

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Комп'ютерних наук
(повна назва)
Кафедра Штучного інтелекту
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти перший (бакалаврський)

Розробка системи аналізу сейсмічної активності на основі
стохастичних моделей
(тема)

Виконав:
здобувач четвертого року навчання,
групи ІТШ-22-4

Валентин РИБІНСЬКИЙ
(власне ім'я, прізвище)

Спеціальність 122 Комп'ютерні науки
(код і повна назва спеціальності)

Тип програми освітньо-професійна
Освітня програма Штучний інтелект
(повна назва освітньої програми)

Керівник ст.викл. Олександр СТЬОПІН
(посада, власне ім'я, прізвище)

Допускається до захисту

Завідувач кафедри ШІ

Лариса ЧАЛА
(підпис) (власне ім'я, прізвище)

2026 р.

Харківський національний університет радіоелектроніки

Факультет _____ Комп'ютерних наук _____

Кафедра _____ Штучного інтелекту _____

Рівень вищої освіти _____ перший (бакалаврський) _____

Спеціальність _____ 122 Комп'ютерні науки _____
(код і повна назва)

Тип програми _____ освітньо-професійна _____

Освітня програма _____ Штучний інтелект _____
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____

(підпис)

«_____» _____ 20 ____ р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

здобувачеві _____ Рибінському Валентину Максимовичу _____
(прізвище, ім'я, по батькові)

1. Тема роботи _____ Розробка системи аналізу сейсмічної активності на основі стохастичних моделей _____

затверджена наказом університету від 5 червня 2026 р. № 488Ст

2. Термін подання здобувачем роботи до екзаменаційної комісії 14 липня 2026 р.

3. Вихідні дані до роботи Науково-технічні публікації з сейсмології та стохастичного моделювання, відкриті сейсмічні каталоги USGS та EMSC, стандарт FDSN Event Web Service, документація Unity, SQLite та бібліотек Zenject і UniTask

4. Перелік питань, що потрібно опрацювати в роботі _____

1) Аналіз предметної галузі _____

2) Методи збору та обробки сейсмічних даних _____

3) Стохастичні моделі аналізу та прогнозування _____

4) Програмна реалізація _____

КАЛЕНДАРНИЙ ПЛАН

[illegible]

Дата видачі завдання 8 червня 20 26 р.

Здобувач

(підпис)

Керівник роботи

ст.викл. Олександр СТЬОПН
(посада, власне ім'я, прізвище)

РЕФЕРАТ

Пояснювальна записка: 51 с., 5 рис., 1 табл., 1 дод., 20 джерел.

БЕКТЕСТИНГ, ВІЗУАЛІЗАЦІЯ ДАНИХ, ЗАКОН ГУТЕНБЕРГА-РІХТЕРА, ЗЕМЛЕТРУСИ, МАРКОВСЬКИЙ ЛАНЦЮГ, ПРОГНОЗУВАННЯ, ПУАССОНІВСЬКИЙ ПРОЦЕС, СЕЙСМІЧНА АКТИВНІСТЬ, СЕЙСМІЧНА НЕБЕЗПЕКА, СТОХАСТИЧНІ МОДЕЛІ, SQLITE, UNITY.

Об'єкт дослідження – процес аналізу сейсмічної активності та просторово-часового розподілу землетрусів.

Предмет дослідження – стохастичні моделі та методи імовірнісного оцінювання сейсмічної небезпеки, а також засоби збору, обробки й візуалізації сейсмічних даних.

Мета роботи – розробка системи аналізу сейсмічної активності на основі стохастичних моделей, що забезпечує збір і обробку сейсмічних даних, оцінювання інтенсивності сейсмічного процесу, побудову імовірнісного прогнозу сейсмічної небезпеки з подальшою перевіркою його якості та візуальне представлення результатів.

Методи дослідження – методи математичного та стохастичного моделювання (марковський ланцюг, пуассонівський процес), закон Гутенберга-Ріхтера й метод максимальної правдоподібності для оцінювання його параметрів, методи просторово-часової агрегації та дедуплікації даних, методи ретроспективної верифікації прогнозу, а також методи візуалізації даних із використанням тривимірної інтерактивної моделі у середовищі Unity.

ABSTRACT

Bachelor's thesis contains: 51 pp., 5 fig., 1 tabl., 1 ann., 20 references.

BACKTESTING, DATA VISUALIZATION, EARTHQUAKES, FORECASTING, GUTENBERG-RICHTER LAW, MARKOV CHAIN, POISSON PROCESS, SEISMIC ACTIVITY, SEISMIC HAZARD, SQLITE, STOCHASTIC MODELS, UNITY.

Object of research – the process of analyzing seismic activity and the spatio-temporal distribution of earthquakes.

Subject of research – stochastic models and methods of probabilistic seismic hazard assessment, as well as tools for collecting, processing, and visualizing seismic data.

Aim of the work – development of a seismic activity analysis system based on stochastic models that provides collection and processing of seismic data, estimation of the intensity of the seismic process, construction of a probabilistic seismic hazard forecast with subsequent verification of its quality, and visual representation of the results.

Research methods – methods of mathematical and stochastic modeling (Markov chain, Poisson process), the Gutenberg-Richter law and the maximum likelihood method for estimating its parameters, methods of spatio-temporal aggregation and deduplication of data, methods of retrospective forecast verification, as well as data visualization methods using an interactive three-dimensional model in the Unity environment.

ЗМІСТ

Перелік умовних позначень, символів, одиниць, скорочень і термінів	7
Вступ.....	8
1 Аналіз предметної галузі	10
1.1 Особливості сейсмічної активності та її основні параметри	10
1.2 Аналіз існуючих рішень.....	14
1.3 Постановка задачі	18
2 Методи збору та обробки сейсмічних даних	22
2.1 Джерела даних та стандарт FDSN.....	22
2.2 Зведення даних до єдиного формату та дедуплікація.....	23
2.3 Просторово-часова агрегація та побудова часових рядів.....	24
3 Стохастичні моделі аналізу та прогнозування	26
3.1 Марковська модель станів сейсмічної активності	26
3.2 Закон Гутенберга-Ріхтера та оцінювання параметрів.....	27
3.3 Пуассонівська модель сейсмічної небезпеки.....	29
3.4 Метрики верифікації прогнозу.....	30
4 Програмна реалізація	32
4.1 Інструменти та архітектура системи.....	32
4.2 Організація зберігання даних	34
4.3 Реалізація обробки та моделей.....	35
4.4 Візуалізація результатів	39
Висновки.....	46
Перелік джерел посилання	48
Додаток А Відомість кваліфікаційної роботи.....	51

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

CSEP – Collaboratory for the Study of Earthquake Predictability – спільнота з вивчення передбачуваності землетрусів;

EMSC – European-Mediterranean Seismological Centre – Європейсько-Середземноморський сейсмологічний центр;

ETAS – Epidemic-Type Aftershock Sequence – модель епідемічного типу послідовності афтершоків;

FDSN – International Federation of Digital Seismograph Networks – Міжнародна федерація цифрових сейсмографічних мереж;

PSHA – Probabilistic Seismic Hazard Analysis – імовірнісний аналіз сейсмічної небезпеки;

SQL – Structured Query Language – мова структурованих запитів;

USGS – United States Geological Survey – геологічна служба США;

UTC – Coordinated Universal Time – всесвітній координований час.

ВСТУП

Сейсмічна активність є одним із найважливіших природних процесів, що безпосередньо впливає на безпеку населення, інфраструктури та навколишнього середовища. Землетруси можуть спричиняти значні руйнування, економічні втрати та загрозу життю людей, особливо в регіонах із підвищеною тектонічною активністю. У зв'язку з цим аналіз сейсмічних подій, виявлення закономірностей у їх просторово-часовому розподілі та оцінювання ймовірності подальшої активності є актуальними задачами сучасної сейсмології.

Сучасні сейсмічні каталоги містять великі обсяги даних про зареєстровані землетруси: час виникнення події, географічні координати епіцентру, глибину, магнітуду та інші параметри. Такі дані мають складну просторово-часову структуру, оскільки сейсмічні події не виникають повністю випадково, а часто утворюють кластери, пов'язані з головними поштовхами та післяшоковими послідовностями. Через це для їх аналізу недостатньо лише простих статистичних методів або візуального перегляду карти подій.

Одним із перспективних напрямів дослідження сейсмічної активності є використання стохастичних моделей, зокрема моделей точкових процесів. Такі моделі дозволяють описувати виникнення землетрусів як випадковий процес у часі та просторі, враховуючи як фонову сейсмічність, так і вплив уже зареєстрованих подій на появу нових. Особливе місце серед таких підходів займає ETAS-модель, яка широко застосовується для опису післяшовкових послідовностей та оцінювання інтенсивності сейсмічного процесу [2], [3].

Водночас задача прогнозування землетрусів залишається складною та обмеженою. Точне передбачення часу, місця й магнітуди майбутнього землетрусу є практично недосяжним у загальному випадку, тому сучасні дослідження орієнтуються переважно на імовірнісне прогнозування

сейсмічної активності. Такий підхід полягає не у визначенні точної майбутньої події, а в оцінюванні ймовірності підвищеної активності в певному регіоні або часовому інтервалі. Сучасні роботи також активно поєднують стохастичні моделі з методами машинного та глибокого навчання для аналізу великих сейсмічних даних [1], [5].

Окрему роль у дослідженні сейсмічної активності відіграє аналіз часових рядів. Кількість подій за певний проміжок часу, зміна середньої магнітуди, частота післяшоків та інші параметри можуть бути подані у вигляді часових залежностей. Аналіз таких рядів дозволяє виявляти тенденції, періоди підвищеної активності, а також оцінювати якість прогнозних моделей. Для цього можуть застосовуватися статистичні методи, моделі з довгою пам'яттю, ARIMA-підходи та нейромережеві архітектури [6], [7].

Не менш важливою складовою є візуалізація результатів аналізу. Сейсмічні дані мають просторову природу, тому їх зручно подавати у вигляді карт, теплових зон, часових графіків та інтерактивних 3D-сцен. Використання середовища Unity дає змогу створювати наочні інтерактивні модулі для відображення сейсмічних подій, їх параметрів та динаміки у часі. Це підвищує зрозумілість результатів аналізу та робить систему більш придатною для практичного використання [10], [11], [12].

Таким чином, розробка системи аналізу сейсмічної активності на основі стохастичних моделей є актуальним завданням, що поєднує методи математичного моделювання, статистичного аналізу, обробки часових рядів, прогнозування та візуалізації даних. Така система може бути використана для дослідження закономірностей у сейсмічних каталогах, оцінювання інтенсивності сейсмічного процесу та наочного представлення результатів у вигляді графіків, карт і тривимірних візуалізацій.

1 АНАЛІЗ ПРЕДМЕТНОЇ ГАЛУЗІ

1.1 Особливості сейсмічної активності та її основні параметри

Сейсмічна активність є природним геофізичним процесом, що проявляється у виникненні коливань земної поверхні внаслідок раптового вивільнення енергії у надрах Землі. Основною формою прояву сейсмічної активності є землетруси, які виникають переважно в зонах тектонічних розломів, на межах літосферних плит або в областях підвищеного геодинамічного напруження. Дослідження таких процесів має важливе значення для оцінювання сейсмічної небезпеки, планування інфраструктури та зменшення ризиків для населення.

Сейсмічні події характеризуються набором параметрів, які фіксуються у спеціалізованих каталогах землетрусів. До основних параметрів належать час виникнення події, географічні координати епіцентру, глибина гіпоцентру, магнітуда, а також регіон або сейсмічна зона, у якій було зафіксовано подію. Саме ці характеристики формують основу для подальшого статистичного аналізу, побудови часових рядів, оцінювання інтенсивності сейсмічного процесу та візуалізації просторового розподілу землетрусів.

Особливістю сейсмічних даних є їх просторово-часова природа. Землетруси не можна розглядати лише як незалежні окремі події, оскільки в багатьох випадках вони утворюють послідовності, пов'язані між собою у часі та просторі. Наприклад, після сильного головного поштовху часто спостерігається серія післяшоків, кількість і частота яких поступово зменшуються з часом. Така кластерна структура є однією з причин використання стохастичних моделей, які дозволяють описувати не лише сам факт виникнення події, а й імовірність появи наступних подій залежно від попередньої сейсмічної активності.

Для аналізу сейсмічної активності важливе значення має поняття магнітуди. Магнітуда відображає енергію, що вивільняється під час землетрусу, і є одним із ключових показників сили події. У сейсмічних каталогах зазвичай спостерігається значна кількість слабких землетрусів і набагато менша кількість сильних подій. Такий нерівномірний розподіл ускладнює аналіз, оскільки модель повинна враховувати як масові слабкі події, так і рідкісні, але потенційно небезпечні сильні землетруси.

Окрім магнітуди, важливим параметром є глибина землетрусу. Неглибокі землетруси зазвичай мають більший вплив на поверхню, тоді як глибокі події можуть відчуватися на значних відстанях, але часто мають інший характер поширення сейсмічних хвиль. У межах розроблюваної системи глибина може використовуватися як додаткова ознака для фільтрації, класифікації або візуального подання подій.

Ще одним важливим аспектом є часовий розподіл сейсмічних подій. Кількість землетрусів за певний проміжок часу може змінюватися нерівномірно: у деякі періоди спостерігається відносно затишшя, а в інші – різке зростання активності. Такі зміни можуть бути пов'язані з післяшокковими процесами, регіональними тектонічними особливостями або зміною умов спостереження. Через це сейсмічні дані доцільно аналізувати як часові ряди, що дозволяє виявляти тенденції, піки активності та періоди підвищеного ризику. Методи аналізу часових рядів активно застосовуються для дослідження довготривалих залежностей у сейсмічних даних та прогнозування їх подальшої динаміки [6], [7].

Сейсмічні події характеризуються набором параметрів, які фіксуються у спеціалізованих каталогах землетрусів. До основних параметрів належать час виникнення події, географічні координати епіцентру, глибина гіпоцентру, магнітуда, а також регіон або сейсмічна зона, у якій було зафіксовано подію (рисунок 1.1)

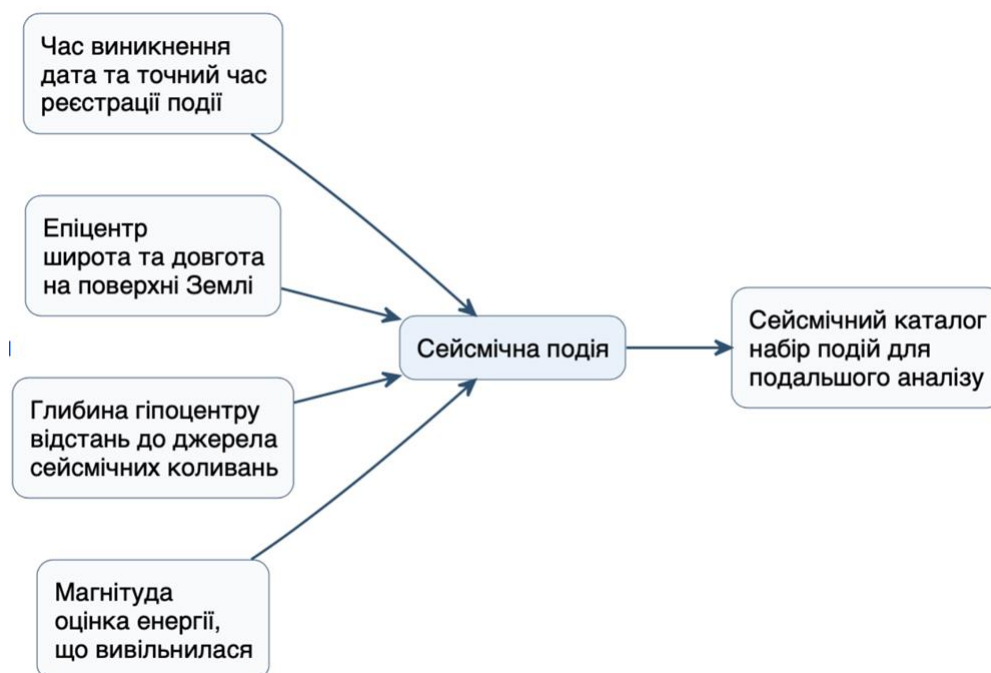


Рисунок 1.1 – Особливості сейсмічної активності та її основні параметри

Однією з фундаментальних емпіричних закономірностей сейсмології є розподіл частота-магнітуда, відомий як закон Гутенберга-Ріхтера. Він установлює лінійну залежність між магнітудою подій та десятковим логарифмом їх кількості:

$\lg N(\geq M) = a - bM$, де $N(\geq M)$ – число землетрусів магнітудою не меншою за M за визначений період; коефіцієнт a характеризує загальну сейсмічну продуктивність регіону, а коефіцієнт b (b -value) відображає співвідношення між кількістю слабких і сильних подій. Для більшості сейсмічно активних зон значення b близьке до одиниці [13]. Зазначена закономірність має ключове значення для імовірного оцінювання сейсмічної небезпеки, оскільки дозволяє екстраполювати частоту рідкісних сильних землетрусів на основі спостережень за численними слабкими подіями.

Практичне застосування закону Гутенберга-Ріхтера ускладнюється неповнотою сейсмічних каталогів у діапазоні малих магнітуд. Через обмежену чутливість та нерівномірне просторове розміщення сейсмічних

станцій частина слабких подій залишається незареєстрованою, унаслідок чого емпіричний розподіл частота–магнітуда відхиляється від лінійного нижче певного порогу. Цей поріг визначають як магнітуду повноти каталогу (M_c) – найменшу магнітуду, починаючи з якої в межах досліджуваного регіону та періоду реєструються практично всі сейсмічні події [14]. Оцінювання параметрів моделі за подіями, слабшими за (M_c), спричиняє систематичне заниження коефіцієнта b , тому визначення магнітуди повноти є обов'язковим етапом попередньої обробки даних перед статистичним аналізом. У межах розроблюваної системи величина M_c задається як параметр і використовується для відсікання неповної частини каталогу під час оцінювання інтенсивності сейсмічного процесу.

Важливою кількісною характеристикою сейсмічного режиму регіону є інтенсивність сейсмічного процесу – середня кількість подій заданого магнітудного діапазону за одиницю часу. У поєднанні із законом Гутенберга-Ріхтера інтенсивність дає змогу переходити від спостережуваної частоти слабких подій до оцінки очікуваної частоти сильних землетрусів, магнітуда яких перевищує заданий рівень. На основі інтенсивності визначають період повторюваності – середній інтервал часу між подіями магнітудою не меншою за задану, а також імовірність виникнення принаймні однієї такої події протягом обраного часового горизонту. Саме ці величини становлять основу імовірнісного оцінювання сейсмічної небезпеки [15] та використовуються в розроблюваній системі для побудови прогнозних карт.

Сучасні підходи до аналізу сейсмічної активності дедалі частіше орієнтуються не на точне передбачення окремого землетрусу, а на імовірнісне оцінювання майбутньої активності. Це пов'язано з тим, що землетруси є складними природними процесами, які залежать від великої кількості факторів і мають значний рівень невизначеності. Тому більш реалістичним завданням є побудова моделей, які дозволяють оцінити

інтенсивність сейсмічного процесу, визначити області підвищеної активності та проаналізувати ймовірність появи нових подій у певному часовому або просторовому інтервалі [1].

Для наочного подання сейсмічної активності доцільно використовувати карти епіцентрів, графіки зміни кількості подій у часі, гістограми магнітуд, теплові карти та інтерактивні 3D-візуалізації. Такі засоби дозволяють краще зрозуміти структуру даних, виявити просторові кластери та оцінити зміну активності в часі. Візуалізація є важливою складовою системи аналізу, оскільки вона робить результати математичного моделювання більш зрозумілими для користувача та спрощує інтерпретацію отриманих результатів [10].

Таким чином, сейсмічна активність є складним просторово-часовим процесом, для аналізу якого необхідно враховувати координати подій, час їх виникнення, магнітуду, глибину та взаємозв'язки між подіями. Наявність випадковості, кластеризації та часової залежності робить доцільним використання стохастичних моделей, які можуть описувати сейсмічний процес з урахуванням невизначеності та взаємного впливу подій. Саме ці особливості визначають необхідність розробки системи, що поєднує статистичний аналіз, математичне моделювання, прогнозування та візуалізацію сейсмічних даних.

1.2 Аналіз існуючих рішень

Для аналізу сейсмічної активності використовуються різні підходи, які можна умовно поділити на статистичні методи, стохастичні моделі, методи аналізу часових рядів, алгоритми машинного навчання та засоби візуалізації даних. Кожен із цих напрямів має власні переваги та обмеження, тому під час розробки системи доцільно враховувати їх комплексно.

Найпростішим підходом до аналізу сейсмічних даних є статистичне дослідження каталогів землетрусів. У цьому випадку аналізуються такі

параметри, як кількість подій за певний період, розподіл магнітуд, глибина гіпоцентрів, просторове розміщення епіцентрів та частота виникнення подій у різних регіонах. Такі методи дозволяють отримати загальне уявлення про характер сейсмічної активності, однак вони не завжди враховують залежність між подіями та складну просторово-часову структуру сейсмічного процесу.

Більш розвиненим напрямом є використання стохастичних моделей, які дають змогу розглядати виникнення землетрусів як випадковий процес у часі та просторі. Однією з найбільш поширених моделей такого типу є ETAS-модель, яка описує сейсмічність як поєднання фонові активності та подій, що можуть породжувати післяшоки. У сучасних дослідженнях ETAS-модель розширюється за допомогою байєсівських методів та напівпараметричних підходів, що дозволяє краще враховувати невизначеність параметрів і просторову неоднорідність сейсмічності [2], [3].

Окремий напрям пов'язаний із підвищенням обчислювальної ефективності стохастичних моделей. Оскільки реальні сейсмічні каталоги можуть містити значну кількість подій, оцінювання параметрів моделей потребує значних обчислювальних ресурсів. У роботі [4] розглянуто підхід *simulation-based inference* для моделі ETAS, який дозволяє виконувати масштабоване оцінювання параметрів без прямого використання класичної функції правдоподібності. Це є важливим для практичних систем, які мають працювати з великими обсягами сейсмічних даних.

Поряд із моделями точкових процесів для аналізу сейсмічної активності застосовують марковські моделі, у яких стан сейсмічного режиму регіону дискретизують на скінченну кількість рівнів активності (наприклад, низький, середній і високий), а зміну активності в часі описують матрицею ймовірностей переходів між цими станами. Такий підхід дозволяє оцінювати ймовірність переходу регіону в стан підвищеної активності, а також визначати стаціонарний розподіл, що характеризує довгострокову частку часу перебування в кожному стані. Перевагою

марковських моделей є прозорість та інтерпретованість результатів порівняно з методами машинного навчання, оскільки матриця переходів безпосередньо відображає динаміку сейсмічного режиму [16].

Окремим напрямом, орієнтованим безпосередньо на оцінювання сейсмічної небезпеки, є імовірнісний аналіз сейсмічної небезпеки (probabilistic seismic hazard analysis, PSHA). У його основі лежить подання виникнення сильних землетрусів як пуассонівського процесу, інтенсивність якого визначають за законом Гутенберга–Ріхтера, екстрапольованим на діапазон великих магнітуд. Такий підхід дозволяє обчислювати ймовірність перевищення заданого рівня магнітуди в межах певного регіону та часового горизонту, а також відповідний період повторюваності [15]. На відміну від детермінованого прогнозування конкретної події, PSHA дає просторовий розподіл рівня небезпеки, який зручно подавати у вигляді карт. Саме цей підхід покладено в основу побудови прогнозних карт у розроблюваній системі.

Поряд зі стохастичними моделями активно застосовуються методи аналізу часових рядів. Сейсмічну активність можна подати у вигляді послідовності значень, наприклад кількості землетрусів за добу, тиждень або місяць. Для таких даних використовуються авторегресійні моделі, методи виявлення довготривалих залежностей, ARIMA-підходи та інші статистичні методи прогнозування. Зокрема, у роботі [6] розглянуто ефекти довгої пам'яті в сейсмічних даних, а в роботі [7] виконано порівняння класичних методів аналізу часових рядів із методами глибокого навчання.

Сучасні дослідження також приділяють значну увагу методам машинного та глибокого навчання. Такі підходи дозволяють автоматично виявляти складні закономірності у великих масивах сейсмічних даних. Наприклад, LSTM-мережі можуть використовуватися для аналізу послідовностей подій у часі, оскільки вони здатні враховувати попередні стани часового ряду [8]. Також застосовуються гібридні архітектури, які

поєднують згорткові нейронні мережі та рекурентні моделі для аналізу часових залежностей у сейсмічних даних [9].

Разом із тим, використання методів машинного навчання для прогнозування землетрусів має певні обмеження. Сейсмічні події є складними природними явищами, тому точне передбачення часу, місця та магнітуди майбутнього землетрусу залишається надзвичайно складним завданням. Через це сучасні підходи частіше спрямовані не на детерміноване передбачення конкретної події, а на імовірнісне прогнозування рівня сейсмічної активності або інтенсивності сейсмічного процесу в певному регіоні [1], [5].

Окремої уваги потребує формалізована перевірка якості прогнозів сейсмічної активності. Оскільки прогнози формулюються в імовірнісних термінах, для їх оцінювання застосовують методи верифікації, зокрема діаграму надійності (reliability diagram), яка порівнює передбачену ймовірність із фактичною частотою настання подій, та оцінку Брайєра (Brier score), що вимірює середньоквадратичне відхилення ймовірнісного прогнозу від фактичного результату. Систематичний підхід до незалежного тестування прогнозних моделей на прихованих даних розвивається в межах міжнародної ініціативи CSEP (Collaboratory for the Study of Earthquake Predictability) [17]. Застосування таких методів дозволяє кількісно оцінити не лише точність, а й калібрування прогнозу, що є необхідним для об'єктивного порівняння моделей.

Важливою складовою сучасних систем аналізу сейсмічної активності є візуалізація результатів. Сейсмічні дані мають просторовий і часовий характер, тому їх доцільно подавати у вигляді карт епіцентрів, теплових карт, часових графіків, 3D-сцен та інтерактивних панелей. У роботі [10] розглянуто використання інтерактивної візуалізації для складних сейсмічних даних і моделей. Для побудови більш наочних програмних рішень можуть застосовуватися ігрові рушії, зокрема Unity, що дозволяє

створювати інтерактивні тривимірні сцени для представлення результатів моделювання [11], [12].

Таким чином, аналіз існуючих підходів показує, що для розробки ефективної системи доцільно поєднувати кілька напрямів: статистичний аналіз каталогів землетрусів, стохастичне моделювання сейсмічного процесу, аналіз часових рядів, імовірнісне прогнозування та візуалізацію результатів. Саме така комбінація дозволяє не лише досліджувати історичні сейсмічні дані, а й оцінювати інтенсивність майбутньої активності та подавати результати у зручній для користувача формі. У даній роботі обрано поєднання марковської моделі станів сейсмічної активності, імовірнісного аналізу небезпеки на основі закону Гутенберга–Ріхтера та пуассонівського процесу, а також формалізованої верифікації прогнозу на незалежних даних.

1.3 Постановка задачі

Аналіз сейсмічної активності є важливим завданням у сфері геофізичних досліджень, оскільки землетруси мають складну просторово-часову природу та можуть становити значну небезпеку для населення, інфраструктури й навколишнього середовища. Сучасні сейсмічні каталоги містять великі обсяги інформації про час виникнення подій, координати епіцентрів, глибину та магнітуду землетрусів. Проте для ефективного використання цих даних необхідні спеціалізовані методи аналізу, які дозволяють виявляти закономірності, оцінювати інтенсивність сейсмічного процесу та подавати результати у зручній для користувача формі.

Традиційний аналіз сейсмічних даних часто обмежується побудовою карт епіцентрів, графіків кількості подій у часі та статистичним описом магнітуд. Такі підходи є корисними для первинного дослідження, однак вони не завжди дозволяють повною мірою врахувати випадковість, кластеризацію та взаємний вплив сейсмічних подій. У зв'язку з цим

актуальним є застосування стохастичних моделей, зокрема моделей точкових процесів, які дозволяють розглядати сейсмічну активність як випадковий процес у часі та просторі [2], [3].

Особливе значення для даної роботи має застосування прозорих та інтерпретованих стохастичних моделей, придатних як для практичної реалізації, так і для кількісної перевірки якості прогнозу. Сейсмічний режим регіону подають у вигляді послідовності дискретних станів активності, зміну яких у часі описують марковським ланцюгом із матрицею ймовірностей переходів [16]. Для оцінювання рівня сейсмічної небезпеки застосовують імовірнісний підхід на основі закону Гутенберга–Ріхтера та пуассонівського процесу, що дає змогу обчислювати ймовірність виникнення сильних подій і період їх повторюваності [15]. Вхідними даними для марковської моделі слугують часові ряди кількості подій, побудовані за визначені інтервали часу [6].

Метою роботи є розробка системи аналізу сейсмічної активності на основі стохастичних моделей, яка забезпечує збір та обробку сейсмічних даних, аналіз просторово-часового розподілу землетрусів, оцінювання інтенсивності сейсмічного процесу, побудову імовірнісного прогнозу сейсмічної небезпеки та візуалізацію отриманих результатів.

Для досягнення поставленої мети необхідно виконати наступні задачі:

- проаналізувати предметну область дослідження, визначити основні параметри сейсмічних подій та особливості їх просторово-часового розподілу;
- дослідити існуючі підходи до аналізу сейсмічної активності, зокрема стохастичні моделі, марковські ланцюги, імовірнісний аналіз сейсмічної небезпеки та методи верифікації прогнозів;
- обрати джерела сейсмічних даних, забезпечити збір подій із відкритих каталогів та зведення різнорідних даних до єдиного формату;
- реалізувати попередню обробку сейсмічних даних, зокрема дедуплікацію подій, зафіксованих різними джерелами, фільтрацію за

магнітудою та часом, а також збереження у локальній базі даних для подальшого аналізу;

- побудувати просторово-часову агрегацію даних і часові ряди сейсмічної активності за визначеними регіонами та інтервалами часу;

- реалізувати марковську модель сейсмічної активності з дискретними станами та оцінити матрицю ймовірностей переходів між ними;

- реалізувати оцінювання параметрів закону Гутенберга–Ріхтера та побудову імовірнісного прогнозу сейсмічної небезпеки на основі пуассонівського процесу;

- виконати ретроспективну перевірку (бектестинг) отриманих прогнозів на незалежних даних із застосуванням діаграми надійності та оцінки Брайєра;

- розробити засоби візуалізації сейсмічних даних і результатів прогнозування у вигляді інтерактивної тривимірної карти із використанням середовища Unity [11], [12];

- реалізувати програмний прототип системи аналізу сейсмічної активності та оцінити результати його роботи.

У межах розроблюваної системи вхідними даними є сейсмічні каталоги відкритого доступу, що містять інформацію про зареєстровані землетруси. Дані збираються з двох незалежних джерел – Геологічної служби США (USGS) та Європейсько-Середземноморського сейсмологічного центру (EMSC), – після чого зводяться до єдиного формату з усуненням дублікатів. Основними параметрами кожної події є дата й час виникнення, широта, довгота, глибина та магнітуда. На основі цих даних система забезпечує фільтрацію подій, просторово-часову агрегацію, побудову часових рядів, оцінювання параметрів стохастичних моделей, імовірнісне прогнозування сейсмічної небезпеки та візуалізацію отриманих результатів.

Очікуваним результатом роботи є програмна система, яка дозволяє виконувати комплексний аналіз сейсмічної активності на основі стохастичних моделей. Така система поєднує збір і попередню обробку сейсмічних даних, математичне моделювання, імовірнісне прогнозування та візуалізацію результатів, а якість отриманих прогнозів підтверджується ретроспективною перевіркою на незалежних даних. Це дає змогу не лише досліджувати закономірності сейсмічних подій, а й оцінювати достовірність побудованих прогнозів і представляти результати у зрозумілій графічній формі.

2 МЕТОДИ ЗБОРУ ТА ОБРОБКИ СЕЙСМІЧНИХ ДАНИХ

Ефективність системи аналізу сейсмічної активності значною мірою визначається якістю вхідних даних та коректністю їх попередньої обробки. У цьому розділі розглянуто методи отримання сейсмічних даних із відкритих каталогів, зведення різномірної інформації до єдиного формату з усуненням дублікатів, а також просторово-часової агрегації подій, що є основою для подальшого стохастичного моделювання.

2.1 Джерела даних та стандарт FDSN

Первинним етапом аналізу сейсмічної активності є отримання даних про зареєстровані землетруси із сейсмічних каталогів. Сейсмічний каталог – це структурований набір записів про сейсмічні події, кожен із яких містить час виникнення, географічні координати епіцентру, глибину гіпоцентру та магнітуду. Провідними постачальниками таких даних у відкритому доступі є Геологічна служба США (United States Geological Survey, USGS) та Європейсько-Середземноморський сейсмологічний центр (European-Mediterranean Seismological Centre, EMSC), які підтримують глобальні та регіональні каталоги, що безперервно поповнюються [18], [19].

Різні сейсмологічні центри історично використовували власні формати зберігання та розповсюдження даних, що ускладнювало їх спільне використання. Для розв'язання цієї проблеми Міжнародною федерацією цифрових сейсмографічних мереж (International Federation of Digital Seismograph Networks, FDSN) було розроблено стандартизований веб-протокол доступу до сейсмічних даних – FDSN Event Web Service. Цей стандарт визначає єдиний формат запитів та відповідей, завдяки чому дані з різних каталогів можна отримувати уніфікованим способом [20].

Запит до сервісу формується як HTTP-звернення з набором параметрів, що обмежують вибірку подій: початковий і кінцевий момент

часу, мінімальна та максимальна магнітуда, а також географічні межі регіону. У відповідь сервіс повертає перелік подій у структурованому форматі, зокрема GeoJSON, де кожна подія містить час, координати, глибину та магнітуду. Оскільки і USGS, і EMSC підтримують стандарт FDSN, збір даних з обох джерел здійснюється за єдиною схемою, що спрощує реалізацію та зменшує кількість специфічного для джерела коду.

Важливою особливістю сервісів FDSN є обмеження на максимальну кількість подій, що повертаються в межах одного запиту. Для великих часових діапазонів це унеможлиблює одночасне отримання всього каталогу, тому збір історичних даних виконують послідовно, розбиваючи весь період на часові вікна. У разі перевищення допустимого обсягу вікно рекурсивно зменшують, доки кількість подій у ньому не відповідатиме обмеженням сервісу. Такий підхід забезпечує повне та коректне отримання каталогу за тривалий проміжок часу.

2.2 Зведення даних до єдиного формату та дедуплікація

Дані, отримані з різних каталогів, попри спільний стандарт доступу, можуть відрізнятися представленням окремих полів. Зокрема, час події може подаватися у вигляді кількості мілісекунд від початку епохи або у форматі ISO 8601, а назви й одиниці окремих параметрів – різнитися між джерелами. Тому перед аналізом необхідне зведення всіх записів до єдиної внутрішньої моделі даних.

Уніфікована модель події містить фіксований набір полів: час виникнення у форматі всесвітнього координованого часу (UTC), широту та довготу епіцентру, глибину гіпоцентру, магнітуду, тип магнітуди, а також ідентифікатор джерела. Приведення часу до єдиної шкали та узгодження одиниць вимірювання є обов'язковою умовою коректного порівняння й об'єднання подій із різних каталогів.

Оскільки один і той самий землетрус часто реєструється одночасно кількома мережами спостереження, об'єднання даних із двох джерел призводить до появи дублікатів – записів, що описують ту саму фізичну подію. Наявність таких дублікатів спотворює статистичні оцінки, зокрема завищує кількість подій, тому необхідним етапом обробки є дедуплікація.

Виявлення дублікатів ґрунтується на просторово-часовій близькості подій: два записи з різних джерел вважають однією подією, якщо різниця в часі їх виникнення, відстань між епіцентрами та різниця магнітуд одночасно не перевищують заданих порогів. Відстань між епіцентрами обчислюють як довжину дуги великого кола на поверхні сфери за формулою Гаверсина:

$$a = \sin^2(\Delta\varphi/2) + \cos \varphi_1 \cdot \cos \varphi_2 \cdot \sin^2(\Delta\lambda/2),$$

$$d = 2R \cdot \operatorname{atan2}(\sqrt{a}, \sqrt{1-a}), \quad (2.1)$$

де φ_1, φ_2 – широти двох точок;

$\Delta\varphi, \Delta\lambda$ – різниці широт і довгот відповідно;

R – середній радіус Землі.

На відміну від повного видалення дублікатів, доцільним є підхід, за якого обидва записи зберігаються, але пов'язуються спільним ідентифікатором кластера. Це дозволяє надалі порівнювати оцінки параметрів однієї події, отримані різними мережами, і водночас уникати подвійного врахування під час статистичного аналізу. Одне з джерел при цьому обирають як пріоритетне для представлення події.

2.3 Просторово-часова агрегація та побудова часових рядів

Каталог окремих подій є надто деталізованим для регіонального аналізу сейсмічного режиму. Для виявлення просторових закономірностей та побудови часових рядів активності дані піддають просторово-часовій

агрегації, під час якої окремі події групують за географічними комірками та часовими інтервалами.

Просторову агрегацію виконують шляхом розбиття поверхні Землі на регулярну сітку в географічних координатах. За еківектангулярного розбиття з кроком Δ градусів індекси комірки, до якої належить подія з координатами (φ, λ) , визначають як:

$$i_{lat} = \lfloor (\varphi + 90) / \Delta \rfloor, \quad i_{lon} = \lfloor (\lambda + 180) / \Delta \rfloor. \quad (2.2)$$

Слід зауважити, що за рівного кроку в градусах площа комірок зменшується з наближенням до полюсів; для регіонального аналізу цим ефектом можна знехтувати, проте за потреби застосовують рівноплощинні сітки.

Часову агрегацію виконують поділом усього періоду спостережень на послідовні інтервали фіксованої тривалості Δt . Індекс інтервалу, до якого належить подія з часом t , обчислюють як:

$$k = \lfloor (t - t_0) / \Delta t \rfloor, \quad (2.3)$$

де t_0 – початковий момент відліку. Для кожної комірки та кожного часового інтервалу підраховують кількість подій, а також агреговані характеристики магнітуди.

Результатом агрегації є, для кожної комірки, послідовність значень кількості подій за послідовні часові інтервали – часовий ряд сейсмічної активності. Такі ряди є основним входом для марковської моделі, у якій рівень активності дискретизують на стани. Окремо для кожної комірки формують сумарні характеристики магнітуд, необхідні для оцінювання параметрів закону Гутенберга–Ріхтера.

Просторово-часова агрегація суттєво зменшує обсяг даних, які безпосередньо беруть участь у моделюванні: замість мільйонів окремих подій аналізують компактний набір рядів за комірками. Це підвищує обчислювальну ефективність та створює основу для застосування стохастичних моделей, розглянутих у наступному розділі.

3 СТОХАСТИЧНІ МОДЕЛІ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ

У цьому розділі розглянуто математичний апарат, покладений в основу аналізу та прогнозування сейсмічної активності: марковську модель для опису зміни рівня активності регіону, закон Гутенберга-Ріхтера й пуассонівський процес для імовірнісного оцінювання сейсмічної небезпеки, а також метрики, за якими визначають якість отриманих прогнозів.

3.1 Марковська модель станів сейсмічної активності

Марковський ланцюг – це стохастичний процес зі скінченною множиною станів, у якому ймовірність переходу в наступний стан залежить лише від поточного стану й не залежить від передісторії (марковська властивість). Стосовно сейсмічної активності стани ланцюга відповідають рівням активності регіону, а переходи – зміні цього рівня в часі [16].

Для побудови ланцюга рівень активності кожного часового інтервалу дискретизують на скінченну кількість станів. У розробленій моделі використано три стани – низький, середній та високий, – межі яких визначають за терцилями розподілу кількості подій у межах самої комірки. Застосування відносних, індивідуальних для комірки, порогів забезпечує, що поняття «висока активність» відповідає підвищеній активності саме для цього регіону, а не глобальному масштабу.

Динаміку переходів між станами описують матрицею переходів P розміром $N \times N$, елементи якої є умовними ймовірностями переходу зі стану i у стан j .

Оцінку цих ймовірностей отримують за спостереженою послідовністю станів як відношення кількості відповідних переходів:

$$P_{ij} = \frac{n_{ij}}{\sum_k n_{ik}} \quad (3.1)$$

де n_{ij} – кількість спостережених переходів зі стану i у стан j сума в знаменнику дорівнює загальній кількості переходів зі стану i . Кожен рядок матриці задовольняє умову нормування, тобто матриця є рядково-стохастичною.

Прогноз розподілу станів на n кроків уперед обчислюють як добуток початкового розподілу ймовірностей на n -й степінь матриці переходів:

$$\pi^{(t+n)} = \pi^{(t)} P^n, \quad (3.2)$$

де $\pi^{(t)}$ – вектор-рядок імовірностей перебування в кожному стані в момент t . Це дозволяє оцінювати ймовірність переходу регіону в стан підвищеної активності на задану кількість інтервалів уперед.

Довгострокову поведінку ланцюга характеризує стаціонарний розподіл π – розподіл, що не змінюється під дією матриці переходів:

$$\pi = \pi P, \quad \sum_i \pi_i = 1 \quad (3.3)$$

Стаціонарний розподіл інтерпретують як частку часу, яку регіон у довгостроковій перспективі проводить у кожному зі станів. Переважання діагональних елементів матриці переходів свідчить про інерційність сейсмічного режиму – схильність до тривалих періодів однакового рівня активності, що відповідає відомій кластеризації сейсмічних подій у часі.

3.2 Закон Гутенберга-Ріхтера та оцінювання параметрів

Основою імовірнісного оцінювання сейсмічної небезпеки є закон Гутенберга-Ріхтера (1.1), який пов'язує магнітуду подій із логарифмом їх кількості. Параметр a характеризує загальну продуктивність регіону, а параметр b визначає співвідношення між кількістю слабких і сильних подій.

Практичне застосування закону потребує оцінювання цих параметрів за наявним каталогом.

Коефіцієнт b оцінюють методом максимальної правдоподібності за формулою Акі:

$$b = \frac{\lg e}{\bar{M} - M_c + \Delta M/2} \quad (3.4)$$

де $\lg e \approx 0,4343$ – десятковий логарифм основи натурального логарифма;

$\bar{M}$ – середня магнітуда подій, не слабших за магнітуду повноти каталогу M_c ;

ΔM – крок дискретизації магнітуди (зазвичай 0,1).

Доданок $\Delta M/2$ враховує округлення магнітуд і зменшує зміщення оцінки.

Коефіцієнт a визначають за загальною кількістю подій, не слабших за магнітуду повноти:

$$a = \lg N(\geq M_c) + b M_c \quad (3.5)$$

де $N(\geq M_c)$ – кількість подій магнітудою не менше M_c .

Коректність обох оцінок безпосередньо залежить від правильного визначення магнітуди повноти M_c , оскільки врахування неповної частини каталогу зміщує оцінку коефіцієнта b у бік заниження [14].

Оцінені параметри дозволяють екстраполювати частоту подій на діапазон великих магнітуд, для яких прямих спостережень недостатньо. Саме ця можливість є основою для переходу до пуассонівської моделі прогнозу [13].

3.3 Пуассонівська модель сейсмічної небезпеки

Виникнення сильних землетрусів у межах регіону моделюють як пуассонівський процес – потік випадкових подій, що відбуваються незалежно одна від одної з деякою середньою інтенсивністю. Такий підхід є основою класичного імовірнісного аналізу сейсмічної небезпеки [15].

Річну інтенсивність подій цільової магнітуди M_t визначають екстраполяцією за законом Гутенберга-Ріхтера:

$$\lambda(\geq M_t) = \lambda(\geq M_c) \cdot 10^{-b(M_t - M_c)}, \quad \lambda(\geq M_c) = \frac{N(\geq)}{T} \quad (3.6)$$

де $\lambda(\geq M_c)$ – річна інтенсивність подій, не слабших за магнітуду повноти;

T – тривалість періоду спостережень у роках.

Імовірність виникнення принаймні однієї події магнітудою не менше M_t протягом часового горизонту H років обчислюють за розподілом Пуассона:

$$P = 1 - e^{-\lambda H}, \quad (3.7)$$

де $\lambda = \lambda(\geq M_t)$. Ця величина безпосередньо характеризує рівень сейсмічної небезпеки регіону на заданий проміжок часу.

Оберненою до інтенсивності характеристикою є середній період повторюваності:

$$T_r = \frac{1}{\lambda} \quad (3.8)$$

який визначає середній інтервал часу між подіями цільової магнітуди.

Слід зазначити, що припущення про незалежність подій, покладене в основу пуассонівської моделі, є спрощенням: реальна сейсмічність

характеризується кластеризацією, зокрема післяшовковими послідовностями. Це зумовлює потребу в кількісній перевірці якості прогнозу, розглянутій у наступному підрозділі.

3.4 Метрики верифікації прогнозу

Оскільки прогноз сейсмічної небезпеки має імовірнісний характер, його якість неможливо оцінити простим порівнянням «справдилося / не справдилося». Для об'єктивного оцінювання застосовують формалізовані метрики верифікації, а сам прогноз перевіряють ретроспективно – на даних, які не використовувалися під час його побудови [17].

Наочним засобом оцінювання є діаграма надійності, на якій за віссю абсцис відкладають прогнозовану ймовірність, а за віссю ординат – фактичну частоту настання подій для груп прогнозів із близькими значеннями ймовірності. Розташування точок поблизу діагоналі свідчить про добре калібрування прогнозу, тобто відповідність заявлених імовірностей реальним частотам.

Кількісною мірою якості імовірнісного прогнозу є оцінка Брайєра – середній квадрат відхилення прогнозованої ймовірності від фактичного результату:

$$BS = \frac{1}{N} \sum (p_i - o_i)^2 \quad (3.9)$$

де p_i – прогнозована ймовірність для i -ї комірки;

o_i – фактичний результат (1, якщо подія сталася, і 0 – якщо ні);

N – кількість комірок. Менше значення оцінки відповідає точнішому прогнозу.

Для порівняння прогнозу з тривіальним (кліматологічним) прогнозом, що використовує лише середню частоту подій, застосовують оцінку майстерності Брайєра:

$$BSS = 1 - \frac{BS}{BS_{\text{ref}}}, \quad BS_{\text{ref}} = \bar{p}(1 - \bar{p}) \quad (3.10)$$

де $\bar{p}$ – базова (середня) частота настання подій. Додатне значення BSS свідчить, що модель перевершує тривіальний прогноз, а значення, близьке до одиниці, – про високу якість прогнозу.

Застосування описаних метрик дозволяє не лише встановити факт працездатності моделі, а й кількісно оцінити ступінь її переваги над базовим прогнозом та виявити систематичні відхилення калібрування. Результати такої перевірки для розробленої системи наведено в наступному розділі.

4 ПРОГРАМНА РЕАЛІЗАЦІЯ

4.1 Інструменти та архітектура системи

Систему реалізовано у вигляді настільного застосунку на ігровому рушії Unity з використанням мови програмування C# (платформа .NET). Вибір Unity зумовлений потребою в інтерактивній тривимірній візуалізації сейсмічних даних на моделі земної кулі та можливістю виконання застосунку на настільних платформах.

Для забезпечення модульності та керованості коду використано низку бібліотек. Впровадження залежностей (dependency injection) реалізовано за допомогою контейнера Zenject, що дозволяє розв'язувати залежності між компонентами без жорсткого зв'язування. Асинхронні операції – мережеві запити та обчислення – виконано на основі бібліотеки UniTask, яка забезпечує неблокуючу роботу без зупинки основного потоку. Для зберігання даних застосовано вбудовану базу даних SQLite через об'єктно-реляційну бібліотеку sqlite-net.

Архітектуру системи побудовано за багатошаровим принципом із чітким розподілом відповідальності між шарами збору даних, зберігання, аналітики та візуалізації. Кожен функціональний модуль оформлено за патерном «Представлення – Медіатор – Сервіс»: представлення відповідає лише за відображення й генерує події користувача, медіатор пов'язує представлення з бізнес-логікою через підписку на ці події, а сервіси реалізують обчислення. Такий поділ спрощує супровід і розвиток системи.

Приклад реалізації медіатора, що пов'язує панель фільтрів із шаром візуалізації, наведено в лістингу 4.1. Медіатор отримує залежності через конструктор, підписується на подію застосування фільтра в методі Initialize та звільняє підписку в Dispose.

Лістинг 4.1 – Медіатор, що пов'язує представлення з логікою

```

public class EarthquakeMediator : IInitializable,
IDisposable
{
    private readonly EarthquakeFilterView _filterView;
    private readonly EarthquakeLayerController _layer;
    private readonly EarthquakeVisualConfig _config;

    public EarthquakeMediator(EarthquakeFilterView
filterView,
                                EarthquakeLayerController
layer,
                                EarthquakeVisualConfig
config)
    {
        _filterView = filterView;
        _layer = layer;
        _config = config;
    }

    public void Initialize() => _filterView.ApplyRequested
+= OnApplyRequested;
    public void Dispose()    => _filterView.ApplyRequested
-= OnApplyRequested;

    private void OnApplyRequested()
    {
        _layer.Apply(new EarthquakeVisualFilter
        {
            FromUtc = _filterView.FromUtc,
            ToUtc    = _filterView.ToUtc,
            MinMagnitude = _config.DefaultMinMagnitude
        });
    }
}

```

4.2 Організація зберігання даних

Зібрані сейсмічні дані зберігаються у вбудованій реляційній базі даних SQLite, файл якої розміщується в каталозі даних застосунку. Використання вбудованої бази замість повторного звернення до зовнішніх сервісів при кожному аналізі забезпечує накопичення історії спостережень, можливість повторного використання даних, а також незалежність аналітики від доступності мережі. SQLite обрано як таку, що не потребує окремого сервера, ефективно працює з великими обсягами даних і виконує запити за індексованими полями за мілісекунди. Загальний обсяг зібраного каталогу становить близько 2 млн подій.

Кожна подія зберігається у вигляді запису уніфікованої моделі, структуру якої наведено в лістингу 4.2. Первинним ключем запису є рядок виду «джерело:ідентифікатор події», що робить повторний імпорт ідемпотентним: повторне завантаження тієї самої події не створює дублікату, а оновлює наявний запис. Для полів, за якими виконуються вибірки (час, магнітуда, ідентифікатор кластера), створено індекси, що прискорює запити.

Лістинг 4.2 – Модель запису сейсмічної події:

```
public class EarthquakeRecord
{
    [PrimaryKey] public string Id { get; set; } //
    "{Source}:{SourceEventId}"
    [Indexed] public string Source { get; set; }
    public string SourceEventId { get; set; }
    [Indexed] public long OriginTimeUnixMs { get; set; }
}

    public double Latitude { get; set; }
    public double Longitude { get; set; }
    public double DepthKm { get; set; }
    [Indexed] public double Magnitude { get; set; }
```

Продовження лістингу 4.2

```

        public string MagnitudeType { get; set; }
        [Indexed]    public string ClusterId { get; set; }    //
зв'язок дублікатів
    }

```

Зіставлення класу з таблицею бази даних виконує об'єктно-реляційна бібліотека `sqlite-net` за допомогою атрибутів `PrimaryKey` та `Indexed`. Запис або оновлення подій здійснюється операцією `INSERT OR REPLACE`, що разом із природним ключем забезпечує коректне об'єднання даних із різних джерел без появи дублікатів.

Ключовим рішенням для забезпечення продуктивності аналітики є виконання агрегації даних на боці бази даних. Оскільки каталог містить мільйони записів, їх завантаження в оперативну пам'ять для подальшого групування було б неефективним. Натомість просторово-часову агрегацію виконують безпосередньо SQL-запитом з оператором `GROUP BY`, який повертає вже згруповані за комірками та часовими інтервалами дані (лістинг 4.3).

Лістинг 4.3 – Агрегація подій на боці бази даних:

```

SELECT CAST((Latitude + 90.0) / ? AS INTEGER) AS ClatIndex,
       CAST((Longitude + 180.0) / ? AS INTEGER) AS
ClonIndex,
       CAST((OriginTimeUnixMs - ?) / ? AS INTEGER) AS
BinIndex,
       COUNT(*) AS EventCount,
       MAX(Magnitude) AS MaxMagnitude
FROM EarthquakeRecord
WHERE Magnitude >= ?
GROUP BY ClatIndex, ClonIndex, BinIndex;

```

4.3 Реалізація обробки та моделей

Збір даних із каталогів здійснюється через веб-сервіс `FDSN` послідовними часовими вікнами. Оскільки сервіс обмежує кількість подій

на один запит, реалізовано адаптивне розбиття: у разі перевищення ліміту (відповідь із кодом 400) вікно рекурсивно ділиться навпіл, доки кількість подій у ньому не відповідатиме обмеженню сервісу (лістинг 4.4).

Лістинг 4.4 – Адаптивне розбиття часового вікна під час збору

```
private async UniTask FetchInto(List<EarthquakeRecord> acc,
                                DateTime start, DateTime
end, CancellationToken ct)
{
    try
    {
        var window = await _source.FetchWindowAsync(start,
end, _config.MinMagnitude, ct);
        acc.AddRange(window);
    }
    catch (EarthquakeRequestException ex) when
(ex.ResponseCode == 400 &&
                                (end -
start) > TimeSpan.FromDays(1))
    {
        var mid = start.AddTicks((end - start).Ticks / 2);
        // ділимо вікно навпіл
        await FetchInto(acc, start, mid, ct);
        await FetchInto(acc, mid, end, ct);
    }
}
```

Об'єднані дані з двох джерел проходять етап дедуплікації відповідно до критеріїв, описаних у підрозділі 2.2. Для кожної події одного джерела серед подій іншого шукають найближчу за часом, відстанню та магнітудою; відстань між епіцентрами обчислюється за формулою Гаверсина (2.1). Реалізацію пошуку відповідника наведено в лістингу 4.5; знайдені відповідники пов'язуються спільним ідентифікатором кластера.

Лістинг 4.5 – Пошук відповідника під час дедуплікації

```
private EarthquakeRecord FindMatch(List<EarthquakeRecord>
primary, EarthquakeRecord candidate)
{
    EarthquakeRecord best = null;
    double bestDistance = double.MaxValue;

    foreach (var p in InTimeWindow(primary, candidate,
_timeWindowMs))
    {
        if (Math.Abs(p.Magnitude - candidate.Magnitude) >
_magnitudeDelta) continue;
        double d = GeoDistance.HaversineKm(p.Latitude,
p.Longitude,
candidate.Latitude, candidate.Longitude);
        if (d <= _distanceKm && d < bestDistance) {
bestDistance = d; best = p; }
    }
    return best;
}
```

Марковську модель будують за часовим рядом станів комірки. Рівень активності кожного інтервалу класифікують на три стани за терцилями, після чого підраховують переходи між станами й нормують кожен рядок матриці згідно з формулою (3.1). Відповідний фрагмент реалізації наведено в лістингу 4.6.

Лістинг 4.6 – Побудова матриці переходів марковського ланцюга

```
int[,] counts = new int[n, n];
for (int i = 1; i < states.Length; i++)
    counts[(int)states[i - 1], (int)states[i]]++;           //
підрахунок переходів
```

Продовження лістингу 4.6

```
double[,] transition = new double[n, n];
for (int from = 0; from < n; from++)
{
    int rowSum = 0;
    for (int to = 0; to < n; to++) rowSum += counts[from,
to];

    for (int to = 0; to < n; to++)
        transition[from, to] = rowSum > 0 ?
(double)counts[from, to] / rowSum : 1.0 / n;
}
```

Оцінювання сейсмічної небезпеки комірки поєднує оцінку коефіцієнта b за формулою Акі з пуассонівським прогнозом. За середньою магнітудою подій, не слабших за магнітуду повноти, обчислюють коефіцієнт b , далі – річну інтенсивність цільової магнітуди, імовірність її перевищення протягом заданого горизонту та період повторюваності (лістинг 4.7).

Лістинг 4.7 – Оцінювання b -value та пуассонівського прогнозу

```
double meanMagnitude = sumMagnitude / eventCount;
double bValue = LOG10_E / (meanMagnitude -
completenessMagnitude + MAGNITUDE_BIN / 2); // (3.4)

double rateAtCompleteness = eventCount / catalogYears;
double lambda = rateAtCompleteness *
    Math.Pow(10.0, -bValue * (targetMagnitude
- completenessMagnitude)); // (3.6)

double probability = 1.0 - Math.Exp(-lambda *
horizonYears); // (3.7)

double returnPeriod = 1.0 / lambda;
// (3.8)
```

Приклад проміжних результатів обчислення параметрів моделей для найактивніших комірок, що виводяться під час роботи системи, наведено на рисунку 4.1.

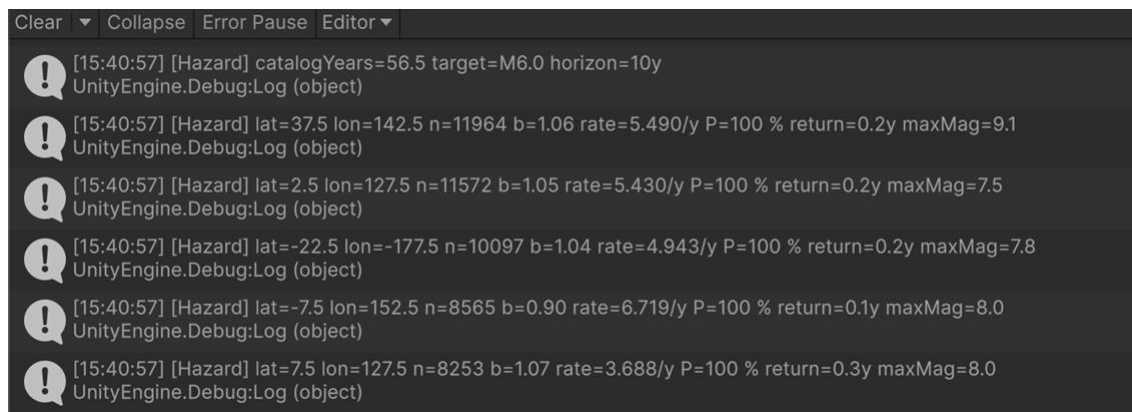


Рисунок 4.1 – Приклад виводу результатів обчислення у консолі середовища розробки

Отримані для кожної комірки параметри – стани й матриця переходів, коефіцієнт b , імовірність перевищення цільової магнітуди та період повторюваності – формують основу для візуалізації прогнозу, розглянутої в наступному підрозділі.

4.4 Візуалізація результатів

Візуальну частину системи побудовано на основі тривимірної моделі земної кулі – сфери з нанесеною картографічною текстурою, яку окрема камера відображає у текстуру, виведену на екран. Користувач обертає глобус перетягуванням, що дозволяє оглядати будь-який регіон. Інтерфейс містить дві вкладки, які перемикають режими відображення: перегляд зареєстрованих подій та аналіз сейсмічної небезпеки.

У режимі перегляду подій епіцентри землетрусів відображаються як набір частинок на поверхні глобуса. Кожну подію розміщують на сфері за її

географічними координатами; перетворення координат у тривимірну точку виконується згідно з формулою проєкції, наведеною в підрозділі 2.3, і реалізовано в лістингу 4.8. Колір і розмір частинки кодують магнітуду (від жовтого до червоного) або глибину події, що дозволяє візуально оцінити просторовий розподіл і силу землетрусів.

Лістинг 4.8 – Проєкція географічних координат на поверхню сфери

```
public static Vector3 LatLonToLocal(double latDeg, double
lonDeg, float radius,
                                float
meridianOffsetDeg, bool invertLongitude)
{
    double lon = lonDeg + meridianOffsetDeg;
    if (invertLongitude) lon = -lon;

    double latRad = latDeg * Math.PI / 180.0;
    double lonRad = lon * Math.PI / 180.0;
    double cosLat = Math.Cos(latRad);

    float x = (float)(radius * cosLat * Math.Sin(lonRad));
    float y = (float)(radius * Math.Sin(latRad));
    float z = (float)(radius * cosLat * Math.Cos(lonRad));
    return new Vector3(x, y, z);
}
```

Загальний вигляд глобуса з відображеними епіцентрами наведено на рисунку 4.4.

У режимі аналізу поверх глобуса відображається карта сейсмічної небезпеки. На відміну від окремих частинок, її подано у вигляді суцільного меша, що огинає сферу: кожна комірка сітки – чотирикутник, кути якого спроєктовано на поверхню. Колір комірки визначається логарифмом річної інтенсивності подій цільової магнітуди (від зеленого до червоного); застосування логарифмічної шкали замість імовірності зумовлене тим, що

для активних зон імовірність наближається до одиниці й втрачає розрізнявальну здатність. Фрагмент обчислення кольору комірки наведено в лістингу 4.9. Отриману карту небезпеки наведено на рисунку 4.5.

Лістинг 4.9 – Визначення кольору комірки за логарифмом інтенсивності

```
float t = Mathf.Clamp01((Mathf.Log10(rate) - lowLog) /
(highLog - lowLog));
Color color = t < 0.5f
    ? Color.Lerp(green, amber, t * 2f)
    : Color.Lerp(amber, red, (t - 0.5f) * 2f);
```

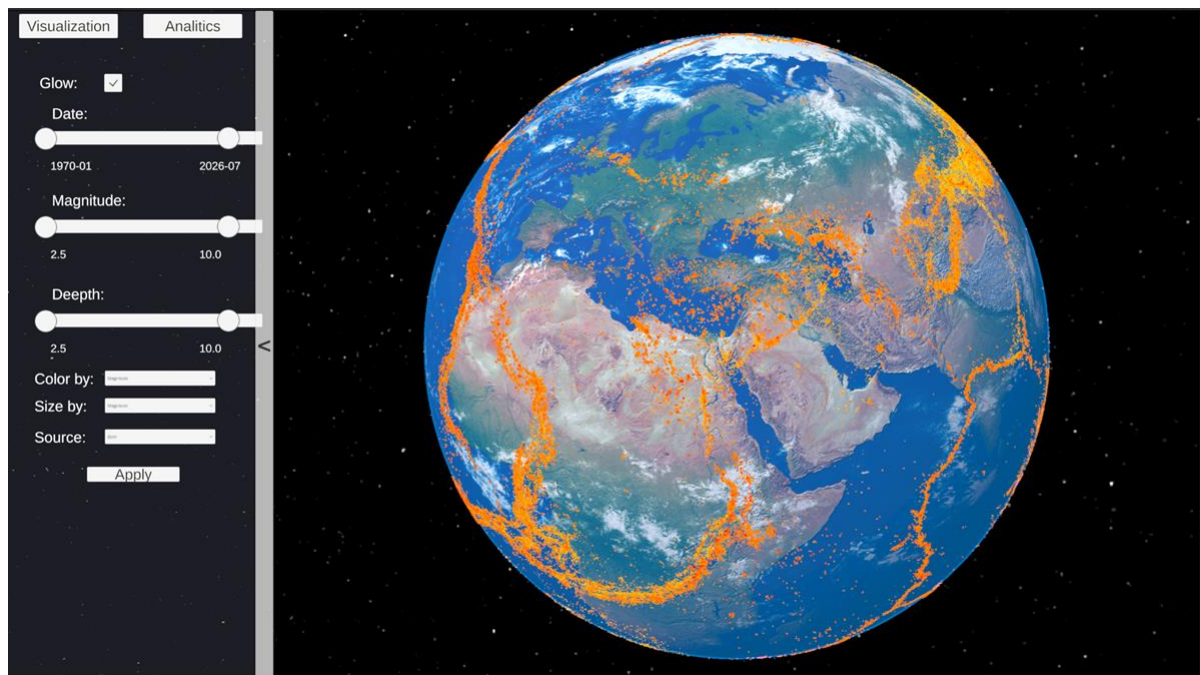


Рисунок 4.2 – Відображення епіцентрів землетрусів на тривимірній моделі Землі (режим перегляду подій)

Система підтримує інтерактивний аналіз окремих регіонів. Після вибору комірки натисканням на глобусі з наступним визначенням відповідної комірки сітки відкривається панель регіону, що містить оцінені для нього параметри: коефіцієнт b , імовірність перевищення цільової

магнітуди, період повторюваності, а також матрицю переходів марковського ланцюга й поточний стан активності. Приклад такої панелі наведено на рисунку 4.6.

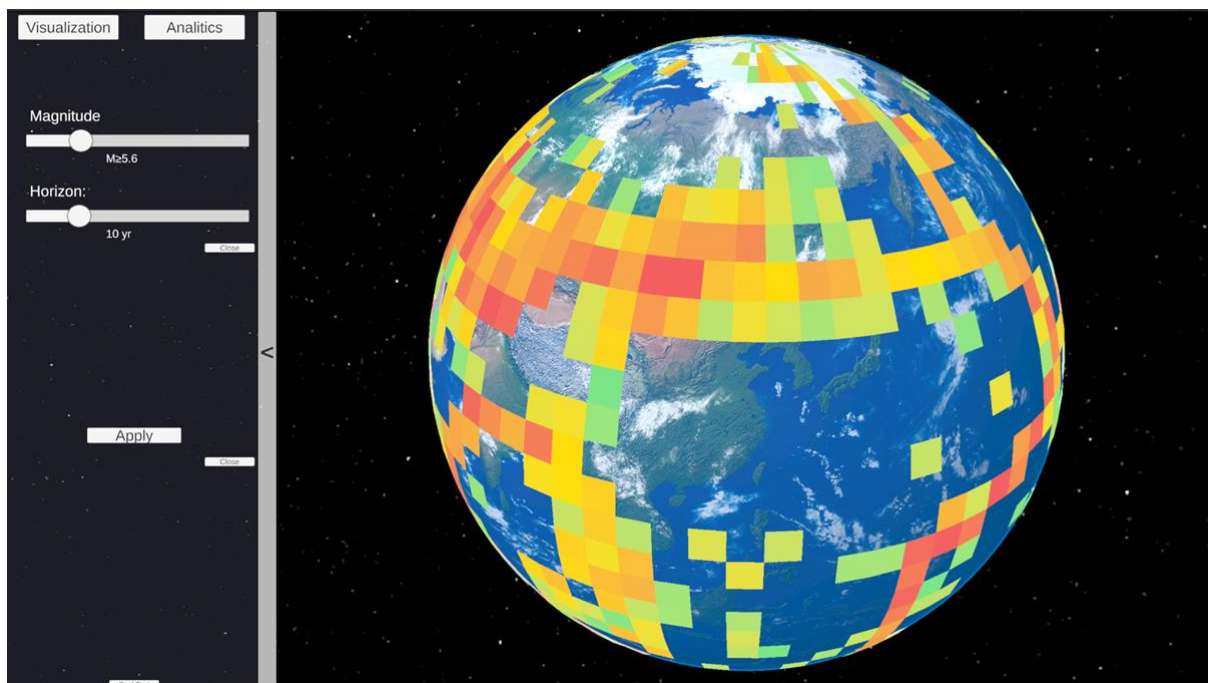


Рисунок 4.3 – Карта сейсмічної небезпеки (режим аналізу)

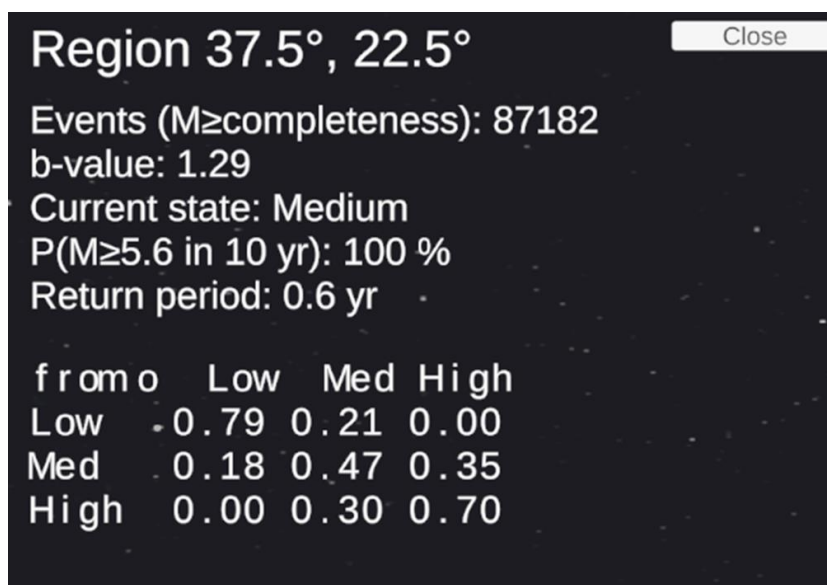


Рисунок 4.4 – Панель аналізу регіону з матрицею переходів марковського ланцюга

Поєднання карти небезпеки з деталізованою панеллю регіону дозволяє користувачеві переходити від загальної просторової картини до кількісних характеристик прогнозу для конкретної зони. Об'єктивну оцінку якості побудованого прогнозу наведено в наступному підрозділі.

4.5 Оцінювання якості прогнозу

Для об'єктивного оцінювання якості прогнозу застосовано ретроспективну перевірку за методикою, описаною в підрозділі 3.4. Каталог розділено в часі на навчальну та тестову частини: параметри моделі оцінювали лише за навчальними даними, після чого прогноз порівнювали з подіями тестового періоду, які не використовувалися під час навчання межа поділу, напр. навчання 1970–2015, тест 2015–2026.

Коректність оцінювання параметрів підтверджується отриманими значеннями коефіцієнта b . Для найактивніших сейсмічних зон його значення близькі до одиниці (таблиця 4.1), що відповідає класичному діапазону закону Гутенберга–Ріхтера й свідчить про адекватність оцінки. Максимальні зафіксовані в цих зонах магнітуди відповідають відомим сильним землетрусам.

Таблиця 4.1 – Оцінені параметри для найактивніших сейсмічних зон

Регіон	Координати (φ , λ)	Кількість подій	Коефіцієнт b	Макс. магнітуда
Японія	37,5° пн., 142,5° сх.	11 964	1,06	9,1
Індонезія (Молукка)	2,5° пн., 127,5° сх.	11 572	1,05	7,5

Продовження таблиці 4.1

Регіон	Координати (φ , λ)	Кількість подій	Коефіцієнт b	Макс. магнітуда
Тонга–Кермадек	22,5° пд., 177,5° зх.	10 097	1,04	7,8
Соломонові острови	7,5° пд., 152,5° сх.	8 565	0,90	8,0
Філіппіни	7,5° пн., 127,5° сх.	8 253	1,07	8,0

Кількісні результати ретроспективної перевірки такі. Оцінка Брайєра для прогнозу становила 0,120, базова частота настання подій – 29%, кількість перевірених комірок – 1084. Оцінка майстерності Брайєра склала 0,41, що є додатною величиною й означає перевагу побудованого прогнозу над тривіальним кліматологічним прогнозом приблизно на 41%. Це підтверджує, що модель має реальну прогностичну цінність, а не відтворює середню частоту подій.

Калібрування прогнозу ілюструє діаграма надійності (рисунок 4.5). Для низьких значень прогнозованої ймовірності точки розташовані поблизу діагоналі, що свідчить про добре калібрування. Водночас для середніх і високих значень спостерігається систематична переоцінка ймовірності: прогнозована ймовірність перевищує фактичну частоту.

Це пояснюється припущенням про незалежність подій, покладеним в основу пуассонівської моделі, – реальна сейсмічність характеризується кластеризацією (післяшовковими послідовностями), унаслідок чого зони з інтенсивною активністю в навчальному періоді отримують завищену оцінку безпеки.

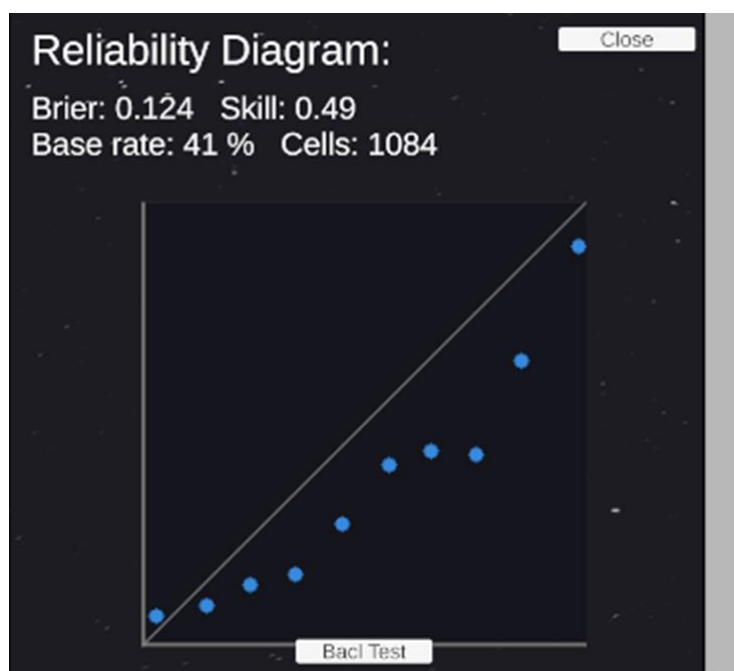


Рисунок 4.5 – Діаграма надійності прогнозу (reliability diagram)

Таким чином, ретроспективна перевірка підтвердила працездатність побудованого прогнозу: додатне значення оцінки майстерності свідчить про його перевагу над базовим прогнозом. Виявлену систематичну переоцінку в зонах підвищеної активності можна усунути застосуванням процедур декластеризації або моделей із урахуванням тригерингу подій, що є напрямом подальшого вдосконалення системи.

ВИСНОВКИ

Кваліфікаційна робота демонструє доцільність і практичну цінність застосування стохастичних моделей для аналізу та імовірнісного прогнозування сейсмічної активності, поєднуючи складний математичний апарат сейсмології з доступним інтерактивним інструментом. Розроблена система заповнює прогалину між теоретичними моделями оцінювання сейсмічної небезпеки та їх практичним застосуванням, перетворюючи громіздкі обчислення над мільйонами подій на наочні прогнозні карти й кількісні характеристики ризику для конкретного регіону, зрозумілі кінцевому користувачеві.

Роботу завершено низкою суттєвих технічних досягнень, що охоплюють увесь шлях від сирих даних до перевіреного прогнозу. Реалізовано автоматизований збір сейсмічних каталогів із двох незалежних відкритих джерел – USGS та EMSC – через уніфікований веб-сервіс FDSN з адаптивним розбиттям часових вікон, що дало змогу повністю отримати історичний каталог обсягом близько 2 млн подій. Різноманітні записи зведено до єдиної моделі, а дублікати тієї самої фізичної події, зареєстрованої різними мережами, усунено дедуплікацією за просторово-часовою близькістю з обчисленням відстані за формулою Гаверсина. Застосування вбудованої бази SQLite з агрегацією на боці бази забезпечило накопичення історії спостережень і високу продуктивність аналітики незалежно від доступності мережі.

Ядром системи є поєднання прозорих та інтерпретованих стохастичних моделей. Марковську модель із трьома станами активності, межі яких визначаються за терцилями розподілу для кожної комірки окремо, застосовано для опису динаміки сейсмічного режиму; переважання діагональних елементів матриці переходів підтвердило його інерційність, що узгоджується з відомою кластеризацією подій. Оцінювання сейсмічної небезпеки реалізовано на основі закону Гутенберга-Ріхтера з обчисленням

коефіцієнта b методом максимальної правдоподібності за формулою Акі та подальшим пуассонівським прогнозом імовірності перевищення цільової магнітуди й періоду повторюваності. Отримані для найактивніших зон значення коефіцієнта b , близькі до одиниці, відповідають класичному діапазону закону й підтверджують коректність оцінювання.

Важливим досягненням є формалізована перевірка якості прогнозу, що вирізняє роботу з-поміж суто описових підходів. Ретроспективна перевірка на незалежному тестовому періоді дала оцінку Брайєра 0,120 за базової частоти подій 29 % на 1084 комірках та оцінку майстерності Брайєра 0,41, що означає перевагу побудованого прогнозу над тривіальним кліматологічним приблизно на 41 % і свідчить про його реальну прогностичну цінність. Водночас діаграма надійності виявила систематичну переоцінку небезпеки в зонах підвищеної активності, зумовлену припущенням про незалежність подій у пуассонівській моделі, – цей результат чесно окреслює межі застосовності обраного апарату.

Практичну цінність системи посилює інтерактивна візуалізація на ігровому рушії Unity: тривимірна модель земної кулі з режимами перегляду епіцентрів і аналізу сейсмічної небезпеки, картою небезпеки та панеллю деталізації окремого регіону з його параметрами й матрицею переходів. Це дозволяє користувачеві переходити від загальної просторової картини до кількісних характеристик прогнозу для конкретної зони, роблячи результати математичного моделювання доступними для інтерпретації.

Загалом розроблена система підтверджує працездатність запропонованого підходу й поєднує збір і обробку сейсмічних даних, стохастичне моделювання, імовірнісне прогнозування та візуалізацію в єдиному програмному рішенні, якість якого підтверджено кількісною перевіркою на незалежних даних. Перспективним напрямом подальшого вдосконалення є усунення виявленої переоцінки небезпеки в активних зонах шляхом декластеризації каталогу або застосування моделей з урахуванням тригерингу подій, зокрема ETAS.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Beroza G. C., Segou M., Mostafa Mousavi S. Machine learning and earthquake forecasting—next steps. *Nature communications*. 2021. Vol. 12, no. 1. URL: <https://doi.org/10.1038/s41467-021-24952-6> (date of access: 12.07.2026).
2. GP-ETAS: semiparametric Bayesian inference for the spatio-temporal epidemic type aftershock sequence model / C. Molkenhuth et al. *Statistics and computing*. 2022. Vol. 32, no. 2. URL: <https://doi.org/10.1007/s11222-022-10085-3> (date of access: 12.07.2026).
3. Petrillo G., Zhuang J. Bayesian earthquake forecasting approach based on the epidemic type aftershock sequence model. *Earth, planets and space*. 2024. Vol. 76, no. 1. URL: <https://doi.org/10.1186/s40623-024-02021-8> (date of access: 12.07.2026).
4. Stockman S., Lawson D. J., Werner M. J. SB-ETAS: using simulation based inference for scalable, likelihood-free inference for the ETAS model of earthquake occurrences. *Statistics and computing*. 2024. Vol. 34, no. 5. URL: <https://doi.org/10.1007/s11222-024-10486-6> (date of access: 12.07.2026).
5. Towards a deep learning approach for short-term data-driven spatiotemporal seismicity rate forecasting / F. Dervisi et al. *Earth, planets and space*. 2025. Vol. 77, no. 1. URL: <https://doi.org/10.1186/s40623-025-02241-6> (date of access: 12.07.2026).
6. Long memory effects and forecasting of earthquake and volcano seismic data / M. C. Mariani et al. *Physica A: statistical mechanics and its applications*. 2020. Vol. 559. P. 125049. URL: <https://doi.org/10.1016/j.physa.2020.125049> (date of access: 12.07.2026).
7. Earthquake magnitude prediction in Turkey: a comparative study of deep learning methods, ARIMA and singular spectrum analysis / H. Öncel Çekim et al. *Environmental earth sciences*. 2023. Vol. 82, no. 16. URL: <https://doi.org/10.1007/s12665-023-11072-1> (date of access: 12.07.2026).

8. WEarthquake prediction based on spatio-temporal data mining: an LSTM network approach / Q. Wang et al. *IEEE transactions on emerging topics in computing*. 2020. Vol. 8, no. 1. P. 148–158. URL: <https://doi.org/10.1109/tetc.2017.2699169> (date of access: 12.07.2026).

9. Nugroho H. A., Subiantoro A., Kusumoputro B. A model fusion of 1D CNN with NARX for accurate earthquake time series prediction. *2024 4th international conference on robotics, automation and artificial intelligence (RAAI)*, Singapore, Singapore, 19–21 December 2024. 2024. P. 340–344. URL: <https://doi.org/10.1109/raai64504.2024.10949569> (date of access: 12.07.2026).

10. Interactive visualization of complex seismic data and models using bokeh / C. Chai et al. *Seismological research letters*. 2018. Vol. 89, no. 2A. P. 668–676. URL: <https://doi.org/10.1785/0220170132> (date of access: 12.07.2026).

11. Portable interactive visualization of large-scale simulations in geotechnical engineering using Unity3D / R.-X. Wang et al. *Advances in engineering software*. 2020. Vol. 148. P. 102838. URL: <https://doi.org/10.1016/j.advengsoft.2020.102838> (date of access: 12.07.2026).

12. Mint: integrating scientific visualizations into virtual reality / S. Geringer et al. *Journal of visualization*. 2024. URL: <https://doi.org/10.1007/s12650-024-01011-y> (date of access: 12.07.2026).

13. Gutenberg B., Richter C. F. Frequency of earthquakes in California*. *Bulletin of the seismological society of America*. 1944. Vol. 34, no. 4. P. 185–188. URL: <https://doi.org/10.1785/bssa0340040185> (date of access: 12.07.2026).

14. Wiemer S. Minimum magnitude of completeness in earthquake catalogs: examples from Alaska, the Western United States, and Japan. *Bulletin of the seismological society of America*. 2000. Vol. 90, no. 4.

P. 859–869. URL: <https://doi.org/10.1785/0119990114> (date of access: 12.07.2026).

15. Cornell C. A. Engineering seismic risk analysis. *Bulletin of the seismological society of America*. 1968. Vol. 58, no. 5. P. 1583–1606. URL: <https://doi.org/10.1785/bssa0580051583> (date of access: 12.07.2026).

16. Cavers M., Vasudevan K. Spatio-temporal complex Markov chain (SCMC) model using directed graphs: earthquake sequencing. *Pure and applied geophysics*. 2014. Vol. 172, no. 2. P. 225–241. URL: <https://doi.org/10.1007/s00024-014-0850-7> (date of access: 12.07.2026).

17. The collaboratory for the study of earthquake predictability: achievements and priorities / D. Schorlemmer et al. *Seismological research letters*. 2018. Vol. 89, no. 4. P. 1305–1313. URL: <https://doi.org/10.1785/0220180053> (date of access: 12.07.2026).

18. API Documentation - Earthquake Catalog. *USGS Earthquake Hazards Program*. URL: <https://earthquake.usgs.gov/fdsnws/event/1/> (date of access: 12.07.2026).

19. European-Mediterranean Seismological Centre. *EMSC*. URL: <https://www.emsc-csem.org/> (date of access: 12.07.2026).

20. FDSN Web Service Specification: fdsnws-event. Version 1.2. *International Federation of Digital Seismograph Networks*. 2019. URL: <https://www.fdsn.org/webservices/fdsnws-event-1.2.pdf> (date of access: 12.07.2026).