

Федеральное государственное автономное
образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
Институт космических и информационных технологий
Кафедра прикладной математики и компьютерной безопасности

УТВЕРЖДАЮ
Заведующий кафедрой
_____ А. А. Кытманов
«_____» 2018 г.

БАКАЛАВРСКАЯ РАБОТА
01.03.04 – Прикладная математика
Визуализация данных геомониторинга для решения прогностических
задач

Руководитель _____ профессор, д.т.н. К.В. Симонов

Выпускник _____ А.Э.о Исмаилов

Красноярск 2018

РЕФЕРАТ

Выпускная квалификационная работа по теме «Визуализация данных геомониторинга для решения прогностических задач» содержит 37 страниц текстового документа, 2 приложения, 12 использованных источников.

ПРЕОБРАЗОВАНИЕ ФУРЬЕ, МЕТОД ГЛАВНЫХ КОМПОНЕНТ, МЕТОД УПРУГИХ КАРТ, СПЕКТРАЛЬНАЯ ПЛОТНОСТЬ, ГЕОДИНАМИЧЕСКИЙ МОНИТОРИНГ, КАТАЛОГ ЗЕМЛЕТРЯСЕНИЙ.

Объектом исследования являются каталоги крупнейших землетрясений и данные с сейсмоприемников, расположенных на юге Камчатки.

Предметом исследования является методы и алгоритмы визуализации многомерных данных.

Целью работы является повышение точности анализа и интерпретации данных геомониторинга сильных землетрясений в рамках полигонных исследований на основе методов аппроксимации и визуализации пространственно-временных рядов.

Задачи:

1. Изучение методов визуализации данных наблюдений для построения соответствующих моделей данных
2. Адаптация алгоритмического обеспечения и выполнение расчетов по визуализации каталога сильного землетрясения на основе метода упругих карт.
3. Аппроксимация и визуализация результатов спектрального анализа записей сейсмических колебаний с помощью логарифмических функций.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ.....	4
1 Визуализация данных	6
1.1 Визуализация многомерных данных.....	6
1.2. Классификация методов визуализации данных.....	7
1.3 Алгоритмы и формулы визуализации многомерных данных	9
1.4 Информационное моделирование с помощью упругих карт	10
1.5 Программное обеспечение	11
1.6 Пример применения метода упругих карт	13
2. Спектральный метод анализа данных.....	17
2.1 Теория спектрального анализа.....	17
2.2 Геодинамический мониторинг сейсмических сигналов	20
2.3 Анализ сейсмических данных геомониторинга.....	22
2.4 Интерпретация десятичного логарифма	25
2.5 Сравнение и анализ результатов расчетов	26
ЗАКЛЮЧЕНИЕ	29
Список используемых источников.....	30
Приложение А	31
Приложение Б	34

ВВЕДЕНИЕ

Одним из главных проблем в моделировании цунами, оценке высот волн вдоль побережья и прогноз цунами являются начальные условия в источнике цунами. Проблема актуальна на сегодняшний день и не вполне решена в силу малой возможности заблаговременного наблюдения в очаговой области. Именно современные данные сейсмологических наблюдений с их широчайшим покрытием позволяют с достаточно высокой точностью моделировать начальные условия [1].

Существует большое количество информационных данных, оформленных в виде больших массивов. Это таблицы экспериментальных исследований данных наблюдений за каким-либо объектом или явлением, данных социологических опросов и т.д. Число строк (параметров) в таких таблицах – десятки тысяч, а число столбцов (признаков) – несколько тысяч. Таблицы данных давно являются основной областью приложений для методов визуализации данных.

Визуализация данных – это представление данных в виде, который обеспечивает наиболее эффективную работу человека по их изучению. Находит широкое применение в научных и статистических исследованиях, в новостных сводках и аналитических обзорах, связанных с визуализацией информации, инфографикой, визуализацией научных данных и т.д.

К задаче визуализации данных сводится проблема представления в наглядной форме данных эксперимента или результатов теоретического исследования. Графики и диаграммы плохо справляются с задачей визуализации, когда возникает необходимость изобразить более трех параметров.

Целью работы является повышение точности анализа и интерпретации данных геомониторинга сильных землетрясений в рамках полигонных исследований на основе методов аппроксимации и визуализации пространственно-временных рядов.

Поставленная цель определила необходимость решения следующих задач:

1. Изучение методов визуализации данных наблюдений для построения соответствующих моделей данных.
2. Адаптация алгоритмического обеспечения и выполнение расчетов по визуализации каталога сильного землетрясения на основе метода упругих сеток.
3. Аппроксимация и визуализация результатов спектрального анализа записей сейсмических колебаний с помощью логарифмических функций.

Визуализация каталога данных сильных землетрясений осуществлялась с использованием метода упругих карт. Аппроксимация и визуализация результатов спектрального анализа данных регистрации сейсмических колебаний проводилась на основе применения логарифмических функций.

В работе используются данные каталогов крупнейших землетрясений, а также данные геолого-геофизических исследованиях, проведённых НП «ЭЦ РОПР» в рамках полигонных исследований на территории юга Камчатки. Расчеты выполнялись в соответствии заданиями А.А. Кабанова – руководителя группы сводной обработки НП «ЭЦ РОПР».

1 Визуализация данных

1.1 Визуализация многомерных данных

Задача визуализации информации в широком смысле слова насчитывает десятки тысяч лет. Наскальные изображения, первые карты местности, современная картография и климатология, томографические исследования, экономика регионов, системы хранения и поиска информации в сетях, распознавание образов, навигация в мобильной робототехнике, многовариантный анализ при проектировании технических систем – диапазон задач, которые объединены общим понятием визуализации.

Будем рассматривать методы визуализации применительно к обработке массивов данных в виде чисел, действительных или комплексных, на непрерывной или дискретной шкале, полученных в процессе проведения натурных или вычислительных экспериментов. Тем не менее, нельзя отрицать важность междисциплинарных связей, когда решение задачи одной предметной области плодотворно решается методами и средствами из другой области. Поэтому необходим обзор современных подходов к визуализации в широкой постановке.

Визуализация данных в самом простейшем случае предполагает изображение зависимости некоторой функции от одного параметра и может быть решена даже без применения вычислительных средств. Более сложной и практически важной задачей является получение функциональных зависимостей от нескольких и даже многих параметров. В двумерном случае для визуализации достаточно построить трёхмерный график, однако при большем количестве аргументов визуализация становится весьма проблематичной [2]. Поэтому усилия современных разработчиков направлены именно на решение визуализации многомерных данных. Это можно считать четвёртым аспектом проблемы визуализации данных.

Первым аспектом задачи визуализации является её практическое назначение, а именно, облегчение восприятия данных для их последующего анализа, например, получения закономерностей влияния параметров исследуемой системы. В этом смысле визуализация по существу является поддержкой принятия решений. Поэтому имеется неразрывная связь между прикладной задачей и способом визуализации. Для визуализации могут использоваться как первичные данные, так и данные, прошедшие предварительную обработку. Эти второй аспект визуализации. Его следует отличать от третьего аспекта — разработки собственно программных алгоритмов и кодов для реализации визуализации на средствах вычислительной техники [2].

1.2. Классификация методов визуализации данных

Обширнейший круг методов визуализации, отличающихся друг от друга теоретическими основами, алгоритмами, программными реализациями, и, конечно, областью использования, нуждается в классификации, которая в целом пока не была построена именно в силу великого разнообразия перечисленных составляющих.

Первый вид классификации может быть сделан по видам используемых данных. Данные, в том числе и те, что используются для визуализации, всегда могут быть представлены как совокупности, т. е. множество кортежей или многокомпонентных объектов данных. Некоторые совокупности могут быть интерпретированы как функции.

Разнообразие объектов и средств визуализации, можно объединить концепцией, предложенной Бертином. Суть её в том, что каждый график состоит из маркеров, которые позиционированы на плоскости и имеют 6 визуальных варьируемых атрибутов: размер, форма, ориентация, цвет, текстура, значение. Эти атрибуты также имеют тип шкалы и уровень (количество распознаваемых предметов). Имеются и другие свойства, относящиеся к

восприятию образов человеком. Классификация может быть сделана по визуализируемым объектам (рис. 1.1).

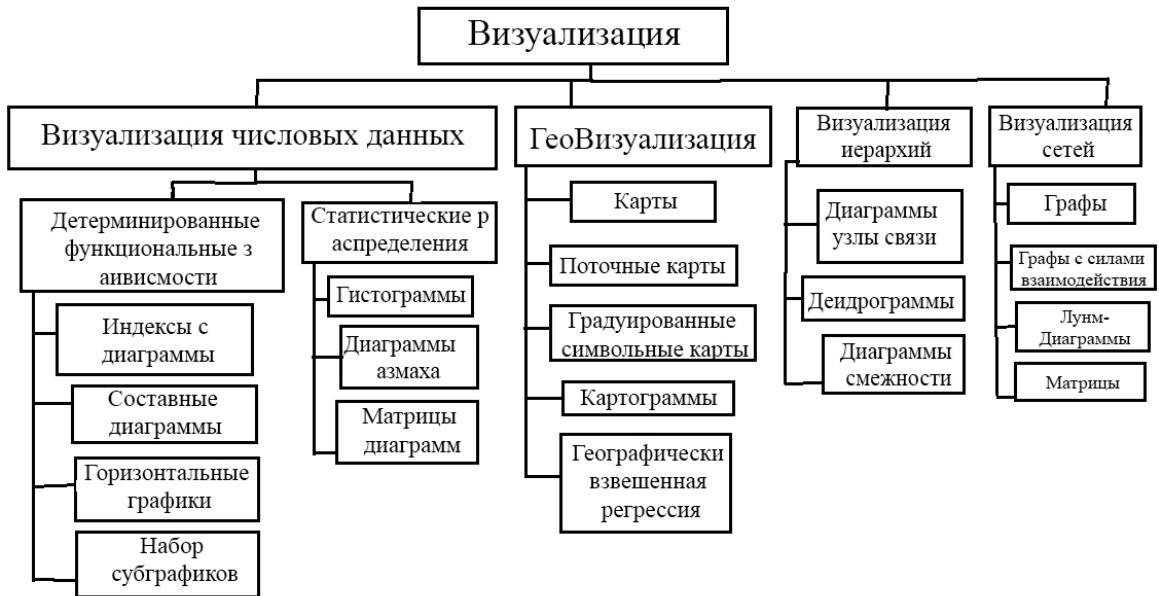


Рис. 1.1 – Классификация по объектам визуализации

Важно отметить, что исходные данные для визуализации часто подвергаются дополнительным преобразованиям. Это связано и с необходимостью извлечения информации для получения новых характеристик, изучаемого процесса или явления и собственно проблемами подготовки данных для лучшего восприятия. Поэтому третья классификация основана на способах преобразования координат и данных (рис. 1.2).



Рис. 1.2 – Классификация видов визуализации по способу преобразования координат и данных

Несмотря на то, что история современной визуализации по существу насчитывает не одно десятилетие, а в широком толковании и не одно столетие, совершенствование технологий визуализации происходит непрерывно и в наше время.

1.3 Алгоритмы и формулы визуализации многомерных данных

Метод главных компонент

Метод главных компонент применяется для снижения размерности пространства наблюдаемых векторов, без существенной потери информации. Предпосылкой метода главных компонент является нормальный закон распределения многомерных векторов. В методе главных компонент случайные величины линейных комбинаций находятся характеристическими векторами ковариационной матрицы. Основными компонентами являются ортогональные системы координат, в которых дисперсии характеризуют их статистические свойства.

Пусть задано начальное множество векторов X линейного пространства Lk . Применение метода главных компонент позволяет перейти к базису пространства Lm ($m \leq k$), таким образом, что: первая компонента (первый вектор базиса) соответствует направлению, по которому дисперсия векторов исходного множества максимальна. Направление второй компоненты (второго вектора базиса) выбирается таким образом, чтобы дисперсия начальных векторов вдоль нее была максимальной при условии ортогональности к первому вектору базиса.

Аналогично определяются и другие векторы базиса. В результате, направления векторов базиса выбираются таким образом, чтобы максимизировать дисперсию исходного множества вдоль первых компонент, называемых главными компонентами. Оказывается, что основная изменчивость векторов исходного множества векторов представлена первыми несколькими компонентами и становится возможным, отбрасывая оставшиеся компоненты, перейти в пространство меньшей размерности.

Результатом применения МГК является вычисление матрицы W размера $m \times k$, осуществляя проекцию векторов пространства L_k на подпространство, растянутое на главных компонентах:

$$Y = W * (X - \mu), Y \in L_m, X \in L_k,$$

где X – вектор из исходного множества, Y – координаты вектора в подпространстве главных компонент, μ – математическое ожидание вектора X начального множества.

Главные компоненты (векторы базиса), выбираемые с помощью МГК, обладают следующим свойством: обратная проекция вектора Y в L_k дает минимальную ошибку реконструкции (минимальное расстояние до образа вектора Y). Нужно отметить, что корректное применение МГК возможно лишь при предположении о нормальном распределении векторов исходного набора.

1.4 Информационное моделирование с помощью упругих карт

Мы построили карту данных, карту остатков, карту вторых остатков и т.д. Последовательность таких карт называется *информационной моделью данных*.

При моделировании данных с помощью многообразий на вход модели подается вектор пространства x , а на выходе удаляется имитированный вектор того же пространства X . Схема похожа на архитектуру нейронной сети, но существенное различие заключается в том, что среди компонентов $x_1, x_2, x_3, \dots, x_n$ вектора x могут быть пустые значения, а на выходе эти значения будут восстановлены.

Таким образом, можем произвольно разделить набор характеристик на вход и выход. Подавая на вход вектор, в который заполнены компоненты входного вектора, в выходные оставлены пустыми, на выходе можно снять восстановленные значения выходных признаков (рис. 1.3). Это открывает возможности для выявления взаимосвязей признаков, решения задач прогнозирования и построения регрессионных зависимостей между признаками [3].

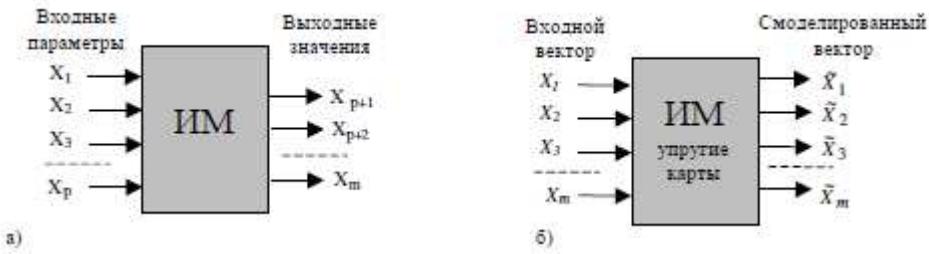


Рис. 1.3 – Два типа информационных моделей.

- а) стандартная нейросетевая схема моделирования;
- б) моделирование данных с помощью многообразий.

Визуализация данных.

Основной особенностью и преимуществом построения двумерных информационных моделей является возможность визуального представления данных и ошибок описания данных моделью. Любая из построенных карт позволяет визуально анализировать распределение самих данных или ошибок в описании [4].

Группирование объектов.

Используя карты, можем разделить объекты на группы. Это можно сделать визуально, оценивая компактность. Если деление на группы невозможно использовать. На карте накладывается двумерная непрерывная цветовая карта. Получите определенный цвет и точки. Полученные цвета могут использоваться, например, если точки данных могут быть размещены на географической карте. Тогда аналогичные значения знака точки на карте будут иметь одинаковые цвета.

1.5 Программное обеспечение

В программе Vida Expert имеется внутренние иерархии объектов [5]. Некоторые из них соответствуют тем объектам, с которыми оперирует исследователь на практике, другие объекты являются контейнерами,

содержащими и упорядочивающими объекты исследования. На рисунке 1.4 изображены объекты системы и отношения между ними.

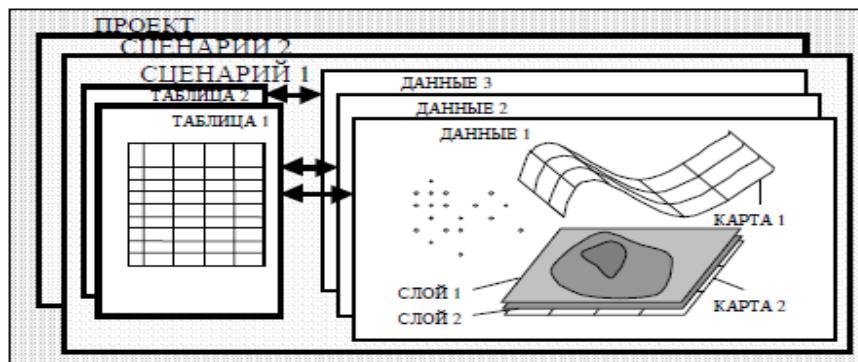


Рис. 1.4 – Внутренняя структура объектов программы Vida Expert

Загрузив таблицу данных, пользователь на основе таблицы, выбирая необходимые числовые поля и указывая способ их нормировки, создаёт объект данных, которые содержат числовой массив всех значений выбранных признаков. В дальнейшем данные сохраняют связь с таблицей, на основе которого он был создан. Используя одну и ту же таблицу, можно создавать различные массивы данных, выбирая разные наборы строк, признаков и способы их нормировки.

На основе данных можно создать настроенные карты, которые хранятся в нём как в контейнере. Они содержат всю необходимую информацию о положении узлов сетки в пространстве.

Для визуализации данных на основе методов создаются информационные слои. В программе имеется 4 вида слоёв, это слой точек данных, сетки, раскрасок и объектов. Каждый слой имеет характеристику. В программе реализовано 4 варианта видов:

- 1) на координатные плоскости;
- 2) на плоскость главных компонент;
- 3) во внутренних координатах карты ;
- 4) во внутренних координатах карты.

В 3-ем пункте карта предстаёт в виде равномерной сетки узлов, точки данных размещены в соответствии с их проекциями на карту. В 4-ом – пункте карта изображается в виде криволинейной сетки. Отдельный диалог позволяет настраивать криволинейную развёртку так, чтобы расстояние между узлами сетки на плоскости как можно точнее соответствовали расстоянию между узлами в исходном пространстве. На карту может быть наложена цветовая гамма, зависящая от выбранного признака и т.д [6].

1.6 Пример применения метода упругих карт

В данном разделе рассматривается использование методов визуализации, а именно, метода упругих карт в одной из актуальных задач геофизического мониторинга: задаче автоматического моделирования глубины и магнитуды землетрясения. Данное исследование важно и включено в работу для демонстрации того, каким образом методы визуализации данных могут обеспечить «подсказку» исследователю в предметной области.

Для примера анализируются данные землетрясения, произошедшего 8 сентября 2017 г. в Тихом океане у южного побережья Мексики. Землетрясение имело магнитуду 8.1 баллов. Эпицентр подземных толчков находился примерно в 100 км к юго-западу от города Пьеджиапан и на глубине 35 км (рис. 1.5 – 1.6).



Рис. 1.5 – Местоположение эпицентра землетрясения у берегов Мексики
8 сентября 2017 г.

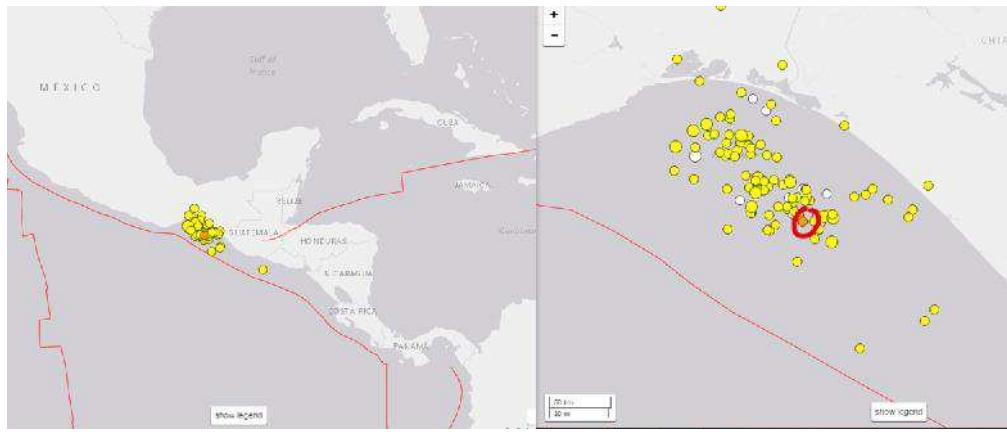


Рис. 1.6 – Эпицентры афтершоков с 8 по 12 сентября. Данные с сайта Геофизической службы США. Точка доступа: <http://earthquake.usgs.gov>.

Всего за 4 дня произошло 126 сейсмических событий. Все эти данные обработаны и импортированы непосредственно в программу (рис. 1.7).

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	time,latitude,longitude,depth,mag,magType,net,gap,dmin,rms,id,updated,place,type,horizontalError,depthError,magError,magNst,status,locationSource,magSource																		
2	2017-09-09T05:17:21.970Z,15.3517,-93.0999,67.854,mb,,206,2,203.9,96,us,us2000air7,2017-09-09T05:28:59.040Z,"84km SSW of Tres Picos, Mexico","earthquake",14,24.4,0.138,14,reviewed,us,us																		
3	2017-09-10T05:51:28.150Z,15.1196,-94.2119,45.354,mb,,161,2,308.8,67,us,us2000aj75,2017-09-10T06:34:44.040Z,"107km SSW of Paredon, Mexico","earthquake",9.3,14.7,0.077,45,reviewed,us,us																		
4	2017-09-11T04:48:12.710Z,15.2066,-93.7966,78.544,mb,,205,2,142.0,77,us,us2000ajuh,2017-09-11T05:29:46.040Z,"78km SSW of Tres Picos, Mexico","earthquake",14.6,21.2,0.108,23,reviewed,us,us																		
5	2017-08-09T09:05:13.330Z,14.4507,-92.6844,41.24.1,mb,,169.0,961.0,83,us,us2000aau,2017-09-07T12:33:27.025Z,"41km SW of Puerto Madero, Mexico","earthquake",5.9,12.6,0.08,42,reviewed,us,us																		
6	2017-08-24T05:09:53.650Z,14.4105,-93.3417,10.4,1,mb,,175,1,292.1,11.8,us,us2000ab7f,2017-09-09T05:34.040Z,"104km WSW of Puerto Madero, Mexico","earthquake",5.3,2.0,0.061,73,reviewed,us,us																		
7	2017-08-26T05:51:57.950Z,15.4615,-92.1793,175.22,4.1,mb,,122.0,668.0,68,us,us2000aeq2,2017-09-13T11:50:32.0,402Z,"11km NNE of Motozintla de Mendoza, Mexico","earthquake",7.9,8.5,0.158,11,reviewed,us,us																		
8	2017-09-01T13:43:52.100Z,15.6132,-93.1129,100.15,4.1,mb,,198.1,583.0,91,us,us2000adu7,2017-09-09T07:56:36.0,402Z,"14km ESE of Pijijiapan, Mexico","earthquake",13.6,17.9,0.117,20,reviewed,us,us																		
9	2017-09-08T16:12:19.190Z,15.3669,-94.2376,61.33,4.1,mb,,182.2,595.1,1,us,us2000aixn,2017-09-09T18:11:21.0,402Z,"83km SSE of Paredon, Mexico","earthquake",10.8,11.7,0.046,132,reviewed,us,us																		
10	2017-09-08T23:22:32.040Z,15.0785,-94.5585,47.75,4.1,mb,,206,2,445,1,36,us,us2000aicg,2017-09-16T02:59:27.0,402Z,"110km SSE of Paredon, Mexico","earthquake",9.4,19.6,0.077,46,reviewed,us,us																		
11	2017-09-09T00:23:20.330Z,16.0661,-94.3911,39.18,4.1,mb,,181,2,879,1.14,us,us2000aihy,2017-09-15T22:58:22.0,402Z,"33km SSE of Ixhuatan, Mexico","earthquake",6.5,16.4,0.086,37,reviewed,us,us																		
12	2017-09-09T01:14:06.760Z,15.3883,-94.006,45.61,4.1,mb,,211.2,393,1.08,us,us2000aiaic,2017-09-09T01:28:14.0,402Z,"72km SSE of Paredon, Mexico","earthquake",11.9,23.2,0.2,14,reviewed,us,us																		
13	2017-09-09T07:00:48.530Z,15.7397,-94.4832,55.79,4.1,mb,,199.2,903.0,82,us,us2000alpa,2017-09-09T07:17:09.0,402Z,"57km SSE of San Francisco del Mar, Mexico","earthquake",13.6,21.6,0.106,24,reviewed,us,us																		

Рис. 1.7 – Интерфейс работы с данными

Как видно из таблицы на рисунке, время переведено в секунды. Начало отсчёта 00:00 часов 8 сентября. Таким образом, землетрясение с максимальной амплитудой произошло в 4:49:21 (на 17361 с).

Далее приступаем к построению пространственной карты. На рисунке 1.8 а построена карта по исходным параметрам выбранным мною: долгота, широта и время. Следующий шаг – визуализация этих данных в двумерном пространстве главных компонент (рис. 1.8 б).

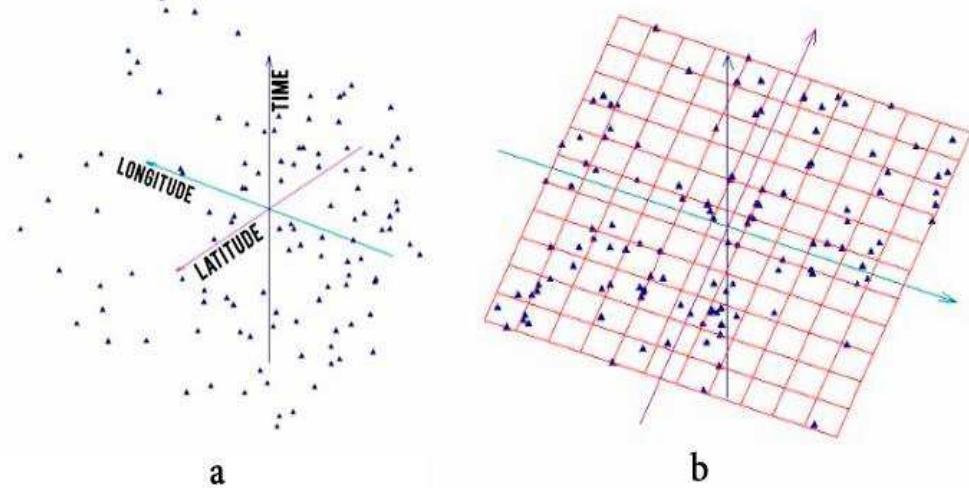


Рис. 1.8 – а) первоначальный вид карты; б) 3D представление массива данных в пространстве главных компонент

Можно изменить цвета определённых точек или атрибут. В данном случае выбрано в качестве особых точек – по дням. Первый день отмечен на карте красным, второй – желтым, третий зеленым и четвертый день – соответственно, синим цветом (рис. 1.9 – 1.11).

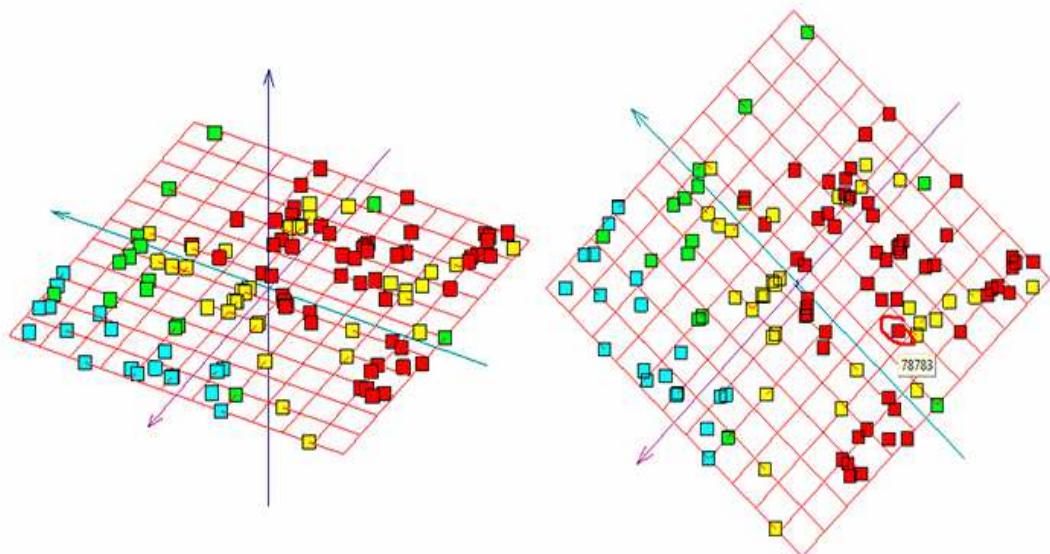


Рис. 1.9 – Расположение массива данных на «подложке» из главных компонент, компоненты массива кодируются цветом

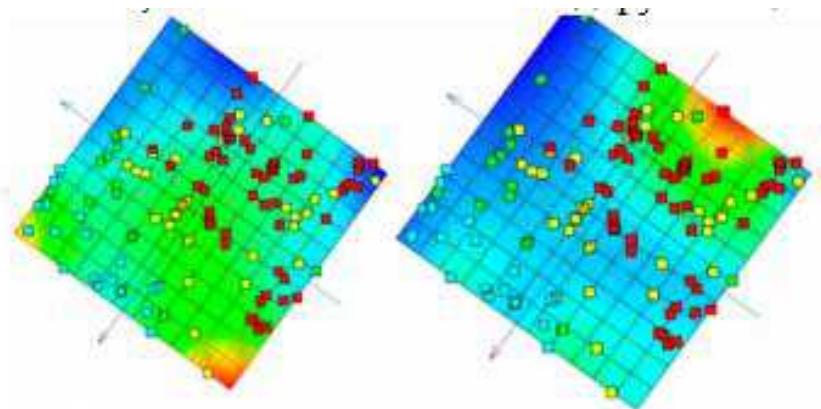


Рис. 1.10 – Результаты информационного моделирования данных с дополнительным указанием зависимости от: времени и глубины землетрясения

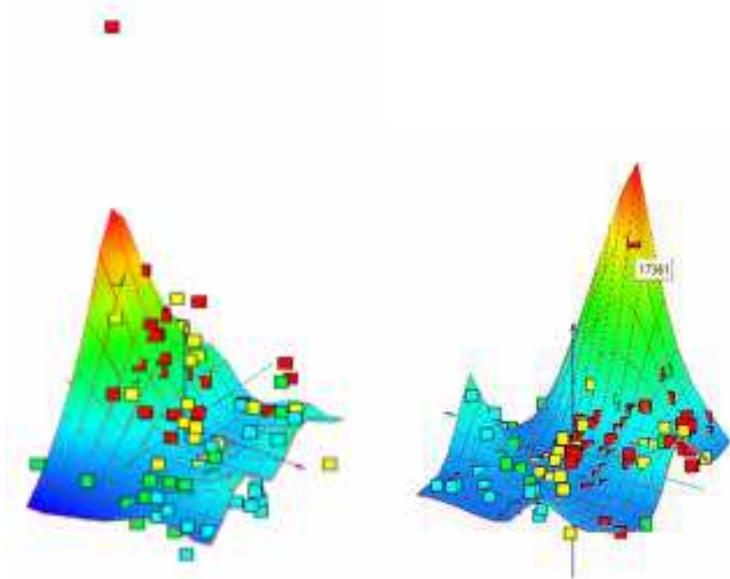


Рис. 1.11 – Трёхмерное представление результатов информационного моделирования данных

Особенностью предложенных методов отображения данных является то, что данные могут содержать пробелы, отсутствующие или некорректные значения отдельных объектов, такие точки данных могут на карту наносится. Также карта служит информационной моделью данных. Модель может решить важную задачу заполнения пробелов. Эта способность может быть использована для составления правдоподобного прогноза поведения системы в условиях, заданных исследователем.

2. Спектральный метод анализа данных

2.1 Теория спектрального анализа

Спектральный анализ – метод обработки сигнала, характеризующий частотный состав измеряемого сигнала. Математической основой этого метода является преобразование Фурье, который соединяет временной или пространственный сигнал с его представлением в частотной области [8].

Обработка сигналов включает в себя задачи аудио и речевого анализа, сигналы в которых возникает проблема управления потоком данных в современных вычислительных системах.

Во время обработки сигналов, приходится решать задачи двух видов – обнаружения и оценивания. Задача измерения значений параметров называется задачей оценивание. Очень часто сигнал может быть зашумлен, на него могут накладываться другие сигналы. Поэтому сигнал обычно разлагают по базисным составляющим пространства сигналов.

Наибольший интерес для многих приложений представляют периодические сигналы. Вполне естественно, что используются синус и косинус, такое разложение можно выполнить с помощью классического преобразования Фурье.

Спектральная оценка, полученная из конечной записи данных, характеризует некоторое предположение об истинной функции, которая может быть получена, если запись данных бесконечной длины. Поэтому поведение и характеристики спектральных оценок следует описывать статистически. Общими статистическими критериями качества оценки являются ее смещение и дисперсия.

Области комбинированных графиков спектральной оценки, где экспериментально определено большое значение дисперсии, указывают на то, что спектральные признаки, видимые в спектре одной реализации, не могут считаться статистически значимыми.

С другой стороны, особенности совмещенных спектров в тех областях, где эта дисперсия мала, может быть соотнесено с реальными частотными составляющими анализируемого сигнала с высокой надежностью.

В случае коротких записей данных часто не удается получить несколько спектральных оценок, и статистический анализ отдельных спектральных оценок, полученных по коротким записям данных, в общем случае представляет собой не простую задачу.

Спектром совокупности данных $y(x)$ называют некоторую функцию другой координаты (или координат, если речь идет о многомерном спектре) $F(\omega)$, полученную в соответствии с определенным алгоритмом. Примерами спектров являются преобразование Фурье, спектр мощности, вейвлет-преобразование.

Рассмотрим основные понятия Фурье-спектра. Преобразование Фурье состоит в представлении сигнала $y(x)$ в виде бесконечной суммы синусоид вида $F(\omega) \cdot \sin(\omega x)$. Функция $F(\omega)$ называется преобразованием Фурье, или интегралом Фурье, или Фурье-спектром сигнала. Ее аргумент ω имеет смысл частоты соответствующей составляющей сигнала. Обратное преобразование Фурье переводит спектр $F(\omega)$ в исходный сигнал $y(x)$.

Согласно определению,

$$F(\omega) = \int_{-\infty}^{\infty} y(x) * \exp(-i\omega x) dx \quad (1)$$

Из определения видно, преобразование Фурье является комплексной величиной, даже если сигнал действительный.

Преобразование Фурье имеет большое значение для различных математических приложений, и для него разработан эффективный алгоритм, называемый алгоритмом БПФ (быстрое преобразование Фурье). Он популярен, благодаря своей ультра-экономичности и практически во всех математических пакетах организован в виде подпрограммы.

Алгоритм БПФ имеет сильное ограничение, которое на практике не является критичным. Дело в том, что аргумент прямого Фурье-преобразования, т.е. объем выборки $y(x_i)$, должен иметь ровно 2^n элементов (n – любое целое число). Соответственно, результатом работы алгоритма БПФ является вектор с $1+2n-1$ элементами. Если число данных не совпадает со степенью 2, то для запуска алгоритма БПФ достаточно дополнить недостающие элементы нулями.

Рассмотрим наиболее типичную для физического эксперимента ситуацию расчета Фурье-спектра действительного сигнала [8]. Чтобы смысл преобразования Фурье был более понятен, используем в качестве модельных данных дискретизацию следующего детерминированного сигнала (рис. 2.1-2.2):

$$y(x) = 1 \cdot \sin(2\pi \cdot 0.05 \cdot x) + 0.5 \cdot \sin(\pi \cdot 0.1 \cdot x) + 0.1 \cdot \sin(2\pi \cdot 0.5 \cdot x). \quad (2)$$

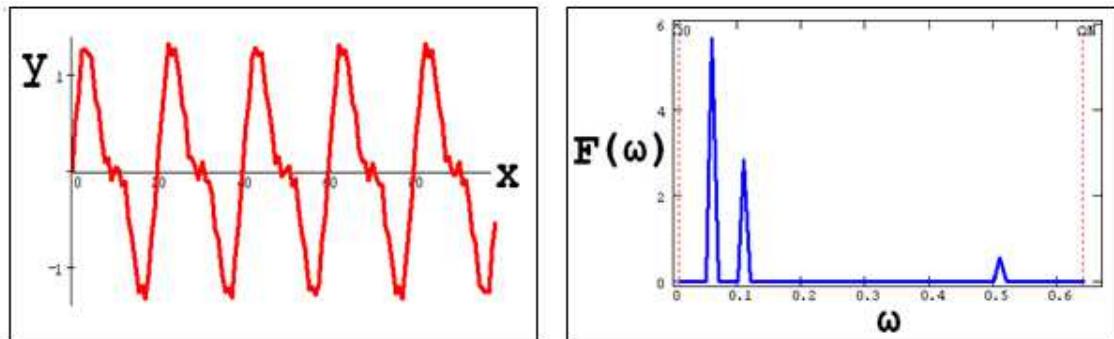


Рис. – 2.1 Модельные данные

Рис. – 2.2 Выборочный Фурье-спектр

На рисунке 2.1 представлены результаты работы алгоритма БПФ в виде модуля Фурье-спектра $|F(\omega)|$, поскольку, повторимся, сам спектр является комплексным. Полезно сравнить полученные амплитуды и местоположение пиков спектра на рисунке 2.2 с определением синусоид в формуле (2).

Показательно, что, если подвергнуть полученное абсолютное значение Фурье-спектра (рис. 2.2) обратному преобразованию Фурье, возможность которого также предоставляет алгоритм БПФ, то профиль исходного сигнала будет реконструирован правильно, но окажется сдвинутым на определенное расстояние вдоль оси x (рис. 2.3).

Это происходит из-за того, что для абсолютной величины комплексного спектра отсутствует информация об относительной фазе отсчетов данных. В остальном, сигнал $y(x)$ восстановлен с большой точностью, что характерно для плавного изменения сигнала. Если в качестве входных данных обратного преобразования Фурье использовать комплексный Фурье-спектр, то совпадение будет полным.

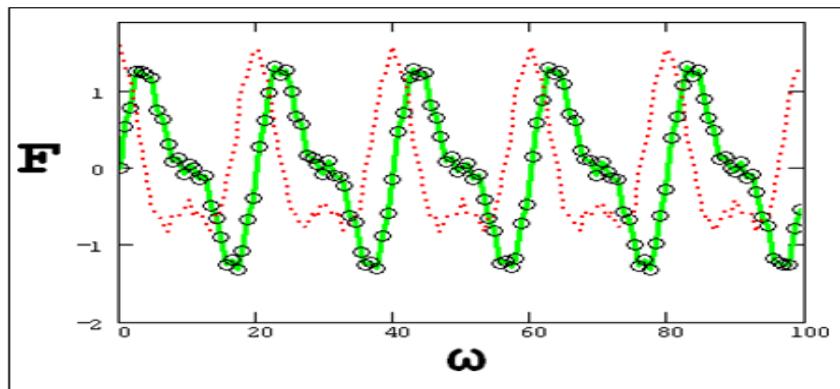


Рис. 2.3 – Обратное преобразование Фурье комплексного (кружки) и действительного (точки) Фурье-спектра

2.2 Геодинамический мониторинг сейсмических сигналов

В течение 2007-2014 гг. на геодинамических полигонах на п-ове Камчатка проводился мониторинг НДС геологической среды комплексом геолого-геофизических методов – сейсмический мониторинг, регистрация естественного импульсного электромагнитного поля [9].

Важным критерием при выборе территории для проведения наблюдений является предшествующая сейсмическая обстановка в ее пределах, а также сведения о высокой геодинамической активности исследуемой площади.

В таблице 2.1 представлена характеристика основных параметров геодинамических полигонов МПР РФ в Сибири и на Дальнем востоке. Также на рисунке 2.4 приведена схема геодинамических полигонов расположенных в на юге Камчатки.

Таблица 2.1 – Состав мониторинговых сетей

Округ РФ	Пункты сейсмического мониторинга (шт.)
Дальневосточный ФО	21
Сибирский ФО	11

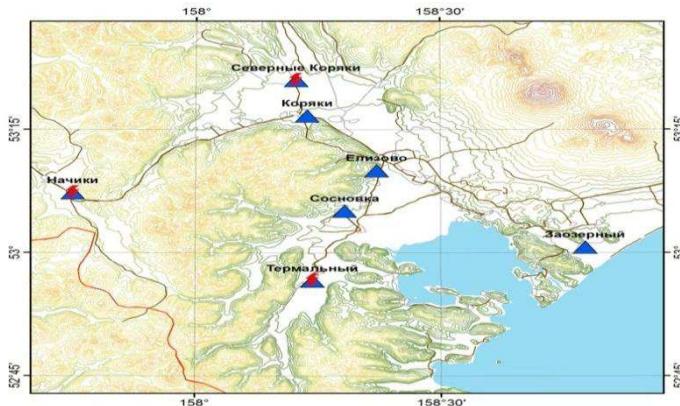


Рис. 2.4 – Карта-схема наблюдательной сети геофизического мониторинга на Петропавловск-Шипунском полигоне

Научно-методические основы и технологические решения, используемые при мониторинге НДС геологической среды на геодинамических полигонах прошли многолетнюю проверку. Сейсмические сети оснащены трехкомпонентными сейсмоприемниками с одной вертикальной и двумя горизонтальными компонентами.

Компоненты сейсмоприемника ориентированы по сторонам света. Ориентация сейсмоприемников следующая: горизонтальная компонента X – на север, горизонтальная компонента Y - на восток, вертикальная компонента Z – вверх.

Важными параметрами сейсмодатчиков являются: преобразуемый параметр движения основания: смещение, скорость или ускорение; чувствительность датчика и динамический диапазон; рабочий диапазон периодов датчика.

Для геодинамического мониторинга применяются инерционные сейсмометры, которые измеряют движение основания земли относительно инерционной точки подвижной массы. Они более просты в использовании и более чувствительны к сигналам, возбуждаемым землетрясениями.

При проведении сейсмического мониторинга рекомендуется использовать короткопериодные сейсмометры, они измеряют сигналы приблизительно в диапазоне 0,1-50 Гц с собственной частотой 0,2-1 Гц. и имеют стабильные характеристики для интервала частот, характерных для землетрясения.

Этот тип датчика способен покрыть динамический диапазон главных сейсмических событий, от регистрации самого низкого уровня сейсмического шума до разрушительных сейсмических событий [10].

2.3 Анализ сейсмических данных геомониторинга

Частотная характеристика сейсмической системы определяется относительной амплитудой и фазой, на которые эта система реагирует при движении грунта в зависимости от частоты его колебаний. Определение амплитудно-частотной характеристики системы необходимо в основном для оценки магнитуды сейсмического мониторинга или проведения сейсмического мониторинга с использованием динамических характеристик сейсмических волн [11].

Требования к амплитудно-частотной и фазово-частотной характеристикам, а также к контролю других аппаратурных параметров определяются соответствующими ведомственными инструкциями.

Рекомендуемый диапазон рабочих частот для регистрации объемных волн составляет 0.5-10 Гц. Такой диапазон должен обеспечить регистрацию, как высокочастотных местных, так и низкочастотных далеких землетрясений.

Для решения задач, связанных с регистрацией местных землетрясений, больше подходит такая амплитудно-частотная характеристика всего

сейсмического тракта, которая обеспечивает во всем диапазоне рабочих частот пропорциональность амплитуды записи величине скорости смещения грунта; для задач, связанных с регистрацией далеких землетрясений для изучения глубинного строения геологической среды, больше пригодна аппаратура с амплитудно-частотной характеристикой пропорциональной смещению грунта.

Задача: Изучение характера спектра исходного сигнала, а также, выделение особенностей на различных частотах с последующей интерпретацией.

Данные с сейсмоприемника, регистрирующего результаты (6 дней), всего 2154000 строк. Размер файла 42 мегабайт. Обрабатываются три компоненты: Z, X, Y (табл. 2.3)

Таблица 2.3 – Фрагмент данных с сейсмодатчика

Время	Z	X	Y
2153994	-51157	-6529	36327
2153995	-50760	-7805	36550
2153996	-50744	-5443	36510
2153997	50874	-5159	36349

Построение графиков спектральной плотности реализуется в специальном программном обеспечении написанной на языке программирования R (см. Приложение А).

Алгоритм работы программы:

1. Загрузка данных.
2. Определяется нижняя и верхняя граница для вывода спектра.
3. Выделяется первый столбец.
4. Вычисляется спектр.
5. Создается png файл.
6. В ней отрисовывается результат расчета спектра.

Этот процесс повторяется ровно три раза, с учетом одного изменения. Выбирается следующий столбец в пункте 3.

Процесс построения трех графиков в этом случае составляет 12 минут, выводятся частоты с 0,2 до 10 Гц.

На рисунках 2.5-2.7 показаны результаты построение спектра трех компонент: Z, Y, X.

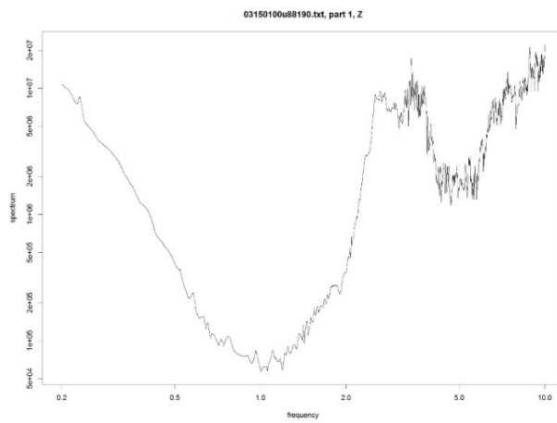


Рис. 2.5 – Результат построения спектра для координаты Z

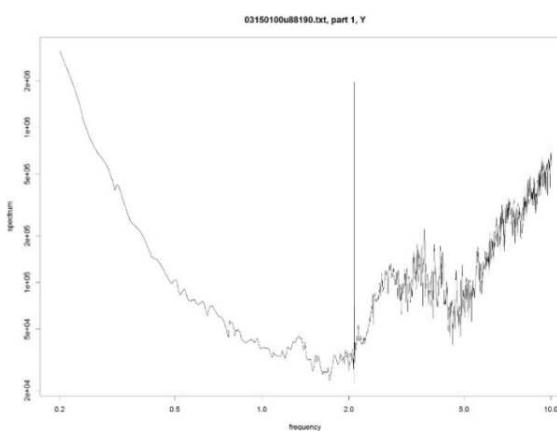


Рис. 2.6 – Результат построения спектра для координаты Y

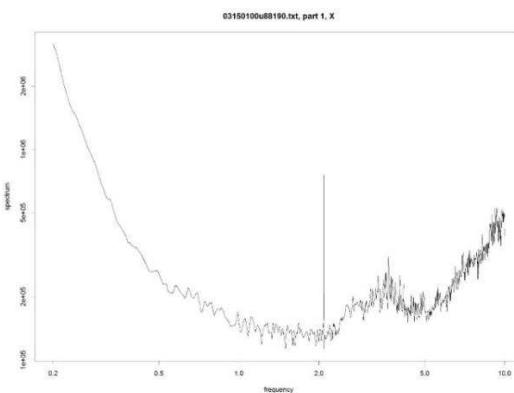


Рис. 2.7 – Результат построения спектра третьего датчика X

2.4 Интерпретация десятичного логарифма

Также решается задача интерпретации для сигналов и построение спектральной плотности с применением логарифмических функций (десятичного логарифма). Написана программа на языке программирования Python (см. Приложение А), которая находит логарифмы значений, работает и с отрицательными числами (табл. 2.4).

Таблица 2.4 – Фрагмент логарифмированных данных

Время (строка)	Z	X	Y
2153994	5.9388	5.3236	5.4092
2153995	5.939	5.321	5.4096
2153996	5.939	5.3258	5.4095
2153997	5.9389	5.3264	5.4092

Далее строим графики спектральной плотности. Процесс построение занял примерно в 5-6 раз больше времени, чем со стандартными значениями. На рисунках 2.8-2.10 показаны результаты построение спектра логарифмированных данных.

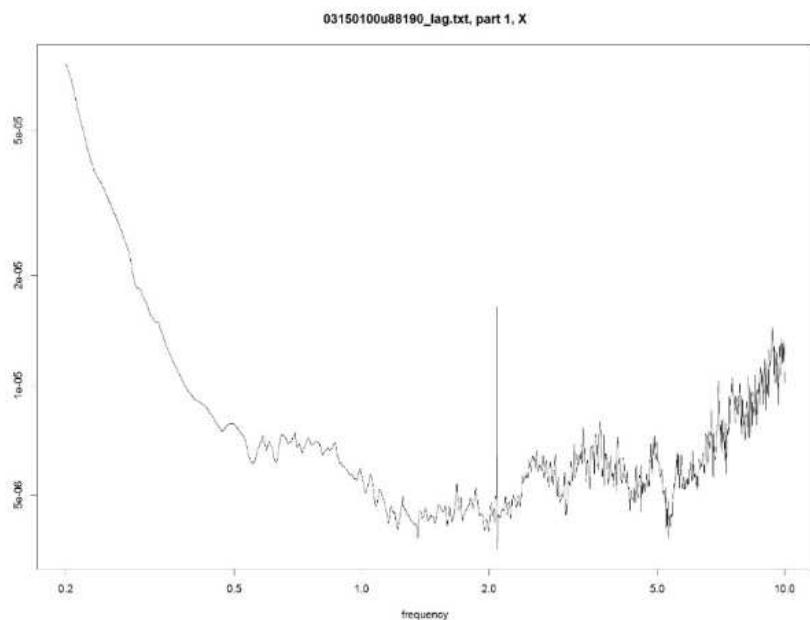


Рис. 2.8 – Результат построения спектра для координаты X

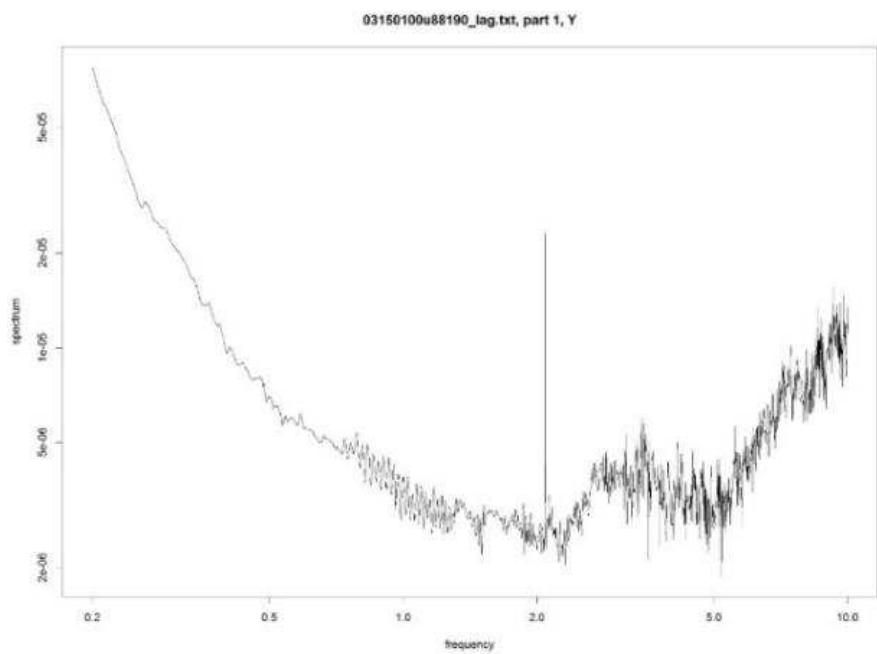


Рис. 2.9 – Результат построения спектра для координаты Y

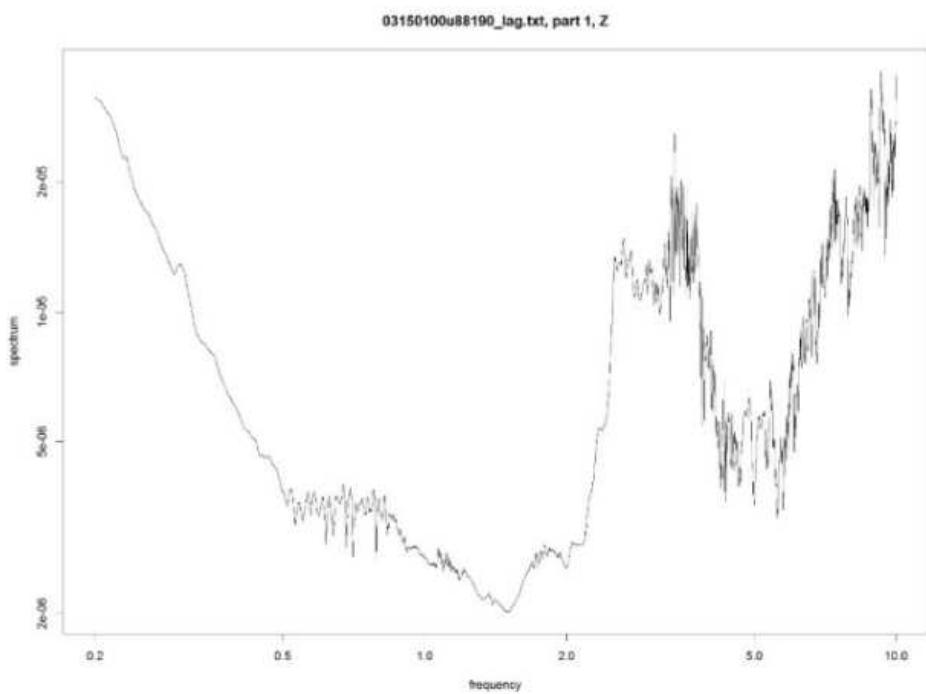


Рис. 2.10 – Результаты построения спектра для координаты Z

2.5 Сравнение и анализ результатов расчетов

На рисунках 2.10-2.12 приведено сравнение спектров стандартных значений с новыми логарифмированными значениями (отмечены красным цветом).

Сравнение стандартных и логарифмированных значений (датчик Z)

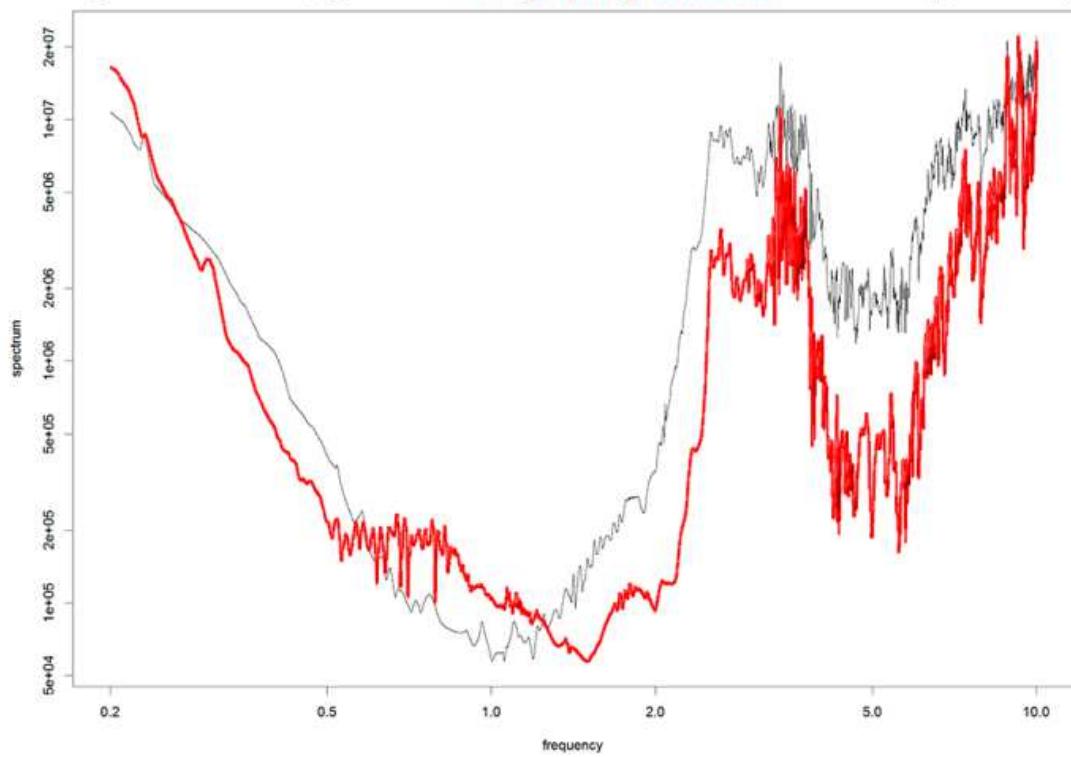


Рис. 2.10 – Сравнение компоненты Z

Сравнение стандартных и логарифмированных значений (датчик Y)

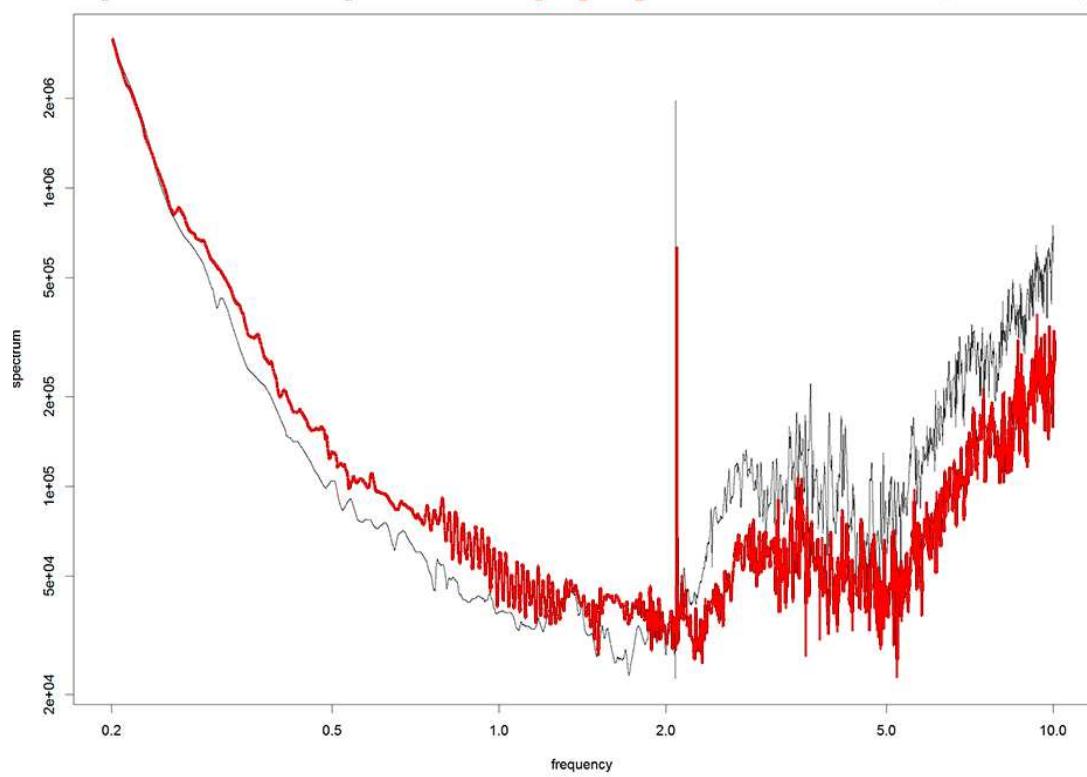


Рис. 2.11 – Сравнение компоненты Y

Сравнение стандартных и логарифмированных значений (датчик X)

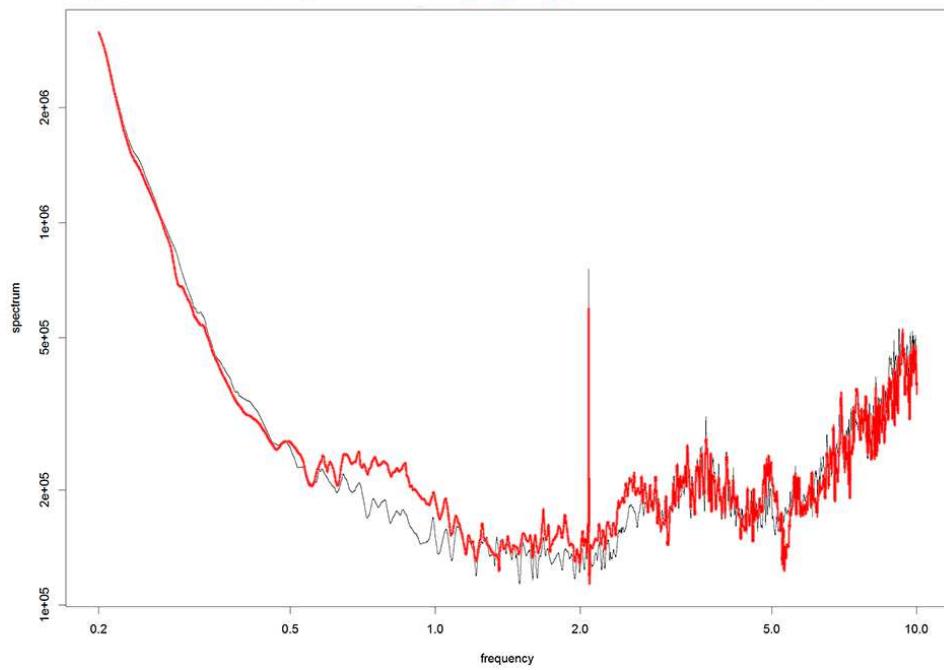


Рис. 2.12 – Сравнение компоненты X

Спектр одномерного сигнала представляет собой поверхность в трехмерном пространстве. Способы визуализации спектра могут быть самыми различными. В работе использован способ логарифмирование данных, что позволяет проследить изменения амплитуд на разных масштабах во времени, а также выявить картину локальных экстремумов этих поверхностей («холмов» и «впадин»), так называемый «скелет» структуры анализируемого процесса [12].

ЗАКЛЮЧЕНИЕ

В работе рассмотрено несколько методов визуализации многомерных данных с пропусками («плохие» данные). Приведены основные виды классификации методов визуализации, их сильные и слабые стороны, области применимости.

Рассмотрена задача построения упругой карты для произвольного набора данных. На основе изученного метода построения упругих карт приведен пример обработки большого массива данных афтершоков сильного землетрясения, произошедшего 8 сентября 2017 г. у берегов Мексики.

Процесс обработки указанных данных представляет собой суть информационного моделирования данных. Представлены результаты моделирования, позволяющие с разных сторон посмотреть на данные и выявить основные их особенности.

Представленный метод спектрального анализа применим ко многим классам сигналов и шумов в предположении о стационарности. Вычислительная эффективность метода основана на использовании алгоритма Быстрого Преобразования Фурье.

Изучен метод спектрального анализа, а также возможности применения этого метода в современных вычислительных системах для обработки сейсмических данных геомониторинга в реальном масштабе времени.

Выполнены расчеты данных наблюдений с сейсмоприемников, расположенных на Камчатском полигоне.

Сформулирован ряд требований по вычислительным ресурсам при реальной обработке, сделан анализ длины выборки данных при различном представлении входного сигнала. Получены результаты экспериментальных исследований по оценке спектральной плотности мощности сигнала.

Список используемых источников

1. Арефьев С.С. Эпицентральные сейсмологические исследования. – М.: ИКЦ «Академкнига», 2003. – 375 с.
2. Бендат Д., Пирсол А. Прикладной анализ случайных данных. – М.: Мир, 1989. – 540 с.
3. Лбов Е.С. Методы обработки разнотипных экспериментальных данных. Новосибирск, 1981. – 156 с.
4. Литтл Р., Рубин Д. Статистический анализ данных с пропусками. – М.: Финансы и статистика, 1991. – 336 с
5. Зиновьев А.Ю. Визуализация многомерных данных. – Красноярск: ИПЦ КГТУ, 2000. – 168 с.
6. Загоруйко Н.Г. Методы обнаружения закономерностей. – М.: Наука, 1981.
7. Бат М. Спектральный анализ в геофизике. – М.: Недра, 1980.
8. Гольдин С.В. Линейные преобразования сейсмических сигналов. – М.: Недра, 1974. – 350 с.
9. Лыгин, А. М. Мониторинг напряжённо-деформированного состояния геологической среды в Сибирском и Дальневосточном федеральных округах в 2007-2014 годах: монография, изд-во «Город» / А. М.
10. Сибгатулин, В. Г. Оценка сейсмической опасности юга Центральной Сибири: В. Г. Сибгатулин, К. В. Симонов, С. А. Перетокин. – Красноярск: КНИИГиМС, 2004. – 190 с.
11. Марпл.-мл. С.А. Цифровой спектральный анализ и его приложения. – М.: Мир, 1990. – 584 с.
12. Отнес Р., Эноксон Л. Прикладной анализ временных рядов: основные методы. – М.: Мир, 1982. – 428 с.

Приложение А

Вычисления десятичного логарифма сигнала

Программа, написанная на языке программирования Python для вычисления логарифма значений сигналов (работает и с отрицательными числами).

```
import math
import sys
#импортируем необходимые библиотеки
file_name_full = sys.argv[1]
print(file_name_full)
#вводим название файла
try:
    file_name, extension = file_name_full.split(".")
except:
    file_name = file_name_full
    extension = ""
#программа находит файл с таким названием
def find_min_column(mass,column):
    mass_tmp = []
    for line in mass:
        mass_tmp.append(line[column])
    return min(mass_tmp)
#процесс поиска минимального значения столбца
def all_sum_column(column,value):
    global mass_full
    for line in range(len(mass_full)):
        mass_full[line][column]+=abs(value)+1.0
        mass_full[line][column]=round(mass_full[line][column],3)
#процесс суммирование если имеется отрицательное число.
def print_mass():
    for i in mass_full:
        print(i)

def mass_log():
    global mass_full
    for line in range(len(mass_full)):
        for column in range(len(mass_full[line])):
            mass_full[line][column]=round(math.log10(mass_full[line][column]),4)
    mass_full = []
#берется значение, находится логарифм и возвращается значение на место.
with open(file_name_full,"r") as file:
    for line in file.readlines():
        mass_full.append(list(map(float,line.strip().split("\t"))))

print("[+] Read file")
#сообщение о том что файл загружен
for i in range(3):
    res = find_min_column(mass_full,i)
    if(res<0):
        all_sum_column(i,res)
```

```

#программа проходит по каждому столбцу, находит самое минимальное значение и если оно
меньше нуля запускается процесс суммирование минимального значения ко всему столбцу
mass_log()
print("[+] Logarithm")
#считается логарифм

with open(file_name+"_out."+extension,"w") as file:
    for line in mass_full:
        file.write("\t".join(map(str,line))+"\n")
#сохранение результатов
print("[+] Finished")

```

Вычисления спектральной плотности сигнала

Программа на языке R для построения графиков спектральной плотности мощности по данным сейсмодатчиков. Используется библиотека psd (Adaptive, sine multitaper power spectral density estimation). Обрабатываются три датчика: Z, X, Y. Каждый файл разбивается на четыре равные части, которые соответствуют (примерно) границам часа. Выводятся частоты с 0,2 до 10 Гц.

```

# Используется библиотека psd (Adaptive, sine multitaper power spectral density estimation)
# Обрабатываются три датчика: Z, X, Y.
# Каждый файл разбивается на четыре равные части, которые соответствуют (примерно)
# границам часа.
# Выводятся частоты с 0,2 до 10 Гц.

```

```

# Подключаем библиотеку psd (она должна быть предварительно установлена).
library(psd)

```

```

# Задаем частоту дискретизации.
fr <- 100

```

```

# Получаем список файлов с расширением txt в рабочей директории.
files <- list.files(pattern=".*txt$")

```

```

# Для каждого файла из списка
for (k in 1:length(files)) {
    # читаем его содержимое,
    dat <- read.table(files[k])
    # определяем длину каждой части (для простоты считаем их равными),
    len <- length(dat[,3]) %% 1
    # определяем нижнюю и верхнюю границу для вывода спектра.
    tmin <- floor(0.2*len/fr) + 1
    tmax <- ceiling(10*len/fr) + 1

```

```

# Для каждой из четырех частей.
for (part in 1:1) {

```

```

# Выделяем первый столбец,
Z <- dat[1:len + (part-1)*len ,1]

```

```

# вычисляем спектр,
spZ <- pspectrum(Z,x.frqsamp=fr)
# создаем новый файл png размером 2400 на 1800,
png(file=paste("png",files[k],"-",part,"-psd-Z.png",sep=""), width = 2400, height = 1800,
pointsize=30)
# отрисовываем в него результаты расчета спектра,
plot(spZ$freq[tmin:tmax],spZ$spec[tmin:tmax],main=paste(files[k],"      part      ",",",
Z",sep=""),type="l",cex=1.5 ,log="xy",xlab="frequency",ylab="spectrum")
# закрываем файл png.
dev.off()

# Выделяем второй столбец,
X <- dat[1:len + (part-1)*len ,2]
# вычисляем спектр,
spX <- pspectrum(X,x.frqsamp=fr)
# создаем новый файл png размером 2400 на 1800,
png(file=paste("png",files[k],"-",part,"-psd-X.png",sep=""), width = 2400, height = 1800,
pointsize=30)
# отрисовываем в него результаты расчета спектра,
plot(spX$freq[tmin:tmax],spX$spec[tmin:tmax],main=paste(files[k],"      part      ",",",
X",sep=""),type="l", cex=1.5 ,log="xy",xlab="frequency",ylab="spectrum")
# закрываем файл png.
dev.off()

# Выделяем третий столбец,
Y <- dat[1:len + (part-1)*len ,3]
# вычисляем спектр,
spY <- pspectrum(Y,x.frqsamp=fr)
# создаем новый файл png размером 2400 на 1800,
png(file=paste("png",files[k],"-",part,"-psd-Y.png",sep=""), width = 2400, height = 1800,
pointsize=30)
# отрисовываем в него результаты расчета спектра,
plot(spY$freq[tmin:tmax],spY$spec[tmin:tmax],main=paste(files[k],"      part      ",",",
Y",sep=""),type="l",cex=1.5 ,log="xy",xlab="frequency",ylab="spectrum")
# закрываем файл png.
dev.off()

}

}

```

Приложение Б

Метод упругих карт применен для визуализации данных о катастрофических землетрясениях. На рисунках Б 1- Б 2 приведены результаты визуализации данных каталогов этих землетрясений.

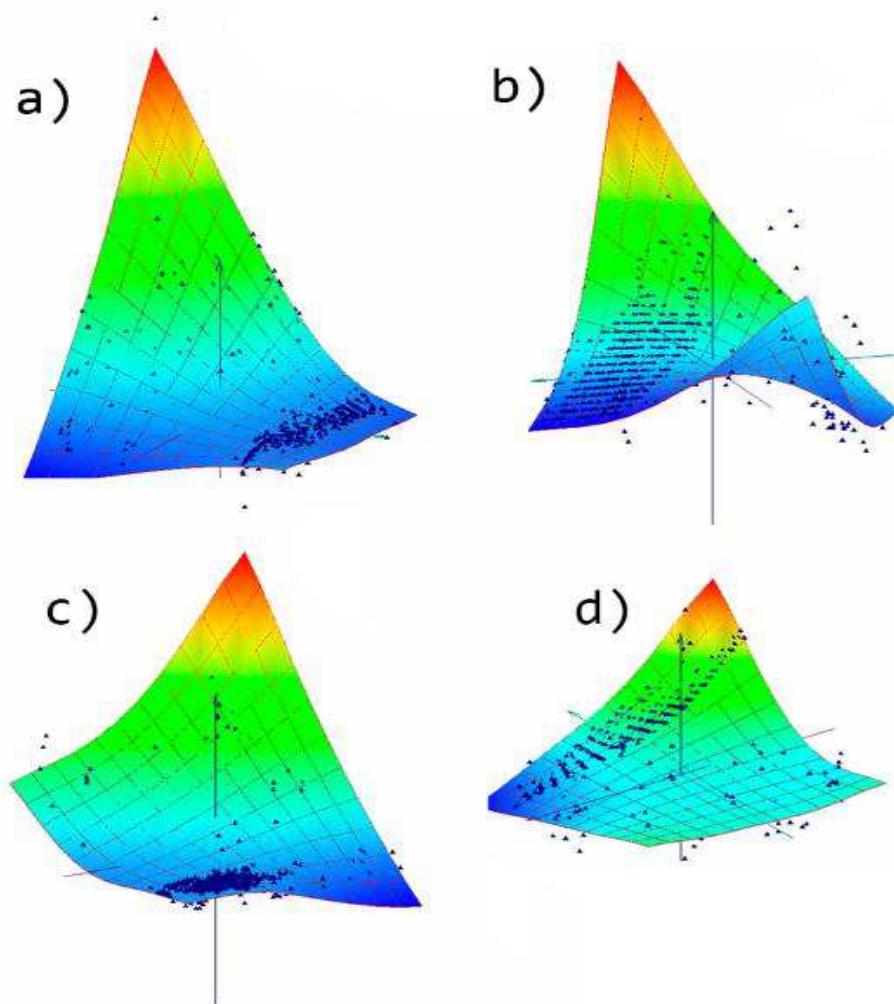


Рис. Б 1 – Результаты визуализации методом упругих карт, афтершоков (последующие 10 дней) Чилийского землетрясения произошедшего 27 февраля 2010 г.: а) зависимость от времени, магнитуды и цветовое кодирование параметра глубины; б) зависимость от времени, глубины и цветовое наложение параметра магнитуды толчка; в) зависимость от координат и цветовое кодирование параметра магнитуды; г) зависимость от координат толчка и цветовое кодирование параметра глубины

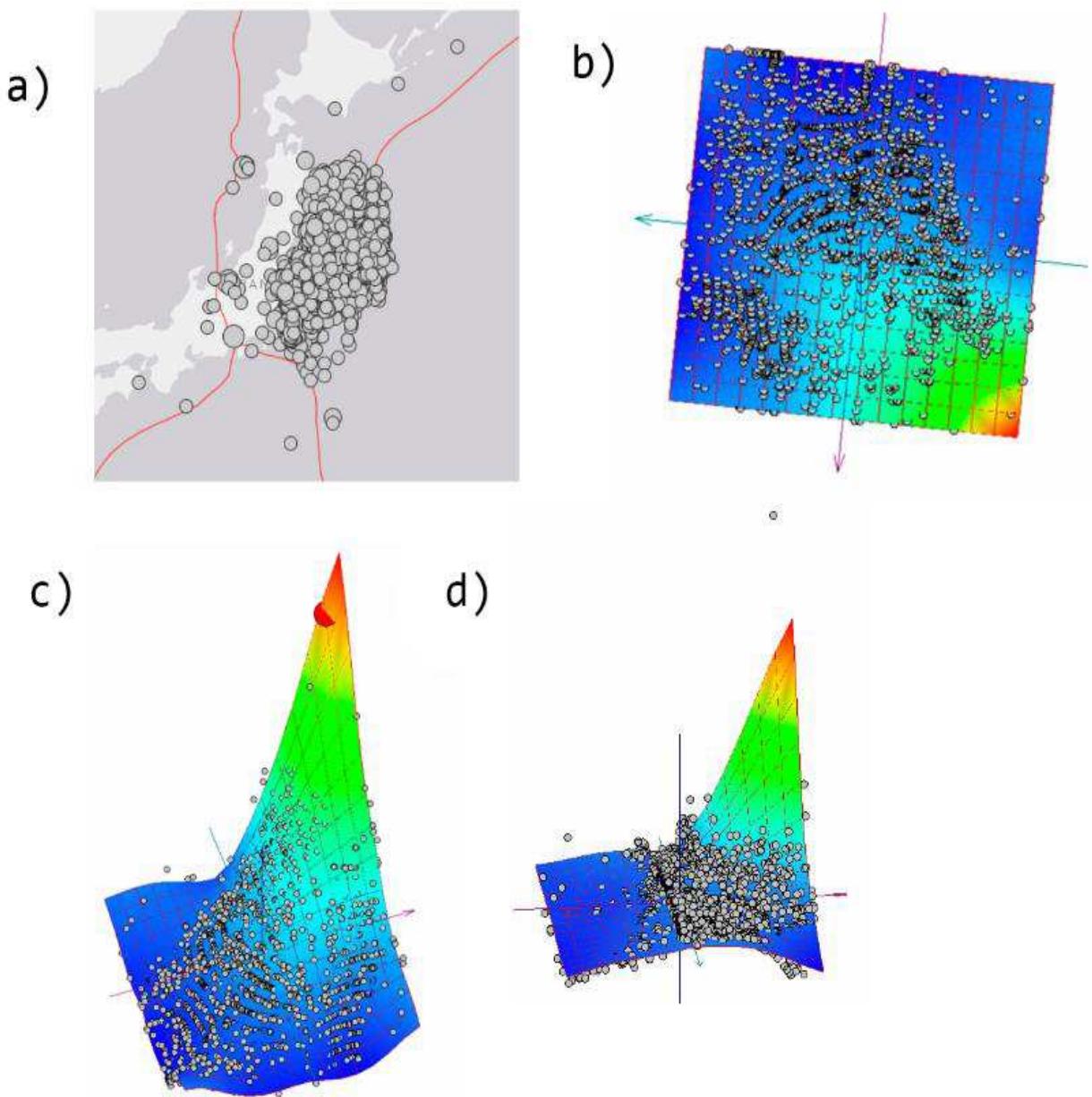


Рис. Б 2 – Результаты визуализации методом упругих карт, афтершоков (последующие 10 дней) землетрясения произошедшего 11 марта 2011 г. у берегов Японии: а) географическая карта землетрясений в день самого мощного сейсмического события; б) 2D модель смоделированной географической карты, зависимость от координат; в) зависимость от координат и цветовое кодирование параметра магнитуды; г) зависимость от координат цветовое кодирование параметра глубины

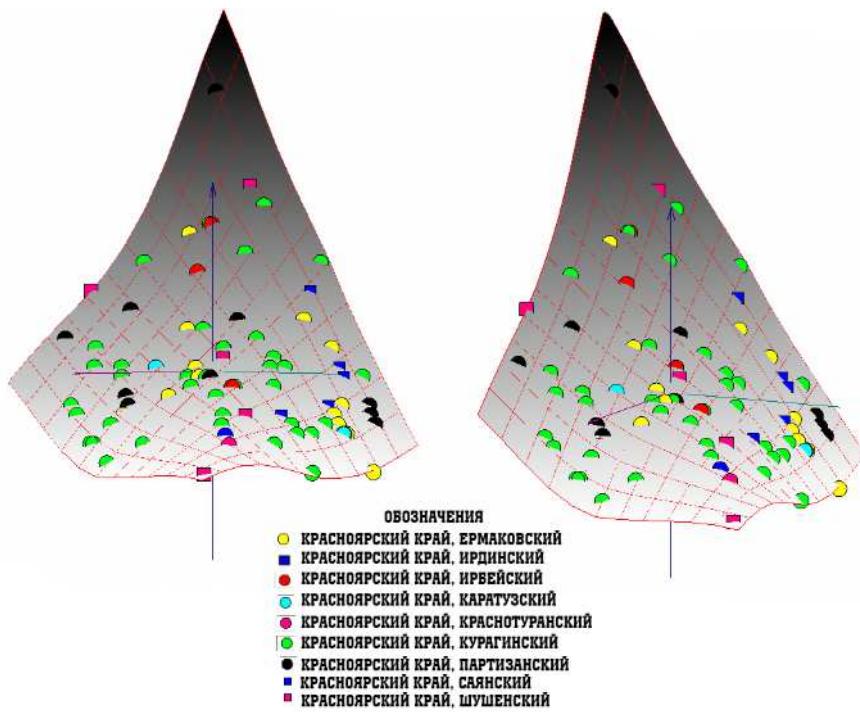


Рис. Б 3 – Результаты визуализации методом упругих карт каталогов землетрясений, произошедших в Красноярском крае за период с 2015 по 2016 гг. (зависимость от времени, магнитуды и параметра h)

Визуализирован также большой объем данных спектрального анализа первичных данных сейсмического мониторинга (за 18 дней) (рис. Б4).

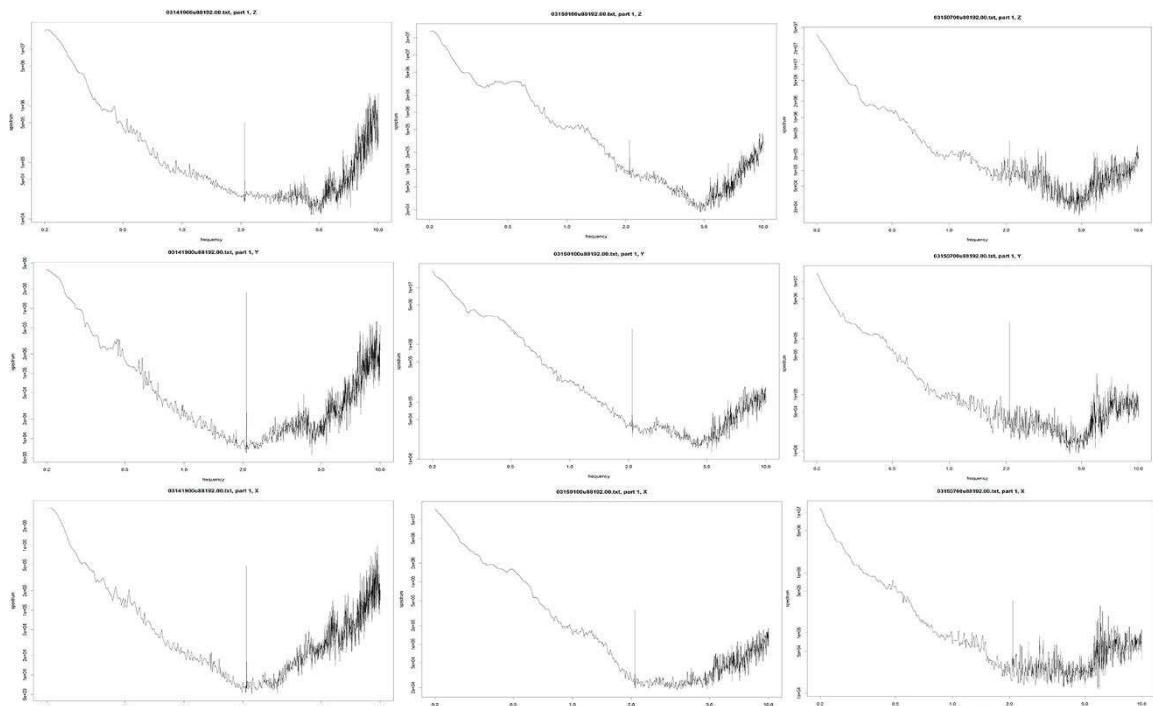


Рис. Б 4 – Результаты построения спектров

На рисунке Б 5 приведены результаты построения спектров для логарифмированных значений.

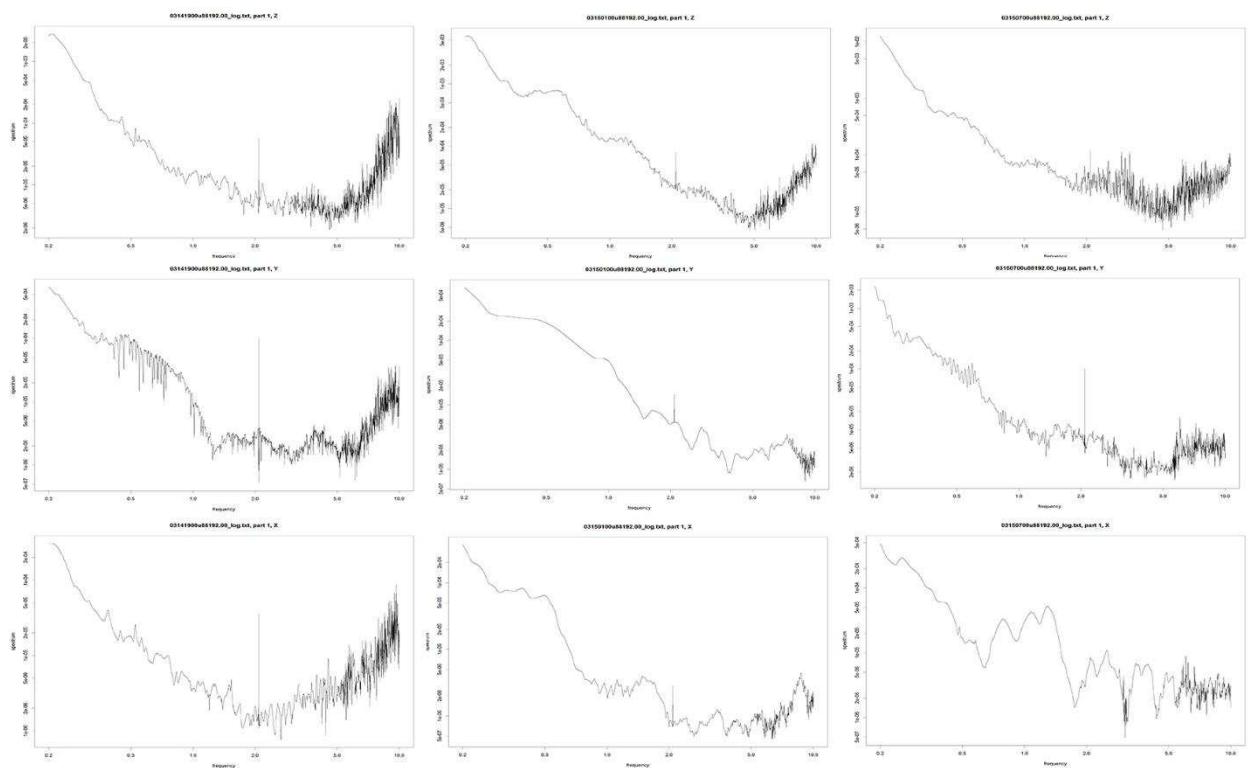
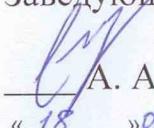


Рис. Б 5 – Результаты построения спектров логарифмированных значений

Федеральное государственное автономное
образовательное учреждение
высшего образования
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»
Институт космических и информационных технологий
Кафедра прикладной математики и компьютерной безопасности

УТВЕРЖДАЮ

Заведующий кафедрой

А. А. Кытманов
«18 »06 2018 г.

БАКАЛАВРСКАЯ РАБОТА

01.03.04 – Прикладная математика

Визуализация данных геомониторинга для решения прогностических
задач

Руководитель  профессор, д.т.н. К.В. Симонов

14.08.2018

А.Э.о Исмаилов


11.06.2018

Выпускник

Красноярск 2018