Spatial–Temporal Point Pattern Analysis of Earthquakes in India and Adjacent Regions (2010–2025)

Mohammad Saqib Ansari

mohammad.ansari@kaust.edu.sa

CEMSE-Statistics
King Abdullah University of Science and Technology

December 8, 2025

What is an earthquake? "An earthquake is an intense shaking of Earth's surface caused by movements in its outer layer." — NASA SpacePlace.

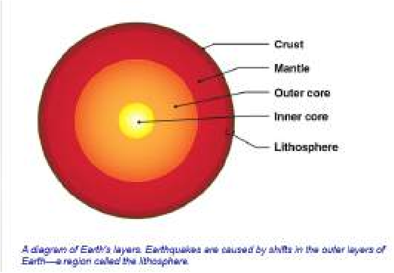
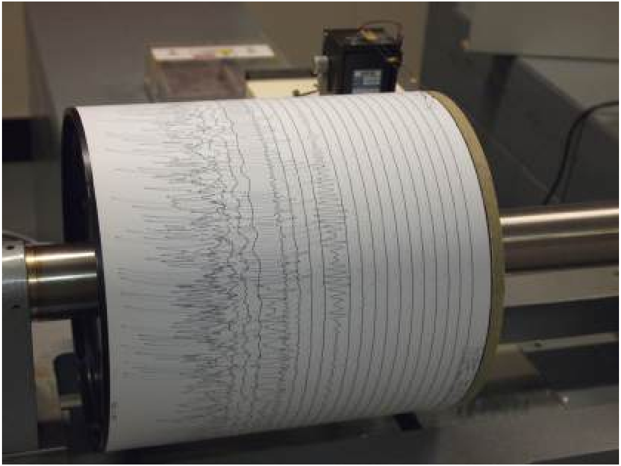


Figure 2: Earth Structure. Source: NASA SpacePlace

Seismogram Example



Seismogram recorded at the Weston Observatory. Source: *(Wikipedia)*

Global Impact of Earthquakes

Earthquakes cause major loss of life and property worldwide. Historically, some of the most devastating events include:

- **2008 Sichuan, China (M 8.0):** Over 70,000 deaths.
- **2003 Algeria (Boumerdès, M 6.8):** 2,200 deaths.
- **1999 Turkey (Marmara, M 7.4):** 17,000 deaths.

Globally, earthquakes account for **about 60,000 deaths annually**, with nearly **90% occurring in developing countries** (OECD, 2008).

Most fatalities result from **building collapse**. Economic losses are also severe—for example, the **1999 Marmara earthquake caused over $5 billion in damages** (World Bank, 2005).

2001 Bhuj Earthquake

The Indian EXPRESS

JOURNALISM OF COURAGE

In Bhuj's rebuilding, the Gujarat approach is widely looked at as a model for reconstruction. (Express Archive Photo)

As the nation celebrated its 68th Republic Day, Gujarat mourned the 16th anniversary of the worst disaster that struck the state on January 26, 2001. Gujarat's historic earthquake killed over 20,000 people, injuring 1,66,000, destroying nearly 4,00,000 homes. The shock waves spread over 700 km; 21 districts were affected and 6,00,000 people left homeless. While many believed that Gujarat would take years to

News excerpt from *(The Indian Express)*

Photo by Arko Datta (AFP/Getty Images)

Photo by Paula Bronstein (Hulton Archive/Getty Images)

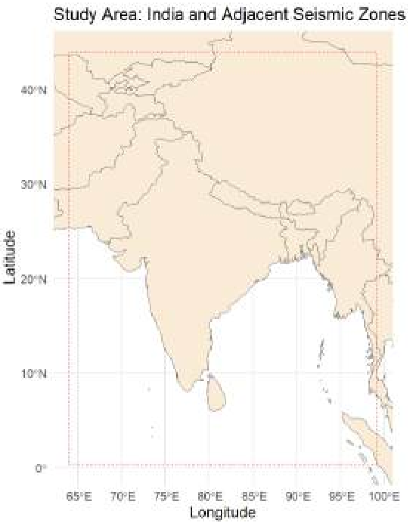
Source: Britannica Kids

Study Area and Dataset

The dataset for this study was obtained from the United States Geological Survey (USGS) Earthquake Catalog, covering India and surrounding seismic zones.

- **Geographical coverage:** Latitudes 0.40°N – 43.98°N, Longitudes 63.99°E – 99.14°E.
- **Time period:** 2010–2025.
- **Number of events:** 14,336 earthquake occurrences.
- **Attributes included:** Magnitude, depth, latitude, longitude, date/time, etc.
- **Additional data:** Mean temperature, elevation, and precipitation obtained using the `terra` and `get_data` packages in R.

Study Area



Click here for 3D Earthquake Plot

Objectives

This study aims to analyze the spatio-temporal patterns of earthquakes in India and adjacent regions (2010–2025) with the following objectives:

- Explore spatial and temporal trends of earthquakes through exploratory data analysis (EDA).
- Investigate clustering, randomness, or regularity of epicenters using quadrant count analysis, Ripley's K-function, LGCP and Kernel Density Estimation (KDE).
- Examine spatio-temporal behavior to detect changes in seismic activity over time.
- Identify seismic hotspots and high-risk zones via intensity mapping and spatial visualization.
- Interpret findings in the context of regional tectonics for hazard assessment and risk mitigation.

Methodology

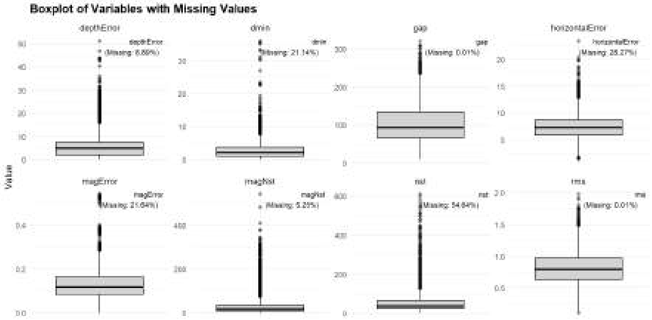
1. **Exploratory Data Analysis (EDA)**
2. **Temporal Analysis**
3. **Spatial Analysis**
4. **Point Pattern Analysis**
5. **Modeling Using Statistical and ML models**

1. Exploratory Data Analysis (EDA)

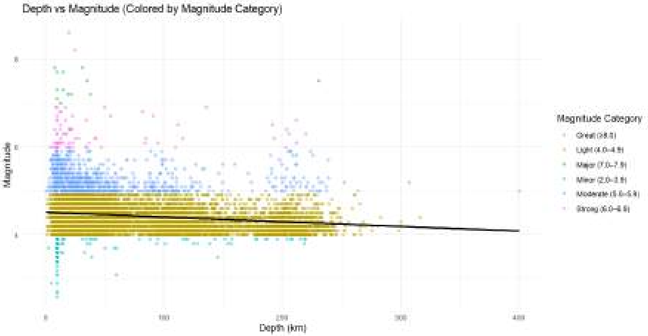
The earthquake dataset (2010–2025) contains 14,336 events. For this analysis, we focus on the key variables: **time, latitude, longitude, depth, and magnitude**.

- **Longitude:** 63.99°E – 99.14°E, median 85.95°E
- **Latitude:** 0.41°N – 43.98°N, median 30.06°N
- **Depth (km):** 0.6 – 400.57, median 25.61 km
- **Magnitude:** 2.6 – 8.6, median 4.4
- **Time period:** 2010-01-01 02:22:23 UTC – 2025-11-26 05:28:38 UTC

1. Missing Values



1. Depth vs Magnitude



1. Earthquake Depth Classification

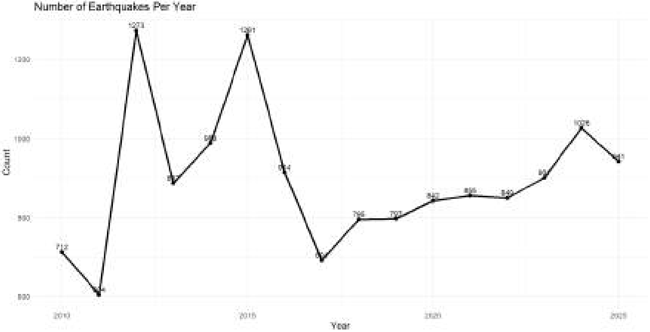
Earthquakes were classified into three categories based on focal depth:

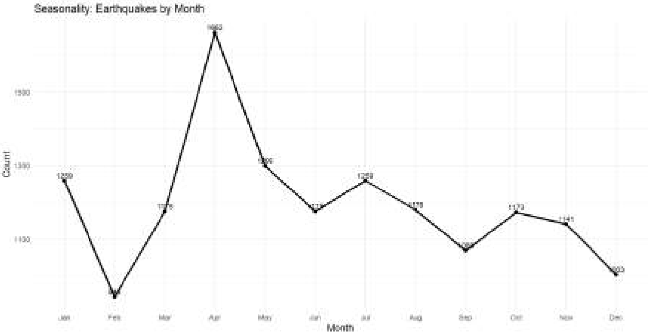
- **Shallow:** ≤ 70 km
- **Intermediate:** 70–300 km
- **Deep:** > 300 km

Depth Category	Count
Shallow	10595
Intermediate	3736
Deep	5

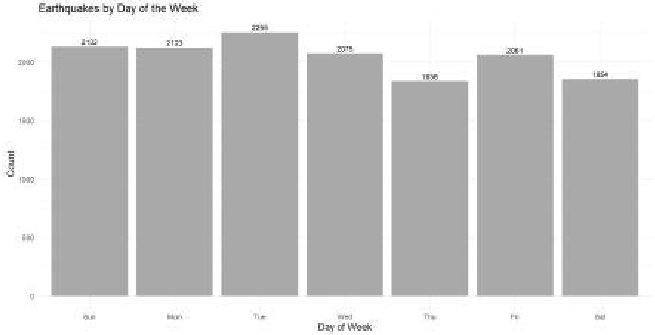
Classification summary from the dataset.

2. Yearly Earthquake Count

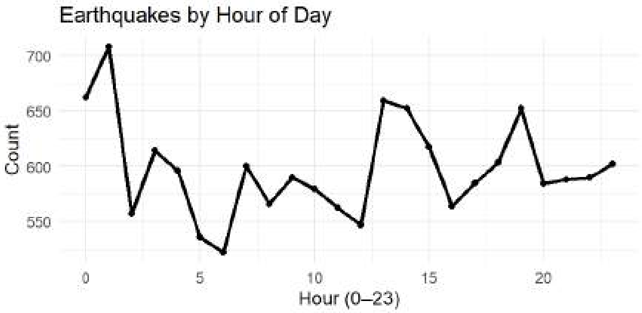



Seasonality: Earthquakes by Month
Count
1500
1300
1100
1173
1141
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep
Oct
Nov
Dec
Month

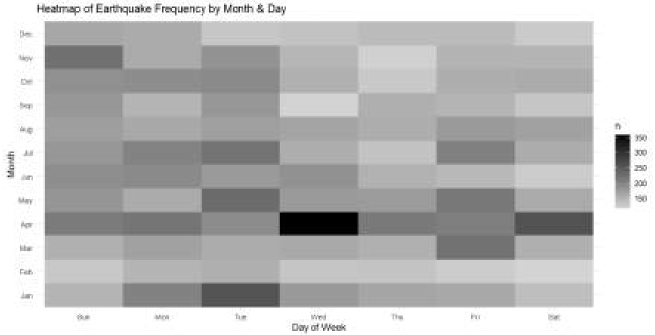
2. Trend by Days



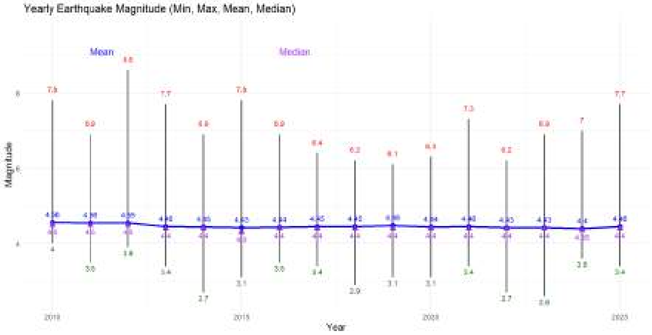
2. Overall Hourly Trend



2. Monthly-Hourly Trend

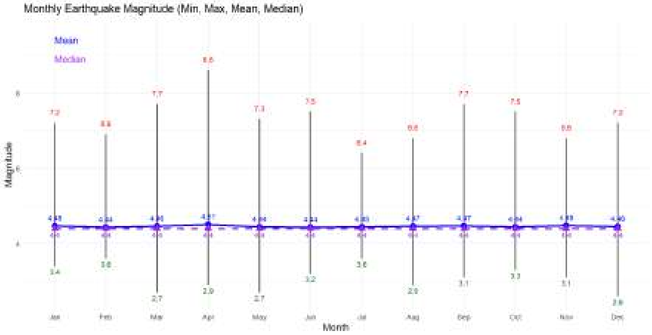


جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of
Science and Technology

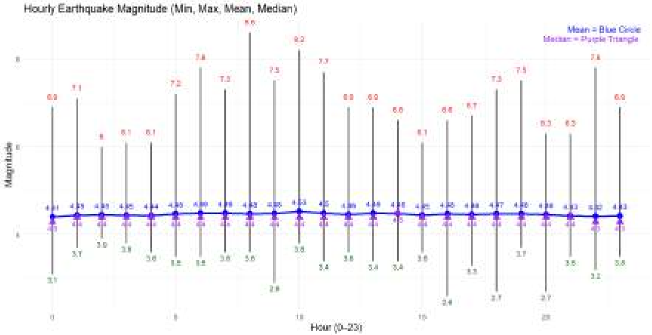


2. Monthly Magnitude Trend

جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology

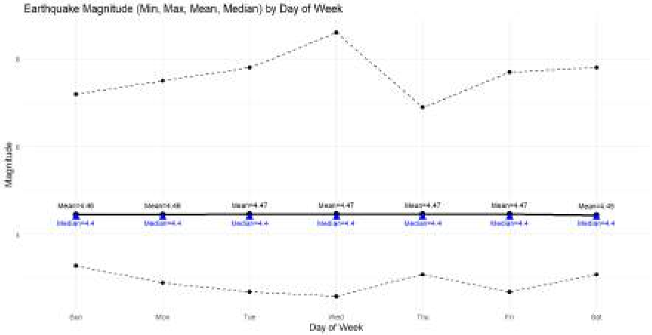


2. Hourly Magnitude Variation

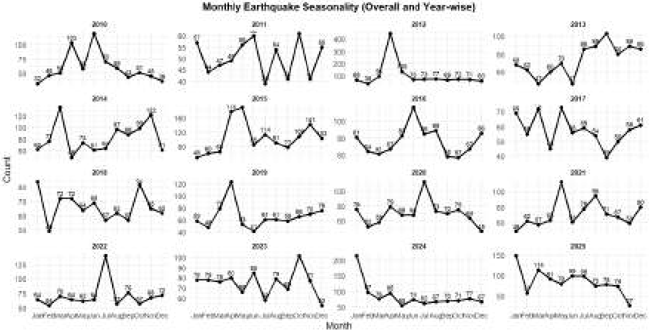


2. Magnitude Variation by Day of Month

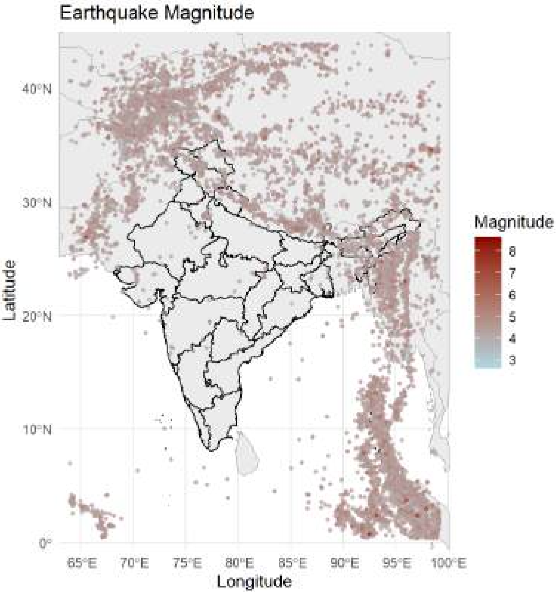
جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of
Science and Technology



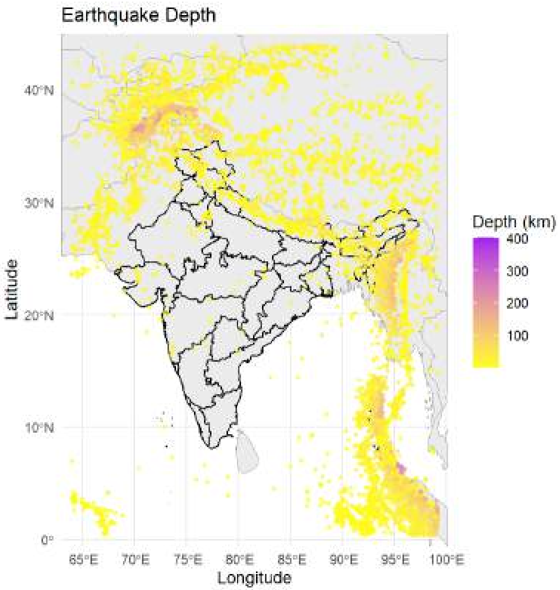
جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology



3.1 Magnitude Spatial Map



3.1 Depth Spatial Map



Hotspot analysis identifies statistically significant spatial clusters of earthquake activity using the **Getis–Ord** G_i^* statistic.

$$G_i^* = \frac{\sum_{j=1}^n w_{ij} x_j - \bar{X} \sum_{j=1}^n w_{ij}}{\sqrt{\left(\frac{\sum_{j=1}^n w_{ij}^2 - \frac{(\sum_{j=1}^n w_{ij})^2}{n}}{n-1} \right) \sum_{j=1}^n (x_j - \bar{X})^2}}$$

Variables:

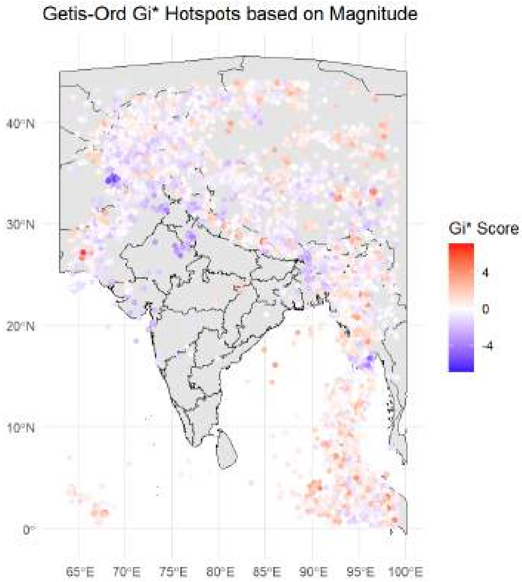
- x_j, w_{ij}
- $\bar{X}$, n

Meaning:

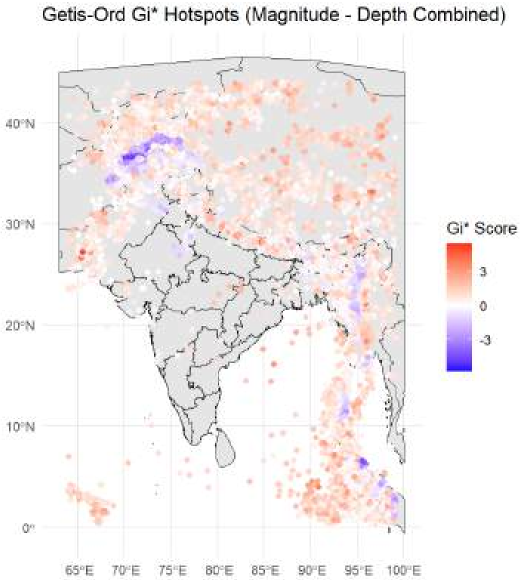
- x_j: attribute at location j
- w_{ij}: spatial weight
- $\bar{X}$: mean of x_j
- n: total locations

Theoretical Range: $G_i^* \in (-\infty, +\infty)$, positive = hotspot, negative = coldspot, near 0 = no clustering.

3.2 Gi* Hotspot Analysis: Magnitude



3.2 Gi* Hotspot Analysis: Combined



4.1 Quadrant Count Analysis

King Abdullah University of Science and Technology

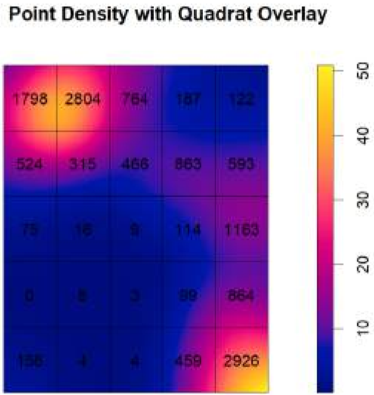
Quadrant (quadrat) analysis examines the spatial distribution of earthquake epicenters. The study region is divided into $n \times n$ quadrants and a Chi-square test is applied to test **Complete Spatial Randomness (CSR)**.

$$\chi^2 = \sum_{i=1}^{n^2} \frac{(O_i - E_i)^2}{E_i}$$

Where:

- O_i = observed number of points in quadrant i
- E_i = expected number of points in quadrant i under CSR
- n = number of divisions along each axis

Result: All Chi-square tests returned $p = 0$, strongly rejecting CSR; earthquake points are **clustered**.



Quadrant Analysis showing earthquake counts per 5×5 grid with intensity background from KDE.

Ripley's K-function is used to assess spatial point patterns by comparing the observed clustering of earthquake epicentres with what would be expected under Complete Spatial Randomness (CSR).

- $K(r)$ measures how many events occur within distance r of a typical point.
- $K_{\text{obs}}(r)$ is compared against simulation envelopes from CSR.
- If $K_{\text{obs}}(r)$ lies **above** the envelope $\rightarrow$ **clustering**.
- If below $\rightarrow$ **dispersion**.

Interpretation for this study:

- The envelope is **very thick**, indicating high uncertainty due to (a) strong spatial inhomogeneity and (b) large variation in event intensity across the region.
- $K_{\text{obs}}(r)$ exceeds the CSR envelope at several scales, suggesting **multi-scale clustering** of earthquakes.

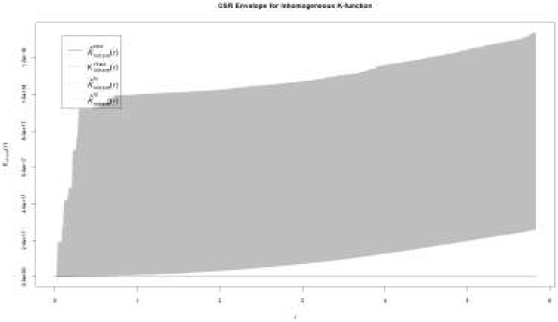


Figure 3: CSR Envelope for Inhomogeneous K-function

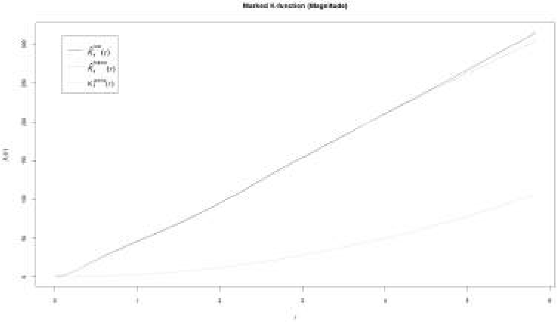


Figure 4: Marked K-function using Magnitude as Mark

4.2 Marked K-function (Depth)

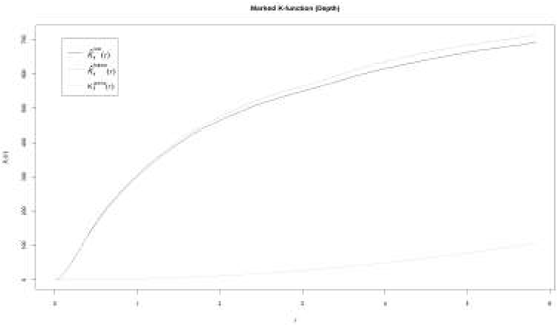


Figure 5: Marked K-function using Depth as Mark

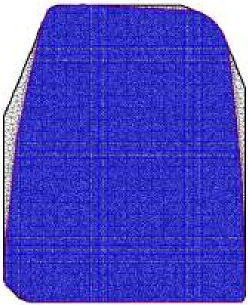
LGCP was used to model the spatial distribution of earthquake events, treating counts in each grid cell as a Poisson process with spatially varying intensity.

$$y_i \sim \text{Poisson}(\mu_i), \qquad \log(\mu_i) = \log(A_i) + \beta_0 + s(\mathbf{s}_i), \tag{1}$$

- A_i: pixel area (offset)
- $s(\mathbf{s}_i)$: latent Gaussian random field capturing spatial correlation

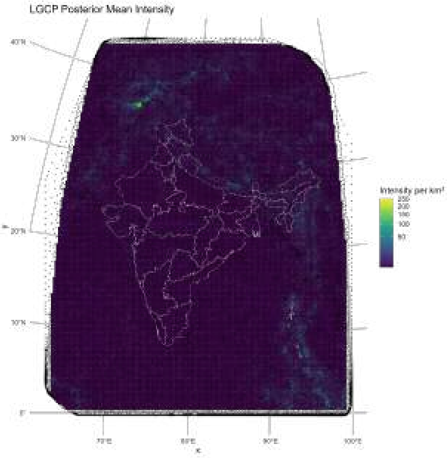
The spatial field was modeled using the **SPDE approach**, and inference was performed via **INLA**, allowing fast Bayesian estimation of large spatial models.

A grid search over candidate **PC priors** (range, marginal SD) for the SPDE parameters was conducted. The best model was selected using the **Deviance Weighted Information Criterion (DWIC)**.



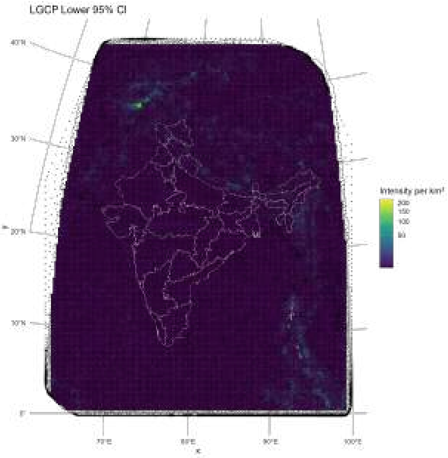
INLA mesh used for fitting the Log-Gaussian Cox Process (LGCP) model.

4.3 LGCP Intensity Estimate — Mean



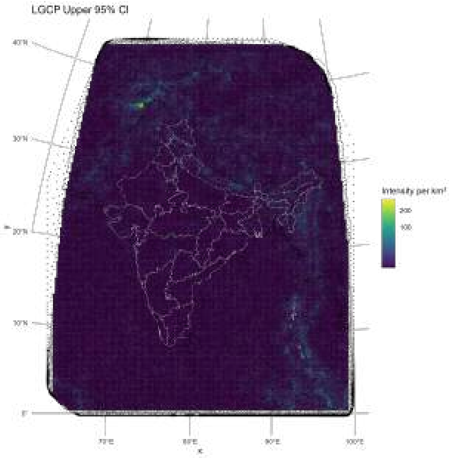
Mean intensity (events per km^2) estimated from the LGCP model using INLA.

LGCP Intensity Estimate — Lower CI



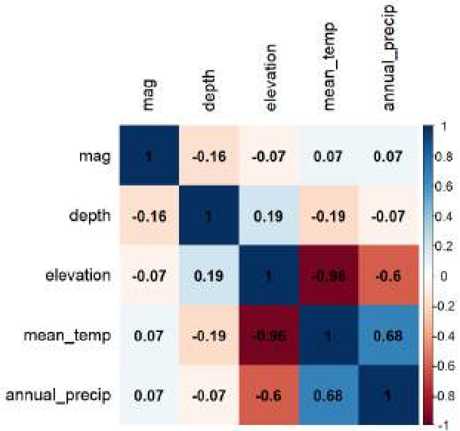
Lower 95% credible interval bound for the LGCP intensity surface.

LGCP Intensity Estimate — Upper CI



Upper 95% credible interval bound for the LGCP intensity surface.

5. Covariate Correlation Plot



	mag	depth	elevation	mean_temp	annual_precip
mag	1	-0.16	-0.07	0.07	0.07
depth	-0.16	1	0.19	-0.19	-0.07
elevation	-0.07	0.19	1	-0.96	-0.6
mean_temp	0.07	-0.19	-0.96	1	0.68
annual_precip	0.07	-0.07	-0.6	0.68	1

Figure 6: Correlation matrix of earthquake covariates.

5.1 ARIMA with covariates

Notation (used throughout):

- Y_t : observed response at time t (magnitude or depth)
- $\mathbf{x}_t = (\mathrm{lat}_t,\ \mathrm{lon}_t)$: covariates
- ε_t : white-noise innovations, $\varepsilon_t \sim N(0, \sigma^2)$

General ARIMA(p, d, q) model:

$$\Phi(B)(1-B)^d Y_t = \Theta(B)\,\varepsilon_t,$$

where

$$\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p, \qquad \Theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q.$$

Regression with ARIMA errors (consistent form):

$$Y_t = c + \mathbf{x}_t^\top \beta + \text{ARIMA}(p, d, q) \text{ error on } Y_t.$$

Model:

$$Y_t = c + \beta_1 \,\text{lat}_t + \beta_2 \,\text{lon}_t + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_5 Y_{t-5} + \theta_1 \varepsilon_{t-1} + \varepsilon_t.$$

This is exactly ARIMA(5,1,1) applied to Y_t with covariates.

Estimated coefficients

Parameter	Estimate	Std. Error
AR(1)	0.0675	0.0091
AR(2)	0.0332	0.0091
AR(3)	0.0084	0.0090
AR(4)	0.0238	0.0090
AR(5)	0.0255	0.0090
MA(1)	-0.9823	0.0028
Latitude	-0.6568	0.7324
Longitude	-24.2613	0.7030

Model fit:
$\sigma^2 = 3130,\ \ell = -72905.7,\ AIC = 145829.4,\ BIC = 145896.9.$

Model:

$$Y_t = \beta_0 + \beta_1 \, \mathrm{lat}_t + \beta_2 \, \mathrm{lon}_t + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \phi_3 Y_{t-3} + \phi_4 Y_{t-4} + \phi_5 Y_{t-5} + \varepsilon_t.$$

(This corresponds to ARIMA(5,1,0) with covariates.)

Estimated parameters

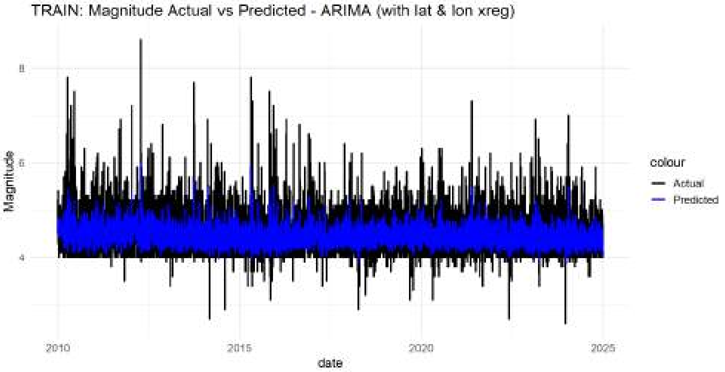
Parameter	Estimate	Std. Error
AR(1)	-0.7903	0.0086
AR(2)	-0.6110	0.0106
AR(3)	-0.4564	0.0112
AR(4)	-0.3132	0.0106
AR(5)	-0.1463	0.0086
Latitude	-0.0163	0.0050
Longitude	0.0481	0.0048
$\sigma^2 = 0.1606, \; \ell = -6754.17, \; AIC = 13524.33, \; BIC = 13584.36$		

ARIMA models fitted using auto-ARIMA with spatial covariates (latitude, longitude).

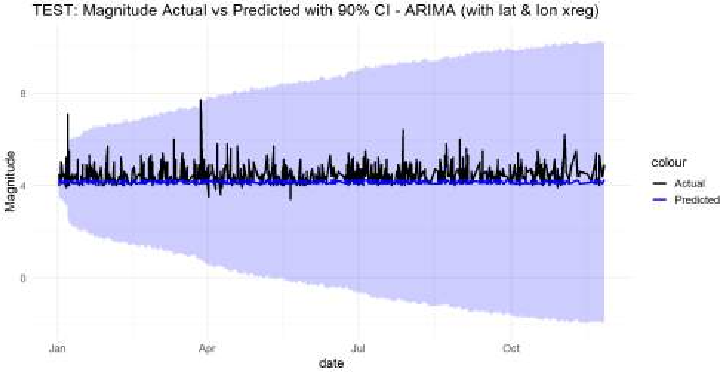
Model	LogLik	AIC	Train RMSE	Test RMSE
ARIMA(5,1,0) – Mag.	−6754.17	13524.33	0.401	0.646
ARIMA(5,1,1) – Depth	−72905.70	145829.40	55.93	56.90

Performance summary of ARIMA models for magnitude and depth.

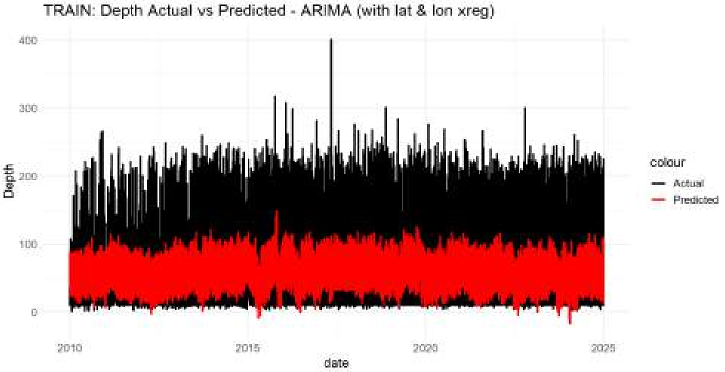
5.1 ARIMA – Prediction (Training)



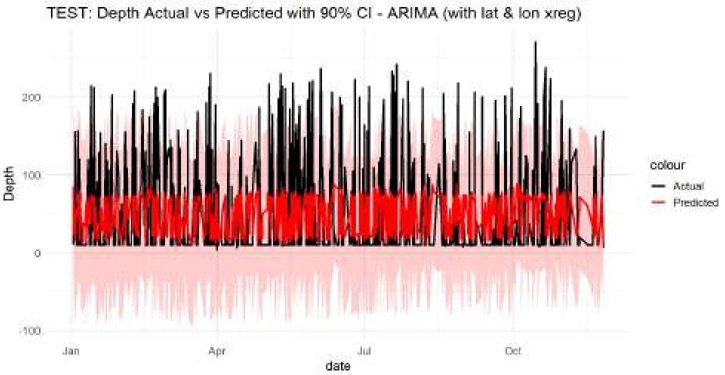
5.1 ARIMA Prediction – Magnitude (Test)



ARIMA Prediction – Depth (Train)



5.1 ARIMA Prediction – Depth (Test)



جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology

Light Gradient Boosting Machine (LGBM): A fast, efficient tree-based boosting algorithm using leaf-wise growth with depth constraints. It is well-suited for large datasets and handles non-linear patterns effectively.

To know more: LGBM Tutorial (GfG)

LGBM Performance Summary

Model	Best Params / Rounds	Train RMSE	Test RMSE
LGBM – Magnitude	63 leaves, lr=0.1, 1000 rounds	0.197	0.427
LGBM – Depth	127 leaves, lr=0.1, 1000 rounds	7.10	27.87

جامعة الملك عبدالله
للعلوم والتقنية
King Abdullah University of
Science and Technology

Past Experience

Top 4 Decile of Model on Train Data

Transformation	Logistic	XgBoost	CatBoost	Light GBM	RF	G NB	ET
Standard Scaler	0.800		0.875	0.855	1.000	0.780	1.000
Min Max	0.800	0.905	0.874	0.855	1.000	0.780	1.000
Log	0.803	0.905	0.874	0.853	1.000	0.789	1.000
Square root	0.799	0.905	0.875	0.856	1.000	0.783	1.000
yeo-jhonson	0.809	0.905	0.875	0.854	1.000	0.789	1.000

Top 4 Decile of Model on Validation Data

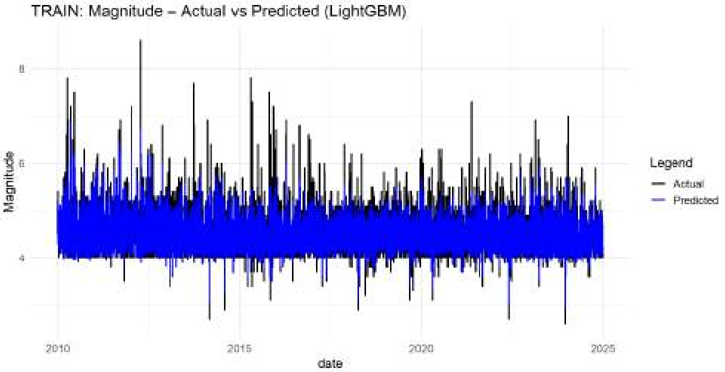
Transform	Logistic	XgBoost	CatBoost	Light GBM	RF	G NB	ET
Standard !	0.794	0.809	0.815	0.819	0.798	0.776	0.790
Min Max	0.794	0.809	0.815	0.819	0.803	0.776	0.777
Log	0.800	0.809	0.815	0.817	0.801	0.789	0.787
Square roo	0.790	0.809	0.815	0.820	0.802	0.781	0.789
yeo-Jhons	0.809	0.809	0.815	0.816	0.796	0.784	0.783

Top 4 Decile of Model on Test Data

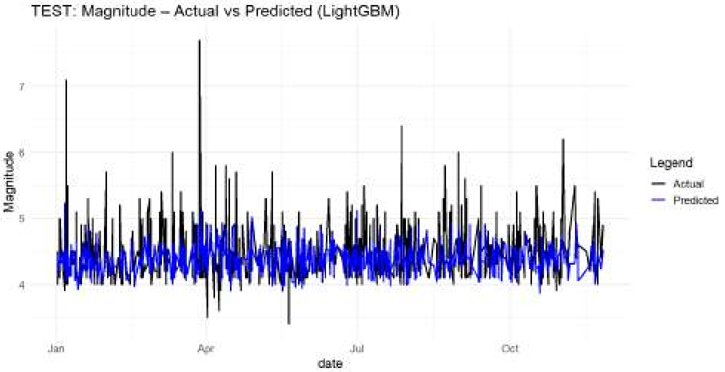
Transformation	Logistic	XgBoost	CatBoost	Light GBM	RF	G NB	ET
Standard Scaler	0.798	0.792	0.822	0.828	0.787	0.776	0.785
Min Max	0.804	0.815	0.822	0.828	0.799	0.775	0.787
Log	0.801	0.815	0.822	0.828	0.800	0.787	0.790
Square root	0.790	0.815	0.822	0.826	0.800	0.780	0.777
yeo-jhonson	0.805	0.815	0.822	0.824	0.800	0.787	0.787

Train-Test-Val Split RMSE from Previous Project

5.2 LGBM Prediction — Magnitude (Train)

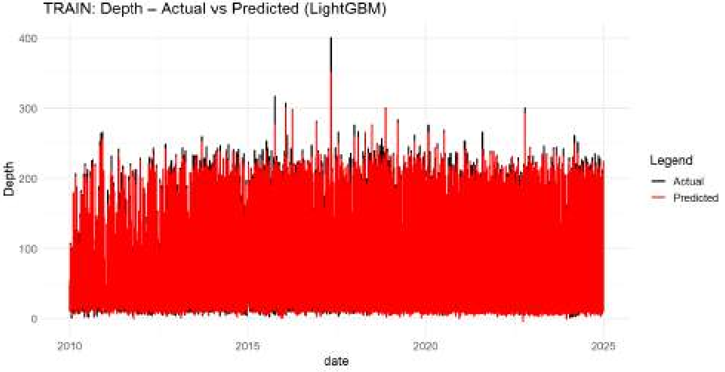


(a) Training Prediction – Magnitude (LGBM)



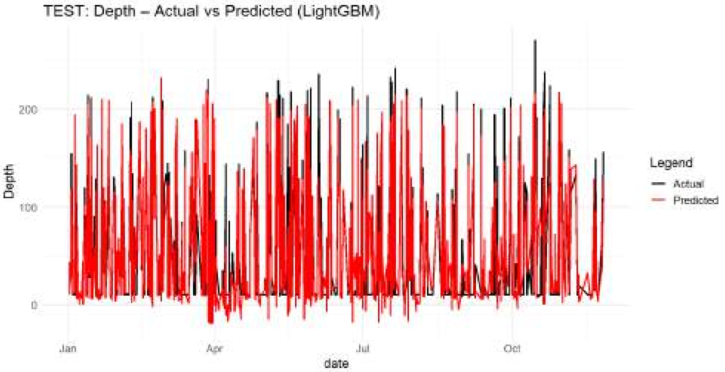
(b) Testing Prediction – Magnitude (LGBM)

5.2 LGBM Prediction — Depth (Train)



(c) Training Prediction – Depth (LGBM)

5.2 LGBM Prediction — Depth (Test)



(d) Testing Prediction – Depth (LGBM)

Conclusion

جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology

- Analyzed spatial and temporal patterns of earthquakes in the region.
- Identified major seismic hotspots and clustering of events.
- Applied ARIMA and LightGBM models for forecasting magnitude and depth.
- LightGBM demonstrated superior predictive performance over ARIMA.
- Insights can aid in risk assessment, disaster preparedness, and urban planning.
- Areas with consistently high seismic intensity should be approached with caution.
- Findings can inform safer construction practices and long-term regional planning.

Limitations

- Analysis limited to the region under study; global patterns not considered.
- Data quality issues: magnitude and depth errors can be significant in some cases.
- Models (ARIMA, LightGBM) are dependent on historical data and may not capture rare extreme events.
- Spatial covariates and environmental factors not fully incorporated in the models.
- Findings are indicative and should be interpreted with caution for decision-making.

Future Scope

- Extend analysis to a larger geographical region or global seismic data.
- Incorporate additional covariates such as soil type, tectonic stress, and fault lines.
- Explore deep learning and hybrid models for improved earthquake prediction.
- Long-term monitoring to detect changes in seismic patterns over decades.
- Use insights to inform urban planning, disaster preparedness, and construction guidelines.

Thank You

References

[1] Robert J. Hijmans. *gedata: Access Geographic Data*. R package version 0.6-7. 2025. URL: https://github.com/rspatial/geodata.

[2] Robert J. Hijmans. *terra: Spatial Data Analysis*. R package version 1.8-87. 2025. URL: https://github.com/rspatial/terra.

[3] U.S. Geological Survey. *Earthquake Depth Classification*. https://www.usgs.gov/programs/earthquake-hazards/determining-depth-earthquake. Accessed: 2025-12-07. 2024.

[4] U.S. Geological Survey. *ANSS Comprehensive Earthquake Catalog (ComCat) — earthquake search results for India / surrounding seismic zones (2010–2025)*. https://earthquake.usgs.gov/earthquakes/map/. Data accessed on 2025-12-06. 2025.

[5] Charles Kenny. *Why Do People Die in Earthquakes? The Costs, Benefits and Institutions of Disaster Risk Reduction in Developing Countries*. Policy Research Working Paper 4823. The World Bank / Global Facility for Disaster Reduction and Recovery (GFDRR), Jan. 2009. URL: https://www.gfdrr.org/sites/default/files/publication/Why%20do%20people%20die%20in%20Earthquakes.pdf.

[6] Paula Moraga. *Spatial Statistics for Data Science: Theory and Practice with R*. CRC Data Science Series. Chapman & Hall/CRC, 2023. ISBN: 9781032633510.

[7] Guolin Ke et al. "LightGBM: A Highly Efficient Gradient Boosting Decision Tree". In: *Advances in Neural Information Processing Systems 30*. 2017, pp. 3146–3154. URL: https://papers.nips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html.

[8] George E. P. Box et al. *Time Series Analysis: Forecasting and Control*. 5th. Hoboken, New Jersey: John Wiley & Sons, 2015. ISBN: 9781118675021.

[9] U.S. Geological Survey. *Determining the Depth of an Earthquake*. https://www.usgs.gov/programs/earthquake-hazards/determining-depth-earthquake. Accessed: 28 November 2025. n.d.

[10] NASA Space Place. *What Is an Earthquake?* https://spaceplace.nasa.gov/earthquakes/en/. Accessed: 27-November-2025. 2023.

[11] Encyclopaedia Britannica Editors. *Earthquake magnitude*. *Encyclopaedia Britannica*. Accessed: 28 November 2025. 2024. URL: https://www.britannica.com/science/earthquake-geology/Earthquake-magnitude.

[12] Arthur Getis and J. Keith Ord. "The Analysis of Spatial Association by Use of Distance Statistics". In: *Geographical Analysis* 24.3 (1992), pp. 189–206. DOI: 10.1111/j.1538-4632.1992.tb00261.x.

[13] Jesper Møller, Anne R. Syversveen, and Rasmus P. Waagepetersen. "Log Gaussian Cox processes". In: *Scandinavian Journal of Statistics* 25.3 (1998), pp. 451–482. DOI: 10.1111/1467-9469.00115. URL: https://onlinelibrary.wiley.com/doi/10.1111/1467-9469.00115.

[14] Håvard Rue, Sara Martino, and Nicolas Chopin. "Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations". In: *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71.2 (2009), pp. 319–392. DOI: 10.1111/j.1467-9868.2008.00700.x. URL: https://onlinelibrary.wiley.com/doi/10.1111/j.1467-9868.2008.00700.x.