

Министерство науки и высшего образования Российской Федерации
Федеральное государственное автономное учреждение
высшего образования
«Национальный исследовательский ядерный университет «МИФИ»
(НИЯУ МИФИ)»

ОТЧЁТ О НАУЧНО-ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЕ

Анализ скорости счета гамма-квантов, регистрируемых в подземном детекторе LVD, с целью предсказания сильных сейсмических событий

Москва 2026

СОДЕРЖАНИЕ

Введение	3
1. Детектор большого объёма (LVD)	4
1.1. Описание установки	4
1.2. Методика регистрации	5
1.3. Гамма-кванты от распада радона – фон эксперимента	6
1.4. Предвестники землетрясений	7
2. Прогнозирование сейсмических событий	8
2.1. Постановка задачи	8
2.2. Описание данных	9
2.3. Обработка и анализ данных	10
2.3.1. Первичная обработка	10
2.3.2. Определение интервала прогнозирования	11
2.3.3. Поиск корреляций	12
2.3.4. Поиск паттернов	13
2.4. О постановке задачи машинного обучения	19
2.5. Формирование признаков	20
2.6. Обучение моделей	25
2.7. Поиск проблем и рассмотрение альтернативных подходов	26
Заключение	27
Список литературы	28

ВВЕДЕНИЕ

Актуальность

Расположенный в подземной лаборатории Гран-Сассо, Италия, сцинтилляционный детектор LVD регистрирует гамма-кванты от естественной радиоактивности материалов конструкции и грунта, окружающего установку. Гамма-кванты (с энергией до 3 МэВ) рождаются от распадов радиоактивных радона и торона, от цепочек распадов Урана и Тория. Известно, что перед мощными землетрясениями происходит усиленный выход подземного радона из микротрещин грунта.

Исследование вариаций скорости гамма-квантов с использованием накопленной статистики архивных данных за более чем 20 лет и онлайн мониторирование выхода радона поможет изучать, как предвестники землетрясений, так и фон в экспериментах по поиску нейтрино и частиц темной материи в низкофоновой лаборатории Гран-Сассо.

Цель исследований

Целью работы является на основе экспериментальных данных по скорости счета гамма-квантов, регистрируемых детектором LVD, от естественной радиоактивности, построить с применением машинного обучения модель для обнаружения предвестниковой фазы землетрясений.

Задачи исследований

- Изучить работу детектора LVD, его характеристики и физику экспериментальных исследований.
- Изучить особенности формирования экспериментальных данных по программе исследований вариаций гамма-квантов.
- Проанализировать экспериментальные данные для выделения характерных для предвестников землетрясений форм сигнала.
- Найти подход к решению задачи по поиску предвестников землетрясений с точки зрения машинного обучения.

1. ДЕТЕКТОР БОЛЬШОГО ОБЪЁМА (LVD)

1.1. Описание установки

Детектор LVD (Large Volume Detector) расположен на глубине 3650 м в.э. в зале А лаборатории Гран-Сассо (рис. 1). Установка сконструирована для регистрации нейтрино, возникающих при коллапсах звездных ядер в нашей Галактике.

Детектор содержит 1000 тонн жидкостного сцинтиллятора на основе уайт-спирта (C_nH_{2n}), который заполняет 840 счетчиков объемом 1.5 м³ каждый. Счетчики имеют форму параллелепипеда (100×100×150 см³) и изготовлены из нержавеющей стали. На верхней грани каждого счетчика установлены три фотоумножителя, работающие в режиме совпадения. Счетчики размещены по 8 штук в металлические модули-контейнеры (портатанки), из которых формируются вертикальные колонны. Пять колонн из 7 портатанков объединяются в одну башню, таким образом, одна башня содержит 35 модулей с 280 счетчиками (рис. 2).

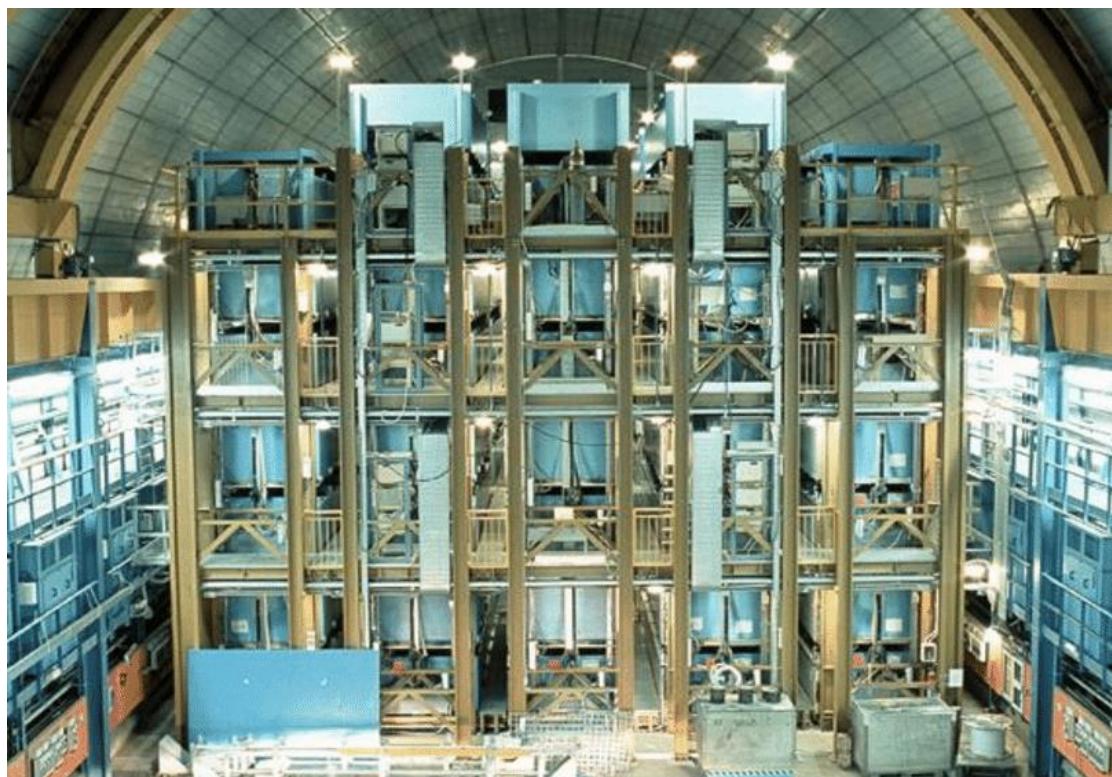


Рис. 1. Детектор LVD в лаборатории Гран-Сассо

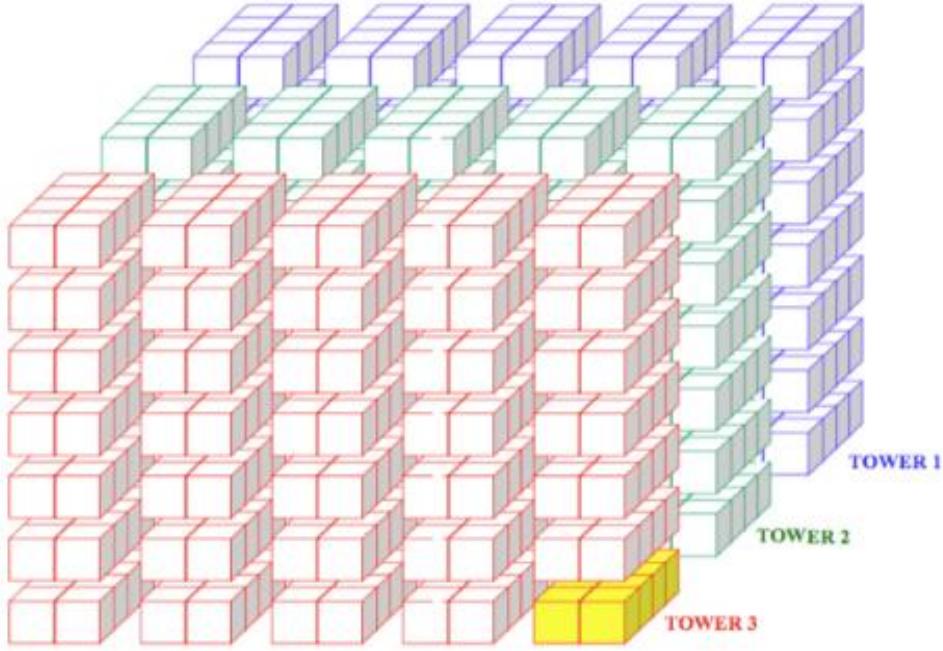


Рис. 2. Пространственная конфигурация счётчиков в составе детектора

1.2. Методика регистрации

Для разных научных задач в детекторе имеется два энергетических порога: верхний $E_{HET} = 4$ МэВ, и нижний $E_{LET} = 0.5$ МэВ. Данные низкоэнергетического порога используются для исследования фона эксперимента и контроля работы счетчиков [1].

Установка LVD способна регистрировать нейтрино всех типов по реакциям взаимодействия с ядрами веществ, входящих в её структуру – водородом, углеродом, железом. Основной реакцией взаимодействия антинейтрино является обратный бета-распад (реакция ОБР): $\bar{\nu}_e + p \rightarrow e^+ + n$, $E_e = E_\nu - 1.3$ МэВ, которая создает два детектируемых сигнала: первый сигнал вызывается позитроном, следующий за ним – захватом нейтрона протоном: $n + p \rightarrow d + \gamma$ ($E_\gamma = 2.2$ МэВ, среднее время захвата около 185 мкс) или железом, входящим в структуру установки: $n + {}^{56}_{26}Fe \rightarrow {}^{57}_{26}Fe + \Sigma\gamma$ ($E_\gamma \sim 7$ МэВ, $\tau \sim 110$ мкс).

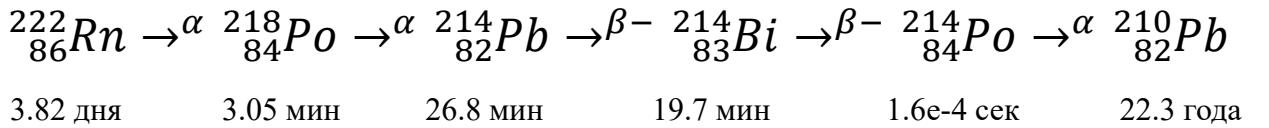
Фоном детектора LVD при поиске нейтрино являются атмосферные мюоны (средняя энергия которых около 280 ГэВ, скорость счёта мюонов на

счетчик $\sim 10^{-4}$ сек $^{-1}$ счётчик $^{-1}$), и естественная радиоактивность скального грунта и материалов установки (для внутренних счетчиков первой башни (~ 45 сек $^{-1}$ счётчик $^{-1}$).

1.3. Гамма-кванты от распада радона – фон эксперимента

Основным источником фона на установке LVD является радон. Естественный радиоактивный газ образуется в результате распада радия, который, в свою очередь, является продуктом распада урана-238. Благодаря своим уникальным свойствам радон служит важным индикатором в различных геологических исследованиях. Этот газ постоянно генерируется в горных породах в процессе радиоактивного распада, что делает его присутствие неизменным в любом горном массиве. Уменьшение концентрации радона происходит по двум основным причинам: а) в результате его распада (период полураспада радона составляет 3.825 дня), и б) из-за миграции газа из массива в атмосферу. Однако новая генерация радона возникает из цепочки распадов урана, что обеспечивает постоянное среднее содержание радона в грунте, определяемое концентрацией урана (радия) в данном массиве.

Установка LVD способна регистрировать гамма-кванты, возникающие при распадах дочерних ядер радона $^{222}_{86}Rn$. Наиболее вероятная цепочка радиоактивных распадов выглядит следующим образом:



Гамма-излучение в основном создается ядрами $^{214}_{83}Bi$, которые, распадаясь, превращаются в $^{214}_{84}Po$ с характерным временем распада $T=19.7$ мин. Энергетический спектр гамма-излучения охватывает диапазон от 0.6 до 2.5 МэВ.

Данные детектора, используемые для исследования фона (E_{LET}), формируются следующим образом: каждые 10 минут все сцинтилляционные

счётчики башни запускаются импульсом от генератора для регистрации E_{LET} , импульсов в течение 10 секунд. Ввиду невысоких энергий гамма-кванты проникают в счётчик главным образом через верхнюю грань, толщина металла которой 4 мм. Боковые и нижняя грани экранированы дополнительными 20 мм металла.

При обработке данных отбираются наиболее стабильно работающие в течение длительного времени счётчики. Суммарный темп счета нормируется на один счетчик в герцах (Hz) и измеряется автоматически. Количество отсчетов, обусловленных радоном, для 70 внутренних счетчиков одной башни составляет примерно 300000 за 10 секунд. Это обеспечивает чувствительность метода 5% на уровне 3σ , что означает, что за 10-секундный интервал измерений установка сможет зафиксировать 5%-е отклонение концентрации с достоверностью 99.7%. Данные суммируются в бине за 1 час. Общее число отсчетов в одном бине составляет около 184000 (420 отсчетов за 10 секунд \times 73 счетчика \times 6 запусков в час), что приводит к относительной ошибке измерений в 0.2%.

1.4. Предвестники землетрясений

Некоторые исследования [2]-[4] указывают на связь аномальных выбросов радона с крупными землетрясениями, что может свидетельствовать о деформации земной коры. Радиоактивные свойства радона делают его мощным индикатором движения земной коры и потенциально эффективным маркером для изучения процессов, связанных с подготовительной фазой землетрясений.

Исследования, направленные на выявление связи между поведением радоновых полей и сейсмической активностью, проводятся по всему миру на протяжении многих десятилетий. В экспериментальных данных LVD наблюдается отклик установки на увеличение сейсмической активности. В некоторых случаях фиксируется рост концентрации радона за 2-3 дня до

землетрясения, что внушает надежду на возможность предсказания предстоящих событий.

Измерения температуры, влажности и давления в зале установки показали, что условия эксплуатации детектора стабильны как в краткосрочной (в течение суток), так и в долгосрочной перспективе (в течение года) [3]. Это создает благоприятные условия для выявления резких изменений концентрации радона и, возможно, для определения предвестников землетрясений.

Однако прогнозирование землетрясений требует предсказания с определенной достоверностью трех характеристик: силы, времени и места очага землетрясения. В настоящее же время проводятся исследования по выявлению источников фоновых радоновых событий [5] и определению характеристик вариаций концентрации радона под землей, которые коррелируют со временем сильных землетрясений.

2. ПРОГОНОЗИРОВАНИЕ СЕЙСМИЧЕСКИХ СОБЫТИЙ

2.1. Постановка задачи

Прогнозирование силы, времени и места землетрясения на основе лишь радоновых данных является задачей если не нерешаемой, то по крайней мере очень сложной. Для её решения нужно применять комплексный подход и подключать данные из других источников. В рамках имеющейся информации о временных рядах имеет смысл решать задачу по прогнозированию самого факта наличия землетрясения.

Если опираться на имеющиеся исследования и полагать, что существует некоторая взаимосвязь между поведением радона незадолго до события и самим событием, задача прогнозирования землетрясения сводится к задаче определения предвестниковой фазы.

Таким образом, стоит задача по определению предвестниковой фазы землетрясения на основе временных рядов скорости счёта гамма-квантов и давления в лаборатории.

2.2. Описание данных

Используемые данные можно поделить на два типа:

- Лабораторные данные – информация, полученная с датчиков установки LVD;
- Данные о землетрясениях.

Лабораторные данные представляют из себя временные ряды давления и скорости счета гамма-квантов (рис. 3), усредненные за час работы детектора и нормированные на число стабильно работающих счетчиков (приведенные в секунду).

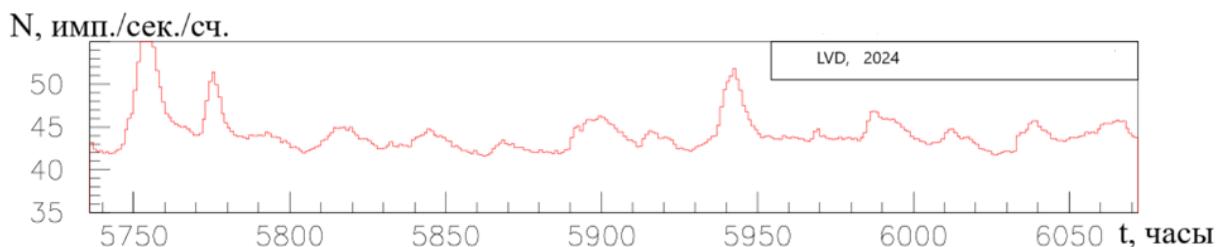


Рис.3. Пример временного ряда скорости счета гамма-квантов на LVD за 2024 год, время в часах с начала года, бин – 1 час.

Данные о землетрясениях получены из следующих источников:

1. Камчатский филиал ФИЦ ЕГС РАН (EMSD) [6];
2. Глобальная база данных значительных землетрясений NCEI/WDS (NOAA) [7];
3. Департамент землетрясений национального института геофизики и вулканологии Италии (INGV) [8];
4. Геофизическая служба РАН (GSRAS) [9].

Они представляют из себя таблицу, в которой содержится информация о каждом событии: дата и время, широта, долгота, глубина, магнитуда, интенсивность, геодезическое расстояние [10] от LVD.

2.3. Обработка и анализ данных

2.3.1. Первичная обработка

На первом этапе данные о землетрясениях из разных источников были приведены к одному формату, объединены и очищены от дубликатов. Далее произведена фильтрация по магнитуде – отсеиваются слабые события, – а также по дистанции от лаборатории (в исходной таблице есть информация о землетрясениях практически со всего земного шара). Значения обоих фильтров являются параметрами для экспериментов. В рамках описанного далее анализа были заданы следующие величины:

- Минимальная магнитуда $M_{min} = 3.5$;
- Максимальное расстояние от установки $R_{max} = 1500$ км.

Распределение отфильтрованных землетрясений по географической карте показано на рисунке 4.

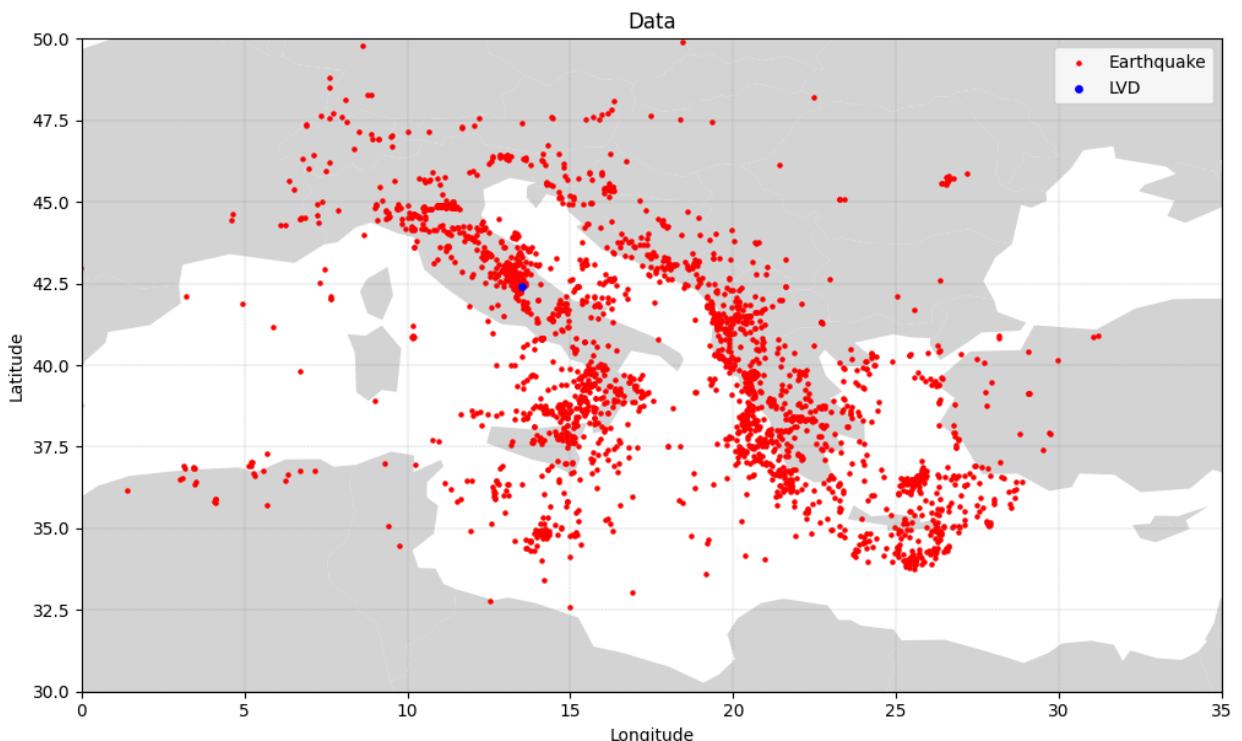


Рис.4. Распределение сейсмических событий по карте

Временные ряды давления и скорости счёта гамма-квантов (далее – радон или радоновые данные) были объединены и усреднены каждый по трём башням. Объединение происходило по условию пересечения диапазонов дат, чтобы ни для одного из рядов на концах не оставалось пустых значений.

2.3.2. Определение интервала прогнозирования

Для корректной формализации задачи нужно, помимо прочего, определить диапазон прогнозирования. Иными словами, необходимо ограничить интервала времени до землетрясения, в котором возникает предвестниковая фаза. Поскольку существующие исследования указывают на рост радона, в основу подхода к решению данной задачи легло выявление его пиковых значений (рис. 5). Параметры алгоритма для их поиска так же подбираются из соображений экспериментатора. В данном случае минимальная дистанция между пиками была 12 точек (часов), а их минимальная амплитуда – половина среднеквадратического отклонения, посчитанного по всему набору данных.

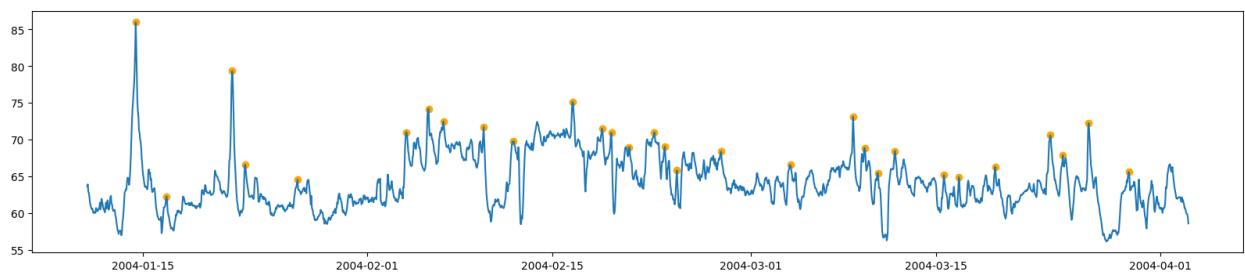


Рис. 5. Пример выявленных пиков радона

Далее для каждого пика было найдено ближайшее (в будущем) землетрясение, по двум датам рассчитан лаг. Одним из критериев подбора адекватных параметров для поиска пиков было равенство между числом лагов и количеством землетрясений. В результате, было получено распределение лагов по количеству дней (рис. 6), а также рассчитаны среднее

значение и среднеквадратическое отклонение, которые составили 5.68 и 7.02 дня соответственно. Таким образом:

- σ – 7 дней – включают 73.9% лагов;
- 2σ – 14 дней – включают 89.3% лагов;
- 3σ – 21 день – включает 95.1% лагов.

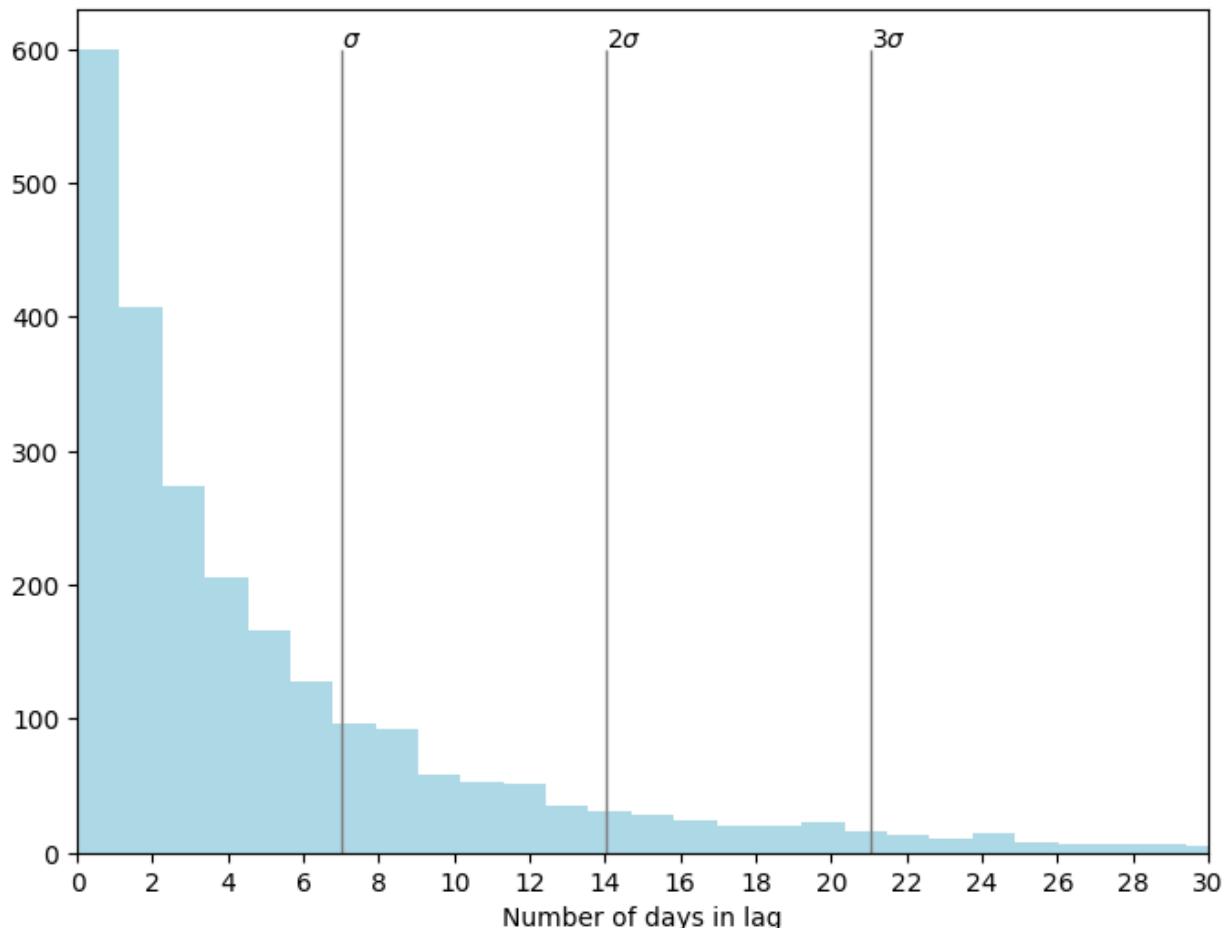


Рис. 6. Распределение лагов по количеству дней

Результат не является строгим и однозначным, но даёт некоторое представление о том, в каком диапазоне имеет смысл искать паттерны, соответствующие предвестникам землетрясений.

2.3.3. Поиск корреляций

Также был произведён расчёт корреляции между значениями радона и фактом землетрясения. Формирование вектора значений для последнего

производилось следующим образом: для каждой даты проверялось, есть ли в ближайшие 96 часов после неё землетрясения, результат проверки принимал значение 0 или 1. Далее рассчитывалась корреляция при разных сдвигах (± 14 дней) между датами двух векторов значений (рис. 7).

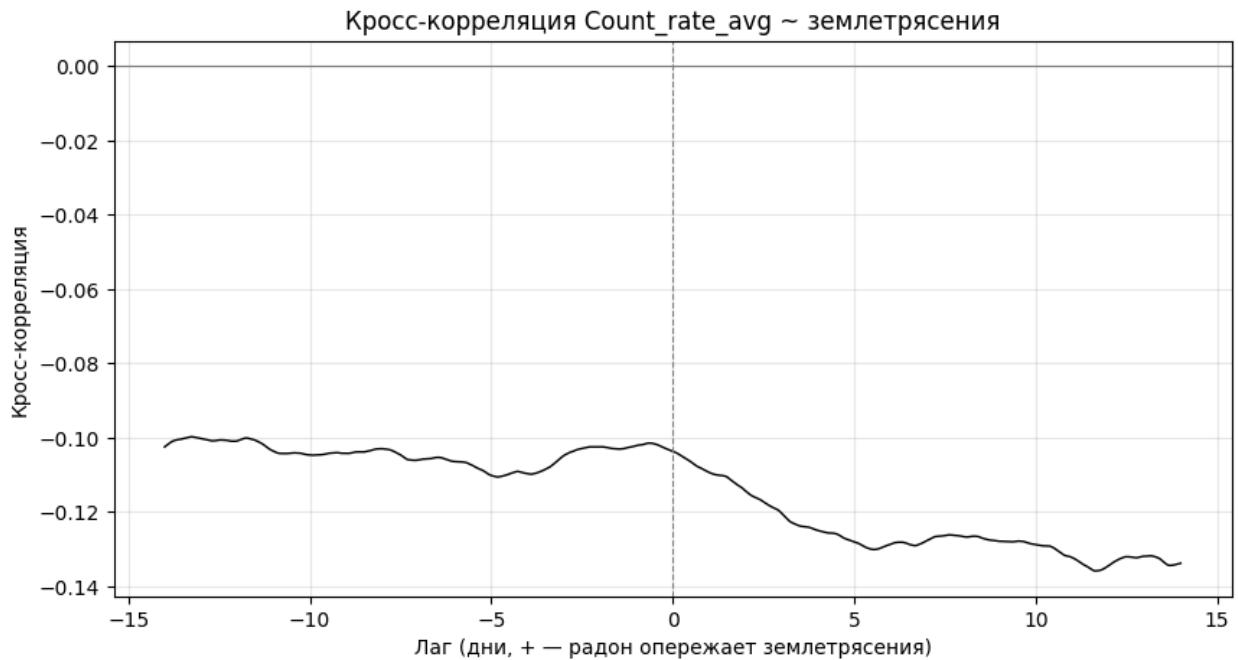


Рис. 7. Корреляция между значениями радона и землетрясениями

Линейная зависимость между фактом землетрясения и значением радона ожидаемо отсутствует.

2.3.4. Поиск паттернов

Основным этапом анализа исходных данных является поиск паттернов временных рядов. Если известно о некотором аномальном поведении радона вблизи землетрясений, имеет смысл проверить, опираясь на ранее полученные значения лагов, существует ли типичный шаблон изменения радона незадолго до событий, например: рост перед, спад после, аномальные флюктуации и т. д.

Методика следующая: для каждого землетрясения берётся окно $\pm \Delta$ дней (в данном случае ± 14) вокруг даты события, из радонового ряда вырезается

соответствующий фрагмент и нормируется относительно среднего в окне уровня радона. Далее все эти фрагменты усредняются между собой, получается «средняя форма сигнала» до и после землетрясения (рис. 8).

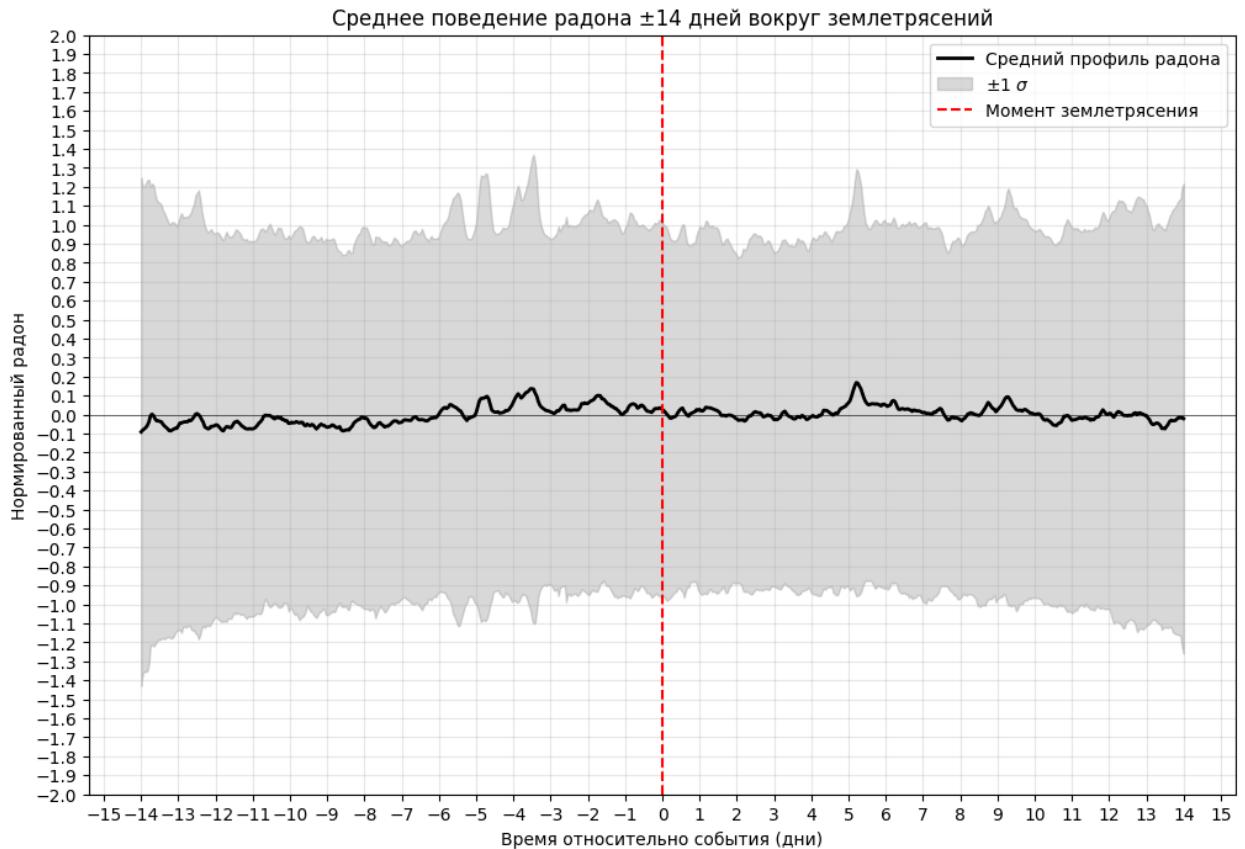


Рис. 8. Средняя форма сигнала радона в окне ± 14 вокруг даты землетрясения

На графике можно видеть три выделяющихся пика за 5.6, 4.8 и 3.5 дня до события. Это характерный паттерн, который в дальнейшем будет использоваться при построении модели, однако до этого полученный результат проверяется на устойчивость и значимость.

Первая проверка произведена на случайных подмножествах землетрясений, каждое размером 10% от всего массива данных (рис. 9).

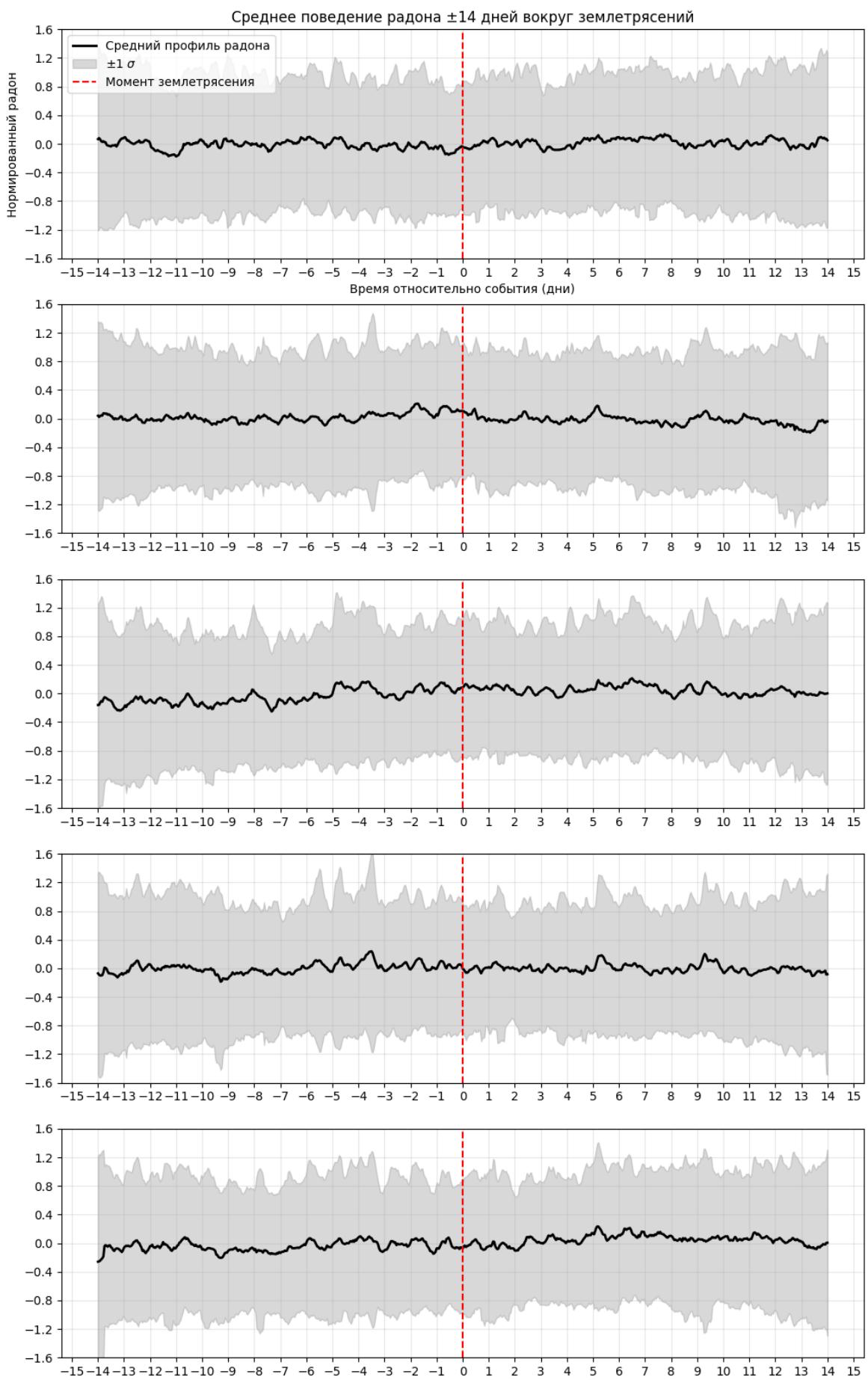


Рис. 9. Проверка паттерна радона на устойчивость

Как можно видеть, паттерн наблюдается на выборках 10% от общего числа землетрясений, однако не всегда. Основной и самый стабильный по подмножествам пик появляется за 3.5 дня до события. Помимо этого, в ожидаемом диапазоне пиков среднее значение радион зачастую явно выше, чем в предыдущие дни.

Вторая проверка произведена на случайном подмножестве дат без землетрясений, его размер будет равен размеру всей выборки дат с землетрясениями (рис. 10).

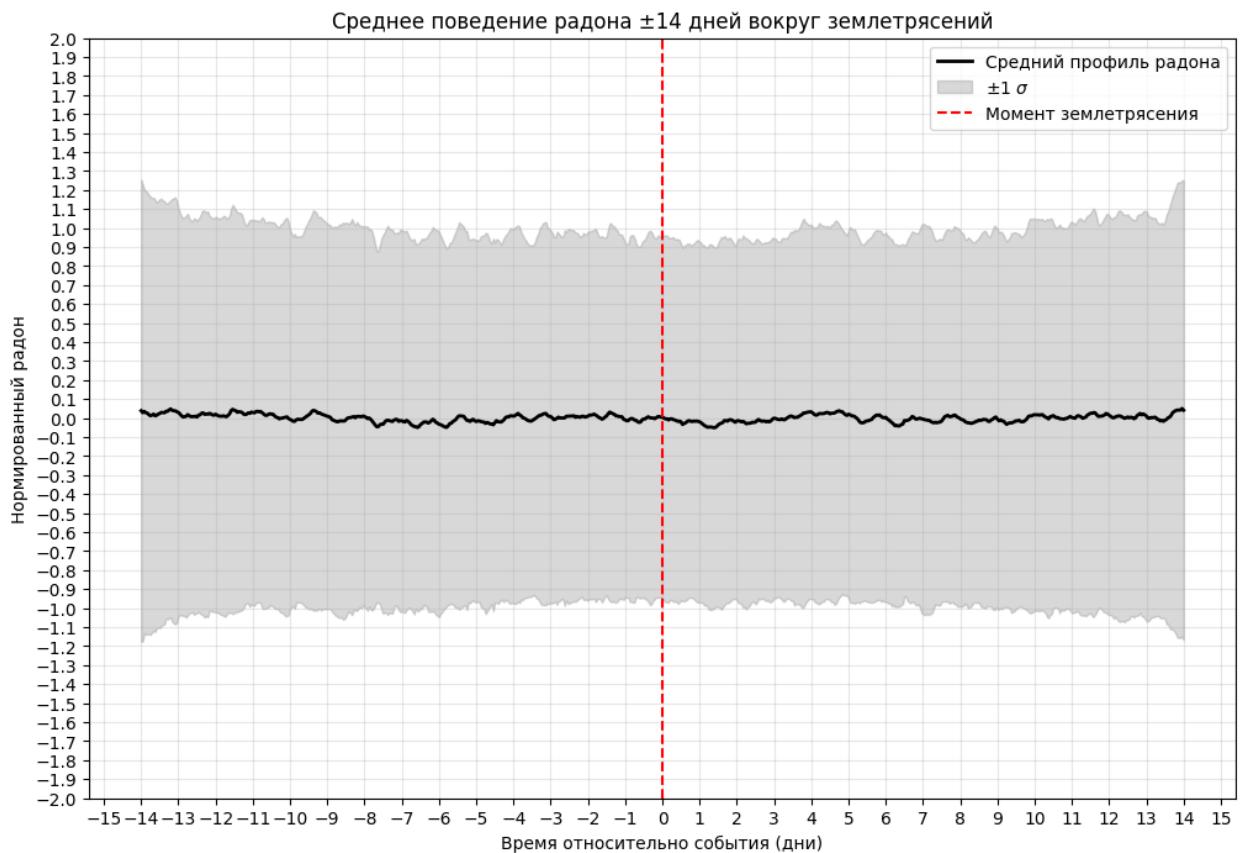


Рис. 10. Проверка паттерна радион на значимость

При размере случайных выборок дат без землетрясений, сопоставимом с общим числом землетрясений, выделенный паттерн, как и какое-либо аномальное поведение, не наблюдается. На основе этого можно допустить, что выделенный сигнал является статистически значимым.

Все вышеизложенные операции по поиску и проверке паттернов аналогичным образом проделаны с временным рядом давления.

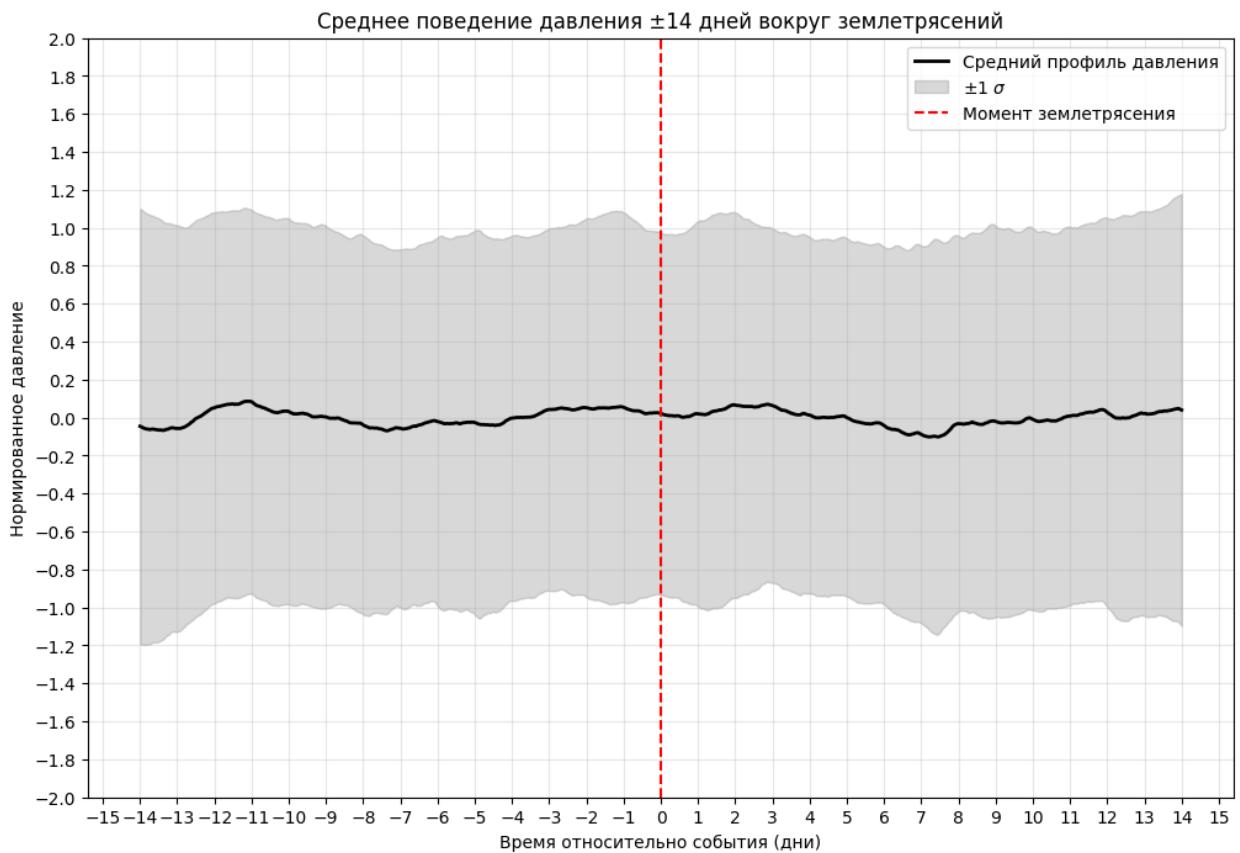


Рис. 11. Средняя форма сигнала давления в окне ± 14 вокруг даты землетрясения

На рис. 11 виден паттерн в виде гладкого пика за 1-2 дня до землетрясения в конце тренда на рост длительностью приблизительно 6 дней. Проверка на устойчивость по случайным подмножествам дат землетрясений на рис. 12. Видно, что за 1-3 дня до события наблюдается небольшой пик.

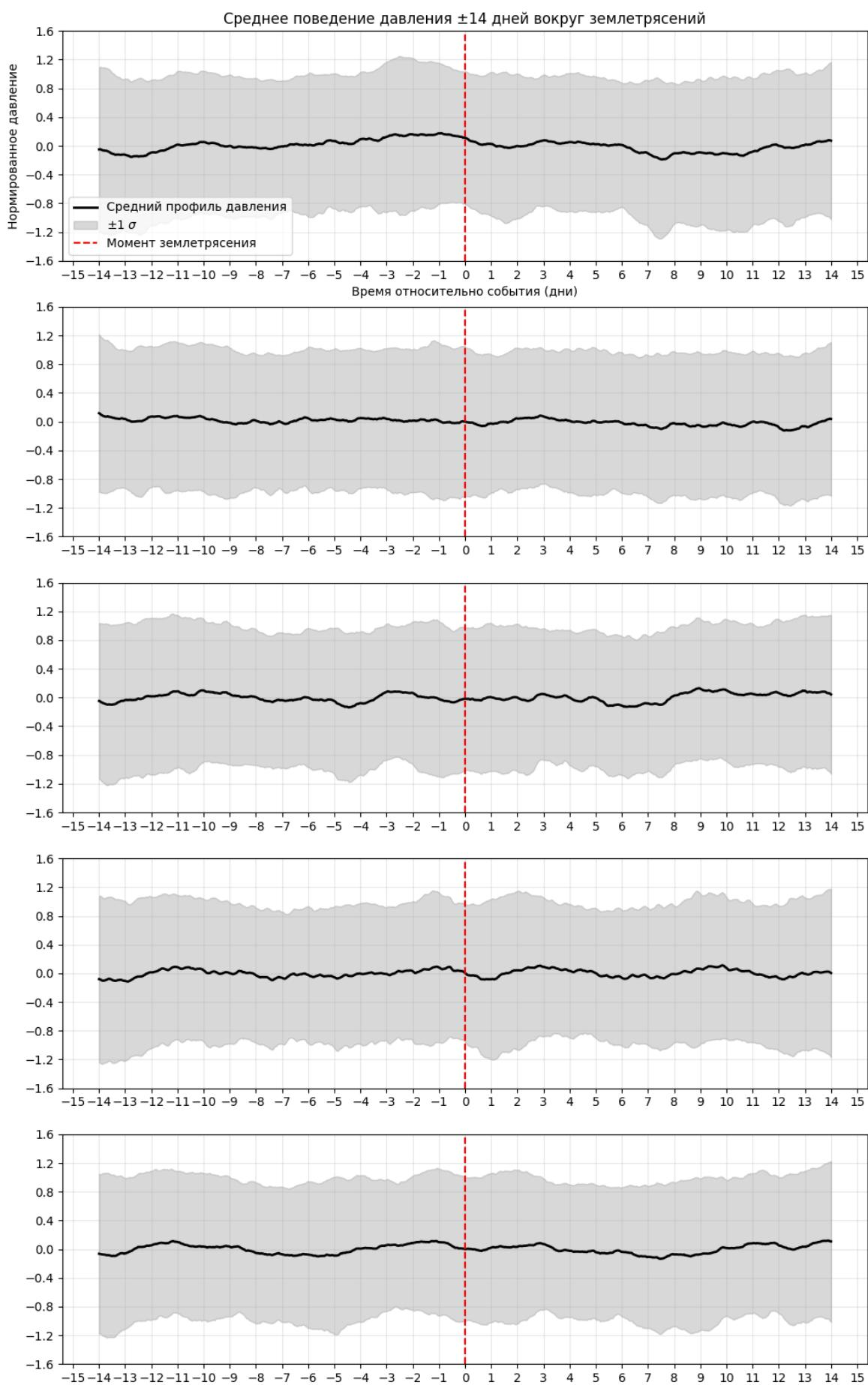


Рис. 12. Проверка паттерна давления на устойчивость

На рис. 13 видно, что при усреднении по датам без событий структура сигнала гораздо стабильнее той, которая наблюдается вокруг землетрясений.

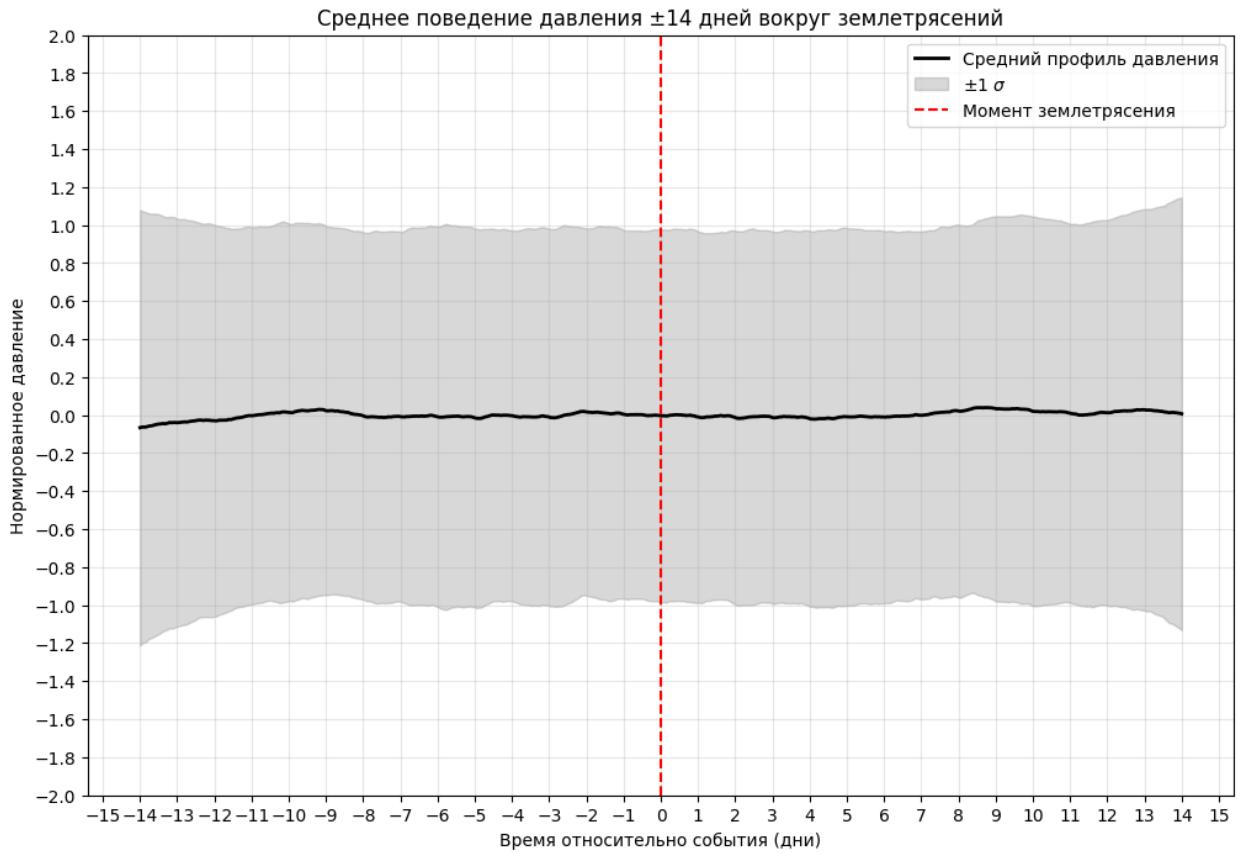


Рис. 13. Проверка паттерна давления на значимость

2.4. О постановке задачи машинного обучения

В ходе решения задачи изначальная постановка была сформулирована как: «оценка вероятности возникновения землетрясения с $M \geq n$ в радиусе $R \leq m$ км в течение следующих k суток», где $n = M_{min}$ – минимальная магнитуда, $m = R_{max}$ – максимальное расстояние от LVD. Тем не менее, такая формулировка не является строгой, поскольку при использовании в анализе данных, полученных из паттернов временных рядов, предшествующих событию, фактически не прогнозируется некая вероятность землетрясения, а даётся некоторая оценка (score) состояния, в котором на данный момент находится система.

2.5. Формирование признаков

При решении поставленной задачи можно использовать множество разных подходов, каждый формируя и проверяя эмпирически.

Перед началом извлечения признаков из паттернов временных рядов главным шагом стал переход от почасовых данных к посutoчным. Такое изменение масштаба уменьшает шум, но вместе с тем теряется часть информации, поэтому в процессе производится агрегация по суточным окнам с извлечением: максимума, минимума, среднего, среднеквадратичного отклонения, дисперсии, медианы. Важно при этом решить, к какой границе суточного интервала будут приписываться полученные значения. Лучшим решением является выбрать левую границу, поскольку в таком случае паттерны не будут сдвигаться направо, за время события, т.е. в ту область, куда модели нельзя «заглядывать». Кроме того, это имеет смысл и с практической точки зрения: при мониторинге в настоящем времени каждое новое наблюдение будет идти в счёт временной отметки, находящейся в начале суток, и ждать их конца для обновления прогноза не нужно.

После такого перехода имеет смысл проверить, что произошло с ранее выявленными паттернами (рис. 14, 15). Можно видеть, что ранее выявленные явные пики стали выглядеть, как заметный тренд на ускоренный рост радона, картина для давления практически не изменилась.

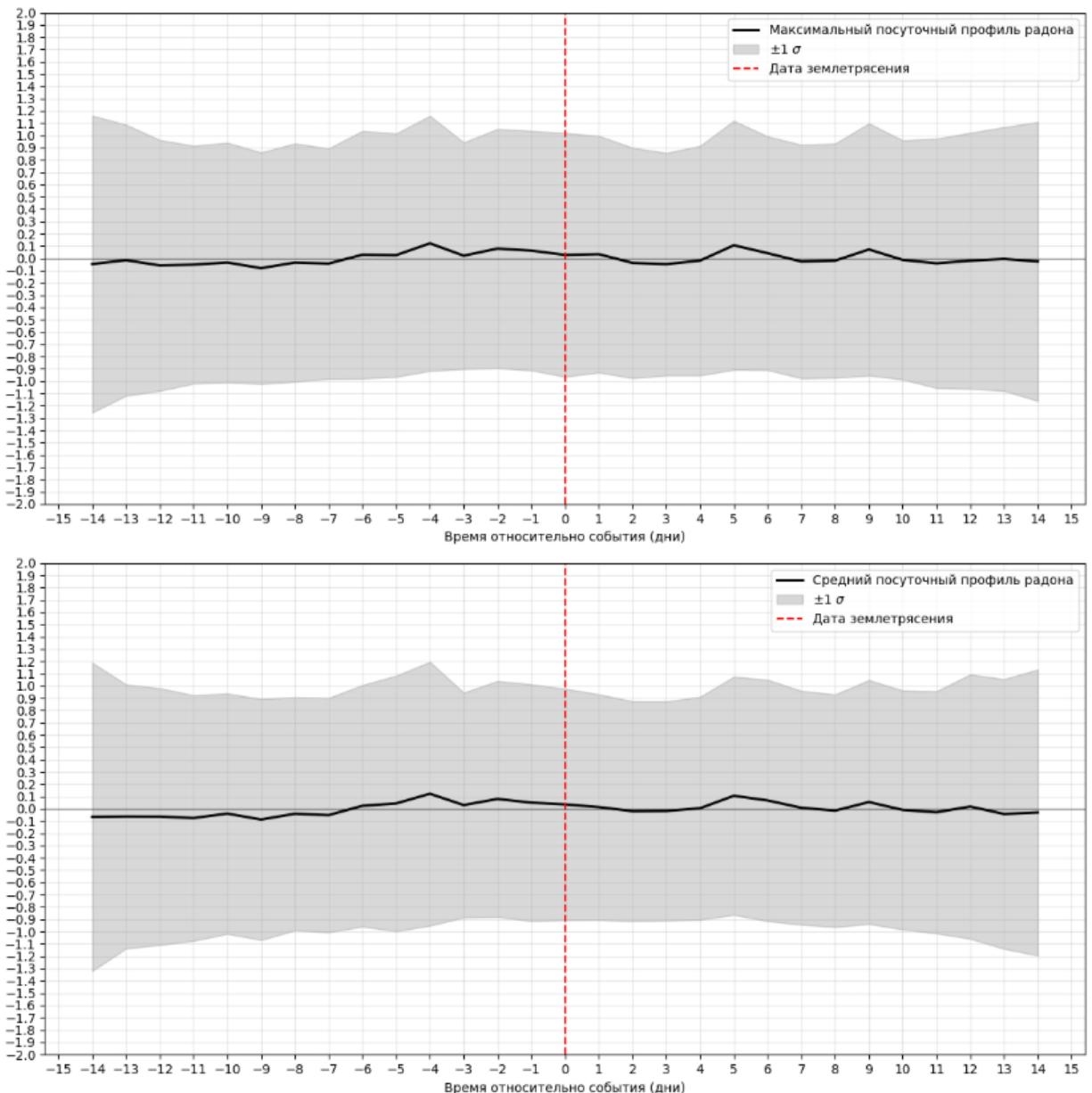


Рис. 14. Средняя форма агрегированного сигнала радона в окне ± 14 вокруг даты землетрясения

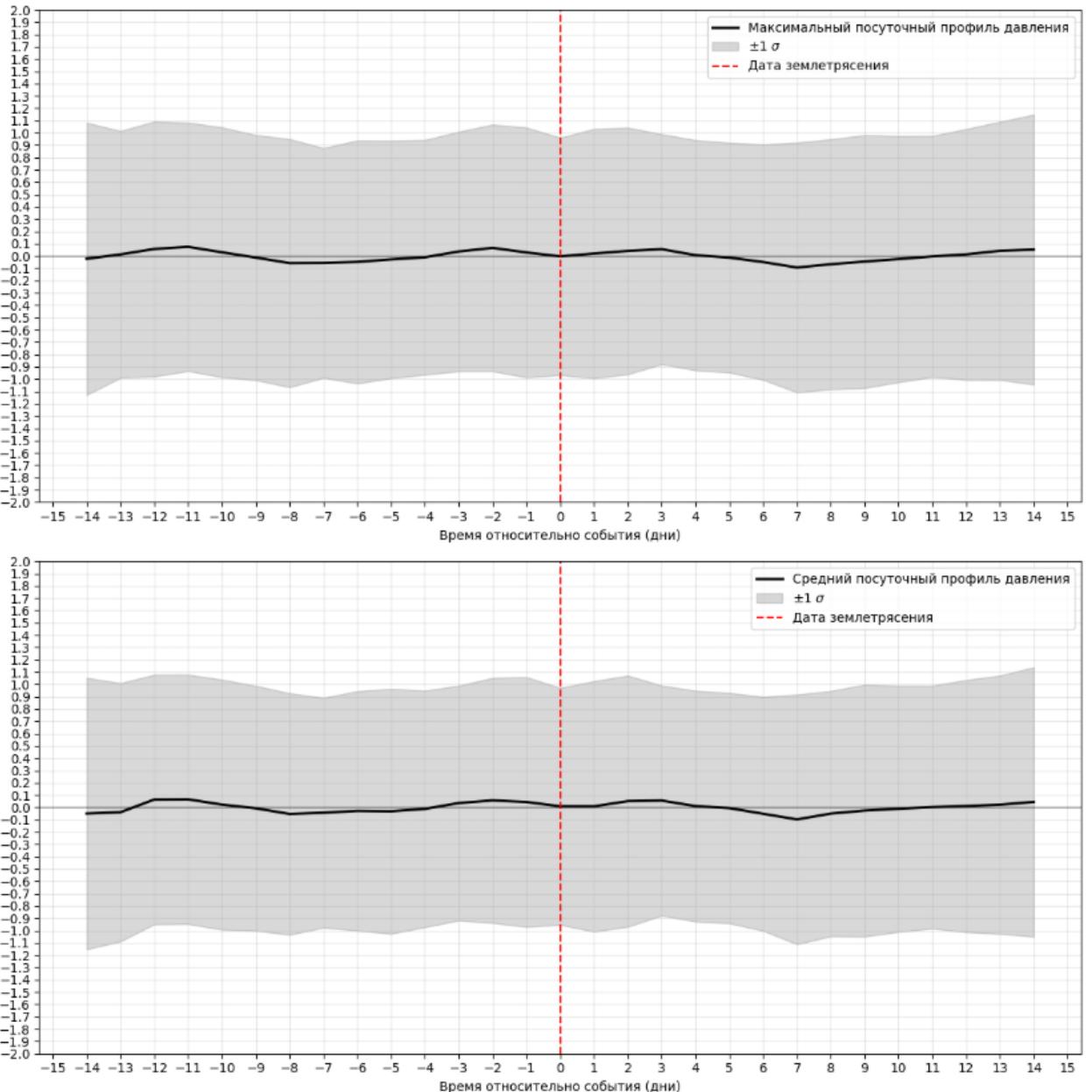


Рис. 15. Средняя форма агрегированного сигнала давления в окне ± 14 вокруг даты землетрясения

С точки зрения машинного обучения, решается задача бинарной классификации в том смысле, что модель должна ответить на вопрос, будет ли в ближайшие k суток землетрясение или нет. Таким образом, формирование целевых меток происходит по следующему принципу:

$$y_t = 1, \text{ если } \exists \text{ дата землетрясения } \in [t + 1, t + k]$$

где k определяет, сколько дней до события будут соответствовать 1; если обозначенное выше условие не выполняется, данным за текущий момент y_t приписывается метка 0.

Итак, основным пространством экспериментов с учётом ранее обсуждённых условностей являлся подбор параметра k и вместе с тем поиск подхода к формированию признаков из обнаруженных паттернов. Выбор интервал около 7 дней и более перед землетрясением приводил сильному дисбалансу в сторону целевого класса. Опытным путём было выявлено, что при $k = 2$ соотношение классов в выборке достигает баланса (1 к 1). Это важно, поскольку при явном выделении мажоритарного класса модель склонна ошибаться в его сторону, так как с точки зрения алгоритма это может быстрее минимизировать ошибку, несмотря на очевидную, с точки зрения человека, проблему такого поведения.

Поскольку паттерн вокруг землетрясения статичен, а при $k = 2$ метка ставится на две даты, предшествующие дате землетрясения, нужно было выработать подход, который при формировании признаков учитывал бы сдвиг за на один день. Основной массив признаков из временных рядов выделялся с помощью скользящих окон, поэтому оставалось корректно подобрать сдвиг назад относительно текущего значения (для которого формируется признак) и ширину окна. Таким образом, в первом базовом решении из исходных временных рядов были созданы признаки:

- Максимум радона (сдвиг – 2, ширина окна – 5);
- Число шагов до максимума радона (сдвиг – 2, ширина окна – 5);
- Максимум давления (сдвиг – 0, ширина окна – 4);
- Число шагов до максимума давления (сдвиг – 0, ширина окна – 4);
- Возрастающий тренд (линейная интерполяция) радона (сдвиг – 3, ширина окна – 2);
- Убывающий тренд радона (сдвиг – 1, ширина окна – 2);
- Тренд давления (сдвиг – 0, ширина окна – 5);
- Отношение возрастающего тренда радона к тренду давления.

Несмотря на то, что признаки формируются на основе выделенных закономерностей, в них практически не содержится информации об относительности значений. То есть, в них отсутствует информация о том, какие значения наблюдаются на более ранних временных отметках относительно текущих (иными словами, почти отсутствует как бы «рекуррентность», которая здесь крайне важна). Учёт этого обстоятельства будет рассмотрен ниже.

В рамках условной второй части экспериментов было задано $k = 1$, то есть каждому землетрясению соответствовала одна дата с меткой 1 перед ним. Поскольку землетрясений больше общего количества дат в наборе данных даже после агрегации по суткам, при $k = 1$ формировался дисбаланс в сторону нулевого класса. Тем не менее, такое решение интерпретируется лучше, чем $k = 2$ и, что важнее, не накладывает ограничений на формирование признаков, которые образуются из-за сдвига целевых значений относительно статичного паттерна. С таким подходом к созданию целевой переменной были выделены следующие признаки:

- Тренд радона (сдвиг – 3, ширина окна – 3);
- Монотонность тренда радона (сдвиг – 3, ширина окна – 3) – отношение дней роста к общему количеству дней в окне;
- Изменение тренда радона – рассчитывается как разница между трендом со сдвигом 3 и шириной окна 3 и трендом со сдвигом 6 и шириной окна 3;
- Находится ли максимум (сдвиг – 0, ширина окна – 7) в окне 2-4 до текущей даты – бинарный признак;
- Выступ пика – рассчитывается как разница между максимальным и медианным значением (сдвиг – 0, ширина окна – 7), отнесённая к межквартильному интервалу за тот же период;
- Спад после пика – разница между пиковым значением и значением за 2 дня до текущего, отнесённая к значению в пике;

- Скорость спада – разница между значением в точке предполагаемого пика и в следующей после этого точке спада;
- Интегральное превышение радона – сначала рассчитывается медиана за 14 дней, затем разница между медианой и текущим значением, после этого с помощью суммирования находится превышение уровня относительно медианного за последние 7 дней;
- Волатильность радона – рассчитывается как отношение среднеквадратичных отклонений за 7 дней и за 14 дней;

Аналогичные признаки с другими окнами и сдвигами рассчитываются для давления. В качестве совместных признаков рассчитываются:

- Совпадение трендов – проверка на равенство знаков трендов радона и давления в одной точке;
- Пик радона раньше пика давления – бинарный признак;
- Отношение интегральных превышений радона и давления.

Все вышеперечисленные признаки рассчитывались по среднесуточным значениям.

2.6. Обучение моделей

В качестве основных метрик использовались precision и recall, для возможности анализа результата рассчитывалась матрица ошибок. В качестве основных моделей были выбраны:

- Логистическая регрессия;
- Метод опорных векторов;
- Градиентный бустинг.

Метрики precision/recall по каждой из моделей для двух подходов, описанных выше, приведены в таблице:

	Логистическая регрессия	Метод опорных векторов	Градиентный бустинг
1	0.481 / 0.882	0.493 / 1.00	0.505 / 0.655
2	0.364 / 0.536	0.373 / 0.00	0.379 / 0.209

В первой серии экспериментов с самыми базовыми признаками при $k = 2$ первые две модели дали надежду на то, что методы на основе деревьев могут заметно улучшить результат, поскольку в задаче важна именно одновременная реализация нескольких паттернов – логическое условие. Результат действительно стал лучше, но не настолько, чтобы имело смысл продолжать работать с гиперпараметрами модели – стало очевидно, что проблема может быть в малом количестве и простоте признаков. Таким образом был сформирован второй подход с $k = 1$, описанный ранее.

Несмотря на, казалось, большую осмысленность, вторая серия экспериментов, вопреки ожиданиям, показала результат даже хуже предыдущего. Если в первом случае модели склонны были делать выбор в сторону класса 1, несмотря на исходный баланс, то во втором случае ситуация противоположная – выбор в сторону класса 0, ставшего мажоритарным, при этом усугубилась проблема различимости классов.

2.7. Поиск проблем и рассмотрение альтернативных подходов

Одним из очевидных возможных недостатков, повлиявших на результат, может быть фундамент решения – агрегация исходных данных по суткам. Безусловно, это сглаживает шум сигнала, но вместе с тем и лишает временной ряд паттернов, которые изначально были обнаружены на масштабе часов. Имеет смысл попробовать, если не оставлять в первоначальном виде, агрегировать данные по меньшим окнам и переходить к менее «грубому» масштабу. Например, не по 24 часа, а по 12 или 6. Это может способствовать сохранению ключевых значений временного ряда и при этом сократить шум.

Кроме того, в процессе анализа акцент делался на исходные данные, но работа с извлечёнными из них признаками практически не осуществлялась. Следует перед обучением модели проанализировать результаты извлечения признаков более обстоятельно: группировка по целевой метке и расчёт статистик, попарная визуализация признаков для выявления взаимосвязей, метод главных компонент и пр.

Также практически все признаки рассчитывались на основе суточного среднего. То есть большая часть значений, полученных в ходе агрегации, не была использована, так как модели обучались только на сформированных признаках.

Имеет смысл в качестве альтернативы попробовать небольшую рекуррентную нейронную сеть, поскольку они спроектированы как раз для задач, где важен контекст. В данной работе была попытка контекст учесть через формирование определённых признаков на основе предыдущих значений.

ЗАКЛЮЧЕНИЕ

В ходе анализа были выявлены и подтверждены ключевые идеи, лежащие в основе исследования. Несмотря на неудовлетворительный итоговый результат, поскольку были использованы лишь несколько из множества других возможных подходов, остаётся большой простор для дальнейших экспериментов и работы с данными.

Список литературы

1. Изучение вариаций низкоэнергетического фона с помощью подземного эксперимента LVD" Известия РАН Серия физическая, 2019, том 83, № 5, с. 673–675.
2. Агафонова Н.Ю. и др. "Предсказание землетрясений с помощью нейтринного телескопа LVD (Италия) под землей". Вестник РУДН. стр. 81-86, LVII Всероссийская конференция по проблемам динамики, физики частиц, физики плазмы и оптоэлектроники = LVII All-Russia Conference on Problems in Dynamics, Particle Physics, Plasma Physics and Optoelectronics: материалы конференции. Москва, РУДН, 17–21 мая 2021 г. – Москва: РУДН, 2021. – 270 с. (ISBN 978-5-209-11093-4) .
3. Агафонова Н.Ю. и др. "Отклик детектора LVD на землетрясения в центральной Италии". Известия РАН. Серия физическая Т.85, N11, 1661-1665 (2021) DOI: 10.31857/S036767652111003X [Перевод - N.Yu. Agafonova, V.V. Ashikhmin, E.A. Dobrynina, R.I. Enikeev, A.S. Malgin, O.G. Ryazhskaya, N.A. Filimonova, I.R. Shakyrjanova, V.F. Yakushev, "Responses of the LVD Detector to Earthquakes in Central Italy", Bull. of the Rus. Acad. of Sci.: Physics, 2021, Vol. 85, No.11, pp.1320–1324, DOI: 10.3103/S1062873821110034].
4. Н.Ю. Агафонова, Е. А. Добрынина, Н. А. Филимонова (от имени коллаборации LVD) "Регистрация распадов радона с помощью нейтринного детектора LVD". ВМУ. Серия 3. ФИЗИКА. АСТРОНОМИЯ. 78(1), 2310202 (2023), 4 стр. (труды конференции Ядро-2022).
5. Агафонова Н.Ю., Добрынина Е.А., Филимонова Н.А., Шакирьянова И.Р. "Отклик детектора LVD на сильные землетрясения в Средиземноморском регионе".
6. <https://sdis.emsd.ru/info/earthquakes/catalogue.php>
7. <https://www.ngdc.noaa.gov/hazel/view/hazards/earthquake/search>
8. <https://terremoti.ingv.it/>
9. <http://www.gsras.ru/new/wf/>

10. <https://geopy.readthedocs.io/en/stable/index.html#module-geopy.distance>