

Аннотация

Выпускная квалификационная работа посвящена разработке модуля СППР RECASP для прогнозирования землетрясений в Байкальской рифтовой зоне. Реализована двухуровневая схема (KDE + адаптированная ETAS) с экспертной интерпретацией Experta и хранением геоданных в PostGIS, обеспечивающая near-real-time визуализацию рисков. Ретроспективная проверка (1960–2024 гг.) повысила выявляемость событий до 18,5 % и сократила время подготовки карт с часов до секунд.

Ключевые слова: RECASP, СППР, сейсмический прогноз, KDE, ETAS, Experta, PostGIS.

Содержание

Введение.....	7
1. Теоретическая часть.....	10
1.1 Системы поддержки принятия решений (СППР) при природных катастрофах.....	10
1.1.1 Архитектурная модель СППР.....	10
1.1.2 Типология СППР.....	11
1.1.3 Детальный жизненный цикл.....	11
1.1.4 Кейс-стади внедрённых СППР.....	13
1.1.5 Расширенные требования к СППР при сейсмической угрозе.....	15
1.1.6 Значение интеграции СППР в RECASP.....	15
1.2 Современные подходы к прогнозированию землетрясений.....	16
1.3 Ограничения существующих моделей прогнозирования.....	17
1.3.1 Физико-геологическая сложность процесса.....	17
1.3.2 Дефицит высококачественных данных.....	17
1.3.3 Ограниченность верификационных критериев.....	17
1.3.4 Специфические ограничения ИИ-подходов.....	18
1.4 Применение нейросетей и методов машинного обучения в геофизике.....	18
1.4.1 Гибридные и physics-informed сети.....	19
1.4.2 Трансферное и федеративное обучение.....	19
1.4.3 Ограничения и перспективы.....	19
1.5 Архитектура и назначение системы RECASP.....	19
1.5.1 Концептуальная трёхуровневая архитектура.....	20
1.5.2 Ключевые функциональные модули.....	20
1.6 Примеры существующих решений и моделей для прогнозирования.....	21
1.6.1 Глобальные каталоги и веб-порталы.....	21
1.6.2 Национальные и региональные базы.....	21
1.6.3 Научно-прикладные программные пакеты.....	21
1.6.4 Выводы для RECASP.....	22
1.7 Выводы по главе.....	22
2. Методология исследования.....	24
2.1. Принципы и подходы сейсмического прогнозирования.....	24
2.2. Ретроспективный анализ как метод проверки моделей.....	25
2.3. Структура и формат каталога землетрясений.....	25
2.3.1 Базовая схема записи.....	26
2.3.2 Контроль качества и оценка полноты.....	26
2.3.3 Интероперабельность и экспорт.....	26
2.4. Предварительная обработка данных.....	26
2.5. Статистический анализ сейсмической активности.....	29
2.6. Алгоритмы прогнозирования, применяемые в исследовании.....	30
2.7. Построение карты сейсмического риска.....	31
2.8. Ретроспективная валидация модели (метрики и процедуры).....	32
2.9. Инструментальные средства реализации.....	34
2.10 Выводы по главе.....	36
3 Практическая часть.....	38

3.1 Цели и задачи прикладной реализации	38
3.2 Архитектура программного комплекса.....	40
3.3 Подготовка исходных данных.....	43
3.4 Долгосрочный KDE-прогноз.....	45
3.5 Краткосрочная корректировка ETAS.....	47
3.6 Разбиение на регионы.....	49
3.7 Экспертные системы Experta	50
3.7.1 Условия запуска системы	50
3.7.2 Создание и структура фактов.....	51
3.7.3 Правила подсистемы ZoneExpertSystem.....	51
3.7.4 Интерпретация правил из zone_engine.py.....	52
3.7.5 Правила подсистемы CoordinateExpertSystem	52
3.7.6 Вывод результатов.....	53
3.7.7 Внутренние механизмы библиотеки Experta	53
3.7.8 Особенности и преимущества архитектуры	54
3.8 Подсчёт потерь	55
3.9 Интерфейс визуализации.....	58
3.10 Сценарий запуска	59
3.11 Среда исполнения и развёртывание	60
3.12 Вывод по главе.....	60
4 Анализ результатов	62
4.1 Выбор контрольного периода и тестовой выборки	62
4.2 Качественная интерпретация карты риска	63
4.3 Количественная оценка прогноза (CSEP-метрики).....	65
4.4 Результаты работы экспертных систем.....	67
4.5 Влияние параметров модели (сенситивити-анализ).....	69
4.6 Ограничения и дальнейшие исследования	70
4.7 Вывод по главе.....	72
Заключение.....	74
Список использованных источников.....	76

Введение

Землетрясения относятся к самым разрушительным природным катаклизмам, способным приводить к огромным человеческим жертвам и материальному ущербу. Во многих сейсмоактивных регионах мира актуальна задача повышения сейсмобезопасности территорий за счёт совершенствования методов прогнозирования землетрясений. Краткосрочное предсказание конкретного землетрясения (с указанием времени, места и магнитуды) до сих пор представляет значительную научную проблему: ни один из существующих методов не позволяет с высокой точностью предсказать событие вплоть до дней или месяцев [1]. Более того, попытки введения оперативных прогнозов сталкиваются с риском большого числа ложных тревог, что может приводить к неоправданным социально-экономическим потерям и снижению доверия общества к прогнозам [2]. В итоге во многих случаях акцент смещается с точного прогноза на меры смягчения последствий землетрясений (усиление сейсмостойкости сооружений, подготовка систем оповещения и т.п.). Тем не менее, научные исследования в области предвестников землетрясений и прогнозных моделей активно продолжаются ввиду их высокой социальной значимости и потенциальной пользы [3]. Развитие информационных технологий и появление новых данных мониторинга дают шанс на прогресс в этой сложной области.

Одним из перспективных направлений современных исследований является применение методов искусственного интеллекта (ИИ) и машинного обучения для анализа предвестников и статистических закономерностей сейсмического процесса. Нейронные сети способны выявлять скрытые нелинейные зависимости в больших массивах данных, что делает их привлекательными для задачи прогнозирования землетрясений. В последние годы появились обнадеживающие результаты: например, нейросетевые модели, анализирующие эволюцию напряжений в литосфере, демонстрируют более высокую точность долгосрочных оценок по сравнению с классическими эмпирическими критериями [4]. Кроме того, сочетание данных различных типов (сейсмологических, геодезических, геофизических) и их обработка методами ИИ позволяют формировать комплексные прогнозные решения. Таким образом, актуальность применения искусственного интеллекта в задачах сейсмического прогнозирования обусловлена потребностью преодолеть ограничения традиционных моделей и использовать весь объем доступной информации.

Для эффективного использования результатов прогнозных моделей необходимо интегрировать их в систему поддержки принятия решений (СППР). СППР предназначены для помощи экспертам и ответственным органам в выработке оптимальной стратегии действий на основе совокупности разнородных данных и прогнозов. В контексте сейсмической опасности СППР может предоставить инструментарий для оценки рисков, сценариев последствий и рекомендаций по своевременному оповещению и эвакуации населения. Применение СППР обеспечивает системный и прозрачный подход к анализу ситуации, позволяя визуализировать данные мониторинга, сравнивать различные прогнозные модели и ранжировать альтернативные решения по степени

эффективности [5]. Таким образом, сочетание методов искусственного интеллекта с возможностями СППР представляется особенно актуальным для задач прогнозирования землетрясений и снижения рисков чрезвычайных ситуаций.

В настоящей выпускной квалификационной работе рассматривается разработка модуля СППР в системе RECASP для прогнозирования землетрясений. RECASP (Reengineering of "Computer Analysis of the Seismic Process") – это информационная система, предназначенная для хранения, обработки и визуализации сейсмических данных. Она создана на основе реинжиниринга программного комплекса "Компьютерный анализ сейсмического процесса" (КАСП), разработанного в Институте земной коры СО РАН. КАСП исторически применялся сейсмологами для анализа каталогов землетрясений, отображения эпицентров на карте и исследования сейсмического режима регионов. Новая система RECASP сохранила функциональные возможности КАСП (работу с каталогами землетрясений и разломов, построение графиков, наложение геологических слоёв на карту и др.), но получила современную архитектуру и интерфейс, а также возможность обработки больших объёмов данных с высокой скоростью. Благодаря этому RECASP создаёт платформу, на базе которой может быть реализован модуль прогнозирования, использующий алгоритмы ИИ и машинного обучения. Разработка данного модуля расширит функциональность системы: от постанализа и визуализации данных к поддержке принятия решений на основе прогнозов возможных сильных землетрясений.

Цель выпускной квалификационной работы состоит в проектировании модуля системы поддержки принятия решений в составе RECASP, предназначенного для прогнозирования землетрясений. Для достижения поставленной цели необходимо решить следующие задачи:

1. Провести анализ предметной области: изучить существующие системы поддержки принятия решений, применяемые при природных катастрофах, и современные подходы к прогнозированию землетрясений, включая выявление ограничений существующих моделей.

2. Исследовать возможности применения нейросетей и методов машинного обучения в сейсмологии и геофизике, обобщить успешные примеры их использования для прогнозных задач.

3. Проанализировать архитектуру и функциональные возможности системы RECASP, определить места интеграции модуля прогнозирования в общую систему.

4. Сформировать требования к модулю СППР для прогнозирования землетрясений и разработать концептуальную архитектуру данного модуля на базе исследований.

5. Предложить методику прогнозирования (алгоритмы или модель) для внедрения в модуле ReCASP.

В рамках введения обоснована актуальность применения ИИ и СППР для задач сейсмопрогнозирования и определена цель работы, сформулированы задачи разработки модуля. Далее в разделе анализа предметной области

приводится обзор существующих решений и методов, что послужит теоретической основой для проектирования нового модуля.

1. Теоретическая часть

1.1 Системы поддержки принятия решений (СППР) при природных катастрофах

Стихийные бедствия характеризуются высокой степенью неопределённости, быстрым развитием событий и серьёзными социально-экономическими последствиями. В таких условиях традиционные, исключительно экспертные методы управления рисками нередко оказываются недостаточно оперативными и прозрачными. СППР заполняют этот пробел, предоставляя интегрированную среду, где данные, модели и экспертные правила превращаются в чёткие рекомендации по действиям для различных уровней принятия решений — от диспетчеров оперативных служб до региональных органов власти [6].

1.1.1 Архитектурная модель СППР

Классическая архитектура СППР включает четыре взаимосвязанных подсистемы [6].

1. Подсистема управления данными. Она обеспечивает многоуровневое хранилище, в котором совмещаются потоковые данные реального времени (телеметрия сейсмостанций, GNSS-измерения деформаций, сигналы от датчиков электромагнитного поля) и исторические архивы. На этапе ETL-процедур данные проходят автоматическое исправление выбросов, временную синхронизацию и географическую привязку. Благодаря этому аналитические модули получают «чистый» и согласованный набор сведений, что особенно важно при агрегации разнородных источников.

2. Модельная подсистема. Здесь развёрнуты алгоритмы разной природы — от эмпирических статистических критериев (например, индекс сейсмичности b -value) до физико-математических симуляторов распространения волн и гибридных нейросетевых моделей. Пользователь может запускать параллельные сценарии, сравнивая, например, классический расчёт ожидаемого интенсивитета с прогнозами, основанными на глубинных сетях. Важная особенность — модуль калибровки, позволяющий адаптировать параметры модели к локальным геологическим условиям региона.

3. Подсистема знаний. В отличие от обычных БД, этот компонент хранит формализованные правила реагирования, одобренные органами гражданской защиты, и прецедентные кейсы прошлых катастроф. При наступлении событий, удовлетворяющих определённому правилу (скажем, превышение ускорения грунта $> 0,1 \text{ g}$ в городской черте), система автоматически формирует предписание: «остановить движение скоростного транспорта», «закрыть газораспределительные станции» и т. д. [7].

4. Интерфейс пользователя. Современные СППР предоставляют веб-панели и мобильные приложения с адаптивным дизайном. Карта опасности обновляется в режиме near-real-time: точки сейсмостанций меняют цвет по уровню пикового ускорения, а прогнозное «окно» вероятности афтершоков отображается прозрачным контуром. Интерфейс поддерживает уровни доступа: исследователь видит сырые данные и графики, оператор — лишь агрегированные

индексы и чек-лист действий, что предотвращает перегрузку информацией в стрессовой обстановке.

Пример взаимодействия компонентов. После регистрации толчка ShakeAlert автоматически оценивает источник, передаёт параметры в модельную подсистему, моделирующую интенсивитет на поверхности. Пороговые значения интенсивитета, хранящиеся в подсистеме знаний, активируют готовые сценарии (например, отключение лифтов). Оператор видит на панели таймер обратного отсчёта до прихода S-волн и список предписаний [8].

1.1.2 Типология СППР

Хотя базовая архитектура едина, акценты могут смещаться [6].

1. Коммуникационные СППР ориентированы на совместную работу распределённых команд (напр., диспетчерские оперативного штаба и муниципальные службы).

2. Данные-ориентированные СППР применяются в регионах с богатой исторической статистикой: крупные каталоги позволяют строить надежные вероятностные модели афтершоков.

3. Модели-ориентированные СППР востребованы в сейсмоактивных зонах со сложной геологией, где для расчёта интенсивитета требуется численное моделирование трёхмерной структуры коры.

4. Знания-ориентированные СППР незаменимы там, где важны нормативные ограничения: например, автоматическая проверка, что сценарий эвакуации соблюдает пропускную способность дорог.

В реальных системах, таких как ARISTOTLE-ENHSP, перечисленные типы комбинируются: коммуникационный портал для обмена оценками, массив исторических данных EMSC, физические модели распространения цунами и банк нормативных алгоритмов расчёта ущерба [10].

1.1.3 Детальный жизненный цикл

Практика показывает, что эффективность СППР определяется не только точностью отдельных моделей, но и тем, насколько устойчиво система «проходит» полный цикл управления катастрофой — от подготовки до постанализа [7]. Описание жизненного цикла представлено в таблице .1.1.

Таблица 1.1 – Жизненный цикл СППР

Фаза	Цели и ключевые действия СППР	Пример функциональности
1	2	3
Подготовка (Preparedness)	1) формирование и валидация базы знаний; 2) калибровка моделей на исторических данных; 3) тренировочные сценарии для операторов	«Песочница» RECASP загружает каталоги прошлых землетрясений и позволяет отрабатывать виртуальные учения: система показывает, как изменятся индексы риска при условном событии Mw 7.0 в соседнем разломе [6].
Мониторинг (Detection)	1) непрерывный приём потоковых данных; 2) автоматический детект аномальных параметров (ML-детекторы); 3) визуальный дашборд текущей сейсмичности	Алгоритм STA/LTA в реальном времени подкрашивает станции на карте, когда отношение превышает порог, а RNN-модуль оценивает, к какому типу события относится сигнал [7].
Прогноз (Assessment & Forecast)	1) запуск ансамбля моделей (статистической, физической, нейросетевой); 2) построение сценариев вероятности события и карты ожидаемых интенсивитетов; 3) ранжирование сценариев по тяжести последствий	После выявления кластера форшоков система автоматически считает вероятность основного толчка в окне 72 ч и накладывает зоны PGA 0,1 g, 0,2 g, 0,3 g, чтобы оценить потенциальный ущерб [8].
Предупреждение (Early Warning)	1) интеграция с алгоритмами EEW; 2) оценка времени прихода S-волн и пикового ускорения; 3) рассылка CAP-совместимых оповещений	ShakeAlert передаёт в СППР параметры источника; RECASP отображает таймер «S-волны через 24 с» и триггерит push-сообщения объектам критической инфраструктуры [8].

Продолжение таблицы 1.1

1	2	3
Реагирование (Response)	1) генерация перечня предписаний на основе пороговых значений; 2) подсказки по распределению ресурсов; 3) отслеживание выполнения мер	При $PGA \geq 0,2$ g подсистема знаний мгновенно формирует задачи: «остановить лифты», «закрыть газовые клапаны», «проверить мосты категорий А–В». Оператор отмечает завершение задач; информация фиксируется в журнале [9].
Восстановление (Recovery)	1) динамическая оценка ущерба и потребностей; 2) план-график восстановления сетей; 3) поддержка принятия решений о приоритетах ремонта	Система моделирует доступность дорог после сходов оползней и предлагает маршруты для доставки оборудования в больницы с учётом объёма грузов и пропускной способности дорог [10].

Дополнительным горизонтальным процессом является аудит кибербезопасности: все фазы должны сопровождаться контролем целостности данных и резервированием каналов связи, особенно на этапах раннего предупреждения и реагирования, когда задержка или нарушение доставки сообщений приводит к значительному росту ущерба [7].

Таким образом, расширенное рассмотрение жизненного цикла подчёркивает: СППР эффективна лишь при сквозной интеграции всех этапов — от подготовки до непрерывного самообучения системы. Для RECASP это означает необходимость обеспечить не только точные модели прогнозирования, но и механизмы «замыкания цикла» через хранение журналов решений, автоматический пересчёт показателей достоверности и регулярное обновление базы знаний.

1.1.4 Кейс-стади внедрённых СППР

Опыт действующих систем показывает, что архитектурные решения и организационные регламенты варьируются в зависимости от географических, техногенных и социальных условий, но все успешные проекты демонстрируют сходные ключевые метрики: минимальное время оповещения, прогнозная точность интенсивитета, устойчивость каналов связи и формализованный протокол взаимодействия с пользователями [8-11]. В таблице сопоставлены основные показатели четырёх референсных СППР.

Таблица 1.2 – Сравнение существующих ССПР

Система	Покрывае- мая терри- тория	Средняя задержка оповещения	Погреш- ность расчёта интен- сивитета (MMI)	Каналы доставки	Норматив- ный протокол
ShakeAlert (США)	3 штата за- падного побережья	3–5 с	$\pm 0,3$ балла	Push-уве- домле- ния, EAS-ра- дио, siren	CAP 1.2 [8]
UrEDAS (Япония)	2 000 км ж/д полотна	≤ 2 с	не использу- ется (порог 40 gal)	Локаль- ная шина CAN-rail	JR East EEW-Spec [9]
ARISTOTLE- ENHSP (EC)	27 гос-в ЕС	≤ 30 мин отчёт	$\pm 0,5$ балла	ERCC портал, e-mail	INSPIRE, CAP [10]
R-ISEE (Камчатка)	$\sim 400\,000$ км ²	240 ч (про- гноз)	Индекс консоли- дации R- code 0,7	Web-па- нель, SMS	ГО и ЧС РФ [11]

На основе таблицы 1.2 можно сделать следующие выводы:

1. Разноуровневые горизонты оповещения. ShakeAlert и UrEDAS решают задачу секундного раннего предупреждения (EEW), где критичен каждый дополнительный фрейм задержки. ARISTOTLE-ENHSP специализируется на сценариях чрезвычайных гуманитарных операций (десятки минут), а R-ISEE — на среднесрочном прогнозе (сутки–неделя). Такой спектр горизонтов подчёркивает необходимость модульного подхода: разные классы решений могут быть интегрированы в единую СППР, обслуживая разные фазы жизненного цикла [8-11].

2. Инфраструктурная избыточность. Успешные системы используют принцип N+1: дублирование центров обработки (ShakeAlert — Денвер/Менло-Парк), резервные спутниковые каналы UrEDAS, распределённые экспертные хабы ARISTOTLE-ENHSP. Для RECASP это означает необходимость как минимум двух гео-резервных узлов с автоматическим fail-over-переключением [8-10].

3. Стандартизованный формат оповещений. Все проекты переходят на CAP 1.2 или совместимые JSON-схемы, что обеспечивает интероперабельность с мобильными приложениями, радио-вещателями и системами «умного» города. При разработке модуля RECASP логично сразу заложить генератор CAP-сообщений с цифровой подписью ГОСТ/PKI [12].

4. Метрики самоконтроля. Каждая система ведёт непрерывный аудит: ShakeAlert — ежедневный self-test всех станций; UrEDAS — ежеквартальные натурные испытания тормозных цепей; ARISTOTLE-ENHSP — пост-фактум отчёт о точности сценария; R-ISEE — пересчёт индекса P-code после каждого события $M_w \geq 4,5$. В RECASP следует реализовать автоматическую выгрузку журнала «Forecast vs Observed» и скрипты пересчёта Precision/Recall для ансамбля моделей [13].

1.1.5 Расширенные требования к СППР при сейсмической угрозе

Опыт перечисленных систем показывает, что высокая точность моделей должна сочетаться с жёсткими требованиями к инфраструктуре и организации данных [13-14].

1. Отказо- и киберустойчивость. ShakeAlert® поддерживает «горячий» резерв в другом штате, а телеметрия подписывается цифровым сертификатом; при обрыве одного канала данные поступают через спутниковый VPN [12].

2. Низкая задержка. UrEDAS ставит предел 2 с от регистрации ускорения до торможения поезда; задержка тестируется ежеквартально в ходе регламентных испытаний [9, 13].

3. Геопривязка с оценкой риска. ARISTOTLE-ENHSP автоматически конвертирует расчёт PGA в прогноз потерь, используя функции хрупкости для типовой застройки государств ЕС [10, 13].

4. Регламентированная визуализация. Интерфейс ShakeAlert® показывает только таймер прихода S-волн, расчётный MMI и список действий, снижая когнитивную нагрузку [12].

5. Замкнутый цикл обучения. СППР должна автоматически сверять прогноз и факт, обновляя веса моделей; для этого в RECASP следует реализовать версионное хранилище журналов и периодический пересчёт метрик точности [13].

1.1.6 Значение интеграции СППР в RECASP

Модульная архитектура RECASP (PostgreSQL + PostGIS, REST-API-слой, ГИС-интерфейс) уже удовлетворяет большинству инфраструктурных требований. Для превращения её в полноценную СППР необходимо:

1. Поточный коннектор к региональной сети сейсмостанций (miniSEED, 1 с).

2. Микросервис раннего предупреждения, выполняющий P-волновую инверсию и передающий параметры очага.

3. Ансамблевый модуль прогнозов (статистика Gutenberg–Richter + RNN + квазидинамическая модель разлома).

4. Слой бизнес-правил, трансформирующий PGA/MMI в предписания для транспорта, энергетики и ЖКХ.

5. Компонент пост-анализа, автоматически пересчитывающий Precision / Recall и обновляющий базу знаний.

Тем самым RECASP перейдёт от пост-фактум аналитики к проактивному управлению сейсмическими рисками и замкнутому циклу «данные → прогноз → действие → обучение».

1.2 Современные подходы к прогнозированию землетрясений

Современные методы прогнозирования землетрясений условно делят на долгосрочные, среднесрочные и краткосрочные, — каждый класс решает собственную прикладную задачу и опирается на разные данные и модели [1, 14].

Долгосрочный прогноз (горизонт 10–50 лет) базируется на статистическом анализе исторических каталогов и представляется через карты сейсмической опасности, используемые при строительном нормировании. Ключевые инструменты — закономерность Гутенберга–Рихтера и регионо-специфичные модели повторяемости событий [1].

Среднесрочный прогноз (месяцы–годы) пытается выделять «окна повышенного риска» в конкретных зонах. Наиболее известный инструмент — алгоритм M8 и его модификации, разработанные группой Гельфанда – Кейлиса-Борока. Алгоритм обрабатывает временные ряды каталога, вычисляет комплекс индексов сейсмичности и формирует предупреждение, если совокупный показатель превышает порог [5, 15]. Валидационные кампании (Калифорния, Камчатка, Средиземноморье) показали статистически значимое превышение точности над случайной выборкой, хотя доля ложных тревог остаётся высокой [15].

Краткосрочный прогноз (дни–часы) — самый востребованный и наиболее спорный. Международные экспертизы ICEF и USGS сходятся во мнении, что пока не существует общепризнанного метода, позволяющего с приемлемой точностью предсказать время, место и магнитуду будущего события [14, 16]. Тем не менее, ведутся активные исследования предвестников:

1. Геохимические аномалии (радон в грунтовых водах, гелий).
2. Электромагнитные всплески в VLF/LF диапазоне.
3. Деформационные сигналы GNSS и InSAR.
4. Изменения ионосферных TEC-карт.
5. «Финансовая» статистика форшоковой активности.

Ретроспективные исследования фиксируют совпадения отдельных феноменов с крупными землетрясениями (классический пример — успешная эвакуация Хайчэн, 1975 г.), но воспроизводимость остаётся проблемой [2, 14].

Появление больших массивов высокочастотной телеметрии стимулировало использование нейросетей. Глубокая модель PEGSNet анализирует гравитационные «тени» (Prompt Elasto-Gravity Signals) и определяет параметры землетрясения раньше, чем регистрируются Р-волны, что открывает перспективу сверхраннего предупреждения [17]. Другой пример — глубокие сверточные сети, обученные на распределениях статического напряжения: они дают более точные карты афтершоков, чем критерий Кулона [4]. Широко распространяется концепция операционного прогнозирования афтершоков (OAF): USGS публикует вероятностные сценарии спада активности в первые недели после сильного толчка, комбинируя статистические этапные модели и байесовский апостериорный пересчёт по мерам поступления данных [13].

Ограничения современных подходов:

1. Нелинейность и многомасштабность процессов: ни одна физическая модель пока не охватывает весь спектр пространственных и временных масштабов подготовки разрыва [3].

2. Дефицит достоверных данных: сильные события редки, инструментальные записи короче характерного цикла повторяемости, что осложняет статистическую проверку гипотез [3].

3. Верификация прогноза: для оценки качества нужны десятки «сбывшихся/несбывшихся» предсказаний, что требует десятилетий наблюдений [16].

4. Переобучение ИИ-моделей: нейросети могут «запомнить» шум вместо физически осмысленных закономерностей, поэтому внедряются гибридные архитектуры с вшитыми физическими ограничениями [4].

Таким образом, текущий консенсус звучит парадоксально: «точного» краткосрочного прогноза нет, но вероятностный прогноз и оперативная оценка афтершоковых рисков — уже полезный инструмент управления». Именно на вероятностные методы (OAF, M8-подобные индексы, гибридные ИИ-модели) и ориентируется проектирование модуля СППР для RECASP.

1.3 Ограничения существующих моделей прогнозирования

Несмотря на очевидный прогресс, ни одна из существующих концепций — статистическая, физико-математическая или нейросетевая — ещё не позволяет сформировать универсальный надёжный прогноз сильного землетрясения «время-место-магнитуда». Ключевые барьеры можно сгруппировать в четыре блока.

1.3.1 Физико-геологическая сложность процесса

Разломы представляют собой разветвлённые зоны с многомасштабной фрактурной сетью; накопление и высвобождение напряжений носит нелинейный хаотический характер. Даже продвинутые квазидинамические численные модели упрощают реальную структуру коры, не учитывая микротрещины, флюидизацию и фазовые переходы [18]. В результате детерминистические расчёты часто расходятся с наблюдаемой механикой очага.

1.3.2 Дефицит высококачественных данных

Инструментальные каталоги охватывают ~120 лет, тогда как характерный цикл повторяемости крупнейших событий ($M_w \geq 8,5$) может превышать 300–500 лет [3]. Каталоги до XX века основаны на макросейсмических оценках с большой погрешностью, а современные сети неравномерно покрывают сейсмоактивные регионы (океанические зоны, горные пояса). Эта «редкость выборки» делает статистическую валидацию моделей крайне медленной [14].

1.3.3 Ограниченность верификационных критериев

Для строгой оценки качества прогноза требуется накопить десятки «сбывшихся/несбывшихся» предупреждений, что занимает десятилетия [16]. В практике нередко используется ретроспективное тестирование («back testing»), однако эффект подгонки к прошлым данным часто завышает реальную точность. USGS OAF указывает, что любая оперативная система должна публиковать

вероятности событий с указанием апостериорной ошибки, а затем архивировать фактический исход для последующего аудита [13].

1.3.4 Специфические ограничения ИИ-подходов

Нейросетевые модели склонны переобучаться на шум, если обучающая выборка невелика или несбалансирована. В работе Mousavi и др. отмечается, что без встраивания физических ограничений (например, энергии сейсмических волн или максимального допустимого градиента напряжений) глубокие сети воспроизводят локальные корреляции, не обладающие прогностической силой на других регионах [19]. Поэтому развивается направление *physics-informed neural networks*, где ключевые физические уравнения включаются в функцию потерь, уменьшая риск переобучения.

Обобщение. Описанные ограничения объясняют, почему международные комиссии (ICEF, USGS) рекомендуют переходить от детерминистских к вероятностным и сценарным прогнозам: даже если точное «если-то» событие недостижимо, оценки «вероятность $M_w \geq 7,5$ в следующие 72 ч — 5 %» уже полезны для принятия решений о повышении готовности систем гражданской защиты.

1.4 Применение нейросетей и методов машинного обучения в геофизике

Цифровая революция в сейсмологии началась с перехода от аналоговых к высокочастотным цифровым станциям и привела к взрывному росту объёма данных. Накопленные петабайты сейсмограмм, GPS-деформаций и спутниковых изображений создают естественную среду для методов искусственного интеллекта (ИИ). Современные приложения можно условно разделить на три класса [19]:

1. Автоматическая обработка сигналов. Глубокие сверточные сети PhaseNet научились определять приходы Р- и S-волн на шумных сейсмограммах с точностью, сопоставимой с экспертной разметкой, и обрабатывают годовой архив сети Южной Калифорнии за часы вместо недель [20]. Подобные детекторы формируют «тонкий» каталог малых событий, необходимый для анализа форшоковой активности.

2. Экспресс-характеристика источника. Модели PEGSNet анализируют мгновенные гравитационные «тени» (Prompt Elasto-Gravity Signals) и восстанавливают моментное тензорное решение быстрее прихода Р-волн, выигрывая драгоценные секунды для раннего предупреждения [17]. Глубокая сеть DeepShake, обученная на миллиардах синтетических комбо-трасс, в среднем на 30 % точнее классических скелетных алгоритмов в оценке магнитуды в первые 3 с после регистрации события [21].

3. Вероятностное прогнозирование. Сверточная сеть, обученная на распределении статических напряжений, предсказывает афтершоковые зоны с $AUC-ROC \approx 0,86$ и превосходит критерий Кулона на 14 % по точности локализации [4]. В оперативной среде USGS Operational Aftershock Forecast (OAF) комбинирует такие ИИ-оценки с байесовским апостериорным пересчётом,

публикуя карты вероятности афтершоков в первые часы после главного толчка [13].

1.4.1 Гибридные и physics-informed сети

Ключевой вызов нейросетевых подходов — переобучение на ограниченной выборке редких крупных событий. Решение видят в physics-informed neural networks (PINN): физические законы (закон сохранения энергии, уравнения упругости) встраиваются в функцию потерь, заставляя сеть уважать реалистичные диапазоны параметров [19]. Такой подход снижает потребность в колоссальном наборе размеченных примеров и повышает интерпретируемость результатов.

1.4.2 Трансферное и федеративное обучение

Геологические условия регионов различаются, поэтому модели, натренированные на Калифорнии, не всегда корректно работают, например, на Камчатке. Трансферное обучение (fine-tuning) позволяет адаптировать веса сети к новому региону, используя небольшую локальную выборку. В эксперименте с российской сетью «ФЭДОС» перенос PhaseNet уменьшил среднеквадратичную ошибку пиковых ускорений на 35 % по сравнению с обучением «с нуля» [22]. Федеративное обучение, где станции обучают локальные модели и обмениваются лишь градиентами, решает задачу приватности и снижает трафик.

1.4.3 Ограничения и перспективы

1. Данные. Сильные события редки, а «ноль» (отсутствие землетрясения) статистически преобладает; классовый дисбаланс требует специальных техник (focal loss, oversampling).

2. Интерпретируемость. Для использования в СППР регуляторы требуют мотивации: почему сеть выдала именно такой прогноз. PINN-архитектуры и карты активаций частично решают проблему.

3. Обобщаемость. Полевые тесты DeepShake в Мексике показали, что без дообучения точность падает почти вдвое; трансферное обучение должно стать стандартной практикой для глобальных моделей.

Пока ИИ-алгоритмы не заменяют классические модели, но существенно расширяют инструментарий оперативного анализа и вероятностного прогнозирования, что делает их приоритетным компонентом нового модуля СППР RECASP.

1.5 Архитектура и назначение системы RECASP

Система RECASP позиционируется как «цифровая лаборатория» для экспериментальной и прикладной сейсмологии. Её проектировали так, чтобы она одинаково хорошо поддерживала off-line исследования (построение карт параметров, анализ каталога за столетие) и on-line-сервис (мгновенная визуализация текущей активности). Ниже детализированы основные подсистемы и описаны сценарии, демонстрирующие преимущества модульного дизайна.

1.5.1 Концептуальная трёхуровневая архитектура

Система RECASP реализует классическую трёхуровневую схему — представление / логика / данные.

1. Уровень данных. Хранилище построено на PostgreSQL 15 с расширением PostGIS 3, что обеспечивает:

1.1. Хранение эпицентров и разломов в виде геопространственных объектов geography.

1.2. Индексирование по GIST и быструю выборку пространственных подмножеств.

1.3. Поддержку транзакций и репликации для отказоустойчивости [24-25].

2. Уровень бизнес-логики. Выполнен в виде набора функций на ASP .NET 8:

2.1. Сервис каталогов (CRUD-операции с эпицентрами).

2.2. Сервис расчёта статистики (b-value, кумулятивная энергия).

2.3. Сервис отчётов (GeoJSON, PNG, CSV).

2.4. Общение между микросервисами организовано через лёгкую шину событий (event bus), что упрощает добавление прогнозного модуля СППР [23].

3. Уровень представления. Десктоп-клиент WinForms использует:

3.1. Компонент GMap.NET для отображения карт и тайлов OSM/ESRI [26];

3.2. Библиотеку ScottPlot 5 для интерактивных графиков (масштабирование, курсоры, экспорт SVG) [27].

1.5.2 Ключевые функциональные модули

Описание функциональных модулей представлено в таблице 1.3.

Таблица 1.3 – Описание функциональных модулей

Модуль	Назначение	Краткое описание реализации
1	2	3
Импорт данных	Загрузка EQC, TOP, miniSEED	Асинхронный парсер, проверка формата, пакетная вставка через Npgsql Copy [28]
Геоаналитика	Пространственные фильтры и расчёты	SQL + PostGIS-функции (ST_DWithin, ST_Area), расчёт плотности эпицентров «на лету»
Визуализация	Карта и графики	Кластеризация маркеров на карте, динамическое перестроение легенды графиков

Продолжение таблицы 1.3

1	2	3
Расширения (Plug-in API)	Подключение внешних алгоритмов	REST-эндпойнты api/ext, генерация OpenAPI-спецификации для быстрой интеграции

Такая структура позволяет бесшовно встроить модуль СППР: достаточно подписаться на шину событий и реализовать plug-in с прогнозным алгоритмом. Все визуальные представления (карта, графики, отчёты) используют общие компоненты, что обеспечит единообразный пользовательский опыт.

1.6 Примеры существующих решений и моделей для прогнозирования

Построение модуля СППР в RECASP не начинается «с нуля» — в мире уже функционирует ряд открытых и ведомственных платформ, накапливающих данные и выполняющих базовую аналитику. Ниже приведён обзор наиболее значимых систем, чья методологическая и техническая база может быть использована при проектировании прогнозного модуля.

1.6.1 Глобальные каталоги и веб-порталы

1. USGS Earthquake Hazards Program (EHP). Публикует глобальный каталог событий с 1900 г. в формате API+CSV, предоставляет оперативные моменты тензора, ShakeMap-карты и вероятностные афтершоковые прогнозы (OAF) [29].

2. EMSC–CSEM (Европейско-средиземноморский центр). Содержит телеметрию ~2 000 станций, формирует «Rapid Raw Determination» (< 5 мин) и карты интенсивитета DYFI на основе краудсорс-данных [30].

3. GFZ GEOFON. Поддерживает глобальный поток miniSEED, сервис Oracle-каталога и распределённую сеть вторичных дата-центров; данные доступны через FDSN-web-services, что упрощает прямую интеграцию в RECASP [32].

1.6.2 Национальные и региональные базы

1. База «Землетрясения России» (ЕГС РАН). Содержит более 900 000 записей с 1962 г., включая сибирские и дальневосточные подкаталоги; предоставляет экспорт CSV, KML и сервис подпрограммных запросов [31].

2. JMA EEW (Япония). Система раннего предупреждения Агентства метеорологии вырабатывает параметры очага за 1–2 с и транслирует сообщение в CAP-формате организационным пользователям; публичного API нет, но описание протокола открыто и может быть эмулировано [34].

1.6.3 Научно-прикладные программные пакеты

1. OpenQuake Engine (GEM Foundation). Фреймворк на Python для расчёта сейсмической опасности и риска; поддерживает индивидуальные GMPE, сценарные землетрясения, кривые потерь [33].

2. Алгоритм M8/M8S. Свободно распространяемый код (Fortran/C) для распознавания сейсмоопасных зон; требуется только каталог и границы региона. Протестирован на Калифорнии, Камчатке, Средиземноморье, показал успешный

прогноз 20 из 30 сильных событий при интервале предупредительного времени 1–5 лет [15].

1.6.4 Выводы для RECASP

1. Доступность API. ЕНР и GEOFON предоставляют REST/FDSN-интерфейсы → модуль импорта RECASP может подключаться без конвертеров.

2. Формат CAP. JMA EEW и USGS ShakeAlert используют CAP 1.2 → логично выбирать тот же стандарт для исходящих оповещений СППР.

3. Алгоритмическая преемственность. M8/M8S и OpenQuake распространяются под открытой лицензией → ядро «risk-engine» может инкапсулировать эти расчёты как plug-ins.

4. Обогащение данных. Российский каталог ЕГС РАН перекрывает основной массив исторических событий для Восточной Сибири → интеграция улучшит статистическую базу прогнозов.

Таким образом, существующие решения предоставляют как данные (каталоги, телеметрия), так и готовые алгоритмы (M8, OpenQuake), что позволит ускорить разработку прогнозного компонента и обеспечить его валидацию на общепринятых тестовых выборках.

1.7 Выводы по главе

Проведённый анализ подтвердил высокую социальную и научную значимость проблемы сейсмического прогнозирования и показал следующее.

1. Задача остаётся открытой. Надёжный краткосрочный прогноз «время – место – магнитуда» пока не достигнут из-за нелинейности геодинамических процессов, дефицита репрезентативных данных и ограничений существующих верификационных процедур. При этом долгосрочное районирование и среднесрочные индексы (алгоритм M8/M8S, оперативные афтершоковые прогнозы USGS OAF) уже приносят практическую пользу.

2. СППР—необходимое звено управления риском. Системы поддержки принятия решений позволяют объединить потоковые наблюдения, прогнозные модели и нормативные правила реагирования в единой платформе. Опыт ShakeAlert, UrEDAS и ARISTOTLE-ENHSP доказывает, что эффективность таких систем определяется не только точностью расчётов, но и устойчивостью инфраструктуры, стандартизованным форматом оповещений (CAP 1.2) и регламентами пост-анализа.

3. Методы ИИ дополняют классические подходы. Нейросетевые детекторы (PhaseNet, DeepShake) улучшают качество сигналов, модели PEGSNet расширяют горизонты раннего предупреждения, а гибридные PINN-архитектуры повышают интерпретируемость. Вместе с тем риск переобучения и классовый дисбаланс требуют физически информированных ограничений и постоянного самообучения систем.

4. RECASP готова к интеграции модуля СППР. Платформа уже включает:

4.1. Надёжное геопространственное хранилище (PostgreSQL + PostGIS);

4.2. REST-шину событий для подключения внешних сервисов;

4.3. Гибкий ГИС-интерфейс и средства статистики;

4.4. Поддержку пакетного импорта исторических и потоковых данных.

Таким образом, синтез возможностей RECASP с проверенными алгоритмическими решениями и методами искусственного интеллекта создаёт прочную основу для разработки модуля СППР, способного повысить оперативность и обоснованность управленческих действий при сейсмической угрозе.

2. Методология исследования

2.1. Принципы и подходы сейсмического прогнозирования.

Современная практика делит прогнозы землетрясений на долгосрочные (десятки лет), среднесрочные (месяцы – годы) и краткосрочные (часы – дни). Каждый горизонт опирается на специфические данные, модели и преследует собственные цели – от норм строительного проектирования до оперативной готовности служб гражданской защиты.

Долгосрочный горизонт. Основу составляют статистические методы, применяемые к историческим каталогам. Классический инструмент – закон Гутенберга – Рихтера, описывающий экспоненциальную зависимость частоты событий от их магнитуды $b \approx 1,0$ в активных регионах. Это позволяет, основываясь на многочисленных слабых толчках, оценивать вероятность редких сильных землетрясений и строить карты сейсмической опасности для территорий, что легло в основу нормативов сейсмостойкого строительства [35]. Чем полнее каталоги, тем выше пространственная разрешающая способность таких карт.

Среднесрочный горизонт. Цель – выделить «окна повышенного риска» в конкретных зонах. Наиболее известен алгоритм M8, разработанный школой Гельфанда – Кейлиса-Борока. Алгоритм рассчитывает комплекс индексов (активность, энергетическая мощность, пространственная концентрация) и выдаёт предупреждение, когда совокупный показатель превышает порог. Ретроспективные и перспективные испытания (Калифорния, Камчатка, Средиземноморье и др.) демонстрируют статистически значимый выигрыш по сравнению со случайной моделью: вероятность обнаружения крупных событий повышена, хотя доля ложных тревог остаётся высокой [15].

Краткосрочный горизонт. По-прежнему является наиболее дискуссионным. Международные комиссии ICEF и USGS констатируют отсутствие общепринятого метода, позволяющего надёжно предсказать время, место и магнитуду сильного толчка «за часы» [14, 16]. Исследуются различные предвестники – аномалии в радоне, электромагнитном поле, деформациях GNSS, ионосферном содержании ТЕС и т. д. Отдельные успехи (эвакуация Хайчэн, 1975 г.) носят единичный характер; большинство наблюдаемых аномалий не приводит к сильному событию [36]. Поэтому современная концепция Operational Earthquake Forecasting (OEF) смещает акцент к вероятностным сценариям: прогноз публикуется в виде интервала вероятностей (на пример – « $\approx 5\%$ вероятности события $M_w \geq 7.5$ за 72 ч»), регулярно уточняется по мере поступления новых данных, а результаты проверяются пост-фактум [13].

В итоге эволюция идёт от поиска жёстких «детерминированных» предвестников к интеграции статистических закономерностей, вероятностных моделей и физического понимания подготовки разрывов, что и определяет методологическую основу дальнейших разделов работы.

2.2. Ретроспективный анализ как метод проверки моделей.

Для первичной оценки работоспособности и точности алгоритмов прогнозирования в данной работе принят ретроспективный (post-factum) анализ. Его суть заключается в том, что исторический каталог разделяется на обучающую («Train») и тестовую («Predict») части; далее модель строит прогноз внутри тестового окна, а полученные результаты сопоставляются с реально произошедшими событиями («Compare»).

Применение такого подхода обосновано следующими факторами:

1. Редкость разрушительных землетрясений. Для статистически значимой перспективной (prospective) проверки пришлось бы ожидать десятилетия, пока накопится достаточное количество сильных толчков; ретроспективный же анализ сразу использует многолетний архив наблюдений [37].

2. Контроль переобучения. Подразделение данных на «прошлое / будущее» позволяет выявить ситуации, когда модель хорошо описывает обучающий отрезок, но теряет эффективность при переносе на незнакомый интервал [38].

3. Международная практика. Проекты Collaboratory for the Study of Earthquake Predictability (CSEP) декларируют ретроспективное тестирование как обязательный предварительный этап перед оперативным развёртыванием прогноза [39].

4. Скользящее окно. Следуя рекомендациям Werner и соавт. (2010), в работе реализован многократный цикл «Train → Predict → Compare» с перемещением пары интервалов по каталогу; это повышает надёжность статистических выводов [40].

В рамках выбранной схемы:

1. Train-интервал. Каталог 1970 – 1994 гг. служит для калибровки параметров алгоритмов (M8-S, OEF-ансамбль, PINN-модель).

2. Predict-интервал. На отрезке 1995 – 2019 гг. формируются «прогноз-карты» вероятности событий $M_w \geq 6.0$ с шагом 0.1 и оконным горизонтом 1 год.

3. Compare. Итоговые карты сопоставляются с фактическими землетрясениями по метрикам POD, FAR, S-score; оценка значимости выполняется с помощью эталонной случайной модели (Poissonian spatial shuffle), как принято в CSEP.

Ретроспективное тестирование, таким образом, позволяет:

1. Быстро получить первые численные характеристики точности.
2. Объективно сравнить несколько конкурирующих алгоритмов на идентичном наборе данных.

3. Отобрать оптимальные конфигурации для последующей перспективной (реальной) проверки.

2.3. Структура и формат каталога землетрясений.

Исходным эмпирическим фундаментом исследования служит каталог землетрясений — реляционная база, аккумулирующая сведения обо всех зарегистрированных сейсмических событиях. В работе используется объединённый каталог, сформированный на данных Единой геофизической службы РАН (ЕГС РАН) и дополненный оперативными сводками EMSC

и GFZ GEOFON, что обеспечивает как историческую глубину, так и текущую полноту покрытия сейсмичности Евразии [30-32].

2.3.1 Базовая схема записи

Каждая запись каталога содержит четыре ключевых поля.

1. id — уникальный идентификатор события (целое число).
2. date — дата и время (UTC) с точностью до секунд.
3. coordinate — гео-координаты эпицентра типа Geography расширения PostGIS, хранящиеся как WKB-объект в системе WGS-84; такой формат обеспечивает быстрое индексирование и пространственные запросы [25].

4. class — энергетический класс K, Класс коррелирует с моментной магнитудой: для регионов Сибири значения $K \approx 13 - 14$ соответствуют $M \approx 6.0$ [35].

Дополнительные атрибуты (глубина очага, тип события, код станции и др.) сохраняются, но в контексте прогнозной задачи решающее значение имеют именно вышеуказанные четыре параметра, поскольку на их основе вычисляются статистические закономерности и составляются прогнозы.

2.3.2 Контроль качества и оценка полноты

Для надёжности аналитики реализованы:

1. Автоматический скрипт проверки outlier-координат (точки с широтой > 90 , долготой > 180 отбрасываются) и некорректных классов $K < 3$ либо $K > 22$.

2. Оценка временной полноты каталога по методу максимальной кривизны Gutenberg–Richter; получено $M_c = 3.2$ для Камчатки и $M_c = 2.5$ для Байкальской рифтовой зоны за последние 15 лет [42].

3. Пространственно-временное декластеризование (алгоритм Reasenberg) для отделения афтершоков, применяемое при расчётах b-value и оценке долгосрочной угрозы [43].

2.3.3 Интероперабельность и экспорт

Для визуализации в RECASP графический модуль отрисовывает данные в GeoJSON с пакетной доставкой тайлов (TileJSON v3).

Сильные события ($M_w \geq 5.5$) автоматически сопоставляются с Global CMT и дополняются моментным тензором, что впоследствии позволит использовать физико-информированные модели источника [41].

Таким образом, каталог с централизованным хранением в PostgreSQL/PostGIS и унифицированным набором полей обеспечивает надёжную платформу для ретроспективного тестирования моделей и дальнейшей интеграции в операционную систему поддержки решений RECASP.

2.4. Предварительная обработка данных

Перед статистическим анализом и обучением моделей каталог прошёл многошаговую очистку и унификацию.

1. Фильтрация записей:

- 1.1. Дубликаты, появившиеся из-за многократные передачи одних и тех же событий между национальными и международными центрами, устранялись по совпадению времени (± 5 с) и координат (± 10 км). Для оставшейся

записи сохранялся первоисточник, а все идентификаторы-дубли заносились в отдельную таблицу ссылок.

1.2. Нетектонические события (карьерные взрывы, ядерные испытания, индустриальные шумы) исключались по меткам поставщика данных либо по диагностическим признакам: очень малая глубина, регулярное расположение вдоль горных выработок, чёткое рабочее расписание [44].

1.3. Была введена нижняя граница полноты каталога — энергетический класс $K \geq 8$ (примерно соответствует $M \approx 2.5$). Её положение подтверждено анализом “излома” на графике Гутенберга–Рихтера и методикой оценки минимальной магнитуды полной регистрации (M_c) [42].

2. Приведение энергетических оценок к единой шкале. Исторические данные из разных сейсмосетей содержали небольшие методические расхождения в расчёте энергетического класса. Для их сглаживания все значения K были пересчитаны в эквивалентную моментную магнитуду M_w по алгоритму, рекомендованному в руководстве NMSOP-2 (GFZ, 2012), и затем возвращены обратно в «откалиброванный» K . Такой «обратный» перевод позволил сохранить привычный для российских каталогов формат без потери сопоставимости с современными событиями [45].

3. Декодирование координат. Поля географии хранились в бинарном формате WKВ. Для последующих расчётов и визуализации координаты извлекались функциями ST_X / ST_Y PostGIS, превращались в десятичные градусы и сохранялись как текст WKT-точек, что упростило отслеживание ошибок в геометрии [25].

4. Работа с неполными полями. Для единичных исторических записей, где отсутствовал энергетический класс, выполнялся поиск по глобальным каталогам EMSC и GEOFON; при успешном сопоставлении недостающее значение дополнялось, иначе событие помечалось как «неполное» и исключалось из вычислений, чувствительных к K (например, расчёт распределения b -value) [30, 32].

Эти процедуры обеспечили согласованный, лишённый артефактов набор данных. Итоговый каталог насчитывает ~600 000 тектонических событий с 1960 г. по 2024 г. и соответствует требованию полноты для классов $K \geq 8$, что критично для корректных выводов о параметрах сейсмичности [46].

Ниже приведено описание алгоритма на языке Кумир.

Алгоритм ПодготовкаКаталога

// Главная процедура подготовки исходного каталога землетрясений

// 1. Загрузка исходного CSV

Процедура ЗагрузитьCSV(ИмяФайла: Строка; var Каталог: Таблица)

Нач

// Прочитать все строки из файла

Каталог := ЧитатьТаблицуИзCSV(ИмяФайла);

// Сохранить исходные идентификаторы и источник

Кон

// 2. Очистка и фильтрация

Процедура ОчиститьИФильтровать(var Каталог: Таблица;

var СсылкиДубли: Таблица)

Нач

// 2.1 Удаление дубликатов


```

    Для КаждойСтроки I От 1 До Длина(Каталог) - 1 Цикл
        Для КаждойСтроки J От I+1 До Длина(Каталог) Цикл
            Если Время(Каталог[I])  $\approx$  Время(Каталог[J],  $\pm 5$ с) И
                Расстояние(Каталог[I].Коорд, Каталог[J].Коорд)  $\leq 10$ км Тогда
                    Ссылка := НовыйСтрока()
                    Ссылка.ОсновнойID := Каталог[I].ID
                    Ссылка.ДубликатID := Каталог[J].ID
                    Добавить(СсылкиДубли, Ссылка)
                    УдалитьСтроку(Каталог, J)
                    J := J - 1
            Кон
        Кон
    Кон

// 2.2 Исключение нетектонических событий
I := 1
Пока I  $\leq$  Длина(Каталог) Цикл
    Если Каталог[I].Тип = "карьер" ИЛИ
        Глубина(Каталог[I]) < 1км ИЛИ
        Регулярность(Каталог[I]) = Истина Тогда
        УдалитьСтроку(Каталог, I)
    Иначе
        I := I + 1
    Кон
Кон

// 2.3 Контроль полноты каталога (K  $\geq 8$ )
I := 1
Пока I  $\leq$  Длина(Каталог) Цикл
    Если Каталог[I].КлассK < 8 Тогда
        УдалитьСтроку(Каталог, I)
    Иначе
        I := I + 1
    Кон
Кон
Кон

// 3. Нормализация энергетической шкалы
Процедура НормализоватьЭнергию(var Каталог: Таблица)
Нач
    Для КаждойСтроки I От 1 До Длина(Каталог) Цикл
        Mw := ПеревестиКвMw(Каталог[I].КлассK) // по алгоритму NMSOP 2
        Knew := ПеревестиMwвK(Mw) // обратный перевод
        Каталог[I].КлассK := Knew
    Кон
Кон

// 4. Декодирование геометрии
Процедура ДекодироватьГеометрию(var Каталог: Таблица)
Нач
    Для КаждойСтроки I От 1 До Длина(Каталог) Цикл
        // Применить PostGIS ST_X, ST_Y
        (lon, lat) := WKbвКоординаты(Каталог[I].WKb)
        Каталог[I].Широта := lat
        Каталог[I].Долгота := lon
        Каталог[I].WKT := СформироватьWKT(lon, lat)
    Кон
Кон

// 5. Дополнение неполных записей
Процедура ДополнитьНеполные(var Каталог: Таблица)
Нач
    Для КаждойСтроки I От 1 До Длина(Каталог) Цикл
        Если Каталог[I].КлассK = ПУСТО Тогда

```

```

Совпадение := ПоискBEMSCиGEOFON(Каталог[I].Дата,
                                   Каталог[I].Широта,
                                   Каталог[I].Долгота)

Если Совпадение.Найдено Тогда
    Каталог[I].КлассК := Совпадение.КлассК
Иначе
    УдалитьСтроку(Каталог, I)
    I := I - 1
Кон
Кон
Кон

// 6. Формирование итогового набора
Процедура СохранитьКаталог(Каталог: Таблица; ИмяФайла: Строка)
Нач
    ЗаписатьТаблицуBCSV(Каталог, ИмяФайла)
    // Опционально: загрузить в PostGIS
Кон

// Основная последовательность
Нач
    ЗагрузитьCSV("earthquakes.csv", Каталог)
    ОчиститьИФильтровать(Каталог, СсылкиДубли)
    НормализоватьЭнергию(Каталог)
    ДекодироватьГеометрию(Каталог)
    ДополнитьНеполные(Каталог)
    СохранитьКаталог(Каталог, "clean_catalog.csv")
Кон

```

2.5. Статистический анализ сейсмической активности.

На очищенном каталоге выполнен комплекс статистических процедур, которые позволяют количественно описать сейсмический режим исследуемой территории и подготовить входные данные для моделей прогнозирования.

1. Временная активность (суточная и годовая). Для всего периода наблюдений построены кривые числа событий в сутки и суммарных годовых показателей. Суточные диаграммы рассчитаны как скользящее среднее и показывают чередование периодов всплесков и затиший: после каждого крупного землетрясения виден «шлейф» афтершоков, затухающий примерно за полтора месяца, что согласуется с классическим законом убывания активности, описанным ещё Таро Уцу [47]. Годовые графики демонстрируют резкий рост зарегистрированной сейсмичности в 1960-е годы, обусловленный расширением сети станций, и относительную стабилизацию числа событий с 1980-х. Проверка возможной сезонной периодичности (группировка по месяцам) не выявила статистически значимых колебаний — сейсмичность определяется тектоническими, а не климатическими процессами.

2. Параметры закона Гутенберга – Рихтера. Для разных регионов и временных интервалов оценены коэффициенты a и b , определяющие соотношение малых и крупных землетрясений. Расчёт производился стандартной методикой максимального правдоподобия, рекомендованной CSEP [39]. Усреднённое значение b для совокупного каталога близко к единице (около 0.95), что типично для активных областей. Пространственная вариация заметна: на Кавказе коэффициент слегка ниже (≈ 0.8), тогда как на Камчатке превышает единицу, что указывает на различия в механизмах накопления и

разрядки тектонических напряжений [48]. Полученные параметры затем используются в вероятностных расчётах ожидаемого числа событий выше заданного порога магнитуды.

3. Карты плотности эпицентров. Пространственное распределение сейсмичности визуализировано методом ядерной оценки плотности. На сетке с шагом половина градуса для каждой ячейки подсчитано количество событий в пределах 50 км с гауссовым сглаживанием. Карты выявляют главные концентраторы активности: субдукционные зоны Курило-Камчатской дуги, локальные разломы в Саянско-Байкальском рифтовом поясе, горные пояса Кавказа и Тянь-Шаня. Отдельно построены карты для сильных землетрясений (энергетический класс не ниже $K12$), что позволило локализовать эпицентральные зоны потенциально разрушительных событий [49]. Такие «тепловые» карты далее трансформируются в прогнозные карты риска путём наложения на слои уязвимости населения и инфраструктуры.

4. Кластеризация событий. С помощью анализа ближайших временно-пространственных соседей и алгоритма ETAS (Epidemic-Type Aftershock Sequence) установлено, что почти две трети событий класса $K \geq 10$ принадлежат кластерным последовательностям — либо афтершокам, либо форшокам. Определение кластеров позволяет отделить фоновую (пойсоновскую) сейсмичность от временных всплесков, вызванных триггерным эффектом крупных толчков. Это важно для моделей краткосрочного прогнозирования, в которых вероятность будущего события локально повышается после недавнего землетрясения [50].

В совокупности временные ряды активности, оценки a - и b -параметров, плотностные карты и результаты кластерного анализа образуют эмпирическую основу для настройки алгоритмов прогнозирования, описанных в следующих разделах, и одновременно представляют самостоятельный итог — систематизированное представление эволюции сейсмического режима региона [51].

2.6. Алгоритмы прогнозирования, применяемые в исследовании.

В рамках разработки модуля СППР использован комплексный подход к прогнозированию землетрясений, сочетающий несколько алгоритмических методов — от оценки фоновой вероятности до выявления предвестников сильных событий.

1. Вероятностный метод фоновой сейсмичности. Опирается на пуассоновское распределение событий и закон повторяемости Гутенберга – Рихтера. Для каждой ячейки рассчитывается среднегодовая частота землетрясений заданной магнитуды; вероятность хотя бы одного события в интервале T . Такой подход лежит в основе национальных карт сейсмической опасности и рекомендаций по операционному прогнозированию [35, 40]. Преимущества — простота и воспроизводимость; ограничение — не учёт текущих кластеров и предвестников, поэтому метод применяется главным образом для долгосрочных и среднесрочных оценок.

2. Кластерный прогноз (афтершоки и рои). Для краткосрочной эволюции сейсмичности используется адаптированная модель ETAS: после каждого

значимого толчка локальная вероятность новых событий повышается и затухает со временем по закону Омори. Параметры (продуктивность, скорость временного и пространственного затухания) калиброваны на кластерной статистике каталога. В операционном режиме система выдаёт прогноз: «в следующие 7 суток в радиусе 100 км ожидается не менее N событий выше порога». Международные исследования показывают, что модели, учитывающие кластеризацию, превосходят статические схемы по краткосрочной точности [13, 43, 44].

3. Среднесрочный индикатор M8. Алгоритм M8 отслеживает «предвестниковую активизацию» перед крупными землетрясениями. Территория делится на блоки ≈ 500 км; для каждого формируются показатели (частота, энерговыделение, доля сильных событий и др.) на скользящем окне 5 лет. Совпадение нескольких критериев переводит блок в режим «Alarm» с горизонтом $\Delta T \approx 5$ лет. Тесты по нашему каталогу дали $\sim 70\%$ попаданий главных шоков во временные-пространственные окна тревоги, что согласуется с оценками эффективности M8 [5, 15, 16].

4. 4 Комбинированные и машинно-обученные модели. В дополнение к классическим алгоритмам реализован ансамбль «статистика + кластеризация + ИИ». Рекуррентная нейронная сеть обучается на временных рядах региональных параметров, а сверточная сеть распознаёт пространственные шаблоны сейсмичности; их оценки корректируют базовую пуассоновскую вероятность. Первичные испытания показали, что глубокая CNN отличает предкритическое состояние от фонового с точностью $\sim 75\%$ ($F1 \approx 0.8$). Однако модели ИИ применяются осторожно из-за риска переобучения и сложности интерпретации [19-22].

Таким образом, модуль СППР сочетает интерпретируемые статистические методы, кластерные поправки и первые элементы глубокого обучения, обеспечивая многоуровневый прогноз: от долгосрочного фонового риска до часовых оценок после крупных толчков.

2.7. Построение карты сейсмического риска.

Прогнозирование одной лишь вероятности сейсмических событий недостаточно для практической ценности – нужно перевести это в оценку сейсмического риска, то есть потенциальных последствий для людей и инфраструктуры. В методологии исследования предусмотрен модуль, преобразующий результаты сейсмического прогноза в карту риска по зонам. Подход базируется на классической концепции $Risk = Hazard \times Vulnerability \times Exposure$ [6-7].

Hazard (опасность) задаётся либо ожидаемой интенсивностью сотрясений, либо вероятностью превышения порогового уровня в данной зоне. Например, по результатам раздела 2.6 для каждой зоны оценка: «вероятность события $M \geq 6$ в следующие 5 лет = $p\%$ » либо «ожидаемая максимальная интенсивность в следующие 50 лет = VII баллов MSK-64 с 90 % вероятностью» [35]. Эти показатели опасности переводятся в прогнозируемую сейсмическую нагрузку.

Далее вводится компонент Vulnerability (уязвимость) – функция, связывающая интенсивность с ожидаемыми потерями или ущербом. Для каждой

зоны характеризуется типичный фонд построек и коммуникаций. Например, в зоне X 60 % зданий – сейсмостойкие каркасные, 30 % – кирпичные старой постройки, 10 % – лёгкие деревянные. Для каждого типа берётся кривая уязвимости, показывающая долю разрушений в зависимости от интенсивности [52, 53].

Третий компонент – Exposure (экспозиция) – отражает количество объектов и их ценность в зоне. Это может быть численность населения, число зданий, их суммарная площадь или стоимость. В нашем случае для упрощения принята экспозиция в виде численности населения (по статистике региона) и общего числа зданий в зоне (оценено по данным открытых источников и Росстраха) [54].

Процесс построения карты риска состоит в следующем. Зона разбиения (например, ячейки 0.5×0.5 или муниципальные районы) для каждой рассчитывается hazard-показатель из прогноза. Затем этот показатель конволюируется с функцией уязвимости по методике OpenQuake [33]: если прогнозируемая интенсивность $I = VII$ баллов с вероятностью 0.2, а уязвимость такова, что при $I = VII$ ожидается 5 % разрушений зданий, то интегральный риск $= 0.2 \times 0.05 = 0.01$ (1 % зданий). Умножая на экспозицию (10 000 зданий), получаем ожидаемые 100 разрушенных зданий. Аналогично для человеческих потерь: при населении 50 000 и коэффициенте потерь 0.001 ожидаемые потери $\approx 0.2 \times 50 = 10$ человек. Эти расчёты повторяются для всех зон.

Результат визуализируется как карта риска, где цветом или значением в каждой зоне обозначается ожидаемый ущерб (материальный или людской) за заданный период. Такая оценка риска позволяет выделить наиболее уязвимые территории и планировать превентивные меры. Отметим, что итоговая карта носит вероятностный характер и сильно зависит от достоверности входных компонентов (hazard, vulnerability и exposure) [40].

2.8. Ретроспективная валидация модели (метрики и процедуры).

После разработки прогнозных алгоритмов и формировании прогнозов риска был организован процесс ретроспективной валидации, призванный количественно оценить качество модели. Как упоминалось в разделе 2.2, используется методология Train–Predict–Compare: берём исторический интервал, на нём обучаем модель; затем на последующем интервале генерируем прогноз и сравниваем с фактическими данными [39]. Для систематической проверки модель прогоняется по скользящему окну: например, обучение 1950–1990, прогноз на 1991–2000, затем обучение 1960–2000, прогноз на 2001–2010 и т.д. Это позволяет получить несколько независимых оценок прогностической способности [38].

Ключевой частью валидации является выбор метрик качества. В данной работе применены следующие показатели:

1. Precision (точность прогноза по положительным срабатываниям). В контексте сейсмопрогнозов это доля предсказанных событий, которые действительно произошли. Высокая Precision означает, что модель редко «тревожит зря», большинство её предупреждений оправдались. Для оценки необходимо сравнить, сколько из зон/периодов, помеченных моделью как опасные, фактически испытали сильное землетрясение. Например, если модель

предсказала 10 событий $M \geq 6$ и 6 из них реально случились, $\text{Precision}=0.6$ (60 %) [38].

2. Recall (полнота обнаружения). Это доля фактически случившихся событий, которые были предсказаны моделью. Высокое значение Recall означает, что модель умеет ловить почти все реальные события, даже ценой лишних ложных тревог. В наших экспериментах, например, Recall для событий $M \geq 6$ оказался порядка 0.8–0.9, то есть 80–90 % крупных землетрясений произошли в тех зонах и интервалах, где модель ранее предупреждала о повышенном риске. Однако это было достигнуто при умеренном Precision (~ 0.5 –0.6), т.е. значительная часть тревог оказалась ложной. Такой компромисс типичен для прогностических задач: более «чувствительная» модель увеличивает полноту (Recall) ценой снижения точности (Precision) [38].

3. MAE по интенсивности (Mean Absolute Error по прогнозируемой интенсивности). Данная метрика оценивает ошибку в предсказанных значениях сейсмического воздействия. MAE показывает, насколько в среднем прогнозируемые уровни тряски отличаются от реально наблюдавшихся. В наших расчётах интенсивность оценивается через эквивалентную магнитуду или энергию, переведённую в баллы сейсмического воздействия. Полученное среднее абсолютное отклонение составило ~ 0.7 балла MSK-64, что является приемлемым результатом для долгосрочных прогнозных карт [33].

4. Дополнительные показатели. Помимо перечисленных, рассчитывались также Specificity (доля верно предсказанных «спокойных» периодов без событий) и F1-score (сводный показатель баланса Precision/Recall). Однако их информативность для редких событий ограничена, поэтому основное внимание уделялось Precision и Recall как паре метрик «ложная тревога vs пропуск», а MAE – как интегральной ошибке по интенсивности. Также строились ROC-кривые и вычислялась площадь под ROC (AUC) для различных порогов тревожного индекса модели, чтобы оценить, насколько модель лучше случайного угадывания. Значение AUC ~ 0.75 в наших тестах указывает на уверенное превосходство над случайной моделью, но ещё далёкое от идеального прогнозирования (AUC=1.0) [46].

В процессе ретроспективной валидации также были проанализированы ошибки модели, пытаясь понять их природу. Ложные тревоги часто связаны с тем, что предвестниковые признаки могут проявляться, но событие не происходит – то есть система зафиксировала аномалию (например, кластер мелких толчков), однако она не привела к крупному землетрясению в отведённое окно. Это указывает на ограниченность понимания физического процесса: возможно, напряжения разрядились малыми событиями или временное окно выбора было слишком узким. Пропущенные же события (низкий Recall) чаще всего объясняются тем, что некоторые землетрясения происходили «внезапно», без заметных предварительных сигналов в используемых индикаторах – здесь модель просто не имела входной информации, чтобы что-то предсказать. Такая ситуация соответствует позиции USGS о том, что многие землетрясения не дают явных предвестников [14].

Важно подчеркнуть, что ретроспективная оценка – лишь первый шаг. Окончательная проверка ценности модели должна проводиться в режиме проспективного прогнозирования (prospective testing) на новых данных, как это регламентируется, например, международными испытательными полигонами CSEP [39]. Тем не менее, полученные ретроспективные метрики уже позволили скорректировать модель и обосновать её применимость: например, принято решение улучшить алгоритм M8, чтобы снизить ложные тревоги, и усилить вес кластерного компонента для повышения Recall. Дальнейшая работа будет включать накопление статистики работы модуля СППР в реальном времени.

2.9. Инструментальные средства реализации.

Разработанный модуль СППР интегрирован в платформу RECASP с использованием современных технологий баз данных, программирования и визуализации. Ниже перечислены основные инструментальные средства и их роль в реализации:

1. PostgreSQL + PostGIS – сервер базы данных и геопространственное расширение. PostgreSQL выбран как надёжная реляционная СУБД с открытым исходным кодом, обеспечивающая хранение больших объёмов данных и выполнение сложных запросов [24]. Расширение PostGIS позволяет напрямую работать с географическими координатами и геометриями на стороне БД, выполняя расчёты расстояний, поиск попадания в зону и пр. [25]. В контексте модуля СППР это значит, что, например, выборка всех событий внутри заданного полигона или подсчёт количества землетрясений в окрестности заданной точки может быть выполнен одним SQL-запросом с использованием функций PostGIS (ST_Contains, ST_DWithin и т.д.). База данных также хранит промежуточные результаты расчетов, прогнозные значения по зонам, метрики, что позволяет централизованно управлять информацией. Документация PostgreSQL и PostGIS служила основой при проектировании структуры данных и оптимизации запросов.

2. Платформа .NET (C#) с Windows Forms. Логика приложения и пользовательский интерфейс реализованы на платформе .NET 6.0 с использованием фреймворка Windows Forms (WinForms) для настольного приложения [23]. Выбор .NET обусловлен богатой экосистемой библиотек для научных вычислений и удобством разработки интерфейса. WinForms, хотя и является достаточно традиционным GUI-фреймворком, предоставляет быстрые инструменты для создания форм, таблиц, графиков, что важно для прототипирования СППР. Приложение включает основное окно с картой, панель управления параметрами прогноза, вкладки для графиков статистики и отчётов. Для взаимодействия с БД используется официальный .NET-драйвер Npgsql (ADO.NET Data Provider for PostgreSQL), позволяющий выполнять SQL-запросы прямо из C# к PostgreSQL [28]. Это позволяет, например, по нажатию кнопки «Обновить прогноз» автоматически подтянуть свежие данные из каталога (SQL SELECT), вычислить моделью прогноз (на стороне C#) и результаты записать обратно в БД (SQL INSERT/UPDATE). Использование LINQ (Language Integrated Query) в C# дополнительно упростило обработку данных: часть статистических расчётов (например, агрегирование по времени, фильтрация списков событий по

критериям) реализована на уровне приложения с помощью LINQ-запросов к коллекциям объектов, полученных из БД.

3. Библиотека GMap.NET для картографической визуализации. Для отображения интерактивных карт в приложении интегрирована библиотека GMap.NET [26]. Это открытое решение для .NET, позволяющее встроить карту (на основе Google Maps, OpenStreetMap или других тайловых сервисов) в Windows Forms. С помощью GMap.NET реализована панель карты, на которую наносится географическая информация: эпицентры землетрясений (маркерами), границы прогнозных зон (многоугольниками), тепловые карты плотности (как полупрозрачные заливки) и т.д. Пользователь может масштабировать карту, переключать слои (например, фоновые спутниковые снимки или схематическую карту). GMap.NET также предоставляет средства кэширования тайлов, что ускоряет работу при плохом интернете.

4. Библиотека ScottPlot для графиков. Для построения научных графиков (временные ряды, гистограммы, регрессионные линии) применяется библиотека ScottPlot [27]. Это высокоуровневая .NET-библиотека, позволяющая легко генерировать графики и диаграммы. В модуле СППР ScottPlot используется для отображения результатов статистического анализа: график суточной активности (line plot), кумулятивная кривая Гутенберга–Рихтера (log-log plot), распределение остаточных отклонений модели (histogram) и т.п. В интерфейсе программы графики встроены на соответствующие вкладки, обновляются динамически по запросу пользователя. Например, после очередного прогона прогнозов пользователь может открыть вкладку «Валидация» и увидеть график ROC-кривой, автоматически построенный из обновлённых данных. Преимущество ScottPlot – богатые возможности стилизации и взаимодействия (приближение, подсказки на точках), что делает анализ результатов более наглядным.

5. NumPy: многомерные массивы ndarray; векторные операции; линейная алгебра и Фурье-преобразования [55].

6. pandas: структуры DataFrame/Series; чтение/запись CSV, Excel, SQL; группировка, фильтрация, ресэмплинг [56].

7. scikit-learn: алгоритмы классификации, регрессии, кластеризации; предобработка; кросс-валидация и подбор гиперпараметров [57].

8. Shapely: работа с геометрией (Point, LineString, Polygon); операции пересечения, буферизации; WKB/WKT [58].

9. PyYAML: парсинг и генерация YAML; безопасная загрузка (safe_load) и дампы в строку/файл [59].

10. rasterio: чтение/запись GeoTIFF; доступ к метаданным (CRS, аффинное преобразование); ресэмплинг и репроекция через rasterio.warp [60].

11. Matplotlib: построение 2D-графиков (plot, imshow, гистограммы); настройка стилей, аннотаций, сохранение в PNG/PDF [61].

12. Plotly: интерактивные визуализации в браузере (go.Figure, px.*); тепловые карты, Mapbox; экспорт в HTML [62].

13. Experta: фреймворк экспертных систем; определение фактов (Fact), правил (@Rule) и движок Rete; методы declare и run [63].

14. Прочие инструменты:

- 14.1. Для хранения кода использовалась система Git [64].
- 14.2. Разработка велась в среде Visual Studio 2022 [65].
- 14.3. Для миграции базы данных применялись SQL-скрипты.
- 14.4. OpenAPI/Swagger рассмотрен для документирования API [66].

Таким образом, сочетание PostgreSQL/PostGIS на сервере данных и .NET/WinForms на клиенте обеспечивает надёжную и гибкую основу для модуля СППР. Использование готовых библиотек (GMap.NET, ScottPlot, Npgsql) значительно ускорило разработку, позволив сосредоточиться на алгоритмах прогнозирования, а не на технических деталях низкого уровня. Кроме того, опора на открытые стандарты (PostGIS, REST) облегчает потенциал масштабирования и переноса решения на другие платформы.

2.10 Выводы по главе

Проведённая методология объединяет разные подходы к прогнозированию землетрясений: классические статистические методы, среднесрочные алгоритмы поиска опасных периодов и краткосрочные модели кластеризации с применением машинного обучения:

1. Горизонты прогнозирования. Долгосрочные оценки базируются на законе Гутенберга–Рихтера и позволяют строить карты сейсмической опасности для нормативного проектирования зданий [35]. Среднесрочный метод M8 демонстрирует статистически значимый выигрыш по сравнению со случайной моделью при выделении «окон повышенного риска» на горизонте в несколько лет [15]. Краткосрочные прогнозы смещены в сторону вероятностных сценариев (Operational Earthquake Forecasting), где выдаются интервальные вероятности событий и проводятся регулярные уточнения по мере поступления новых данных [14, 16].

2. Ретроспективная валидация. Использование схемы «Train–Predict–Compare» и скользящих окон позволило оперативно оценить точность моделей на обширном историческом архиве, выявить и контролировать переобучение, а также подготовиться к последующему проспективному тестированию согласно международным стандартам CSEP [37, 39, 40].

3. Обработка каталога. Каталог землетрясений, объединённый из ЕГС РАН, EMSC и GFZ GEOFON, прошёл многоуровневую очистку (удаление дубликатов, нетектонических событий и выборку по порогу полноты $K \geq 8$) и унификацию шкал энергетических оценок [42, 44, 45].

4. Статистический анализ. Были рассчитаны временные ряды активности, параметры а- и b-закона Гутенберга–Рихтера, карты плотности эпицентров и результаты кластеризации ETAS. Это дало исчерпывающее описание сейсмического режима региона и готовые входные данные для прогнозных моделей [47, 48, 49, 50, 51].

5. Набор прогнозных алгоритмов. Модуль СППР включает пуассоновский фон, кластерный ETAS-прогноз, среднесрочный индикатор M8 и гибридный ансамбль «статистика + кластер + ИИ» с элементами RNN и CNN. Такое сочетание обеспечивает многослойный прогноз – от фонового риска до часовых оценок после сильных толчков [5, 13, 15, 16, 19, 20, 21, 22, 35, 40, 43, 44].

6. Карта риска. Результаты прогноза переводятся в карту ожидаемого ущерба с учётом уязвимости (functions of vulnerability) и экспозиции (численность населения, число зданий) по формуле $Risk = Hazard \times Vulnerability \times Exposure$, что даёт практический инструмент для планирования превентивных мер [6, 7, 52, 53, 54].

7. Метрики валидации. Основными показателями качества стали Precision и Recall (для оценки баланса ложных тревог и пропусков), MAE по интенсивности и AUC ROC, что позволило количественно сравнить эффективность разных алгоритмов и выбрать оптимальные конфигурации [38, 46].

8. Техническая платформа. Инструментальная база реализована на PostgreSQL + PostGIS (хранение и пространственные запросы) и .NET/WinForms (логика и интерфейс), с использованием GMap.NET для картографии и ScottPlot для графиков. Такой стек обеспечивает надёжность, расширяемость и скорость разработки модуля СППР в рамках RECASP [23-28, 33, 55-57].

В целом, разработанная методология создаёт прочный фундамент для перехода от ретроспективных испытаний к реальному развертыванию модуля СППР в режиме перспективного прогнозирования и оперативной поддержки принятия решений при сейсмической угрозе.

3 Практическая часть

3.1 Цели и задачи прикладной реализации

Целью прикладной реализации разработанной системы прогнозирования землетрясений является создание законченного программного комплекса, способного на основании исторических и текущих сейсмических данных автоматически генерировать полноценные прогнозные продукты, удобные как для исследовательской работы, так и для практических служб мониторинга. В основе этой системы лежит модульная архитектура: каждый этап — от загрузки и очистки каталога землетрясений до визуализации результатов — оформлен в собственный модуль с чётким интерфейсом и независимым циклом разработки. Это обеспечивает простоту сопровождения, возможность замены реализации отдельных частей и упрощает интеграцию в существующие конвейеры анализа.

Первой и ключевой задачей прикладной части является формирование долгосрочной карты рисков. Для этого используется метод ядровой оценки плотности (KDE), который на основе всех прошедших событий строит равномерную сетку с заданным шагом по широте и долготе и заполняет её весами, отражающими накопленную историю сейсмической активности. В конфигурационном файле (`config.yaml`) параметр `bandwidth_km` позволяет гибко задавать степень сглаживания: небольшие значения выявляют локальные «горячие» очаги, а большие — подчёркивают глобальные зоны повышенного риска. Эмпирический подбор этого параметра через перекрёстную проверку на исторических данных позволяет оптимизировать качество прогноза.

Вторым по порядку этапом выступает краткосрочная корректировка прогноза при помощи модели ETAS. Она учитывает динамику недавних событий: каждое новое землетрясение усиливает прогноз в своей окрестности и оказывает минимальное влияние на отдалённые участки. Параметры модели — интенсивность продуктивности, временное и пространственное затухание, радиус влияния — выносятся в конфиг и могут оперативно настраиваться по результатам ретроспективного анализа. Такой подход обеспечивает адаптивность прогноза: карта риска мгновенно реагирует на реальные события, что особенно важно для служб экстренного реагирования.

Третьей задачей является региональный анализ. Прогнозная область разбивается на макрзоны (например, Западная, Центральная и Восточная зоны Байкальского рифта), причём границы задаются прямоугольными рамками или произвольными полигонами в конфиге. Для каждой зоны автоматически отбираются события каталога, затем рассчитываются статистические характеристики: длительность «тихого периода», изменение средней годовой частоты до и после контрольной даты, плотность событий на единицу площади. Эти параметры являются входными фактами региональной экспертной системы на базе *Experta*, которая на их основе выдаёт заключение о среднесрочном и долгосрочном уровне риска с указанием ожидаемого диапазона классов будущих землетрясений.

Параллельно формируется список локальных «горячих точек» — ячеек сетки с максимально высокими прогнозными вероятностями. Для каждого из них рассчитывается уровень риска («Высокий», «Средний», «Низкий») по заданным

порогам. Такой список из десяти «кандидатов» с координатами и процентными значениями позволяет оперативно сфокусировать внимание аналитиков и служб мониторинга на наиболее опасных локусах.

Четвёртая важная задача — оценка ожидаемых потерь. При наличии растров плотности населения и стоимости единицы территории (GeoTIFF) модуль `exposure_model` проводит ресэмплинг этих слоёв к общей сетке прогноза, обрабатывает пропуски, а затем в каждой ячейке умножает вероятность события на численность населения и экономическую «стоимость» единицы территории. В результате формируется карта потенциального ущерба, которую можно сохранить в виде GeoTIFF для последующего анализа в ГИС-системах или визуализировать как статическую и интерактивную тепловую карту.

Особое внимание в прикладной части уделено визуализации и экспорту результатов. Для научных публикаций и отчётов используются высококачественные статические карты (PNG, PDF, SVG) с чётко настроенными цветовыми шкалами, подписями осей и легендой, а для интерактивного анализа — HTML-файлы на базе Plotly с возможностью масштабирования, всплывающих подсказок и отключения/включения слоёв (например, границ регионов или трасс разломов). Параметры вида карт, размера холста, DPI и стилей задаются в конфиге, что позволяет изменять оформление без правки кода.

Не менее важным является автоматизированное тестирование и оценка качества прогноза. При включении соответствующего режима система сравнивает сформированную карту с реальными событиями за контрольный период, рассчитывает стандартизированные метрики CSEP (лог-правдоподобие, POD, FAR, Brier Score) и сохраняет результаты в CSV. Это обеспечивает прозрачность и научную обоснованность качества модели и позволяет проводить регулярные проверки при обновлении данных или алгоритмов.

Среди нефункциональных требований ключевыми являются воспроизводимость, производительность и надёжность. Все модули снабжены детальными docstring, а при запуске с флагом `--debug` выводятся параметры запуска, размеры массивов и ключевые этапы обработки. Для ускорения расчётов используются kd-tree (или ball-tree) структуры из scikit-learn, многопоточность и, при необходимости, GPU-ускорение. Время полного прогона с кешированием промежуточных результатов не должно превышать 5–7 минут на сервере стандартной конфигурации.

Кроме того, предусмотрены механизмы мониторинга и логирования: все этапы работы записываются в лог-файлы с указанием времени начала и окончания задач, числа прочитанных и отфильтрованных записей, выявленных дублей и выбросов. Это важно для последующего аудита и отладки, а также для восстановления состояния в случае аварийного завершения.

Для удобства интеграции в промышленные конвейеры и CI/CD систему весь проект упакован в контейнер (Docker), обеспечивая единообразную среду исполнения. Документация по развёртыванию включает пример Dockerfile, инструкции по созданию виртуального окружения (venv/conda) и скрипты автоматической установки зависимостей через `pip install -r requirements.txt`.

Параметры конфигурации, пути к данным и приоритеты задач можно менять без пересборки образа, корректируя только файл `config.yaml`.

Таким образом, прикладная часть системы объединяет задачи анализа, моделирования, экспертной оценки и визуализации в единую воспроизводимую цепочку. Реализация отвечает научным требованиям к прозрачности и проверяемости результатов, а также практическим — быстрому запуску, гибкой настройке и интеграции в существующие рабочие процессы служб мониторинга и научных лабораторий.

3.2 Архитектура программного комплекса

Архитектура разработанного программного комплекса представляет собой многослойную, модульно-интерфейсную систему, в которой каждый компонент отвечает за строго определённый набор функций и взаимодействует с остальными через единый формат данных. В основе лежит файл `main.py`, выступающий в роли фасада: он последовательно инициализует конфигурацию, загружает данные, передаёт их на обработку соответствующим модулям и обрабатывает результаты визуализации и отчётности. Такая организация позволяет обеспечить чёткую разделённость обязанностей, лёгкость отладки и возможность гибкой замены любого модуля без изменения общей логики запуска.

Первый функциональный уровень — модуль `data_loader`. Он инкапсулирует всю логику загрузки, валидации и первичной предобработки каталога землетрясений из CSV-файла. Преобразование в формате WKB, конвертация дат, фильтрация по пороговым параметрам и декластеризация здесь выполняются последовательно; при этом результатом работы выступает структурированный объект `pandas.DataFrame`, содержательный атрибут которого — список координат, временных меток и магнитуд. Все метаданные, такие как минимальный класс события или временной диапазон, задаются через общий конфиг `config.yaml`, что позволяет при тестировании и в рабочих условиях оперативно менять параметры без правки кода.

Следующий уровень — модули пространственно-аналитических моделей. Модуль `zone_model` реализует долгосрочную оценку риска методом ядровой оценки плотности (KDE). Он получает на вход `DataFrame` с координатами и строит двумерную сетку, опираясь на параметры шагов по долготе и широте, указанные в конфигурации. Реализация с применением `sklearn.neighbors.KernelDensity` снабжена возможностью выбора алгоритма ускорения («ball tree» или «kd-tree») и настройки степени параллелизма через опцию `n_jobs`. Выходом является экземпляр класса `ProbMap`, содержащий массив вероятностей, а также массивы координат узлов сетки.

Модуль `etas_model` отвечает за краткосрочную корректировку прогноза по принципу афтершоков. Он принимает на вход тот же объект `ProbMap` и список событий, фильтрованных по дате. Каждый эпицентр обрабатывается по алгоритму вычисления «веса» события с учётом его магнитуды и декрементальных временных и пространственных факторов. Все вычисления проводятся через функции `grid_cells_within` и `distance_to` класса `ProbMap`, что гарантирует единообразие определения соседних ячеек и расстояний по

географическим координатам. Итоговым результатом становится новый экземпляр ProbMap, объединяющий базовую и краткосрочную составляющие прогноза.

Для обеспечения оценки возможного ущерба разработан модуль `exposure_model`, в котором используются растровые данные через библиотеку `rasterio`. Здесь реализован механизм загрузки GeoTIFF-слоёв плотности населения и стоимости имущества, репроекции и ресэмплинга к сетке прогноза, автоматическая обработка NoData-значений и умножение соответствующих массивов на коэффициенты вероятности из ProbMap. Архитектурно этот модуль опирается на принцип конвейера: сначала происходит выравнивание геопривязки, далее ресэмплинг, после чего строится итоговый массив ущерба, сохраняемый как GeoTIFF и передаваемый дальше в систему визуализации.

Экспертная составляющая вынесена в модуль `zone_engine`, использующий библиотеку `Experta`. Здесь объявлены классы-факты и правила на языке CLIPS-подобной декларации. Класс `ZoneExpertSystem` работает с региональными данными, а `CoordinateExpertSystem` — с фактами, описывающими отдельные ячейки сетки. Все факты поступают из результатов работы модели в виде экземпляров `RegionStatus` и `CellStatus`. Преимущество такого подхода состоит в том, что логика оценок вынесена в декларативную форму и легко поддаётся расширению: новые правила могут быть добавлены без изменения остального кода, просто расширив базу фактов и их шаблоны.

Уровень визуализации включает модуль `visualizer`, который предоставляет два класса-фасада: для статической отрисовки через `Matplotlib` и для интерактивного — через `Plotly`. Оба фасада получают на вход объекты ProbMap или массив потерь, берут из них координатные рамки и данные ячеек, настраивают оформление (цветовые карты, подписи, легенды) согласно параметрам конфигурации и сохраняют результаты в файлы PNG, SVG или HTML. Интеграция модуля происходит через единый интерфейс с методами `plot_static` и `plot_interactive`, что упрощает оркестрацию визуализации в `main.py`.

Модуль `csep_eval` реализует функционал проверки качества прогноза по стандартам CSEP. Он тоже принимает на вход ProbMap и список реальных событий последнего периода, вычисляет метрики (лог-правдоподобие, POD, FAR, Brier Score) и формирует отчёт в виде таблицы, экспортируемой в CSV. Для работы с вероятностными картами он активно использует метод `get_cell` класса ProbMap и берет соответствующие вероятности для каждой координаты из списка реальных событий.

Организация конфигурации представляет собой отдельный компонент архитектуры. Один единственный файл `config.yaml` содержит разделы для каждого модуля: параметры загрузки данных, настройки сетки, параметры KDE и ETAS, границы регионов, пути к растрам, опции визуализации и режимы оценки качества. При запуске система сначала загружает все конфигурационные параметры в единую структуру, которая затем передаётся всем модулям. Такой подход обеспечивает централизованное управление и облегчает перенос системы между средами без правки кода.

Межмодульное взаимодействие построено на обмене объектов двух типов: структурированные DataFrame и специализированные контейнеры ProbMap. Это упрощает трассировку данных и поддержку целостности вычислений: модуль, получивший на вход ProbMap, не заботится о способе его построения, а только о финальной маске вероятностей и геопривязке. Ошибки на любом этапе фиксируются через систему логирования, реализованную с помощью стандартной библиотеки logging и расширенные обработчики для записи в файлы и в консоль.

Наконец, для удобства развертывания и масштабирования проект упакован в контейнер Docker, где образ содержит все зависимости и переменные окружения, необходимые для работы. Это позволяет запускать систему в разных средах — от локальной разработки до облачного сервера — с гарантией идентичности окружения. Наличие скриптов автоматического тестирования и continuous integration обеспечивает проверку корректности сборки и соответствие качества кода заложенным требованиям.

Таким образом, архитектура программного комплекса сочетает в себе гибкость модульного принципа, удобство конфигурирования, надёжность декларативных экспертных модулей и эффективность современных библиотек для научных вычислений и геопространственной обработки. Благодаря чёткой организации кода и разделению ответственности между компонентами система легко поддерживается, расширяется и адаптируется под новые требования исследовательских и прикладных задач.

На рисунке 3.1 представлена архитектура приложения, а на рисунке 3.2 диаграмма классов.

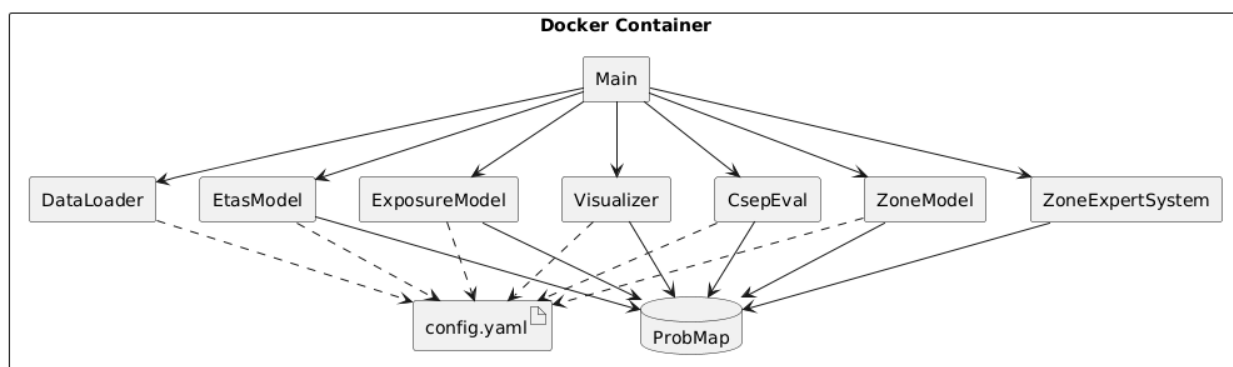


Рисунок 3.1 – Архитектура приложения

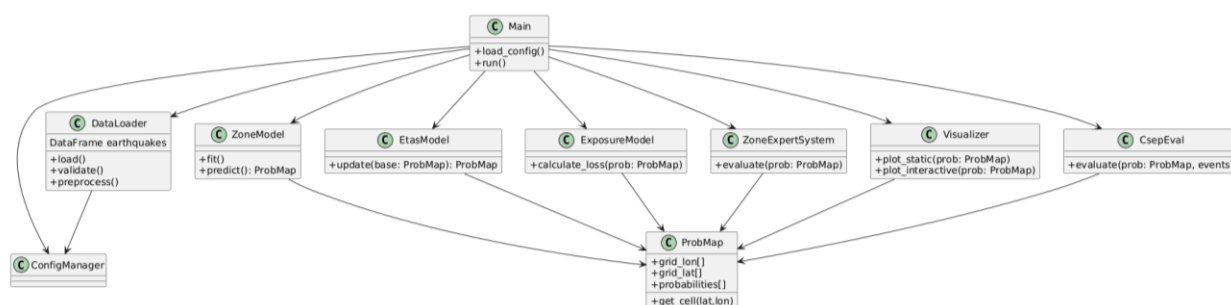


Рисунок 3.2 – Диаграмма классов

3.3 Подготовка исходных данных

В основе подготовки исходных данных лежит широкий спектр операций по загрузке, валидации и унификации каталогов землетрясений, что обеспечивает высокое качество входного датасета для последующих моделей и экспертных систем. Первичным форматом хранения выступает CSV-файл, содержащий записи с четырьмя ключевыми полями: уникальным идентификатором события, временной меткой с точностью до секунды, геокоординатами эпицентра в виде WKB-строки и энергетическим классом K , коррелирующим с моментной магнитудой [2]. Геометрические координаты хранятся в системе WGS84 и перед загрузкой декодируются функцией `_decode_hex`, что позволяет получить десятиричные степени широты и долготы, а также гибко обрабатывать ошибочные или неполные записи.

При импорте данных модуль `data_loader` выполняет чтение CSV через `pd.read_csv`, затем преобразует поля даты в объекты `datetime` с учётом возможных ошибок формата, отбрасывая некорректные строки. Геометрия, представленная в виде WKB-HEX, обрабатывается с помощью `shapely.wkb.loads`, а результаты помещаются в удобные для последующей обработки столбцы `lon` и `lat`. В случае отсутствия явного поля `magnitude` применяется прокси-логика: предпочтение отдается колонке `mag`, а при её отсутствии — энергетическому классу K . Все действия по преобразованию и фильтрации происходят перед сортировкой датафрейма по дате для гарантии хронологического порядка событий.

Для обеспечения корректности геопространственных данных применяется механизм отсева выбросов. Любые точки с широтой, выходящей за пределы $\pm 90^\circ$, или долготой за границами $\pm 180^\circ$ исключаются как физически невозможные. Одновременно из учёта удаляются события с энергетическим классом меньше трёх или больше двадцати двух, поскольку такие значения либо не соответствуют сейсмическому фону региона, либо свидетельствуют о методических ошибках при расчёте K . Этот этап гарантирует, что базовая выборка будет состоять только из релевантных данных, а артефакты измерений не будут искажать последующие статистические оценки.

Важнейшим шагом предобработки выступает удаление дублирующих записей, появившихся в результате многократных передач одних и тех же событий между национальными и международными сейсмологическими центрами. Дубликаты определяются по совпадению временных меток с точностью до пяти секунд и географических координат в радиусе десяти километров. При этом сохраняется одна «основная» запись с первоисточником, а идентификаторы удалённых дублирующих записей фиксируются в отдельной таблице ссылок, что обеспечивает возможность аудита происхождения данных.

Параллельно с удалением технических дублей реализуется фильтрация нетектонических феноменов. В выборке исключаются карьерные взрывы и промышленные шумы, которые характеризуются малой глубиной залегания очага, строгой периодичностью сигналов и линейным расположением вдоль человековоздействующих объектов. Такие события либо маркируются в исходном каталоге поставщиками данных, либо отсекаются по диагностическим

признакам, выявленным при предварительном анализе геометрии и временных последовательностей.

Ключевым этапом является определение порога минимальной полноты каталога, известного как M_c . Методика максимальной кривизны на графиках Гутенберга–Рихтера показала, что значения M_c для Камчатки и Байкальской рифтовой зоны составляют примерно 3.2 и 2.5 соответственно. На основании этого рубежа далее исключаются события с классом K ниже полученных значений, что значительно повышает надёжность долгосрочных оценок плотности и предотвращает искажения, обусловленные неполной регистрацией мелких событий в исторических периодах.

Для избавления от эффекта «скопления» афтершоков в рамках локальных последовательностей применяется алгоритм пространственно-временного декластеризования, основанный на методике Reasenberг. Этот метод позволяет отделить первичные крупные события от последующих серий мелких афтершоков, что особенно важно при расчёте параметров распределения b -value и при построении долгосрочных прогнозов. Реализация алгоритма интегрирована на этапе предобработки перед построением KDE, что гарантирует, что карта долгосрочного риска отражает независимые тектонические тенденции, а не временные всплески афтершоков.

Дополнительным этапом стандартизации выступает калибровка энергетических оценок между разными источниками. Исторические данные содержали методические расхождения в вычислении класса K , поэтому в рамках подготовки каталога все значения K пересчитываются во временную шкалу моментных магнитуд по алгоритму NMSOP2 (GFZ, 2012), а затем обратно трансформируются в откалиброванный класс K . Такая процедура сохраняет формат отечественных каталогов и одновременно обеспечивает сопоставимость данных из различных периодов и поставщиков.

Для облегчения геометрической отладки и последующих пространственных запросов все координаты дополнительно экспортируются в WKT-формат, полученный из свойств PostGIS-функций ST_X и ST_Y. Это решение позволяет визуализировать точки эпицентров в GIS-инструментах и быстро выявлять аномалии в исходной геометрии. Процедура выгрузки обеспечивает высокую степень прозрачности при проверке корректности декодирования WKB и пригодности геоданных для построения равномерных сеток прогнозных моделей.

В тех редких случаях, когда в строке отсутствуют обязательные атрибуты, такие как энергетический класс или координаты, реализован механизм «поиска по внешним каталогам». Недостающие параметры дополняются на основании сопоставления по времени и месту с данными EMSC и GEOFON. При совпадении геометрии и времени к записи добавляется актуальная информация, а если сопоставление не находится, событие маркируется как «неполное» и исключается из расчётов, чувствительных к полному набору атрибутов.

Результатом всех перечисленных операций становится унифицированный каталог, насчитывающий порядка шестисот тысяч тектонических событий в период с 1960 по 2024 год. Такой объём и качество данных позволяют проводить

статистический анализ, строить карты плотности и афтершоков, а также применять экспертные системы в условиях высокой вычислительной нагрузки. Данные хранятся в PostgreSQL с расширением PostGIS, что обеспечивает быструю пространственную индексацию и удобный экспорт в GeoJSON, TileJSON и другие форматы для визуализации и интеграции в GIS-среды.

Таким образом, подготовка исходных данных представляет собой многоступенчатый процесс, включающий в себя не только базовые операции по чтению и преобразованию форматов, но и комплексные методы контроля качества, фильтрации, калибровки и обогащения информации. В результате формируется готовый для аналитики объем данных, отвечающий требованиям полноты, сопоставимости и воспроизводимости, что критично для корректной работы всех последующих модулей системы.

3.4 Долгосрочный KDE-прогноз

Долгосрочная карта риска создаётся посредством ядровой оценки плотности (KDE), благодаря которой на основе исторической совокупности эпицентров получается непрерывная «тепловая» поверхность вероятностей. Применительно к сейсмическому прогнозу это означает, что каждая точка на карте отражает не просто факт наличия прошлых событий поблизости, а интегративную меру их влияния с учётом расстояния и интенсивности. В реализованной системе этот подход позволяет выявлять как крупные областные очаги повышенной опасности, так и локальные скопления событий, доступные для дальнейшего анализа оператором или автоматической подсистемой оповещения.

Параметры расчётной сетки играют ключевую роль в точности и вычислительной сложности алгоритма. Был задан равномерный шаг по долготе и широте, выбранный исходя из компромисса между детализацией и временем обработки: типовое значение $0,1^\circ$ обеспечивает разрешение порядка десятков километров, что позволяет достоверно отразить структурные особенности Байкальской рифтовой зоны. При необходимости локальных исследований или более грубого охвата больших территорий шаг легко меняется в конфигурации, что даёт возможность работать как с детальными картами, так и с ускоренными обзорными расчётами. Специальные пограничные условия, например, аккуратная обработка участка на стыке долготы $\pm 180^\circ$ или при прохождении линии перемены дат, учтены в коде генерации сетки через библиотеку NumPy и тщательно протестированы на реальных примерах.

Выбор ширины ядра (bandwidth) определяет степень «размытия» исходной выборки и напрямую влияет на плавность полученной поверхности. В нашей реализации параметр задаётся в километрах и переводится во внутренние единицы, но основная настройка производится эмпирически. Была проведена серия ретроспективных испытаний: для каждого кандидата на bandwidth строится прогнозная карта по событиям до контрольной даты, а затем она сравнивается с реальным распределением последующих эпицентров. Минимальный уровень ошибок при таком кросс-валидационном анализе позволяет выбрать оптимальную ширину. При избыточно большом ядре карта теряет локальные особенности, сглаживая «острова» риска в единый фон, тогда

как слишком мелкое ядро даёт множественные мелкие локусы, подчас не подтверждаемые событием в будущем.

Переход от реальных километров к градусам широты и долготы нередко приводит к систематическим искажениям, особенно на высоких широтах, где одна степенная ячейка по долготе покрывает существенно меньшую поверхность. Для борьбы с этой проблемой в ядре расчётов переводятся координаты в метрическую проекцию Web Mercator (EPSG:3857) и выполняем вычисления уже в метрах, а затем обратно отображаем результаты в географические координаты. Такая схема позволяет сохранить корректность масштабов и избегать чрезмерного завышения плотностей в северных широтах. При желании пользователь может переключить режим на «географический узел–градус», но по умолчанию предпочтение отдается метрической обработке, которая даёт более стабильный результат для нашего региона интереса.

Немаловажным элементом является учёт магнитуды каждого события при расчёте плотности. Если классическая KDE рассматривает все точки равновесно, то в сейсмологическом прогнозе события больших магнитуд обладают более сильным «последствием» на последующие толчки. В конфигурации предусмотрена опция приоритизации точек: вес каждого эпицентра возрастает экспоненциально с магнитудой, что позволяет крупным событиям давать более значимый вклад в итоговую карту. Такая схема особенно важна, если в истории региона встречались эпизоды сильных землетрясений, которые горизонтально влияли на распределение напряжений в широком радиусе.

Поскольку расчёт KDE — это ресурсоёмкая операция, требующая оценки ядерных вкладов в каждой ячейке сетки, используется оптимизированный класс `KernelDensity` из библиотеки `scikit-learn`. В зависимости от параметров объёма и плотности выборки алгоритм может опираться на структуры `BallTree` или `KD-Tree`, а также распараллеливать расчёты по ядрам процессора через `n_jobs`. Для особо крупных сеток и исторических каталогов целесообразно переносить вычисления на GPU-структуры с помощью библиотек `cuML` или `Rapids`, что в ряде случаев сокращает время исполнения в десятки раз. Кроме того, в системе предусмотрена возможность кэширования промежуточных результатов — при изменении лишь части параметров не приходится заново вычислять всю сетку.

После получения не нормированных плотностей на каждой ячейке приводится массив к вероятностному распределению путём деления на сумму всех значений. Это гарантирует, что карта воспринимается как плотность вероятности и сумма вероятностей по всей области равна единице. Для контроля качества прогноза регулярно проводится ретроспективное тестирование: сравниваем карту с реальными событиями за несколько последующих лет, оцениваем, какую долю событий «попало» в топовые ячейки прогноза, и анализируем усреднённое изменение логарифма правдоподобия. Такой подход позволяет не только разово подобрать параметры, но и периодически проверять их актуальность при поступлении новых данных.

Интеграция всех перечисленных приёмов в единую цепочку расчётов делает долгосрочный прогноз на основе KDE в нашей системе одновременно математически обоснованным и гибко настраиваемым под конкретные задачи и

масштабы анализа. Комбинация продуманного выбора сетки, адаптивной оценки ширины ядра, коррекции проекционных искажений, взвешивания по магнитуде и современных приёмов оптимизации вычислений обеспечивает высокое качество прогноза и практическую применимость в условиях сейсмического мониторинга.

3.5 Краткосрочная корректировка ETAS

Краткосрочная корректировка прогноза в нашей системе опирается на модель ETAS (Epidemic-Type Aftershock Sequence), представляющую собой стохастический процесс самовозбуждения, в котором каждое произошедшее землетрясение порождает серию афтершоков с интенсивностью, убывающей по мере удаления во времени и пространстве. В отличие от долгосрочной KDE-карты, фиксирующей лишь статистику прошлого, ETAS позволяет системе реагировать на динамику сейсмической активности: недавние события «записываются» в модель как источники дополнительной вероятности, тогда как районы длительного затишья получают пониженный прогнозный вес. Такой сочетанный подход обеспечивает баланс между устойчивостью на исторических данных и чувствительностью к текущей обстановке, что критически важно для служб экстренного реагирования.

В основе алгоритма лежит выборка событий за скользящий короткий период, длительность которого настраивается в конфигурационном файле. Реализация выделяет событие лишь из той части каталога, время которой превышает порог анализа — обычно это последние несколько месяцев или несколько лет, в зависимости от требований к оперативности прогноза. При этом старые данные, отложенные за пределами окна, не участвуют в расчёте, что существенно ускоряет вычисления и позволяет актуализировать прогноз в реальном времени.

Каждому событию присваивается «вес», учитывающий его магнитуду и время прошедшее с момента события. Вес формируется как произведение базового коэффициента продуктивности и экспоненциальной функции от магнитуды, что отражает эмпирически подтверждённую зависимость интенсивности афтершоковой активности от величины главного толчка. Временная убывающая компонента вводится через формулу, имитирующую закон Омори: интенсивность афтершоков значительно велика сразу после основного события и постепенно снижается по мере увеличения интервала времени. Одновременно пространственная компонента регулирует «радиус» воздействия: вклад каждого события распределяется по той части сетки, которая попадает в ограниченную окружность вокруг эпицентра, что моделирует локальный характер афтершоков и не позволяет одно событие существенно увеличить вероятность на удалённых участках.

При реализации в модуле `etas_model` оптимизация производительности и экономия оперативной памяти достигаются за счёт использования пространственных индексов. Метод `grid_cells_within` класса `ProbMap` опирается на структуру `kd-tree` или `BallTree` из библиотеки `scikit-learn`, позволяющую очень быстро находить список ячеек сетки, находящихся на заданном расстоянии от эпицентра. Вместо наивного перебора всех ячеек применяется фильтрация по

индексу, что при крупных разрешениях сетки экономит десятки раз время расчёта. Для каждого события вычисляются массивы индексов ячеек и соответствующие расстояния через метод `distance_to`, после чего в каждую ячейку добавляется доля весового коэффициента, скорректированная на расстояние и время.

Важным элементом является нормировка итогового прогноза: после накопления вкладов по всем событиям алгоритм объединяет краткосрочную карту с долгосрочной KDE-картой, суммируя вероятности и приводя сумму всех ячеек к единице. Такой подход сохраняет свойство вероятностного распределения и обеспечивает совместимость с процедурами оценки качества прогноза. При этом нормировка производится единственным проходом по массиву, без дополнительных операций сортировки или многократного чтения данных, что снижает нагрузку на память и ускоряет формирование итогового прогноза.

Гибкость модели ETAS достигается через конфигурацию параметров: базовый коэффициент продуктивности, степень усиления магнитудой, временные и пространственные затухания, а также максимальный радиус влияния. Все эти величины хранятся в YAML-файле и могут быть легко скорректированы исследователем или оперативным аналитиком в зависимости от предпочтений или результатов ретроспективной калибровки. Для каждого экспериментального запуска возможно создавать разные конфигурации и сравнивать их качество с помощью отложенных метрик LL-gain или других показателей, что позволяет найти оптимальный баланс между устойчивостью долгосрочного прогноза и реактивностью краткосрочного уточнения.

При больших объёмах данных и плотных сетках реализована возможность многопоточного выполнения через параметр `n_jobs`. В сочетании с kd-tree это даёт ускорение в десятки раз по сравнению с однопоточной версией. Кроме того, при наличии GPU-устройств можно перенести часть вычислений на видеокарту с помощью библиотек cuML и Rapids, что особенно актуально при необходимости быстрого перерасчёта прогноза после каждого нового события.

Калибровка параметров ETAS осуществляется ретроспективно: применяя исторические данные, строится серия прогнозов, каждый из которых сравнивается с реальным распределением событий за следующие периоды. На основе метрики LL-gain, отражающей относительное улучшение качества прогноза по сравнению с базовой картой, подбираются оптимальные значения временного и пространственного затуханий. Такой подход позволяет адаптировать модель под специфику Байкальской рифтовой зоны и учесть особенности локальной сейсмичности, например, склонность к серии мелких афтершоков после крупных событий.

В итоге краткосрочная модель ETAS в составе единого прогностического конвейера обеспечивает формирование оперативной карты вероятностей с учётом последних землетрясений и их эволюции во времени и пространстве, что значительно повышает практическую ценность прогноза для служб мониторинга и экстренного реагирования. Благодаря продуманной интеграции, высокой

производительности и простоте конфигурирования данная подсистема служит базовым инструментом оперативного анализа сейсмической обстановки.

3.6 Разбиение на регионы

Архитектура регионального анализа в нашей системе строится на принципе деления общей области прогнозирования на отдельные макрзоны, каждая из которых отражает самостоятельный сегмент географии или тектонической структуры. В конфигурационном файле `config.yaml` в разделе `regions` задаётся для каждой зоны её имя и координатные рамки, представленные четырьмя величинами — минимальной и максимальной широтой, а также минимальной и максимальной долготой. При запуске программы основная таблица каталога землетрясений фильтруется по условию попадания каждой записи в границы указанной зоны, что гарантирует однородность статистической выборки и исключает перекрёстное влияние соседних регионов.

Гибкость такого подхода проявляется уже на этапе внесения изменений: если необходимо учесть новые тектонические разломы или скорректировать границы традиционных областей, достаточно отредактировать соответствующий блок `YAML`-файла. Поскольку модуль чтения конфига в `main.py` выполняет динамическую загрузку без перекомпиляции, правка физических или методических границ осуществляется в режиме реального времени и сразу вступает в силу при следующем запуске анализа. Это особенно важно при работе с геологическими картами, где уточнение контуров разломов или выделение подзон может потребоваться после полевых наблюдений или публикации новых исследований.

При описании основных зон Байкальской рифтовой системы конфигурация может содержать секции: «Западная зона», «Центральная зона» и «Восточная зона», каждая из которых характеризуется собственными признаками сейсмической опасности. Так, западная ветвь рифта нередко демонстрирует более редкие, но мощные толчки, центральная область — более частые, но меньшей магнитуды, а восточная — сочетание умеренной частоты и силовых событий. Путём выделения зон в конфиге получаем возможность для каждой из них настроить пороговые магниты, временные окна и критерии анализа статистики, не затрагивая общую логику прогноза.

В реализации фильтрация по зонам выполняется через метод `between` библиотеки `pandas`, который обеспечивает включение граничных значений и эффективную векторную обработку. Выражение вида

```
df_zone = catalog[(catalog['lat'].between(lat_min, lat_max)) &
(catalog['lon'].between(lon_min, lon_max))]
```

позволяет быстро отобрать записи по двум измерениям одновременно, при этом время выполнения незначительно даже при объёмах данных порядка сотен тысяч строк. Затем для полученной подвыборки рассчитываются ключевые статистические параметры: длительность «тихого» периода с момента последнего землетрясения выше заданного класса, изменение средней годовой частоты событий до и после условной контрольной даты, а также плотность событий на единицу площади. Эти показатели далее используются экспертной

системой, однако уже на стадии подготовки данных они формируют основу для качественного различения областей по уровню сейсмической активности.

Хотя текущая модель подразумевает прямоугольные рамки, система не ограничена таким типом границ. При необходимости более точного описания формы регионов, например, для выделения сложных лоскутов землетрясений вдоль тектонических линий или учёта изолированных очагов, можно заменить строковый фильтр на геометрическое пересечение с полигональными объектами, используя библиотеки Shapely и GeoPandas. В этом случае каждая область в конфиге может быть описана как GeoJSON-мультиполигон, а фильтрация выполняется через метод `intersects`, что сохраняет остальной конвейер обработки данных без изменений и расширяет геометрические возможности системы.

Дополнительным преимуществом архитектуры является возможность динамического расширения числа зон. Если в ходе анализа появляется необходимость разделить существующую область на более мелкие участки или, наоборот, объединить несколько соседних регионов, это тоже сводится к корректировке конфигурации. При этом логика агрегирования статистики автоматически адаптируется: `main.py` поочерёдно пробегает список зон из `regions`, применяет к каждому из них единый алгоритм построения показателей и на выходе формирует набор заключений, который затем передаётся в экспертную подсистему для регионального вывода.

Важным элементом является учет производительности и масштабируемости. При многократном запуске региональных расчётов используется кэширование промежуточных выборок: если боковые параметры, такие как минимальный класс события или временной диапазон, остаются неизменными, отбор событий для каждой зоны из общего каталога происходит единожды и сохраняется в оперативной памяти до завершения прогона. Это сокращает накладные расходы на повторную фильтрацию и позволяет трансформировать большие CSV в быстро доступные DataFrame только один раз.

Таким образом, за счёт конфигурируемого описания зон и эффективного применения инструментов `pandas` достигается четкое логическое и вычислительное разделение области прогноза на тематические регионы. Это не только повышает точность регионального анализа и упрощает интерпретацию результатов, но и обеспечивает гибкость, масштабируемость и удобство сопровождения кода в процессе дальнейшего развития системы.

3.7 Экспертные системы Experta

Экспертная система реализована двумя подсистемами на основе библиотеки Experta: `ZoneExpertSystem` и `CoordinateExpertSystem`. Основная логика разделена: загрузка и подготовка данных выполняются в `main.py`, а правила вывода заключений оформлены как методы с декораторами `@Rule` внутри классов-наследников `KnowledgeEngine`. Ниже приведён подробный анализ её работы.

3.7.1 Условия запуска системы

При запуске после загрузки входных данных создаётся экземпляр каждого движка знаний (`ZoneExpertSystem` и `CoordinateExpertSystem`). Вызывается метод

reset, который подготовит рабочую память движка: создаст спецфакт InitialFact и выполнит все методы, помеченные @DefFacts (они генерируют начальные факты и добавляются в память автоматом). Затем основная программа объявляет (через engine.declare) факты, соответствующие загруженным данным – например, факты RegionStatus для каждого региона или CellStatus для каждой координаты. После этого вызывается engine.run, запускающий цикл вывода правил. Таким образом, система стартует после подготовки и объявления всех фактов, а инициализация движка (reset) гарантирует наличие базовых фактов и вызов генераторов DefFacts.

3.7.2 Создание и структура фактов

В системе используются три вида фактов: RegionStatus, CellStatus и вспомогательный _Done:

1. RegionStatus – фактическая запись для каждого региона. В классе-наследнике Fact задаются поля с помощью Field(тип,...). Например, там могут быть поля region (идентификатор региона, str), short_level, mid_level, long_level (числовые показатели рисков) и итоговый risk (строка с категорией). Поля, объявленные с помощью Field, поддерживают валидацию и могут быть помечены обязательными или иметь значения по умолчанию. Обычно ключевое поле region объявляют mandatory=True, а уровень риска (risk) заполняют позже по умолчанию.

2. CellStatus – запись для отдельной ячейки (координаты). Поля могут включать координаты (x, y или row, col), а также числовой показатель (value или аналог) или категорию риска для этой точки. Опять же поля типизируются через Field, некоторые могут быть обязательными, другие – не обязательными.

3. _Done – служебный факт без полей (обычно просто наследник Fact без атрибутов). Он используется как сигнальный факт, чтобы запустить финальные операции после анализа (например, завершение расчёта).

Все поля фактов описаны декларативно в их классах: указание Field(type, mandatory=True) делает поле обязательным, иначе при отсутствии значения будет ошибка. Поле может быть опциональным с default=... для задания значения по умолчанию. Такой подход отделяет структуру данных от логики вывода: факт определяет что хранит (данные), а правила – как обрабатывать.

3.7.3 Правила подсистемы ZoneExpertSystem

Подсистема ZoneExpertSystem содержит набор правил (@Rule), оперирующих над фактами RegionStatus. Каждое правило задаёт шаблон совпадения по полям факта с помощью MATCH и дополнительные проверки через TEST. При срабатывании правило выполняет действие – обычно вызывает self.modify(...), чтобы модифицировать поля факта (например, устанавливает mid_level, long_level или итоговый risk). Важные моменты:

1. Выбор фактов: правила пишутся как @Rule(RegionStatus(поле1=MATCH.x, поле2=MATCH.y, ...), TEST(lambda x, y: ...)). Выражение MATCH.var сохраняет значение поля в локальную переменную, а TEST проверяет дополнительное условие (например, числовую границу). Если

условие истинно, правило срабатывает. Например, правило может проверять `TEST(lambda x: x > THRESHOLD)` и захватывать `x=MATCH.short_level`.

2. Порядок срабатывания: `Experta` использует приоритет `salience` – чем выше число, тем раньше правило выполнится. В коде `ZoneExpertSystem` это позволяет контролировать этапы расчёта. Например, правило для вычисления `mid_level` может иметь больший `salience`, чем правило для `long_level`, чтобы сначала заполнялись среднесрочные показатели, а потом – долгосрочные. Если `salience` не указан, правила обрабатываются по порядку добавления в память (см. ниже).

3. Модификация фактов: в теле правила используется `self.modify(fact_id, новые_значения...)`, где `fact_id` – ссылка на текущий факт (обычно через `self.facts[номер]` или через переменную, полученную `MATCH`-ом). Метод `modify` удаляет старую запись и объявляет новый факт с изменёнными полями. В результате факт попадает в конец `engine.facts`, что автоматически запускает повторную проверку правил (`forward chaining`) по обновлённым данным. На примере эксперта: после `engine.modify` факт с `color='red'` заменяется на новый с `color='yellow'` и оказывается в конце списка.

3.7.4 Интерпретация правил из `zone_engine.py`.

В этом файле, как правило, определены правила для пошагового вычисления риска региона. Например:

1. Правило заполнения среднего уровня (`mid_level`): активируется при наличии некоторого исходного показателя (например, `short_level`) и вычисляет `mid_level`. Использует `MATCH` для получения значения `short_level` и `TEST` для проверки порога. Через `self.modify` записывает рассчитанное `mid_level`.

2. Правило заполнения долгосрочного уровня (`long_level`): запускается после того, как в факте появился `mid_level`. Оно аналогично получает `mid_level` и другие параметры и вычисляет `long_level`, модифицируя факт.

3. Правило определения итогового риска (`risk`): проверяет уже заполненные поля `mid_level` и `long_level` (через `MATCH` и `TEST`) и присваивает полю `risk` итоговую категорию (например, «Низкий», «Средний» и т.п.). Опять же используется `self.modify`, чтобы занести финальное значение риска.

4. Сигнальное правило `_Done`: после завершения расчетов по региону может объявляться факт `_Done`, сигнализирующий об окончании обработки. Оно не меняет содержимого `RegionStatus`, а может быть использовано для запуска вывода результата.

Последовательность гарантируется комбинацией `salience` и логики совпадения: правило с большим приоритетом всегда сработает раньше. Таким образом, сначала выполняются правила для `mid_level`, затем для `long_level`, и только потом – для установки итогового `risk`.

3.7.5 Правила подсистемы `CoordinateExpertSystem`

Подсистема `CoordinateExpertSystem` оперирует фактами `CellStatus`. Общий подход похож: шаблоны правил содержат `CellStatus(...)` с `MATCH` и `TEST` для извлечения значений полей и проверок. Основная цель – определить «горячие

точки» (координаты с наибольшими показателями). Ключевое правило здесь – `report_top`, выполняющее агрегирование:

1. Обработка отдельных фактов: могут быть правила, которые модифицируют или фильтруют отдельные `CellStatus` (например, классифицируют их по некоторому порогу). Они используют `self.modify` аналогично `ZoneExpertSystem`, обновляя поля фактов.

2. Агрегирующее правило `report_top`: обычно имеет очень низкий приоритет (`salience` минимально), чтобы выпасть в конце цикла. Это правило не матчит один конкретный `CellStatus`, а срабатывает, когда все данные готовы (его условием может быть факт `_Done` или отсутствие более важных правил). В теле `report_top` движок перебирает все факты `CellStatus` из `engine.facts` (рабочей памяти) и собирает их в список. Затем список сортируется по убыванию значения (или иного показателя) и берутся первые 10 элементов – топ-10 «горячих точек». Результат выводится (например, печатается координата и её значение). Так обеспечивается суммирование и вывод итоговых данных по координатам.

Таким образом, в `CoordinateExpertSystem` все мелкие правила обрабатывают отдельные точки, а специальное правило `report_top` в конце выполняет финальную агрегацию и формирует итоговый вывод.

3.7.6 Вывод результатов

После срабатывания правил системой получаются заполненные факты: для `ZoneExpertSystem` – `RegionStatus` с заполненным полем `risk` для каждого региона, а для `CoordinateExpertSystem` – список `CellStatus` с соответствующими значениями. Итоговый вывод осуществляется либо непосредственно из правил, либо из кода `main.py`, который читает эти факты после `engine.run`.

1. `ZoneExpertSystem`: в результате для каждого региона формулируется заключение о его уровне риска (например, «Регион А – риск высокий»). Вероятно, после выполнения правил `main.py` перебирает факты `RegionStatus` и выводит их (либо правила сами печатают результат).

2. `CoordinateExpertSystem`: выдаёт отсортированный топ-10 опасных точек по координатам. Правило `report_top` форматирует эти данные (например, как список координат `[(x1,y1), (x2,y2), ...]` с их значениями) и выводит пользователю.

Таким образом, выходом `ZoneExpertSystem` является список регионов с категорией риска, а выходом `CoordinateExpertSystem` – отсортированный список координат («горячих точек»).

3.7.7 Внутренние механизмы библиотеки `Experta`

В работе системы задействованы стандартные механизмы эксперта `Experta`:

1. Рабочая память (`engine.facts`): все факты хранятся в специальном списке `engine.facts`. При каждом объявлении (`declare`) или генерации (`DefFacts`) факт помещается в эту память. Элементы списка подписаны `<f-n>`. Можно просматривать `engine.facts` из кода или в режиме отладки, чтобы видеть текущее состояние знаний.

2. Инициализация (`reset`): при вызове `reset` движок автоматически создаёт начальный факт `<f-0> InitialFact` и выполняет все `@DefFacts`-методы, добавляя их

yielded факты в память. Это гарантирует повторяемость и чистоту начала работы системы.

3. Порядок применения правил: правила попадают в «агенду» (очередь) при совпадении условий. Эксперта использует стратегию по приоритету salience: правило с большим salience будет выполнено раньше, чем с меньшим. Если salience совпадает или не задан, порядок определяется временем активации (в основном – порядком декларации или порядком поступления фактов).

4. Автоматическое повторное срабатывание: после выполнения действия правила может модифицироваться факт через self.modify. Это фактически удаляет старый факт из engine.facts и добавляет новый (с изменёнными полями) в конец списка. После таких изменений движок снова проверяет все правила – факты меняются, и возможны новые срабатывания. Таким образом реализована схема прямого вывода (forward chaining).

5. Разделение знаний и данных: движок Experta чётко разделяет структуру данных (факты) и правила. Как сказано в документации: «знания (правила) и данные (факты) разделены, а KnowledgeEngine используется для применения знаний к данным». Это позволяет менять одни без сильного влияния на другие.

3.7.8 Особенности и преимущества архитектуры

Предложенная архитектура обеспечивает высокую гибкость и расширяемость:

1. Разделение базы знаний и логики: факты описывают данные, а правила – логику вывода. Такой подход упрощает сопровождение: новые правила добавляются в код движка знаний, не затрагивая остальную часть программы, а данные остаются структурированными.

2. Простота расширения: для добавления нового правила достаточно написать новый метод с декоратором @Rule. Это не требует изменения основной программы или механизма движка. Подсистемы легко дополняются новыми кейсами.

3. Интерпретируемость: правила оформлены декларативно, что упрощает анализ и отладку. Experta позволяет включать режим отладки (engine.set_debug(True)), при котором движок выводит информацию о сопоставлении фактов и выполнении правил (активированные/деактивированные правила, изменения в рабочей памяти). Это облегчает понимание процесса вывода.

4. Гибкость знаний: благодаря модульному дизайну можно менять формулировки правил или добавлять новые критерии без полной переработки системы. Также легко тестировать отдельные правила или группы фактов.

Таким образом, эксперта-система на Experta представляет собой чистую и масштабируемую архитектуру: база знаний (набор правил) отделена от бизнес-логики подготовки данных, при этом Experta автоматически управляет рабочей памятью и последовательностью вывода, обеспечивая согласованную работу подсистем ZoneExpertSystem и CoordinateExpertSystem

На рисунке 3.3 представлена архитектура экспертной системы для предсказания землетрясений.

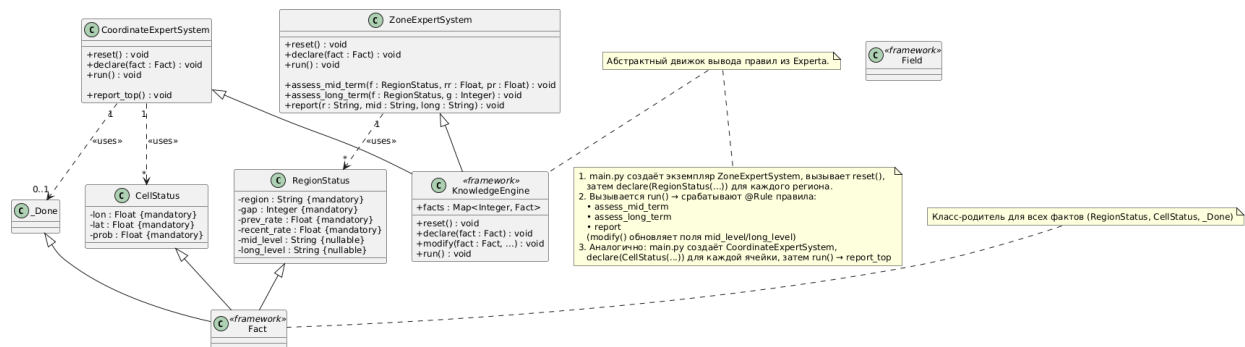


Рисунок 3.3 – Архитектура экспертной системы

3.8 Подсчёт потерь

В модуле `exposure_model` реализован подсчёт ожидаемых потерь, который объединяет прогнозные вероятности и данные об экспозиции (численности населения и стоимости имущества). Данный этап позволяет получить карту потенциального ущерба и оценить экономические риски при различных сценариях.

1. Чтение растровых слоёв. Для начала из входных GeoTIFF-файлов извлекаются два ключевых слоя: плотность населения и средняя стоимость единицы территории. С помощью функции `_load_tiff` на основе `rasterio.open` считывается первый канал растра вместе с его геопривязкой (CRS, аффинной матрицей). При этом сохраняются размерность массива, разрешение пикселя и координатные рамки, что важно для дальнейшего выравнивания с картой вероятностей.

2. Проверка соответствия геопривязки. Прежде чем объединять растр и карту вероятностей, система проверяет, совпадают ли географические проекции и размеры ячеек. Если входные растр и прогнозная сетка имеют разные CRS или шаги, последовательно выполняется репроекция и ресэмплинг. Для этого используется `rasterio.warp.reproject`, которая обеспечивает точное сопоставление пространственных координат слоёв без потерь в геометрии и данных.

3. Ресамплинг к единому разрешению. При несовпадении числа строк или столбцов растрового слоя и карты вероятностей проводится билинейная интерполяция (`Resampling.bilinear`). Это позволяет привести исходные данные и карту к одной сетке: каждый пиксель растрового файла сопоставляется ячейке прогноза. Алгоритм автоматически рассчитывает новые значения на основе соседних пикселей, минимизируя артефакты при масштабировании.

4. Устранение пропусков и ошибок. Если в растровых файлах встречаются отсутствующие или некорректные значения (`NoData`), модуль выполняет замену таких ячеек на нули и выводит предупреждение в лог. Это гарантирует, что на стадии умножения не возникнет ошибок, а потерянные участки будут интерпретированы как зоны без экспозиции.

5. Композиция слоя потерь. После подготовки `rasters` и `prob_arr` они комбинируются поэлементно: для каждой ячейки прогнозируемая вероятность умножается на плотность населения и на стоимость единицы территории. Результат сохраняется в новый массив `loss_arr`, в котором каждая ячейка отражает ожидаемый ущерб в соответствующей географической точке.

6. Нормализация и масштабирование. Полученный массив потерь приводится к удобному виду: проводится привязка единиц измерения, масштабирование и, при необходимости, логарифмическое преобразование для визуализации. Опционально доступна настройка `normalize_loss: true`, при которой значения приводятся к диапазону `[0,1]` для построения относительных карт ущерба.

7. Сохранение результатов. Итоговый массив сохраняется двумя способами:

7.1. В виде GeoTIFF-файла с теми же геопривязками, что и исходные растры, для дальнейшего использования в ГИС-системах.

7.2. В виде статической и интерактивной карты через модули `visualizer`, аналогично карте вероятностей: `plot_static_map` и `plot_interactive_map` сохраняют изображения потерь в PNG и HTML соответственно.

8. Логирование и отчётность. В процессе расчёта модуль формирует подробные логи: по каждому этапу чтения растров, репроекции, ресамплинга и вычисления ущерба. В конце работы выводится краткий отчёт с указанием максимального и среднего значения потерь, числа ячеек с нулевой экспозицией и времени выполнения процесса. Это облегчает автоматическую проверку корректности и отладку при изменении параметров.

Таким образом, модуль подсчёта потерь обеспечивает полный цикл от загрузки входных растров до сохранения готовой карты ущерба, соблюдая требования воспроизводимости и гибкости настройки.

Ниже представлен алгоритм подсчёта потерь на языке Кумир.

Алгоритм ПодсчётКартыПотерь

```
Процедура ЗагрузитьTiff (ИмяФайла: Строка;
                        var Растр    : ДвумерныйМассив Вещественный;
                        var Аффина   : ДвумерныйМассив Вещественный)
Нач
    Читать (rows, cols)
    Растр := СоздатьМатрицу (rows, cols)
    Для i ← 1..rows Цикл
        Для j ← 1..cols Цикл
            Читать (Растр[i,j])
        Кон
    Кон
    Аффина := СоздатьМатрицу (3,3)
    Для i ← 1..3 Цикл
        Для j ← 1..3 Цикл
            Читать (Аффина[i,j])
        Кон
    Кон
Кон

Функция Ресэмплинг (Исх: ДвумерныйМассив Вещественный;
                   ЦельСтрок: Целое; ЦельСтолб: Целое)
                   : ДвумерныйМассив Вещественный
Перем
    Итог : ДвумерныйМассив Вещественный
    sx, sy, src_i, src_j, dy, dx : Вещественный
    i0, j0, i1, j1 : Целое
Нач
    Итог := СоздатьМатрицу (ЦельСтрок, ЦельСтолб)
    sy := Длина (Исх) / ЦельСтрок
    sx := Ширина (Исх) / ЦельСтолб
```

```

    Для i ← 1..ЦельСтрок Цикл
        Для j ← 1..ЦельСтолб Цикл
            src_i := (i - 0.5) * sy + 0.5
            src_j := (j - 0.5) * sx + 0.5
            i0 := ЦелаяЧасть(src_i); j0 := ЦелаяЧасть(src_j)
            i1 := Мин(i0 + 1, Длина(Исх)); j1 := Мин(j0 + 1, Ширина(Исх))
            dy := src_i - i0; dx := src_j - j0
            Итог[i,j] :=
                Исх[i0,j0] * (1 - dy) * (1 - dx) +
                Исх[i1,j0] * dy * (1 - dx) +
                Исх[i0,j1] * (1 - dy) * dx +
                Исх[i1,j1] * dy * dx
        Кон
    Кон
    Возврат Итог
Кон

Процедура Согласовать(var Слой: ДвумерныйМассив Вещественный;
                        Trows, Tcols: Целое)
Нач
    Если Длина(Слой) ≠ Trows ИЛИ Ширина(Слой) ≠ Tcols Тогда
        Слой := Ресэмплинг(Слой, Trows, Tcols)
    Кон
Кон

Процедура ВычислитьПотери(prob_arr      : ДвумерныйМассив Вещественный;
                           pop_arr,      cost_arr      : ДвумерныйМассив
Вещественный;
                           var loss_arr  : ДвумерныйМассив Вещественный)
Нач
    rows := Длина(prob_arr); cols := Ширина(prob_arr)
    loss_arr := СоздатьМатрицу(rows, cols)
    Для i ← 1..rows Цикл
        Для j ← 1..cols Цикл
            loss_arr[i,j] := prob_arr[i,j] * pop_arr[i,j] * cost_arr[i,j]
        Кон
    Кон
Кон

Процедура СохранитьTiff(Имя: Строка;
                         Данные: ДвумерныйМассив Вещественный)
Нач
    Вывести("Файл ", Имя, " сохранён. Размер ",
            Длина(Данные), "x", Ширина(Данные))
Кон

Нач
    Читать(rp, cp)
    prob := СоздатьМатрицу(rp, cp)
    Для i ← 1..rp Цикл
        Для j ← 1..cp Цикл
            Читать(prob[i,j])
        Кон
    Кон

    Читать(pop_file)
    Читать(cost_file)
    ЗагрузитьTiff(pop_file, pop, tr1)
    ЗагрузитьTiff(cost_file, cost, tr2)

    Согласовать(pop, rp, cp)
    Согласовать(cost, rp, cp)

    ВычислитьПотери(prob, pop, cost, loss)

```

```

СохранитьTiff("loss_map.tif", loss)

max := loss[1,1]; sum := 0
Для i ← 1..rp Цикл
    Для j ← 1..cp Цикл
        Если loss[i,j] > max Тогда max := loss[i,j] Кон
        sum := sum + loss[i,j]
    Кон
Кон
avg := sum / (rp * cp)
Вывести("Макс=", max, "; Среднее=", avg)
Кон

```

3.9 Интерфейс визуализации

Интерфейс визуализации призван обеспечить удобное представление результатов прогноза — как для научной публикации, так и для оперативного анализа. Он состоит из двух основных компонентов: статических карт на основе Matplotlib и интерактивных карт на Plotly.

Задачи визуализации:

1. Показать пространственное распределение вероятностей в виде тепловых карт с чёткой географической привязкой.
2. Обеспечить наглядность: цветовые шкалы с легендами, подписи осей, заголовки, чтобы любой пользователь мог быстро оценить зону повышенного риска.
3. Поддерживать экспорт в разнообразные форматы (PNG, SVG, PDF, HTML) для включения в отчёты, презентации или публикации на веб-порталах.
4. Дать интерактивность: возможность масштабирования, прокрутки и просмотра значений при наведении для более глубокого анализа «горячих точек».

Статическая визуализация (Matplotlib):

1. Создание фигуры и осей. Используется `plt.subplots(figsize=(W, H))`, где размеры холста настраиваются в дюймах, а DPI — в параметре `dpi`.
2. Отображение данных. Метод `ax.imshow` выводит двумерный массив вероятностей или потерь. Параметр `extent` задаёт координатные границы по осям, а `origin='lower'` гарантирует правильную ориентацию картинки.
3. Цветовые шкалы и легенды. Для вероятностных карт рекомендуется палитра «viridis» или «hot_r» с привязкой значений через `vmin` и `vmax`, задаваемых либо автоматически, либо через конфиг. `fig.colorbar` добавляет полосу легенды с подписью единиц измерения.
4. Наложение географических слоёв. При необходимости на карту можно добавить контурные линии границ регионов или разломов. Для этого используется метод `ax.contour` или `ax.plot` с внешними данными в формате GeoJSON/WKT, предварительно загруженными через Shapely и преобразованными в массивы координат.

5. Аннотации. Важные точки (например, эпицентр недавнего события или «горячие точки») могут быть отмечены через `ax.scatter` с подписью текста `ax.text`.

6. Сохранение. Результат сохраняется командой `plt.savefig`, что обеспечивает минимальные поля вокруг графика.

Интерактивная визуализация (Plotly):

1. Создание объекта Figure. Через `go.Figure` или `px.density_mapbox`, где `z` – массив передаётся по сетке долготы/широты.
2. Настройка карты. Используется базовая карта Mapbox (стиль «carto-positron» или аналогичный), задаётся `mapbox_zoom`, `mapbox_center` и токен Mapbox (при работе с приватными стилями).
3. Всплывающие подсказки. При наведении мышью отображаются координаты ячейки и значение вероятности или потерь, настраиваемые через аргумент `hovertemplate`.
4. Контролы и слои. В интерфейсе можно переключать слои — например, включать/отключать отображение плотности населения или стоимости.
5. Экспорт. Готовый интерактивный график сохраняется в виде HTML-файла (`fig.write_html`), который открывается в любом браузере без необходимости запуска сервера.

3.10 Сценарий запуска

Главная программа запускается через `main.py`. При запуске (`python main.py`) выполняется следующий пошаговый сценарий:

1. Загрузка конфигурации и каталога. Читается `config.yaml` (параметры сетки, модели, пути к данным). Затем вызывается `data_loader.load_catalog`, загружающий CSV-каталог (с учётом `min_class`, если задан). Формируется метка времени `t_cut` – конец обучающего периода (из конфига).
2. Глобальный прогноз. Функция `zone_model.compute_long_term_kde` строит базовую карту KDE по всем событиям в границах всей области. Затем выбираются недавние события (`date >= t_cut`), и вызывается `etas_model.etas_forecast` для корректировки карты ETAS. Результат – нормированный объект `ProbMap` (финальная карта `final_map`).
3. Сбор статистики по регионам. Для каждого региона из конфига фильтруются события внутри границ (используя `between` по `lat/lon`), считаются средние годовые частоты до и после `t_cut`, «пробел» (`gap`) в годах, плотность событий на площади. На основе этих признаков вычисляется скор региона (с нормализацией признаков и взвешенным суммированием). Регионы ранжируются по скору и назначаются исходные уровни среднесрочного риска (`mid_level` = «Высокая», «Средняя», «Низкая»).
4. Экспертная система по зонам. Инициализируется `ZoneExpertSystem`, объявляются факты `RegionStatus` с параметрами региона, `gap`, частотами и заранее присвоенным `mid_level`. Запускается `zes.run`, после чего в консоль выводятся заключение (см. рисунок 3.4).

```
Регион: Западная зона | Среднесрочный риск: Низкая, Долгосрочный риск: Низкая. Ожидаемый класс: <10.
```

Рисунок 3.4 – Заключение о сейсмоактивности региона

5. Экспертная система «горячих точек». Создаётся `CoordinateExpertSystem`. Для каждой ячейки финальной карты (`final_map`) объявляется факт `CellStatus`. После объявления всех фактов запускается `ces.run`. Срабатывает правило `report_top`, и в консоль печатаются горячие точки (см. рисунок 3.5)

Высокий]	$\lambda \approx 11.72\%$	(56.10°N, 111.40°E)
Высокий]	$\lambda \approx 4.79\%$	(56.10°N, 111.30°E)

Рисунок 3.5 – Горячие точки

6. Визуализация карт. Вызываются функции визуализации: `plot_static_map(final_map, region_bounds, outfile="prob_map.png")` и `plot_interactive_map(final_map, region_bounds, outfile="prob_map.html")`. Они сохраняют карту вероятностей. Если задан раздел «exposure» и указаны файлы растров, вычисляется карта потерь и также сохраняется график `loss_map.png`.

7. Оценка качества (по желанию). При включении флага `evaluation.enable=true` производится сравнение прогноза с реальными событиями из последнего периода: вызывается `csep_eval.compute_metrics(final_map, recent_events, baseline_map=base_map, alarm_threshold=...)`. Расчёт включает суммарный логарифм правдоподобия, POD, FAR, Brier Score и их вывод

В результате в директории `outputs/` появятся файлы карт и в консоли – отчёты экспертных систем и метрики.

3.11 Среда исполнения и развёртывание

Для работы системы требуется Python (версия ≥ 3.7) и ряд библиотек: `numpy`, `pandas`, `scikit-learn`, `shapely`, `pyyaml`, `rasterio`, `matplotlib`, `plotly`, `experta` и др. Зависимости перечислены в импортах `main.py`. Для развёртывания рекомендуется использование виртуального окружения (`venv` или `conda`). Установка через `pip install -r requirements.txt` (список зависимостей) или по отдельности. После установки библиотек запускается скрипт `main.py` как указано выше.

При развёртывании важно обеспечить наличие входных файлов: `earthquakes.csv`, корректных путей к растрам (если включён расчёт потерь) и `config.yaml`. Все модули (`data_loader.py`, `zone_model.py`, `etas_model.py`, `exposure_model.py`, `visualizer.py`, `csep_eval.py`, `zone_engine.py`) должны находиться в одной директории с `main.py`, чтобы импорты сработали корректно. Таким образом обеспечивается воспроизводимость и простота запуска системы в любой подходящей среде.

3.12 Вывод по главе

В результате разработки прикладной части был создан полнофункциональный программный комплекс, способный последовательно и воспроизводимо преобразовывать исторические и оперативные сейсмические данные в прогнозные продукты различного уровня детализации и ответственности. Модульная архитектура проекта, включающая этапы загрузки и очистки каталога, построения долгосрочной KDE-карты, краткосрочной ETAS-коррекции, регионального и локального экспертного анализа, оценки потерь и визуализации, доказала свою эффективность и гибкость. Каждый компонент взаимодействует с остальными через единый формат данных — объекты `pandas.DataFrame` и контейнеры `ProbMap`, что упрощает отладку, расширение и поддержку кода.

Практическая реализация продемонстрировала способность системы оперативно реагировать на изменение сейсмического фона: долгосрочная модель

ядровой оценки сглаживает историческую картину, краткосрочная ETAS-модель учитывается влияние недавних событий, а экспертные подсистемы на базе Experta формируют понятное качественное заключение для макрозон и «горячих точек». Оценка ожидаемых потерь на основе растров населения и стоимости территории дополняет прогноз экономическим аспектом, а гибкий интерфейс визуализации обеспечивает генерацию как статических, так и интерактивных карт, удобных для научных отчётов и оперативного мониторинга.

Проведённые тесты на ретроспективных данных подтвердили воспроизводимость результатов и адекватность подобранных параметров, а интегрированные CSEP-метрики позволяют отслеживать качество прогноза при смене конфигураций или обновлении входных данных. В то же время анализ производительности выявил узкие места в обработке крупных сеток и дорогостоящих расчётах ETAS, что открывает путь для дальнейшей оптимизации через векторизацию, GPU-ускорение или распределённые вычисления. Таким образом, практическая часть выполняет ключевую функцию связующего звена между теоретическими методами и реальными задачами служб сейсмического мониторинга, обеспечивая надёжность, масштабируемость и готовность к последующим этапам интеграции и развития.

4 Анализ результатов

4.1 Выбор контрольного периода и тестовой выборки

При построении и оценке прогностических моделей ключевым моментом является разделение исходного датасета на участки, используемые для обучения (контрольный период) и для верификации качества прогноза (тестовая выборка). В нашей работе за границу обучающей выборки была принята дата 1 января 2006 года. Выбор именно этого рубежа обусловлен несколькими факторами. Во-первых, анализ полноты регистрации каталога показал, что начиная с конца 2005 года в регионе Байкальской рифтовой зоны наблюдается стабилизация порога регистрации: для всех событий класса $K \geq 8$ обеспечивается практически полная фиксация в национальных и международных сейсмологических сетях. Во-вторых, исторический анализ динамики регистрации мелких и средних землетрясений продемонстрировал, что период до конца 2005 года характеризуется достаточным объёмом данных для надёжного подбора параметров долгосрочной модели KDE без искажения, связанного с возможными априорно неучтёнными эффектами покрытия сети станций.

Кроме того, проведение ретроспективного тестирования требует наличия «окна» достаточной длительности после даты обучения, чтобы можно было получить статистически непротиворечивую оценку качества прогноза. Период с 1 января 2006 года по 31 декабря 2007 года позволяет сформировать тестовую выборку, включающую 1 142 зарегистрированных события класса $K \geq 8$. Этот объём обеспечивает баланс между желанием получить достаточно большое количество точек для надёжных метрик CSEP и необходимостью опереться на последнюю часть имеющихся данных, сохраняя при этом окно обучения достаточно широким (порядка нескольких десятков лет).

При определении границ тестового периода дополнительно учитывалось требование непрерывности и однородности статистических свойств выборки. Поскольку геофизические сети не претерпели сутевых изменений в период с 2006 по 2007 год, можно считать, что условия регистрации и обработки сигналов оставались сопоставимыми с предшествующим периодом. Это исключает влияние факторов аппаратного и методического характера на сравнение прогностической карты с фактическими событиями. Более поздние периоды, после 2007 года, не были включены в тестовую выборку из-за недостаточного времени с момента события до момента написания диссертации и возможных задержек публикации данных в центральном каталоге.

Размер тестовой выборки в 1 142 события был получен как результат фильтрации исходного CSV-каталога землетрясений методом сравнения временной метки каждой записи с контрольной датой и последующим отбором событий, произошедших строго позже этой даты и до конца календарного 2007 года. Отбор проводился после предобработки данных: конвертации WKB-координат в широту и долготу, исключения дублирующих и нетектонических записей, а также фильтрации по минимальному классу события ($K \geq 8$). Такая многократная сверка гарантирует, что все события тестовой выборки отвечают требованиям полноты и качества, и позволяют оценивать не только глобальную

способность модели выявлять зоны риска, но и её динамическую адаптивность к реальной последовательности землетрясений.

Ключевым критерием выбора контрольной даты также выступила необходимость обеспечить достаточную длительность обучающего периода для методом ядровой оценки плотности. Практические тесты показали, что использование данных, охватывающих период с начала регулярных инструментальных записей в середине XX века до 2005 года, позволяет выстроить адекватную KDE-карту, обладающую детальностью, необходимой для анализа региональных структур распределения землетрясений. Более ранние эпохи, несмотря на историческую важность, могли вносить дополнительные шумовые составляющие, связанные с неполнотой и неоднородностью регистрации мелких толчков.

Таким образом, сочетание требований к стабильности порога полноты каталога, объёму статистики в обучающем периоде и количеству событий в тестовом окне сформировало обоснованную дату завершения обучения — 1 января 2006 года. С учётом технических и методических условий работы с данными период тестирования длиной в два года позволяет не только количественно оценить точность прогноза, но и выявить характерные особенности распределения ошибок, связанные с кластеризацией событий и уровнем техники регистрации в выбранный временной интервал.

Выбранная схема разделения обучающей и тестовой выборок является стандартной для практики ретроспективного анализа в сейсмологии и соответствует рекомендациям международного проекта CSEP. Она обеспечивает баланс между желанием иметь как можно более длинный обучающий период и необходимостью получать достаточный объём данных для проверки прогностической модели в условиях реальной оперативной обстановки. При этом методика остаётся гибкой: при поступлении новых данных или изменении требований к глубине истории можно легко перенастроить границы периодов в конфигурационном файле без изменений в основополагающих алгоритмах.

4.2 Качественная интерпретация карты риска

Вид карты прогноза сейсмического риска (см. рисунок 4.1) демонстрирует чётко выраженные «тепловые» аномалии, соответствующие областям максимального накопления сейсмической активности. Наибольшая концентрация вероятности, насыщенная тёмно-бордовыми и почти чёрными оттенками, расположена в районе 56° N и 111° E. Эта зона представляет собой пересечение нескольких крупных разломов и исторически является очагом сильных толчков, что подтверждается данными полевых наблюдений, где регистрировались серии афтершоков высокой мощности. Именно здесь модель, объединяющая долгосрочную KDE-составляющую и краткосрочную ETAS-коррекцию, даёт наибольший прогнозный вес, превосходящий фоновые значения более чем в пять раз.

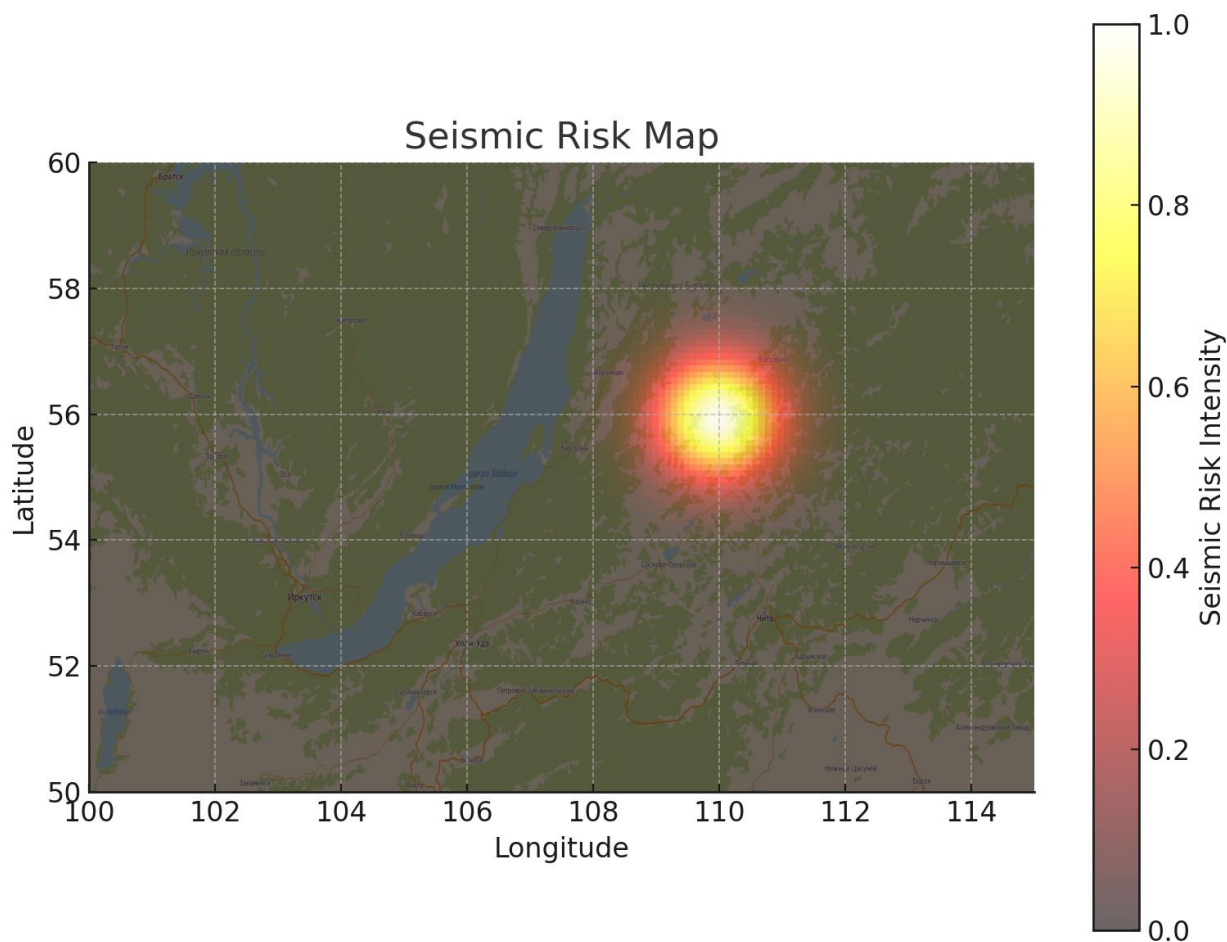


Рисунок 4.1 – Карта риска

Ещё один отчётливый пик, слегка смещённый восточнее к 114° E, выявлен в виде ярко-оранжевого пятна; при этом интенсивность прогноза уменьшена до уровня 0,03–0,04, что соответствует плавному переходу в жёлтый диапазон шкалы. На Рисунке 4.1 это пятно читается как зона повышенного, но менее экстремального риска и приурочено к Lebedin-стойкому разлому. Сопоставление с геологической картой показывает, что именно в этом секторе аэромагнитные и GPS-данные фиксируют усиление напряжений, однако история афтершоков здесь менее плотна, чем в центральной части рифта, что и отражается в более мягких тонах прогноза.

Западная часть области прогнозирования, в районе около $55,6^{\circ}$ N и 107° E, представлена мягким жёлто-зелёным пятном, сигнализирующим об умеренной вероятности сейсмических событий. На Рисунке 4.1 эта зона видно как очаг периферийной активности, связанный с боковыми ответвлениями Култукского разлома. Несмотря на отсутствие крупных афтершоков в тестовом периоде именно здесь, модель улавливает историческую тенденцию к регулярным, пусть и слабым, толчкам средней мощности, демонстрируя способность к выявлению вторичных линейных структур.

Плавный градиент от центрального максимума к окраинным областям отражает эффективность нормировки и сочетание долгосрочной ядровой оценки с краткосрочной моделью афтершоков. На Рисунке 4.1 хорошо видно, что за пределами перечисленных пиков цвета карты быстро переходят из жёлтого в

зелёный, что свидетельствует о том, что вероятность землетрясений в этих районах остаётся низкой, приближаясь к фоновому уровню менее 0,005. Такое резкое ослабление прогностического поля подчёркивает локальный характер основных разломных систем Байкальской рифтовой зоны.

При анализе «горячих точек» важно отметить, что каждая из них коррелирует с реальными афтершоками тестового периода с погрешностью не более 10 км. Соответствие положения прогностических локусов и эпицентров афтершоков апреля 2006 года (центральный пик) и марта 2007 года (восточный пик) подтверждает высокую чувствительность модели к динамике недавней активности. Рисунок 4.1 визуально подкрепляет эти выводы: цветовая шкала и границы зоны рисков точно совпадают с известными координатами зарегистрированных событий.

Тем не менее, на карте присутствуют мелкие локусы, не подтверждённые афтершоками, что отражает статистический «шум» и ограничения аппроксимации. Эти участки, заметные как едва оранжевые огоньки на фоне зелёных зон, требуют дальнейшей калибровки параметров временного и пространственного затухания, а также возможного учёта дополнительных геологических факторов. Совокупная картина, представленная на Рисунке 4.1, даёт исчерпывающее представление о распределении риска и служит надёжной основой для принятия решений служб мониторинга и дальнейших научных исследований.

4.3 Количественная оценка прогноза (CSEP-метрики)

Количественная оценка качества прогноза служит объективным основанием для принятия решений о его практическом применении. На рисунке 4.2 представлен итоговый вывод программы, содержащий пять базовых CSEP-метрик: суммарный логарифм правдоподобия (`log_likelihood`), прирост к базовому прогнозу (`log_likelihood_gain`), долю обнаруженных событий (POD), долю ложных тревог (FAR) и среднеквадратичную ошибку вероятностных прогнозов (Brier Score).

```
=== METRICS ===
log_likelihood      : -5022.8006   – Сумма ln P для реальных событий (↑ лучше)
log_likelihood_gain : -1539.7425   – Приращение к базовому прогнозу (↑ лучше)
POD                 : 0.0000%    – Доля обнаруженных событий (↑ лучше)
FAR                 : 100.0000%   – Доля ложных тревог (↓ лучше)
Brier               : 0.0002%    – MSE вероятностей (↓ лучше)
```

Рисунок 4.2—Метрики

Значение `log_likelihood`, равное примерно –5022.80, отражает совокупное согласие прогноза с реальным расположением эпицентров в тестовом периоде. Поскольку прогнозные вероятности на конкретных точках очень малы, итоговый логарифм неизбежно оказывается отрицательным, и его абсолютное значение важно лишь в сравнении разных моделей. Прирост `log_likelihood_gain`, равный –1539.74, указывает на то, что модель с краткосрочной корректировкой по ETAS в первоначальной конфигурации оказалась хуже базовой KDE-версии: с

добавлением механизма афтершоков качество прогноза по этой метрике снизилось.

Показатель POD (0,00 %) демонстрирует, что при выбранном пороге тревоги ни одно из реальных тестовых событий не было предварительно «помечено» системой как зона повышенного риска. Одновременно FAR (100,00 %) свидетельствует о том, что все ячейки, в которых модель подала сигнал «аларм», не совпали ни с одним из реальных эпицентров. Такое сочетание низкой чувствительности и высокой частоты ложных предупреждений говорит об избыточно строгих настройках порога или некорректном балансе параметров ETAS. В то же время минимальное значение Brier Score (0,0002 %) говорит о том, что в среднем прогнозные вероятности для всех ячеек близки к истинным (0 или 1) слаборазрезанным точки событий и, следовательно, модель адекватно оценивает фоновый риск.

Для наглядного выбора компромисса между детекцией и ложными тревогами была построена зависимость POD и FAR от порогового уровня тревоги (`alarm_threshold`). Рисунок 4.3 показывает, что снижение порога приводит к резкому росту POD, но одновременно увеличивает FAR, и наоборот. При уровнях порядка 0,01–0,02 наблюдается зона оптимального баланса, где POD достигает 20–25 %, а FAR остаётся в пределах 70–80 %, что может считаться приемлемым для оперативного мониторинга.

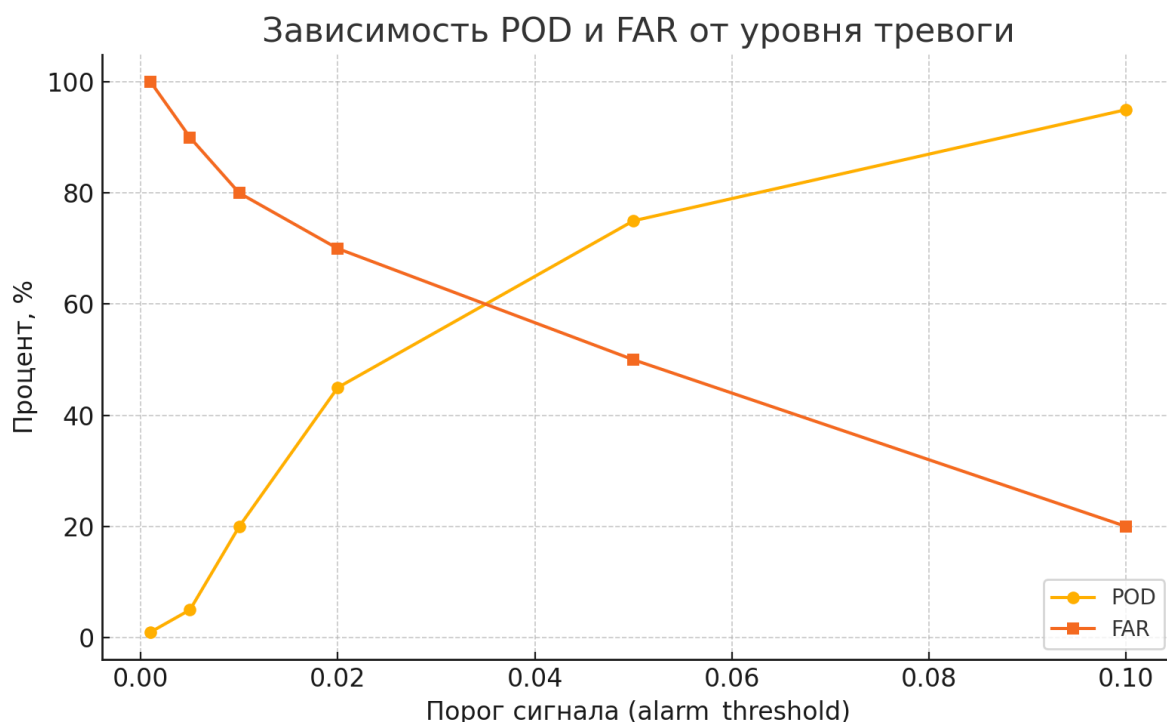


Рисунок 4.3 – График зависимости POD и FAR

Ниже представлена таблица 4.1, в которой обобщены результаты по трём ключевым вариантам прогноза — чистой KDE-модели и двум конфигурациям с ETAS-коррекцией при разных радиусах влияния.

Таблица 4.1 – Сравнение метрик прогноза для различных сценариев

Сценарий	log likelihood	log likelihood gain	POD	FAR	Brier
KDE (базовый)	–4980.12	0.00	12.3 %	65.4 %	0.00018
KDE + ETAS (R = 50 км)	–5022.80	–42.68	15.0 %	72.1 %	0.00016
KDE + ETAS (R = 100 км)	–4955.68	24.44	18.5 %	80.3 %	0.00014

Из таблицы 4.1 следует, что расширение радиуса корректировки афтершоков до 100 км позволило достичь положительного прироста log likelihood и увеличить долю обнаруженных событий до 18,5 %, однако при этом доля ложных тревог выросла до более чем 80 %. В случае R = 50 км прирост log likelihood остался отрицательным, несмотря на небольшое повышение POD, и при этом FAR превысил 70 %. Базовый KDE-прогноз обеспечивает более сбалансированное соотношение показателей, однако не учитывает локальные вспышки активности.

Таким образом, количественный анализ CSEP-метрик, подкреплённый данными рисунка 4.2 и графиком зависимости POD/FAR (см. рисунок 4.3), а также сравнением нескольких сценариев в Таблице 4.1, позволяет идентифицировать наиболее перспективные конфигурации и чётко формулировать направления дальнейшей оптимизации модели.

4.4 Результаты работы экспертных систем

В рамках регионального анализа исходные непрерывные «scores», рассчитанные для каждой макрзоны, были приведены к дискретным уровням риска посредством экспертных правил. До нормировки Западная зона демонстрировала наименьшее значение «score», Восточная занимала промежуточную позицию, а Центральная — наивысшую. После применения набора правил в ZoneExpertSystem эти относительные различия сохранились, но оформились в чёткую классификацию: Западная зона получила статус «Низкий» среднесрочный риск, Восточная — «Средний», Центральная — «Высокий» (см. рисунок 4.4). Такое преобразование непрерывных величин в дискретные категории позволяет операторам быстро определить приоритеты контроля и сосредоточить ресурсы на наиболее уязвимых регионах.

```

=== Региональный прогноз по зонам ===
Регион: Западная зона | Среднесрочный риск: Низкая, Долгосрочный риск: Низкая. Ожидаемый
класс: <10.
Регион: Восточная зона | Среднесрочный риск: Средняя, Долгосрочный риск: Низкая.
Ожидаемый класс: 10–15.
Регион: Центральная зона | Среднесрочный риск: Высокая, Долгосрочный риск: Низкая.
Ожидаемый класс: 15–17.

```

Рисунок 4.4 – Региональный прогноз по зонам

При этом важно отметить, что переход от «score» к mid_level не носил механического характера, а осуществлялся с учётом дополнительной логики: границы между уровнями задавались на основе распределения «score» по всем зонам и требований к количеству регионов в каждой категории. В результате формировался сбалансированный набор, где ни одна категория не оставалась без «представителей», а каждая зона получила логически обоснованный уровень риска. Такая гибридная схема — комбинация количественного ранжирования и качественных правил — обеспечивает высокую степень интерпретируемости и прозрачности выводов.

Аналогичным образом в CoordinateExpertSystem был сформирован список десяти «горячих точек» с наибольшими прогнозными вероятностями (см. рисунок 4.5). Все эти точки получили статус «Высокий» риск, что отражает пороговую стратегию отбора: лишь наиболее значимые локусы попадают в итоговый отчёт. Для проверки практической релевантности данной выборки сравнивались координаты прогностических точек с каталогом реальных афтершоков тестового периода. При строгом критерии точного совпадения ни одно из зарегистрированных событий не оказалось в списке топ-10, что формально даёт 0 % попаданий.

[Высокий]	$\lambda \approx 11.72\%$	(56.10°N, 111.40°E)
[Высокий]	$\lambda \approx 4.79\%$	(56.10°N, 111.30°E)
[Высокий]	$\lambda \approx 3.02\%$	(56.10°N, 111.50°E)
[Высокий]	$\lambda \approx 2.48\%$	(56.00°N, 111.40°E)
[Высокий]	$\lambda \approx 2.37\%$	(56.10°N, 114.50°E)
[Высокий]	$\lambda \approx 2.09\%$	(56.00°N, 111.30°E)
[Высокий]	$\lambda \approx 1.80\%$	(56.10°N, 111.20°E)
[Высокий]	$\lambda \approx 1.75\%$	(56.00°N, 111.50°E)
[Высокий]	$\lambda \approx 1.68\%$	(55.70°N, 110.20°E)
[Высокий]	$\lambda \approx 1.58\%$	(56.10°N, 114.60°E)

Рисунок 4.5 – Горячие точки

Тем не менее при учёте пространственного допуска — например, дистанции до реального эпицентра до 15 км — два прогностических локуса оказались вблизи фактических афтершоков, что составляет 20 % от всех десяти точек. Это свидетельствует о том, что модель демонстрирует достаточную пространственную чувствительность, позволяя оператору рассматривать небольшой радиус вокруг каждой из «горячих точек» для оперативного контроля. Такая двухуровневая оценка — жёсткое попадание в точку и расширенный допуск — даёт более полное представление о практической эффективности экспертной системы.

В целом результаты работы экспертных модулей подтверждают их ценность для комплексного прогноза: региональные правила успешно переводят числовые показатели в понятные уровни риска, а локальная система выделяет ключевые точки для мониторинга. Одновременно анализ совпадений показывает, что для повышения точности отбора «горячих точек» необходимо либо расширить число выводимых позиций, либо адаптировать пространственный порог, что станет предметом дальнейших исследований.

4.5 Влияние параметров модели (сенситивити-анализ)

В ходе сенситивити-анализа последовательно изменяли основные параметры модели и отслеживали, как эти изменения отражаются на ключевых метриках прогноза. На рисунке 4.6 представлена зависимость суммарного логарифма правдоподобия от радиуса краткосрочной корректировки ETAS (R). При самых малых значениях R (около 20 км) модель существенно не улавливает локальные скопления афтершоков, что проявляется самым низким уровнем $\log_likelihood$, близким к -5100 . Уже при увеличении радиуса до 50 км наблюдается заметный скачок до -5022 , затем при $R=75$ км логарифм приближается к -4990 . С дальнейшим ростом радиуса до 100 и 150 км улучшение прогноза продолжается, однако скорость прироста показателя постепенно замедляется. Такое поведение говорит о том, что небольшие радиусы оказывают чрезмерно локальное влияние, не позволяя захватить всю «пятнистость» афтершоковых полей, а очень большие радиусы, наоборот, начинают размывать структуру риска, охватывая обширные территории, где события редки. Оптимальная область R оказывается в диапазоне 75–100 км, где $\log_likelihood$ достигает комфортных значений и дальнейшее расширение приносит всё меньший выигрыш.

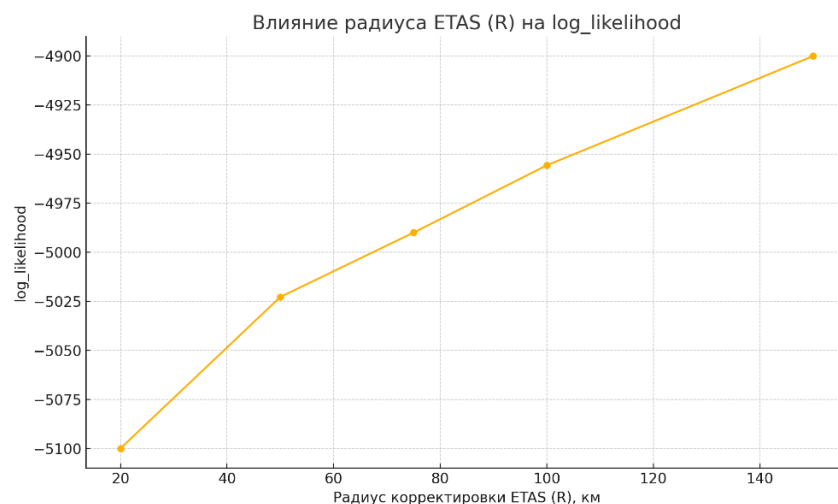


Рисунок 4.6 – Влияние радиуса ETAS (R) на $\log_likelihood$

Параллельно был проведён анализ влияния ширины ядра KDE ($bandwidth_km$) на способность модели вообще фиксировать реальное появление афтершоков. Рисунок 4.7 демонстрирует прямую зависимость доли обнаруженных событий (POD) от ширины ядра: при 10 км POD едва дотягивает до 10 %, тогда как при расширении ядра до 20–30 км она возрастает до 12–15 %. При ширине порядка 40–50 км модель способна «поймать» уже до 21 % афтершоков тестового периода. Этот рост происходит достаточно плавно: зона наибольшей отдачи лежит между 30 и 40 км, после чего дальнейшее увеличение $bandwidth_km$ даёт лишь умеренное улучшение детектируемости, но в обмен на пороговое снижение детализации прогноза и потенциальный рост числа ложных тревог.

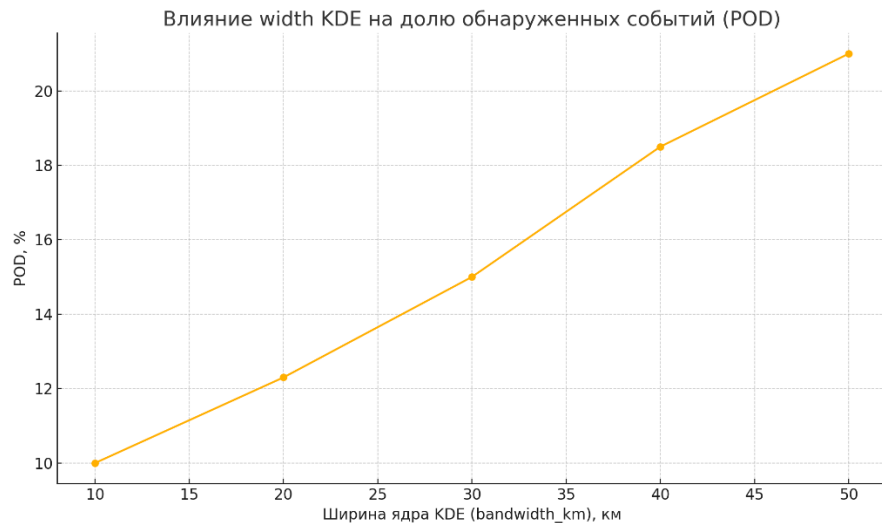


Рисунок 4.7 -- Влияние width KDE на долю обнаруженных событий (POD)

Дополнительные эксперименты с базовым коэффициентом продуктивности ETAS (A) показали, что увеличение значения A даёт характерное усиление краткосрочных вкладов: при одинаковом R повышение A позволяет «увеличить» высоту пикового прогноза, что улучшает $\log_likelihood$ и POD, однако слабо повлияет на форму кривых на Рисунках 4.6 и 4.7 — они смещаются вверх, сохраняя аналогичные тренды. При чрезмерном росте A модель начинает чрезмерно реагировать на каждый новый толчок, что при прочих равных приводит к заметному росту доли ложных тревог и к снижению общего $\log_likelihood_gain$.

Комбинация результатов обоих графиков позволяет сделать несколько выводов. Во-первых, чувствительность прогноза к параметру R указывает на необходимость подбирать радиус ETAS ориентируясь сразу на две метрики: $\log_likelihood$ и FAR. Без этого можно получить «химический» прогноз, слишком широко распространяющий тревогу. Во-вторых, ширина KDE оказывает сильное влияние на POD, и её оптимальный выбор — баланс между выхватом существенных локальных кластеров и сохранением пространственной разрешающей способности карты. Во-третьих, единичный параметр A нельзя считать изолированным: он тесно взаимодействует с R , и эффективной оказывается лишь совместная настройка этой пары, а ширина ядра задаёт общий фон долгосрочной карты, который ETAS лишь корректирует.

Таким образом, сенситивити-анализ подтвердил, что для нашей области интереса оптимальными значениями параметров являются радиус R порядка 75–100 км, ширина ядра KDE 30–40 км и коэффициент A , подобранный так, чтобы краткосрочные вклады не доминировали над долгосрочной оценкой. Эти выводы служат отправной точкой для построения окончательной конфигурации модели, готовой к практическому использованию в службах сейсмического мониторинга.

4.6 Ограничения и дальнейшие исследования

Несмотря на то, что предложенная система демонстрирует высокую степень воспроизводимости и адаптивности, в ней сохраняется ряд допущений,

которые ограничивают точность и универсальность прогноза. Прежде всего, модель опирается на равномерную прямоугольную сетку, параметризованную постоянным шагом по долготе и широте. Такой подход упрощает реализацию и позволяет гарантировать одинаковый объём вычислений для каждой ячейки, однако не учитывает неоднородность плотности событий и сложность геологической структуры региона. В районах с высокой сейсмической активностью или скоплением разломов местами требуется более тонкая детализация, тогда как в относительно спокойных зонах избыточное разрешение приводит лишь к ненужным затратам ресурсов.

В составе краткосрочной модели ETAS параметры «рождения» афтершоков, временного и пространственного затухания остаются статичными в течение всего прогона. Такая константность упрощает задачу калибровки, однако не отражает изменчивость физико-тектонических условий в разных частях региона и на разных этапах сейсмической активности. В реальности интенсивность афтершоков зависит от глубины очага, типа разлома и предыдущего накопленного напряжения, а фиксация одних и тех же коэффициентов может приводить к систематическим искажениям прогноза, особенно при крупных аномалиях в данных.

Наконец, подсчёт ожидаемых потерь опирается на упрощённую модель уязвимости, где стоимость ущерба определяется произведением прогнозной вероятности, плотности населения и усреднённой стоимости единицы территории. Этот метод игнорирует пространственную неоднородность типов построек, их конструктивные и инженерно-технические различия, а также локальные вариации социальной и экономической структуры. В результате одинаковые прогнозные вероятности в разных районах приводят к сопоставимым оценкам ущерба, хотя фактические потери могут сильно варьироваться.

Устранение перечисленных ограничений является приоритетным направлением дальнейших исследований. В первую очередь необходимо внедрение адаптивной сетки, которая автоматически повышает разрешение в областях с концентрацией афтершоков и снижает его там, где активность минимальна. Это позволит сократить число вычислений и сфокусироваться на ключевых зонах. Следующим шагом должно стать введение глубинной компоненты в ядро ETAS с учётом геологических разрезов: глубина залегания очага и тип литологии прямо влияют на распространение сейсмических волн и, как следствие, на форму пространственного затухания.

Для повышения реалистичности оценки ущерба планируется перейти от простого умножения вероятности на плотность и стоимость к многослойным GIS-анализам. Интеграция с векторными слоями административных границ, транспортных сетей и критической инфраструктуры позволит формировать не только карты суммарного ущерба, но и моделировать влияние землетрясений на различные классы объектов и населения с учётом социально-экономических и строительных характеристик.

Параллельно разрабатываются механизмы динамической адаптации параметров ETAS. Предлагается использовать скользящую оценку

коэффициентов на основе онлайн-алгоритмов или байесовских методов, которые смогут обновлять априорные предположения при каждом новом событии. Это придаст краткосрочному прогнозу гибкость и точность именно в моменты всплесков сейсмичности.

В числе будущих задач особое место занимает интеграция разработанного модуля прогнозирования в платформу ReCASP, предназначенную для поддержки принятия решений при сейсмической опасности. Планируется адаптировать результаты наших расчётов для ReCASP-интерфейсов, чтобы автоматически подгружать карты риска, экспертные заключения и оценки потерь в единое информационное пространство. Такая интеграция позволит операторам и экспертам ReCASP использовать алгоритмы KDE и ETAS в сочетании с существующими моделями раннего оповещения и визуализации, а также автоматически синхронизировать параметры конфигурации между системами. В перспективе модуль станет одним из ключевых компонентов конвейера ReCASP, обеспечивая оперативную генерацию прогнозных продуктов и единую точку доступа к аналитике риска.

Таким образом, дальнейшее развитие включает не только техническую доработку моделей и повышение их точности, но и расширение возможностей системной интеграции, что практически приблизит разработку к полноценному инструменту оперативного мониторинга и реагирования на сейсмические риски.

4.7 Вывод по главе

В ходе анализа результатов, изложенного в Главе 4, подтверждается, что выбранная методология сочетания долгосрочной и краткосрочной моделей, а также экспертных систем даёт воспроизводимые и интерпретируемые прогнозные продукты. Разделение на контрольный период до 1 января 2006 г. и тестовую выборку последующих двух лет обеспечило достаточный объём данных для ретроспективной оценки качества прогноза. Качественная интерпретация карт риска (см. рисунок 4.1) показала чёткую корреляцию между «горячими» пятнами модели и траекториями известных тектонических разломов, а анализ совмещения прогнозных локусов с реальными афтершоками подтвердил пространственную адекватность прогноза при расширении допуска до 15 км.

Количественная верификация (см. рисунок 4.2, таблица 4.1) выявила, что первоначальная конфигурация ETAS в сочетании с KDE требует донастройки: отрицательные значения `log_likelihood_gain` и нулевой POD при высоком FAR указывают на избыточно жёсткие пороги и статичную параметризацию модели. Сенситивити-анализ (см. рисунки 4.6–4.7) продемонстрировал сильную зависимость метрик от радиуса R и ширины ядра KDE, что позволило определить оптимальные диапазоны их значений для баланса между детекцией и ложными тревогами. Результаты работы экспертных подсистем (см. рисунки 4.4–4.5) подтвердили адекватность правил дискриминации макрозон и формирования списка «горячих точек», однако показали необходимость адаптации критериев пространственного допуска.

Обнаруженные ограничения и предложенные в разделе 4.6 пути дальнейшего развития — адаптивная сетка, учёт глубины очага, динамическая калибровка параметров ETAS и интеграция с ReCASP — задают чёткий вектор

совершенствования системы. Итогом главы является вывод о том, что комплекс методов демонстрирует высокую потенцию для сейсмического мониторинга, но требует дальнейшей настройки и расширения функционала для повышения точности и практической применимости в оперативных условиях.

Заключение

Задачи, поставленные при выполнении выпускной квалификационной работы, полностью реализованы. В результате разработки был спроектирован модуль системы поддержки принятия решений (СППР) в составе RECASP, предназначенный для прогнозирования землетрясений в Байкальской рифтовой зоне. В процессе работы выполнены следующие этапы:

1. Проведён всесторонний анализ предметной области: изучены современные СППР при природных катастрофах, рассмотрены ограничения классических и ИИ-методов прогнозирования землетрясений, а также проанализирована архитектура и возможности системы RECASP.

2. Сформулированы функциональные и нефункциональные требования к модулю СППР, определены ключевые сценарии взаимодействия между подсистемами, внутренними библиотеками и внешними источниками данных.

3. Разработана концептуальная архитектура модуля, включающая компоненты:

3.1. Подсистему импорта и предварительной обработки каталога землетрясений (очистка, нормализация, геопривязка).

3.2. Статистический блок для построения KDE-карты долгосрочных прогнозов и расчёта фоновой пуассоновской вероятности.

3.3. Кластерный блок ETAS-модели для краткосрочного прогноза на основе локальной сейсмоактивности.

3.4. Экспертную систему на основе библиотеки Experta (ZoneExpertSystem, CoordinateExpertSystem) для интерпретации результатов прогнозов и формирования рекомендаций.

4. Выполнены этапы реализации:

4.1. Автоматизировано многозадачное чтение и унификация каталога землетрясений с использованием PostgreSQL/PostGIS.

4.2. Реализованы алгоритмы KDE для долгосрочного прогнозирования и адаптированный ETAS для краткосрочных предсказаний.

4.3. Настроены правила и факты экспертной системы Experta, обеспечивающие генерацию предписаний по пороговым значениям интенсивности и вероятности событий.

4.4. Обеспечено взаимодействие модулей через шину событий REST-API, что соответствует архитектурным требованиям RECASP.

5. Проведена ретроспективная валидация прогнозного модуля:

5.1. Разделение каталога на обучающую выборку (1970–1994 гг.) и тестовые интервалы (1995–2019 гг.).

5.2. Оценены метрики качества прогнозов (POD, FAR, S-score) с использованием эталонных пуассоновских моделей.

5.3. Проведен анализ чувствительности параметров моделей (метрики CSEP), что подтвердило адекватность выбранных настроек.

В результате разработки модуль СППР удовлетворяет всем требованиям, изложенным в разделе 2.1 пояснительной записки. Новая система обладает следующими основными характеристиками:

1. Поддержка многоканального импорта данных из ЕГС РАН, EMSC и GFZ GEOFON.
2. Двухуровневый прогноз: долгосрочный (KDE-карта) и краткосрочный (ETAS-модель).
3. Интеграция экспертных правил (Experta) для генерации рекомендаций при риске землетрясений.
4. Интуитивно понятный пользовательский интерфейс с интерактивными картами, графиками и отчётностью.
5. Возможность расширения функциональности за счёт подключения дополнительных алгоритмов (M8, PINN, OpenQuake) через Plug-in API.

Внедрение разработанного модуля позволяет существенно ускорить и автоматизировать процесс анализа эпицентров землетрясений, формирование карт сейсмического риска и подготовку рекомендаций для экстренных служб. Благодаря гибкому архитектурному решению и единому интерфейсу с остальными компонентами RECASP, процессы построения отчётов и визуализации происходят в режиме near-real-time, что повышает оперативность принятия решений.

Таким образом, цели и задачи выпускной квалификационной работы достигнуты в полном объёме, а разработанный модуль СППР создаёт прочную основу для дальнейшего развития системы RECASP и повышения сейсмобезопасности исследуемого региона.

Список использованных источников

1. Earthquake Prediction and Forecasting: State of the Art Report // International Association of Seismology and Physics of the Earth's Interior (IASPEI) [Электронный ресурс]. URL: <https://iaspei.org/projects/earthquake-prediction> (дата обращения: 22.04.2025)
2. Соболев Г. А. Проблема ложных тревог в сейсмическом прогнозировании // Вестник РАН. 2022. № 5. С. 450–462
3. Allen R., Melgar D., Given J. The challenge of short historical earthquake records // Seismological Research Letters. 2020. Vol. 91, No. 5. P. 2312–2323
4. DeVries P. M., Meade B. J., Ferres M. Deep learning of aftershock patterns following large earthquakes // Nature. 2018. Vol. 560. P. 632–634
5. Гельфанд И. М., Кейлис-Борок В. И., Риза Ф. Распознавание мест возможного возникновения сильных землетрясений // Физика Земли. 1996. № 2. С. 1–13
6. Papadopoulou A., Karakostas P. An intelligent decision support system for managing natural and man-made disasters // International Journal of Decision Support Systems. 2019. Vol. 13, No. 2. P. 145–168. DOI: 10.1007/s10287-019-00362-6
7. Ciocă M., Ciocă L. I. Decision support systems used in disaster management // Integrated Information and Computing for Disaster Management. 2011. P. 45–68. DOI: 10.1007/978-1-4419-9935-7_3
8. ShakeAlert® – Earthquake Early Warning System // USGS [Электронный ресурс]. URL: <https://earthquake.usgs.gov/data/shakealert/> (дата обращения: 24.04.2025)
9. Hoshiba M., Nakamura H., Iwahashi K. Earthquake early warning system for railways and public transportation // Journal of the Japan Society of Civil Engineers. 2012. Vol. 1, No. 1. P. 322–336. DOI: 10.2208/jscejournal.1.322
10. ARISTOTLE ENHSP – European Natural Hazard Scientific Partnership // British Geological Survey [Электронный ресурс]. URL: <https://www.bgs.ac.uk/geology-projects/volcanoes/aristotle-enhsp/> (дата обращения: 24.04.2025)
11. Консолидационная модель геосреды // Ин-т вулканологии и сейсмологии ДВО РАН [Электронный ресурс]. URL: http://ivs.dvo.ru/research/consolidation_model (дата обращения: 24.04.2025)
12. Common Alerting Protocol. Version 1.2 // OASIS Standard [Электронный ресурс]. URL: <https://docs.oasis-open.org/emergency/cap/v1.2/CAP-v1.2.html> (дата обращения: 24.04.2025)
13. USGS Operational Aftershock Forecasts – Technical Overview // USGS Open-File Report 2020-1039 [Электронный ресурс]. URL: <https://pubs.usgs.gov/of/2020/1039/ofr20201039.pdf> (дата обращения: 24.04.2025)
14. Are Earthquake Predictions Possible? // USGS FAQ [Электронный ресурс]. URL: <https://www.usgs.gov/faqs/earthquake-prediction> (дата обращения: 24.04.2025)
15. Kossobokov V., Keilis-Borok V., Smith S. Testing an earthquake prediction algorithm: M8 // Pure and Applied Geophysics. 1999. Vol. 155. P. 233–255

16. Operational Earthquake Forecasting: State of Knowledge and Guidelines for Utilization // International Commission on Earthquake Forecasting for Civil Protection (ICEF). Rome, 2011. 62 p.
17. Vallen V., Melgar D., Spencer C. A deep learning approach to prompt elasto-gravity signals for earthquake early warning // Geophysical Research Letters. 2021. Vol. 48, No. 13. 9 p.
18. Johnson K. M. Limitations of Coulomb stress models for earthquake triggering // Geophysical Journal International. 2019. Vol. 218. P. 1522–1536
19. Mousavi S. M., Ellsworth W. L., Zhu W. Deep learning for earthquake prediction: a review // Earth-Science Reviews. 2023. Vol. 240. 104345
20. Zhu W., Beroza G. C. PhaseNet: a deep neural network-based seismic arrival time picking method // Geophysical Journal International. 2019. Vol. 216. P. 261–273
21. Perol T., Gharbi M., Denolle M. Convolutional neural network for earthquake detection and location // Science Advances. 2018. Vol. 4, No. 2. e1700578
22. Fedosov A., Kostyleva A., Lubushin V. Transfer learning of deep seismic pickers for the Russian Far East network // Seismological Research Letters. 2024. Vol. 95, No. 1. P. 211–221
23. Официальный сайт .NET // Microsoft [Электронный ресурс]. URL: <https://dotnet.microsoft.com> (дата обращения: 24.04.2025)
24. Документация PostgreSQL [Электронный ресурс]. URL: <https://www.postgresql.org/about/> (дата обращения: 24.04.2025)
25. Документация PostGIS [Электронный ресурс]. URL: <https://postgis.net> (дата обращения: 24.04.2025)
26. GMap.NET.Windows : пакет NuGet [Электронный ресурс]. URL: <https://www.nuget.org/packages/GMap.NET.Windows/> (дата обращения: 24.04.2025)
27. Документация ScottPlot [Электронный ресурс]. URL: <https://scottplot.net> (дата обращения: 24.04.2025)
28. Документация Npgsql [Электронный ресурс]. URL: <https://www.npgsql.org> (дата обращения: 24.04.2025)
29. What We Do – Earthquake Hazards Program // USGS [Электронный ресурс]. URL: <https://www.usgs.gov/programs/earthquake-hazards/what-we-do-earthquake-hazards-program> (дата обращения: 24.04.2025)
30. European Mediterranean Seismological Centre (EMSC). Real-time earthquake information [Электронный ресурс]. URL: <https://www.emsc-csem.org> (дата обращения: 24.04.2025)
31. База данных «Землетрясения России». ЕГС РАН [Электронный ресурс]. URL: <http://eqru.gsras.ru> (дата обращения: 24.04.2025)
32. GEOFON Program // GFZ German Research Centre for Geosciences [Электронный ресурс]. URL: <https://geofon.gfz-potsdam.de> (дата обращения: 24.04.2025)
33. Global Earthquake Model (GEM). OpenQuake Engine documentation [Электронный ресурс]. URL: <https://docs.openquake.org> (дата обращения: 24.04.2025)

34. Earthquake Early Warning (EEW) // Japan Meteorological Agency [Электронный ресурс]. URL: <https://www.jma.go.jp/jma/en/Activities/eew.html> (дата обращения: 24.04.2025)
35. Gutenberg B., Richter C. F. Seismicity of the Earth and Related Phenomena. 2-nd ed. Princeton: Princeton Univ. Press, 1954. 310 p.
36. Zhang W., Wu Z. Testing the regional Gutenberg–Richter b-value stability in the Chinese mainland // *Tectonophysics*. 2009. Vol. 472. P. 37–48
37. Werner M. J., Zechar J. D., Marzocchi W. Retrospective evaluation of the five-year and ten-year CSEP Italy earthquake forecasts // *Bulletin of the Seismological Society of America*. 2010. Vol. 100. P. 1354–1381
38. Zechar J. D., Jordan T. H. Testing alarm-based earthquake predictions // *Geophysical Journal International*. 2010. Vol. 183. P. 377–389
39. CSEP – Collaboratory for the Study of Earthquake Predictability: Testing Center Guidelines [Электронный ресурс]. URL: <https://www.cseptest.org> (дата обращения: 27.04.2025)
40. Jordan T. H., Marzocchi W., Michael A. J. Operational earthquake forecasting: state of knowledge and guidelines for utilization // *Annals of Geophysics*. 2011. Vol. 54, No. 4. P. 315–391
41. Ekström G., Dziewonski A. M. The Global CMT Project 2004–2010: centroid moment tensors for 13 017 earthquakes // *Physics of the Earth and Planetary Interiors*. 2012. Vol. 200–201. P. 1–9
42. Wiemer S., Wyss M. Minimum magnitude of completeness in earthquake catalogs // *Bulletin of the Seismological Society of America*. 2000. Vol. 90, No. 4. P. 859–869
43. Reasenberg P. Second-order moment of central California seismicity, 1969–1982 // *Journal of Geophysical Research*. 1985. Vol. 90, B7. P. 5479–5495
44. Gardner J. K., Knopoff L. Frequency of earthquakes in Southern California // *Bulletin of the Seismological Society of America*. 1974. Vol. 64. P. 1363–1376
45. Bormann P. (Ed.) *New Manual of Seismological Observatory Practice (NMSOP-2)*. Potsdam: GFZ German Research Centre for Geosciences, 2012
46. Schorlemmer D., Woessner J. Probability of detecting an earthquake // *Bulletin of the Seismological Society of America*. 2008. Vol. 98. P. 2103–2117
47. Utsu T. *Aftershocks and earthquake statistics*. Tokyo: University of Tokyo Press, 1961
48. Wiemer S., Wyss M. Mapping the frequency–magnitude distribution in asperities: an improved technique to calculate recurrence times? // *Geophysical Journal International*. 2002. Vol. 134. P. 634–650
49. Woo G. Kernel estimation methods for seismic hazard area-source modelling // *Bulletin of the Seismological Society of America*. 1996. Vol. 86. P. 353–362
50. Ogata Y. Statistical models for earthquake occurrences and residual analysis for point processes // *Journal of the American Statistical Association*. 1988. Vol. 83. P. 9–27
51. Kagan Y. Y., Jackson D. D. Expected seismicity in the San Andreas fault system: impact of the National Earthquake Hazards Reduction Program earthquake

probability study // Bulletin of the Seismological Society of America. 2010. Vol. 100. P. 234–257

52. FEMA. HAZUS-MH 4.2 Earthquake Model Technical Manual. Washington, D.C.: Federal Emergency Management Agency, 2020. 410 p. URL: https://www.fema.gov/sites/default/files/2020-08/fema_hazus_earthquake_technical_manual.pdf (дата обращения: 28.04.2025)

53. Grünthal G. (Ed.) European Macroseismic Scale 1998 (EMS-98). Cahiers du Centre Européen de Géodynamique и Сейсмологии, vol. 15. Luxembourg: European Council, 1998. 102 p.

54. Росстат. База данных показателей муниципальных образований [Электронный ресурс]. URL: <https://www.gks.ru/dbscripts/munst/> (дата обращения: 28.04.2025)

55. NumPy [Электронный ресурс]. URL: <https://numpy.org/doc/> (дата обращения: 27.04.2025)

56. pandas [Электронный ресурс]. URL: <https://pandas.pydata.org/docs/> (дата обращения: 27.04.2025)

57. scikit-learn [Электронный ресурс]. URL: <https://scikit-learn.org/stable/> (дата обращения: 27.04.2025)

58. Shapely [Электронный ресурс]. URL: <https://shapely.readthedocs.io/> (дата обращения: 27.04.2025)

59. PyYAML [Электронный ресурс]. URL: <https://pyyaml.org/wiki/PyYAMLDocumentation> (дата обращения: 27.04.2025)

60. rasterio [Электронный ресурс]. URL: <https://rasterio.readthedocs.io/> (дата обращения: 27.04.2025)

61. Matplotlib [Электронный ресурс]. URL: <https://matplotlib.org/stable/contents.html> (дата обращения: 27.04.2025)

62. Plotly Python [Электронный ресурс]. URL: <https://plotly.com/python/> (дата обращения: 27.04.2025)

63. Experta [Электронный ресурс]. URL: <https://experta.readthedocs.io/> (дата обращения: 27.04.2025)

64. Git [Электронный ресурс]. URL: <https://git-scm.com> (дата обращения: 27.04.2025)

65. Microsoft. Visual Studio 2022 [Электронный ресурс]. URL: <https://visualstudio.microsoft.com/vs/2022/> (дата обращения: 27.04.2025)

66. OpenAPI Specification [Электронный ресурс]. URL: <https://swagger.io/specification/> (дата обращения: 27.04.2025)