The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 3, No 12, December 2025
PP 1-14

Seismic Patterns Beyond Magnitude: Depth–Magnitude Inversion, Spatial Clusters, and Tsunami Associations

Nissar Ahmed Pinjari[1], Samruddhi Rasal[2]
[1]Researcher, Seattle, WA, United States
[2]MIT School of Computing, MIT ADT University, Pune, India

Corresponding Author *: Samruddhi Rasal ,samruddhi.rasal03@gmail.com

Abstract

The early warning systems of earthquakes and tsunamis are usually based on quick magnitude measurements, but the depth of the event and the area covered by monitoring can have significant changes to risk interpretation. The paper integrates an already analyzed seismic report based on 782 records and 13 features such as magnitude, depth, and tsunami occurrence and geographic coordinates to describe seismic behavior and tsunami-related variability using spatio-temporal and statistical data. Distributional analysis indicates that the earthquakes magnitudes follow a central tendency of near 6.8, with significant high-end outliers, reflecting the fact that there are frequent moderate earthquakes and infrequent but with high-consequence earthquakes. Correlational results highlight depth as an essential explanatory variable of magnitude behavior in this data set and that the depth magnitude relationship does not follow typical independence expectations and that there are indications of a depth threshold about which magnitude behavior varies more abruptly. Spatial patterning also shows that seismic activity is not evenly distributed but in clusters, giving credence to the fact that monitoring and hazard modelling should be region specific. The yearly tsunami analysis is associated with the increased earthquake activity, and 2011 was a bright year of tsunami related seismicity, which means that tsunami-related seismicity can be related to cascading or high seismic events. The report also points out that there is an operational asymmetrical monitoring infrastructure whereby the signal measures are relatively consistent with widely distributed counts of stations, which suggests that there is an uneven observational capacity that could affect the reliability of detection. Taking together, these results encourage depth-aware risk modeling, hotspot-based monitoring, and combined seismic-tsunami surveillance plans which encompass the incorporation of the depth, location clustering, and temporal anomaly cues instead of the magnitude itself.

Keywords:

Seismic Risk Assessment, Depth–Magnitude Relationship, Spatio-Temporal Analysis, DBSCAN Clustering, Seismic Hotspots

1. Introduction

Earthquakes are one of the most consistent natural risks since they happen with little warning and potentially cause secondary risks that have cascading effects, most notably the tsunamis in coastal areas and in subduction-zone locations. Quick magnitude estimation is commonly adopted by rapid hazard communication in practice, but in operational practice, magnitude is an imperfect characterization of the potential impact when the depth, location, and coverage by monitoring change significantly. This is of particular concern to tsunami risk, wherein shallow offshore earthquakes may result in disproportional effects when compared to more profound earthquakes of the same magnitude. Consequently, data-driven analyses which collectively examine event features and observational setting are necessary to enhance seismic risk deciphering and solidify early apprehensions and countermeasures.

The present paper summarizes the results of a previously analyzed seismic report that was built on a set of 782 earthquakes records characterized by 13 variables, such as magnitude, depth, tsunami occurrence, and geographic coordinates. It is aimed at describing the statistical and spatio-temporal structure of seismicity as modeled in the dataset and pointing out patterns that have a direct impact on hazard assessment. In particular, the paper will focus on magnitude and tsunami distributions, assess annual variability of earthquake activity, determine spatial clustering of earthquakes, and assess the correlation between depth and magnitude to determine whether there is any depth-sensitive behavior in the observed seismic sequences or not.

The main reason why this work is needed is the existence of a depth-dependent magnitude structure which questions simplistic assumptions of feature independence. Evidence in the report shows that depth is closely correlated with magnitude behavior and that a threshold like change exists between shallow and deep depths and indicates that seismic characterization and screening processes can be enhanced by explicitly representing depth-sensitive signals. Simultaneously, spatial patterning suggests that the presence of earthquakes is non-uniform, meaning it is clustered, and region-specific methods of monitoring and modeling earthquakes are justified. Temporal results also suggest that the years when there are tsunami events are also the ones where the earthquake activity is high and it peaked in 2011 suggesting that tsunami-related periods may be the ones where seismic events are heightened or cascading.

Besides the physical event patterns, the outcomes indicate operational reasons that regard monitoring infrastructure. Rather than the uniformity in the number of stations, the report reveals more reliable signal measure values indicating an uneven observational ability which can influence the point of detection and certainty in the speed of characterization. In general, the results encourage depth-sensitive seismic risk modeling, hotspots-based surveillance and

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 3, No 12, December 2025
PP 1−14

combined earthquake-tsunami surveillance measures that value depth, spatial localization as well as temporal peculiarities instead of magnitude.

2. Literature Survey

Early studies on churn prediction primarily relied on statistical methods and traditional data mining techniques. Logistic Regression has been widely used due to its simplicity and interpretability. Tsai and Lu (2009) demonstrated the use of Logistic Regression in predicting customer churn, highlighting its effectiveness in identifying key predictors and providing interpretable results. Similarly, Burez and Van den Poel (2009) explored class imbalance in churn prediction, showcasing how balancing techniques can improve model performance.

With the advancement in computing power and the availability of large datasets, more complex models like Decision Trees and ensemble methods gained popularity. Breiman (2001) introduced Random Forests, an ensemble method that constructs multiple decision trees during training and outputs the mode of the classes for classification. This method improves predictive performance and robustness against overfitting. Liaw and Wiener (2002) further demonstrated the application of Random Forests in various domains, including churn prediction, emphasizing its accuracy and stability.

Idris, Rizwan, and Khan (2012) applied Random Forests to predict churn in the telecom industry, highlighting its superior performance compared to single-tree methods. The use of ensemble methods like Random Forests and Gradient Boosting has become a standard approach in churn prediction due to their ability to handle large, complex datasets and improve prediction accuracy.

3. Background

Earthquakes are sudden releases of accumulated strain Earthquakes refer to rapid discharge of built-up strain energy in the crust of the earth, which in most cases are along active plate boundaries and fault systems. The resultant seismic activities have the potential to create vibration at the grounds at extensive areas causing destruction of buildings, lifelines, and other vital infrastructure. Although the physical cause is tectonic, the extent of the impact is a product of source factors and exposure to society. Since the urbanization is growing into seismically unstable areas, even the same physical phenomenon is going to have a considerably different result based on population density, building norms, and emergency preparedness. Due to this reason, recent seismic risk evaluation methods have given greater emphasis on the amalgamation of occasion physics and geographical setting, as well as operational restrictions, instead of using an individual intensity marker.

Magnitude is the most used summary measure of earthquake size although it is not a full description of hazard. Magnitude measures the amount of energy released, but the potential to cause damage depends heavily on depth, distance to populated locations, soil conditions at the location, and there should be a rupture mechanism. Practically, depth causes the seismic waves to attenuate less until they reach the surface and shallow events tend to produce bigger surface shaking than the same magnitude events at a deeper depth. This is particularly necessary when it comes to fast assessment processes, where decision-makers are frequently forced to make decisions based on incomplete information and estimates that are time-constrained. It is also required that a depth-sensitive interpretation can be used to provide more reliable early-stage screening and prioritization.

Tsunamis are another complex issue since not all earthquakes result in the creation of tsunamis even when they are of large magnitude. The generation of a tsunami normally involves the seafloor movement, which is contingent on the site of the earthquake (offshore or inland), the regime of depths, the fault geometry and the vertical element of movement. In turn, combination of features is the best way of modeling tsunami risk conditionally, instead of considering tsunami likelihood as a homogeneous function of magnitude. This is operationally significant since warning systems need to trade between emergencies and accuracy: excessively broad tsunami alerts will cause expensive false alarms and undermine their credibility, whereas excessively strict criteria may cause delays to the warning system when infrequent but disastrous events occur.

Early warning and hazard response systems are based on the monitoring networks, which indicate the seismic signals and estimate the location, depth and magnitude of the earthquakes in near real-time. Nevertheless, quality monitoring is not even within regions. The coverage of sensors and good telemetry is enjoyed in some regions whereas the sparse network of stations, patchy maintenance and poor data quality are experienced in others. Such variations can influence the detection latency, localization accuracy, and uncertainty limits, which, subsequently, have an impact on the quality of reliable rapid characterization and downstream decision-making. The identification of observational capacity as an element of the hazard context is thus important in interpreting patterns in historical data, as well as when converting analytics into operational advice.

Analytically, earthquake datasets are usually skewed and skewed. Middle of the package regimes contain a large share of the events, and a small number of extreme events preoccupy the upper end and can frequently lead to the highest hazard. This framework questions modeling principles based on normality or equal variance and encourages more rigorous statistical methods based on distributional diagnostics, sensitive interpretation, and nonparametric logic. Equally, earthquakes are not uniformly distributed; they are likely to occur in tectonic boundaries and within seismically active belts, forming hotspots that bear unequal risk. It is also handy to find all these clusters to interpret scientifically as well as to design monitoring and mitigation plans.

The use of spatial clustering techniques is especially useful when studying hazards due to the ability to identify high densities of events without prescribing regular cluster forms. The density-based clustering method (unsupervised) can isolate the hotspots areas and diffuse events, which are consistent with the natural clustering of seismicity in the physical world. Together with correlation analysis and depth-magnitude inspection, these techniques encourage a more comprehensive view of the interaction of physical characteristics with the spatial context, and they can prevent overly simplistic ones drawn based on one or two variable summaries.

Another important parameter of seismic risk is temporal variability. The presence of earthquakes may vary on an annual basis with aftershock sequences earthquakes, changes in the stress of a region and clustering effects, and the periods of the tsunami may coincide with the high seismicity. Although coincidence over time does not necessarily mean causality, the detection of abnormal periods is operationally significant as it may be used to inform preparedness and resource policy and the formulation of surveillance strategies that react to regime transitions but not just to incidents. Data This paper will employ an earthquake dataset of 782 event records that have been compiled and characterized using 13 variables. The data is arranged on an event-by-event basis with each row corresponding to one earthquake observation and the columns reflecting a combination of physical parameters, spatial environment, and parameters to do with monitoring. The report specifies that the variables included will pertain to the magnitude and depth of the earthquake, an outcome measure concerning the tsunami and geographic coordinates (latitude and longitude), and other fields that are utilized to determine the measurement context and monitoring infrastructure. The main objective of exploiting this dataset is not to reproduce a comprehensive geophysical simulation but derive sound empirical patterns that help in the interpretation of hazards, variability analysis associated with tsunamis and monitoring implications.

There is a fundamental group of variables, which are considered as the hazard descriptors such as magnitude and depth. The central intensity measure is examined and analyzed as magnitude and summarized based on a robust distributional statistic providing insight into the central tendency and tails. The depth is also considered as a major contextual feature since it can significantly change the interpretation of hazards, and it also is directly analyzed in its connection with magnitude. Spatial context is presented in latitude and longitude, which allows us to map the density of events and identify the regions of seismicity concentration. The tsunami variable is considered as a nominal outcome and divides the data into tsunami-related and non-tsunami-related events periods, which enables the analysis to determine if the tsunami-tagged periods or conditions are uniformly distributed or they are concentrated in specific situations.

Besides the event physics and location, the dataset also has the variables associated with monitoring as explained in the report, especially those of signal and number of stations. Those are areas where the consistency of observational capacity and measurement are tested as being consistent across the dataset or they are significantly different across regions or even contexts of events. This is significant since the lack of station coverage can affect the detection latency, location precision and certainty in fast characterization which are operationally significant in early warning systems. The variability of the station the report puts focus on encourages the interpretation of the hazard signals and the observational capacity as opposed to assuming that all events are measured under the similar conditions of monitoring.

Since this paper was written based on an already-prepared report instead of the raw data processing pipeline, the dataset is assumed to be ingested and formatted to be analyzed. Consequently, the analysis is aimed at open interpretation of statistical evidence provided, distributional structure, clustering tendency and time trends, as opposed to asserting complete control over the quality of raw data. In places where it is pertinent, the paper formally considers observational coverage (say, number of stations) as an aspect of the analytic coverage and recognizes that the variability of infrastructure may serve to bias the quality of characterizing events and may bias inter-regional or inter-temporal comparisons.

4. Methodology

It is analyzed in a framework of report workflow including descriptive statistics, distributional diagnostics, correlation analysis, spatial clustering, and temporal aggregation to identify patterns of interest to earthquake hazard interpretation and variability associated with tsunamis. This involves a process of identifying the baseline descriptive characteristics of magnitude and other variables of interest by using summary statistics and visualization based on distributions. Magnitude behavior is viewed in a distributional perspective to find locations where the concentration of events occurs, whether the distribution is heavy-tailed or symmetric, and whether unusual events have a material divergence out of the normal regime. The interquartile range (IQR) method of identifying outliers is a non-parametric tool that is robust and offers a non-normative method of identifying extreme values. To further confirm distributional

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 3, No 12, December 2025
PP 1−14

assumptions, a normality test (JarqueBeratest) is used and makes it possible to explicitly indicate violations of the Gaussian form and rationalize approaches that are resistant to skewness and tail risk.

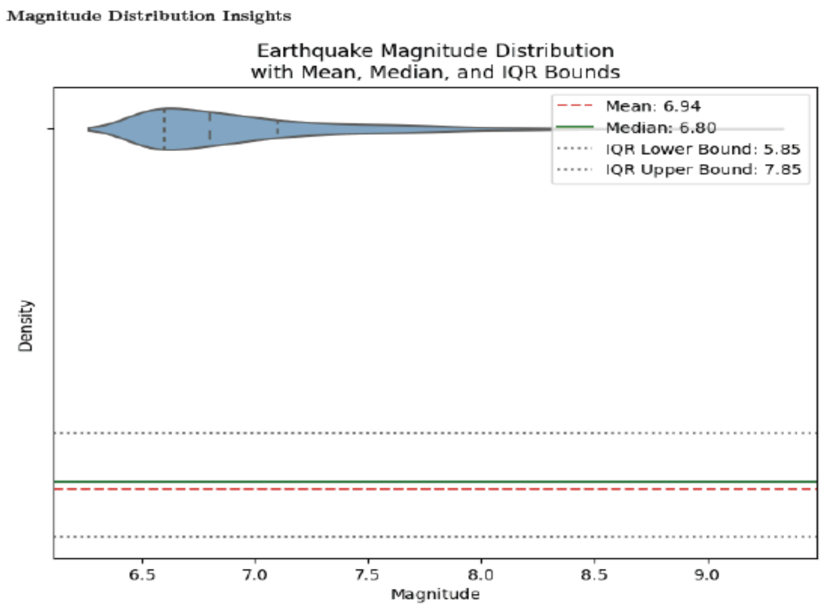

Fig.1 Magnitude Distribution Insights

The incidence of tidal waves is nominal data with the risk of unequal distribution of classes. The report-based workflow measures the relativity of the prevalence of tsunami-related and non-tsunami-related events and takes the detected imbalance, as a sign that the risk of tsunamis is not evenly distributed by event. An inferential framing is a simple example of a framing that is applied to assess the significance of the imbalance in the presence of usual assumptions about categorical outcomes[13,16,17]. The main purpose of using this component is to encourage conditional interpretation of hazards, i.e. tsunami screening must include physical and contextual cues as opposed to uniformity in the probability applied to individual earthquakes.

Correlation analysis is used to assess the relationships between continuous variables and Pearson correlation is the main linear measure of association, which is backed up by heat map-based visualization to determine the predominant relationships. Special focus is made on the depth magnitude relationship since the report suggests a depth-dependent magnitude structure that does

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 3, No 12, December 2025
PP 1−14

not follow the general independence assumptions. The piece of work integrates correlation evidence with scatter-based inspection and regression-like interpretation to determine whether the depth-magnitude relation is likely to be linear at all depths or whether it varies across regimes. To capture behavior of regimes, the report makes a reference to a nonlinear regression interpretation that defines a critical shallow-depth threshold (around 15 km) in which changes in magnitude behavior become more pronounced with depth[1,5,7]. They employ it as an operational insight: they encourage depth-conscious hazard modelling and posit that depth can be a useful modifier in speedy assessment pipelines and tsunami-based screening.

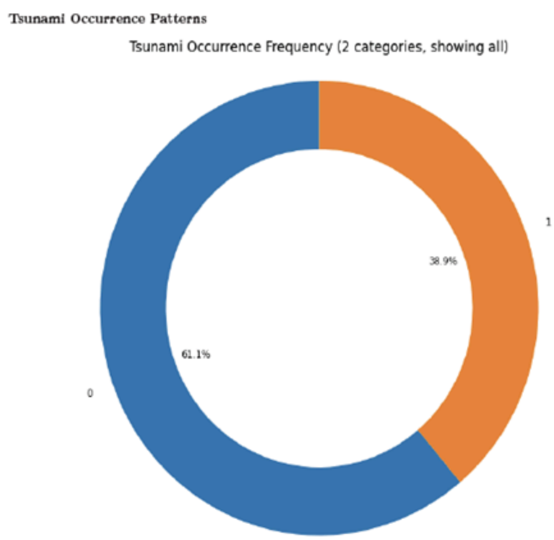

Fig.2 Tsunami Occurrence Patterns

Spatial structure is examined through unsupervised clustering of geographic coordinates through DBSCAN. The technique is suited to seismicity since it determines dense regions of activity without having to specify the number of clusters a priori and can deal with irregular shapes of clusters as found in tectonic environments[9.12]. DBSCAN also automatically considers sparse observations to be noise, which is used to differentiate between concentrated seismic hotspots and dispersed events. These clustering findings are discussed as the fact that seismicity is not evenly distributed across geography, and thus the different areas should be monitored and localized hazard modeling. Temporal behavior is analyzed by year-by-year aggregation of earthquake occurrences and comparison between tsunami tagged and non-tsunami tagged periods. It is this step that is utilized to identify whether tsunami-related years show a consistent increase of earthquake activity and whether any year is an outlier in its behavior based on a

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 3, No 12, December 2025
PP 1–14

higher seismic activity. The report brings to the fore a sharp peak year in the tsunami-related seismicity, and this time scale deviation is identified as an incentive in integrated earthquake-tsunami monitoring approaches that involve both time regime identification and no single-event magnitude limits[2,4].

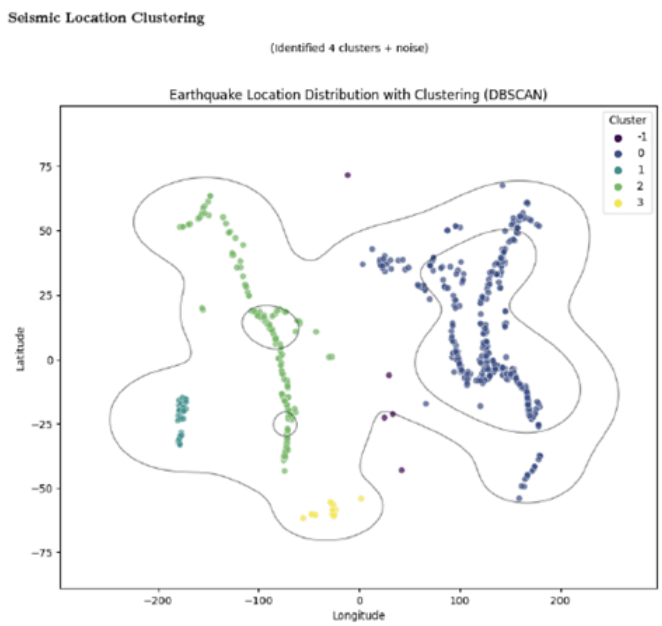

Fig 3. Earthquake Location Distribution

Lastly, the operational monitoring lens is incorporated in the study as it analyses the distribution of variables of the measurement, namely, the variable signal and the number of stations[14,15]. The workflow also compares the dispersion and skewness of these fields to determine whether there is a tendency of measurement practice to be standardized and a variable display of infrastructure deployment. The number of stations is viewed as an indicator of observational capacity and the right-skewed distribution, and extreme outliers are seen as indicators of in-homogenous monitoring coverage [3,6,8]. The report-based methodology further compares the adequacy of the coverage to a given industry standard, and measures prevalence of shortfall in different regions, providing incentive to use coverage-conscious uncertainty in decision-making on early warning and risk assessment.

5. Result Analysis

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 3, No 12, December 2025
PP 1−14

The magnitude distribution of the data demonstrates that the seismic activity is concentrated on moderate and large earthquakes as opposed to being evenly distributed among the smaller ones. The center of interest of most observations lies in a comparatively few-banded range around magnitude 6.8 with smaller numbers expanding into the extreme range. This structure implies heavy-tailed risk profile whereby frequent, moderate-intensity occurrences of events are present with rarer but potentially high-impact earthquakes. Hazardwise, this is important since summary averages may miss tail risk, and preparedness planning needs to take into consideration a small set of events that may take a preponderation of total damage and disruption. The presence of a tsunami does not follow a similar distribution within the data set and is represented as an inequitable phenomenon which means that earthquakes associated with a tsunami are not a random sample but a subset of all earthquakes. This is in keeping with the operational perception that tsunami risk cannot be represented as a fixed probability to all earthquakes but rather conditionally on a basis of physical and contextual cues like depth regime and space position. This in effect encourages integrated screening logic with tsunami alerts remaining based on a mix of features and not magnitude.

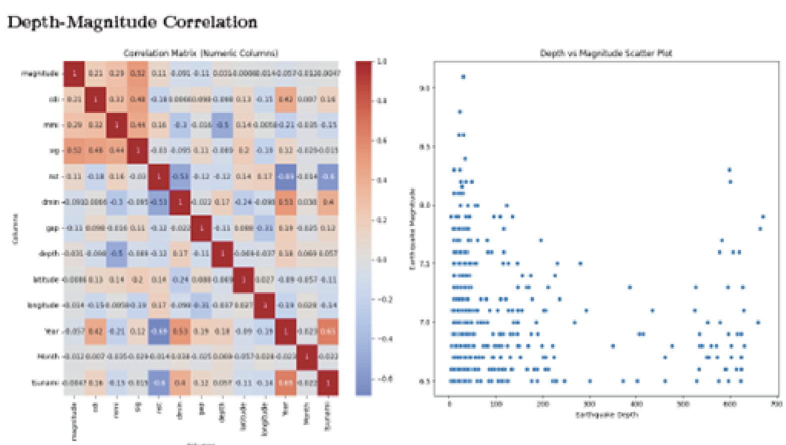

Fig.4 Depth Magnitude Correlation

The depth-dependent magnitude behavior structure is among the most significant discoveries. The statistical evidence that has been reported demonstrates that depth is closely related to magnitude and that relationship is not linear in nature and that there is a low-depth limit beyond which behavior in magnitude changes more aggressively. This depth-magnitude pattern directly relates to risk interpretation since depth will have effects on the characteristics of ground shaking and the potential of occurrence of secondary hazard in coastal areas. Explicit addition of depth to

The Journal of Computational Science and Engineering (TJCSE)
ISSN 2583-9055 (Media Online)
Vol 3, No 12, December 2025
PP 1–14

seismic characterization can thus enhance prioritization and remove misclassification in fast assessment processes.

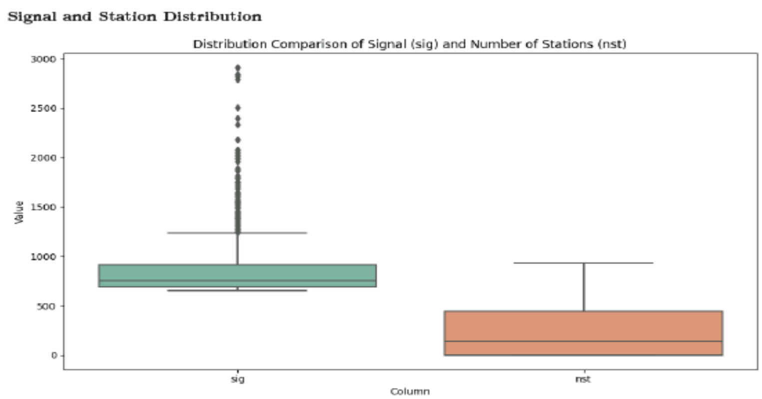

Fig.5 Signal and Station Distribution

Spatial patterning also suggests that seismicity is not distributed evenly over geographic space but is clustered instead. The unsupervised clustering of the latitudes and the longitudes only identifies few dense clusters that are an indication of the seismic hotspots, with the few scattered points that may not be the cores of a cluster. This confirms that the assessment and monitoring strategies towards hazards must be region specific: high monitoring density and localized modeling assumptions and specific mitigation investment are needed in hotspots regions whereas low monitoring density and localized detection and response strategies may be needed in low density regions to achieve similar reliability.

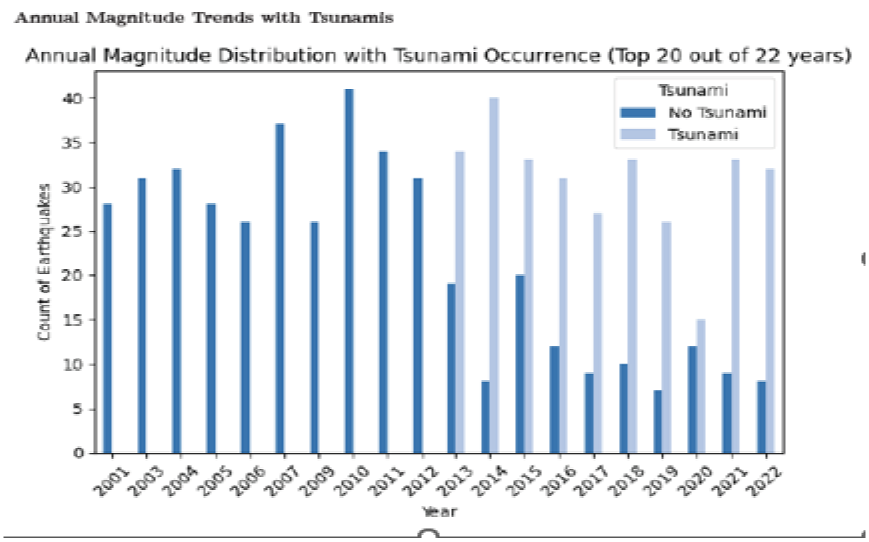

Fig.6 Tsunamis Annual Magnitude trends

The temporal analysis is used to connect the tsunami-tagged periods to the increase in seismic activity, resulting in an apparent peak year of tsunami-related seismicity. Although this tendency cannot be determined causal, it implies that tsunami incident can be correlated with times of high or cascading seismic activity that are operationally applicable in readiness planning. This would encourage the management of tsunami-related periods as increased vigil regimes in which the frequency and probability of impact could be increased.

Lastly, there is an operational disproportion in the surveillance infrastructure, as noted in the report. Signal-related values seem to be relatively stable whereas station count metrics are relatively different and highly skewed. This randomness suggests that the observational capacity between regions is uneven, and this may affect the latency of detection, location accuracy, and rapid characterization confidence. In practice, these findings would be useful both to enhance coverage in under-instrumented regions and to address coverage-aware uncertainty in both earthquake characterization and tsunami warning.

Spatial patterning also suggests that seismicity is not distributed evenly over geographic space but is clustered instead. The unsupervised clustering of the latitudes and the longitudes only identifies few dense clusters that are an indication of the seismic hotspots, with the few scattered points that may not be the cores of a cluster. This confirms that the assessment and monitoring strategies towards hazards must be region specific: high monitoring density and localized modeling assumptions and specific mitigation investment are needed in hotspots regions whereas low monitoring density and localized detection and response strategies may be needed in low density regions to achieve similar reliability.

Conclusion:

This study demonstrates that earthquake hazard assessment cannot rely on magnitude alone, as depth-dependent magnitude behavior, spatial clustering, and temporal tsunami associations significantly influence risk interpretation. The findings highlight the need for depth-aware, region-specific, and integrated seismic–tsunami monitoring frameworks to improve early warning accuracy and hazard preparedness.

References

1. Pawar, R., Ghumbre, S., Deshmukh, R., & Kolhe, K. (2020). A hybrid approach towards improving performance of recommender system using matrix factorization techniques. International Journal of Future Generation Communication and Networking, 13(4), 467-477.
2. Kale, D. R., Mane, T. S., Buchade, A., Patel, P. B., Wadhwa, L. K., & Pawar, R. G. (2024, October). Federated Learning for Privacy-Preserving Data Mining. In 2024 International Conference on Intelligent Systems and Advanced Applications (ICISAA) (pp. 1-6). IEEE.
3. Limkar, S., Baser, S., Jhamnani, Y., Shinde, P., Jithu, R., & Chinchmalatpure, P. (2019, June). Chatbot: Integration of Applications Within Slack Channel. In International Conference on Intelligent Computing and Communication (pp. 551-558). Singapore: Springer Singapore.
4. Vayadande, K., Bhosle, A. A., Pawar, R. G., Joshi, D. J., Bailke, P. A., & Lohade, O. (2024). Innovative approaches for skin disease identification in machine learning: A comprehensive study. Oral Oncology Reports, 10, 100365.
5. Bird, P., & Kagan, Y. Y. (2004). Plate-tectonic analysis of shallow seismicity: Apparent boundary width, beta, corner magnitude, coupled lithosphere thickness, and coupling in seven tectonic settings. Bulletin of the Seismological Society of America, 94(6), 2380–2399. https://doi.org/10.1785/0120030107
6. Gutenberg, B., & Richter, C. F. (1954). Seismicity of the Earth and associated phenomena (2nd ed.). Princeton University Press.
7. Helmstetter, A., Kagan, Y. Y., & Jackson, D. D. (2007). High-resolution time‐independent grid‐based forecast for M ≥ 5 earthquakes in California. Seismological Research Letters, 78(1), 78–86. https://doi.org/10.1785/gssrl.78.1.78
8. Kanamori, H. (1977). The energy release in great earthquakes. Journal of Geophysical Research, 82(20), 2981–2987. https://doi.org/10.1029/JB082i020p02981
9. Lay, T., & Wallace, T. C. (1995). Modern global seismology. Academic Press.
10. Okada, Y. (1985). Surface deformation due to shear and tensile faults in a half-space. Bulletin of the Seismological Society of America, 75(4), 1135–1154.
11. Dixit, B., Pawar, G., Gayakwad, R., Joshi, M., Mahajan, R., & Chinchmalatpure, S. V. (2023). Challenges and a novel approach for image captioning using neural network and searching techniques. International Journal of Intelligent Systems and Applications in Engineering, 11(3), 712-720.

12. Scholz, C. H. (2019). The mechanics of earthquakes and faulting (3rd ed.). Cambridge University Press. https://doi.org/10.1017/9781316681473
13. Shearer, P. M. (2009). Introduction to seismology (2nd ed.). Cambridge University Press.
14. Limkar, S., Hunashimarad, S., Chinchmalatpure, P., Baj, A., & Patil, R. (2020). Potential of robust face recognition from real-time CCTV video stream for biometric attendance using convolutional neural network. In Intelligent Data Engineering and Analytics: Frontiers in Intelligent Computing: Theory and Applications (FICTA 2020), Volume 2 (pp. 11-20). Singapore: Springer Singapore.
15. Stein, S., & Wysession, M. (2003). An introduction to seismology, earthquakes, and Earth structure. Blackwell Publishing.
16. Chinchmalatpure, P., Chinchmalatpure, S., & Konde, S. (2025). Determinants of cyber incident severity: Exploratory evidence and benchmark prediction of financial loss. International Journal of Scientific Research in Engineering and Technology (IJSRET), 11(1), 1–6.
17. Wald, D. J., & Allen, T. I. (2007). Topographic slope as a proxy for seismic site conditions and amplification. Bulletin of the Seismological Society of America, 97(5), 1379–1395. https://doi.org/10.1785/0120060267