



# Point Pattern Data Analysis

Spatial–Temporal Point Pattern Analysis of Earthquakes in India and  
Adjacent Regions (2010–2025)  
King Abdullah University of Science and Technology  
CEMSE Division – PhD Statistics

Mohd Saqib Ansari  
[mohammad.ansari@kaust.edu.sa](mailto:mohammad.ansari@kaust.edu.sa)

8 December, 2025

## Contents

<b>1</b>	<b>Abstract</b>	<b>2</b>
<b>2</b>	<b>Introduction</b>	<b>2</b>
2.1	Objective . . . . .	2
<b>3</b>	<b>About the Data</b>	<b>3</b>
<b>4</b>	<b>Methodology</b>	<b>3</b>
4.1	Exploratory Data Analysis and Preprocessing . . . . .	3
4.2	Temporal Analysis . . . . .	4
4.3	Spatial Analysis . . . . .	4
4.3.1	Choropleth Analysis . . . . .	4
4.3.2	Hotspot Analysis . . . . .	4
4.4	Point Pattern Analysis . . . . .	7
4.4.1	Quadrant Count Analysis . . . . .	7
4.4.2	Ripley's K-function Analysis . . . . .	7
4.4.3	Log-Gaussian Cox Process (LGCP) Modeling Using INLA . . . . .	7
4.5	Modeling the earthquake . . . . .	9
4.5.1	ARIMA Models . . . . .	9
4.5.2	LightGBM (LGBM) . . . . .	10
4.6	Limitations and Future Scope . . . . .	11

# 1 Abstract

*In this study, we analyze the spatio-temporal patterns of earthquakes in India and its adjacent regions. We employed choropleth maps for visualization and performed hotspot analysis using local statistics to identify regions of high seismic activity. Quadrant count analysis and Ripley's K-function were used to investigate spatial clustering and point pattern characteristics. Furthermore, we modeled earthquake magnitude and depth using ARIMA and LightGBM models to explore temporal trends and predictive patterns. The results provide insights into the spatial distribution and temporal dynamics of earthquakes in the region, highlighting areas of elevated seismic risk and demonstrating the utility of combining geospatial analysis with predictive modeling.*

## 2 Introduction

Earthquakes are natural phenomena involving sudden ground shaking caused by movement in Earth's outermost layer. Earth has four main layers: a solid crust, a hot and nearly solid mantle, a liquid outer core, and a solid inner core. The crust and upper mantle form the lithosphere, which is divided into tectonic plates that drift slowly over the viscous mantle. Their movement creates stress, and when this stress exceeds the strength of rocks, it results in fractures called faults. Earthquakes originate at these faults, and the point where they begin is called the epicentre. Their time, location, and intensity are measured using a seismometer<sup>1</sup>.

Point pattern analysis provides a robust statistical framework for studying the spatial and spatiotemporal distribution of earthquake epicentres. Since earthquakes cluster along plate boundaries and fault zones, this analysis helps identify seismic hotspots and understand clustering behaviour, supporting disaster risk assessment and urban planning.

Earthquakes also cause significant loss of life and property, especially in densely populated or poorly constructed regions.<sup>2</sup>

### 2.1 Objective

The study aims to analyze earthquake patterns using geostatistical methods learned in *Spatial Data Science with R (STAT2300)*. Specifically, it focuses on:

- Exploring spatial and temporal trends of earthquakes through EDA.
- Investigating clustering, randomness, or regularity of epicenters using quadrant count analysis, Ripley's K-function, and Kernel Density Estimation.
- Examining spatio-temporal behavior to detect aftershocks and changes in seismic activity.
- Identifying seismic hotspots and high-risk zones via intensity mapping and spatial visualization.
- Interpreting findings in the context of regional tectonics for hazard assessment and risk mitigation.

---

<sup>1</sup>NASA Space Place. 2023.

<sup>2</sup>Charles Kenny. 2009.

### 3 About the Data

The dataset for this study was obtained from the United States Geological Survey (USGS) Earthquake Catalog<sup>3</sup>, covering India and surrounding seismic zones (latitudes 0.40°N–43.98°N, longitudes 63.99°E–99.14°E) for 2010–2025. It contains 14,336 earthquake events, each described by several attributes, some of which are:

- **time**: Timestamp (UTC) of the earthquake occurrence.
- **latitude**: North–south coordinate of the epicenter.
- **longitude**: East–west coordinate of the epicenter.
- **depth**: Depth (km) of the earthquake origin.
- **mag**: Earthquake magnitude, representing released energy.

With spatial (latitude, longitude), temporal (time), and physical (magnitude, depth) attributes, this dataset is suitable for spatial and spatio-temporal point pattern analysis, providing insights into seismic clustering, temporal distribution, and fault zone activity.

In addition, for exploratory analysis, global elevation<sup>4</sup> and climate variables (mean annual temperature and precipitation)<sup>5</sup> were extracted at earthquake epicenters to explore potential correlations with magnitude and depth.

### 4 Methodology

The methodological framework of this study follows a structured workflow. It begins with Exploratory Data Analysis (EDA), using descriptive statistics and visualizations to understand the distribution, variability, and characteristics of earthquake data. Temporal analysis identifies trends, seasonal patterns, and clustering of events over time, while spatial analysis explores the geographic distribution of epicenters and detects potential hotspot regions. Building on this, a spatio-temporal analysis examines how seismic activity evolves across both space and time. Finally, advanced modeling techniques, including ARIMA and LightGBM, are applied to forecast earthquake magnitude and depth, and point pattern analysis methods, such as Kernel Density Estimation and Ripley’s K-function, assess the degree of clustering or randomness in earthquake occurrences.

#### 4.1 Exploratory Data Analysis and Preprocessing

In the Exploratory Data Analysis (EDA), we observed that the dataset consists of 14,336 observations with 22 variables, and the summary of missing values is presented in Fig 1a. For point pattern analysis, the spatial accuracy is essential, and fortunately, the location variables (latitude and longitude) are complete without missing values. Regarding magnitude reliability, approximately 95%, 60%, and 12% of the events have errors exceeding 0.05, 0.10, and 0.20 (error range 0 to 1), respectively, indicating that while minor errors are prevalent, large magnitude errors are relatively limited. Out of all recorded earthquakes, 10,595 were shallow, 3,736 were intermediate, and only 5 were deep events<sup>6</sup>.

---

<sup>3</sup>U.S. Geological Survey. 2025.

<sup>4</sup>Robert J. Hijmans. 2025.

<sup>5</sup>Robert J. Hijmans. 2025.

<sup>6</sup>U.S. Geological Survey. 2024.

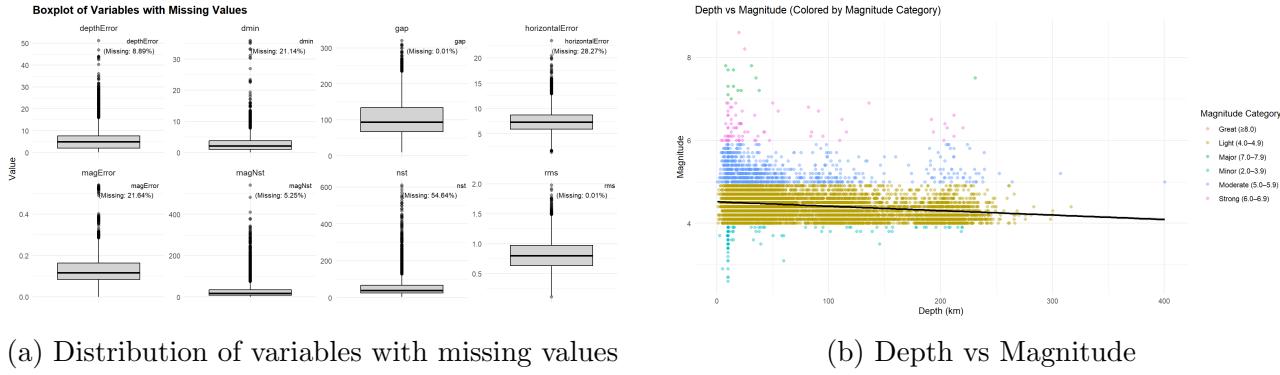


Figure 1: Comparison of Missing Values and Depth-Magnitude Relationship

7

## 4.2 Temporal Analysis

In the temporal analysis, Fig. 2a illustrates the annual earthquake distribution, where the years 2012 and 2015 recorded the highest number of earthquake events, while 2011 reported the lowest. The overall monthly earthquake pattern in Fig. 2b shows that April had the highest frequency (1662 events) between 2010 and 2025, whereas February had the lowest (945 events). The seasonality pattern across years, shown in Fig. 7b, does not reveal any consistent or recurring monthly trends, indicating irregular peaks across different years. The distribution across weekdays in Fig. 2c also does not exhibit any specific trend. However, the overall hourly trend in Fig. 2d clearly suggests that the highest number of earthquakes occurred around 01:00 hour, while the lowest counts were recorded around 06:00 hour. Additionally, Fig. 2e presents a month–weekday heatmap, highlighting subtle variations in earthquake occurrences across different days and months.

## 4.3 Spatial Analysis

In this section, we performed two main types of spatial analysis: *choropleth analysis* and *hotspot analysis*.

### 4.3.1 Choropleth Analysis

In the choropleth analysis, we visualized the spatial distribution of earthquake locations along with their magnitude ranges (Figure 3a). Similarly, we plotted the spatial distribution of earthquake depths (Figure 3b). It helps to visualise the variation in earthquake depths across the region and identify areas with consistently shallow or deep events.

### 4.3.2 Hotspot Analysis

Hotspot analysis is used to identify statistically significant spatial clusters of earthquake activity. The Getis–Ord  $G_i^*$ <sup>8</sup> statistic detects locations where high (hotspots) or low (coldspots) values of magnitude or combined magnitude–depth occur more frequently than would be expected under spatial randomness. These clusters help reveal the spatial concentration and intensity of seismic events, as illustrated in Figure 4.

The Getis–Ord  $G_i^*$  statistic for a location  $i$  is calculated as:

<sup>8</sup>Arthur Getis and J. Keith Ord. *Geographical Analysis*. 1992.

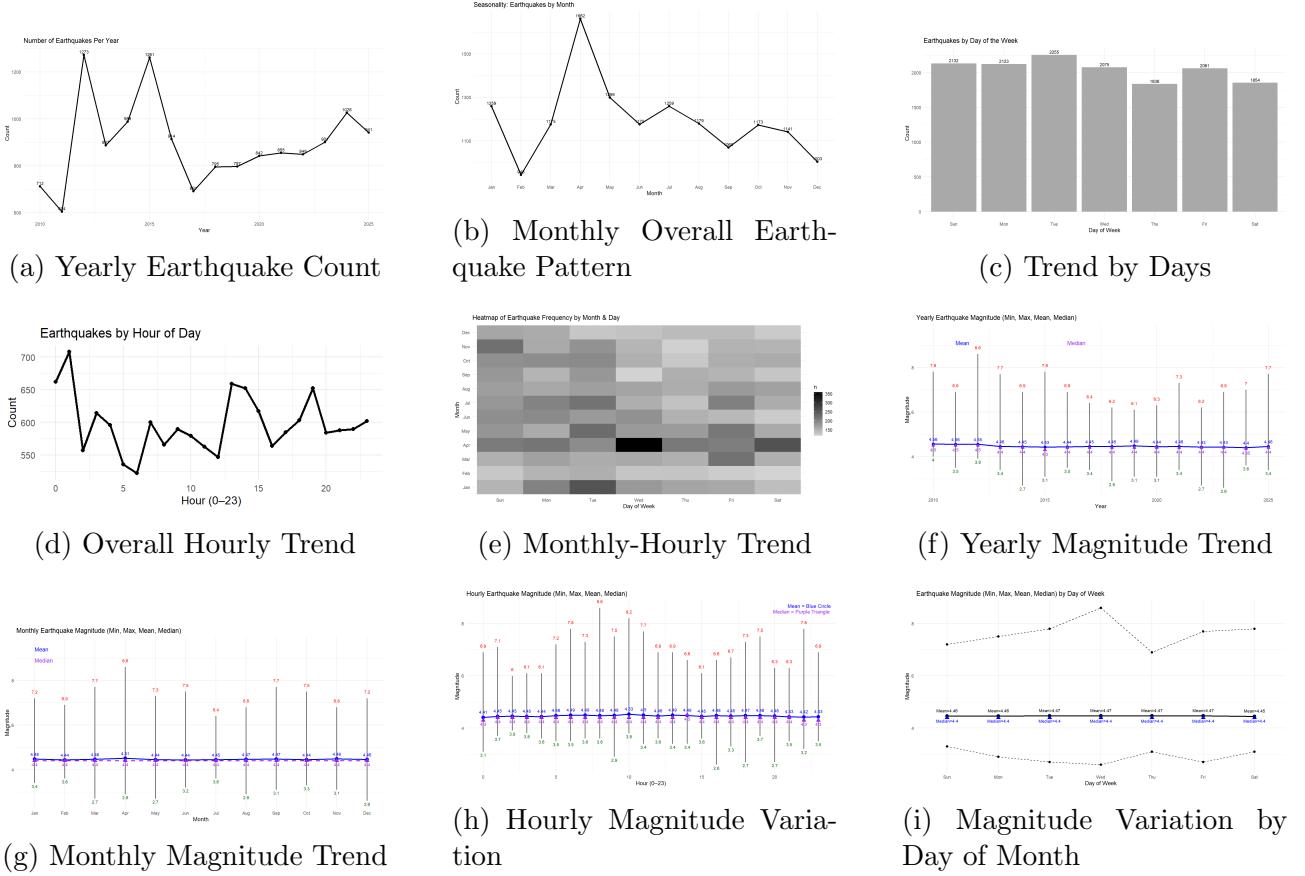


Figure 2: Comprehensive  $3 \times 3$  Visualization of Earthquake Temporal and Magnitude Patterns Across Different Scales

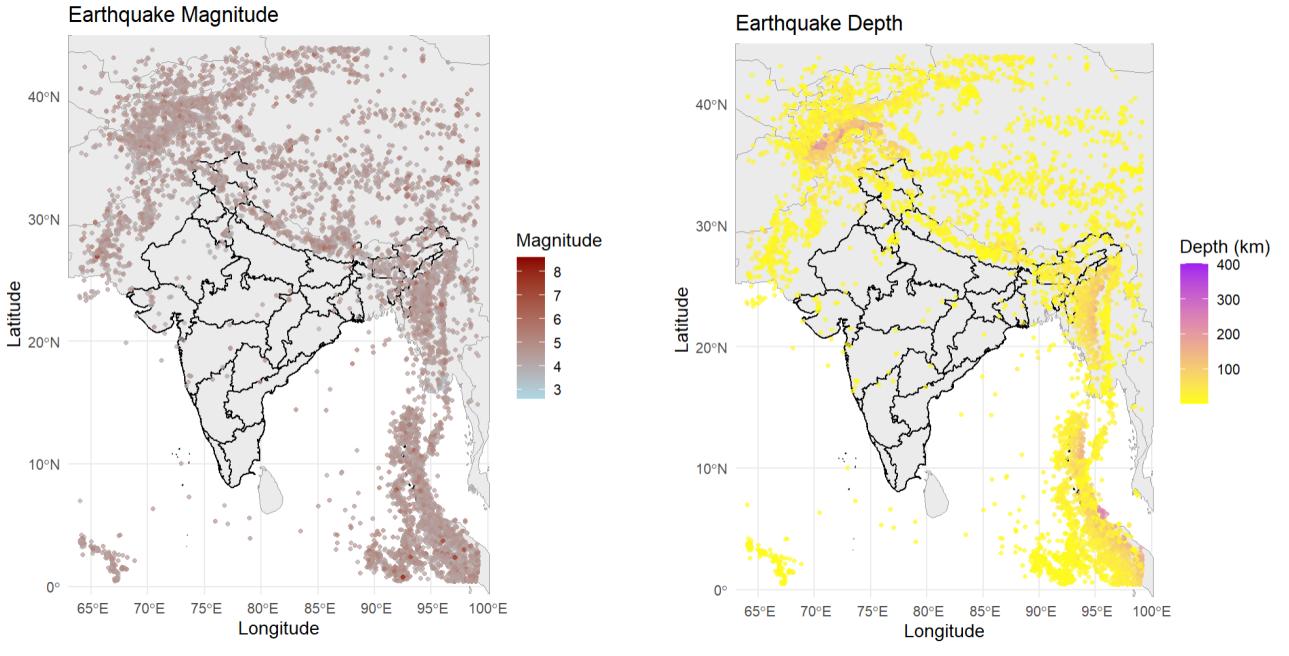


Figure 3: Spatial visualization of earthquake magnitude and depth in the region.

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

where:

- $x_j$  is the attribute value at location  $j$  (e.g., magnitude),
- $w_{ij}$  is the spatial weight between locations  $i$  and  $j$ ,
- $\bar{X}$  is the mean of all  $x_j$ ,
- $n$  is the total number of locations.

The  $G_i^*$  statistic is a z-score and theoretically ranges from  $-\infty$  to  $+\infty$ , with large positive values indicating significant hotspots, and large negative values indicating significant coldspots. Values near zero indicate no significant clustering.

In our analysis:

- The  $Gi^*$  statistic for magnitude-based hotspot analysis ranges from  $-6.86$  to  $7.06$ , indicating the presence of both statistically significant hotspots (large positive  $G_i^*$ ) and coldspots (large negative  $G_i^*$ ), while the median value close to zero suggests that most locations do not exhibit strong clustering.

- For the combined magnitude-depth  $Gi^*$  analysis, the observed range ( $-5.83$  to  $5.47$ ) and a higher median value ( $0.52$ ) indicate comparatively stronger positive clustering when depth information is incorporated. This implies that combining depth with magnitude enhances the detection of meaningful seismic hotspots, particularly those associated with deeper high-magnitude events.

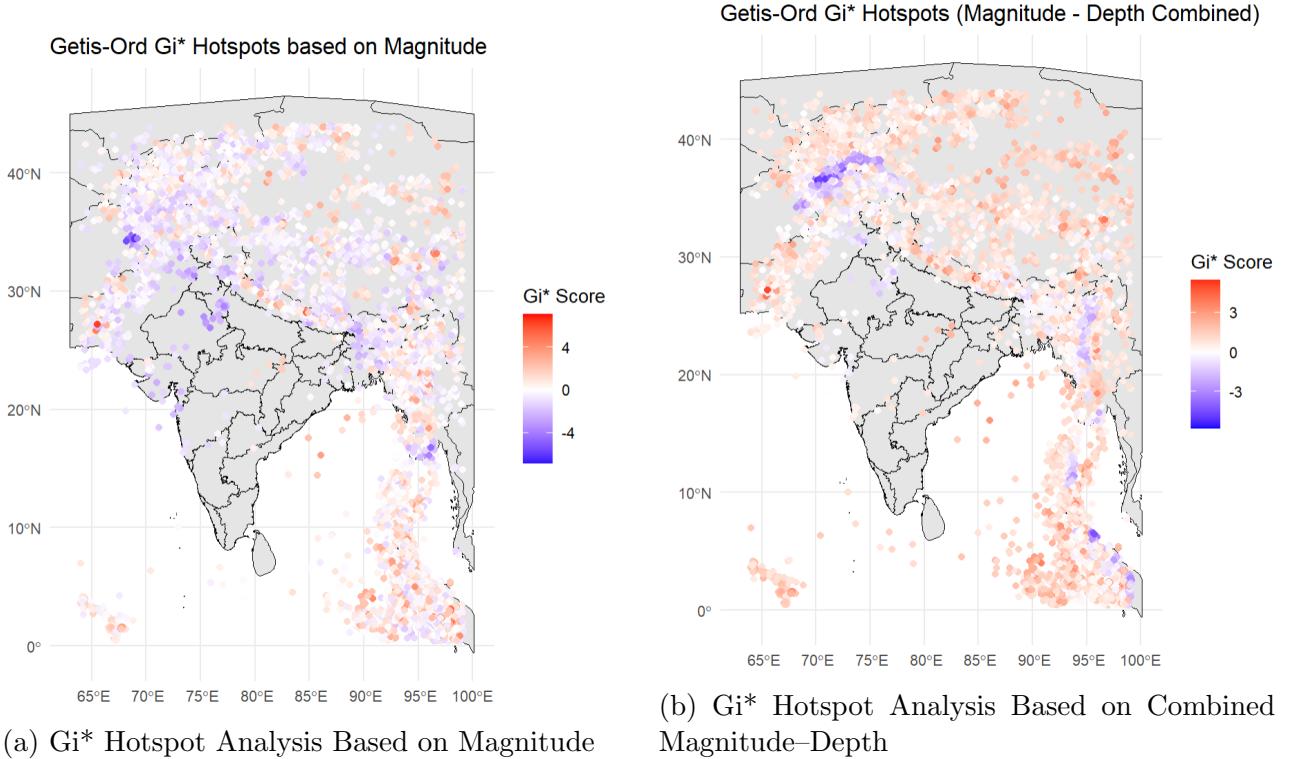


Figure 4: Spatial Hotspot Analysis using Getis-Ord  $Gi^*$

## 4.4 Point Pattern Analysis

### 4.4.1 Quadrant Count Analysis

We performed a quadrant (quadrat) analysis<sup>9</sup> to examine the spatial distribution of earthquake epicenters in the study region. The region was divided into different numbers of quadrants ( $n = 3, 4, 5, 10, 15$ ) and a Chi-square test for Complete Spatial Randomness (CSR) was performed for each case. All tests returned  $p$ -values of 0, strongly rejecting CSR and confirming that the earthquake points are **clustered** rather than randomly distributed.

The quadrant counts for  $n = 5$  are illustrated in Figure 5a.

### 4.4.2 Ripley's K-function Analysis

Ripley's K-function<sup>10</sup> is used to assess **spatial clustering or dispersion of points** across multiple distance scales compared to complete spatial randomness (CSR).

In Figure 5b, the observed inhomogeneous K-function ( $K_{\text{obs}}^{\text{inhom}}$ ) remains close to zero, while the simulation envelope reaches very high values, indicating high variability and caution in interpreting departures.

For depth (Figure 5d), the homogeneous Poisson K-function ( $K_f^{\text{pois}}$ ) lies below the isotropic K-function ( $K_f^{\text{iso}}$ ), showing significant small-scale clustering, while the transformed K-function ( $K_f^{\text{trans}}$ ) is slightly above  $K_f^{\text{iso}}$ , indicating moderate dispersion.

For magnitude (Figure 5c), the K-function increases roughly linearly with distance, suggesting a more uniform spatial distribution, unlike depth which shows clustering at shorter distances.

Overall, the analysis indicates that **earthquake depths are strongly clustered**, while **magnitudes are more evenly distributed**.

### 4.4.3 Log-Gaussian Cox Process (LGCP) Modeling Using INLA

The spatial distribution of earthquake events was modeled using a **Log-Gaussian Cox Process (LGCP)**<sup>11</sup>, where counts in each pixel are assumed to arise from a Poisson process with a spatially varying intensity:

$$y_i \sim \text{Poisson}(\mu_i), \quad \log(\mu_i) = \log(A_i) + \beta_0 + s(\mathbf{s}_i), \quad (2)$$

where  $A_i$  is the pixel area (offset),  $\beta_0$  is the global intercept, and  $s(\mathbf{s}_i)$  is a latent Gaussian random field representing spatial correlation. The latent field was modeled using the Stochastic Partial Differential Equation (SPDE) approach and inference was performed via **Integrated Nested Laplace Approximation (INLA)**<sup>12</sup>, which provides efficient Bayesian estimation for spatial models.

A grid search over candidate PC priors for the SPDE parameters (range and marginal standard deviation) was performed, and the best prior was selected based on the **Deviance Weighted Information Criterion (DWIC)**:

$$\text{Range} = 100 \text{ km}, \quad \sigma = 0.5$$

---

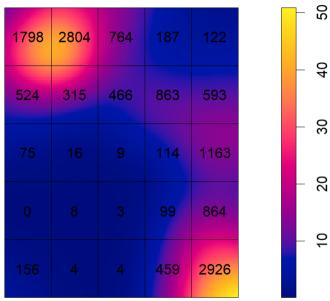
<sup>9</sup>Paula Moraga. *Spatial Statistics for Data Science: Theory and Practice with R*. 2023.

<sup>10</sup>Adrian Baddeley, Jesper Møller, and Rasmus Waagepetersen. *Statistica Neerlandica*. 2000.

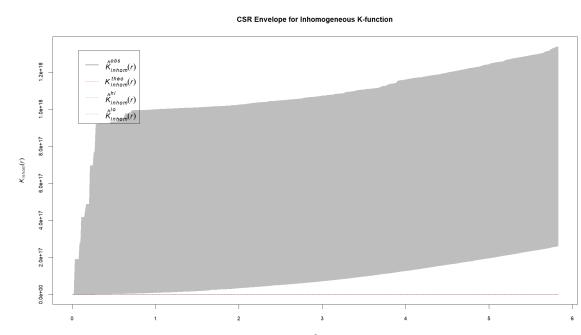
<sup>11</sup>Jesper Møller, Anne R. Syversveen, and Rasmus P. Waagepetersen. *Scandinavian Journal of Statistics*. 1998.

<sup>12</sup>Håvard Rue, Sara Martino, and Nicolas Chopin. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2009.

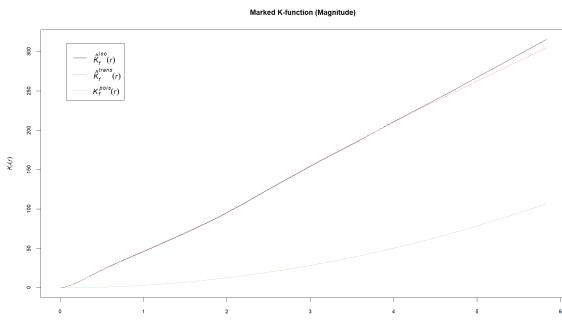
Point Density with Quadrat Overlay



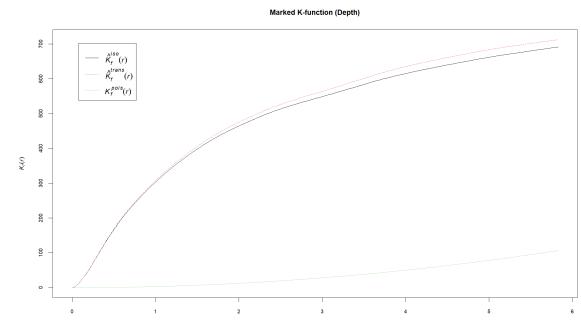
(a) Quadrant Analysis: Earthquake Counts per Grid (5x5)



(b) CSR Envelope Kinhom



(c) Marked K-function (Magnitude)



(d) Marked K-function (Depth)

Figure 5: Point Density with Quad Overlay Spatial clustering analysis using Ripley's K, Marked K functions

Using INLA with this prior, we obtained posterior mean and 95% credible intervals for each pixel, producing spatial intensity maps that highlight earthquake hotspots and quantify uncertainty.

The resulting intensity maps are shown in Figure 6, including the posterior mean, lower, and upper 95% credible interval bounds.

Table 1: Key parameter estimates from the INLA–SPDE LGCP earthquake occurrence model

Parameter	Mean	95% CI (Lower)	95% CI (Upper)
<b>Spatial Hyperparameters</b>			
Range ( $\phi$ )	234,103	214,366	250,816
Stdev of spatial field ( $\sigma$ )	2.85	2.64	3.03
<b>Fixed Effect</b>			
Intercept ( $\beta_0$ )	-10.20	-10.62	-9.78
<b>Model Fit Metrics</b>			
WAIC	691,389.5	—	—
DIC	26,079.96	—	—
Marginal log-likelihood	-13,367.81	—	—

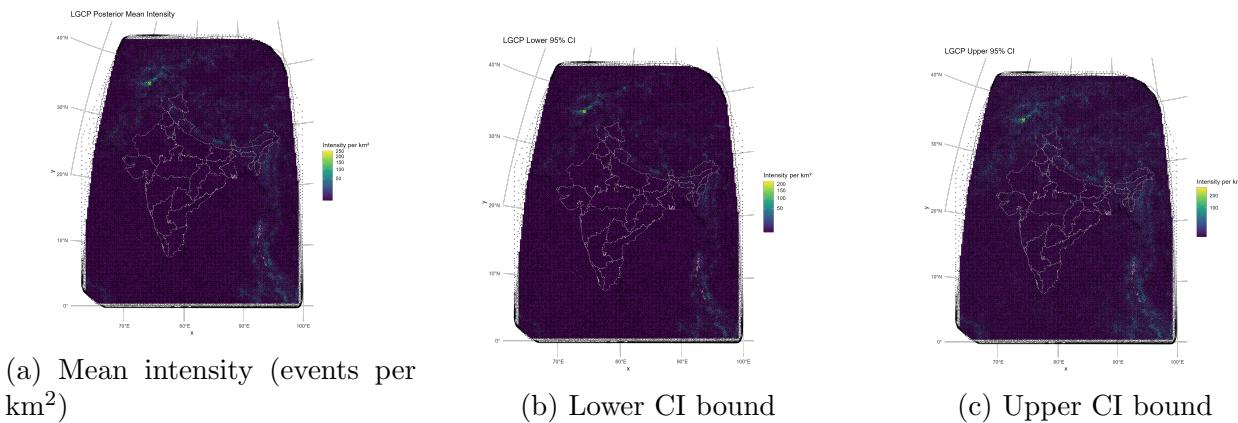


Figure 6: Spatial intensity estimates (per km<sup>2</sup>) from the Log-Gaussian Cox Process (LGCP) fitted using INLA, showing the posterior mean and 95% credible interval bounds.

## 4.5 Modeling the earthquake

A model can be chosen based on the specific requirements. Before modeling, we examined the correlation of earthquake magnitude and depth with environmental and climatic variables such as elevation, mean temperature, and rainfall, but all correlations were below 0.1, as shown in 7a.

### 4.5.1 ARIMA Models

To model the temporal behaviour of earthquake *magnitude* and *depth*, models with covariates (latitude and longitude) were fitted using the training period 2010–2024. The year 2025 was kept aside for out-of-sample testing.

Model selection was based on AIC/BIC and log-likelihood comparison across several ARIMA <sup>13</sup> configurations. The best-performing models were:

<sup>13</sup>George E. P. Box et al. *Time Series Analysis: Forecasting and Control*. 2015.

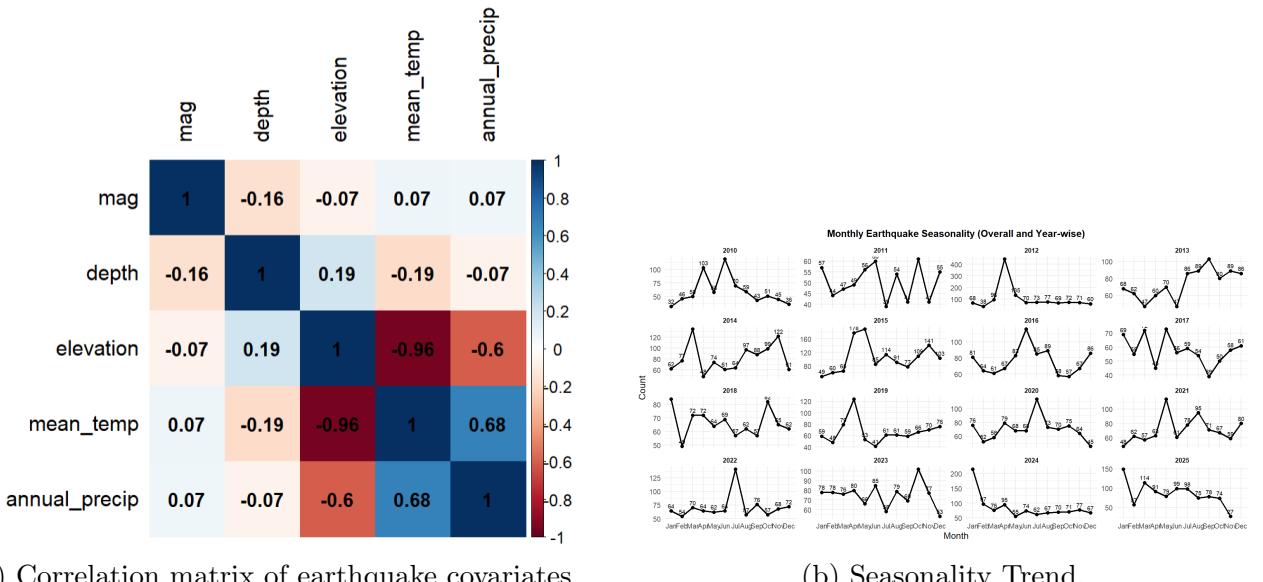


Figure 7: Side-by-side visualization of earthquake covariate correlations and seasonal patterns.

- ARIMA(5,1,0) for **Magnitude** with covariates (**lat**, **lon**)
- ARIMA(5,1,1) for **Depth** with covariates (**lat**, **lon**)

Table 2 summarises the model fit statistics and the performance on both training and testing data.

Table 2: Summary of ARIMA model performance for earthquake magnitude and depth

Model	Log-Likelihood	AIC	Train RMSE	Test RMSE
ARIMA(5,1,0) – Magnitude	-6754.17	13524.33	0.401	0.646
ARIMA(5,1,1) – Depth	-72905.70	145829.40	55.93	56.90

The ARIMA(5,1,0) model shows good predictive capability for magnitude, with moderate increase in RMSE on the test set. Depth prediction is more challenging due to large variance inherent in the data, as reflected in higher RMSE values; however, the ARIMA(5,1,1) model provides the best fit among the tested configurations.

#### 4.5.2 LightGBM (LGBM)

To complement the ARIMA models, a gradient boosting approach using LightGBM (LGBM)<sup>14</sup> was implemented for predicting earthquake magnitude and depth. LGBM is a tree-based boosting algorithm known for high predictive accuracy and fast training. Hyperparameter tuning was performed using grid search, and the best configuration for each model was identified as follows:

- **Magnitude Model:** 63 leaves, learning rate = 0.1, min\_data\_in\_leaf = 20, objective = regression, metric = rmse, 1000 boosting rounds
- **Depth Model:** 127 leaves, learning rate = 0.1, min\_data\_in\_leaf = 20, objective = regression, metric = rmse, 1000 boosting rounds

<sup>14</sup>Guolin Ke et al. 2017.

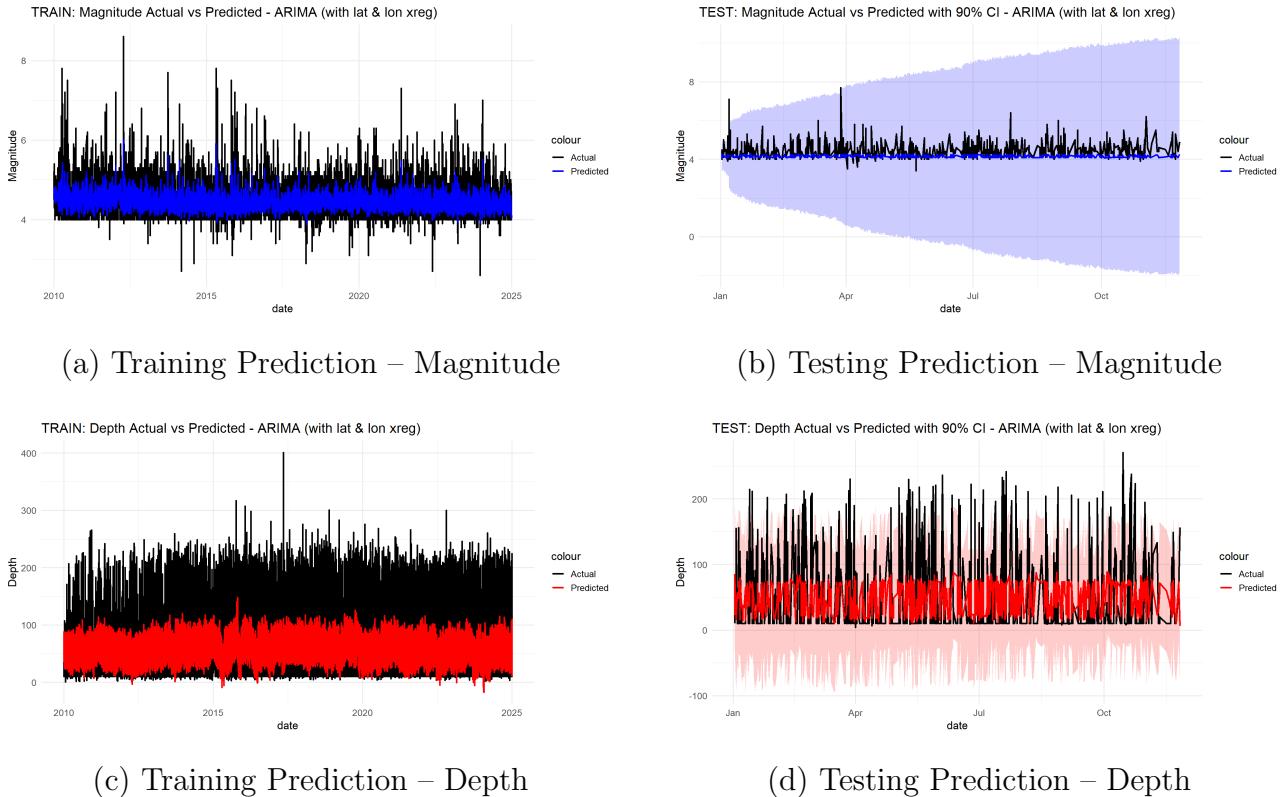


Figure 8: ARIMA model predictions for magnitude and depth on training (2010–2024) and testing (2025) datasets.

Table 3 summarises the model performance for both magnitude and depth using training (2010–2024) and testing (2025) datasets. LGBM demonstrates strong predictive performance, especially for depth, with significantly lower RMSE compared to ARIMA models.

Table 3: LGBM performance for magnitude and depth prediction (corrected)

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

## 4.6 Limitations and Future Scope

This study is limited to earthquake data from India and adjacent regions. In several cases, the USGS catalog indicates that magnitude estimates may contain errors of up to 50%, which introduces uncertainty into both the exploratory analysis and the modeling results. Since the quality of any statistical model depends heavily on the quality of the input data, improving the accuracy of magnitude, depth, and station-based measurements would significantly enhance model reliability.

Future work may expand the spatial domain, incorporate higher-resolution covariates, and analyze a longer temporal period to better capture changes in seismic activity. A more comprehensive comparison of seismic hotspots across multiple decades may also provide deeper insights into long-term tectonic behavior and evolving seismic risk.

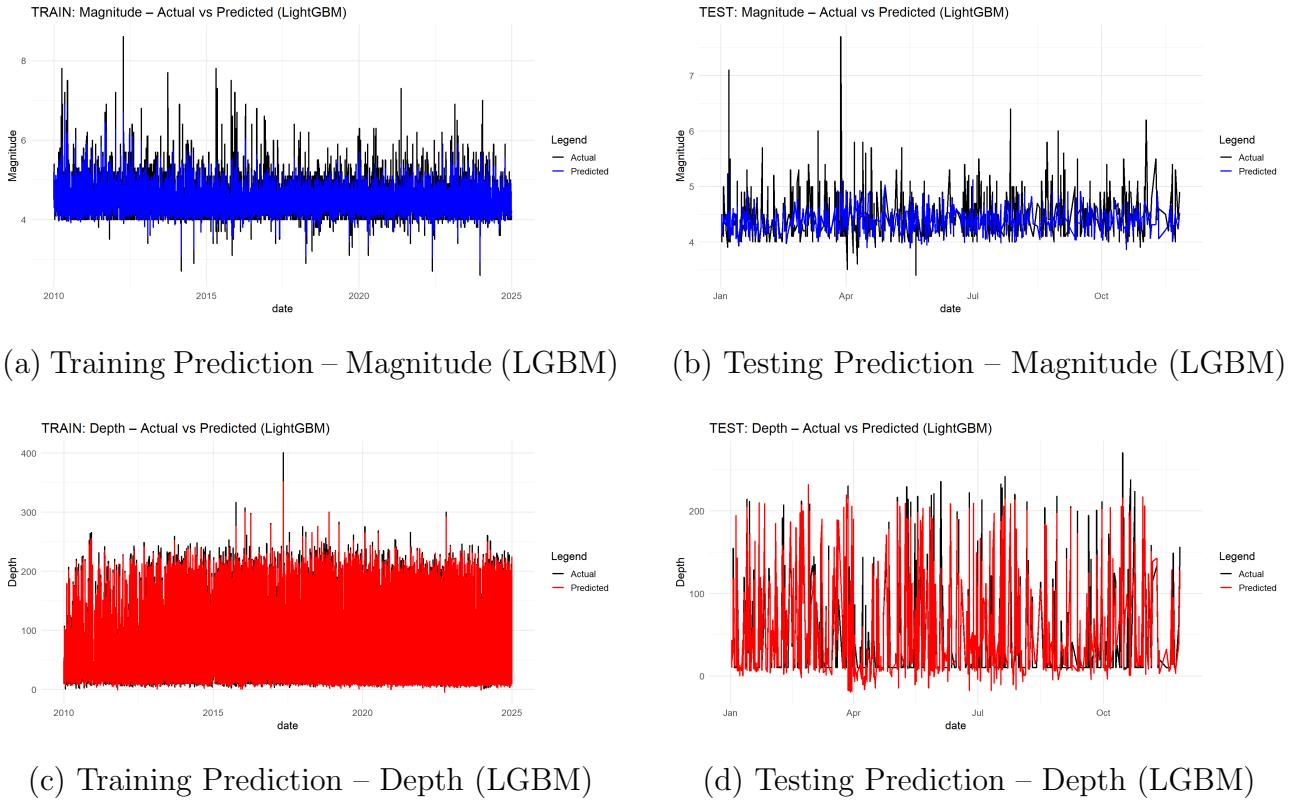


Figure 9: LGBM model predictions for magnitude and depth on training (2010–2024) and testing (2025) datasets.

## 4.7 Conclusion

This study examined the spatial and temporal characteristics of earthquakes in the region, identified major seismic hotspots, and analyzed variations in magnitude and depth. Time-series models, including ARIMA and LightGBM, were applied to forecast earthquake magnitude and depth, with LightGBM showing superior predictive performance.

The findings contribute to a better understanding of seismic activity patterns and may support risk assessment, urban planning, and disaster preparedness. Areas exhibiting consistently high seismic intensity and clustering should be approached with caution, and the insights derived from this analysis can inform safer construction practices and long-term regional planning.

## References

- [1] NASA Space Place. “What Is an Earthquake?”
- [2] 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>.
- [3] U.S. Geological Survey. “ANSS Comprehensive Earthquake Catalog (ComCat) — earthquake search results for India / surrounding seismic zones (2010–2025)”.
- [4] Robert J. Hijmans. *terra: Spatial Data Analysis*. R package version 1.8-87, 2025. URL: <https://github.com/rspatial/terra>.

- [5] Robert J. Hijmans. *geodata: Access Geographic Data*. R package version 0.6-7, 2025. URL: <https://github.com/rspatial/geodata>.
- [6] U.S. Geological Survey. “Earthquake Depth Classification”.
- [7] Encyclopaedia Britannica Editors. “Earthquake magnitude”.
- [8] Arthur Getis and J. Keith Ord. “The Analysis of Spatial Association by Use of Distance Statistics”. *Geographical Analysis* 24, pp. 189–206, 1992. DOI: [10.1111/j.1538-4632.1992.tb00261.x](https://doi.org/10.1111/j.1538-4632.1992.tb00261.x).
- [9] Paula Moraga. *Spatial Statistics for Data Science: Theory and Practice with R*. CRC Data Science Series. Chapman & Hall/CRC, 2023. ISBN: 9781032633510.
- [10] Adrian Baddeley, Jesper Møller, and Rasmus Waagepetersen. “Non- and semi-parametric estimation of spatial interaction in inhomogeneous point patterns”. *Statistica Neerlandica* 54, pp. 329–350, 2000. DOI: [10.1111/1467-9574.00154](https://doi.org/10.1111/1467-9574.00154).
- [11] Jesper Møller, Anne R. Syversveen, and Rasmus P. Waagepetersen. “Log Gaussian Cox processes”. *Scandinavian Journal of Statistics* 25, pp. 451–482, 1998. DOI: [10.1111/1467-9469.00115](https://doi.org/10.1111/1467-9469.00115). URL: <https://onlinelibrary.wiley.com/doi/10.1111/1467-9469.00115>.
- [12] Håvard Rue, Sara Martino, and Nicolas Chopin. “Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations”. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 71, pp. 319–392, 2009. DOI: [10.1111/j.1467-9868.2008.00700.x](https://doi.org/10.1111/j.1467-9868.2008.00700.x). URL: <https://onlinelibrary.wiley.com/doi/10.1111/j.1467-9868.2008.00700.x>.
- [13] George E. P. Box et al. *Time Series Analysis: Forecasting and Control*. 5th. Hoboken, New Jersey: John Wiley & Sons, 2015. ISBN: 9781118675021.
- [14] 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>.