

DOI: <https://doi.org/10.36910/6775-2524-0560-2020-41-30>

УДК 004.89

Проніна Ольга Ігорівна, к.т.н.

<https://orcid.org/0000-0001-7085-8027>

Якін Костянтин Юрійович, магістр

<https://orcid.org/0000-0003-3455-3204>

Державний вищий навчальний заклад «Приазовський державний технічний університет», м. Маріуполь, Україна

МОДЕЛЮВАННЯ КЛАСТЕРИЗАЦІЇ ЗЕМЛЕТРУСІВ ЗА ДОПОМОГОЮ МЕТОДІВ МАШИННОГО НАВЧАННЯ

Проніна О. І., Якін К. Ю. Моделювання кластеризації землетрусів за допомогою методів машинного навчання. Сейсмічний аналіз полягає у вивченні поширення пружних хвиль в товщі порід. Ґрунтуючись на принципах цього аналізу, доцільна побудова моделі, яка б описувала основні метрики поведінки земних ґрунтів під час землетрусів і яка могла б використовуватися для задач прогнозування сейсмічної активності в подальших дослідженнях. Робота присвячена моделюванню кластеризації землетрусів та візуалізації розподілу мейншокових процесів з використанням сучасних методів машинного навчання, таких як стохастичний t-розподіл та кластеризація. Розглянуто основні етапи побудови моделі для більш точного результату. Проведено моделювання на вибірці даних за останні сто років для перевірки адекватності розробленої моделі.

Ключові слова: машинне навчання, сейсмічний аналіз, мейншокові процеси, кластеризація, стохастичний t-розподіл.

Пронина О. И., Якин К. Ю. Моделирование кластеризации землетрясений с помощью методов машинного обучения. Сейсмический анализ заключается в изучении распространения упругих волн в толще породы. Основываясь на принципах этого анализа, целесообразно построение модели, которая описывала основные метрики поведения земных почв во время землетрясений, и которая могла бы использоваться для задач прогнозирования сейсмической активности в дальнейших исследованиях. Работа посвящена моделированию кластеризации землетрясений и визуализации распределения мейншоковых процессов с использованием современных методов машинного обучения, таких как стохастический t-распределение и кластеризация. Рассмотрены основные этапы построения модели для более точного результата. Проведено моделирование на выборке данных за последние сто лет для проверки адекватности разработанной модели.

Ключевые слова: машинное обучение, сейсмический анализ, мейншоковые процессы, кластеризация, стохастическое t-распределение.

Pronina O. I., Yakin K. Yu. Modeling earthquake clustering using machine learning methods. Seismic analysis is to study the propagation of elastic waves in the bedrock. Based on the principles of this analysis, it is advisable to build a model that would describe the basic metrics of the behavior of terrestrial soils during earthquakes and that could be used for tasks of predicting seismic activity in further research. The work is devoted to modeling clustering of earthquakes and visualization of the distribution of main shock processes using modern machine learning methods, such as stochastic t-distribution and clustering. The main stages of model construction for a more accurate result are considered. Modeling on data samples for the last hundred years is carried out to check the adequacy of the developed model.

Keywords: machine learning, seismic analysis, mainshock processes, clustering, stochastic t-distribution.

Постановка наукової проблеми. Землетруси є однією з найбільш руйнівних сил природи, що приносять людству величезні матеріальні збитки і велике число жертв. Значна частина поверхні Землі, включаючи населені пункти з розвиненою інфраструктурою, знаходиться в сейсмоактивних зонах. Мейншоковий процес – процес, який описує характер протікання головного товчка земної поверхні під час сейсмічної активності недр Землі. Для забезпечення безпеки життєдіяльності доцільно проводити сейсмічний аналіз, який полягає у вивченні поширення пружних хвиль в товщі порід. Подібний науковий підхід має низку незаперечних переваг і знаходить своє відображення і застосування для вирішення багатьох реальних задач. Однак, при вивченні природи і механізму більшості динамічних явищ необхідно вирішувати обернені задачі – за кінцевим результатом ефекту відновлювати вихідну модель.

Метою даної роботи є побудова моделі сейсмічних процесів на основі аналізу взаємозв'язків, залежностей і закономірностей різних сейсмічних показників. При цьому також потрібно забезпечити високу точність побудови моделі, для досягнення якої використовувалися новітні програмно-апаратні засоби, реалізовані за допомогою мови програмування загального призначення Python.

Моделювання здійснювалося на основі даних, що містять кількісні показники основних ознак протікання сейсмічних процесів за останні 100 років на території Європи, отримані шляхом експериментальних розрахунків і вимірювань.

Аналіз досліджень. Прогноз сейсмокатастроф включає систему заходів щодо визначення ймовірності виникнення великих землетрусів, їх масштабів і можливих наслідків у зоні їх впливів для конкретного регіону. З вищезазначеного випливає, що прогноз землетрусу базується на вирішенні трьох основних завдань:

- передбачення ймовірного місця або району, де очікується землетрус;
- прогноз імовірною магнітуди (або сили) очікуваної події;
- визначення ймовірного часу очікуваної сейсмокатастрофи.

У даній статті [1] автори детально вивчають процес глибинного ковзання літосферних плит, що викликається земними тріщинами, на основі методів 3D-моделювання, а точніше методу Applied Element Method (АЕМ). Цей метод використовується для вивчення просторового зміни руху ґрунту і основною його перевагою є можливість ініціювання тріщини, яка виникла на основі ушкоджень корінних порід, і поширення тріщини до колапсу. Однак, для отримання більш точних результатів розподілу сейсмічних процесів такого підходу буде замало, тому в роботі [2] представлена модель, заснована на інтеграції чотирьох різних обчислювальних концепцій: нейронна мережа з динамічною затримкою, вейвлет, нечітка логіка і концепція простору реконструйованих станів з теорії хаосу. Модель балансує глобальний і локальний вплив даних навчання і ефективно враховує неточність, існуючу в даних датчика. Експериментальні результати п'ятиповерхового сталевого каркаса використовуються для перевірки обчислювальної моделі і демонстрації її точності і ефективності.

Крім того, стаття [3] пропонує модель прогнозування землетрусів за астрономічними даними на основі системи штучного інтелекту. Моделювання подій здійснювалося за параметром подібності, який є аналогом коефіцієнта кореляції в статистиці. Виявлена залежність параметра подібності від магнітуди, глибини вогнища (гіпофокуса) і числа землетрусів, що відбуваються щодня на нашій планеті, як в місячному, так і в 2-3 денному прогнозі. Обговорюється гіпотеза, що цей ефект обумовлений гравітацією небесних тіл. Для перевірки гіпотези було досліджено вплив дванадцяти гравітаційних комплексів на всю сукупність подій і виділені найбільш інформативні астрономічні ознаки землетрусів. Але в даному випадку мається проблема збору таких ознак, тому що досить велика кількість сильних землетрусів у світі протікає з тривалістю в кілька хвилин або навіть секунд, що значно ускладнює спостереження і фіксацію необхідних вимірювань. Тому для таких ситуацій автори статті [4] проводять моніторинг статистичних властивостей мережі, який може бути корисним для короткострокової оцінки небезпеки виникнення головних ударів при наявності форшоків. Мережа вимірює як глобальний (середній коефіцієнт кластеризації, індекс малого світу), так і локальний (міжцентровий) мейншоки, які можуть бути використані для прогнозування як у часі, так і в просторі. Результати показують статистично значуще збільшення топологічних заходів навколо місця розташування епіцентру приблизно за 2 місяці до головного удару.

В роботі [5] автори на основі каталогу землетрусів Курило-Камчатського регіону за 1737-2007 рр. для Південних Курильських островів створена тривимірна модель зон можливих вогнищ землетрусів для оцінки сейсмічної небезпеки. Сейсмічними методами уточнені границі структурних елементів зон можливих вогнищ землетрусів.

Виклад основного матеріалу й обґрунтування отриманих результатів дослідження. Магнітуда землетрусу – величина, що характеризує енергію, що виділилася при землетрусі в вигляді сейсмічних хвиль [6]. Первісна шкала магнітуди була запропонована американським сейсмологом Чарльзом Ріхтером в 1935 році. Ріхтер запропонував для оцінки сили землетрусу (в його епіцентрі) десятковий логарифм переміщення A (в мікрометрах) голки стандартного сейсмографа Вуда - Андерсона, розташованого на відстані не більше 600 км від епіцентру:

$$M_L = \log A + f \quad (1)$$

де M_L – магнітуда Ріхтера;

f – коригувальна функція, яка обчислюється по таблиці в залежності від відстані до епіцентру [7];

A – амплітуда коливань землі (в мікрометрах).

Однак, запропонований спосіб вимірювання враховував тільки поверхневі хвилі, в той час як при глибинних землетрусах істотна частина енергії виділяється у формі об'ємних хвиль. Протягом наступних кількох десятків років шкала Ріхтера уточнювалася і наводилася у відповідність з новими спостереженнями. Зараз існує кілька похідних шкал, найважливішими з яких є магнітуда об'ємних хвиль (m_b) і магнітуда поверхневих хвиль (M_s):

$$m_b = \log \left(\frac{A}{T} \right) + Q(D, h), \quad (2)$$

$$M_s = \log \left(\frac{A}{T} \right) + 1,66 \log D + 3,3, \quad (3)$$

де m_b – магнітуда об'ємних хвиль;

M_s – магнітуда поверхневих хвиль;

A – амплітуда коливань землі (в мікрометрах);

T – період хвилі (в секундах);

Q – поправна функція;

D – відстань до епіцентру землетрусу;

h – глибина вогнища землетрусу [8].

Виявилось, що ці шкали дають некоректні результати для найбільших землетрусів - при $M \sim 8$ настає насичення. Це пов'язано з тим, що тривалість багатьох дуже сильних землетрусів становить понад 20 с - періоду поверхневих хвиль, використовуваних в M_s . Тому в 1979 році сейсмолог Хіроо Канаморі визнав цей недолік і зробив простий, але важливий крок визначення величини, заснованої на оцінках випромінюваної енергії M_w , де «w» означає роботу (енергію). Дана величина в розрахунках ґрунтується на емпіричній залежності Гутенберга-Ріхтера [9-10]:

$$M_w = \frac{2}{3} \log E_s - 3,2, \quad (4)$$

$$\log E_s \approx 1,5M_s + 4,8, \quad (5)$$

де M_w – енергія землетрусу;

M_s – магнітуда поверхневих хвиль;

E_s – енергія поверхневих хвиль.

Крім усього іншого, в 1991 р. Володимир Тобіас і Рейнхард Миттаг запропонували пов'язати величину поверхневої хвилі з масштабом локальної величини M_L [7], використовуючи:

$$M_s = 1,45M_L - 3,2. \quad (6)$$

де M_s – магнітуда поверхневих хвиль;

M_L – магнітуда Ріхтера.

Таким чином, шляхом приведення перерахованих формул отримуємо наступну залежність:

$$M_w = \frac{2}{3} \log E_s - 3,2 = \frac{2}{3} (1,5(1,45M_L - 3,2) + 4,8) - 3,2 = 1,45M_L - 3,2. \quad (7)$$

де M_w – енергія землетрусу;

M_L – магнітуда Ріхтера;

E_s – енергія поверхневих хвиль.

Отже, можна зробити висновок про те, що залежність енергії, випромінюваної землетрусом, від локальної магнітуди відповідає залежності між магнітудою поверхневих хвиль і цієї ж локальної магнітудою. Однак, незважаючи на ідентичність формул і скалярну рівність, обидві величини описують все ж різну природу явищ.

Моделювання. Процес побудови моделі складається з декількох основних етапів. На етапі підготовки робочої вибірки потрібно провести первинну оцінку наявних даних з подальшим відбором якісно значущих об'єктів. Однак, для оцінки залежності цільової функції від вихідних факторів необхідно не тільки спиратися на математичну модель, але і враховувати інші закономірності числових параметрів. Для цього слід побудувати матрицю чутливості, яка відображає ступінь кореляції між усіма ознаками.

Як можна помітити, на енергію землетрусу, крім показника за шкалою Ріхтера, також в значній мірі впливає і тривалість головного шокowego процесу, хоч це і не було передбачено в матмоделі. Цей факт однозначно має своє відображення і в наукових колах: в історії було зафіксовано кілька випадків слабких по інтенсивності землетрусів, тривалість яких досягала декількох тижнів, внаслідок чого енергія, що випромінює такими ковзаннями літосферних плит, досягала рівня 6 балів. Карта чутливості представлена на рис.1, основні критерії вимірювання md – тривалість землетрусу; mw – енергія землетрусу; $richter$ – магнітуда Ріхтера.



Рис.1 – Карта чутливості взаємодії показників землетрусів

На матриці відображається сильна взаємодія параметрів оцінки сили землетрусу за шкалою Ріхтера та тривалістю сейсмічних процесів. На карті чутливості більш сильна залежність відображається більш світлим кольором.

Для наочності було побудовано графік розподілу ключових величин, після чого вибірка була розбита на 3 групи в залежності від інтенсивності землетрусу: незначні, слабкі та помірні. Даний розподіл показує, що енергія землетрусу прямо пропорційна виділеним ознакам: чим більше тривалість та інтенсивність - тим більше енергія, що випромінюється. Графік розподілу зображений на рис.2.

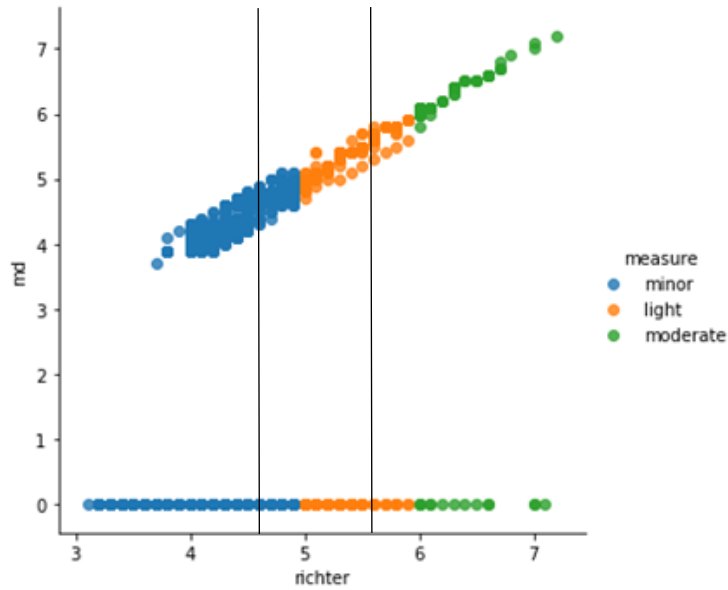


Рис.2 – Розподіл землетрусів за групами потужності

Таким чином, графік дає чітко зрозуміти, що тенденція між зазначеними параметрами дійсно існує в загальному випадку, а також показує, що за силою впливу землетруси можна розділити на 3 умовних кластера (розподіл здійснюється вертикальними лініями). Перша група – землетруси незначної сили до балів 1-5 за шкалою Ріхтера, друга група – це землетруси легкої сили до балів 5-6 за шкалою Ріхтера, та остання група – це землетруси помірної сили до балів 6-8 за шкалою Ріхтера. Тому, для більш глибокого аналізу розподілу в середині груп слід скористатися алгоритмом кластеризації. Але перед цим необхідно провести процес нормалізації даних, який дозволить усунути різницю в розмірності вхідного вектора параметрів, тим самим усуваючи числову залежність. При такому підході відкривається можливість відстежити інші приховані закономірності даної вибірки, які не могли бути розкриті через математичну модель. Розподіл вибірки за кластерами представлено на рис.3.

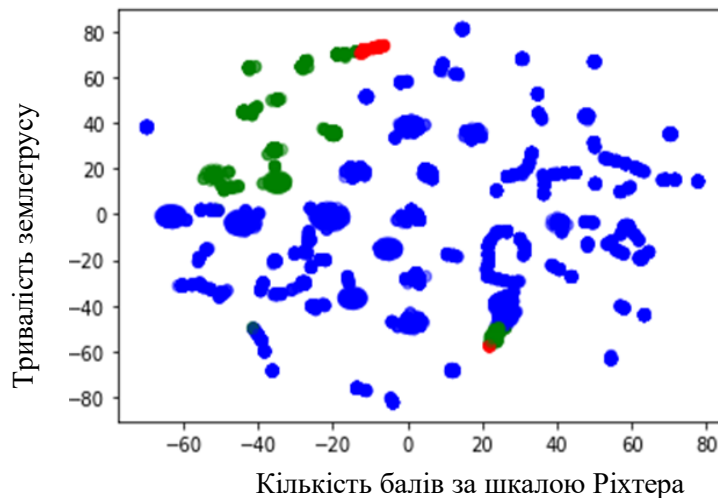


Рис.3 – Кластеризація землетрусів за рівнем сили

Як і очікувалося, процес кластеризації дав дещо інші результати, тому як розподіл вже не є настільки точним, наскільки це описано теоретично. Хоча в цілому загальна тенденція все так же помітна, розташування деяких кластерів на графіку не відповідає очікуваному результату. Причиною

цього може послужити наявність у вибірці математично не обґрунтованих наборів даних, які так само не закріплені у відповідних наукових базах знань. Отже, для підтвердження даної гіпотези слід провести додатковий процес очищення даних і повторну реалізацію алгоритму кластеризації. Результати повторного аналізу можна побачити на рис.4.

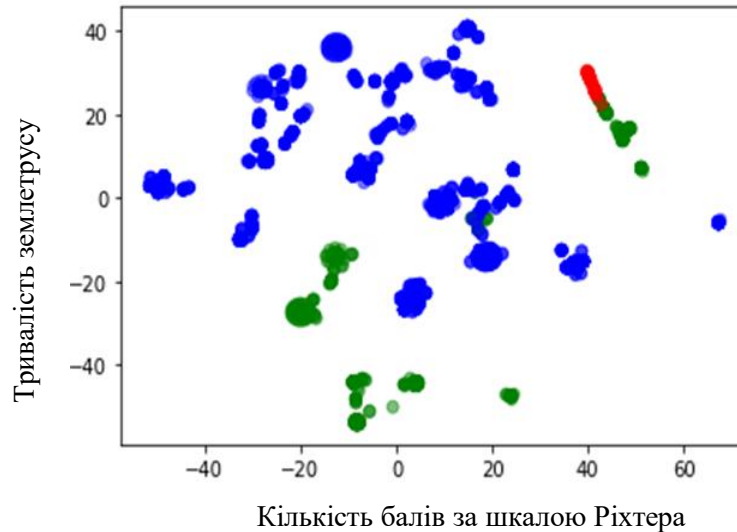


Рис.4 – Кластеризація землетрусів за рівнем сили після очищення вибірки

Таким чином, відсіяні раніше дані можна називати викидами, так як при їх відсутності результати кластеризації зазнали незначних змін. Тому аналіз цієї ситуації вимагає більш глибокої оцінки експертів в області геології і геофізики, що виходить за рамки даної роботи, але може послужити основою для подальших наукових досліджень.

Висновки та перспективи подальшого дослідження. В рамках даної роботи розроблена модель для аналізу сейсмічної активності землетрусів, що дозволила оцінити залежність їх виникнення на основі властивостей, які були виведені шляхом аналізу математичної моделі землетрусів. Результати, отримані в ході роботи, в цілому не суперечать теоретичним відомостям і мають практичну цінність не тільки для прикладних задач, але також і для подальших наукових досліджень. Розроблена модель пояснює основні аспекти та принципи протікання великої безлічі видів землетрусів в залежності від сили і характеру впливу сейсмічних хвиль. За рахунок моделі досягається високий рівень оптимізації процесу аналізу і класифікації виникаючих мейншокових процесів. Таким чином, модель може ефективно використовуватися, в тому числі, і в завданнях прогнозування сейсмічної активності, яке дозволяє підвищити забезпечення безпеки життєдіяльності населення і має економічну вигоду. Крім того, унаслідок неповноти вихідних даних була виділена група об'єктів, значення цільової функції яких принципово відрізняється навіть з урахуванням подібності набору властивостей даних об'єктів з об'єктами сусідніх кластерів. Подібний результат може вимагати більш глибокого і детального вивчення в рамках відповідних проектів.

Список бібліографічного опису

1. Hussain, Mohammad & Ramancharla, Pradeep. (2016). Modelling of Reverse Dip-Slip Faults Using 3D Applied Element Method. *Frontiers in Geotechnical Engineering*. 4. 1. 10.14355/fge.2016.04.001. Gronroos C. Service management and marketing / C. Gronroos. – West Sussex, 2000. – p. 496.
2. Adeli, Hojjat & Asce, F & Jiang, Xiaomo. (2006). Dynamic Fuzzy Wavelet Neural Network Model for Structural System Identification. *Journal of Structural Engineering-asce - J STRUCT ENG-ASCE*. 132. 10.1061/(ASCE)0733-9445(2006)132:1(102).
3. Трунев А. П., Луценко Е. В. Прогнозирование землетрясений по астрономическим данным с использованием системы искусственного интеллекта // Научный журнал КубГАУ - Scientific Journal of KubSAU. 2009. №52. URL: <https://cyberleninka.ru/article/n/prognozirovanie-zemletryaseniy-po-astronomicheskim-dannym-s-ispolzovaniem-sistemy-iskusstvennogo-intellekta> (дата обращения: 12.06.2020).
4. Daskalaki, Eleni & Spiliotis, Konstantinos & Siettos, Constantinos & Minadakis, George & Papadopoulos, Gerassimos. (2016). Foreshocks and Short-Term Hazard Assessment to Large Earthquakes using Complex Networks: the Case of the 2009 L'Aquila

Earthquake. Nonlinear Processes in Geophysics Discussions. 1-20. 10.5194/npg-2015-80.

5. Андреева М.Ю., Патрикеев В.Н. Моделирование зон очагов потенциальных землетрясений для территорий близ Бениоффа зоны // Геодинамика и тектонофизика. 2012. Т. 3. № 1. С. 69–76. DOI: 10.5800 / GT-2012-3-1-0063.

6. Schäfer, Andreas & Wenzel, Friedemann. (2019). Global Megathrust Earthquake Hazard—Maximum Magnitude Assessment Using Multi-Variate Machine Learning Table_1.csv. *Frontiers in Earth Science*. 7. 10.3389/feart.2019.00136.

7. Rouet-Leduc, Bertrand & Hulbert, Claudia & Lubbers, Nicholas & Barros, Kipton & Humphreys, Colin & Johnson, Paul. (2017). Machine Learning Predicts Laboratory Earthquakes. *Geophysical Research Letters*. 44. 10.1002/2017gl074677.

8. Mikumo, Takeshi & Miyatake, Takashi. (2009). Numerical modelling of space and time variations of seismic activity before major earthquakes. *Geophysical Journal of the Royal Astronomical Society*. 74. 559 - 583.

9. Matcharashvili, Teimuraz & Hatano, Takahiro & Chelidze, T. & Zhukova, Natalia. (2018). Simple statistics for complex Earthquake time distributions. *Nonlinear Processes in Geophysics*. 25. 497-510. 10.5194/npg-25-497-2018.

10. Abdelkrim A, Ghorbel C, Benrejeb M, et al. Lmi-based tracking control for takagi-sugeno fuzzy model [J]. *International Journal of Control & Automation*, 2010, № 3(2). – P. 21 – 36.

References

1. Hussain, Mohammad & Ramancharla, Pradeep. (2016). Modeling of Reverse Dip-Slip Faults Using 3D Applied Element Method. *Frontiers in Geotechnical Engineering*. 4.1.10.14355 / fge.2016.04.001. Gronroos C. Service management and marketing / C. Gronroos. - West Sussex, 2000. p. 496.

2. Adeli, Hojjat & Asce, F & Jiang, Xiaomo. (2006). Dynamic Fuzzy Wavelet Neural Network Model for Structural System Identification. *Journal of Structural Engineering-asce*. J STRUCT ENG-ASCE. 132.10.1061. (ASCE) 0733-9445 (2006) 132:1(102).

3. Trunev AP, Lutsenko EV Prediction of earthquakes based on astronomical data using an artificial intelligence system. *Scientific Journal of KubSAU*. 2009. No. 52. URL: <https://cyberleninka.ru/article/n/prognozirovanie-zemletryaseniy-po-astronomicheskim-dannym-s-ispolzovaniem-sistemy-iskusstvennogo-intellekta> (date accessed: 1.10.2020).

4. Daskalaki, Eleni & Spiliotis, Konstantinos & Siettos, Constantinos & Minadakis, George & Papadopoulos, Gerassimos. (2016). Foreshocks and Short-Term Hazard Assessment to Large Earthquakes using Complex Networks: the Case of the 2009 L'Aquila Earthquake. *Nonlinear Processes in Geophysics Discussions*. 1-20. 10.5194 / npg-2015-80.

5. Андреева М.Ю., Патрикеев В.Н. Modeling zones of sources of potential earthquakes for territories near the Benioff zone. *Geodynamics and tectonophysics*. 2012. Т. 3. No. 1. P. 69–76. DOI: 10.5800 / GT-2012-3-1-0063.

6. Schäfer, Andreas & Wenzel, Friedemann. (2019). Global Megathrust Earthquake Hazard – Maximum Magnitude Assessment Using Multi-Variate Machine Learning Table_1.csv. *Frontiers in Earth Science*. 7.10.3389 / feart.2019.00136.

7. Rouet-Leduc, Bertrand & Hulbert, Claudia & Lubbers, Nicholas & Barros, Kipton & Humphreys, Colin & Johnson, Paul. (2017). Machine Learning Predicts Laboratory Earthquakes. *Geophysical Research Letters*. 44.10.1002 / 2017gl074677.

8. Mikumo, Takeshi & Miyatake, Takashi. (2009). Numerical modeling of space and time variations of seismic activity before major earthquakes. *Geophysical Journal of the Royal Astronomical Society*. 74.559 - 583.

9. Matcharashvili, Teimuraz & Hatano, Takahiro & Chelidze, T. & Zhukova, Natalia. (2018). Simple statistics for complex Earthquake time distributions. *Nonlinear Processes in Geophysics*. 25.497-510. 10.5194 / npg-25-497-2018.

10. Abdelkrim A, Ghorbel C, Benrejeb M, et al. Lmi-based tracking control for takagi-sugeno fuzzy model [J]. *International Journal of Control & Automation*, 2010, No. 3 (2). P. 21 - 36.