

ПРИМЕНЕНИЕ МЕТОДОВ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ ДЛЯ ИССЛЕДОВАНИЯ СТАНДАРТНЫХ ОТКЛОНЕНИЙ ОТСЧЕТОВ ГРАВИМЕТРА AUTOGRAV

И.В. Геник

Горный институт УрО РАН, г. Пермь

Аннотация: В статье излагаются результаты, продолжающие исследования 2023 года, связанные с обработкой долговременных записей гравитационного поля, выполненных гравиметрами Autograv. Ранее был создан первый вариант программы «Анализ временных рядов», включающий процедуры ввода и контроля данных и реализующий ряд алгоритмов обработки. Программа была реализована на языке Python с использованием его математических библиотек. Следующий этап в 2024 году связан с созданием и тестированием расширенных процедур обработки и углубленным анализом данных: регрессионный анализ, сглаживание (скользящее среднее с равными и не равными весами, фильтр Савицкого-Голея, фильтр Винера), статистические оценки вида распределения, оценка спектральной плотности мощности и др.

Рассмотренные подходы можно применять не только к временным рядам гравиразведочных данных, но и к пространственным гравиметрическим данным (как к отдельным профилям, так и целым площадям). Широко известным примером использования статистических закономерностей при интерпретации гравиметрических данных является программный комплекс КОСКАД 3D. Другой пример использования статистики связан с анализом результатов анализа региональных геофизических работ (сейсморазведка, гравиразведка, магниторазведка) на нефть и газ и на этой основе прогнозирования возможных результатов на других площадях.

В целом применение различных методов статистической обработки гравиметрических данных создает большой набор признаков для применения методов искусственного интеллекта, что открывает возможности для построения баз знаний, разработки экспертных систем, выполнения машинного обучения с последующей целью в виде создания систем автоматической интерпретации гравиразведочных работ без участия человека.

Ключевые слова: геофизика, гравиразведка, гравиметр, статистические методы, временные ряды, интерпретация, программное обеспечение.

Введение

В статье описывается продолжение исследований, связанных с обработкой и анализом долговременных записей гравиметра Autograv [2] и созданием программы «Анализ временных рядов». В предыдущей работе были рассмотрены: научные направления, которые связаны с анализом долговременных записей [1]; характеристики используемых гравиметров; созданный первый вариант программы анализа временных рядов, реализовавший предварительный анализ временных рядов, включая описательную статистику, исследование вида распределения данных, спектральный анализ, вычисление автокорреляционных функций с построением доверительных интервалов для них.

В данной работе представлены новые результаты, связанные с реализацией (и исследованием эффективности для обработки долговременных записей гравиметра) на языке Python следующих подходов к анализу временных рядов [6-8]: регрессионный анализ, различные методы сглаживания, статистические тестирования вида распределения, оценки спектральной плотности мощности и др.

Подходы к анализу долговременных записей гравиметра, представленные в этой и предыдущей статье, применимы не только к временным, но и к пространственным зависимостям гравитационного поля (гравиметрическим профилям), предоставляя широкий набор критериев для сравнения данных и последующей их классификации. В расширенном и модифицированном виде подходы применимы не только для профильных, но и для площадных гравиметрических работ, позволяя создавать научно обоснованные алгоритмы прогноза результатов гравиметрических работ [3]. Такого рода алгоритмы прогноза могут позволить дать обоснование необходимости гравиразведочных иссле-

дований, их информативности и геологической содержательности, что, в конечном итоге, даст возможность расширить применение гравиразведки.

Методы и подходы

Исходные данные – записи гравитационного поля гравиметром CG-5 Autograv с 18 февраля по 3 апреля 2017 года, всего 63382 отсчета (рис. 1). Анализировались стандартные отклонения отсчетов гравиметра, изменявшиеся от 0.007 до 1.782 мГал, среднее значение равно 0.035 мГал, медиана – 0.031 мГал; отношение максимального и среднего значений около 51.

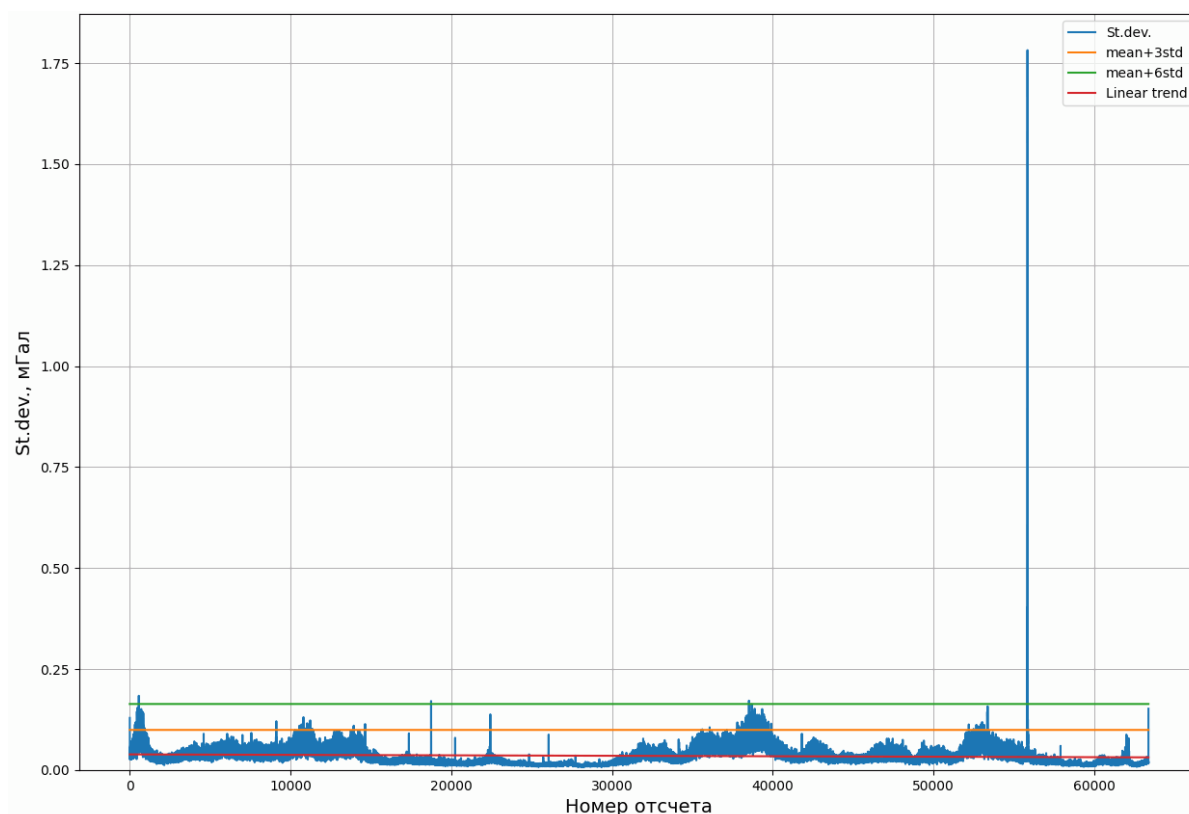


Рис. 1. Временной ряд стандартного отклонения отсчетов гравиметра CG-5 Autograv

В работе были использованы следующие методы и оценки: оценка вида распределения данных; построение регрессионных зависимостей и их ранжирование в зависимости от коэффициента детерминации; тестирование гетероскедастичности случайных ошибок линейной регрессионной модели; оценка разброса данных для сглаживающих фильтров разного вида; исследование стационарности временного ряда; построение периодограмм и оценка спектральной плотности мощности.

В целом представленные выше методы и оценки позволяют детально охарактеризовать временной ряд и получить его трансформанты, что является базой для дальнейших работ для детального сопоставления записей гравиметров, сделанных в разных локациях.

Результаты и обсуждение

Визуально разброс данных можно оценить при построении горизонтальных прямых на уровнях 3σ и 6σ , где σ – стандартное отклонение рассматриваемого набора. В диапазон $[\text{mean}-3\sigma; \text{mean}+3\sigma]$, где mean – среднее значение, в случае нормального распре-

деления попадают 99,73% значений, а в диапазон с 6σ – 99,99966% значений. Для данного временного ряда вне диапазона 3σ – 511 отсчетов ($\approx 0.81\%$), а вне диапазона 6σ – 29 отсчетов (0,045%), что указывает на то, что имеется отклонение от нормального вида распределения.

В математической статистике имеются специальные критерии (тесты) для проверки нормальности распределения. В работе был использован тест Шапиро-Уилка, в котором проверяется нулевая гипотеза о том, что выборка принадлежит нормальному распределению. В результате теста значения p -уровня значимости получилось меньше выбранного α -уровня, и нулевая гипотеза была отклонена. Коэффициент асимметрии для рассматриваемого временного ряда составляет более 12 (правый хвост распределения длиннее левого), а коэффициент эксцесса более 790 (распределение имеет острый пик). Такого рода распределения могут быть приведены к нормальному путем логарифмирования значений. Результат преобразования иллюстрируется рис. 2.

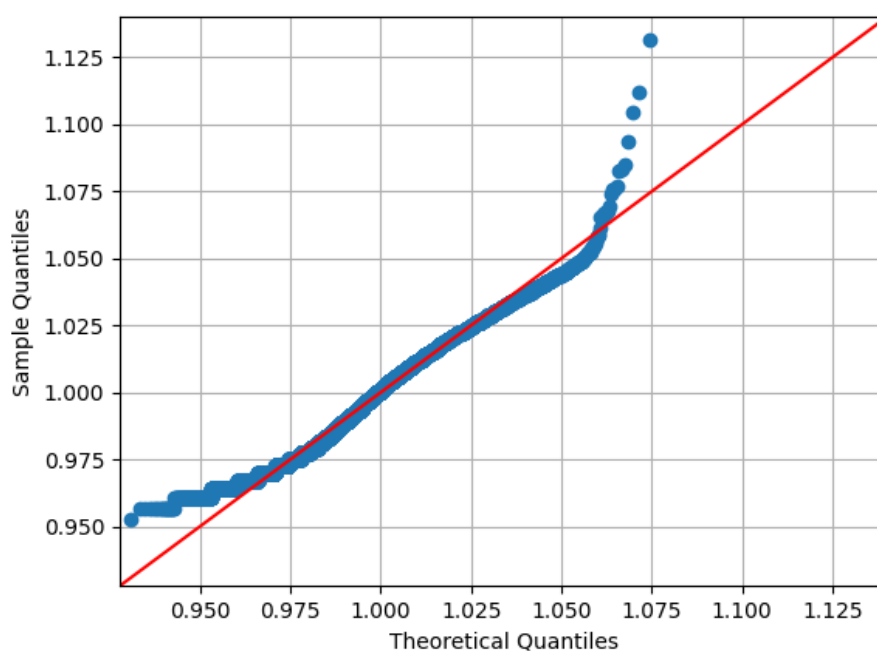


Рис. 2. Диаграмма квантиль-квантиль для набора гравитационных данных и логнормального распределения

На диаграмме представлено сравнение двух распределений вероятностей путем сопоставления их квантилей. В случае, если два сравниваемых распределения подобны, точки будут лежать на линии $Y = X$. В нашем случае основной массив синих точек лежит на этой линии, но наименьшие и наибольшие значения отклоняются от логнормального распределения (красная линия).

Сложный характер получаемого статистического распределения для временного ряда обусловил выполнение дополнительных исследований методом бутстрэп (bootstrap), основанным на многократной генерации выборок методом Монте-Карло на базе имеющегося набора данных. Параметры выполнения: полный временной ряд (63382 отсчета); программа на языке Python; персональный компьютер 2018 года выпуска. Число выборок при оценке среднего значения было 6000, 95% доверительный интервал получен равным 0.0350-0.0353 мГал (при среднем значении для всего набора данных равным 0.035 мГал).

Следующие исследования временного ряда выполнялись методами регрессионного анализа. Были рассмотрены несколько видов регрессионных зависимостей, построенных мето-

дом наименьших квадратов (МНК), но все они характеризовались низкими коэффициентами детерминации R^2 . Например, для линейной зависимости $F(X) = -1.154 \cdot 10^{-7}X + 0.039$ был получен $R^2 \approx 0.01$, при этом прямая имеет очень малый наклон к оси X и можно сказать, что прямая практически параллельна оси X .

Для регрессионных моделей были выполнены тестирование гетероскедастичности случайных ошибок линейной регрессионной модели (Тест Уайта), т.е. исследована степень однородности наблюдений, выражающаяся в неодинаковой дисперсии случайной ошибки (остатка) регрессии. Нулевая гипотеза теста предполагает равномерный разброс остатков. В результате теста получено, что нулевая гипотеза может быть отвергнута, т.е. присутствует гетероскедастичность, которая приводит к росту влияния данных с высокой дисперсией на регрессионную модель.

Следующий этап исследований состоял в оценке эффективности применения различных методов сглаживания для рассматриваемого временного ряда. Актуальность таких оценок связана с указанным выше сложным характером временного ряда. Проверка работы вначале выполнялась для тестовой выборки, созданной генератором случайных чисел. Необходимо было найти фильтр, который, сохраняя форму кривой, обеспечивает наибольшее сглаживание.

Вначале были опробованы равновесовые фильтры скользящего среднего: трех- и семиточечные. Применение фильтров позволило уменьшить величину максимума в интервале отсчетов 55850-55855. В результате работы семиточечного фильтра число точек над уровнем 3σ стало меньше в 4 раза, но число точек над уровнем 6σ не изменилось, т.к. максимум расширился вследствие сглаживания. Далее была протестирована работа фильтров другого вида: неравновесовое скользящее среднее (веса треугольной формы), фильтр Савицкого-Голея, фильтр Винера. Результаты применения фильтров представлены на рис. 3.

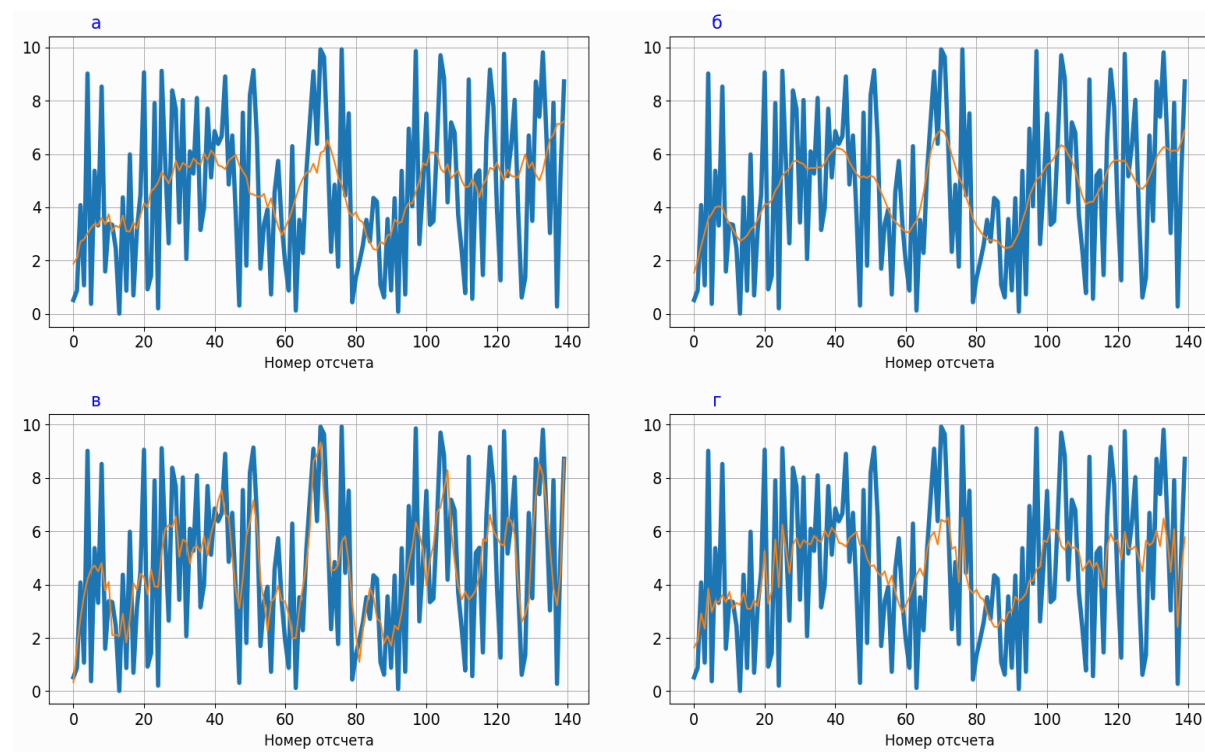


Рис. 3. Результаты применения сглаживающих фильтров для тестовой выборки:
а – скользящее среднее с равными весами; б – скользящее среднее с треугольной весовой функцией;
в – фильтр Савицкого-Голея; г – фильтр Винера

Оценка степени сглаживания выполнялась с использованием коэффициента корреляции k исходной и сглаженной кривой. Для критерия, предусматривающего сохранение формы сигнала, наилучшее сглаживание обеспечивает фильтр Винера ($k \approx 0.56$), имеющий рекурсивный алгоритм, подстраивающийся под параметры фильтруемого сигнала. Фильтр Савицкого-Голея характеризуется $k \approx 0.52$, фильтр скользящего среднего с треугольной весовой функцией – $k \approx 0.47$. Наименьшее $k \approx 0.27$ имеет фильтр скользящего среднего с равными весами. Таким образом, из рассмотренных методов сглаживания винеровская фильтрация наиболее предпочтительна.

Полученные статистические оценки (логнормальное распределение данных, низкие коэффициенты детерминации регрессионных моделей, гетероскедастичность случайных ошибок линейной регрессионной модели) указывают, что весьма вероятен нестационарный характер временного ряда (изменение основных параметров процесса во времени). В связи с этим выполнены два теста на стационарность временного ряда.

Первый – тест Краскела-Уоллиса – проверяет нулевую гипотезу о том, что медианы всех групп, на которые разбита последовательность, равны. Это наиболее простая и одновременно наглядная проверка (рис. 4): для стационарных процессов статистические характеристики выборки должны совпадать с характеристиками всего ряда.

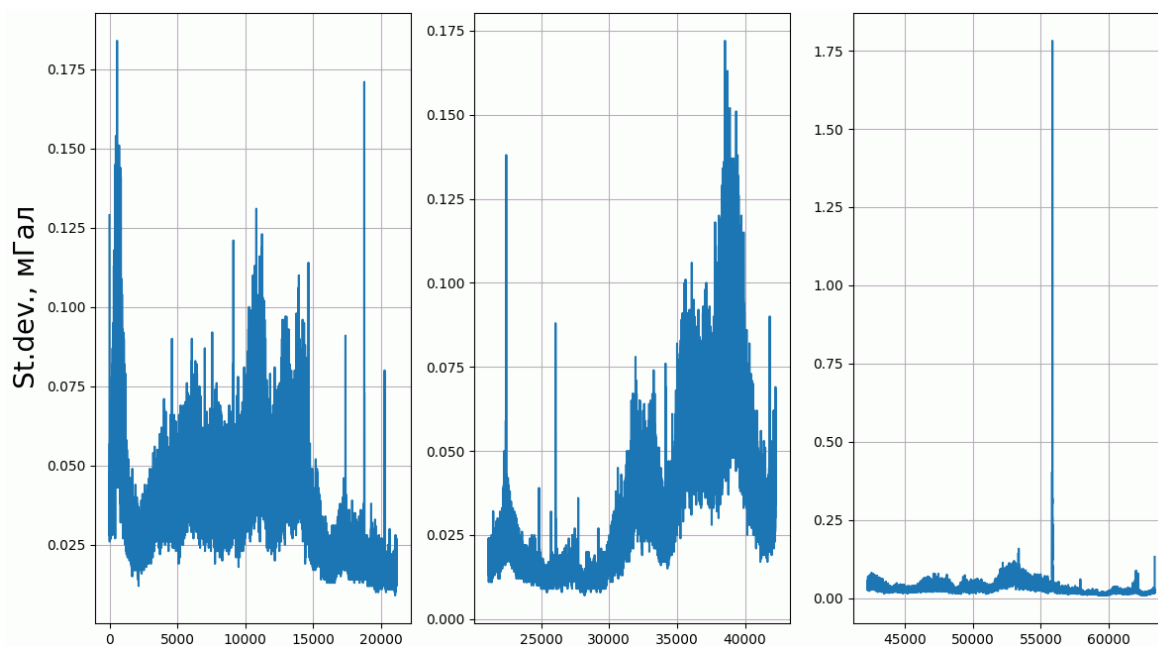


Рис. 4. Выборки временного ряда для выполнения теста Краскела-Уоллиса

В результате выполнения теста получилось, что нулевая гипотеза (равенство медиан вследствие стационарности ряда) отвергается. Результаты второго теста (расширенный тест Дикки-Фуллера) также показали, что гипотезу о нестационарности нельзя отвергнуть.

Нестационарность ряда означает, что при его изучении необходимо выделять и отдельно анализировать разные временные компоненты: трендовую, сезонную и остаточную составляющие. Раздельное исследование последовательно отделяемых компонент позволяет более достоверно оценивать временной ряд и делать более точные выводы: можно предсказывать характер протекающего процесса в долгосрочной перспективе, имея информацию о тренде; зная сезонную составляющую, можно корректировать прогноз; исследование случайной величины позволяют найти лучшие способы ее исключения.

Одним из самых распространенных методов изучения временных рядов является их частотный анализ [7, 8]. При спектральных исследованиях сигналов одной из основных целей является определение спектральной плотности мощности (СПМ). Для определения СПМ могут использоваться прямой или косвенный метод. Прямой метод является, пожалуй, самым простым и связан с вычислением квадрата модуля дискретного преобразования Фурье отдельных участков последовательности, получаемая при этом характеристика называется периодограммой. Стандартные подходы к вычислению периодограмм могут давать значительную зашумленность результата, в связи с чем разрабатывались различные методы расчета, предусматривающие снижение шума. Одним из самых распространенных является метод Уэлча (Welch's method), этот метод использован в данной работе. СПМ для исследуемого временного ряда, полученная по методу Уэлча, представлена на рис. 5. На диаграмме спектральной плотности мощности удалена шумовая составляющая, что позволяет хорошо выделять преобладающие полосы частот: наиболее низкочастотные значения, примерно до 0.09 Гц имеют наибольшую СПМ.

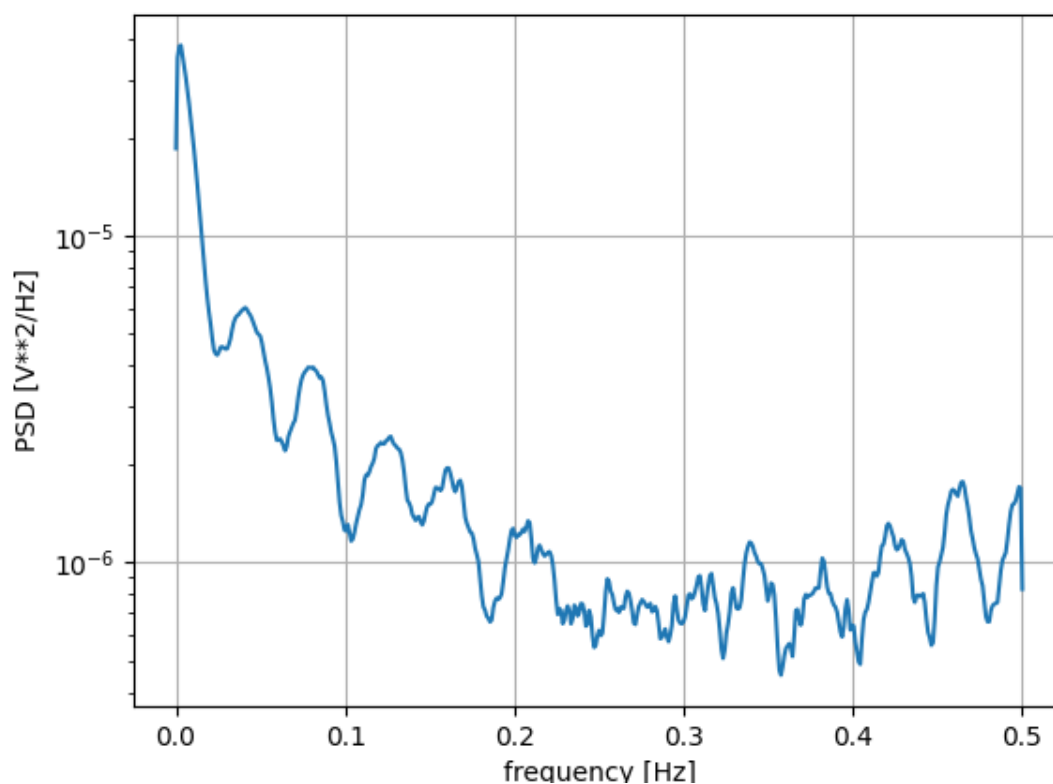


Рис. 5. Диаграмма спектральной плотности мощности, вычисленная по методу Уэлча

Выводы

В работе продолжены исследования, связанные с разработкой методов анализа долговременных записей гравиметров и созданием программы «Анализ временных рядов». Представлены результаты: оценки вида распределения данных; применения регрессионного анализа и оценки его достоверности; работы различных методов сглаживания; оценки стационарности временных рядов; исследования частотного состава сигнала с оценкой спектральной плотности мощности по методу Уэлча.

Представленные подходы дают основу для сопоставления записей гравиметров, построения расширенной классификации временных рядов [4-8] и имеют возможности применения для анализа профильных и площадных гравиметрических данных.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Антонов Ю.В. Высокочастотные колебания неприливных сейсмо-гравитационных пульсаций // Вестник ВГУ. Сер. Геология. – 2022. – № 3. – С. 66-75. – DOI: 10.17308/geology/1609-0691/2022/3/66-75.
2. Геник И.В. Об обработке долговременных записей гравиметра Autograv // Горное эхо. – 2023. – № 4 (93). – С. 56-61. – DOI: 10.7242/echo.2023.4.8.
3. Геник И.В. Применение статистических методов для прогнозирования результатов региональных гравиметрических работ // Стратегия и процессы освоения георесурсов: сб. науч. тр. Вып. 15 / ГИ УрО РАН. – Пермь, 2017. – С. 150-153.
4. Никитин А.А., Петров А.В. Теоретические основы обработки геофизической информации: учеб. пособие. – 3-е изд. доп. – М. ВНИИГеосистем, 2013. – 117 с.: ил.
5. Никитин А.А., Петров А.В., Зиновкин С.В. Развитие статистических приемов обработки и интерпретации геофизических полей в компьютерной технологии КОСКАД 3D // Изв. вузов. Геология и разведка. – 2007. – № 6. – С. 68-74.
6. Хайндман Р., Атанасопулос Д. Прогнозирование: принципы и практика. – М.: ДМК Пресс, 2023. – 458 с.
7. Mills T.C. Applied Time Series Analysis: A Practical Guide to Modeling and Forecasting. – Elsevier, 2019. – 354 p.
8. Nielsen A. Practical Time Series Analysis: Prediction with Statistics and Machine Learning. – Sebastopol: O'Reilly Media, 2019. – 504 p.

УДК 550.834.05

DOI:10.7242/echo.2025.1.7

ИСПОЛЬЗОВАНИЕ СЕЙСМИЧЕСКИХ АТТРИБУТОВ ДЛЯ ИЗУЧЕНИЯ ОСОБЕННОСТЕЙ СТРОЕНИЯ МАССИВА ЛЕДЯНОЙ ГОРЫ

И.Ю. Герасимова

Горный институт УрО РАН, г. Пермь

Аннотация: Представлены результаты анализа сейсмических атрибутов, проведенного с целью локализации ослабленных зон в массиве Ледяной горы, включающем Кунгурскую пещеру. Оценка распределений суммарного коэффициента, сформированного на основе данных многопараметрического моделирования и анализе динамических параметров упругих волн позволяет на качественном уровне выделить участки проявления подземного карста.

Ключевые слова: динамические атрибуты, Кунгурская ледяная пещера, верхняя часть разреза, карстовая полость.

Массив Ледяной горы, включающий Кунгурскую пещеру, используется как естественный полигон для геофизических исследований с середины XX века. Опытнометодические работы направлены на решение разнообразных задач, связанных как с оптимизацией методик и технологий проведения полевых наблюдений, так и с повышением результативности данных, получаемых на этапах обработки и интерпретации. Одним из изучаемых вопросов является локализация полостей Кунгурской пещеры в регистрируемых полях. В статье представлены материалы, полученные в рамках решения указанной задачи, базирующиеся на изменении динамических характеристик упругих волн.

Территория отличается расчлененным рельефом, сложным геологическим и структурно-тектоническим строением [1, 4], а также развитием карстовых и карстово-суффозионных процессов, способствующих образованию провалов и воронок на поверхности массива и формированию в толще пород на разных гипсометрических уровнях пустот размерами от нескольких миллиметров до значительных по объемам пещер. По данным топографического плана Кунгурской ледяной пещеры [2] средняя ширина гrotтов и галерей составляет 7,8 м, а средняя высота – 2,6 м. Подземные полости перекры-