работы условные обозначения . .	116
5.2	Общематематические условные обозначения	117