Российский государственный геологоразведочный университет им. Серго Орджоникидзе

На правах рукописи

Яковлев Павел Викторович

Анализ пространственно-временных особенностей временных рядов GPS для выделения областей интенсивных движений земной коры

Специальность 25.00.10 – Геофизика, геофизические методы поисков полезных ископаемых

Диссертация на соискание ученой степени
кандидата физико-математических наук

Научный руководитель:
доктор физико-математических наук
Любушин А.А.

Москва 2016

Оглавление

Введение

В предлагаемой диссертации рассматриваются вопросы анализа временных рядов мониторинга движений земной поверхности по данным спутниковой геодезии (GPS). Разработаны методы оценки характеристик временных рядов и построения карт их статистических свойств, полученных от сетей наблюдений, насчитывающих несколько тысяч пунктов, расположенных в различных частях мира. Исследуемые данные свободно доступны в сети интернет. Целью анализа рассматриваемых временных рядов является выделение аномальных зон на поверхности Земли, характеризуемых повышенной интенсивностью хаотических движений малых блоков земной коры. Подобного рода движения занимают промежуточную нишу в общем процессе диссипации тектонической энергии в коре между общими медленными движениями плит и быстрыми подвижками между блоками земной коры в результате сейсмического процесса. Следует подчеркнуть, что изучаемый тип движений земной поверхности не поддается регистрации обычными сейсмическими датчиками и одновременно не может быть выделен с помощью традиционного анализа данных спутниковой геодезии, целью которого является сглаживание временных рядов и выделение трендов, показывающих направления движения тектонических плит.

Анализ временных рядов всегда был более чем актуальным вопросом, а с увеличением компьютерных мощностей в этом направлении перед исследователями открываются новые возможности, что лишь повышает интерес к изучению такого статистического материала, как временной ряд. В настоящее время временные ряды можно встретить повсеместно. Связано это с тем, что во многих сферах деятельности человечество научилось следить за различного рода динамическими системами путем создания специальных систем мониторинга. Такие системы мониторинга можно увидеть в геофизике, экономике, сейсмологии, социологии, метеорологии и др. В каких-то сферах человеческой деятельности наблюдения ведутся уже относительно давно, в каких-то недавно, но везде данные собираются с целью изучения природы наблюдаемой динамической системы, выявления структуры временных рядов и их возможного прогнозирования.

Спутниковая система навигации GPS сама по себе появилась относительно давно, еще в 1980-х годах. Несмотря на то, что изначально GPS создавался как инструмент для военных целей, со временем его использование стало доступно для гражданских применений по всему миру. Благодаря высокой точности измерений с помощью спутниковой системы навигации, стало возможным отслеживать малейшие перемещения любых объектов на земной поверхности, в том числе и литосферным плит, что привнесло огромный вклад в такую геологическую теорию как тектоника плит.

Актуальность темы диссертации

Актуальность определяется современным процессом резкого увеличения объема геофизических данных, свободно доступных для исследования. Эта тенденция особенно сильно проявляет себя для данных спутниковой геодезии. В частности, в настоящее время доступны ежесуточные трехкомпонентные временные ряды GPS от почти 11 тысяч пунктов наблюдения, расположенных по всему миру, причем с середины 2013 года они доступны также с шагом по времени 5 минут. Этот объем данных дает геофизике новые возможности для исследования тонкой структуры движений земной коры и, в частности, таких скрытых процессов как «тихие землетрясения».

В данной диссертации основное внимание уделяется анализу нерегулярной составляющей временных рядов, которая при традиционном анализе называется «шумом» и обычно подавляется с помощью операций сглаживания. Хаотический характер изучаемой составляющей сигналов определяет набор статистических инструментов, разработанных и примененных для ее изучения – это исследование выбросов и скачков, а также фрактальных и мультифрактальных характеристик.

Цели и задачи диссертации

Целью данной диссертационной работы является разработка методов и алгоритмов для анализа нерегулярной (шумовой) составляющей временных рядов спутниковой геодезии для выделения на поверхности Земли регионов, характеризуемых повышенной интенсивностью движений малых блоков земной коры. Для достижения этой цели были решены следующие задачи:

1. Разработан энтропийный метод анализа скачкообразной составляющей временных рядов, основанный на использовании введенного автором понятия псевдо-производной.
2. Разработан метод выделения аномальных регионов, основанный на использовании меры интенсивности выбросов во временных рядах.
3. Разработаны методы построения карт распределения по пространству мульти-фрактальных параметров временных рядов GPS на основе анализа данных, содержащих пропуски регистрации.
4. Разработаны методы построения карт распределения максимального собственного числа корреляционной матрицы и спектральной экспоненты шума от заданного числа ближайших станций.

Научная новизна

В настоящей работе вводится такое понятие как псевдо-производная, имеющее определенный ряд свойств, сходных с классическим определением производной. Предлагается конкурентно-способный метод построения ступенчатой аппроксимации временных рядов с использованием псевдо-производной. Преимуществом данного метода является предельная простота реализации, из которой следует высокая скорость вычислений для случая обработки большого числа временных рядов. Рассматриваются также различные модификации метода.

Предлагается метод оценки скачкообразности сигналов, основанный на использовании их кусочно-ступенчатой аппроксимации. Под скачкообразностью здесь подразумевается резкое изменение среднего значения. Помимо скачков, в работе приведено выделение таких особенностей как выбросы с помощью метода взвешенной суммы приращений стандартных отклонений. И скачки, и выбросы не рассматриваются в работе, как случайные неинформативные события, а наоборот, предполагается, что они являются одними из важнейших составляющих сигнала, которые могут нести в себе важную информацию.

Предлагается новый метод оценки сейсмической опасности, основанный на построении карт статистических свойств случайных флуктуаций временных рядов GPS. Главной идеей этого подхода является предположение, численно доказанное в работах Любушина А.А. [25], [26], [27], [28], что перед катастрофой происходит синхронизация шумовых смещений земной поверхности, что является индикатором консолидации блоков земной коры и признаком повышенной сейсмической активности. Области консолидации малых блоков земной коры интерпретируются как зоны повышенных значений максимального собственного числа корреляционной матрицы. Фактически используемая методика заключается в применении метода главных компонент к анализу шумов временных рядов, регистрируемых синхронно на большом числе GPS станций, покрывающих значительную область поверхности Земли.

Методы исследования

В диссертации использовались методы статистического анализа, метод главных компонент, метод псевдо-производных и взвешенной суммы приращений стандартных отклонений, мультифрактальный анализ.

Публикации

Материалы диссертации были опубликованы в восьми печатных изданиях, четыре из которых являются тезисами докладов на конференциях, а остальные четыре в журналах, включенных ВАК РФ в перечень ведущих рецензируемых изданий.

Апробация работы

Основные промежуточные результаты исследований, приведенные в диссертации, докладывались на следующих конференциях:

- XI Международная научно-практическая конференция «Новые идеи в науках о Земле», Москва 2013
- VII Международная конференция студентов, аспирантов и молодых ученых «Молодые – наукам о Земле», Москва, 2014
- XII Международная научно-практическая конференция «Новые идеи в науках о Земле», Москва, 2014
- Проблемный Совет «Сейсмичность Земли, природные и природно-техногенные катастрофы», Москва, 2014
- IUGG 2015 General Assembly, Prague, 2015
- Конференция молодых ученых и аспирантов ИФЗ РАН-2016, Москва, 2016.

Личный вклад автора

Автором выполнены и написаны все расчеты и программы. Он ввел понятие псевдо-производной и предложил основную идею построения кусочно-ступенчатой аппроксимации временных рядов. Метод взвешенной суммы приращений стандартных отклонений для выделения выбросов в сигналах также разработан автором.

Практическая значимость

В диссертации созданы алгоритмы анализа большого числа (нескольких тысяч) временных рядов систем космической геодезии, включающие в себя автоматический контроль качества данных и учет пропусков регистрации, позволяющие визуализировать пространственно-временную картину эволюции ряда информативных свойств поля смещений земной поверхности в виде последовательности карт, оцениваемых в скользящих временных окнах. Эта последовательность карт может быть использована в задачах оценки текущей сейсмической опасности и прогноза землетрясений.

Основные научные результаты, выносимые на защиту

- Метод построения кусочно-ступенчатой аппроксимации временных рядов, основанный на определении псевдо-производной и ее свойствах
- Метод взвешенной суммы приращений стандартных отклонений для выделения выбросов во временных рядах
- Методы выделения аномальных и сейсмически опасных регионов, а также зон интенсивных движений земной коры, основанные на оценке мультифрактальных и статистических параметров временных рядов GPS, выбросов и скачкообразной компоненты

Достоверность научных исследований

Достоверность и обоснованность полученных научных результатов гарантируется строгостью используемого математического аппарата и подтверждается сопоставлением результатов анализа временных рядов с сейсмическим режимом и особенностями геологического строения исследуемых регионов.

Структура и объем диссертации

Диссертация состоит из 3 глав. В первой главе рассмотрены основные задачи и направления анализа временных рядов. Сказано о важности выделения во временных рядах таких аномалий как выбросы и скачки, так как они могут содержать в себе полезную информацию. Кратко изложен вводный материал по мультифракталам, а также по методам вычисления мультифрактальных параметров временных рядов.

Во второй главе предложены три метода анализа временных рядов, в результате выдающих важные статистические характеристики, которые будут использованы для построения карт их пространственного распределения. Введены два новых математических инструмента, таких как псевдо-производная и взвешенная сумма приращений стандартных отклонений. Первый изобретен для оценки скачкообразной составляющей временных рядов, а второй – для определения меры выбросов.

В третьей главе представлены результаты анализа в виде карт пространственного распределения различных статистических величин для двух регионов планеты: острова Японии и западная часть Соединенных Штатов Америки. Произведена интерпретация построенных карт следующих величин:

1. Нормализованной энтропии скачкообразной компоненты сигналов GPS
2. Нормализованной энтропии меры выбросов (взвешенной суммы приращений стандартных отклонений
3. Спектральной экспоненты
4. Отношения максимального собственного значения корреляционной матрицы к сумме всех собственных значений
5. Обобщенного показателя Херста
6. Ширины спектра сингулярности
7. Корреляционной размерности сигналов GPS

8. Логарифма дисперсии шума

Благодарности

Автор искренне благодарен и признателен своему научному руководителю Любушину Алексею Александровичу за помощь на всех этапах выполнения диссертации, публикации статей и за правильную и точную постановку задач, вдохновивших на изобретение нового математического инструмента.

Глава 1. Некоторые методы анализа и статистики временных рядов

1.1. Методы анализа временных рядов

В настоящее время проводится мониторинг самых разных динамических систем, например, таких как сейсмические сигналы, GPS сигналы, изменения котировок на биржевых ранках и т.д. Как правило, результатом мониторинга является дискретный сигнал, называемый временным рядом. Основной задачей при анализе таких временных данных является составление прогноза о дальнейшем поведении системы с целью предсказания появления возможной аномалии (катастрофы).

Такие системы как Глобальная Навигационная Спутниковая Система (ГЛОНАСС) и Global Positioning System (GPS) являются самыми удобными и точными инструментами для отслеживания различных деформаций земной поверхности. В настоящей работе в качестве анализируемых данных выступают именно сигналы GPS.

Существует множество различных направлений в анализе временных рядов, начиная от классических и заканчивая наиболее активно развивающихся по сей день:

1. Спектральный анализ
2. Корреляционно-регрессионный анализ
3. Спектрально-сингулярный анализ
4. Авторегрессионные модели, модели скользящего среднего
5. Мультифрактальный анализ
6. Анализ с точки зрения динамических систем и т.д.

В теории обработки временных рядов первые два подхода имеют достаточно узкий, хотя и полезный, спектр возможностей, так как спектральный анализ в основном сводится, как правило, к анализу распределения энергии сигнала по частотам, а корреляционный анализ – к определению степени связи между двумя или более наблюдаемыми величинами. Оба этих подхода являются достаточно популярными в силу их простоты и широкого распространения в области анализа различных данных. Практически все известные методы, относящиеся к спектральному и корреляционно-регрессионному анализу, также реализованы и оптимизированы во многих программных пакетах, что позволяет с легкостью комбинировать их с другими подходами.

В геофизике также одним из наиболее популярных математических инструментов для исследования геофизических сигналов является спектральный анализ. В качестве геофизических сигналов могут браться сигналы, порожденные различными полями (магнитными, волновыми, электрическими и т.д.). Спектральный анализ позволяет решить одну из фундаментальных задач исследования сигналов: отделить полезную информативную компоненту от помех, что в дальнейшем значительно упрощает процесс интерпретации полученных данных.

Написано огромное количество трудов по спектрально-сингулярному анализу, основанному на методе главных компонент [9], [10], [51]. Данный метод широко используется для решения разного рода задач: выделения трендов, прогнозирование, заполнение пропущенных данных, в том числе в данных геофизического мониторинга [71] и т.д.

Авторегрессионные модели и модели скользящего среднего основаны на определении параметров моделей для воссоздания схожего по структуре временного ряда. Такие подходы заключаются в простой идее: если мы научимся воссоздавать сигналы, то мы научимся прогнозировать дальнейшее развитие наблюдаемой системы. Однако на практике все намного сложнее. Практически всегда нахождение коэффициентов моделей не является достаточным условием для качественного и точного прогноза, так как на реальные динамические системы действует огромное количество факторов, большее число из которых модели не учитывают, тем самым данные подходы позволяют в основном строить более или менее точные краткосрочные прогнозы [17].

Исследование фрактальных свойств временных рядов, основываясь на предположении, что анализируемые сигналы являются самоподобными структурами, является наиболее перспективным и популярным, так как с помощью данного формализма возможно дать количественную характеристику сложности геометрии исследуемых сигналов – фрактальную размерность. Данное направление приобретает все большую популярность в связи с тем, что использование лишь детерминистических законов не всегда хорошо отражает действительное поведение системы. Это случается, когда изучаемый сигнал сильно зашумлен, и некоторые его случайные отклонения могут оказаться настолько значительными, что пренебрегать ими не является правильным решением. В таких случаях говорят о детерминированном хаосе, для изучения которых используются моно- и мультифракталы, позволяющие помимо глобальных численных характеристик оценить также и локальную структуру, и ее особенности. Рассмотрим данное направление исследования временных рядов подробнее.

1.2. Мультифрактальный анализ

Существует несколько определений фракталов – более или менее строгих. Одним из наиболее распространенных менее строгих определений является следующее. Фрактал – это множество с полным или частичным самоподобием (свойством скейлинга) и имеющее дробную размерность. Согласно более строгому, и в то же время более правильному, определению, фракталом называют объекты, фрактальная размерность которых меньше или равна их топологической размерности, и в общем случае, фрактальная размерность вовсе не обязательно дробная. Примерами фрактальных множеств с целой фрактальной размерностью являются кривая Пеано и так называемая кривая дракона – их фрактальные размерности равны 2.

Для описания мультифрактала не достаточно использование одного значения фрактальной размерности, так как в мультифракталах присутствует более одного показателя скейлинга. Одним из самых простых и показательных примеров мультифрактала является мультипликативный процесс, демонстрирующий необходимость введения обобщения фрактальной размерности, так как обычной фрактальной размерности не хватает для описания его свойств.

Чтобы построить данное множество, возьмем единичный отрезок с мерой 1. Далее разобьем его на две равные части и присвоим левой и правой половинам меры p и $1-p$ соответственно. Затем проделаем ту же операцию для каждой из половинок. В итоге получим следующую картину:

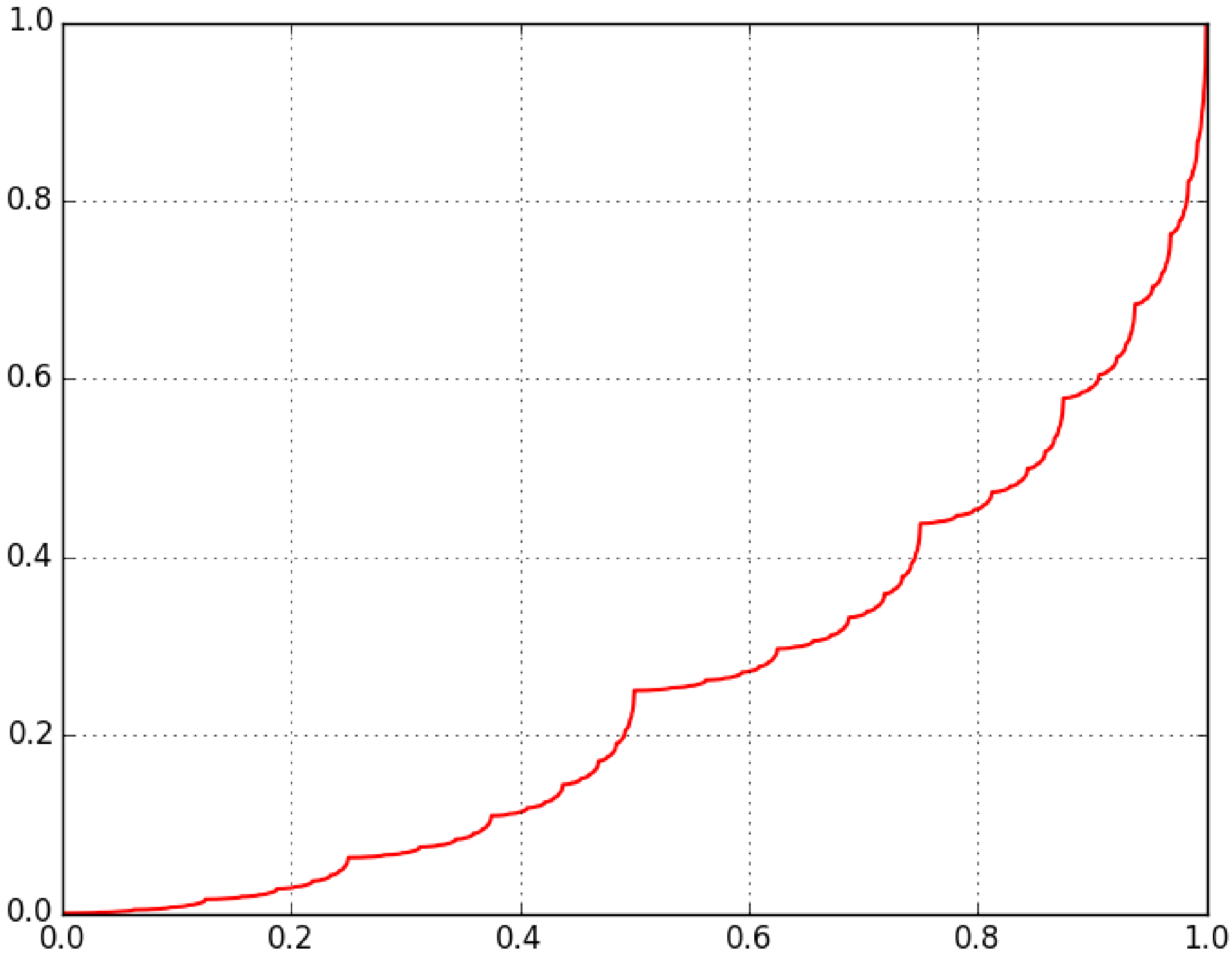


Рисунок 1.2.1. Пример мультипликативного биномиального процесса, $p = 0.25$

Распишем подробнее, что из себя представляет данное множество [30]. Пусть n – номер итерации, тогда на n-й итерации единичный отрезок делится $N = 2^n$ маленьких отрезков. Обозначим

$$t_i^n = \frac{i}{2^n}, \qquad i = 0, \dots, N-1 \tag{1.2.1}$$

границы получившихся малых интервалов. Припишем каждому отрезку меру $\mu_i^n = p^{k_i} \cdot (1-p)^{n-k_i}$, где k_i – число нулей в первых n знаках двоичного разложения числа t_i^n. Таким образом, можно определить меру интервала $[0, t)$:

$$\mu_n(t) = \int_0^t \rho_n(x) dx \tag{1.2.2}$$

Число точек t_i^k, имеющих в первых k знаках своего двоичного разложения одинаковое количество нулей совпадает с биномиальным (отсюда и название процесса) коэффициентом $\binom{n}{k}$. Стоит отметить, что количество интервалов с мерой $p^k \cdot (1-p)^{n-k}$ равно этому же числу. Это связано с тем, что нули двоичного представления чисел t_i^k соответствуют левым половинам интервалов, а единицы – правым. В итоге получается множество, растянутой по горизонтали и вертикали в 2 и $1/(1-p)$ раз соответственно. Множества, части которых соответствуют исходному с точностью до коэффициента подобия, называют самоафинными. В данном случае множители $1/(1-p)$ и 2 являются коэффициентами подобия.

Если просуммировать меры μ_i^n всех отрезков, то получится $\mu_n(t) = 1$ – меру отрезка. Для функции $\mu_n(t)$ выполняется следующее равенство:

$$\mu_n(t) = \begin{cases} p \cdot \mu_n(2t), & 0 \le t \le 0.5 \\ p + (1-p) \cdot \mu_n(2t-1), & 0.5 < t \le 1 \end{cases} \tag{1.2.3}$$

Вычислим размерность Хаусдорфа получившегося фрактального объекта, как предел отношения $\ln N$ и $\ln \frac{1}{r_n}$, где r_n – длина отрезка на итерации под номер n. Получаем:

$$D = \frac{\ln N}{\ln \frac{1}{r}} = \frac{\ln 2^n}{\ln \frac{1}{2^{-n}}} = \frac{\ln 2^n}{\ln 2^n} = 1 \qquad (1.2.4)$$

Таким образом, фрактальная размерность Хаусдорфа не позволит никак охарактеризовать множество, порожденное биномиальным процессом. В таких случаях прибегают к мультифрактальному спектру, позволяющему оценить подмножества исходного множества, обладающими одним и тем же свойством – в данном случае подынтервалы с одной и той же вероятностью $p^k \cdot (1-p)^{n-k}$. Мультифрактальный спектр для мультипликативного процесса с $p = 0.25$ выглядит следующим образом:

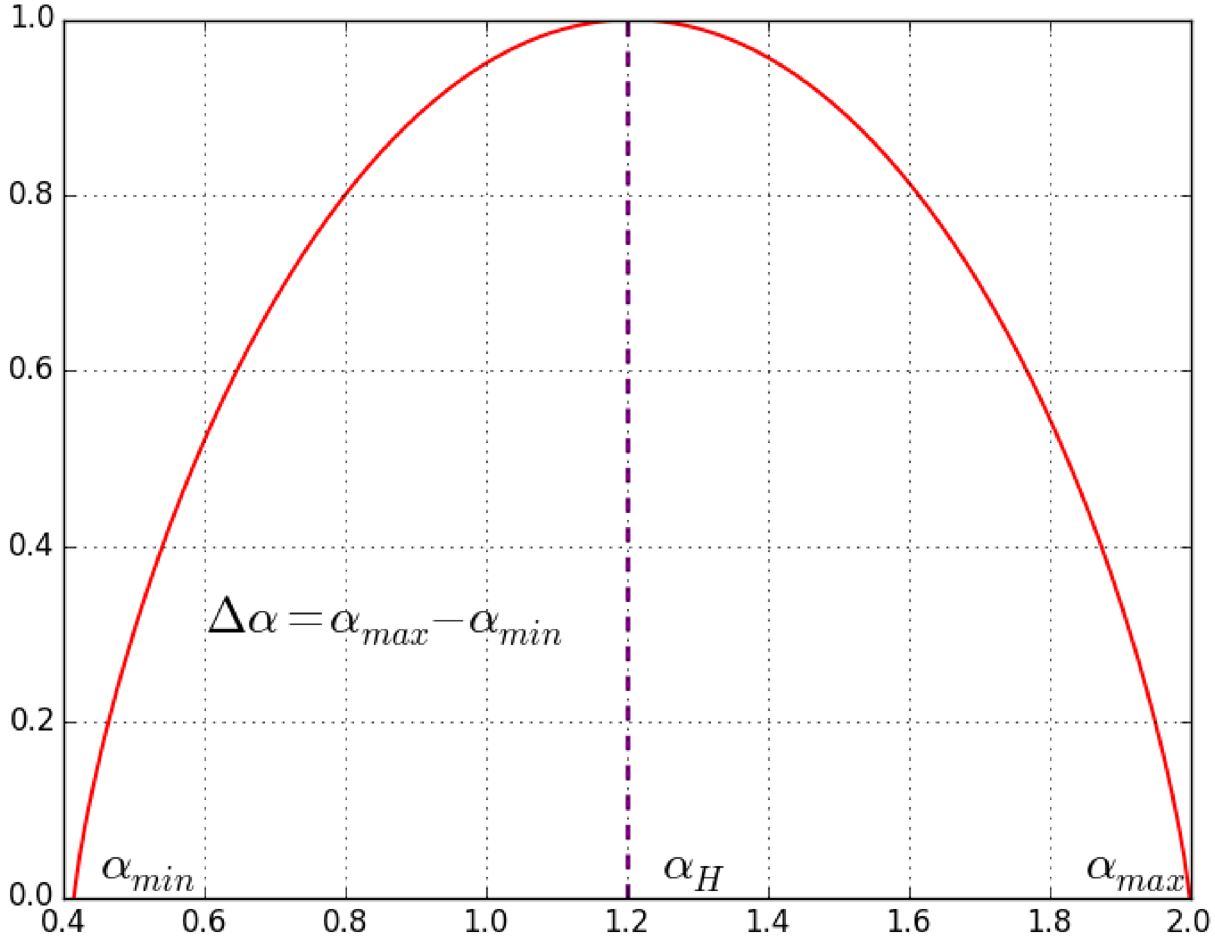


Рисунок 1.2.2. Мультифрактальный спектр мультипликативного процесса, $p = 0.25$

В силу того, что исследуемые временные ряды GPS также являются мультифракталами, то опишем, как на практике обычно находится мультифрактальный спектр, называемый также спектром сингулярности. Существует несколько способов вычисления данного спектра: с использованием непрерывного вейвлет-преобразования (метод ММВП, он же WTMM) или так называемого флуктуационного анализа с предварительным удалением тренда (MFDFA – multifractal detrended fluctuation analysis).

Величина спектра сингулярности имеет очень важное прикладное значение. Сам по себе спектр сингулярности есть фрактальная размерность точек (касаемо предмета анализа настоящей работы) анализируемого сигнала или множества, которые имеют один и тот же показатель (экспоненту) Гельдера-Липшица, характеризующий локальное поведение функции в заданной точке t_0 [37]. В случае, если функция $g(t)$ является n раз дифференцируемой в точке t_0 и не имеет производную порядка $n+1$, аппроксимируем ее полиномом $p_n(t)$ степени n путем разложения в ряд Тейлора в окрестности этой точки, тогда показатель Гельдера-Липшица в точке t_0 есть максимальное значение α, для которого справедливо

$$|g(t)-p_n(t-t_0)| \leq c|t-t_0|^{\alpha}, \qquad (1.2.5)$$

где k – некоторая константа. Таким образом, показатель α – показатель гладкости функции. Чем он выше, тем более гладкой является функция. Помимо гладкости показатель Гельдера-Липшица также указывает на масштабируемость анализируемого сигнала. Чем шире спектр значений гельдеровских экспонент, тем сложнее считается сигнал. Зачастую при анализе временных рядов данная сложность характеризует здоровое состояние наблюдаемой системы. Например, при изучении артериального давления [84] и сердечного ритма плода [11] было установлено, что именно большой диапазон значений экспонент Гельдера-Липшица свидетельствовали о нормальном существовании и развитии исследуемого объекта. Широкий спектр значений α также связывают с повышенной мульти-фрактальностью системы и наоборот.

В основном ременные ряды являются дискретными величинами, и более того не являющиеся дискретными значениями гладких функций, то есть недифференцируемыми в каждый момент времени t, следовательно, у нас не получится получить удовлетворительной аппроксимации полиномом более высоких порядков в окрестности точки $t = t_0$. Поэтому предполагается, что функция помимо регулярной (гладкой) составляющей также содержит слагаемые с нецелочисленным показателем степени времени, отвечающие за хаотичность ее поведения, то есть представима в виде следующей суммы, вытекающей из неравенства (1.2.5):

$$g(t) = p_n(t - t_0) + c|t - t_0|^{\alpha(t_0)} \qquad (1.2.6)$$

Исторически первым методом для оценки спектра сингулярности появился метод WTMM, основанный на использовании непрерывного вейвлет-преобразования (НВП) сигнала, так как именно вейвлет-преобразование хорошо подходит для устранения полиномиальной компоненты из сигнала.

Определим НВП сигнала $g(t)$ по формуле:

$$W(a, s) = \frac{1}{\sqrt{s}} \int_{-\infty}^{+\infty} g(t) \cdot \phi\left(\frac{a - \tau}{s}\right) dt, \qquad (1.2.7)$$

где a, s, ϕ – параметры сдвига по времени, масштаба и вейвлет-функция соответственно, в качестве которой, например, можно использовать функцию Гаусса или ее производные:

$$\phi_n(x) = (-1)^n \cdot \frac{d^n e^{-\frac{x}{2}}}{dx^n}. \qquad (1.2.8)$$

Так, например, вторая производная функции Гаусса, известная также под названием «мексиканская шляпа», позволяет устранить полиномы нулевого (постоянную составляющую) и первого порядков (тренд). Более подробно ознакомиться с теорией вейвлет-анализа сигналов можно, например, в работе [78]. В силу избыточности НВП для выделения сингулярности функции достаточно использовать информацию лишь об экстремумах полученных вейвлет-коэффициентов, построив древовидную структуру – скелета вейвлет-преобразования. Пример такого скелета изображен во второй главе на рисунке 2.9.2. Следуя по линии скелета, соединяющей экстремум на различных уровнях детальности, можно также отследить и поведение сингулярности функции $g(t)$. При малых значениях масштаба выполняется степенной закон, связывающий значения коэффициентов НВП и уровень детальности:

$$|W(a_k, s)| \sim s^{\alpha_k} \qquad (1.2.9)$$

В случае, когда для каждой сингулярности функции показатель Гельдера-Липшица имеет одно и то же значение, то есть $\alpha_k = H$, то это означает что функция является моно-фракталом, однако если они отличаются, то это позволяет нам утверждать, что функция имеет более сложную структуру – мульти-фрактальную.

Чтобы избежать вычислительной ошибки из-за влияния друг на друга близко расположенных нерегулярностей, вместо самих модулей вейвлет-коэффициентов принято для расчетов использовать более точную и надежную оценку в виде частичных сумм $Z(q, s)$, определяемых по формуле

$$Z(q, s) = \sum_{l \in \Omega(s)} |W(a, s)|^q, \qquad (1.2.10)$$

где $\Omega(s)$ – множество всех экстремумов на уровне детальности s, а q в общем случае является любым действительным числом. Однако бывают случаи, когда величины $W(a, s)$ обращаются в 0, и тогда частичная сумма $Z(q, s)$ не может быть определена для отрицательных значений степени q, поэтому используется следующая формула

$$Z(q, s) = \sum_{l \in \Omega(s)} \left(\sup_{s^* \le s} |W(a, s^*)| \right)^q, \qquad (1.2.11)$$

где s^* - некоторое заданное фиксированное значение уровня детальности. Аналогично (1.2.9) величина (1.2.11) при малых значениям s также демонстрирует свойства самоподобия:

$$Z(q, s) \sim s^{\tau(q)}, \qquad (1.2.12)$$

где $\tau(q)$ – угол наклона линии регрессии к зависимости между $\ln(s)$ и $\ln\big(Z(q, s)\big)$, также называемым еще показателем массы. При расчетах принято брать как положительные так и отрицательные значения степени q, чтобы учесть влияние как больших значений модуля $|W(a,s)|$ так и малых соответственно. Связь между показателем массы $\tau(q)$ и мульти-фрактальными характеристиками определяется с помощью преобразования Лежандра:

$$\begin{cases} \alpha = \dfrac{d\tau}{dq} \\ f(\alpha) = q\alpha - \tau(q) \end{cases}, \qquad (1.2.13)$$

где α – показатели Гельдера-Липшица, а $f(\alpha)$ – спектр сингулярности. Значение α, при котором достигается максимум функции $f(\alpha)$ называется обобщенным показателем Херста, $\alpha = \alpha_H$. Шириной спектра сингулярности называется величина $\Delta\alpha = \alpha_{max} - \alpha_{min}$. Одним из преимуществ метода WTMM для оценки спектра сингулярности является возможность оценки спектра как по максимумам коэффициентов вейвлет-преобразования, так и по минимумам в отдельности [72]. Данный подход еще несильно изучен, однако предоставляет собой большой потенциал для анализа характеристик исследуемых систем с других сторон.

В работе [83] показано, что метод MFDFA оказался более надежным и подходящим для вычисления спектра сингулярности. Одним из преимуществ метода MFDFA перед WTMM является более точное определение моно- и бифрактальных структур. Также на спектр сингулярности сильное влияние может оказать выбор материнского вейвлета. Однако зачастую оба метода дают схожий результат. Несмотря на надежность метода MFDFA, он также имеет и минус в сравнении с методом максимумов вейвлет-преобразований – MFDFA не применим для сигналов с большим количеством пропусков в сигналах, так как на определенных масштабах могут найтись интервалы времени, в которых временной ряд всюду равен константе, то есть величина флуктуаций на нем равна 0. Данный факт не позволит возвести нулевой размах в отрицательные степени q, следовательно, данный интервал выпадет из расчетов, что приведет к неправильной оценке спектра сингулярности, в то время как WTMM остается полностью рабочим. Одним из классических примеров, для которого невозможно применить метод MFDFA является так называемая «чертова лестница» – разновидность канторова множества.

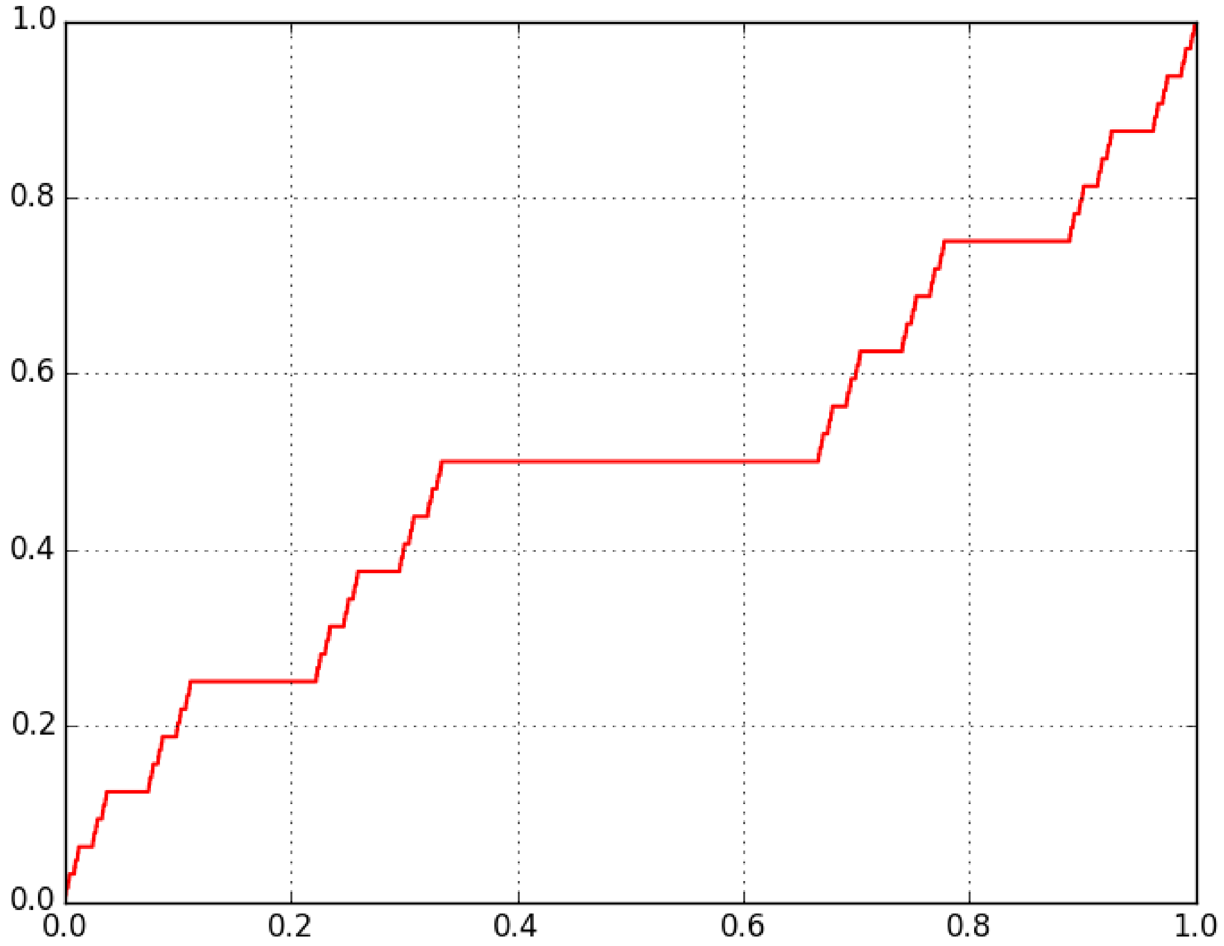


Рисунок 1.2.3. «Чертова лестница»

Исследование мульти-фрактальных свойств и характеристик сигналов GPS проводилось многими авторами. Так, например, работе Захарова В.С. [13] показано, что шумовая компонента исследуемых временных рядов GPS не подлежит устранению, так как не является в значительной степени вредной, а наоборот, хранит в себе важную информацию о состоянии земной коры, под которой происходят сложнейшей природы движения литосферных плит. Основным результатом является установление соответствия между динамическими характеристиками шумовой составляющей сигналов GPS и геодинамическими и геотектоническими характеристиками исследуемых регионов.

Помимо сейсмотектоники сигналы спутниковые системы позволяют производить мониторинг состояния ионосферы. Проведенный в работе [5] анализ позволил установить, что с помощью GPS можно не только производить контроль за изменением состояния ионосферы, но и также локализовать возможные районы предполагаемых землетрясений, так как в период его подготовки происходит уменьшение значения максимума электронной концентрации в слое ионосферы.

Как уже было упомянуто выше, доказан факт полезности и информативности хаотической компоненты во временных рядах GPS. На основе анализа низкочастотного микросейсмического фона, зарегистрированного с помощью сейсмических станций сети F-net, был сделан долгосрочный прогноз [27] предстоящего крупного землетрясения в Японии 11 марта 2011 года в 130 км от населенного пункта Сендай на острове Хонсю, названное после «Великое японское землетрясение». Такое название ему было присвоено из-за невероятной силы и нанесенного ущерба. По данным Геологической службы США магнитуда в эпицентре землетрясения достигла 9, а сам остров Хонсю сместился на 2.4 метра. Данная оценка о смещении была получена на основании смещения одной из станций GPS. Так как станции GPS фиксируют даже самые малые смещения земной поверхности по всем трем координатам, то возник вопрос, возможно ли составить прогноз землетрясения и дать оценку расположения грядущей катастрофы с помощью временных рядов GPS. В качестве одного из основных инструментов в работах Любушина по анализу сейсмических временных рядов использовалась теория мультифракталов, а ключевым исследуемым параметром являлась ширина спектра сингулярности $\Delta\alpha$. Данный спектр сингулярностей вычислялся уже упомянутым выше методом MFDFA. Изложим ниже подробнее этот алгоритм вычисления $\Delta\alpha$ [30], [64].

1. Разобьем временной ряд $X(t), t = 1, \dots, N$ на $N(s) = \left[\frac{N}{s}\right]$ непересекающихся интервалов, где s – число отсчетов в каждом интервале, вычисляемое следующим образом:

$$s = k \cdot 10^{\Delta s},$$

где $\Delta s = \frac{(\log s_{max} - \log s_{min})}{Ns-1}$, $k = 0, \dots, Ns - 1$, Ns – общее число различных длин интервалов, s_{max}, s_{min} – максимальная и минимальная длины интервалов. В общем случае N некратно s, и может остаться неучтенной часть в конце временного ряда. Чтобы избежать этого, повторим ту же процедуру разбиения, начиная с конца сигнала. Таким образом, получим $2N(s)$ временных промежутков. Важно отметить, что для получения лучшей оценки спектра сингулярности необходимо получить равноудаленные значения s в логарифмическом масштабе, и только потом вычислять реальное число отсчетов на каждом интервале. В противном случае, при переходе к логарифмическому масштабу, большая часть рассчитанных показателей Гельдера-Липшица будет сконцентрирована в одной стороне, возле α_{min}, что приведет к смещению «колокола» спектра сингулярности, а, следовательно, в лучшем случае к неверной оценке обобщенного показателя Херста.

2. Воспользовавшись методом наименьших квадратов, удалим тренд на каждом из $2N(s)$ промежутков

$$y_j^{s,m}(t) = y_j^s(t) - p_j^{s,m}(t),$$

где $y_j^s(t) = X\big((j-1)s + i\big)$ для $j = 1, \dots, N(s)$ и $y_j^s(t) = X\Big(N - \big(j - N(s)\big)s + i\Big)$ для $j = N(s) + 1, \dots, 2N(s)$, а m – порядок аппроксимирующего полинома. В случае, когда $m \geq 1$, метод MFDFA также называют MFDFA2, MFDFA3 и т.д. соответственно. При отсутствии явных тенденций в сигнале необходимо удалить постоянную составляющую, то есть вычесть из значений временного ряда его математическое ожидание.

3. Вычислим затем следующую сумму:

$$Z^m(q,s) = \left(\frac{1}{2N(s)} \sum_{j=1}^{2N(s)} \left(\max_{1 \le t \le s} y_j^{s,m}(t) - \min_{1 \le t \le s} y_j^{s,m}(t) \right)^q \right)^{\frac{1}{q}}, \qquad (1.2.14)$$

где в общем случае параметр q может принимать любое действительное число. Аналогично методу WTMM вариация параметра q позволит увидеть, как зависит величина $Z^m(q,s)$ от масштаба, так как при $q > 0$ наибольший вклад в сумму привносят промежутки с большими колебаниями, а при $q < 0$ – промежутки с малыми колебаниями. При $q = 0$, воспользовавшись логарифмическим усреднением, получим следующую формулу для частичной суммы:

$$Z^m(0,s) = \exp\left(\frac{1}{4N(s)} \sum_{j=1}^{2N(s)} \ln\left(\max_{1 \le t \le s} y_j^{s,m}(t) - \min_{1 \le t \le s} y_j^{s,m}(t) \right) \right) \quad (1.2.15)$$

Для слишком больших масштабов ($s > N/4$) величина $Z^m(q,s)$ становится статистически ненадежной, так как число временных промежутков $N(s)$ слишком мало.

4. Определим величину $h(q)$ как коэффициент линейной регрессии между $\log s$ и $\log Z^m(q,s)$: $Z^m(q,s) \sim s^{h(q)}$.
5. Далее мультифрактальный спектр вычисляется с помощью преобразования Лежандра по следующей формуле:

$$F(\alpha) = \max\left(\min_q \left[q\big(\alpha - h(q)\big) + 1 \right], 0 \right), \qquad (1.2.16)$$

где $\alpha = h(q) + qh'(q)$.

В работе [89] доказано, что спектр сингулярности для мультифрактальных объектов имеет выпуклый колоколообразный вид. В случае монофрактала он вообще сходится в точку. Снижение мультифрактальности, то есть уменьшение ширины спектра сингулярности, связывают с увеличением межэлементной связи в системе – синхронизацией, что, в свою очередь, может являться предвестником катастрофы. Таким образом, Любушин А.А. применил данный подход к анализу микросейсмических колебаний и указал на увеличение значения множественной корреляции мультифрактальных параметров, что позволило предсказать подготовку землетрясения в Японии 11 марта 2011 года [28].

Настоящая работа является логическим продолжением работ Любушина А.А., однако в качестве источника информации о сейсмической активности земной поверхности вместо сейсмических станций являются стационарные станции GPS, регистрирующие даже самые малые смещения литосферных плит, и анализ производится без применения мультифрактальной теории. Полученные результаты будут сопоставлены с результатами, получаемыми с помощью фрактального подхода (параграф 3.5).

Сигналы GPS не являются исключением и также, как и многие другие временные ряды, содержат в себе различного рода особенности, такие как скачки, выбросы и резкие изменения трендов. Назовем эти особенности событиями и перейдем к описанию и обзору методов для их поиска в анализируемых сигналах.

1.3. Выбросы

Как правило, выбросом называют некоторый объект, который отличается от других представителей одной и той же выборки по одному или нескольким признакам. Одним из наиболее часто встречающихся в работах исследователей определением выброса является определение, предложенное Хоукинсом [63], которое в переводе звучит следующим образом: выброс – это наблюдение, которое настолько отличается от остальных наблюдений, что кажется, что оно было вызвано каким-либо другим механизмом[1].

---

[1] An outlier is an observation which deviates so much from the other observations as to arouse suspicions that it was generated by a different mechanism (оригинал)

Существует немалое количество методов для поиска выбросов во временных рядах, основанных на статистических тестах, изменении дисперсий и т.д. Сравнительно молодыми методами являются алгоритмы, опирающиеся на расстояния между элементами выборки, другими словами, использующие алгоритмы кластеризации. Ранее среди ученых было схожее мнение о том, что вся полезная информация о сигналах хранится в трендах и их изменениях, и поиск выбросов производится с целью их последующего удаления из исходных данных, так как они считаются случайными событиями, нарушающими общую статистическую картину, в том числе если речь идет об исследовании геофизических сигналов. Однако со временем данное мнение преобразовалось, и теперь все больше исследователей прилагают усилия для анализа именно шумовой компоненты временного ряда, так как изучение выбросов может оказаться полезным:

1. В социологии при анализе результатов опросов появление резких отклонений говорит о наличие интереса к предмету опроса.
2. В естественных науках имеется большое число систем мониторинга, порождающих временные ряды, выбросы в которых зачастую свидетельствуют о скрытых природных или антропогенных факторов, вызывающих подобные аномалии и т.д.

В данной работе в основе анализа также лежит предположение о том, что выбросы не считаются лишней информацией, которая препятствует изучению временных рядов, а обратно, рассматриваются в качестве носителей информации об аномалиях и нарушениях структуры сигналов, которые, согласно определению Хоукинса, вызваны сторонним механизмом. Весь анализ нацелен на изучение этого механизма. Выбросы во временных рядах GPS могут позволить получить больше информации о том, что происходит с земной поверхностью, а самое главное, что происходит непосредственной под ней, ведь они, за исключением тех случаев, когда это всего лишь случайные явления, вызванные погодными условиям, перепадами электричества и т.д., отражают резкие смещения земной коры. Такие смещения происходят в результате выделения всплесков накопленной между литосферными плитами энергии. Иногда в литературе по анализу временных рядов GPS встречается такое понятие, как изменение скорости. Для его объяснения используется упрощенная модель землетрясения.

Пусть на шероховатой поверхности лежит прямоугольное тело массы m, давящее на поверхности с силой $P = mg$, где g – ускорение свободного падения. Через пружину с небольшим коэффициентом жесткости k на него действует сила F, которая перемещает тело с постоянной и очень малой скоростью. При условии, что сила трения тела по поверхности гораздо выше силы трения скольжения, мы будем наблюдать следующую картину, изображенную на иллюстрации ниже:

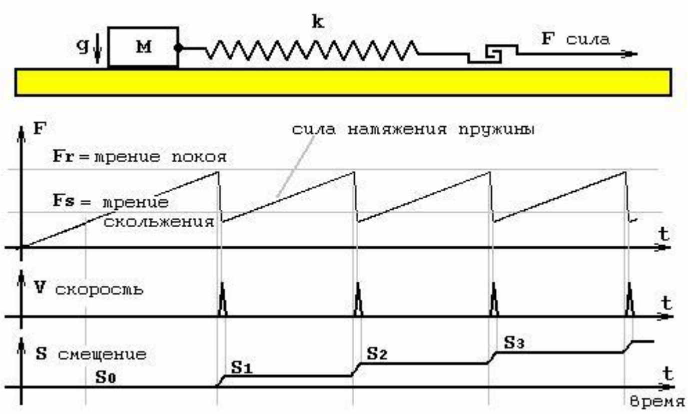


Рисунок 1.3.1. Упрощенная модель землетрясения[2]

На рисунке 1.3.1 мы видим, что график скорости $V(t)$ почти всюду равен нулю в силу того, что тело покоится до тех пор, пока сила, действующая на него, не превысит силу трения о поверхность. Затем происходит резкое смещение под действием сразу трех сил: инерции, натяжения пружины и трения скольжения. Этот процесс происходит аналогичным образом и при движении литосферным плит: вдоль разделяющей плиты поверхности сдвига под действием огромных сил трения происходит разрушение горных пород, что влечет за собой такие резкие смещения. Таким образом, чем больше резких смещений (выбросов), тем менее

[2] Главные движущие силы землетрясений, дрейфа континентов и горообразования. URL: http://geo.web.ru/db/msg.html?mid=1174973

прочной становится земная кора в местах стыка литосферных плит, что повышает вероятность возникновения в этой области землетрясения. Следовательно, предположение о том, что выбросы во временных рядах GPS являются случайными, и решение об их исключении из исходных данных могут привести к потере важной информации о системе, вызывающей подобные аномалии в сигнале.

Говоря про методы выделения выбросов во временных рядах, то их на данный момент имеется огромное количество, начиная от классических методов, некоторые из которых хорошо описаны в [88] до различных автоматизированных тестов. Наиболее подробный список методов для выделения выбросов в сигналах предложен в книге [53]. Однако исследования уже существующих результатов по выделению выбросов во временных рядах GPS не дали положительного результата: гипотез или выводов о природе данных отклонений сделано не было.

1.4. Скачки среднего уровня

Другими часто встречающимся во время анализа событием являются разрывы – резкие скачки среднего уровня. Они так же несут в себе полезную информацию. На рисунке 1.3.1 мы видим, что график смещения – это кусочно-ступенчатая функция. Моменты времени изменения скорости и смещения для обоих графиков совпадают. В реальной жизни график смещений, полученный с действующей GPS станции, не является кусочно-ступенчатой функцией, а содержит помимо естественных колебаний земной поверхности, характерных для состояния «покоя» Земли, еще и шум (белый и так называемый фликкер-шум). Задача поиска скачков среднего уровня в каком-то смысле является частным случаем задачи выделения трендов, так как участок временного ряда с постоянным средним – это тренд, параллельный оси абсцисс.

Для поиска скачков в сигналах существует также огромное число методов, такие как, например, Standard Normal Homogeneity Test (SNHT) [62], отношение

Фишера, U-критерий Манна-Уитни и др. Главной особенностей (положительной или отрицательной) всех этих методов является то, что они все используются для оценки сходств по какому-либо признаку между двумя выборками. Применительно к временным рядам GPS эти алгоритмы в своем первоначальном виде не подходят, так как исследуемый сигнал, как правило, содержит множество скачков и изменений трендов. Для выделения нескольких событий приходится применять данные методы рекурсивно и придумывать критерий остановки его работы.

Однако задачу поиска участков с постоянным средним значением возможно решить и другим способом, построив кусочно-ступенчатую аппроксимацию исходного временного ряда. Такая аппроксимация, вероятнее всего, не будет настолько же однозначно соответствовать выбросам, как это показано на рисунке 1.3.1, в силу идеализации модели, тем не менее полезной статистической характеристикой сигнала является его степень (мера) скачкообразности: чем более скачкообразным является временной ряд GPS, тем выше сейсмическая активность Земли. Наиболее популярными являются методы, основанные на дискретном вейвлет-преобразовании Хаара и непрерывном вейвлет-преобразовании (WTMM). Данные методы построения кусочно-ступенчатых аппроксимаций для временных рядов будут рассмотрены подробно ниже во второй главе. Среди прочих алгоритмов выделения скачкообразной составляющей временных рядов следует упомянуть метод, основанный на использовании статистики «хи-квадрат». В работе [87] предложен алгоритм обнаружения резких изменений среднего значения во временных рядах GPS, основанный на критерии Стьюдента и названный STARS (sequential t-test analysis of regime shifts). Данный алгоритм был изобретен применимо к анализу климатических изменений, и согласно терминологии автора, отметившим, что эти изменения проявляются в виде «режимов», будем использовать термин «режим» для обозначения «ступенек» кусочно-ступенчатой аппроксимации. Алгоритм STARS состоит из следующих этапов:

1. Зададим длину скользящего окна равной l.
2. Определим величину

$$diff = t \cdot \sqrt{\frac{2\sigma_l^2}{l}},$$

где t – это значение функции распределения Стьюдента с $2l - 2$ степенями свобод при заданном уровне значимости p.

3. Вычислим среднее значение $\bar{x}_{R1}$ первых l значений исходной выборки X в качестве оценки для режима $R1$, тогда значения, которые будут относиться к режиму $R2$, должны находиться за порогом $\bar{x}'_{R2} = \bar{x}_{R1} \pm diff$.
4. Далее для каждого нового значения, начиная с отсчета $i = l + 1$, проверяем, выходит ли оно заданный порог $\bar{x}_{R1} \pm diff$. Если нет, то полагаем, что режим остался прежним, и пересчитываем среднее значение $\bar{x}_{R1}$ с учетом нового значения x_i и $l - 1$ предыдущих значений. Если же новое значение превышает заданный порог, полагаем, что режим сместился, и это, возможно, начальная точка j нового режима $R2$.
5. После того как точка j обнаружена, каждое новое значение $x_i, i > j$ используется для подтверждения или опровержения нулевой гипотезы о том, что j – момент изменения режима. Если отклонение $x_i - \bar{x}'_{R2}$ имеет тот же знак, что и в момент появления смещения режима, то это увеличивает вероятность того, что смена режима действительно произошла. Обратное утверждение справедливо, если отклонение имеет противоположный знак. Данное изменение в достоверности того, что смена режима произошла в момент времени $i = j$ отражается в так называемом индексе смены режима RSI (regime shift index), который является накопленной суммой нормализованных отклонений:

$$RSI_{i,j} = \sum_{i=j}^{j+m} \frac{x_i^*}{l\sigma_l}, \qquad m = 0,1,\dots,l-1.$$

Здесь $x_i^* = x_i - \bar{x}'_{R2}$, если сдвиг положительный, и $x_i^* = \bar{x}'_{R2} - x_i$, если сдвиг отрицательный. Если в какой-то момент времени в промежутке $i = [j + 1, j + l - 1]$ индекс RSI принимает отрицательное значение, то переходим к шагу 6, в противном случае, переходим к шагу 7.

6. Отрицательное значение индекса RSI означает, что тест на наличие смены режима в момент времени j не удался. Присваиваем индексу значение 0 и пересчитываем среднее значение $\bar{x}_{R1}$ для проверки значений $x_i, i > j + 1$ на превышение порога $\bar{x}_{R1} \pm diff$, как на этапе 4.
7. Положительное значение индекса RSI означает, что изменение режима в момент времени j оказалось статистически значимым при уровне значимости p. Вычисляем среднее значение $\bar{x}_{R2}$, и с данного отсчета оно становится опорным, так как дальнейшие вычисления будут проводиться относительно него. Поиск следующего смещения к режиму $R3$ начинается с момента времени $i = j + 1$. Данный подход необходим, чтобы убедиться, что временные границы следующего режима определены верно, даже если фактическая продолжительность режима $R2$ меньше, чем l отсчетов. Вычисление продолжаются с шага 4 по шаг 7, пока все данные временного ряда X не будут проанализированы. Если временных рядов несколько, то индекс RSI является арифметическим средним индексов RSI для каждого временного ряда.

Данный алгоритм был придуман, так как другие ранее изобретенные алгоритмы для выделения скачков во временных рядах, например, [55], [67] имели общую проблему: резкое ухудшение результата статистических тестов по краям временных рядов. Это означает, что для их применения должно быть накоплено значительное количество данных, и несмотря на то, что уже имеется достаточно большой объем данных, велика вероятность того, что не так давно произошла смена режима [86].

В 2014 году алгоритм STARS был протестирован на смоделированных ременных рядах, содержащих в себе все особенности (линейные тренды, сезонные циклы, скачки, пропуски и сочетания белого и фликкер шумов) реальных временных рядов, получаемых с помощью спутниковой навигационной системы. В связи с этим метод STARS был модифицирован [85], и результаты анализа показали, что 48% определенных с помощью модифицированного алгоритма STARS резких изменений среднего значения оказались верными.

1.5. Источники данных

Во время исследования использовались различные трехкомпонентные GPS сигналы (содержащие смещения координат по трем координатам: восточным, северным и вертикальным), которые находятся в свободном для скачивания доступе по нижеприведенным ссылкам:

1. ftp://sideshow.jpl.nasa.gov/pub/JPL_GPS_Timeseries/japan/30min_sol/ – временные ряды с шагом по времени 30 минут, взятые с 1248 GPS станций, расположенных на Японских островах в период с 30 января 2011 года по 26 марта 2011 года.
2. http://gf9.ucs.indiana.edu/daily_rdahmmexec/daily/ – временные ряды с шагом по времени 1440 минут (1 день), взятые с 10590 GPS станций из разных регионов в период с 1996 года и обновляются ежедневно по сей день.
3. ftp://gneiss.nbmg.unr.edu/rapids_5min/kenv/ – временные ряды с шагом по времени 5 минут, взятые с 8902 GPS станций из разных регионов в период с 28 февраля 2013 года и обновляются ежедневно по сей день.

Информация о землетрясениях взята с сайта Геологической службы США (http://earthquake.usgs.gov/search).

В настоящей работе по большей части приведены примеры и результаты анализа 30-минутных временных рядов GPS, полученных со станций Японских островов. Они являются более пригодными для прогноза, нежели дневные. Отсутствие других примеров обусловлено малым количеством сильных (магнитудой большей или равной 7.5) землетрясений и отсутствием данных GPS в тех регионах, в которых сильные землетрясения все-таки случаются.

Наблюдения за сейсмической активностью Земли и движением литосферных плит земной поверхности ведут по всему миру. Число только представленных в открытом доступе GPS станций превышает десяток тысяч. Данная сеть станций не является равномерно распределенной, что не позволяет получить качественную картину о всех смещениях земной поверхности, однако имеется некоторое число регионов с очень густой сетью станций, что в результате анализа позволяет получить более четкий и информативный результат. На рисунке 1.5.1 видно, что наиболее густую сеть GPS станций имеют Соединенные Штаты Америки, Европа, Япония, Австралия и Новая Зеландия.

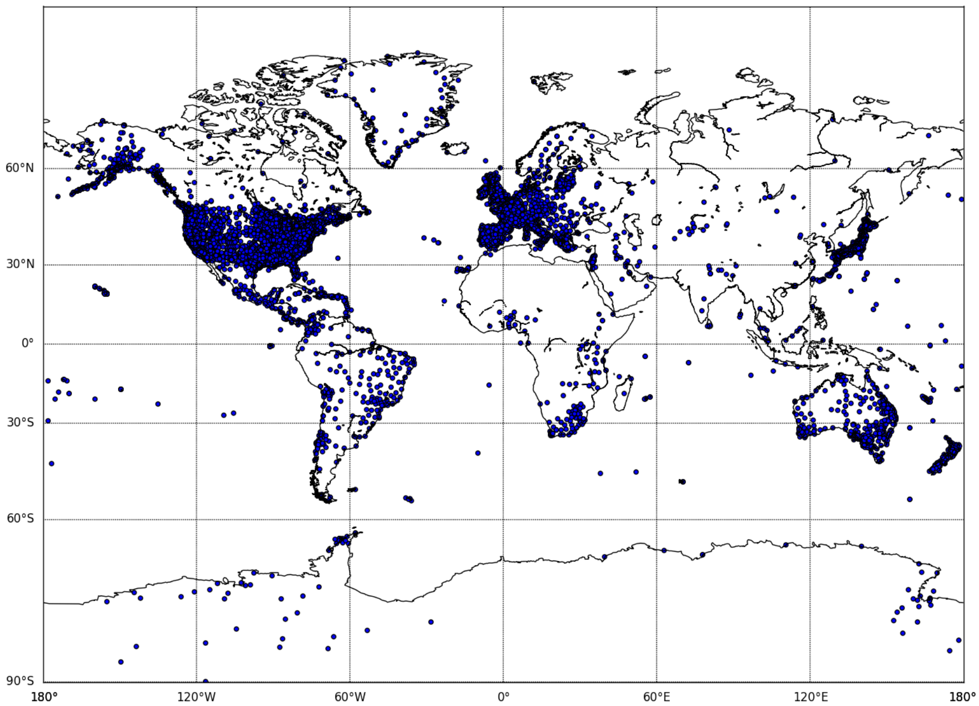


Рисунок 1.5.1. Расположение GPS станций

В течение последних десятков лет населению Земли все чаще приходится сталкиваться с сильными и разрушительными землетрясениями магнитудой, превышающей или равной 8. Тенденция роста количества таких землетрясений заставляет насторожиться. Согласно данным, предоставленным Геологической службой Соединенных Штатов Америки, число сильных толчков увеличивается приблизительно в два с половиной раза каждые десять лет. Ниже в таблице приведена данная статистика:

Таблица 1.5.1. Статистика землетрясений магнитудой ≥ 8 в период с 1974 года по 2014 год

Период	**Количество землетрясений**
1974.01.01 – 1984.01.01	1
1984.01.01 – 1994.01.01	4
1994.01.01 – 2004.01.01	10
2004.01.01 – 2014.01.01	17

Расположение эпицентров землетрясений за 2014 год магнитудой выше 4 изображено на рисунке ниже. Он иллюстрирует известный факт, что в основном землетрясения происходят на стыке литосферных плит. В следствие этого изучение землетрясений сильно усложняется, так как эти стыки в большинстве своем приходятся на океаническую кору, где в настоящий момент установка систем наблюдения невозможна. Наиболее доступными для изучения регионами являются Соединенные Штаты Америки и острова Японии, так как в этих регионах установлено достаточно большое количество стационарных станций GPS, землетрясения тут – сравнительно частое явление. Однако сейсмическая активность США заметно ниже, чем у японских островов. Так же для анализа интересны регионы Новой Зеландии и Южной Америки (Чили), однако, к сожалению, на этих территориях расположено слишком мало GPS станций, что в результате анализа не позволяет получить адекватную и информативную картину.

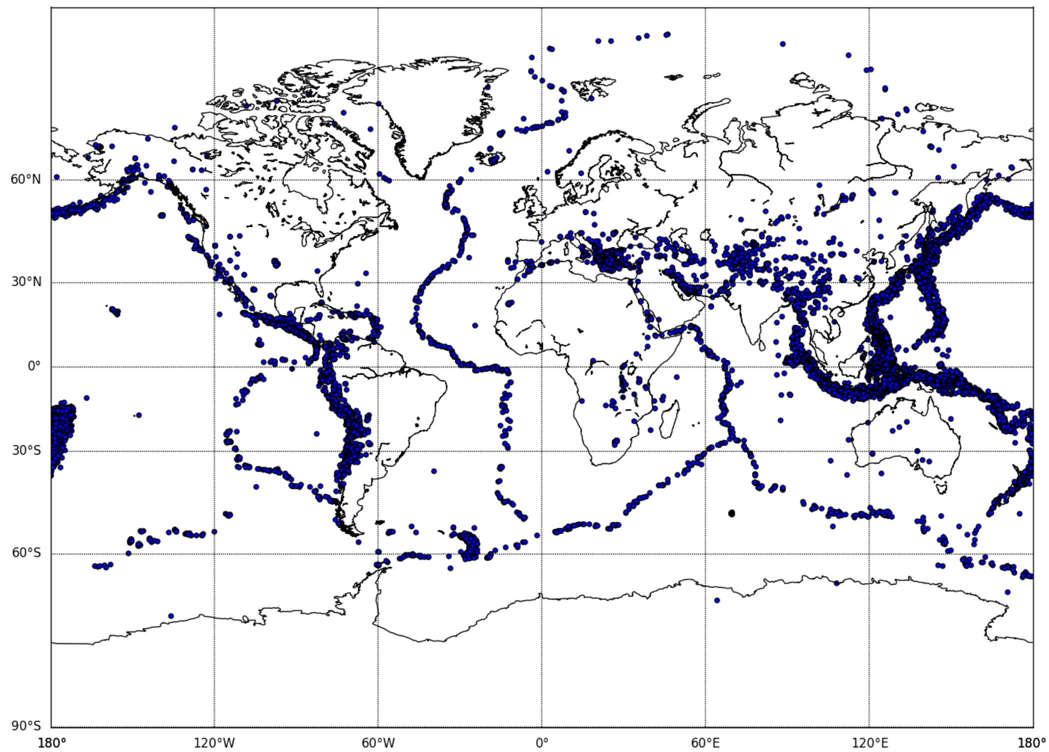


Рисунок 1.5.2. Расположение эпицентров землетрясений магнитудой больше 4 за 2014 год. Число землетрясений 17547.

1.6. Заключение к первой главе

В первой главе рассмотрены основные направления в анализе временных рядов и их задачи. Так же введено понятие «события», в качестве которого в работе рассматриваются выбросы, скачки среднего уровня и изломы трендов. Описаны основные цели выделения данных событий и существующие на сегодняшний день результаты исследований данных, получаемых с помощью спутниковой навигационной системы позиционирования GPS. Также в первой главе упомянута упрощенная модель землетрясения, объясняющая, почему важно не устранять выбросы и скачки из временных рядов, а наоборот, сделать упор на их исследовании. Указаны основные источники данных, используемых во время анализа, и выделены сейсмически активные регионы Земли, представляющие наибольший интерес для изучения.

Глава 2. Методы определения меры выбросов и скачкообразности временных рядов GPS

2.1. Пропуски во временных рядах

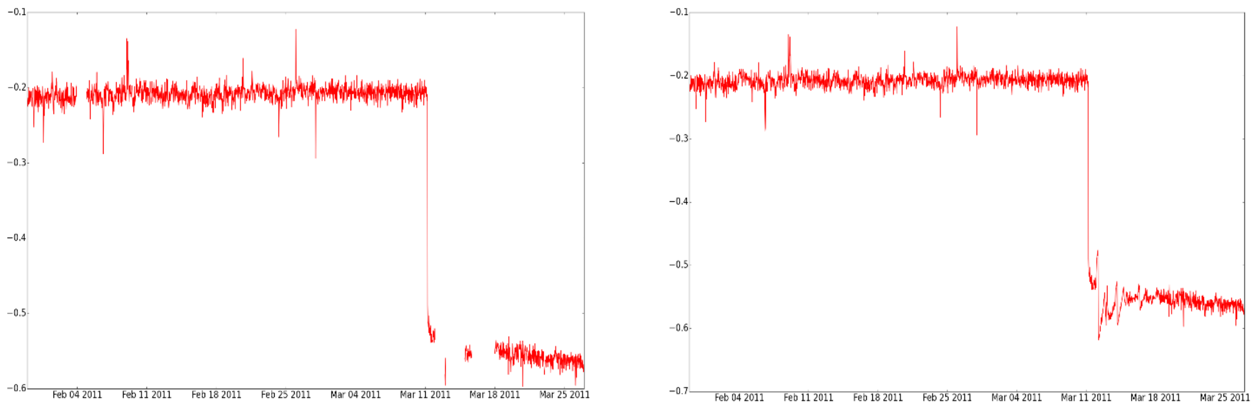
Как правило, при анализе временных рядов нам приходится сталкиваться с отсутствием разного количества данных, вызванного ошибкой регистрации прибора, сбоями в электрической сети и т.п. Временные ряды GPS также не являются исключениями. GPS сигналы, взятые с шагом по времени 1 день, предоставляются уже с заполненными пропусками. Промежутки в данных заменены константой, равной последнему зарегистрированному значению. Никакого другого алгоритма для восполнения набора дневных данных не использовалось, так как дневные временные ряды GPS в основном использовались для построения примеров кусочно-ступенчатых аппроксимаций в силу того, что они являются менее зашумленными, чем 30-минутные и 5-минутные сигналы, и примеры получаются более наглядными.

Промежутки во временных рядах, взятых с временным шагом в 30 минут, заполнены с помощью другого алгоритма. Пусть $X(t), t = 0, \dots, N-1$ – временной ряд, Δt – шаг по времени, а $G_{t_1,t_2}, t_2 > t_1 + \Delta t$ – промежуток в регистрации данных в интервале времени (t_1, t_2), где $t_1 = 0, \dots N-3$, $t_2 = 2, \dots N-1$. Положим $M = t_2 - t_1 - 1$ – количество пропущенных данных. Возьмем M значений слева и справа от пропуска, то есть

$$\begin{aligned} L(t) &= X(t), t = t_1 - M, \dots, t_1, \\ R(t) &= X(t), t = t_2, \dots t_2 + M. \end{aligned} \qquad (2.1.1)$$

Удалим из $L(t)$ и $R(t)$ линейный тренд, найдем полусумму их отклонений от трендов и наложим ее на линейный тренд, являющимся прямой, проходящей через точки $\left(t_1, X_{t_1}\right)$ и $\left(t_2, X_{t_2}\right)$. Получившиеся значения и будут использоваться в качестве восполняющего сигнала. Такой алгоритм хорошо подходит в следующих трех случаях:

1. Когда пропусков не слишком много, тогда восстановленный сигнал визуально практически не отличается от исходного временного ряда.
2. Когда промежутки расположены более, чем на M отсчетов к первому и последнему зарегистрированным значениям.
3. Когда соседние промежутки расположены на расстоянии друг от друга более, чем на M отсчетов.



а) исходный временной ряд   б) восстановленный временной ряд

Рисунок 2.1.1. Пример временного ряда с пропусками.

Такие условия являются довольно сильными, что не позволило использовать данный подход на временных рядах GPS, взятых с шагом по времени 5 минут, поэтому пропуски в этих сигналах восполнены константой так же, как и в дневных.

Так как любое восстановление данных может сильно отличаться от реальных показаний, во время анализа будет использоваться параметр γ_{t_1,t_2} – процент пропущенных отсчетов за промежуток времени (t_1, t_2), поэтому метод восстановления недостающих данных практически никак не влияет на результаты анализа за счет их минимизации параметром γ_{t_1,t_2}, равным в рамках работы 5%.

2.2. **Спектральная экспонента вейвлет-преобразования**

Основываясь на результатах анализа низкочастотного сейсмического шума и предположении [73], что перед крупной катастрофой происходит глобальная синхронизация движений земной поверхности, естественным образом возникает вопрос: возможно ли наблюдать подобный процесс при анализе временных рядов

GPS? Рассмотрим, как коррелировали друг с другом 30-минутные сигналы GPS, полученные с Японских островов в течение месяца перед землетрясением 11 марта 2011 года.

Пусть $X(t),\ t=0,\dots,N-1$ – временной ряд, а c_j^k – коэффициенты вейвлет-преобразования $X(t)$, разложенного в системе ортогональных конечно-базисных функций. Верхний индекс k является номер уровнем детализации вейвлет-разложения, а индекс j соответствует центру временного промежутка. Наибольшее возможное значение уровня детализации m зависит от объема исследуемой выборки. В настоящей работе будем использовать набор из 17 вейвлетов: 10 ортогональных вейвлетов Добеши, порядок которых находится в диапазоне от 2 до 20 (использование более высоких порядков приводит к численной неустойчивости) и 7 так называемым симлетов, являющихся модификацией вейвлетов Добеши, однако имеющих более симметричную форму. Симлеты облают теми же свойствами компактности, ортогональности, полноты и гладкости, как и обычные вейвлеты, однако для порядков в диапазоне от 2 до 6 они совпадают с ортогональным базисом Добеши, в то время как для порядков в диапазоне 8 до 20 можно выявиться некоторые различия в форме базисных функций.

Для выбора оптимального вейвлет-базиса воспользуемся критерием минимума нормализованной энтропии распределения квадрата значений вейвлет-коэффициентов.

$$En=-\sum_{k=1}^{m}\sum_{j=1}^{M_k}p_j^k\ln p_j^k\to min,$$

$$p_j^k=\frac{\left|c_j^k\right|^2}{\sum_{j=1}^{M_k}\left|c_j^k\right|^2}. \qquad (2.2.1)$$

Здесь m – это число уровня детализации, M_k – число вейвлет-коэффициентов на каждом уровне детализации k. Число m зависит от длины временного ряда N.

Например, если $N = 2^n$, тогда $m = n$, $M_k = 2^{n-k}$. Условие $N = 2^n$ необходимо для применения быстрого дискретного вейвлет-преобразования [78]. Если длина N сигнала не является степенью двойки, тогда временной ряд $X(t)$ дополняется нулями до ближайшего значения степени двойки, чтобы получившая длина временного ряда получилась $L = 2^n > N$. В этом случае среди всех 2^{n-k} вейвлет-коэффициентов на уровне детализации k, только $N \cdot 2^{-k}$ соответствуют реальным колебаниям сигнала, в то время как остальные коэффициенты раны нулю. Таким образом только $M_k = N \cdot 2^{-k}$ вейвлет-коэффициентов используется для вычисления нормализованной энтропии.

Метод $(2.2.1)$ позволяет выбрать такой базис для временного ряда $X(t)$, что распределения отличие распределения коэффициентов вейвлет-преобразования будет максимально отличаться от равномерного распределения. В этом случае максимум информации будет сконцентрировано в минимальном количестве вейвлет-коэффициентов. После определения оптимального ортогонального вейвлет-базиса из критерия $(2.2.1)$ можно вычислить средние значения квадратов вейвлет-коэффициентов на каждом уровне детализации:

$$S_k = \frac{1}{M_k} \sum_{j=1}^{M_k} \left| c_j^k \right|^2 \qquad (2.2.2)$$

Средние значения $(2.2.2)$ квадратов ортогональных вейвлет-коэффициентов представляют собой энергию колебаний в соответствии с частотной полосой уровня детализации. Это означает, что значения S_k можно рассматривать в качестве спектральной мощности временного ряда $X(t)$. Частотная полоса уровня детализации с номером k определяется следующим образом:

$$\left[f_{min}^k, f_{max}^k\right] = \left[\frac{1}{2^{k+1}\Delta s}, \frac{1}{2^k \Delta s}\right], \qquad (2.2.3)$$

где Δs – время дискретизации временного ряда. Рассмотрим значения периодов, которые соответствуют центральным частотам полос:

$$T_k = \frac{2}{f_{min}^k + f_{max}^k} = \frac{2\Delta s}{2^{-k} + 2^{-(k+1)}}. \qquad (2.2.4)$$

Таким образом, значения $S_k = S(T_k)$, $k = 1, \dots, m$ схожи с обычным значением спектральной мощности Фурье. Отличием от классической оценки Фурье-спектра является тот факт, что значения (2.2.2) являются более усредненными – поэтому зависимость $S_k = S(T_k)$ более сглаженной. Рассмотрим следующую модель спектральной мощности, полученной с помощью вейвлет-преобразования:

$$\log\big(S(T_k)\big) = b \cdot \log(T_k) + c + \varepsilon_k, \qquad k = 1, \dots, m \qquad (2.2.5)$$

где ε_k – остаточные случайные величины с нулевым средним, являющиеся белым шумом. Параметр b в формуле (2.2.5) может быть назван спектральной экспонентой, и он похож на обычную спектральную экспоненту, которая широко используется для изучения форм спектра мощности. Значение b оценивается из условия метода наименьших квадратов:

$$\sum_{k=1}^{m} \varepsilon_k^2 \to \min_{b,c}. \qquad (2.2.6)$$

2.3. Псевдо-производная

Понятие производной чаще всего используется при математическом моделировании при решении дифференциальных уравнений, однако в анализе временных рядов оно не нашло широкого применения. Зачастую используется переход к приращениям, являющийся в случае временных рядов, переходу к первой производной исходного сигнала. Однако из-за того, что временных ряды, как правило, содержат в себе шумовую компоненту, то использование перехода к приращениям для определения, скажем, участков возрастания или убывания сигнала, максимумов и минимумов, не является пригодным. Для этого необходимо сначала найти тренд временного ряда, представляющий из себя гладкую функцию, и только потом уже использовать производную для описанных выше задач. Чтобы упростить данный процесс, введем понятие псевдопроизводной, которое является некоторым обобщением классического определения производной.

Анализ временных рядов будем проводить в скользящем окне, которое в свою очередь разобьем на два равных подокна L и R – левое и правое соответственно. Данный подход позволит нам более детально взглянуть на то, что происходит с сигналом внутри окна, так как мы будем иметь возможность сравнить состояние «покоя» с теми изменениями, которые происходят в результате появления в окне события. Событием мы назовем одно из трех изменений в сигнале: излом тренда, выброс или скачок.

Результатом анализа должна являться функция, отражающая степень изменчивости исходных данных. Итак, перейдем к вводу определений.

Пусть $X(t)$ – временной ряд, где $t = 0, \dots, N-1$, а $M \geq 2$ – длина скользящего окна. Тогда подокна L и R будут длиной $\left[\frac{M}{2}\right]$ и $M - \left[\frac{M}{2}\right]$ отсчетов соответственно. Таким образом, мы можем собрать некоторую описательную статистику ряда $X(t)$, оперируя следующими величинами

$$L_{max}^{i} = \max_{i \leq t < i + \left[\frac{M}{2}\right]} X(t), \qquad L_{min}^{i} = \min_{i \leq t < i + \left[\frac{M}{2}\right]} X(t), \tag{2.3.1}$$

$$R_{max}^{i} = \max_{i + \left[\frac{M}{2}\right] \leq t < i + M} X(t), \qquad R_{min}^{i} = \min_{i + \left[\frac{M}{2}\right] \leq t < i + M} X(t), \tag{2.3.2}$$

где $i = 0, \dots, N-M-1$. Рекомендуется использоваться в качестве ширины окна четное значение, чтобы анализ был более справедливым, так как и в левом, и в правом подокнах будет одинаковое количество отсчетов временного ряда.

Имея четыре числовых характеристики $(2.3.1)$ и $(2.3.2)$ нам надо связать с помощью них окна L и R. Введем для этого две величины:

$$\Delta ED_i = R_{max}^{i} - L_{min}^{i}, \qquad \Delta ID_i = R_{min}^{i} - L_{max}^{i}, \tag{2.3.3}$$

и назовем их явным и скрытым приращениями соответственно, тогда полным приращением будет являться их сумма.

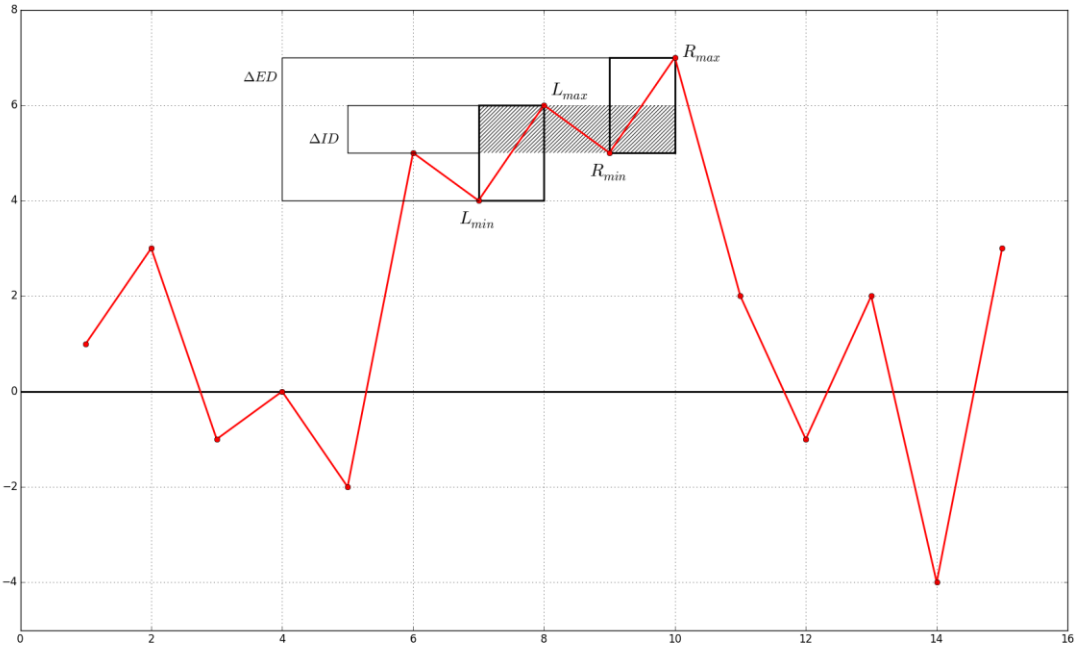


Рисунок 2.3.1. Пример скрытого и явного приращений сигнала.

Псевдо-производной на базе M назовем величину

$$D_i = \frac{\Delta ED_i + \Delta ID_i}{2}, \qquad (2.3.4)$$

которая и будет являться характеристикой изменчивости сигнала. Коэффициент $\frac{1}{2}$ введен, чтобы не учитывать событие дважды.

2.4. Сходства псевдо-производной с производной

Как говорилось выше, понятие псевдо-производной является некоторым обобщением классического определения производной, отсюда и появилось такое название для величины $(2.3.4)$. Рассмотрим три возможных случая для значения D. Так как представленные ниже выкладки справедливы для каждого окна, индекс i в них отсутствует

1. $D > 0 \Rightarrow \frac{R_{max}-L_{min}}{2} > \frac{L_{max}-R_{min}}{2} \Rightarrow \frac{R_{max}+R_{min}}{2} > \frac{L_{max}+L_{min}}{2} \Rightarrow$ середина окна R расположена выше середины окна L, что указывает на рост значений сигнала.
2. $D < 0 \Rightarrow \frac{R_{max}-L_{min}}{2} < \frac{L_{max}-R_{min}}{2} \Rightarrow \frac{R_{max}+R_{min}}{2} < \frac{L_{max}+L_{min}}{2} \Rightarrow$ середина окна L расположена выше середины окна R, что указывает на спад значений

сигнала. Этот случай полностью зеркален предыдущему, за счет того, что скрытое и явное приращения меняются местами.

3. $D = 0 \Rightarrow \frac{R_{max} - L_{min}}{2} = \frac{L_{max} - R_{min}}{2} \Rightarrow \frac{R_{max} + R_{min}}{2} = \frac{L_{max} + L_{min}}{2} \Rightarrow$ середины окон L и R совпадают.

Делаем вывод, что по знаку величины $(2.3.4)$, также, как и по знаку обычной производной, можно судить, возрастают или убывают значения сигнала. Таким же образом можно находить точки экстремума сигнала. Очевидно, что само значение D аппроксимирует величину приращения сигнала в том направлении, в котором он меняется. Напомним, что $i = 0, \dots, N - M - 1$, то есть каждое рассчитанное значение сопоставляется началу окна R, и все значения для этого смещаются по оси абсцисс на $\left[\frac{M}{2}\right]$. Таким образом, мы получаем четыре числовых массива длиной $N - M$ отсчетов каждый.

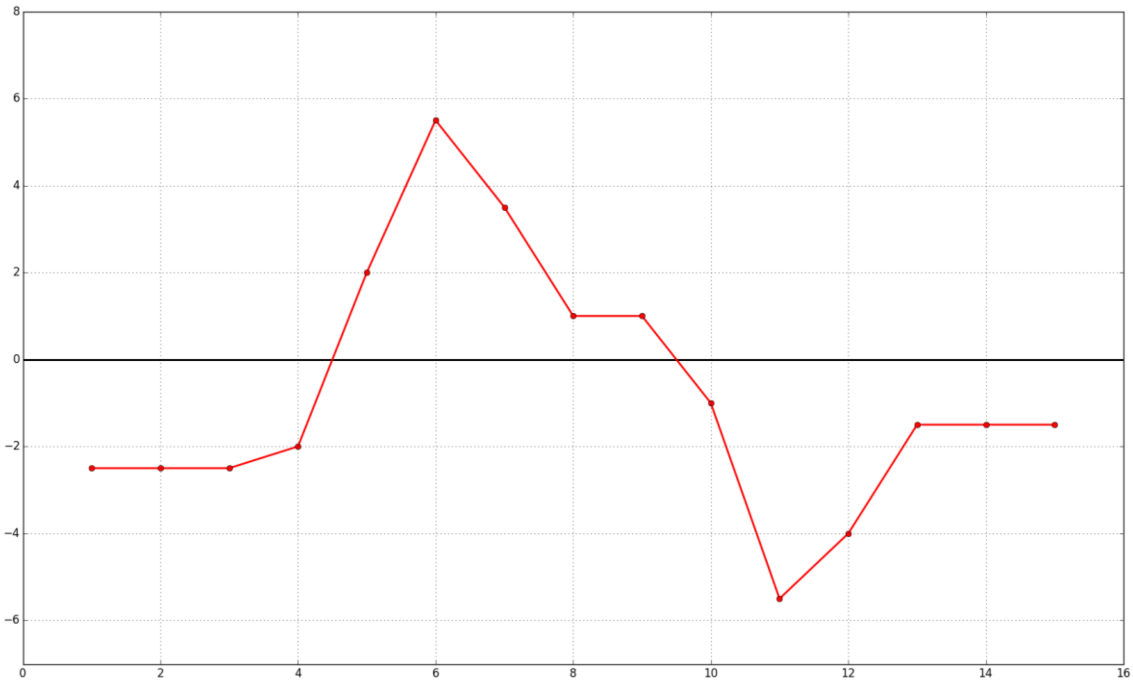


Рисунок 2.4.1. График псевдо-производной для сигнала $X(t)$

Действительно, при $D(t) > 0$, $t \in [5; 10)$ сигнал $X(t)$ возрастает, а при $D(t) < 0, t \in [1,5) \cup [10; 15)$ он убывает. Отсюда можно предположить, что при $D(t) = 0$ сигнал $X(t)$ имеет точки экстремума. Однако в случае дискретных данных мы далеко не всегда можем попасть точно на пересечение графика с осью абсцисс, поэтому точкой экстремума будем считать то значение временного ряда, при котором псевдо-производная меняет знак, то есть в нашем случае это $t = 5$ и $t = 10$ (рисунок 2.4.1). Так как псевдо-производная имеет на M отсчетов меньше, чем исходный сигнал, то крайние значения псевдо-производной предлагается положить равными первому и последнему значениям псевдо-производной, если они отличны от нуля, так как нули производной имеют важное значение при поиске экстремумов. В случае, если первое значение псевдо-производной равно нулю, то положим первые $\left[\frac{M}{2}\right]$ отсчетов равными малому, но ненулевому значению. Аналогично поступаем с последними $M - \left[\frac{M}{2}\right]$ отсчетами, если последнее значение псевдо-производной равно нулю.

В частном случае, когда длина M скользящего окна равна 2 и шаг дискретизации сигнала равен 1, то значения D_i полностью совпадают со значениями конечных разностей, используемых для аппроксимации производной, а также является простым переходом к приращениям, так как $R^i_{max} = R^i_{min} = R$ и $L^i_{max} = L^i_{min} = L \Rightarrow$

$$D_i = \frac{\Delta ED_i + \Delta ID_i}{2} = \frac{R^i_{max} - L^i_{min} + R^i_{min} - L^i_{max}}{2} = \frac{2R - 2L}{2} = X(i+1) - X(i) \qquad (2.4.1)$$

В силу такого сходства псевдо-производной с классическим определением производной, можно говорить о нахождении не только точек экстремума временного ряда, но также будут справедливы рассуждения для определения точек перегиба исследуемых сигналов.

2.5. Свойства псевдо-производной

Наряду со сходством псевдо-производной с производной в определении промежутков возрастания и убывания сигнала, можно найти и другие свойства:

1. Псевдо-производная от константы равна нулю.

2. Псевдо-производная от прямой равна константе.
3. Псевдо-производная скачка (разрыва первого рода с точкой конечно разрыва) равна похожей по виду δ-функции.

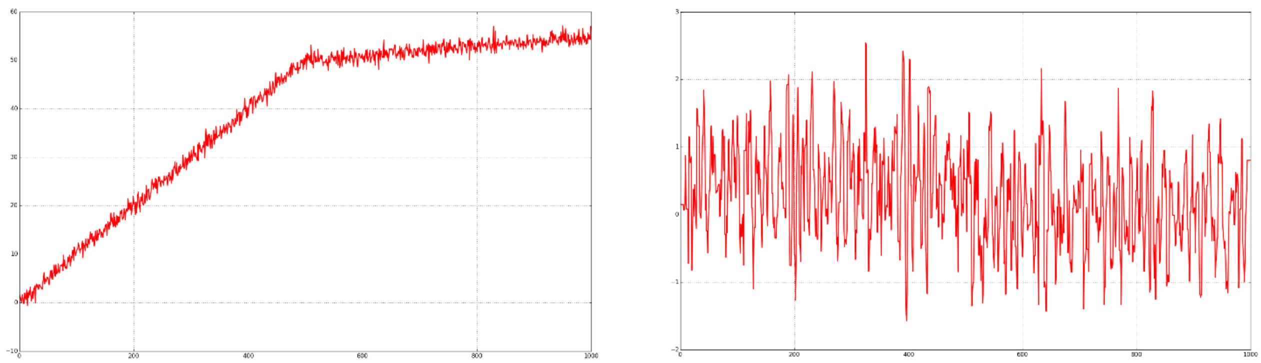
Для наглядности будем рассматривать смоделированные кусочно-линейные функции, состоящие из 1000 отсчетов и содержащие в себе случайную составляющую, в которых в момент времени $t = 500$ происходит одно из перечисленных событий: излом линии тренда, выброс или скачок.

2.6. Псевдо-производная временных рядов с изломами тренда

Пусть $X(t)$ – искусственный сигнал с изломом тренда, определяемый по формуле:

$$X(t) = \begin{cases} 0.1t + \varepsilon, & t < 500 \\ 50 + 0.01t + \varepsilon, & t \geq 500 \end{cases}, \qquad t = 0, \dots, 999$$

где ε – гауссовский белый шум со стандартным отклонением 1. Рассмотрим теперь псевдо-производную получившегося временного ряда (рисунок 2.6.1).



а) исходный сигнал б) псевдо-производная, $M = 10$

Рисунок 2.6.1. Пример псевдо-производной для сигнала, содержащего излома тренда

Чтобы определить значение наклона линии тенденции с помощью псевдо-производной в пределах окна, будем аппроксимировать тренд прямой вида $y = kx$, где k – искомая величина, то есть каждому окну сопоставим прямоугольную область, ширина которой равна M, а высота – полному приращению. Получаем, что угол наклона будет равен отношению полного приращения к длине окна или отношению значения псевдо-производной к половине окна:

$$k_i = \frac{\Delta ED_i + \Delta ID_i}{M} = \frac{2D_i}{M} = \frac{D_i}{\frac{M}{2}} \tag{2.6.1}$$

Из формулы $(2.6.1)$ следует, что когда сигнал не содержит в себе шума, то горизонтальные участки псевдо-производной прямо пропорциональны тангенсу угла наклона линии тенденции, что делает график псевдо-производной похожей на зашумленную ступенчатую функцию. В данном случае горизонтальные участки отсутствуют, поэтому коэффициент наклона линии тенденции пропорционален среднему значению псевдо-производной в интервале. Рассчитав математическое ожидание в промежутках $t < 500$ и $t \geq 500$, мы получим 0.4899 и 0.0578 соответственно. Следовательно, коэффициенты наклона в этих промежутках по формуле $(2.6.1)$ равны 0.098 и 0.0116, что приблизительно равно исходным данным.

Чтобы построить линию тренда непосредственно на графике исходного сигнала и посмотреть на результат аппроксимации в i-ом окне, будем искать прямую вида $y_i = k_i x + b_i$, где k_i определяется тем же способом. В качестве свободного коэффициента b_i можно, например, взять такую величину, при которой искомая прямая будет проходить через точку $A = (x_0, y_0)$, где $x_0 = \frac{(X(i)+X(i+M-1))}{2}$ – середина окна, а $y_0 = \frac{1}{M}\sum_{k=i}^{i+M-1} X(k)\, , i = 0, \dots, N - M - 1$ – математическое ожидание значений сигнала в текущем окне, таким образом:

$$b_i = \frac{1}{M}\sum_{k=i}^{i+M-1} X(k) - k_i \cdot \frac{M}{2} \tag{2.6.2}$$

Очевидно, что если для каждой точки временного ряда найти коэффициенты k и b по формулам $(2.6.1)$ и $(2.6.2)$ соответственно, то в качестве линии тренда мы получим привычную аппроксимацию скользящим средним. Чтобы найти аппроксимацию линии тенденции не только для текущего временного окна шириной, равной базе псевдо-производной, а для произвольного промежутка шириной L отсчетов, необходимо в формуле $(2.6.2)$ заменить M на L. Для нашего смоделированного временно ряда, где $L = 500$, получим:

$$b_1 = \frac{1}{L}\sum_{k=0}^{L-1} X(k) - k_1 \cdot \frac{L}{2} \approx 0.5293,$$

$$b_2 = \frac{1}{L}\sum_{k=L}^{N-1} X(k) - k_2 \cdot \frac{L}{2} \approx 49.5032$$

и графически аппроксимация линии тенденции выглядит следующим образом:

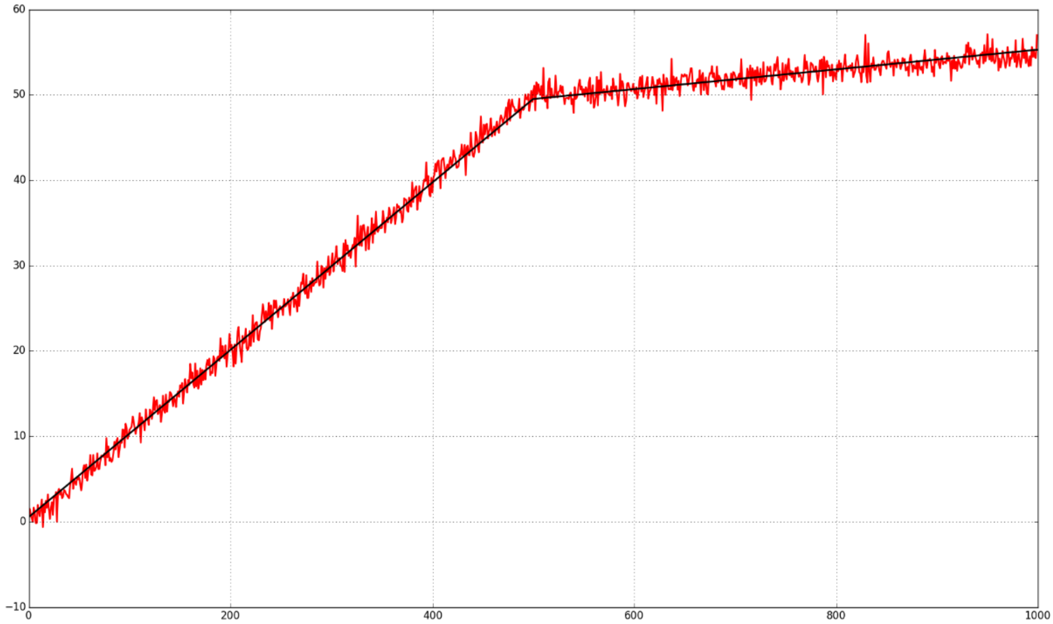


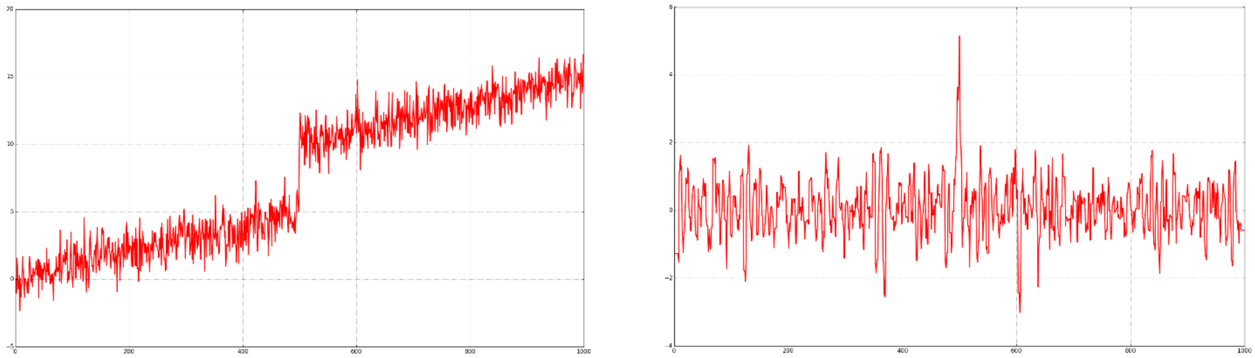
Рисунок 2.6.2. Аппроксимация линии тенденции временного ряда с помощью псевдо-производной

2.7. Псевдо-производная временных рядов, содержащих скачки

Пусть $X(t)$ – искусственный временной ряд, содержащий разрыв первого рода с точкой конечно разрыва (скачок), определяемый по формуле:

$$X(t) = \begin{cases} 0.01t + \varepsilon, & t < 500 \\ 0.01t + 10 + \varepsilon, & t \geq 500 \end{cases}, \qquad t = 0, \dots, 999$$

где ε – гауссовский белый шум со стандартным отклонением 1.



а) исходный сигнал б) псевдо-производная $X(t)$, $M = 10$

Рисунок 2.7.1. Пример псевдо-производной для сигнала, содержащего скачок

Рисунок 2.7.1 демонстрирует, что скачок в сигнале в момент времени $t = 500$ определился в виде выброса. Средние значения псевдо-производной в интервалах $t < 500$ и $t > 500$ колеблются возле значений 0.0683 и 0.0548, что соответствует коэффициентам наклона линии тенденции 0.0137 и 0.011. При условии, что исходный временной ряд не содержит шумовую компоненту, такой вид графика псевдо-производной будет похож на δ-функцию Дирака.

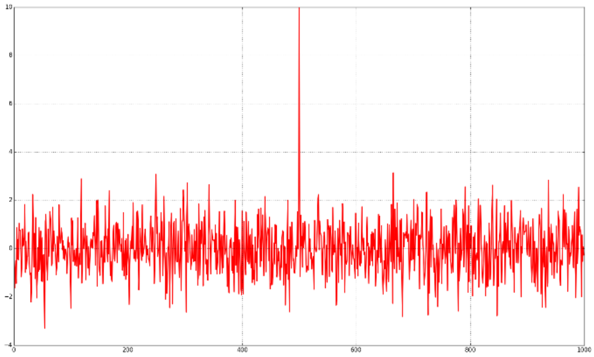
Стоит отметить, что выделение разрывов с использованием только определения псевдо-производной зачастую может не привести ни к каким результатам, так как данный подход работает только при наличии в анализируемом сигнале большого по модулю скачка, что приведет к явному появлению выброса в псевдо-производной. Иначе в случае, когда исходный временной ряд сильно зашумлен, псевдо-производная также будет зашумленной.

2.8. Псевдо-производная временных рядов, содержащих выбросы

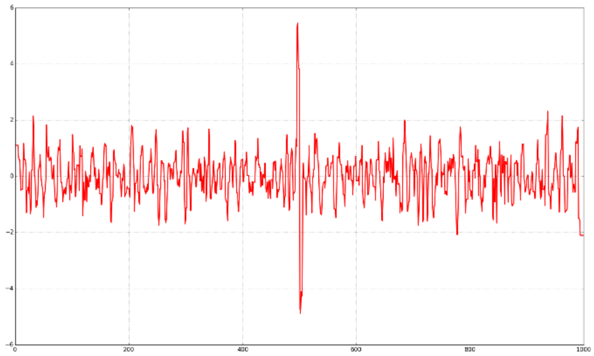
Смоделируем теперь временной ряд $X(t)$, содержащий выброс (для гладких функций он соответствует разрыву первого рода с устранимой точкой разрыва):

$$X(t) = \begin{cases} \varepsilon, & t \neq 500 \\ 10 + \varepsilon, & t = 500 \end{cases}, \qquad t = 0, \dots, 999$$

где ε – гауссовский белый шум со стандартным отклонением 1.



а) исходный сигнал



б) псевдо-производная $X(t)$, $M = 10$

Рисунок 2.8.1. Пример псевдо-производной для сигнала, содержащего выброс

Так как шум содержит множество выбросов малой амплитуды, то определить данный вид события становится сложнее. На графике псевдо-производной (рисунок 2.8.1) можно выделить положительный и отрицательный выбросы в малом интервале времени, содержащем отсчет $t = 500$, напоминающие в случае отсутствия шума кусочно-линейную аппроксимацию производной функции Гаусса и соответствующие смоделированному событию. Первый всплеск говорит нам о том, что в окне R произошло событие, в то время как в окне L все еще «покой», а отрицательный всплеск соответствует возвращению сигнала после события на линию тренда. Если бы выброс имел меньшую амплитуду, то он был бы трудно различим на графике псевдо-производной.

Рассмотрев три примера событий, можно сказать, что все они по-своему определяются с помощью псевдо-производной, что позволяет, взглянув на ее график, определить тип события. Таким образом, выделим следующие три свойства псевдо-производной:

- Переход разрыва в выброс.
- Переход выброса в положительный и отрицательный «всплески.
- Переход изменений угла наклона тренда в ступенчатую функцию.

Зачастую нам в руки попадают сильно зашумленные сигналы, что делает и их псевдо-производную также зашумленной. В таких случаях иногда уже не представляется возможным визуальное определения типа произошедшего в сигнале события. Угол наклона линии тренда сигнала не будет пропорционален значению по оси ординат горизонтальных участков псевдо-производной, так как они в принципе могут отсутствовать. Он пропорционален среднему значению величины D.

На практике встречаются сигналы, более сложные по своей структуре и сильно отличающиеся от тех, что были рассмотрены. В них могут содержаться несколько событий одновременно, что не позволит с такой легкостью определять события, лишь взглянув на график их псевдо-производной. Однако подход совсем не изменяется: с помощью первой псевдо-производной можно находить локальные и глобальные точки экстремума, варьируя параметром M. Чем он больше, тем меньше локальных точек экстремума будет определено. Вторую псевдо-производную по аналогии с производной второго порядка можно использовать для определения точек перегиба анализируемого сигнала, которые также соответствуют максимальным полным приращениям сигнала.

2.9. Кусочно-ступенчатая аппроксимация временного ряда

Кусочно-ступенчатая аппроксимация (КСА) сигнала сама по себе, как правило, является очень грубой аппроксимацией исходных данных. Однако ее можно рассматривать в качестве оценки скелета временного ряда, то есть структуры, хранящей в себе наиболее важную информацию об исходном объекте. Применительно к анализу временных рядов GPS такая аппроксимация может помочь обнаружить резкие изменения среднего значения на фоне шума и дать оценку степени скачкообразности исследуемого сигнала.

Существует несколько методов построения кусочно-ступенчатой аппроксимации, наиболее популярными из которых являются методы с использованием как дискретных, так и непрерывных вейвлет-преобразований. Дискретное вейвлет-преобразование (ДВП) Хаара является примером кусочно-ступенчатой аппроксимации временного ряда в силу вида материнской вейвлет-функции, являющейся кусочно-постоянной функцией. Задается вейвлет Хаара по следующей формуле:

$$\psi = \begin{cases} 1, & 0 \leq x \leq 0.5 \\ -1, & 0.5 \leq x \leq 1 \\ 0, & x \notin [0,1) \end{cases} \qquad (2.9.1)$$

Простота реализации построения кусочно-ступенчатой аппроксимации с помощью дискретного вейвлет-преобразования Хаара является положительной стороной, однако также данный подход имеет несколько отрицательных сторон. Единственным параметром данного метода является количество обнуляемых коэффициентов вейвлев-разложения GPS сигнала. В качестве критерия выбора процента обнуляемых коэффициентов можно использовать алгоритм Донохо-Джонстона [54]. Ниже изображен пример кусочно-ступенчатой аппроксимации методом Хаара, коэффициент Доноха-Джонстона ≈ 0.9856, среднеквадратическая ошибка аппроксимации 3.101.

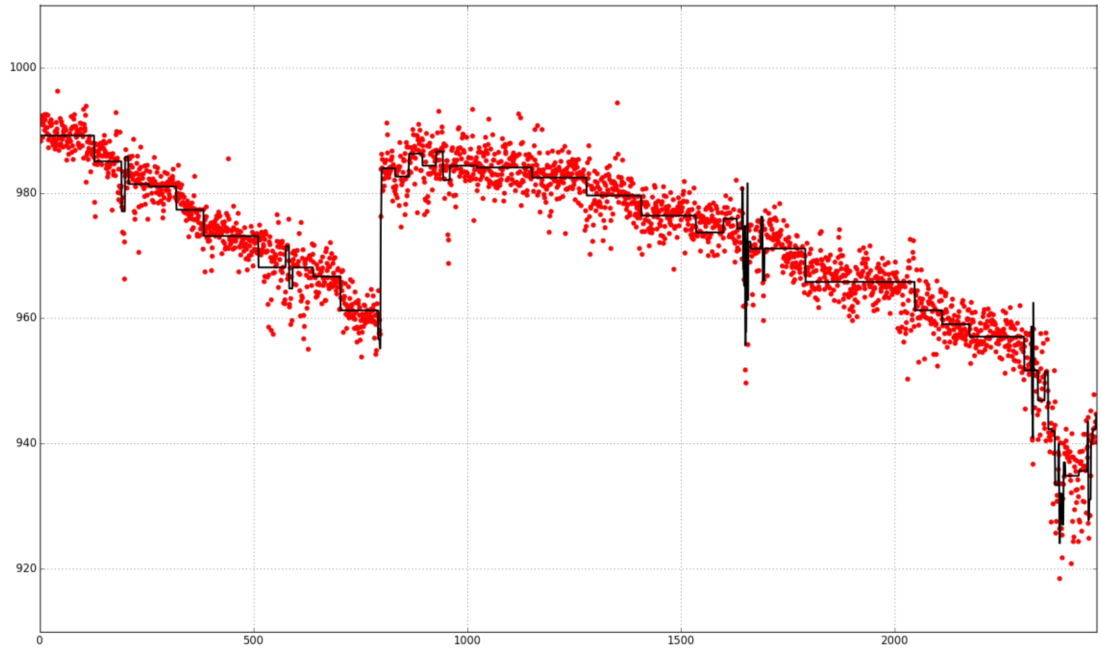


Рисунок 2.9.1. Пример КСА сигнала GPS, построенной с помощью ДВП Хаара.

Другой алгоритм построения кусочно-ступенчатой аппроксимации так же основан на использовании вейвлет-преобразования, только не дискретного, а непрерывного (НВП). Данный метод называется WTMM (wavelet transform modulus maxima), или ММВП (максимум модулей вейвлет-преобразования) [22], [82].

Определим масштабно-зависимое сглаживание сигнала с помощью некоторого ядра усреднения следующей формулой:

$$\bar{X}(s,t) = \frac{\int_{-\infty}^{+\infty} X(t+sv)\cdot\psi_0(v)dv}{\int_{-\infty}^{+\infty}\psi_0(v)dv}, \qquad (2.9.2)$$

где $s > 0$ – масштаб, а $\psi_0(t)$ является некоторой быстро стремящейся к нулю функцией. Для последующих вычислений в качестве нее мы будем использовать функцию Гаусса: $\psi_0(t) = e^{-t}$. Ядро НВП определим по формуле:

$$\psi_n(t) = (-1)^n \cdot \frac{d^n\psi_0(t)}{dt^n} \equiv (-1)^n \cdot \psi_0^{(n)}(t). \qquad (2.9.3)$$

Если воспользоваться формулой интегрирования по частям, то легко получить формулу для производной заданного порядка n от функции $(2.9.1)$:

$$\mathrm{c}_n(t,s) = \frac{1}{n!}\frac{d^n\bar{X}(s,t)}{dt^n} = \frac{\int_{-\infty}^{+\infty} X(t+sv)\cdot\psi_0(v)dv}{s^n\int_{-\infty}^{+\infty}\psi_0(v)dv}, \qquad (2.9.4)$$

где $n!$ – коэффициент Тейлора.

Из названия метода ММВП следует, что вторым этапом является определение точек максимума модуля t_k . Для порядка производной $n \geq 1$ они вычисляется как локальный максимум коэффициентов $(2.9.4)$. В частном случае, при $n = 0$, точки максимума модуля определяются как точки минимумов и максимумов сглаженного сигнала. Объединение всех найденных точек образует скелет непрерывного вейвлет-преобразования. Пример такого скелета приведен на рисунке 2.9.2 ниже.

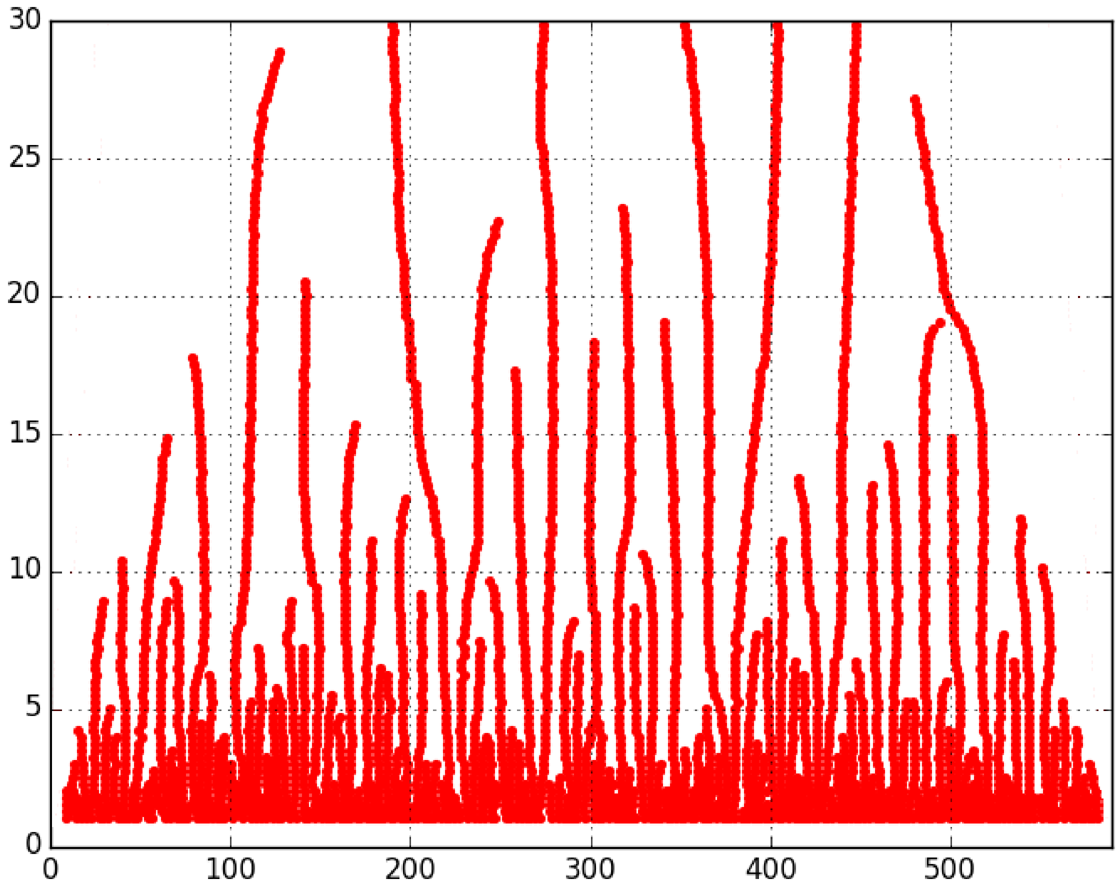


Рисунок 2.9.2. Пример скелета непрерывного вейвлет-преобразования

Затем кусочно-ступенчатую аппроксимацию можно построить по следующей формуле:

$$Stepwise(t) = \mu(\{X(t) \mid t_k \leq t \leq t_{k+1}\}), \qquad k = 0, \dots, M-1 \qquad (2.9.5)$$

где M – общее число локальных экстремумов коэффициентов непрерывного вейвлет-разложения, а μ – усредняющая функция, в качестве которой можно взять либо математическое ожидание, либо медиану.

Если сравнивать по количеству входных параметров данный подход с дискретным преобразованием Хаара, то по сути единственным параметров данного алгоритма является значение масштаба s. Ниже на рисунке 2.9.3 изображен результат работы метода WTMM для порогового значения 34 со среднеквадратической ошибкой аппроксимации 5.5045.

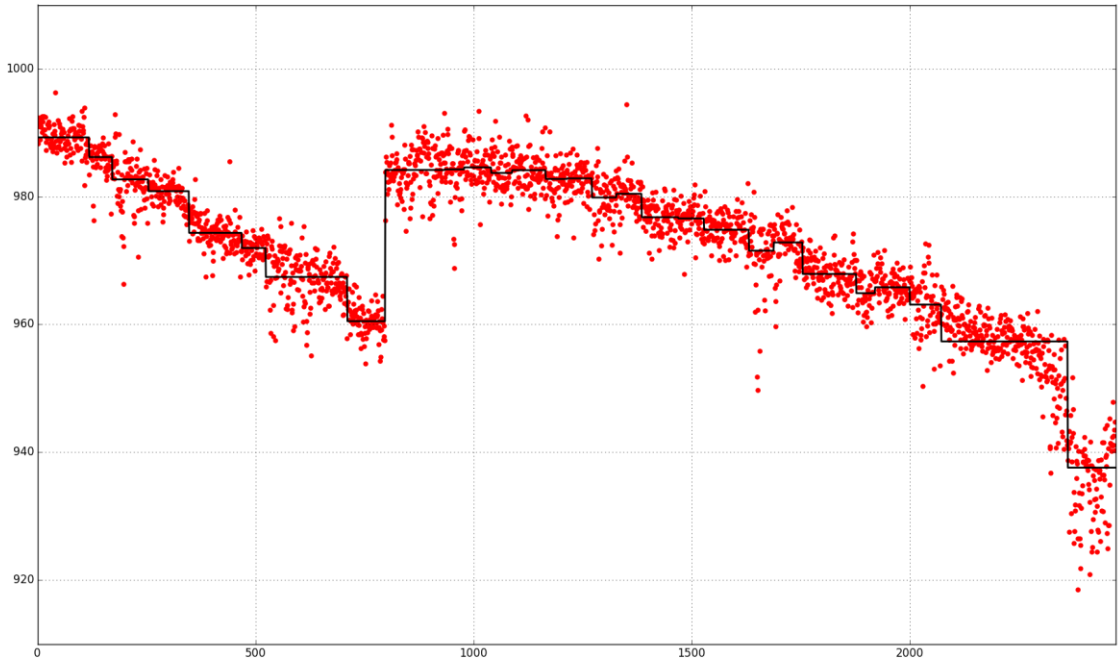


Рисунок 2.9.3. Пример КСА сигнала GPS, построенной с помощью WTMM

В качестве порогового значения масштаба выбрано такое максимальное значение, которое не приводит к резкому увеличению среднеквадратической аппроксимации.

В настоящей работе предлагается новый метод построения кусочно-ступенчатой аппроксимации, основанный на использовании введенного ранее понятия псевдо-производной. Алгоритм состоит из следующих трех шагов:

1. Рассчитать псевдо-производную исходного сигнала.
2. Удалить из псевдо-производной скользящее среднее в окне, имеющим радиус, равный базе псевдо-производной.
3. Найти нули получившегося временного ряда и построить «ступеньки» в интервалах между ними, где каждая ступенька – это медианное значение исходного сигнала в интервалах между нулями псевдо-производной (альтернативный вариант этого шага будет изложен позже).

Рассмотрим подробнее пункты 1-3. Основной целью данного алгоритма является поиск нулей псевдо-производной, потому что они, как и нули производной, указывают на точки экстремума сигнала, то есть на точки, в которых функция меняет свой характер: убывает или возрастает. В случае, когда сигнал содержит наклонный тренд и колебания сигнала относительно него малы, то псевдо-производная может вовсе не пересекать ось абсцисс. Как говорилось выше, коэффициент наклона линии тенденции временного ряда пропорционален среднему значению псевдо-производной, поэтому поиск точек экстремума исходного сигнала произведем путем вычитания скользящего среднего из значений псевдо-производной.

Удалив скользящее среднее и найдя нули получившегося сигнала, мы можем в качестве «ступенек» взять медианное или среднее арифметическое значения сигнала в промежутках между нулями. Поскольку мы рассматриваем дискретное время, то под «нулями» сигнала следует понимать середины двух соседних отсчетов, которые имеют разные знаки. Таким образом, мы получим кусочно-ступенчатую аппроксимацию исходного сигнала. Обозначим через $Stepwise(t)$ построенную таким образом кусочно-ступенчатую аппроксимацию временного ряда, где $t = 1, \dots, N$ – целочисленный временной индекс, N – длина временного ряда в числе отсчетов.

Такой алгоритм построения кусочно-ступенчатой аппроксимации был предложен изначально и изложен в [21]. Можно использовать какой-либо другой вариант построения кусочно-ступенчатой функции на основе псевдо-производной, например:

1. Строить «ступеньки» не в интервалах возрастая и убывания сигнала, а в интервалах, соединяющих середины промежутков знакопостоянства псевдо-производной. Именно такой вариант построения кусочно-ступенчатой аппроксимации будет использоваться далее в работе.
2. Построить кусочно-ступенчатую аппроксимацию по нулям второй псевдо-производной. Так как первая псевдо-производная хранит в себе информацию о коэффициенте наклона линии тенденции в виде среднего значения в выбранном временном промежутке, то вычисление второй псевдо-производной позволяет «неявно» устранить тренды из анализа

колебаний исходного временного ряда. Нули второй псевдо-производной, аналогично второй классической производной, соответствуют экстремумам первой псевдо-производной, то есть точкам перегиба.

Наличие трех разных алгоритмов делает процесс построения кусочно-ступенчатой аппроксимации более гибким и дает возможность подобрать наилучшую аппроксимацию для исследуемого временного ряда. Главной отличительной чертой методов, предложенных выше под пунктами 1-2, является то, что «ступеньки» аппроксимации будут построены в интервалах не между экстремумами исходного сигнала, а между точками перегиба, что в каких-то случаях делает аппроксимацию более верной.

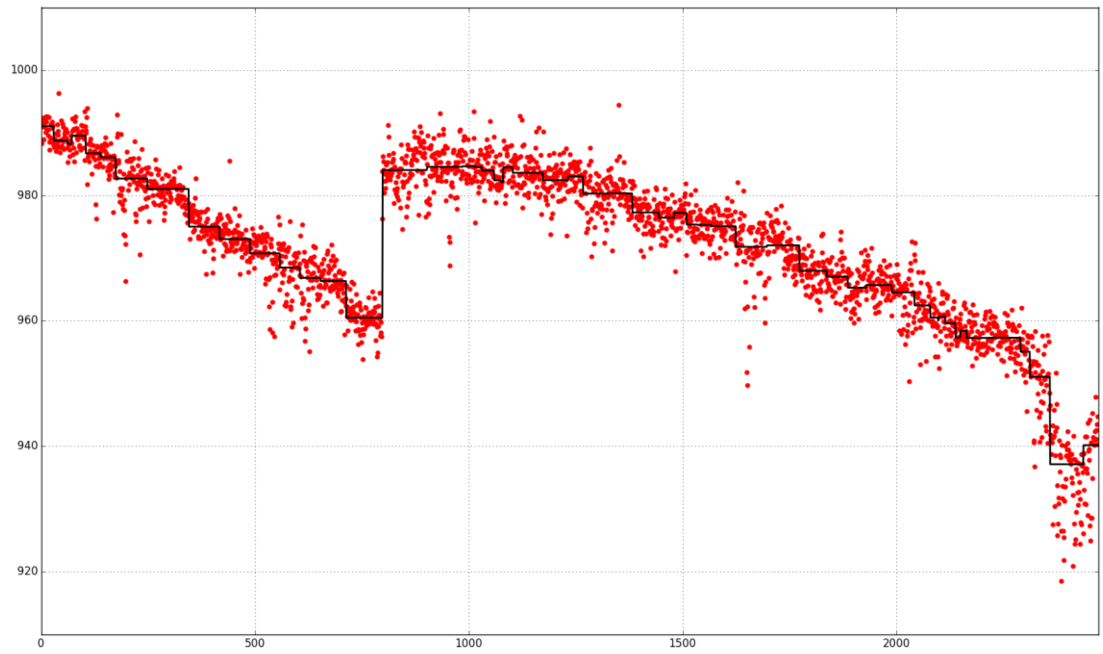


Рисунок 2.9.4. Пример КСА сигнала GPS, построенной методом псевдо-производной

Значение базы псевдо-производной выбрано из тех же соображений, что и пороговое значение масштаба для метода WTMM, и равно 100, среднеквадратическая ошибка 3.6221

Сравнивая описанные выше три метода, можно сделать следующий вывод: все они является однопараметрическими, однако метод Хаара является наименее устойчивым к выбросам, но приводит к наименьшему значению среднеквадратической ошибки, в то время как метод построение кусочно-ступенчатой аппроксимации методом WTMM и с помощью псевдо-производной являются робастными методами в силу алгоритма построения, в котором используется медиана в качестве робастной оценки среднего значения сигнала. Таким образом, метод Хаара не является подходящим для использования оценки скачков среднего уровня, так получившаяся кусочно-ступенчатая аппроксимация может содержать большое число выбросов.

Метод псевдо-производной имеет некоторые преимущества перед методом WTMM: он быстрее, что играет существенную роль при анализе большого числа сигналов, и рост среднеквадратической ошибки происходит медленнее, что позволяет более плавно изменять кусочно-ступенчатую аппроксимацию, варьируя единственным входным параметром – базой псевдо-производной. Какого-либо функционального соответствия между параметром масштаба метода WTMM, количество обнуляемых коэффициентов вейвлет-преобразования Хаара и величиной базы псевдо-производной установить не удалось, в следствие чего нет возможности составить точную сравнительную характеристику времени реализации этих трех алгоритмов. Тем не менее ниже в таблице представлено усредненное время, требуемое для построения кусочно-ступенчатой аппроксимации с 50 разными наиболее подходящими входными параметрами для построения кусочно-ступенчатой аппроксимации: Хаар – процент обнуляемых коэффициентов в отрезке $[0.95, 0.99]$, WTMM – пороговое значение масштаба в отрезке $[10, 30]$, псевдо-производная – величина базы в отрезке $[50, 100]$. Результаты представлены для того временного ряда GPS, аппроксимации для которого вычислялись выше.

Таблица 2.9.1. Сравнение времени (в секундах) работы алгоритмов.

Хаар	**WTMM**	**Псевдо-производная**
1.6651	3.0021	1.4861

Из таблицы 2.9.1 видно, что для заданных диапазонов входных параметров псевдо-производная имеет наилучший результат, однако стоит отметить, что для более малых значений базы время реализации увеличивается, но все же меньше времени, требуемого методом WTMM, приблизительно в 1.5 раза.

2.10. Мера скачкообразности сигнала

Размер базы M псевдо-производной будет влиять на количество аппроксимирующих «ступеней». Чем меньше M, тем более детальной является кусочно-ступенчатая аппроксимация сигнала. В то же время интерес представляют далеко не все скачки кусочно-ступенчатой аппроксимации, а лишь те из них, которые соответствуют наибольшей амплитуде скачка. Однако следует учесть, что на величины скачков кусочно-ступенчатой аппроксимации при заданной базе вычисления псевдо-производной влияет также шум, то есть часть высокоамплитудных скачков кусочно-ступенчатой аппроксимации обусловлена влиянием шума. По этой причине для подавления влияния шума далее используется усреднение большого числа кусочно-ступенчатых аппроксимаций, построенных для множества значений базы M, поскольку в точке, где имеется разрыв сигнала, кусочно-ступенчатая аппроксимация будет иметь разрыв для большинства значений M.

Для этой цели введем два параметра M_{min} и M_{max} – минимального и максимального значений базы M для вычисления псевдо-производной. Вычислим сигнал $\overline{Stepwise}(t)$, равный среднему от всех кусочно-ступенчатых аппроксимаций:

$$\overline{Stepwise}(t) = \frac{1}{M_{max} - M_{min} + 1} \sum_{M=M_{min}}^{M=M_{max}} Stepwise_M(t). \qquad (2.10.1)$$

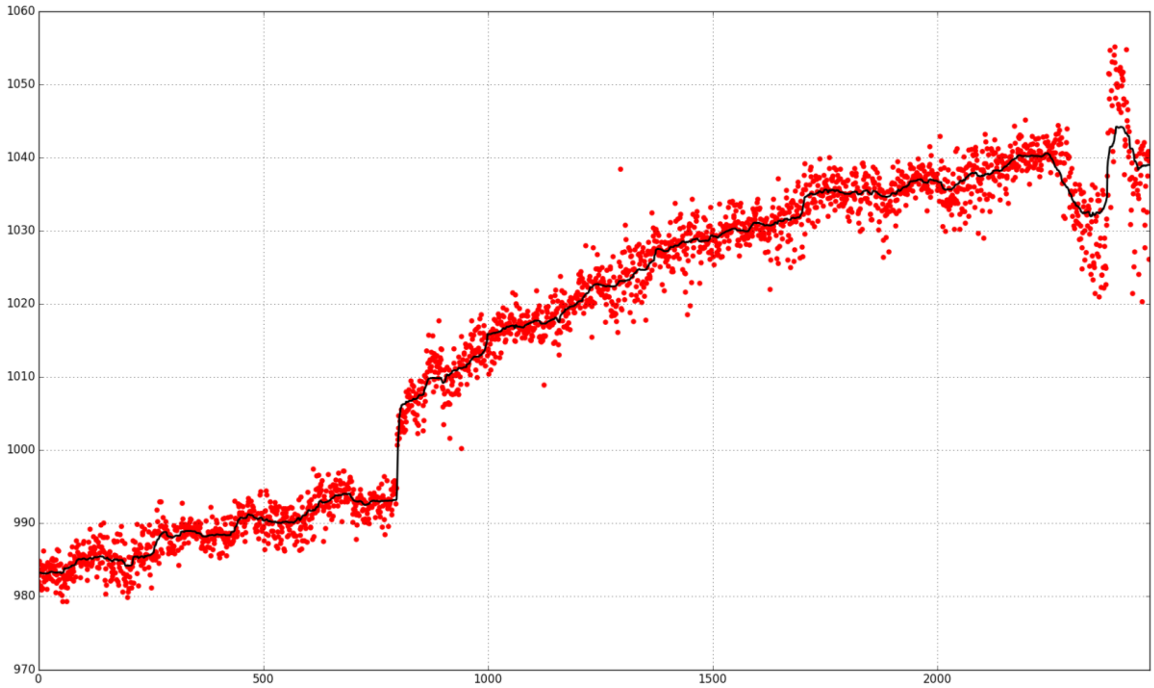


Рисунок 2.10.1. Усредненная КСА. База псевдо-производной изменялась от 10 до 200

Обозначим абсолютное приращение усредненной кусочно-ступенчатой аппроксимации

$$R(t) = |\overline{Stepwise}(t) - \overline{Stepwise}(t-1)| \qquad (2.10.2)$$

Поставим цель сформулировать количественный критерий наличия в зашумленном временном ряду значимой скачкообразной составляющей. Вариации вспомогательного временного ряда $R(t)$ несут в себе информацию о наличии в исходном временном ряде значимых скачков среднего уровня.

Далее в качестве меры скачкообразности временного ряда введем нормализованную энтропию этих превышений:

$$En = -\frac{1}{\ln N^+} \sum\nolimits_t^+ p(t) \cdot \ln\big(p(t)\big), \qquad (2.10.3)$$

где символ $\sum_t^+$ означает суммирование лишь по тем временным индексам t, для которых $R(t) > 0$, N^+ – общее число таких индексов, $p(t) = R(t)/\sum_\tau^+ R(\tau)$ – величины, преобразованные к значениям вероятностей. По построению величина En удовлетворяет ограничению $0 \leq En \leq 1$, и чем ближе ее значение к 1, тем более хаотичными являются вариации величин $R(t)$. Временной ряд, не имеющий значимых скачков, характеризуется хаотичным изменением величин $R(t)$ и, как следствием этого, значения нормализованной энтропии для него выше, чем для ряда, содержащего скачки. Пример такого ряда изображен ниже, $En = 0.8809$.

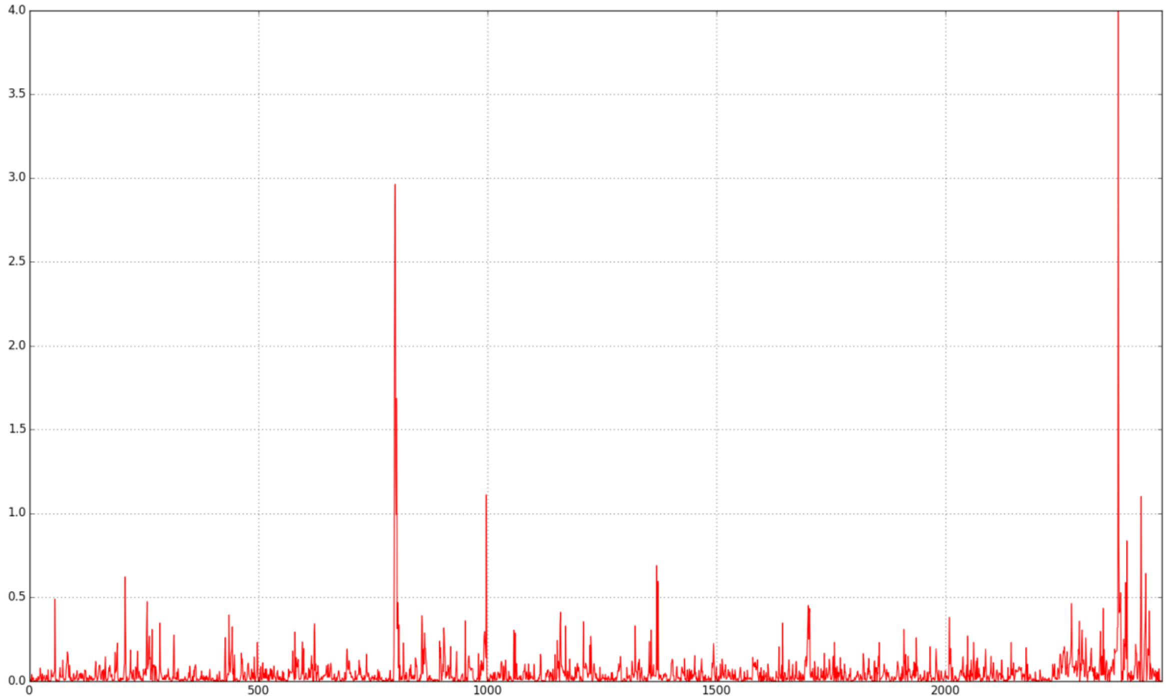


Рисунок 2.10.2. Пример временного ряда $R(t)$

2.11. Мера выбросов во временных рядах

Несмотря на то, что при введении понятия псевдо-производной использовалось одно скользящее окно, фактически анализ проводился в двух следующих друг за другом окнах. Такой подход является очень удобным, так как позволяет учитывать большее количество факторов и их изменения. Схожий подход будет использоваться при поиске выбросов во временных рядах.

Некоторые существующие методы выделения выбросов основаны на определении расстояния от текущего отсчета временного ряда до всех остальных. Чем больше это расстояние, тем выше вероятность, что данный отсчет является выбросом. Однако вычисление расстояний для каждого момента времени при условии, что временные ряды могут достигать сотни тысяч отсчетов, и рядов может быть очень велико, требует большого количества времени и становится непригодным для использования. В следствие чего, возникает идея использовать вместо расстояния какую-либо иную величину, которая так же или похожим образом, будет отражать степень отклонение текущего отсчета от всей выборки. Одной из таких величин является стандартное отклонение, которое очень чувствительно к выбросам. То есть если взять временной ряд, не содержащий выбросов, вычислить стандартное отклонение, а затем вычислить стандартное отклонение, добавив выброс, мы увидим относительно большую разницу. Таким образом, если в каждый момент времени вычислять разницу стандартных отклонений слева и справа от текущего отсчета, то мы сможем сделать предположение о том, является ли значение выбросом или нет. Ниже подробно описан данный алгоритм выделения выбросов во временных рядах.

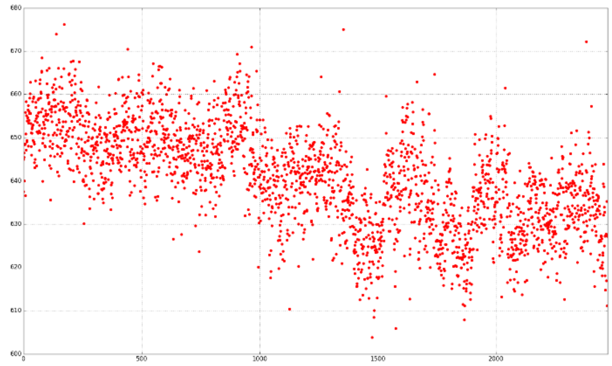
Пусть $X(t)$ — временной ряд, где $t = 0, \dots, N-1$. В каждый момент времени $t = i$ нас будет интересовать, насколько текущее значение сигнала отличается от значений, взятых слева и справа. В качестве сравнительной характеристики рассмотрим следующие разности:

$$\begin{aligned} L_i &= \left|\sigma_L^i - \sigma_L^{i-1}\right| \\ R_i &= \left|\sigma_R^i - \sigma_R^{i+1}\right|, \end{aligned} \qquad (2.11.1)$$

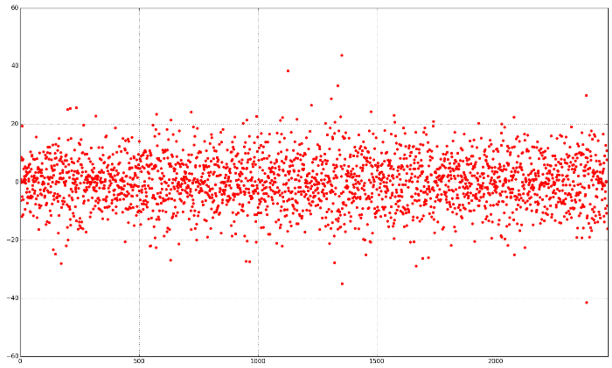
где σ_L^i, σ_R^i – стандартные отклонения, рассчитанные для значений $t \in [0, i]$ и $t \in [i, N-1]$ соответственно. Величины σ_L^{-1} и σ_R^N, выходящие за границы возможных значений t, положим равными нулю. Пояснение этого условия будет изложено позже. Таким образом, в каждый момент времени мы имеем абсолютные значения разностей стандартных отклонений, которые являются более показательными в отличие от обычной разности, так как стандартное отклонение само по себе чувствительно к малым отклонениям. Величины $(2.11.1)$ показывают, сильно ли отклоняется текущее значение временного ряда относительно левой и правой частей сигнала в отдельности. Чтобы определить меру отклонения относительно всего сигнала в совокупности, рассмотрим взвешенную сумму приращений стандартных отклонений $(2.11.1)$:

$$W_t = \frac{t}{N} L_t + \frac{N-1-t}{N} R_t \qquad (2.11.2)$$

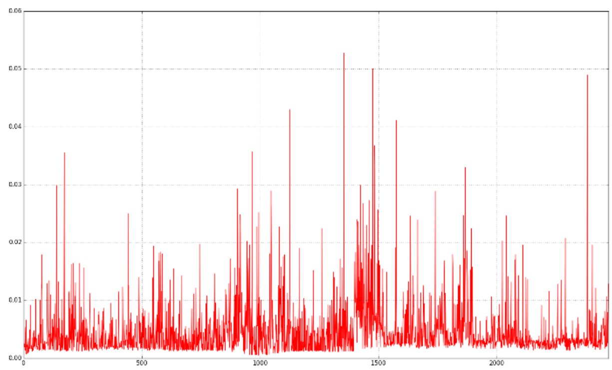
В момент времени $t = 0$, когда возможно сравнение текущего значения только со значениями справа, мы видим, что первое слагаемое суммы $(2.11.2)$ обращается в ноль, поэтому значение величины σ_L^{-1} может быть любым, например, 0. В другом крайнем случае при $t = N - 1$, аналогично, второе слагаемое суммы $(2.11.2)$ равно нулю, поэтому σ_R^N также можно положить равным 0. Пример величины W_t для сигнала GPS (смещения вертикальных координат) со станции G001 изображен ниже.



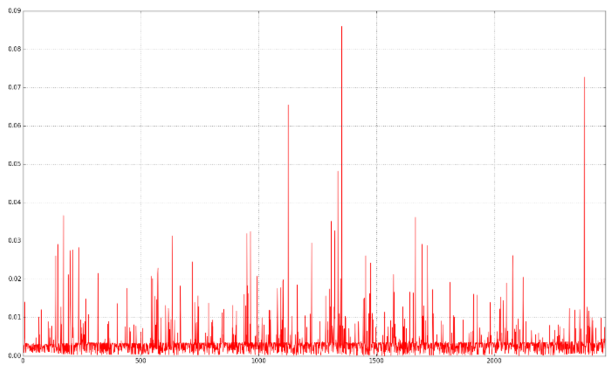
а) Исходный сигнал



б) Приращения исходного сигнала



в) Величина W_t для исходного сигнала



г) Величина W_t для приращений сигнала

Рисунок 2.11.1. Пример величины W_t

Величина W_t чувствительна к наличию трендов во временных рядах, поэтому ее вычисление непосредственно из исходных сигналов может оказаться сильно зашумленным и малоинформативным (в). Однако если вычислять меру выбросов для приращений исходного временного ряда, то можно отчетливо увидеть (г) полосу низких значений, четко отделенную от выбросов. Так как выбросов в сигнале меньше, чем прочих значений (именно поэтому мы их и называем выбросами), то можно отделить их путем построения гистограммы вспомогательного ряда W_t:

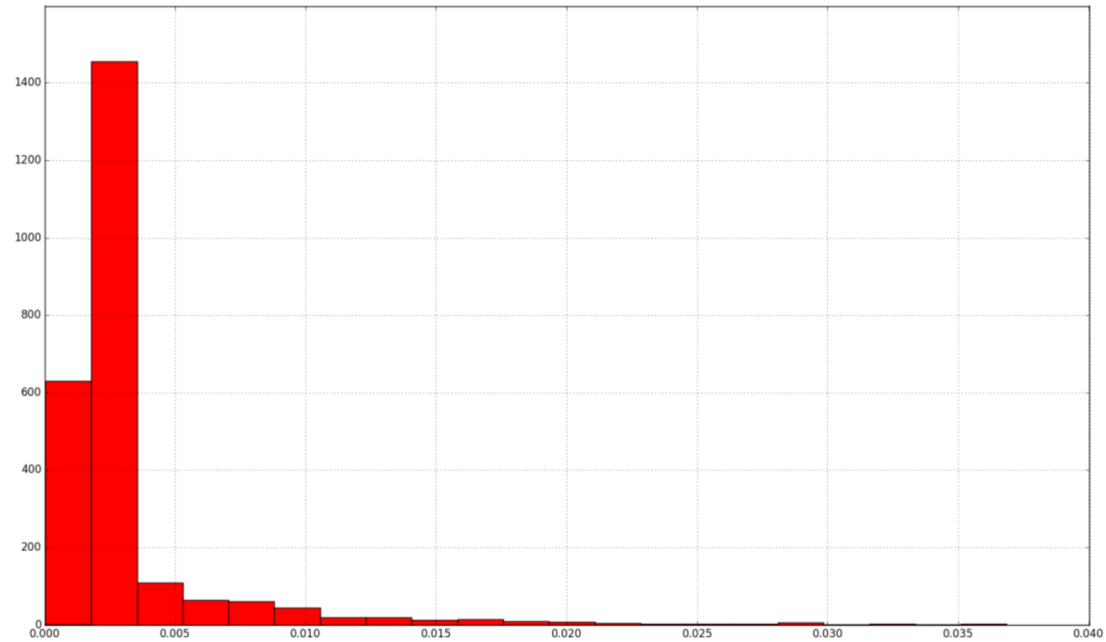


Рисунок 2.11.2. Гистограмма выбросов

Из рисунка 2.11.2 мы видим, что значения ряда W_t, колеблющиеся возле нуля, действительно преобладают, то есть выбросы являются более редким явлением в сигнале. Возьмем в качестве порога него правую границу интервала, в которой гистограмма W_t достигает своего максимума. В случае, если встречается несколько одинаковых максимумов, то выбираем наиболее удаленную от нуля границу интервала гистограммы. Выбросами временного ряда W_t будем считать все отсчеты, значения которых превышают найденный порог. Изменяемым параметром поиска порогового значения является число бинов гистограммы: чем оно выше, тем выше детализация оценки плотности распределения выбросов. Таким образом, может появиться ложный максимум гистограммы. Одной из наиболее распространенных оценок оптимального числа бинов, которая также используется в статье, является [43].

$$n = [\sqrt{N}], \tag{2.11.3}$$

где $[x]$ – обозначает целую часть числа x.

Подбор критерия определения оптимального количества бинов гистограммы выбросов W_t является наиболее важным и сложным этапом анализа. Ниже продемонстрирован результат расчета порогового значения, отделяющего выбросы от резких остальных значений W_t.

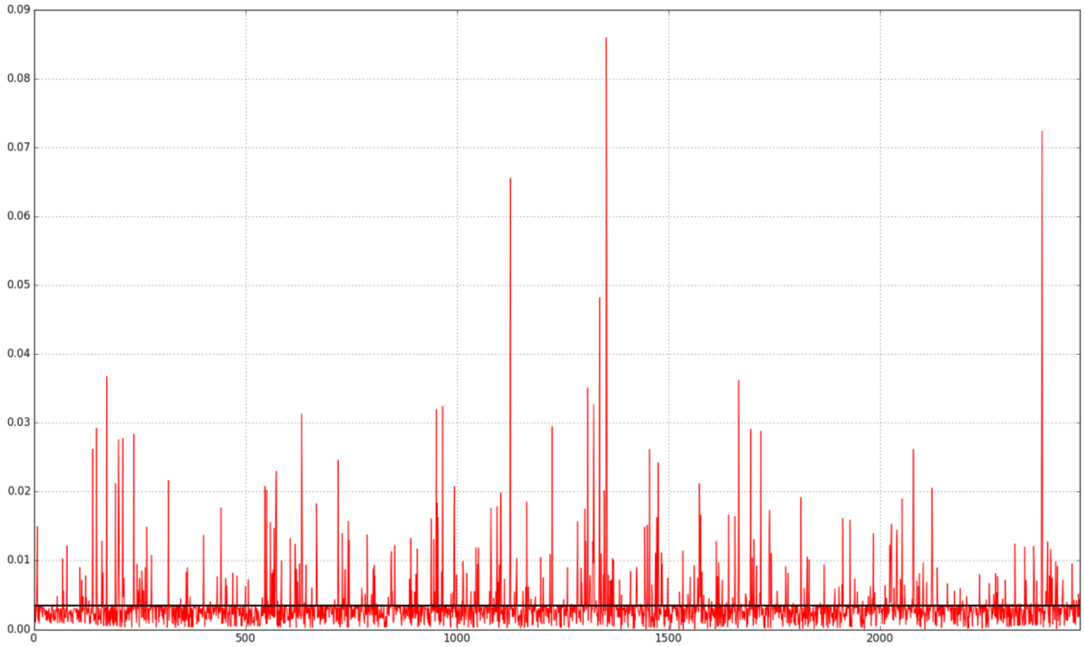
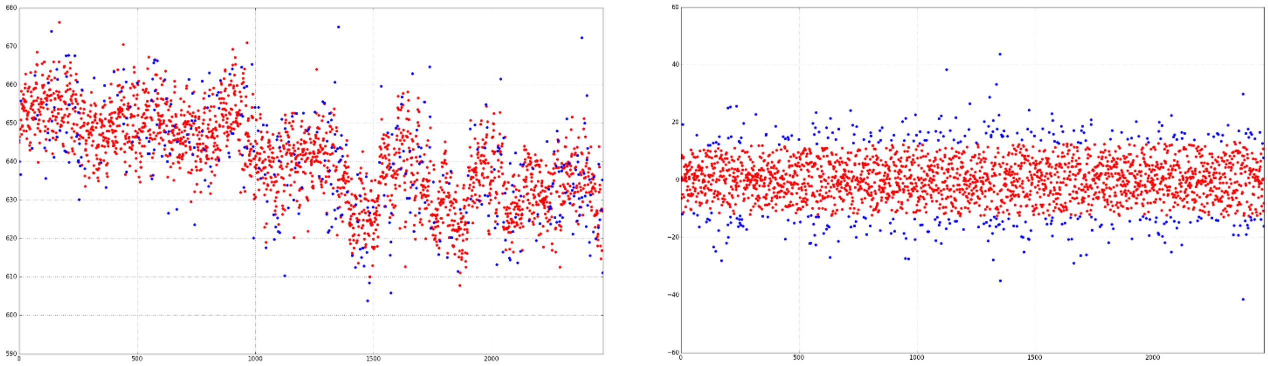


Рисунок 2.11.3. Определение порогового значения для определения выбросов временного ряда

Данный подход не является схожим с классическими методами выделения выбросов, так как основной его целью не является выделения всех выбросов временного ряда, а поиск таких отсчетов, которые наиболее сильно влияют на стандартные отклонения исходного сигнала, основываясь на определении выбросов вспомогательного временного ряда. Поэтому фактически величины W_t могут как содержать в себе те отсчеты, визуально сильно отклоняющиеся от всей выборки, так и не содержать их. Порядок значимости может регулироваться с помощью количества бинов гистограммы выбросов. Ниже изображены графически определенные выбросы величины W_t для вертикальных смещений станции G001.



а) Исходный сигнал б) Приращения исходного сигнала

Рисунок 2.11.4. Пример определения значений, сильно влияющих на стандартное отклонение приращений. Синим цветом обозначены выбросы W_t, а красным – значения, не содержащие выбросов

В общем случае вычисление W_t для приращений исходного временного ряда является более корректным, так как выброс, как правило, – это единичное резкое отклонение, таким образом, это отклонение можно увидеть, перейдя к приращениям. Далее вычисление меры выбросов будет проводиться только для приращений временных рядов.

Покажем, что если временной ряд не является константой, то величина W_t положительна для любого значения t. Пусть $X(t),\ t = 0, \dots, N-1$ – временной ряд. Если $X(t) -$ константа, то $W_t = 0$. Это очевидно, так как все приращения временного ряда также являются константой, то есть $X_{inc}(t) = C = 0 \Rightarrow W_t = \frac{t}{N} \cdot |L_\sigma^t - L_\sigma^{t-1}| + \frac{N-t-1}{N} \cdot |R_\sigma^t - R_\sigma^{t+1}| = 0$, так как $L_\sigma^t = L_\sigma^{t-1} = R_\sigma^t = R_\sigma^{t+1} = 0$.

Предположим, что хотя бы одно значение в момент времени τ ряда $X(t)$ отлично от всех, тогда приращения $X_{inc}(t)$ отлично от нуля в момент времени $\tau - 1$. Таким образом, из формулы $(2.11.2)$ следует, что хотя бы одно из слагаемых положительно, а, следовательно, W_t также положительна, что и требовалось доказать.

Чтобы дать количественную оценку меры выбросов используем также, как и в случае определения меры скачкообразности временных рядов, понятие нормализованной энтропии. Для этого перейдем от самих выбросов к вероятностям:

$$p(t) = \frac{W_t}{\sum_k W_k} \tag{2.11.4}$$

Из доказанного выше утверждения следует, что нормализованная энтропия определена для любого t:

$$En = -\frac{1}{ln(N-1)} \sum_t p(t) \cdot ln\big(p(t)\big) \tag{2.11.5}$$

В аргументе логарифма стоит $N - 1$, так как мы перешли к приращениям, таким образом, число отсчетов уменьшилось на единицу. Значение нормализованной энтропии временного ряда W_t, изображенного на рисунке 2.11.3, равно 0.9454.

Идея использования нормализованной энтропии в качестве одного из основных критериев оценки нарушения спокойного состояния земной поверхности заключается в том, что при подготовке к сильному землетрясению происходит синхронизация сейсмических шумов. Такая синхронизация позволяет говорить о появлении некоторой структуризации временных рядов GPS и об уменьшении степени беспорядочности системы. Таким образом, нас будут интересовать наиболее низкие значения нормализованной энтропии временных рядов GPS, так как именно они соответствуют меньшей степени беспорядка системы.

2.12. Рекуррентные формулы для стандартного отклонения

При анализе длинных временных рядов процесс вычисления W_t может занять очень много времени, если вычислять на каждом этапе стандартное отклонение по определению. Чтобы ускорить данный процесс выведем рекуррентную формулу, что позволит легко находить разности $(2.11.1)$. Обозначим для удобства $X_t = X(t)$ – значения временного ряда и возьмем выражения из формулы $(2.11.1)$, стоящие под знаком модуля, то есть

$$L = \sigma_L^t - \sigma_L^{t-1}, \qquad R = \sigma_R^t - \sigma_R^{t+1}. \tag{2.12.1}$$

Для начала выведем рекуррентную формулу для L.

$$s_{t-1} = \sigma_L^{t-1} = \sqrt{\frac{1}{t}\sum_{k=0}^{t-1}(X_k - m_{t-1})^2},$$

$$s_t = \sigma_L^t = \sqrt{\frac{1}{t+1}\sum_{k=0}^{t}(X_k - m_t)^2}, \tag{2.12.2}$$

где

$$m_{t-1} = \frac{1}{t}\sum_{k=0}^{t-1} X_k\,, \qquad m_t = \frac{1}{t+1}\sum_{k=0}^{t} X_k\,, \qquad \sigma_L^0 = 0$$

математические ожидания, которые также вычисляются по формуле

$$m_t = \frac{tm_{t-1} + X_t}{t+1} = \frac{tm_{t-1}}{t+1} + \frac{X_t}{t+1}, \tag{2.12.3}$$

тогда их разница равна

$$m_t - m_{t-1} = \frac{t}{t+1}m_{t-1} + \frac{1}{t+1}X_t - m_{t-1} = \frac{X_t - m_{t-1}}{t+1} \tag{2.12.4}$$

Возведем в квадрат (2.12.2), чтобы избавиться от квадратного корня и выделим квадрат предыдущего значения стандартного отклонения.

$$\begin{aligned} s_t^2 &= \frac{1}{t+1}\sum_{k=0}^{t}(X_k - m_t)^2 = \frac{1}{t+1}\sum_{k=0}^{t-1}(X_k - m_t)^2 + \frac{(X_t - m_t)^2}{t+1} \\ &= \frac{1}{t+1}\sum_{k=0}^{t-1}(X_k - m_{t-1} + m_{t-1} - m_t)^2 \\ &+ \frac{\left(\frac{t}{t+1}(X_t - m_{t-1})\right)^2}{t+1} = \end{aligned} \tag{2.12.5}$$

$$= \frac{1}{t+1}\left[\sum_{k=0}^{t-1}(X_k - m_{t-1})^2 - 2\sum_{k=0}^{t-1}(X_k - m_{t-1})(m_t - m_{t-1}) + \sum_{k=0}^{t-1}(m_t - m_{t-1})^2\right] + \frac{t^2(X_t - m_{t-1})^2}{(t+1)^3}$$

Из формулы разности математических ожиданий (2.12.4) следует:

$$\begin{aligned} s_t^2 &= \frac{1}{t+1}\left[\sum_{k=0}^{t-1}(X_k - m_{t-1})^2 - \frac{2(X_t - m_{t-1})}{t+1}\sum_{k=0}^{t-1}(X_k - m_{t-1}) \right. \\ &\quad \left. + \frac{1}{(t+1)^2}\sum_{k=0}^{t-1}(X_t - m_{t-1})^2\right] + \frac{t^2(X_t - m_{t-1})^2}{(t+1)^3} = \\ &= \frac{1}{t+1}\left[ts_{t-1}^2 + \frac{t(X_t - m_{t-1})^2}{(t+1)^2}\right] + \frac{t^2(X_t - m_{t-1})^2}{(t+1)^3} = \\ &= \frac{ts_{t-1}^2}{t+1} + \frac{t(X_t - m_{t-1})^2}{(t+1)^3} + \frac{t^2(X_t - m_{t-1})^2}{(t+1)^3} = \\ &= \frac{ts_{t-1}^2}{t+1} + t\left(\frac{X_t - m_{t-1}}{t+1}\right)^2 \end{aligned} \qquad (2.12.6)$$

Таким образом, рекуррентная формула для стандартного отклонения имеет вид

$$\begin{aligned} \sigma_L^t &= \sqrt{\frac{t(\sigma_L^{t-1})^2}{t+1} + t\left(\frac{X_t - m_{t-1}}{t+1}\right)^2} = \\ &= \frac{1}{t+1}\sqrt{t\big((t+1)(\sigma_L^{t-1})^2 + (X_t - m_{t-1})^2\big)} \end{aligned} \qquad (2.12.7)$$

Проведем аналогичные выкладки для R:

$$\begin{aligned} s_t = \sigma_R^t &= \sqrt{\frac{1}{N-t}\sum_{k=t}^{N-1}(X_k - m_t)^2}, \\ s_{t+1} = \sigma_R^{t+1} &= \sqrt{\frac{1}{N-t-1}\sum_{k=t+1}^{N-1}(X_k - m_{t+1})^2}, \end{aligned} \qquad (2.12.8)$$

где

$$m_t = \frac{1}{N-t}\sum_{k=t}^{N-1}X_k, \qquad m_{t+1} = \frac{1}{N-t-1}\sum_{k=t+1}^{N-1}X_k,$$

а σ_R^0 – стандартное отклонение всего временного ряда. Для правого окна также справедливы рекуррентные формулы для математического ожидания

$$m_{t+1}=\frac{(N-t)m_t-X_t}{N-t-1}=\frac{N-t}{N-t-1}m_t-\frac{1}{N-t-1}X_t,$$

тогда их разница равна

$$m_{t+1}-m_t=\frac{N-t}{N-t-1}m_t-\frac{1}{N-t-1}X_t-m_t=\frac{m_t-X_t}{N-t-1}. \qquad (2.12.9)$$

Возведем в квадрат (2.12.8), чтобы избавиться от квадратного корня и выделим квадрат предыдущего значения стандартного отклонения:

$$\begin{aligned}
s_{t+1}^2&=\frac{1}{N-t-1}\sum_{k=t+1}^{N-1}(X_k-m_{t+1})^2\\
&=\frac{1}{N-t-1}\sum_{k=t}^{N-1}(X_k-m_{t+1})^2-\frac{(X_t-m_{t+1})^2}{N-t-1}=\\
&=\frac{1}{N-t-1}\sum_{k=t}^{N-1}(X_k-m_t+m_t-m_{t+1})^2\\
&-\frac{1}{N-t-1}\left(X_t-\frac{N-t}{N-t-1}m_t+\frac{1}{N-t-1}X_t\right)^2= \qquad (2.12.10)\\
&=\frac{1}{N-t-1}\left[\sum_{k=t}^{N-1}(X_k-m_t)^2\right.\\
&\left.-2\sum_{k=t}^{N-1}(X_k-m_t)(m_{t+1}-m_t)+\sum_{k=t}^{N-1}(m_{t+1}-m_t)^2\right]\\
&-\frac{1}{N-t-1}\left(\frac{N-t}{N-t-1}X_t-\frac{N-t}{N-t-1}m_t\right)^2=\\
&=\frac{1}{N-t-1}\left[\sum_{k=t}^{N-1}(X_k-m_t)^2-2\sum_{k=t}^{N-1}(X_k-m_t)(m_{t+1}-m_t)\right.\\
&\left.+\sum_{k=t}^{N-1}(m_{t+1}-m_t)^2\right]-\frac{(N-t)^2(X_t-m_t)^2}{(N-t-1)^3}.
\end{aligned}$$

Из формулы разности математических ожиданий (2.12.9) следует:

$$s_{t+1}^2 = \frac{1}{N-t-1}\left[\sum_{k=t}^{N-1}(X_k - m_t)^2 - 2\frac{m_t - X_t}{N-t-1}\sum_{k=t}^{N-1}(X_k - m_t)\right.$$
$$\left. + \frac{(N-t)(m_t - X_t)^2}{(N-t-1)^2}\right] - \frac{(N-t)^2(X_t - m_t)^2}{(N-t-1)^3} =$$
$$= \frac{1}{N-t-1}\left[(N-t)s_t^2 + \frac{(N-t)(m_t - X_t)^2}{(N-t-1)^2}\right] \qquad (2.12.11)$$
$$- \frac{(N-t)^2(X_t - m_t)^2}{(N-t-1)^3} =$$
$$= \frac{(N-t)s_t^2}{N-t-1} + \frac{(N-t)(m_t - X_t)^2(1-N+t)}{(N-t-1)^3}.$$

Таким образом, рекуррентная формула для стандартного отклонения имеет вид:

$$\sigma_R^t = \sqrt{\frac{(N-t)(\sigma_R^{t-1})^2}{N-t-1} + \frac{(N-t)(m_t - X_t)^2(1-N+t)}{(N-t-1)^3}} \qquad (2.12.12)$$

Мы получили рекуррентные формулы для вычисления стандартных отклонений, что позволит значительно быстрее производить расчет величины W_t. Ниже в таблице приведены результаты сравнения времени (в секундах) вычисления напрямую и с помощью выведенных формул $(2.12.7)$ и $(2.12.12)$, показывающие, что многочисленное вычисление стандартного отклонения можно осуществлять даже для временных рядов, содержащих сотни тысяч отсчетов за приемлемое время:

Таблица 2.12.1. Сравнение времени (в секундах) вычисления стандартного отклонения напрямую и с помощью рекуррентных формул

Число отсчетов	**Напрямую**	**Рекуррентно**
1000	0.456	0.023
5000	5.347	0.115
10000	22.471	0.248
20000	96.032	0.449

2.13. Заключение к второй главе

Во второй главе диссертации рассмотрены три метода анализа временных рядов GPS, основанных на вычислении следующих описательных статистик:

1. Спектральной экспоненты дискретного вейвлет-преобразования;
2. Нормализованной энтропии скачкообразной составляющей;
3. Нормализованной энтропии меры выбросов.

Введено понятие псевдо-производной и описаны ее свойства, имеющие определенные сходства со свойствами классической производной и являющиеся новым инструментом для анализа временных рядов. На основе псевдо-производной предложены метод быстрой оценки коэффициентов наклонов линий тенденций в заданном временном интервале и алгоритмы построения кусочно-ступенчатой аппроксимации сигналов. Произведено численное сравнение результатов работы предложенного алгоритма построения кусочно-ступенчатой аппроксимации с уже имеющимися методами дискретного вейвлет-преобразования Хаара и максимума модуля непрерывного вейвлет-преобразования (WTMM).

В качестве меры скачкообразности временного ряда предложена нормализованная энтропия приращений усредненной кусочно-ступенчатой аппроксимации, так как данная характеристика позволяет оценить степень хаотичности скачков: если скачки становятся менее случайными, то значение нормализованной энтропии становится ниже, чем для полностью хаотичного сигнала.

Аналогичная мера предложена и для оценки выбросов во временных рядах. В качестве вспомогательного ряда выбрана взвешенная сумма приращений стандартных отклонений слева и справа от заданного момента времени. Данный ряд позволяет определить те значения временного ряда, которые на фоне естественного шума выделяются наиболее сильно, следовательно, по значению нормализованной энтропии этого вспомогательного ряда можно судить о систематичности появляющихся в анализируемом сигнале выбросов. Чтобы ускорить процесс вычисления вспомогательного ряда выведены рекуррентные формулы для приращений стандартных отклонений.

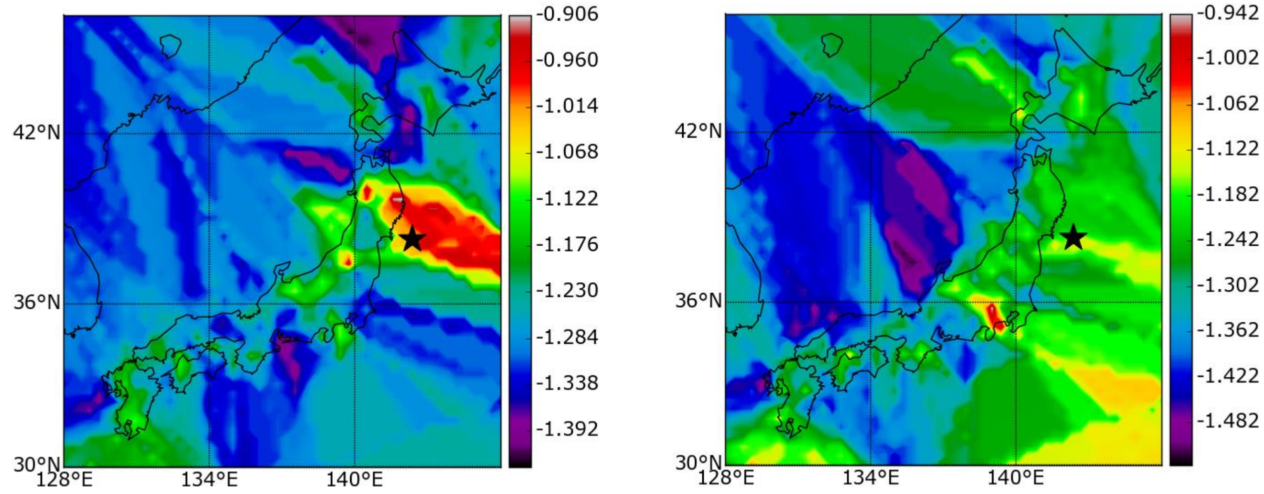
Глава 3. Пространственное распределение статистик временных рядов GPS

В предыдущей главе были описаны три метода анализа временных рядов. Рассмотрим, какие они дают результаты для различных временных рядов GPS. Основным способом визуализации результатов является построение карт пространственного распределения вычисленных статистик для временных рядов GPS в заданном промежутке времени. Для построения карт вводилась прямоугольная область анализируемого региона и покрывалась регулярной сетью из 50x50 узлов. Для некоторых карт значения в узловых точка вычислялись путем простого усреднения значений, полученных для $n = 10$ ближайших станций.

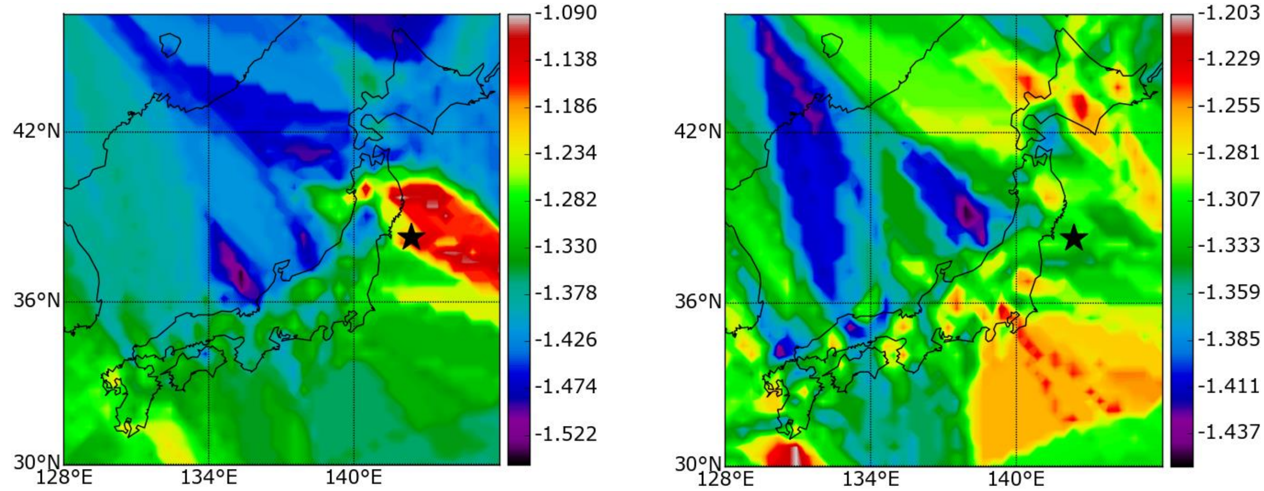
3.1. Построение карт спектральной экспоненты

На рисунке 3.1.1 изображены 6 карт пространственного распределения значений спектральной экспоненты: до и после мега-землетрясения Тохоку 11 марта 2011 года для каждой из трех компонент временных рядов GPS: смещения восточных, северных и вертикальных координат. В период более, чем за месяц до катастрофы, на картах видно, что область подготовки землетрясения выделяется повышенными значениями спектральной экспоненты. Данные величины были рассчитаны после перехода к приращениям, таким образом, высокие значения соответствуют увеличению мощности низкочастотных колебаний. После землетрясения видно, что накопленная энергия распределилась по тихоокеанскому побережью, где еще на протяжении нескольких дней встречались афтершоковые подземные толчки.

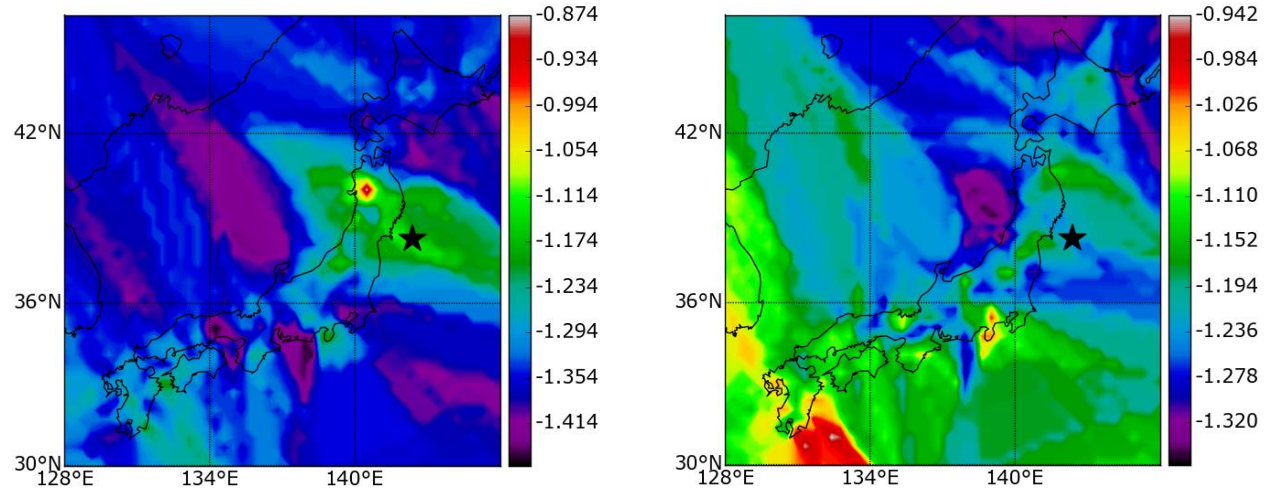
До землетрясения **После землетрясения**



а) смещения восточных координат



б) смещения северных координат



в) смещения вертикальных координат

Рисунок 3.1.1. Карты спектральной экспоненты до и после землетрясения

Вычислив значения спектральной экспоненты для каждой компоненты временного ряда для более четкого выделения особенностей распределения данной статистики воспользуемся методом главных компонент.

Присвоим индекс $\alpha = 1,2,3$ каждому временному ряду смещений восточных, северных и вертикальных координат соответственно, тогда значения спектральной экспоненты в каждом узле сети обозначим b_{ij}^{α}. Вычислим средние значения и дисперсии:

$$s_\alpha = \frac{1}{N_x N_y} \sum_{i,j} b_{ij}^{\alpha},$$
$$\sigma_\alpha^2 = \frac{1}{N_x N_y} \sum_{i,j} \left(b_{ij}^{\alpha} - s_\alpha\right)^2, \qquad (3.1.1)$$

где N_x, N_y – число узлов по долготе и широте соответственно, в нашем случае $N_x = N_y = N = 50$. Построим корреляционную матрицу размерностью 3х3:

$$R = \left(r_{\alpha\beta}\right),$$
$$r_{\alpha\beta} = \frac{1}{N^2} \sum_{i,j} \Delta b_{ij}^{\alpha} \cdot \Delta b_{ij}^{\beta}, \qquad (3.1.2)$$

где $\Delta b_{ij}^{\alpha} = \frac{b_{ij}^{\alpha} - s_\alpha}{\sigma_\alpha}$, $\alpha, \beta = 1,2,3$. Тогда первые главные компоненты вычисляются по формуле

$$P_{ij} = \sum_{\alpha=1}^{3} \Delta b_{ij}^{\alpha} \cdot U_\alpha \qquad (3.1.3)$$

где U_α – компоненты собственного вектора матрицы R, соответствующего максимальному собственному значению. В результате получим следующие карты, изображенные на рисунке 3.1.2.

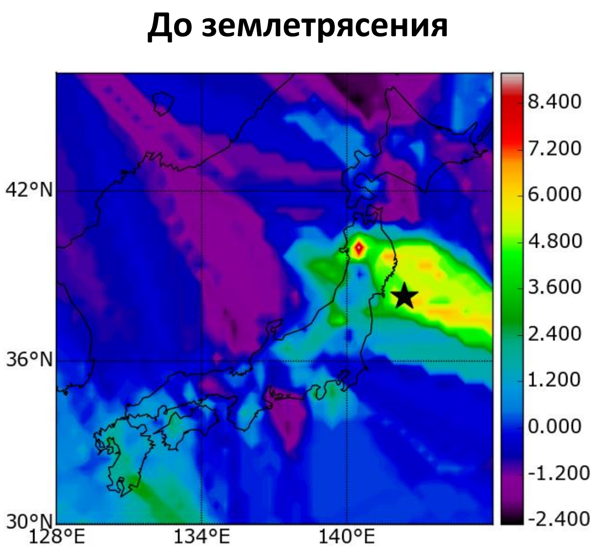


Рисунок 3.1.2. Карты первой главной компоненты значений спектральной экспоненты для флуктуаций компонент сигналов GPS после центрирования и нормировки

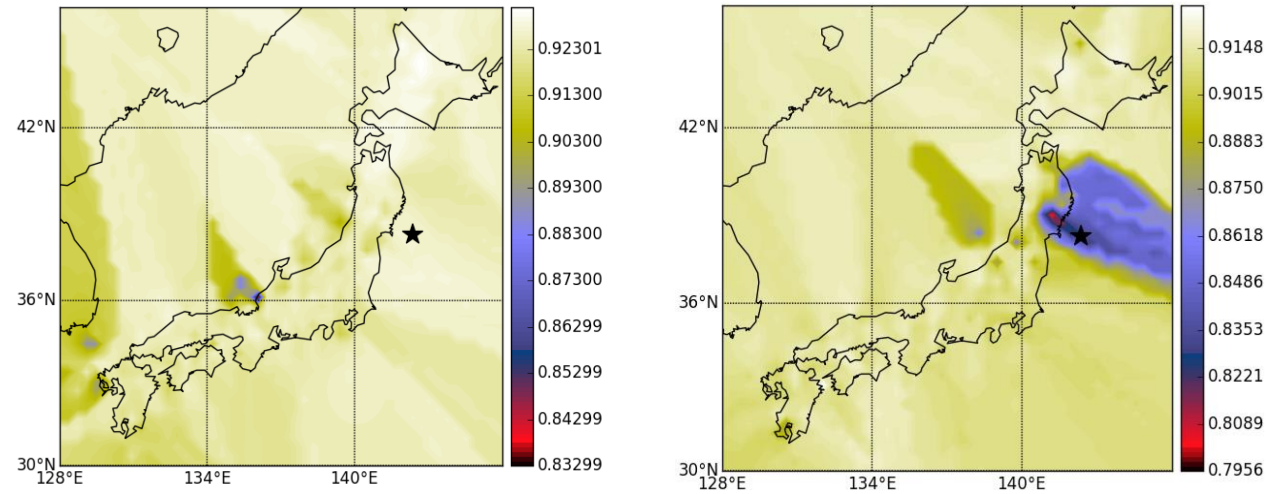
Аномальная зона в области эпицентра будущего землетрясения также характеризуется высоким значением первой главной компоненты. Стоит отметить, что в сейсмически нестабильном регионе в период подготовки землетрясения значения первой главной компоненты колеблются в диапазоне от 4 до 6.5, а после землетрясения величины $(3.1.3)$ резко упали и распределились более равномерно вдоль тихоокеанского побережья.

3.2. Построение карт нормализованной энтропии скачков

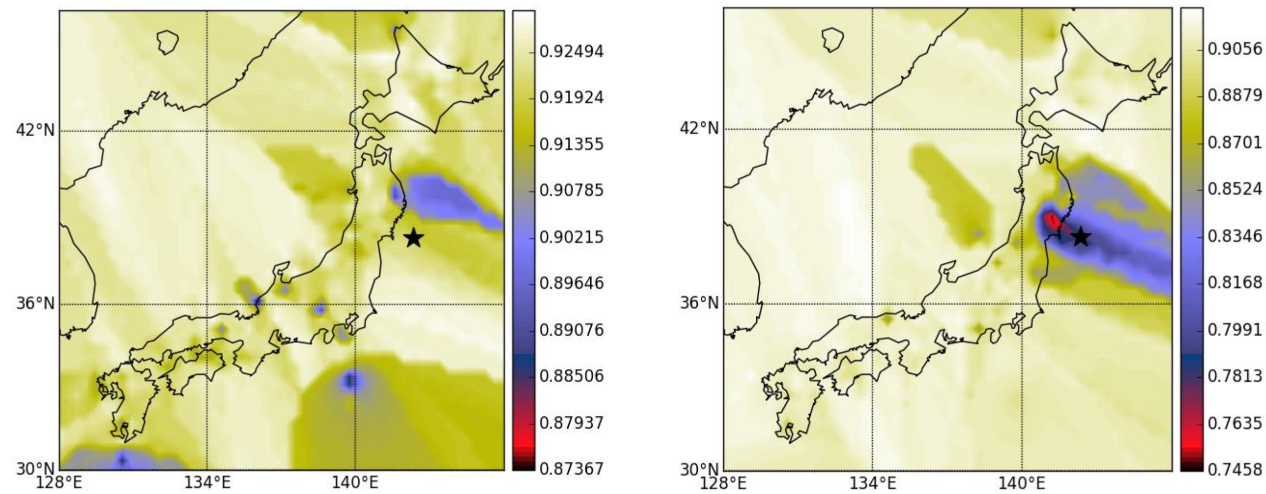
Ниже (рисунок 3.2.1) представлены результаты картирования распределения нормализованной энтропии скачкообразной составляющей временных рядов GPS. В отличие от карт спектральной экспоненты, на карте смещения восточных координат до землетрясения отсутствуют настолько явные предвестники катастрофы, однако на картах для смещений северных и вертикальных координат, можно увидеть область пониженных значений нормализованной энтропии. Таким образом, стоит отметить, что применение для прогнозирования катастрофы только анализа скачкообразной компоненты сигналов GPS по предложенному алгоритму не является верным решением. Однако в совокупности с анализом выбросов (результаты представлены в следующем параграфе) временных рядов GPS полученные карты могут дополнить общую картину и тем самым подтвердить или опровергнуть гипотезу о надвигающейся катастрофе.

Аномалия в области эпицентра после землетрясения объясняется наличием так называемых афтершоков, которые появлялись в этом регионе на протяжении еще нескольких дней после основного толчка.

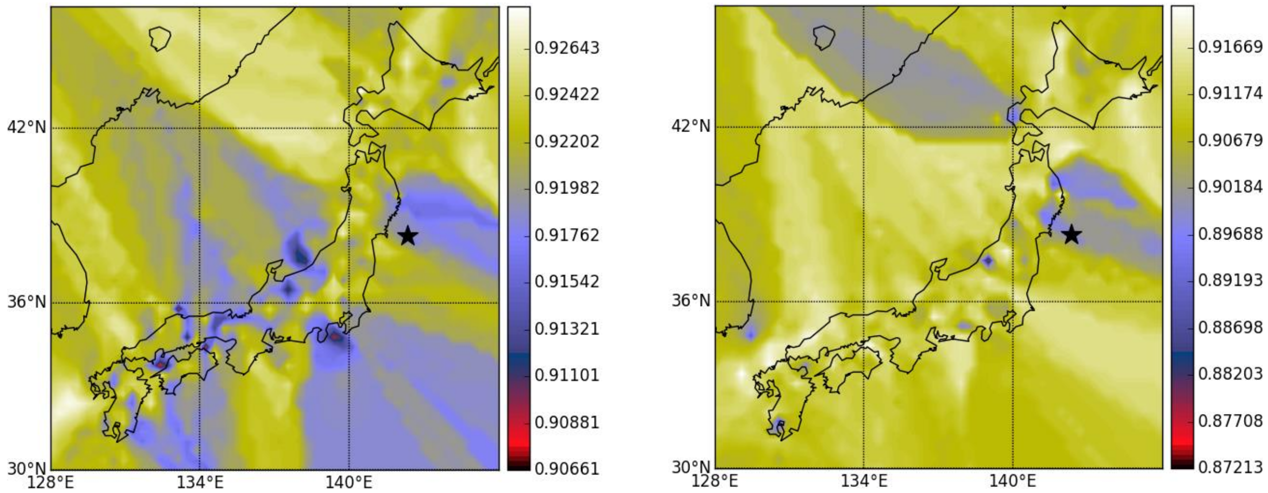
До землетрясения (30.01.2011-09.03.2011) **После землетрясения (12.03.2011-26.03.2011)**



а) Смещение восточных координат



б) Смещение северных координат



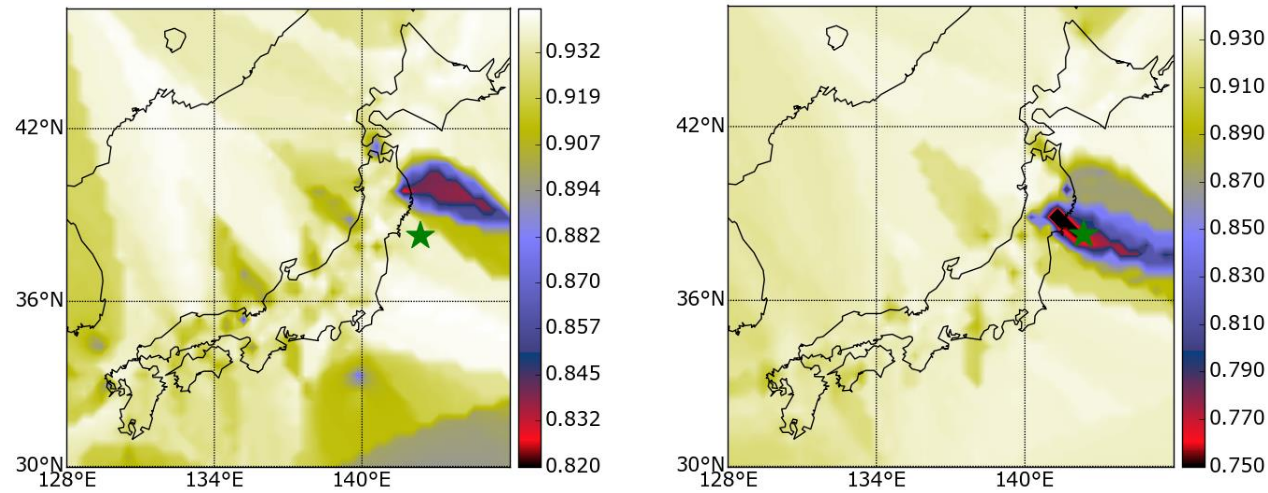
в) Смещение вертикальных координат

Рисунок 3.2.1. Карты нормализованной энтропии скачков до и после землетрясения

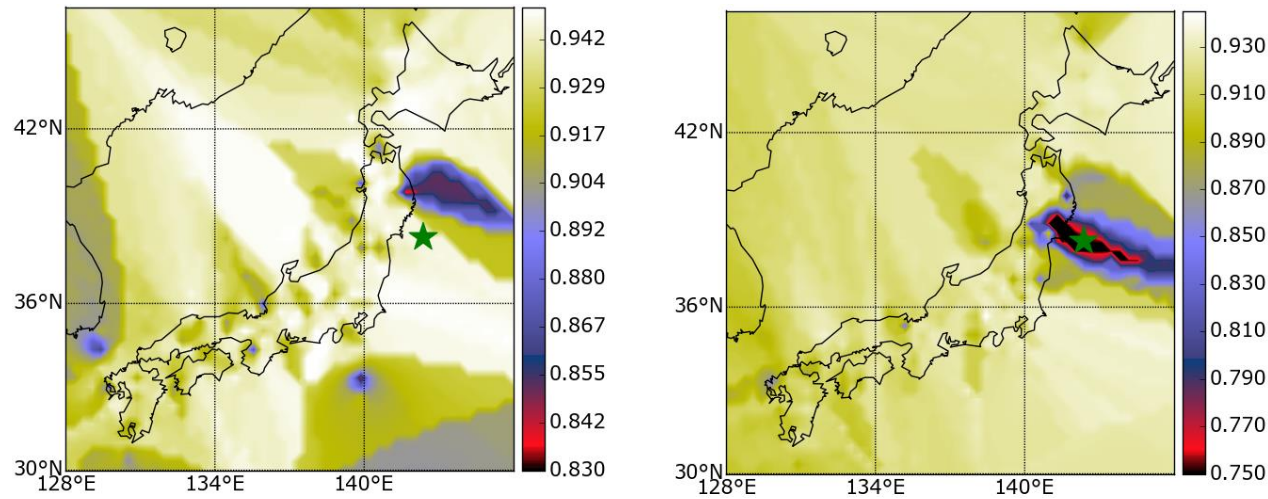
3.3. Построение карт нормализованной энтропии выбросов

Ниже (рисунок 3.3.1) на картах, построенных до землетрясения, для всех координат видно, что «нормальным» значением нормализованной энтропии является величина больше 0.9, однако в окрестности, находящейся в непосредственной близости от эпицентра землетрясения (отмечен черной звездой), заметны пониженные значения, то есть аномальная область тяготеет к эпицентру будущего землетрясения. Таким образом, можно сделать предположение, что если энтропия меньше 0.9, то это может сигнализировать о наличии повышенной сейсмической активности. На картах, построенных после землетрясения, мы видим, что зона пониженных значений сместилось точно в область наиболее сильных толчков во время землетрясения, что подтверждает гипотезу о том, что пониженная нормализованная энтропия статистики W отражает нестабильность наблюдаемой системы. Аномалия после землетрясения также является последствием афтершоковых толчков.

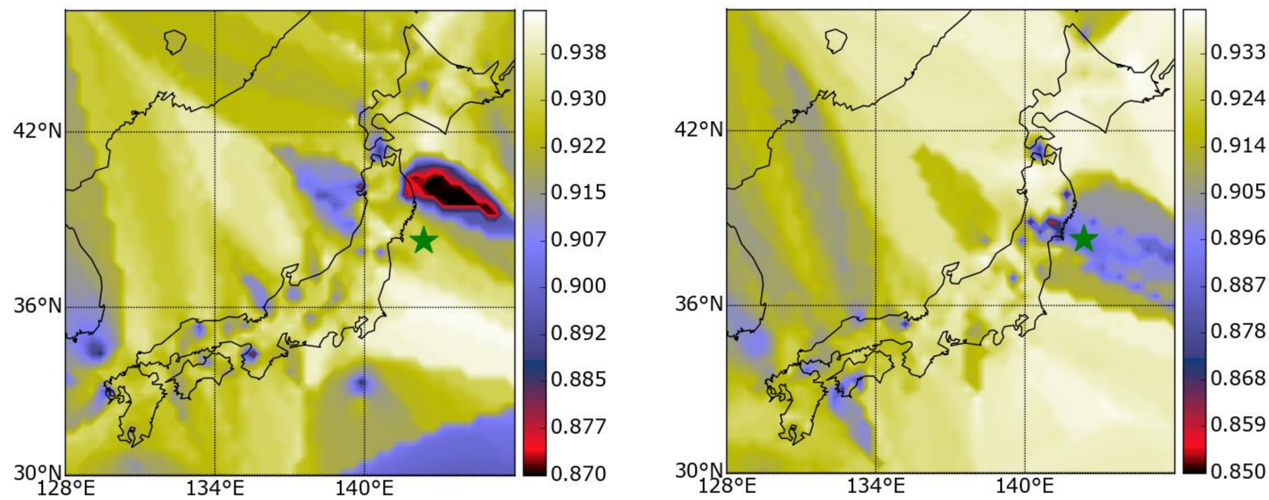
До землетрясения (30.01.2011-09.03.2011) **После землетрясения (12.03.2011-26.03.2011)**



а) Смещение восточных координат



б) Смещение северных координат

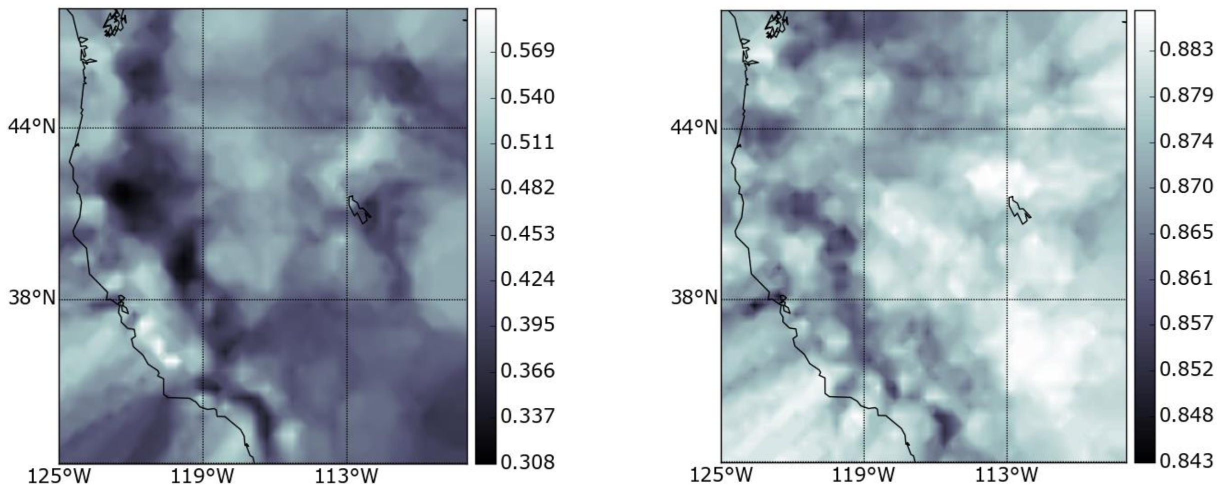


в) Смещение вертикальных координат

Рисунок 3.3.1. Карты нормализованной энтропии выбросов до и после землетрясения

3.4. Выделение зон интенсивных движений земной коры

Картирование собственных значений корреляционных матриц позволяет более четко охарактеризовать уровень синхронизации временных рядов GPS в отличие от использования просто коэффициента корреляции, отражая тем самым общность колебаний в шуме. В частности, на примере дневных данных GPS (смещения вертикальных координат) для западного региона Соединенных Штатов Америки на карте нормированного максимального собственного значения (рисунок 3.4.1а) к сумме всех собственных значений корреляционной матрицы, вычисленной на основе сигналов, полученных с $n = 10$ ближайших к узлу сети станций, после удаления линейного тренда и перехода к приращениям видно, что низкие значения расположены вдоль тихоокеанского побережья, где расположена одна из крупнейших систем разломов, в которой также расположен знаменитый разлом Сан-Андреас. Значения в узлах сети вычислялись путем усреднения нормализованных собственных значений для ближайших станций.



а) Нормализованное собственное значение б) Нормализованной энтропии скачков

Рисунок 3.4.1. Сравнение карт нормированного собственного значения и нормализованной энтропии скачков на примере западного региона США

Если сравнивать данную карту с картой нормализованной энтропии скачков (рисунок 3.4.1б), то можно отметить, что области низких значений для них в значительной мере совпадают. Как следствие данный факт предположительно может указывать на область интенсивных движений земной коры. Таким образом, мы имеем два независимых друг от друга метода, позволяющие выделить данное явление. Взаимное соответствие между получаемыми значениями установить пока что невозможно, так как это требует более глубокого и подробного изучения как на примере западного региона США, так и прочих других, так как каждый отдельно взятый регион имеет свои особенности, свою структуру литосферных плит, а, следовательно, процессы под землей протекают по-разному. Таким образом, если для одного региона распределение статистики является аномальным и предвещающим катастрофу, в другом регионе это может не значить ничего. Тем не менее рисунок 3.4.1, демонстрирует, что два разных метода дают качественно схожие результаты, что позволяет применять их совместно для подтверждения или опровержения полученных результатов.

3.5. Карты мультифрактальных параметров временных рядов GPS

Полученные в разделе 3.4 результаты позволили выделить на западной части США гипотетически наиболее интенсивные движения литосферных плит. В основе данной гипотезы лежат два факта: области выделены низкими значениями, полученными двумя разными методами. Пониженные значения в обоих случаях означают наличие какого-то механизма, являющегося потенциальной угрозой для данного региона планеты.

В первом случае (рисунок 3.4.1а) нарушение коррелированности временных рядов GPS, характерных практически для всей территории США, расположенной на североамериканской плите. О том, что знаменитый тектонический разлом Сан-Андреас является одним из самых сейсмически активных в мире, давно известно, и совпадение низких значений первой главной компоненты корреляционной матрицы в области, расположенной точно вдоль разлома, лишний раз это подтверждает.

Во втором случае (рисунок 3.4.1б) та же самая зона разлома выделена пониженными значениями энтропии скачкообразной компоненты сигналов GPS.

Здесь уже идет речь не о некорреливанности временных рядов GPS, а, скорее, о причине ее понижения. Вызвано данное понижение энтропии наличием в сигналах систематических выбросов, то есть нарушением хаоса, являющего в данном случае нормальным. Хорошее объяснение нормальности хаоса («здоровый хаос») дано в [26]. Таким образом, карта распределения энтропии позволила выделить зону, в которой присутствует собственное поведение земной коры, отличное от остальной территории США, и это поведение генерируется каким-то механизмом, понижающим степени свободы системы. Как следствие, мы имеем другие значения корреляции временных рядов.

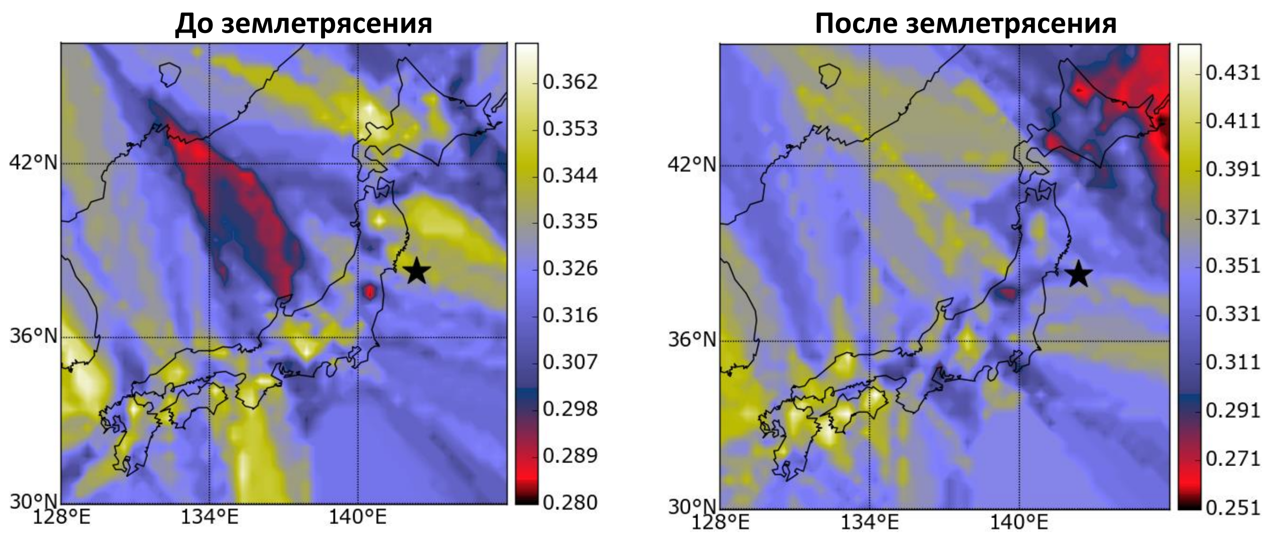
Заговорив на тему хаоса, мы затрагиваем огромный математический раздел теории хаоса, который занимается изучением динамических систем. Одной из основных задач применения данной теории на практике является прогнозирование эволюции той или иной системы во времени, однако на сегодняшний день до сих пор нет математического инструмента для составления реалистичного прогноза землетрясений или извержений вулканов. В основном все предположения опираются на наличие периодичности данных явлений. Если временной период без происшествий подходит к концу, а сейсмическая активность постепенно увеличивается, то составляется прогноз, что скоро случится катастрофа. Подобные прогнозы составлены для США как относительно разлома Сан-Андреас (последнее разрушительное землетрясение было в Сан-Франциско в 1906 году, а периодичность составляет приблизительно 150 лет), так и Йеллоустонского вулкана (последнее извержение случилось 640000 лет назад, а вычисленная периодичность составляет примерно 600000).

Наличие периодичности в изучаемых системах является очень сильным фактором при составлении прогнозов, однако, когда встает вопрос о прогнозе природной катастрофы, погрешность может составлять от нескольких до сотен лет, что не делает прогноз реалистичным. Поэтому, возвращаясь к вопросу об изучении временных рядов с точки зрения теории хаоса, к прогнозу подключают результаты анализа аттрактора системы и его корреляционной размерности, являющейся одной из самых важных количественных характеристик. корреляционная размерность временных рядов, вычисляемая с помощью так называемого корреляционного интеграла. Так, например, в работах [69], [70] результаты применения корреляционного интеграла показали, что корреляционная размерность распределения гипоцентров землетрясений имела аномально низкие значения перед землетрясениями в Японии в 2000 (магнитуда 7.3) и в 2003 годах (магнитуда 8.0). Аналогичный анализ был применен к Камчатскому региону [6], и результаты так же показали резкое снижение значения корреляционной размерности аттрактора приблизительно за полгода до землетрясения 1983 года (магнитуда 7.0). К сожалению, все данные прогнозы сделаны постфактум.

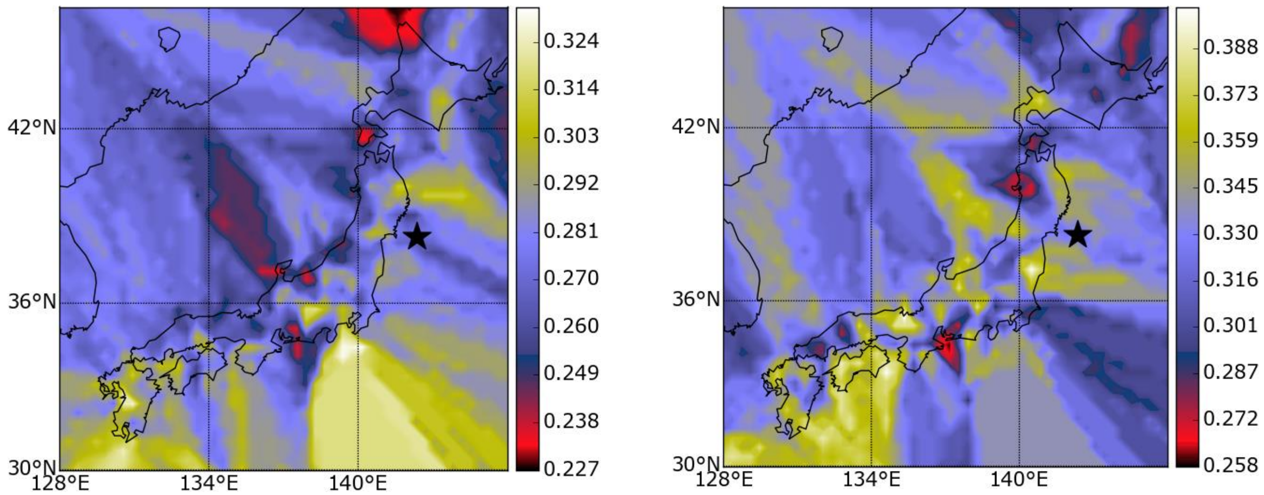
Не только аттракторы имеют фрактальную структуру. Сами временные ряды, по которым данные аттракторы восстанавливаются [14], [57] представляются собой фрактальные объекты, и для них также применимы понятия корреляционной и фрактальной размерностей. Данные размерности тесно связаны друг с другом, так как они являются частными случаями обобщенных фрактальных размерностей Реньи. Спектр этих обобщенных размерностей можно оценить с помощью метода MFDFA, упомянутого в разделе 1.2. Существует также понятие информационной размерности, которая также является частным случаем размерностей Реньи [44]. Такое изобилие различных характеристик обусловлено сложностью фрактальных структур.

Как говорилось выше во вводной части настоящей работы, большой интерес представляют собой также величины α_H и $\Delta\alpha$ – характеристики мульти-фрактального спектра сингулярности. Оценка каждого из них в скользящем временном окне позволяет проследить за динамикой их изменения, однако вопрос связи между каким-либо значением ширины носителя спектра сингулярности с «нормальной» или «здоровой» тектоникой плит пока что остается открытым.

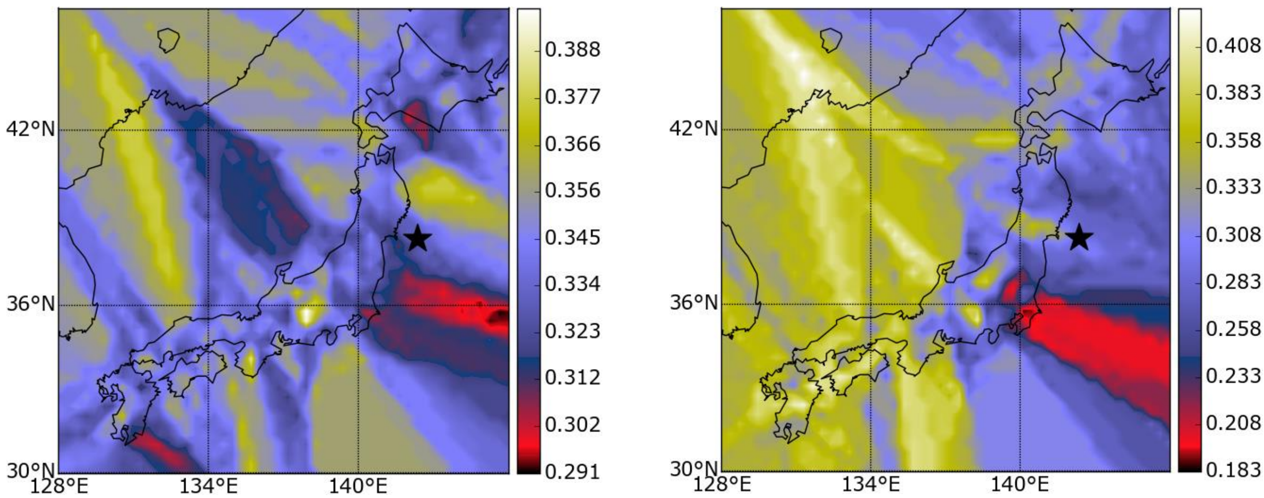
Применительно к исследованию региона Японии [26] низкие значения ширины спектра сингулярности низкочастотных микросейсмических колебаний представляли собой сигналы о приближающейся опасности. Конечно, не каждая область низких значений является предвестником надвигающейся катастрофы, однако если данная область имеет тенденции к увеличению площади покрытия и к падению значений, то этот регион может быть оценен как сейсмически опасный. Временные ряды GPS отличаются от временных рядов сейсмических колебаний, и, как следствие, использование аналогичного подхода оказалось бы, скорее всего, неверным. Целью настоящей работы является не прогноз землетрясений, а выделение регионов, отличающихся повышенной интенсивностью движения литосферных плит. Поэтому если говорить про мультифрактальный анализ рядов GPS, то повышенный интерес представляют высокие не низкие значения спектра сингулярности, а наоборот, высокие. Именно широкий спектр является свидетельством того, что исследуемый сигнал имеет сложную структуру, то есть анализируемый шум является разнообразным. К сожалению, в силу отсутствия большого набора данных нет возможности построить карт распределения спектра сингулярности до землетрясения 2011 года в Японии. Тем не менее взглянем на усредненные карты распределения $\Delta\alpha$ для 30-минутных временных рядов GPS до и после землетрясения после перехода к приращениям.



а) Смещение восточных координат



б) Смещение северных координат

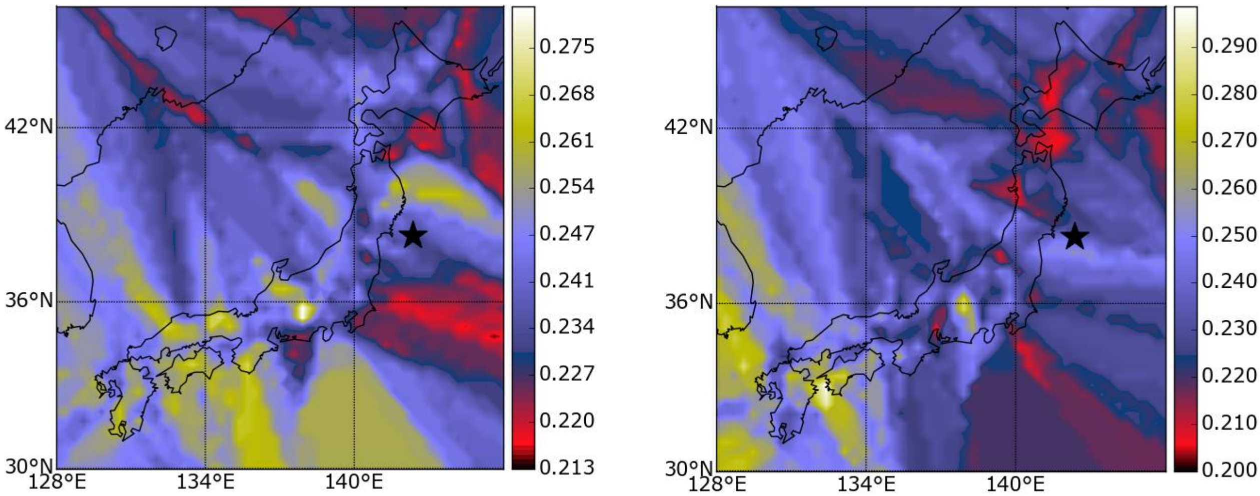


в) Смещение вертикальных координат

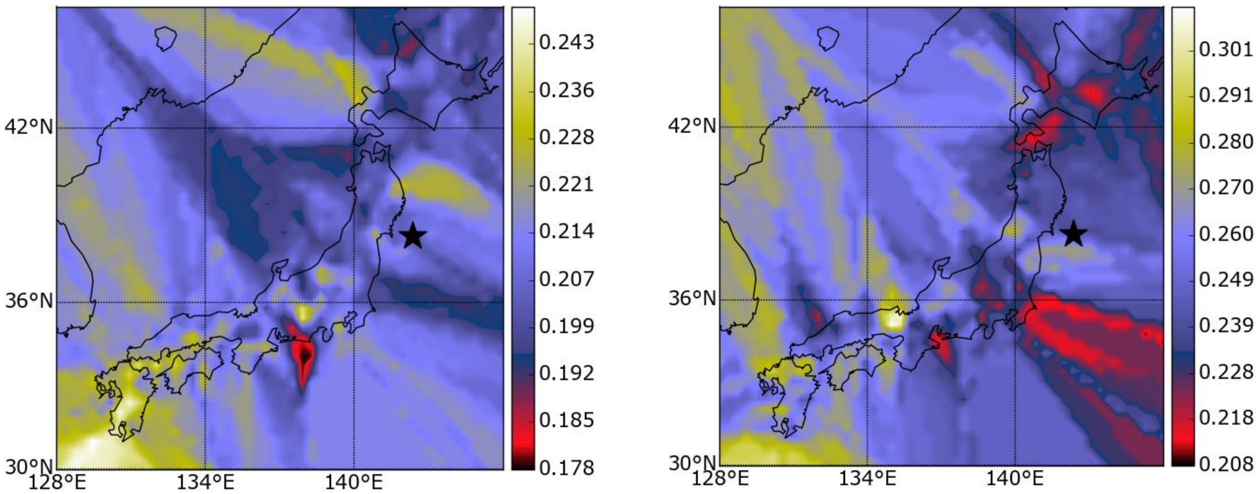
Рисунок 3.5.1. Усредненные карты распределения $\Delta\alpha$ после перехода к приращениям, построенные в скользящем временном окне длиной 672 отсчета (2 недели) со сдвигом в 48 отсчетов (1 сутки)

По результатам, представленным на рисунке 3.5.1, нельзя однозначно определить регион подготовки землетрясения, однако стоит отметить, что в период до землетрясения на каждой из трех карт повышенными значениями выделена та же область, что и определена с помощью картирования распределения статистики выбросов и скачков временных рядов GPS. На картах спектральной экспоненты (рисунок 3.1.1) аномальная область имеется большую площадь и включает, как эпицентр землетрясения, так и прилежащий регион, определяемый остальными методами сейсмически нестабильным. Аналогичную картину можно увидеть и на картах распределения обобщенного показателя Херста (рисунок 3.5.2).

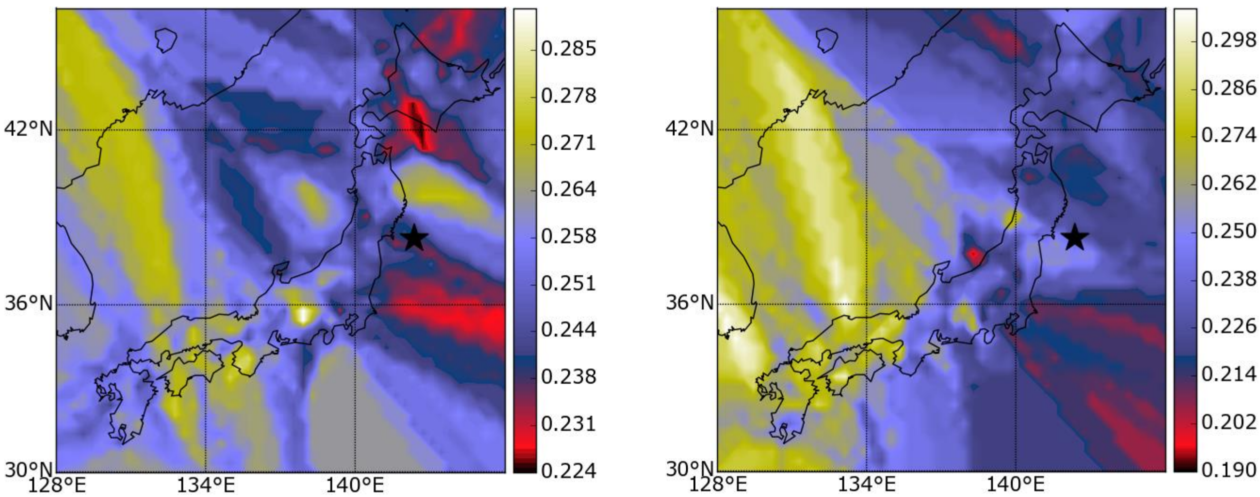
До землетрясения **После землетрясения**



а) Смещение восточных координат



б) Смещение северных координат



в) Смещение вертикальных координат

Рисунок 3.5.2. Усредненные карты распределения обобщенного показателя Херста после перехода к приращениям, построенные в скользящем временном окне длиной 672 отсчета (2 недели) со сдвигом в 48 отсчетов (1 сутки)

Показатель Херста, как правило оценивается с целью определить персистентность (сохранение тенденции изменения временных рядов) и антиперсистентность (уменьшение или изменение имеющейся тенденции в будущем) временного ряда. Термин персистентность пришел к нам в лексикон из английского – persistent, в переводе означающего постоянный, устойчивый. На наличие персистентности временных рядов указывает показатель Херста, чье значение находится в диапазоне $0.5 < \alpha_H < 1$. В случае, когда $0 \leq \alpha_H < 0.5$ – временной ряд антиперсистентен. Если показатель Херста равен 0.5, то принято говорить о случайности временного ряда. В общем случае обобщенный показатель Херста не обязательно находится пределах от 0 до 1.

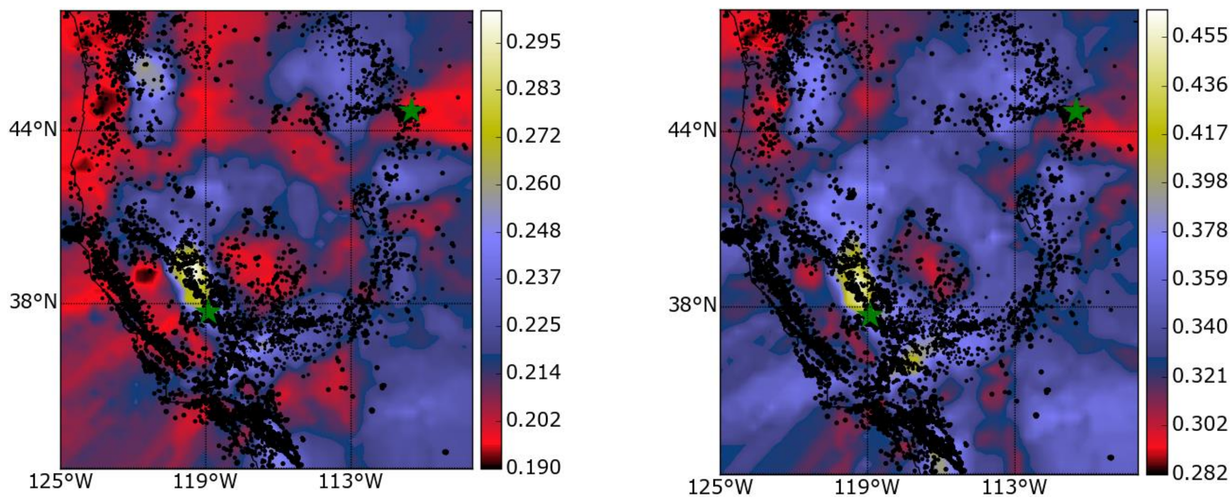
В данном случае на примере сигналов GPS нас не интересует смена тенденции, интерес вызывает другой факт. Временные ряда с высоким показателем Херста являются слабо зашумленными, то есть имеют в своей структуре фактически один тренд, в то время как низкие значения показателя Херста (близкие к нулю) говорят о наличии более ярко выраженной шумовой компоненты. Из рисунка 3.5.2 видно, что значения α_H сильно меньше величины 0.5, то есть говорить о полной случайности блуждания временных рядов GPS после перехода к приращениям не приходится – все они в большинстве своем относятся к антиперсистентным сигналам – зашумленным, что вызывает повышенный интерес к изучению данного шума.

Разобравшись, что несут за собой значения обобщенного показателя Херста, проинтерпретируем еще раз карты, изображенные на рисунке 3.5.2. Во временной промежуток с 30 января 2011 года по 9 марта 2011 область, расположенная к северо-востоку от эпицентра будущего землетрясения, имеет повышенные значения α_H, что означает, что смещения по всем трем координатам, полученные с ближайших станций, установленных на островах, характеризуются меньшей антиперсистентностью (большей случайностью), чем в остальном регионе Японии. Данная случайность, возможно, вызвана появлением новых механизмов в земной коре, порождающих данные смещения. Именно в этой новой случайной составляющей хранится информация о выбросах и скачках, обнаруженных при анализе, в этом же регионе (рисунки 3.2.1 и 3.3.1).

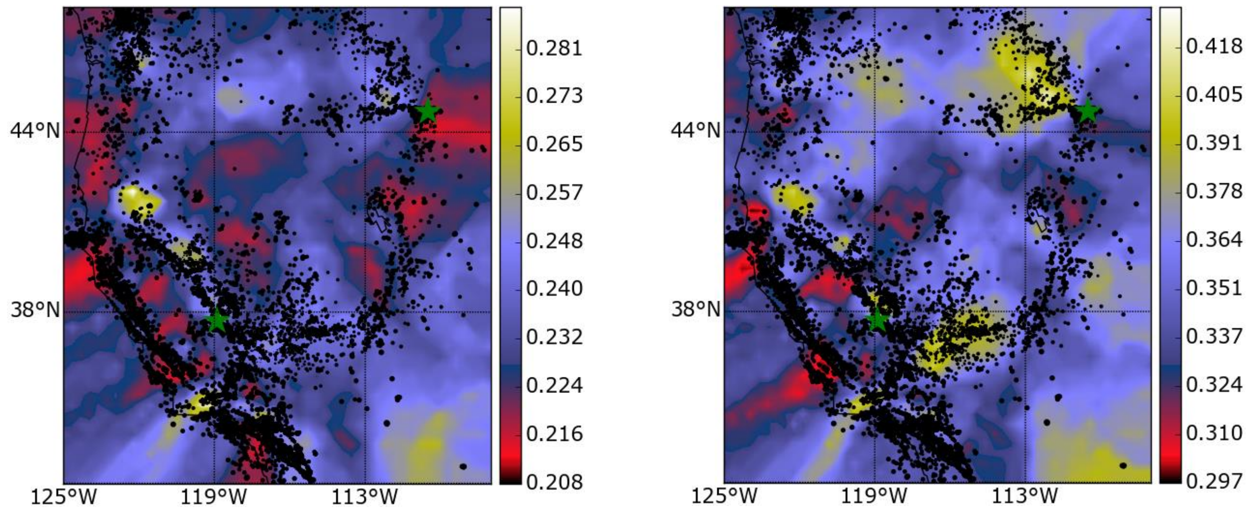
Неясность (зашумленность) картины, а именно наличие протяженных областей тех или иных значений вдали от островов, вызвана тем, что сеть GPS станций в Японии расположена в основном на островах, в силу сложности или невозможности их установки на морском или океаническом пространстве. Таким образом, для акватории все вычисленные значения являются грубой аппроксимацией, что не позволяет однозначно интерпретировать полученные результаты. Другая причина, по которой аномальная зона не выявлена столь же ясно, как на картах выбросов, скачков или спектральной экспоненты (хотя этого следовало ожидать в силу соотношения $\beta = 2\alpha_H - 1$ в случае полного соответствия исследуемых сигналов фрактальным объектам) может быть связана с нехваткой анализируемых данных. Карты, изображенные на рисунках 3.5.1 и 3.5.2, являются усреднением всего лишь 25 карт, построенных в скользящих временных окнах длиной 2 недели со сдвигом по времени 1 сутки. Для отслеживания динамики мульти-фрактальных число таких карт должно быть значительно выше. Данные за более ранний промежуток времени имеются во временных рядах с шагом по времени 1 день, но их слишком мало, и длина временных рядов значительно меньше, что делает их непригодными для данного анализа.

Большее число данных имеется для региона западной части США, однако это временные ряды взяты с шагом дискретизации в 1 сутки. Для этой области удалось построить усреднение 257 карт для каждой из трех координат, каждая из которых получена в скользящем временном окне 730 отсчетов (приблизительно два года) со сдвигом по времени 7 суток. Результаты изображены ниже на рисунке 3.5.3.

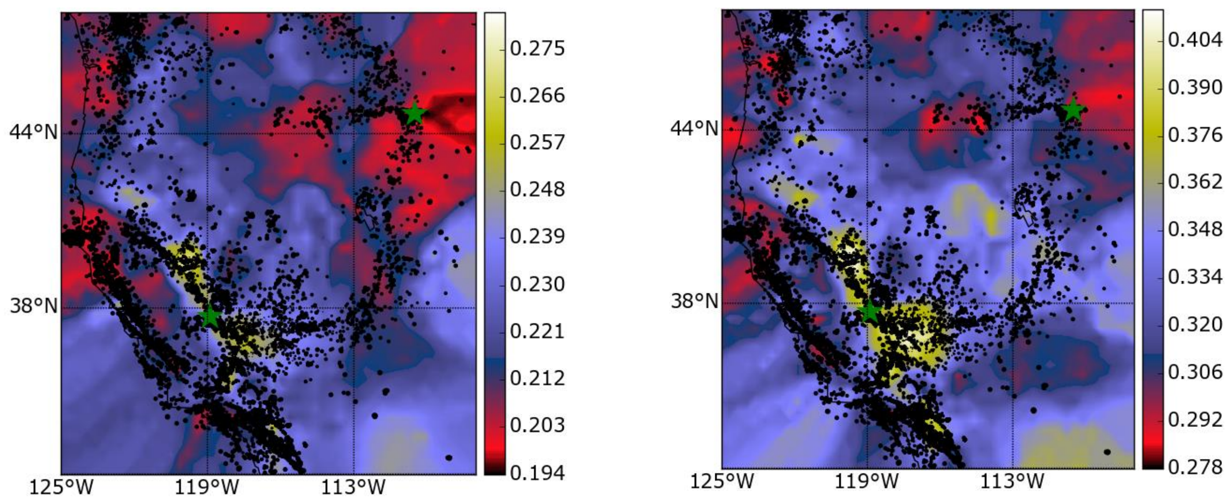
α_H $\Delta\alpha$



а) Смещение восточных координат



б) Смещение северных координат



в) Смещение вертикальных координат

Рисунок 3.5.3. Усредненные карты распределения ширины спектра сингулярности и обобщенного показателя Херста для западной части США.

Для полученных результатов можно построить также и карты первой главной компоненты, как это было проделано для японских островов.

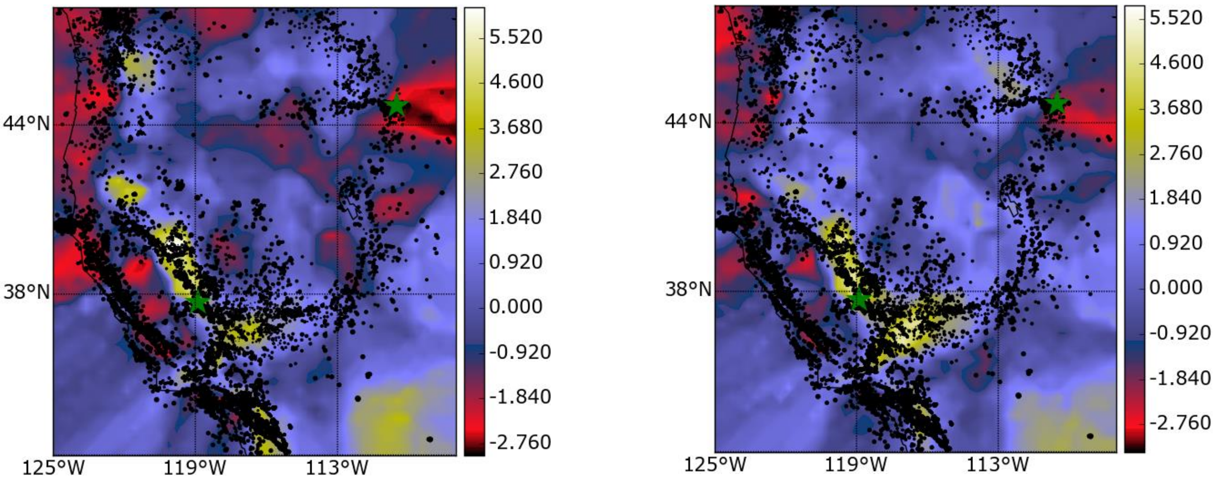


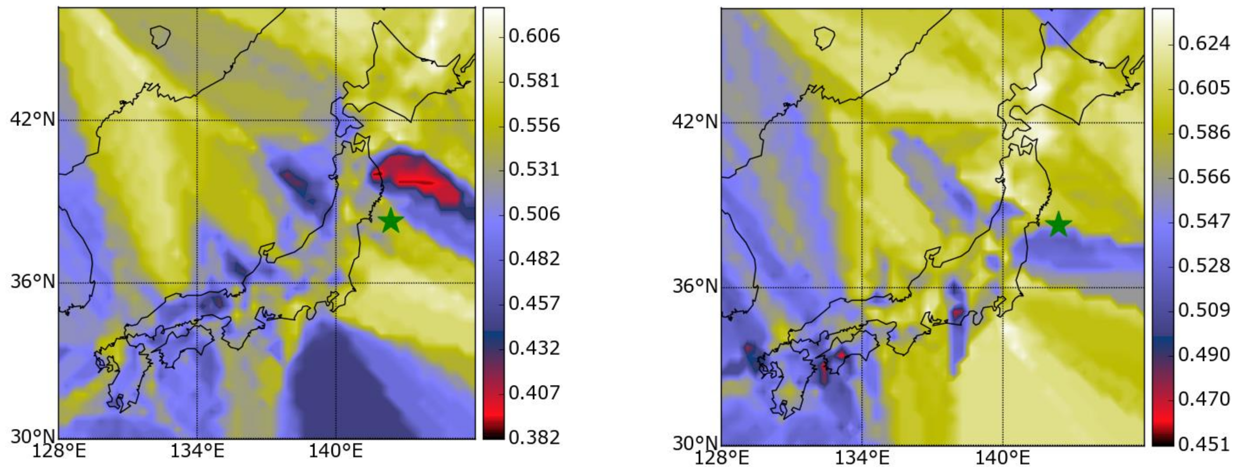
Рисунок 3.5.4. Карты первой главной компоненты для западной части США. Слева – обобщенный показатель Херста, справа – ширина спектра сингулярности.

Черными точками на рисунках 3.5.3 и 3.5.4 обозначены эпицентры всех произошедших в 2015 году землетрясений. Зеленые звездочки соответствуют двум супервулканам – Лонг Велли (-118.9 з.д., 37.7 с.ш.) и Йеллоустон (-110.5 з.д., 44.667 с.ш.). Интерес к вулкану Лонг Велли появился после обнародования наблюдений геологов о том, что в 2015 году он проявлял очень сильную и необычную активность – участились мелкие землетрясения, а почва за этот год поднялась там на 25 см. Дополнительную сейсмическую активность привносит система разломов, контуры которых наглядно отражают эпицентры землетрясений.

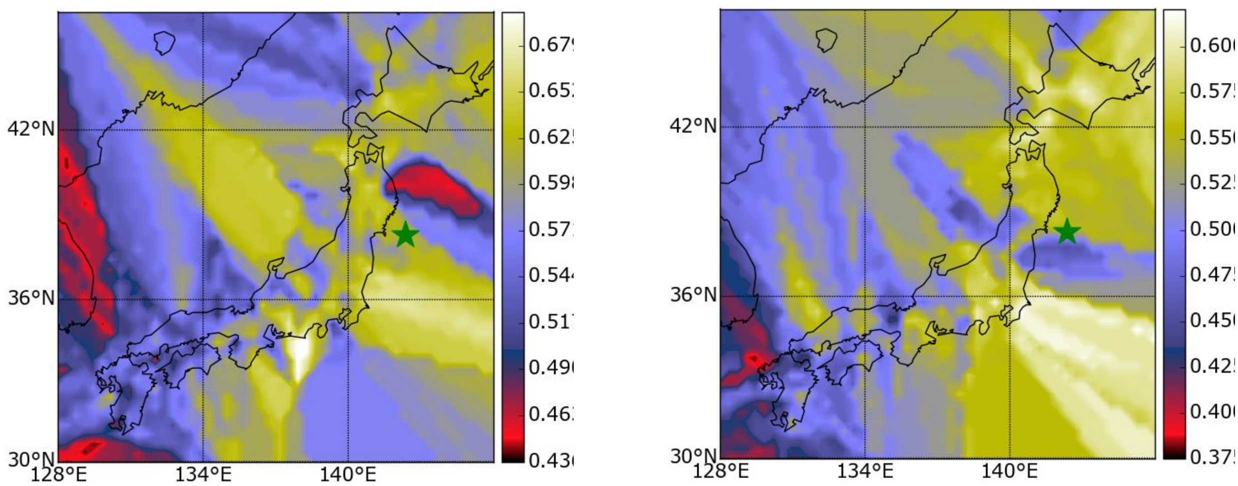
Из рисунков 3.5.3 и 3.5.4 видно, что области повышенных значений обобщенного показателя Херста и ширины спектра сингулярности хорошо ложатся на эпицентры землетрясений, расположенных вдоль разломов, которые проходят вдали от стыка североамериканской и тихоокеанской плит. В первом случае (также, как и в Японии) это говорит о меньшей антиперситенстности сигналов, а увеличенная ширина спектра сингулярности свидетельствует о более сложной локальной структуре временных рядов GPS – многообразии показателей Гельдера-Липшица, нехарактерного для остальной территории США.

В завершение мультифрактального анализа сейсмической активности земной поверхности рассмотрим карты корреляционной размерности временных рядов GPS для США и Японии, построенных уже без скользящего временного окна в период с 1 января 2009 года для США и в период до и после землетрясения в Тохоку для Японии.

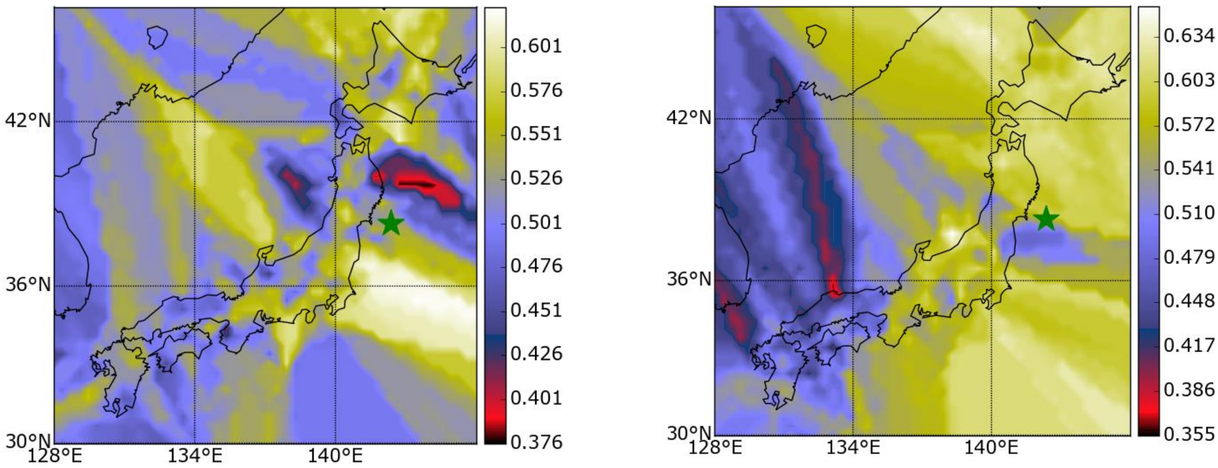
До землетрясения **После землетрясения**



а) Смещение восточных координат



б) Смещение северных координат



в) Смещение вертикальных координат

Рисунок 3.5.5. Карты корреляционной размерности до и после землетрясения в Тохоку

На рисунке 3.5.5 видно, что до землетрясения выделена пониженными значениями корреляционной размерности все та же зона, расположенная к северо-востоку от эпицентра землетрясения. Это означает, что в этот период времени (с 30 января по 9 марта 2011 года) смещения литосферных происходило неравномерно, то есть в выделенной аномальной зоне была повышенная тектоническая активность, что вызывало рост числа скачков (рисунок 3.2.1), выбросов (рисунок 3.3.1) и мощности низкочастотных колебаний (рисунок 3.3.1). После катастрофы, когда накопленное напряжение на стыке плит было разряжено, значения корреляционной размерности немного выровнялись, что говорит о стремлении тектонических смещений распределиться равномерно. Несмотря на то, что корреляционная размерность считалось не для восстановленного аттрактора, а для самих временных рядов GPS, результаты, полученные в работах по анализу распределения гипоцентров землетрясений, подтверждаются [6], [69], [70]. Построим снова карты первой главной компоненты, чтобы увидеть насколько коррелируют друг с другом карты разных компонент временных рядов GPS и тем самым усреднить полученные выше результаты.

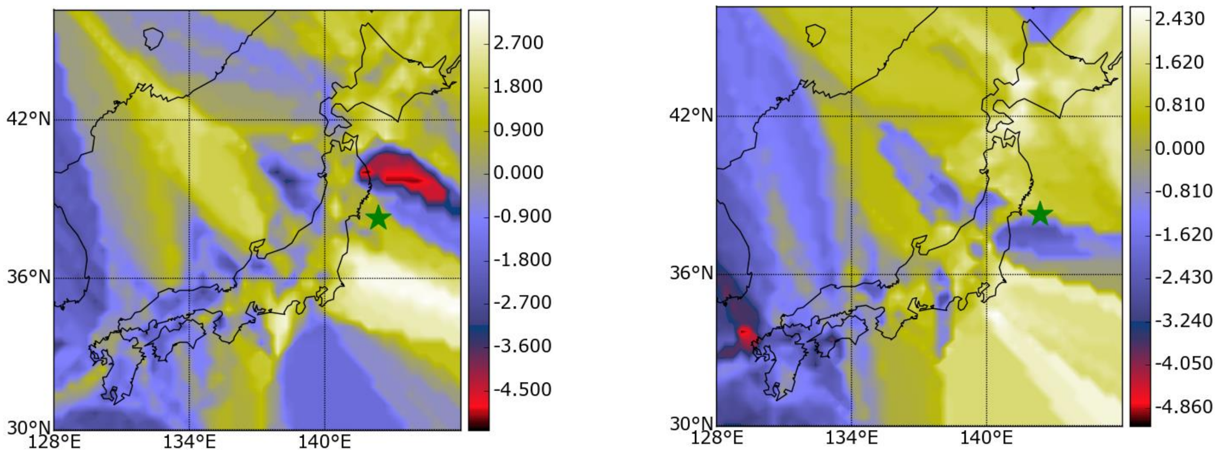
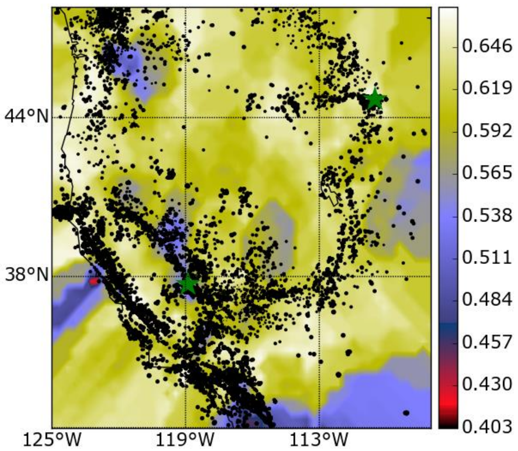
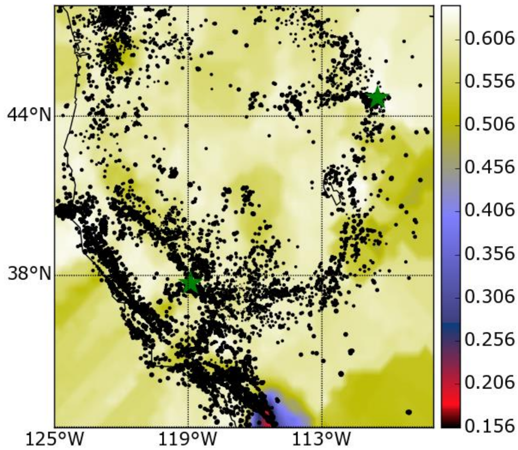


Рисунок 3.5.6. Карты первой главной компоненты для карт корреляционной размерности до и после землетрясения в Тохоку

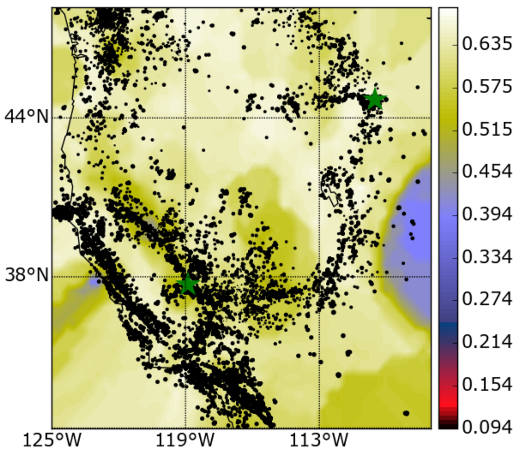
Взглянем теперь на карты корреляционной размерности западной территории США, построенные на основе данных, полученных за последние 7 лет.



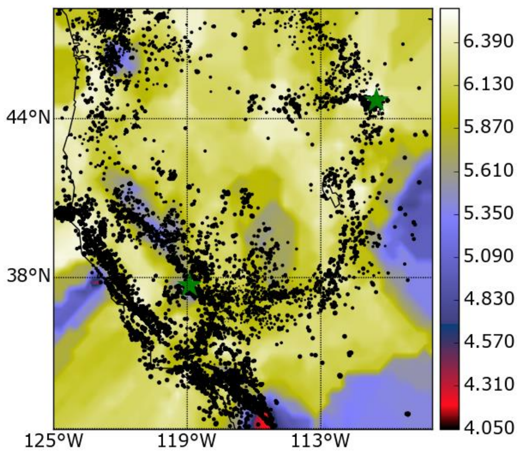
а) Восточные координаты



б) северные координаты



в) вертикальные координаты



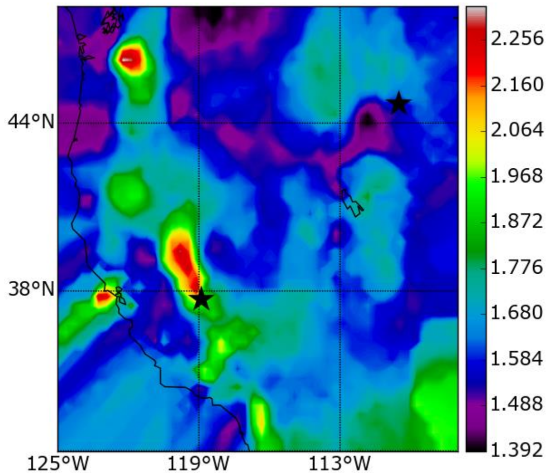
г) первая главная компонента

Рисунок 3.5.7. Карты корреляционной размерности для западной территории США, построенные в период с 1 января 2009 по 1 декабря 2015

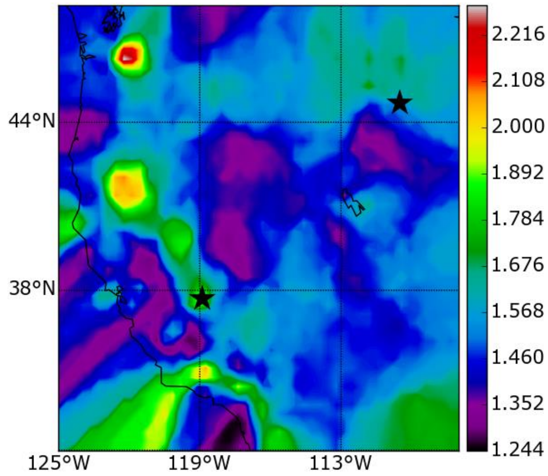
Из рисунка 3.5.7 видно, что сильно низких значений корреляционной размерности нет, откуда можно сделать вывод, что не смотря на повышенную сейсмическую активность в зоне системы разломов, напряжение в них не возрастает пока что до критических размеров, так как оно разряжается ежегодно в виде тысяч мелких землетрясений магнитудой 1-2, однако следует более внимательно относиться к этому региону, так как из карты первой главной компоненты видно, что малая область низких значений корреляционной размерности временных рядов GPS расположена в узле, через которую проходит разлом Сан-Андреас. Но, если учесть, что данная область мала, то можно предположить, что такой результат вызван какой-либо другой причиной.

3.6. Карты логарифма дисперсии шума

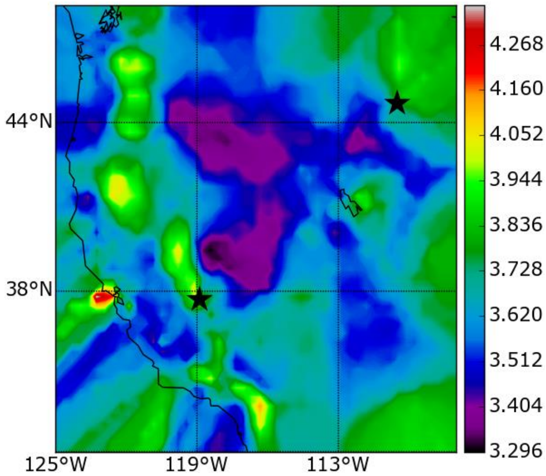
Помимо предложенного выше множества алгоритмов выделения зон с предполагаемыми интенсивными движениями земной коры, аналогичные области в западном регионе США выделяются для каждой компоненты сигнала GPS и на картах более простой статистики: логарифма дисперсии шума. Значения вычислялись после перехода к приращениям, таким образом, каждой станции GPS сопоставлялось значение $\ln \sigma^2$, где σ^2 – дисперсия приращений исходного временного ряда. Интерпретировать данный результат значительно проще, так как высокие значения логарифма говорят о том, что в данном регионе амплитуда колебаний выше, чем на другой территории. Использование логарифма обусловлено тем, что данный подход позволяет сгладить значения, тем самым имеющиеся выбросы среди значений дисперсии не будут сильно портить картину. Ниже на рисунке 3.6.1 продемонстрированы карты пространственного распределения логарифма дисперсии шума, построенные в период с 1 января 2009 года по 1 декабря 2015 года.



а) восточные координаты



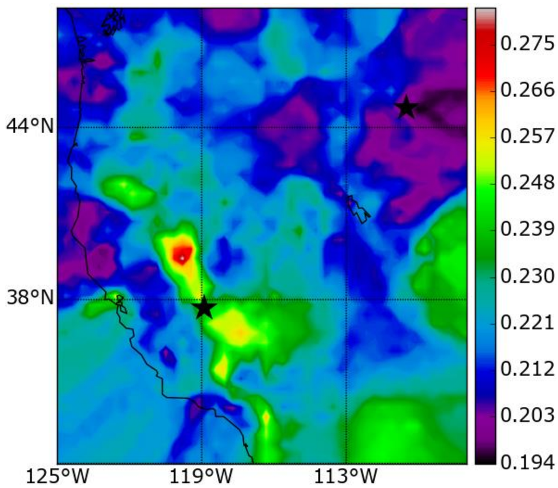
б) северные координаты



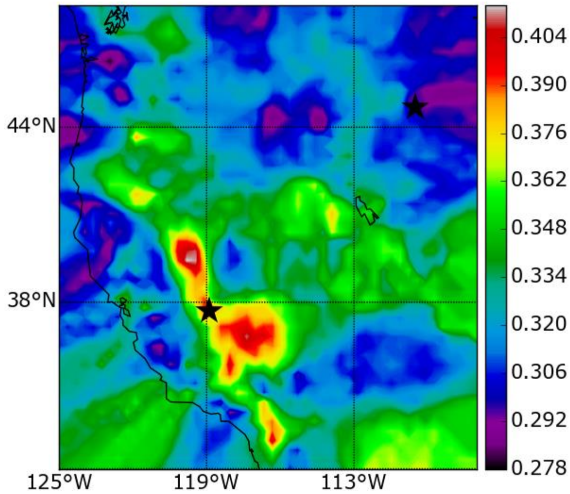
в) вертикальные координаты

Рисунок 3.6.1. Карты логарифма дисперсии шума для западной территории США

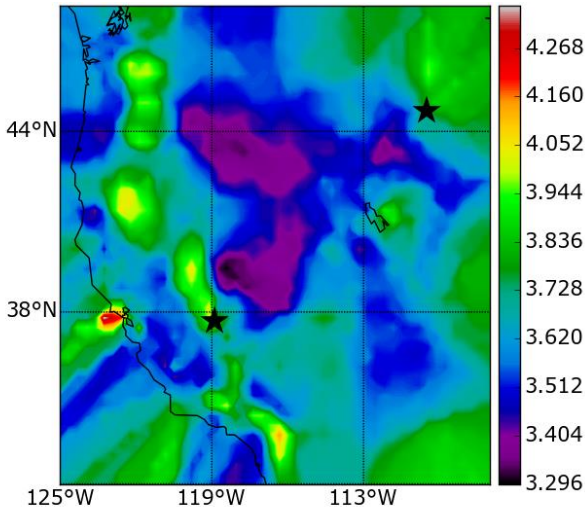
В качестве итоговой картины по проведенной работе можно использовать приведенные ниже карты пяти различных статистических характеристик для вертикальных смещений, демонстрирующие полученные результаты для западной территории США:



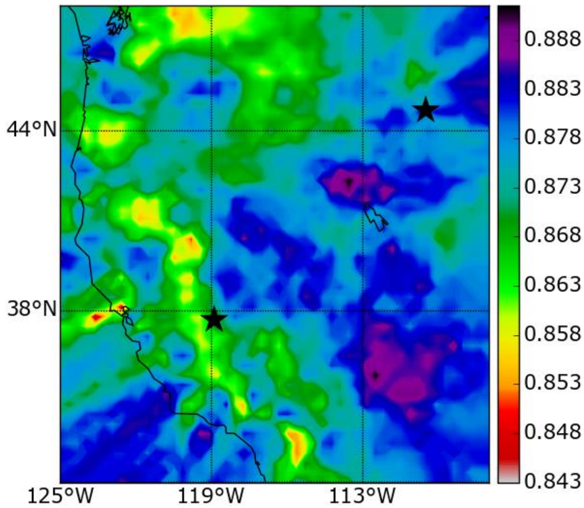
Обобщенный показатель Херста



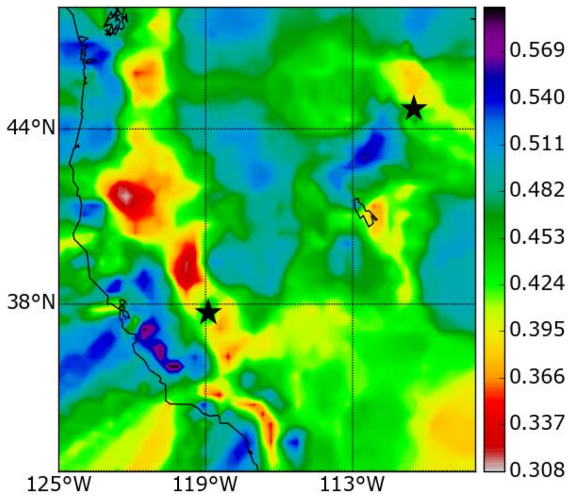
Ширина спекртра сингулярности



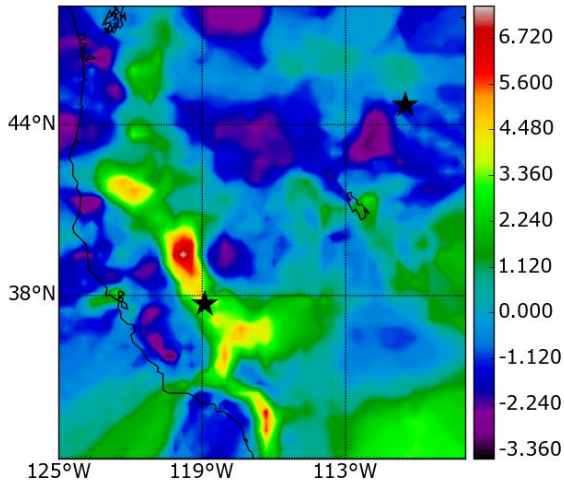
Логарифм дисперсии



Нормализованная энтропия скачков



Отношение максимального собственного значения корреляционной матрицы к сумме всех собственных значений



Первая главная компонента всех полученных карт

Рисунок 3.6.2. Карты пространственного распределения различных статистик для вертикальных координат временных рядов GPS

Заключение

В третьей главе рассмотрены результаты работы предложенных во второй главе методов выделения особенностей во временных рядах GPS с целью выделения наиболее опасных регионов Земли с повышенной сейсмической активности, а также зоны интенсивных смещений земной поверхности. Как показали результаты, построение карт распределения всех предложенных статистик позволило выделить область подготовки землетрясения на основе анализа тридцатиминутных данных GPS. В преддверии землетрясения очень точно удалось выделить эпицентр будущего землетрясения с помощью построения карт спектральной экспоненты вейвлет-преобразования. Для сейсмически нестабильной зоны оказались характерны высокие значений спектральной экспоненты, указывающие на рост низкочастотных колебаний временных рядов GPS после перехода к приращениям.

В подтверждение предположения, что перед началом землетрясения возможно уменьшение хаотичности в поведении земной поверхности, выступили карты нормализованной энтропии скачков и выбросов. Эпицентр на этих картах не был определен так же точно, как на картах спектральной экспоненты, однако, очевидно, что область пониженных значений тяготит к сейсмически неустойчивой зоне. Данное смещение поддается следующему объяснению: основываясь на том, что как на карте выбросов, так и на карте скачков выделена одна и та же область пониженных значений нормализованной энтропии, можно сделать вывод, что данная зона соответствует тому месту, где породы, находящиеся на стыке литосферных плит, начали разрушаться первыми. Глядя на идеализированную модель землетрясения, изображенную на рисунке 1.3.1, видно, что моменту превышения силы трения соответствуют как выброс на графике скорости, так и образование «ступеньки» на графике смещения, таким образом, логично ожидать того же эффекта на примере литосферных плит, где на их стыке расположены породы ограниченной крепости, способные резко разрушиться под действием высокого давления.

Карты выбросов и скачков построены на основе новых методов анализа временных рядов – псевдо-производной и взвешенной суммы приращений стандартных отклонений. Первый является быстрым и удобным методом для построения кусочно-ступенчатой аппроксимации рядов, используя в качестве разрыва точки экстремумов или точки перегиба временного ряда. Точки перегиба возможно найти благодаря сходству свойств псевдо-производной с классическим определением производной. Данный метод оказался быстрее, чем имеющиеся методы построения кусочно-ступенчатой аппроксимации методом WTMM и с использованием дискретного вейвлет-преобразования Хаара. Взвешенная сумма приращений стандартных отклонений позволила сопоставить исходному ряду другой вспомогательный временной ряд, содержащий в себе ярко выраженные выбросы, если таковые имеются. В силу того, что стандартное отклонение очень чувствительно к трендовой составляющей, в настоящей работе приведен анализ временных рядов после перехода к приращениям. Тем не менее данный подход позволил указать на аномально высокую концентрацию выбросов во временных рядах GPS по всем компонентам. Для вычисления взвешенной суммы приращений стандартных отклонений для длинных временных рядов были выведены рекуррентные формулы. В качестве вспомогательного математического инструмента для интерпретации полученных данных о скачках и выбросах было использовано понятие информационной энтропии, позволяющей оценить меру хаотичности сигналов. Аномальными считаются те зоны, для которых значение нормализованной энтропии принимают низкие значения (для японского региона ниже 0.9), что указывает на уменьшение естественного и здорового хаоса в связи с образованием некоторого механизма, порождающего новые локальные структуры временных рядов GPS.

Чтобы оценить адекватность и применимость предложенных методов, в работе также приведены карты максимального собственного значения корреляционных матриц и мультифрактальных параметров – методов, которые в настоящее время широко используются для анализа временных рядов различной природы. Карты максимального собственного значения (а точнее его отношения к сумме всех собственных значений корреляционной матрицы) позволяют выделить участки повышенной или, наоборот, пониженной синхронизации смещений земной поверхности, что продемонстрировано на карте западной территории США. На ней видно, что территория вдоль тихоокеанского побережья, где расположена одна из самых больших систем разломов, характеризуется низкими значениями корреляций – это является следствием трения тектонических плит друг о друга, в то время как на самой североамериканской плите преобладают более высокие собственные значения корреляционной матрицы, так как она движется целиком.

Мультифрактальный анализ использовался в работе для построения карт таких характеристик как обобщенный показатель Херста, ширина спектра сингулярности и корреляционная размерность временных рядов GPS. Карты обобщенного показателя Херста позволили выделить регионы с повышенным числом случайных смещений как для Японии, так и для США, совпадающими с регионами, определенными на картах скачкообразной составляющей сигналов GPS.

Появление этих случайных смещений обусловлено высоким значением ширины спектра сингулярности, свидетельствующим о том, что временные ряды GPS в данной области земной поверхности имеют более сложную локальную структуру.

Карты корреляционной размерности, применительно к Японии, являются еще одним подтверждением того, что в преддверии катастрофы в области, непосредственно близкой к эпицентру землетрясения, имела место аномальная активность земной коры. Пониженные значения корреляционных размерностей за полтора месяца (возможно, и ранее, однако не имеется доступных данных для анализа за более ранний промежуток времени) до землетрясения можно интерпретировать, как неравномерное распределение смещений литосферных плит по всем трех компонентам временных рядов GPS. После землетрясения на картах видно, что накопленное напряжение спало, и смещения коры распределились более равномерно, в то время как карты скачков и выбросов по-прежнему указывают на наличие большого количества случайных резких смещений, вызванных афтершоковыми толчками, которые еще на протяжении нескольких недель появлялись в этом регионе.

На картах корреляционной размерности США также можно разглядеть слабо выраженную зону низких значений, расположенную вдоль тихоокеанского побережья, однако данные значения не являются критически низкими в силу того, что постоянно накапливающееся напряжение в данном регионе разряжается ежегодно в виде сотен или тысяч слабых землетрясений магнитудой 1-2. Зоны низких значений, расположенных в центральной части США, требуют дополнительного изучения как с математической точки зрения, так и с точки зрения геологии для более корректной интерпретации полученных результатов.

Удобным обобщением полученных результатов, для всех построенных в настоящей работе карт являются карт первых главных компонент, так как они позволяют определить, насколько коррелируют друг с другом все компоненты, и получить тем самым некоторый усредненный результат.

Таким образом, предложены методы анализа временных рядов GPS, которые на примере США позволяют выделить зоны интенсивных движений земной коры, а также предоставляют возможность составления прогноза о подготовке сильного землетрясения на примере землетрясения в Японии 11 марта 2011 года. К сожалению, на данный момент нет физической возможности протестировать эти алгоритмы на примере других землетрясений, так как сильные толчки происходят сравнительно редко, а если и происходят, то в регионах (например, землетрясение в Чили 17 сентября 2015 года магнитудой 8.3), где практически не имеются установленные GPS станции (на момент землетрясения всего 20 штук), что делает картину непригодной для интерпретации.

Также важным фактором для анализа предложенными методами является временной промежуток, с которым происходит регистрация данных. Чем он меньше, тем лучше можно получить результат, так как, например, дневные данные являются малополезными для извлечения из них информации о выбросах и скачках, так как данные аномалии происходят за сравнительно короткий промежуток времени. Однако при соблюдении упомянутых условий все предложенные в настоящей работе методы в совокупности представляют новый перспективный инструмент для мониторинга сейсмической активности земной поверхности.

Область дальнейших исследований

Полученные в работе результаты открывают множество направлений для продолжения исследований в области анализа случайной составляющей временных рядов.

Предложенное понятие псевдо-производной является новым и, как следствие, может иметь дополнительные, еще неизученные, полезные свойства и применения помимо быстрого построения кусочно-ступенчатой аппроксимации. Возникает вопрос, существует ли возможность полного или частичного восстановления исходного временного ряда по его псевдо-производной с точностью до константы? Если так, то необходимо ввести понятие псевдо-первообразной. Данный факт может быть полезен, например, в следующем. Напомним, что одним из свойств псевдо-производной является то, что ее среднее значение соответствует коэффициенту углу наклона линейной тенденции на заданном интервале. Для построения кусочно-ступенчатой аппроксимации использовалось так называемое неявное удаление этих тенденций путем вычитания из нее скользящего среднего, таким образом, временной ряд, восстановленный после такого неявного удаления тренда может выступить в качестве оценки остаточного сигнала и должен быть сравнен с другими метода удаления тренда, причем необязательно линейного.

Взвешенная сумма стандартных отклонений также представляется интерес с точки зрения определения выбросов для исходного временного ряда, то есть без перехода к приращениям. Это позволит выделять не только резкие смещения (выбросы), но и более постепенные, которые зарегистрированы в течение некоторого числа последовательных дискретных моментов времени. Это является актуальным вопросом, так как во временных рядах GPS с малым шагом дискретизации число таких плавных скачков сравнительно велико. То же самое можно сказать и про финансовые временные ряды, где значение на одном временном масштабе является выбросом, а на другом – серией однонаправленных выбросов.

Мультифрактальный анализ временных рядов, несмотря на рост его популярности, также имеет расширение его применения. Так, например, научный интерес представляет предложенный в [72] несимметричный анализ спектра сингулярности применительно к временным рядам GPS, то есть рассчитанный вдоль максимумов и минимум непрерывного вейвлет-преобразования в отдельности. Данный подход позволяет работать с удвоенным числом мультифрактальных параметров: обобщенного показателя Херста и спектра сингулярности. Установление связи между ними и картирование пространственного распределения полученных характеристик может помочь изучить структуру сигналов еще более детально.

Физический смысл формы графика спектра сингулярности также является одним из неизученных с точки зрения его значимости. С математической точки зрения все более или менее ясно: в случае, например, мультипликативного биномиального процесса форма «колокола» является симметричной, что говорит о равномерном распределении точек множества, имеющих один и тот же скейлинговый показатель Гельдера-Липшица. Если «колокол» смещен в ту или иную сторону – данное распределение неравномерно, однако вопрос интерпретации этого факта применительно к временным рядам GPS пока что остается открытым.

Список литературы

1. Антонов В.И., Загайнов А.И., Ва ван Куанг. Автоматизированное программное обеспечение для численных мультифрактальных исследований // Научно-технические ведомости СПбГПУ, Т. 2, № 169, 2013.
2. Б.А.Сулейманов, Ф.С.Исмайлов, О.А.Дышин, Н.И.Гусейнова. Мультифрактальный анализ состояния разработки нефтяного месторождения // Нефтегазовое дело, Т. 2, 2012. С. 20-26.
3. Безручко Б.П., Смирнов Д.А. Реконструкция обыкновенных дифференциальных уравнений по временным рядам. Учебно-методическое пособие. Саратов: ГосУНЦ “Колледж”, 2000.
4. Безручко Б.П., Смирнов Д.А. Современные проблемы моделирования по временным рядам // Вестник ХПИ, № 145, 2007. С. 34-45.
5. Бондур В.Н., Смирнов В.М. Мониторинг вариаций ионосферы в период подготовки и прохождения землетрясений по данным спутниковых навигационных систем // 31st International Symposium on Remote Sensing of Environment. Санкт-Петербург. 2005.
6. Воропаев П.В. Изменение во времени корреляционной размернсти гипоцентров землетрясений Камчатки // Проблемы комплексного геофизического мониторинга Дальнего Востока России (тезисы докладов 2-й региональной научно-технической конференции). Петропавловск-Камчатский. 2009. С. 20-24.
7. Гиляров В.Л., Корсуков В.Е., Бутенко П.Н., Светлов В.Н. Применение вейвлет-преобравзования при изучении изменения фрактальных свойств поверхностей аморфных металлов под воздействием механической нагрузки // Физика твердого тела, Т. 46, № 10, 2004. С. 1806-1810.

8. Главные движущие силы землетрясений, дрейфа континентов и горообразования [Электронный ресурс] // Все о геологии: [сайт]. URL: http:// geo.web.ru/db/msg.html?mid=1174973
9. Голяндина Н.Э. Метод «Гусеница»-SSA: анализ временных рядов: Учеб. пособие. СПб: Изд-во СПбГУ, 2004.
10. Голяндина Н.Э. Метод «Гусеница»-SSA: прогноз временных рядов: Учеб. пособие. СПб: Изд-во СПбГУ, 2004.
11. Гудков Г.В., Пенжоян Г.А., Туриченко О.В. Мультифрактальная природа сердечного ритма плода при его различных функциональных состояниях // Вестник новых медицинских технологий, Т. XII, № 3, 2006. С. 101-104.
12. Долгая А.А., Викулин А.В., Акманова Д.Р. О некоторых особенностях временных рядов очагов землетрясений и извержений вулканов // Материалы региональной конференции, посвященной Дню вулканолога, "Вулканизм и связанные с ним процессы". Петропавловск-Камчатский. 2014. С. 163-167.
13. Захаров В.С. Анализ динамических характеристик временных рядов смещений земной поверхности по данным GPS // Электромагнитные волны и электронные системы, № 5, 2004. С. 13-20.
14. Захаров В.С. Анализ корреляционной размерности временных рядов выделения сейсмической энергии // Сборник трудов студентов, аспирантов и преподавателей кафедры общей и прикладной геофизики Университета «Дубна», 2007. С. 76–84.
15. Ильяшенко Ю.С. Аттракторы и их фрактальная размерность. Москва: МЦНМО, 2005.
16. Касимова В.А., Любушин А.А. Перспективы использования данных сети широкополосных сейсмических станций для изучения сигналов синхронизации поля // Материалы X региональной молодежной научной

конференции «Исследования в области наук о Земле». Петропавловск-Камчатский. 2012. С. 115-126.

17. Лукашин Ю.П. Адаптивные методы краткосрочного прогнозирования временных рядов. Москва: «Финансы и статистика», 2003.
18. Любушин А.А., Копылова Г.Н., Касимова В.А., Таранова Л.Н. Мульти-фрактальные характеристики сейсмического шума на Камчатке // Проблемы комплексного геофизического мониторинга Дальнего Востока России. Тр. Четвертой научно-техн. конф. Обнинск. 2013. С. 382-386.
19. Любушин А.А., Копылова Г.Н., Касимова В.А., Таранова Л.Н. О свойствах поля низкочастотных шумов, зарегистрированных на камчатской сети широкополосных сейсмических станций // Вестник КРАУНЦ. Серия: Науки о Земле, № 26, 2015. С. 20-36.
20. Любушин А.А., Яковлев П.В., Родионов Е.А. Многомерный анализ параметров флуктуаций GPS сигналов до и после мегаземлетрясения 11 марта 2011 г. в Японии // Геофизические исследования, Т. 16, № 1, 2015. С. 14-23.
21. Любушин А.А., Яковлев П.В. Энтропийная мера скачкообразной составляющей временных рядов GPS // Физика Земли, Т. 52, № 1, 2016. С. 98-107.
22. Любушин А.А. Анализ данных систем геофизического и экологического мониторинга. Москва: Наука, 2007.
23. Любушин А.А. Карты свойств низкочастотных микросейсм для оценки сейсмической опасности // Физика Земли, № 1, 2013. С. 11-20.
24. Любушин А.А. Микросейсмический шум в минутном диапазоне периодов: свойства и возможные прогностические признаки // Физика Земли, № 4, 2008. С. 17-34.
25. Любушин А.А. Мультифрактальные статистики региональных и глобальных полей низкочастотных микросейсм // Проблемы комплексного

геофизического мониторинга Дальнего Востока России. Тр. Второй научно-техн. конф. Петропавловск-Камчатский. 2009. С. 11-17.

26. Любушин А.А. Прогноз Великого Японского землетрясения // Природа, № 8, 2012. С. 23-33.
27. Любушин А.А. Сейсмическая катастрофа в Японии 11 марта 2011 года. Долгосрочный прогноз по низкочастотным микросейсмам // Геофизические процессы и биосфера, Т. 10, № 1, 2011. С. 9-35.
28. Любушин А.А. Статистики временных фрагментов низкочастотных микросейсм: их тренды и синхронизация // Физика Земли, № 5, 2010. С. 86-96.
29. Любушин А.А. Тренды и ритмы синхронизации мультифрактальных параметров поля // Физика Земли, № 5, 2009. С. 15-28.
30. Любушин А.А. Фрактальный анализ временных рядов. Москва. 2006.
31. Мандельброт Б. Фрактальная геометрия природы. Москва: Институт компьютерных технологий, 2002.
32. Масловская А.Г, Барабаш Т.К., Бурдина А.И. Применение вейвлет-преобразования для цифровой обработки токов переполяризации сегнетоэлектриков с самоподобной структурой // Вестник Амурского государственного университета, № 53, 2011. С. 42-49.
33. Нагорнов О.В., Никитаев В.Г., Простокишин В.М., Тюфлин С.А., Проничев А.Н., Бухарова Т.И., Чистов К.С., Кашафутдинов Р.З., Хоркин В.А. Вейвлет-анализ в примерах. Москва: НИЯУ МИФИ, 2010.
34. Новицкий П. В., Зограф И. А. Оценка погрешностей результатов измерений. 2-е изд. Ленинград: Энергоатомиздат, 1991.
35. Олемской А.И., Борисюк В.Н., Шуда И.А., Багдасарян А.А. Мультифрактальный анализ временных рядов экономических систем // Журнал нано- и электронной физики, Т. 1, № 3, 2009. С. 82-88.

36.Орлов Ю.Н. Оптимальное разбиение гистограммы для оценивания выборочной плотности функции распределения нестационарного временного ряда // Препринты ИПМ им. М.В.Келдыша, № 14, 2013.

37.Павлов А.Н., Анищенко В.С. Мультифрактальный анализ сложных сигналов // Успехи физических наук, Т. 177, № 8, Август 2007. С. 859-876.

38.Павлов А.Н., Павлова О.Н. Анализ корреляционных свойств случайных процессов по сигналам малой длительности // Письма в ЖТФ, Т. 34, № 7, Апрель 2008. С. 71-78.

39.Павлова О.Н., Павлов А.Н. Мультифрактальное описание динамики нефронов // Известия вузов. Прикладная нелинейная динамика, Т. 19, № 2, 2011.

40.Розенберг Г.С., Гелашвили Д.Б., Иудин Д.И. Фрактальности Эйфелевой башни. // Самарская Лука: проблемы региональной и глобальной экологии, Т. 20, № 3, 2011. С. 174-191.

41.Федоров, М. В. Метод идентификации форм распределений малых выборок // Рос. хим. об-ва им. Д. И. Менделеева, Vol. XLVI, No. 3, 2002. pp. 9-11.

42.Хлюпин А.Н., Динариев О.Ю. Фрактальный анализ трехмерной микроструктуры пористых материалов // Журнал технической физики, Т. 85, № 6, 2015. С. 17-22.

43.Чернецкий В.И. Математическое моделирование стохастических систем. Петрозаводск: Издательство Петрозаводского государственного университета, 1994.

44.Шредер М. Фракталы, хаос, степенные законы: миниатюры из бесконечного рая. Ижевск: НИЦ "Регулярная и стохастическая динамика", 2001.

45. Юшко В.А. Применение многомерного статистического анализа для обработки данных электротеллурических наблюдений // Геофизический мониторинг и проблемы сейсмической безопасности Дальнего Востока

России. Труды региональной научно-технической конференции. Петропавловск-Камчатский. 2007. Т. 2. С. 227-231.

46. Яковлев П.В. Анализ многомерных временных рядов финансового мониторинга // XI Международная конференция «Новые идеи в науках о Земле». М.: МГРИ-РГГРУ. 9-12 апреля 2013. Т. 3. С. 268.
47. Яковлев П.В. Анализ шумовой составляющей сигналов GPS // VII Международная научная конференция студентов аспирантов и молодых ученых «Молодые – наукам о Земле». М.: МГРИ-РГГРУ, 15-17 апреля, 2014. Т.2. С. 208
48. Яковлев П.В. Энтропийная мера выбросов во временных рядах GPS // Геофизические исследования, Т. 17, № 1, 2016. С. 37-45.
49. Яковлев П.В., Любушин А.А. Выделение скачкообразной составляющей сигналов GPS путем построения кусочно-ступенчатых аппроксимаций методом псевдо-производных // XII Международная научно-практическая конференция "Новые идеи в науках о Земле". М.: МГРИ-РГГРУ, 2015. Т.2. С. 445-446.
50. Agnieszka Kitlas Golińska. Detrended Fluctuation Analysis in biomedical signal processing: selected examples // Studies in logic, grammar and rhethoric, Vol. 29, No. 42, 2012. pp. 107-115.
51. Alexandrov T. A method of trend extraction using singular spectrum analysis // RevStat, No. 1, April 2009. pp. 1-22.
52. C. Ordonez, J. Martinez, J.R. Rodriguez-Perez, A. Reyes. Detection of outliers in GPS mesurements by using functional-data analysis // Journal of Surveying Engineering, Vol. 137, No. 4, November 2011. pp. 150-155.
53. Charu C. Aggarwal. Outlier Analysis. Springer New York, 2013.
54. David L. Donoho, Iain M. Johnstone. Minimax estimation via wavelet shrinkage, Standord University, Standord, Technical Report 402, 1992.

55. Easterling, D. R., Peterson, T. C. A new method for detecting undocumented discontinuities in climatological time series // International Journal of Climatology, Vol. 15, 1995. pp. 369-377.
56. Ester M., Kriegel H.-P., Sander J., Xu X. A density-based algorithm for discovering clusters in large spatial databases with noise // Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining. June 1996. Vol. 2. pp. 169 - 194.
57. F. Takens, D. Rang, L.S. Young. Detecting stange attractors in turbulence // Lecture Notes. Mathematics, No. 898, 1981. pp. 366-381.
58. Gazeaux, J., et al. Detecting offsets in GPS time series: First results from the detection of offsets in GPS experiment // Journall of Geophysical Research: Solid Earth, Vol. 118, May 2013. pp. 2397–2407.
59. Geoffrey Blewitt, William C. Hammond, Corné Kreemer, Hans-Peter Plag, Seth Stein, Emile Okal. GPS for real-time earthquake source determination and tsunami warning systems // Journal of Geodesy, Vol. 83, No. 3, March 2009. pp. 335-343.
60. Gustavo H. Orair, Carlos H. C. Teixeira, Wagner Meira, Jr., Ye Wang, Srinivasan Parthasarathy. Distance-based outlier detection // Proceedings of the VLDB Endowment, Vol. 3, No. 1-2, September 2010. pp. 1469-1480.
61. Hancong Liu, Sirish Shah, Wei Jiang. On-line outlier detection and data cleaning // Computers and Chemical Engineering, No. 28, 2004. pp. 1635-1647.
62. Hans Alexandersson, Anders Moberg. Homogenization of swedish temperature data. Part I: Homogenity test for linear trends // International journal of climatology, Vol. 17, 1996. pp. 25–34.
63. Hawkins D. Identification of outliers. London: Chapman and Hall, 1980.
64. Jan W. Kantelhardt, Stephan A. Zschiegner, Eva Koscielny-Bunde, Armin Bunde, Shlomo Havlin, H. Eugene Stanley. Multifractal detrended fluctuation analysis of nonstationary time series // Physica A, Vol. 316, No. 1, 2002. pp. 87-114.

65. Jan W. Kantelhardt. Fractal and Multifractal Time Series // Encyclopedia of Complexity and Systems Science, April 2008. pp. 3754-3779.

66. Jessica Lin, Eamonn Keogh, Ada Fu, Helga Van Herle. Approximations to Magic: Finding Unusual Medical Time Series // Computer-Based Medical Systems. Proceedings. 18th IEEE Symposium on. June 2005. pp. 329-334.

67. John R. Lanzante. Resistant, robust and non-parametric techniques for the analysis of climate data: theory and examples, including applications to historical radiosonde station data // International Journal of Climotology, Vol. 16, 1996. pp. 1197-1226.

68. Karanjit Singh, Dr. Shuchita Upadhyaya. Outlier Detection: Applications And Techniques // International Journal of Computer Science Issues, Vol. 9, No. 3, January 2012. pp. 307-323.

69. Kei Murase. A Characteristic Change in Fractal Dimension Prior to the 2003 Tokachi-oki Earthquake (MJ = 8.0), Hokkaido, Northern Japan // Earth Planets Space, Vol. 56, 2004. pp. 401-405.

70. Kei Murase. The Characteristic Changes in Fractal Dimension Preceding the 2000 Tottori-ken Seibu Earthquake // Zisin, Vol. 2, No. 55, 2002. pp. 11-18.

71. Kondrashov D, Chil M. Spatio-temporal filling of missing points in geophysical data sets // Nonlinear Processes in Geophysics, No. 13, May 2006. pp. 151-159.

72. Kyoko Ohashi, Luı´s A. Nunes Amaral, Benjamin H. Natelson, Yoshiharu Yamamoto. Asymmetrical singularities in real-world signals // Physical Review E, No. 68, 2003.

73. Lyubushin A., Yakovlev P. Properties of GPS noise at Japan islands before and after Tohoku mega-earthquake // SpringerPlus, Vol. 3, July 2014. pp. 364-370.

74. Lyubushin A., Yakovlev P. Thin structure of GPS time series noise // 26th General Assembly of International Union of Geodesy and Geophysics (IUGG). Prague, June 22 - July 02, 2015.

75. Lyubushin A.A. Wavelet-based coherence measures of global seismic noise properties // Journal of seismology, April 2015.
76. Lyubushin A.A. Wavelet-based coherence measures of global seismic noise properties // Journal of Seismology, Vol. 19, No. 2, April 2015. pp. 329-340.
77. M. Pesentia, M. Pirasa. A modified forward search approach applied to time series analysis // DITAG - Land, Environment and Geo-Engineering, Vol. XXXVII, 2008. pp. 787-792.
78. Mallat S. A wavelet tour of signal processing. San Diego, London, Boston, N.Y., Sydney, Tokio, Toronto: Academic Press, 1998.
79. Markus M. Breunig, Hans-Peter Kriegel, Raymond T. Ng, Jörg Sander. LOF: identifying density-based local outliers // Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data. SIGMOD. June 2000. Vol. 29. pp. 93-104.
80. Mohito Sugiyama, Karsten M. Borgwardt. Rapid Distance-Based Outlier Detection via Sampling // Advances in Neural Information Processing Systems 26. 2013. pp. 467-475.
81. ODD2 Workshop on Outlier Detection & Description under Data Diversity // ACD SIGKDD International Conference on Knowledge Discovery and Data Mining. 2014.
82. Osorio Ivan, Lyubushin Alexey, Sornette Didier. Automated seizure detection: Unrecognized challenges, unexpected insights // Epilepsy Behav, Vol. 22, No. 1, December 2011. pp. 7-17.
83. Paweł Oświęcimka, Jarosław Kwapień, and Stanisław Drożdż. Wavelet versus detrended fluctuation analysis of multifractal structures // Physical Review E, Vol. 74, No. 1, July 2006.
84. Plamen Ch. Ivanov. Beyond 1/f: Multifractality in Human Heartbeat Dynamics // Nature, No. 399, June 1999. pp. 461-465.

85. S. Bruni, Susanna Zerbini, F. Raicich, M. Errico, E. Santi. Detecting discontinuities in GNSS coordinate time series with STARS: case study, the Bologna and Medicina GPS sites // Journal of Geodesy, Vol. 88, No. 12, 2014. pp. 1203-1214.
86. Sergei Rodionov, James E. Overland. Application of a sequential regime shift detection method to the Bering Sea ecosystem // ICES Journal of Marine Science, No. 62, 2005. pp. 328-332.
87. Sergei Rodionov. A sequential algorithm for testing climate regime shifts // Geophysical research letters, Vol. 31, No. 9, May 2004. P. L09204.
88. Songwon Seo. A Review and Comparison of Methods for Detecting Outliers in Univariate Data Sets // Master's Thesis, April 2006.
89. Stephane Jaffard. Mathematical Tools for Multifractal Signal Processing // Signal Processing for Multimedia, 1999. pp. 111-127.
90. Stephen D. Bay, Mark Schwabacher. Near Linear Time Detection of Distance-Based Outliers and Applications to Security // SIAM Data Mining Conference, Workshop on Data Mining for Counter Terrorism and Security, 2003.
91. Stephen D. Bay. Mining distance-based outliers in near linear time with randomization and a simple pruning rule // Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining. 2003. pp. 29-38.
92. T. Sauer, J.A. Yorke, M. Casdagli. Embedology // Journal of Statistical Physics, No. 65, 1991. pp. 579-616.
93. Varun Chandola, Arindam Banerjee, Vipin Kumar. Anomaly detection: A survey // ACM Computing Surveys (CSUR), Vol. 41, No. 3, July 2009.
94. Xingxing Li, Maorong Ge, Yong Zhang, Rongjiang Wang, Peiliang Xu, Jens Wickert, Harald Schuh. New approach for earthquake/tsunami monitoring using dense GPS networks // Scientific Reports, No. 3, September 2013.

95. Xuan Hong Dang, Barbora Micenková, Ira Assent, Raymond T. Ng. Local Outlier Detection with Interpretation // Machine Learning and Knowledge Discovery in Databases, Vol. 8190, September 2013. pp. 304-320.