Earthquake Analysis And Prediction: Comparing ETAS And USGS Earthquake Data

CS 4980 Capstone Report, 2025

Vishal Kamalakrishnan
Computer Science
The University of Virginia
School of Engineering and Applied Science
Charlottesville, Virginia USA
cjq2cw@virginia.edu

ABSTRACT

Earthquakes pose a significant threat to human lives and infrastructure, providing a need for accurate earthquake forecasting models to help mitigate risk. My capstone research project presents a comprehensive analysis comparing synthetic earthquake simulations generated by the Epidemic-Type Aftershock Sequence (ETAS) model with historical earthquake records from the United States Geological Survey (USGS) spanning 1960 to 2023. I found discrepancies between modeled and observed data by analyzing patterns and energy related to significant earthquakes (magnitude ≥ 6.5). Notably, ETAS simulated data shows higher average energy levels before and after major earthquakes, while USGS data indicates a considerable energy spike just before the main event, which is not observed in ETAS data. Furthermore, cell-grid-based analyses reveal that ETAS tends to centralize energy within specific regions post-event, in contrast to the more dispersed patterns observed in USGS data. These findings highlight limitations in current forecasting models and suggest avenues for improvement through developing more sophisticated spatial grid-based models and incorporating detailed analysis of fault lines.

1. INTRODUCTION

Earthquakes remain among the most destructive and unpredictable natural disasters, posing significant risks to human life and infrastructure, especially in seismically active regions like California. Early and accurate earthquake forecasting remains a scientific and engineering challenge and is critical for minimizing casualties, reducing economic losses, and enabling effective emergency response. The ongoing evolution of seismic hazard models, such as the Epidemic-Type Aftershock Sequence (ETAS) model, offers new opportunities to understand and anticipate earthquake activity.

My project investigates earthquake patterns in California by comparing synthetic ETAS simulations with historical earthquake records from the United States Geological Survey (USGS) between 1960 and 2023. By focusing on spatial and temporal patterns, energy release, and event clustering before and after major earthquakes (magnitude ≥6.5), my project aims to identify discrepancies between modeled and observed earthquake data. The project seeks to uncover limitations in current forecasting approaches and lay the groundwork for more accurate seismic risk assessment. Additionally, I propose several areas for improvement regarding the ETAS model, explain how the data for ETAS is generated, and suggest optimal parameters to align with the recorded seismic data from USGS.

2. RELATED WORKS

Earthquake simulation and forecasting have been extensively studied using both empirical data and statistical models. The USGS Earthquake Catalog has long been a primary source for analyzing seismic patterns. With the rise of deep learning, several models have been developed that employ deep neural networks and other machine learning techniques to predict earthquakes [1]. The ETAS model captures the temporal and spatial distribution of aftershocks following a mainshock [2]. It considers the influence of past earthquakes on future seismicity, providing valuable insights for hazard assessment.

In a previous project titled "Surrogate Modeling for Efficient Earthquake Prediction," I explored the creation of a surrogate model for ETAS data, but I was unable to create a model with sufficiently high accuracy. Prior to this capstone, in a project titled “Earthquake Historical Data Analysis” [4], I analyzed USGS data to detect trends within earthquake magnitudes, multiplicity, and energy release. The project highlighted that while ETAS effectively models the overall pattern of earthquakes, it fails to accurately model USGS data in both earthquake locations and multiplicity. This motivated the exploration of a deeper approach to better understand the discrepancies between ETAS data and USGS data.

3. PROJECT DESIGN

I designed the capstone research project through several analysis notebooks and pipelines to efficiently compare, visualize, and forecast seismic activity in California. The system integrates advanced data processing, statistical modeling, and interactive visualization to deliver insights for earthquake forecasting. The end deliverable was a website that showcased the findings of the research.

The analysis was implemented using Python Jupyter Notebooks utilizing various statistical and machine learning packages. The code is shared publicly on GitHub and is well documented. Each part of the analysis features interactive visualizations built with Plotly to animate earthquake events and show energy distributions across time and space. This allows users to easily observe earthquake patterns and discrepancies between simulated and real data.

The first step in the design process focuses on data transformation and cleaning the records from both the USGS catalog. I filtered records to include events of magnitude > 3 within California from 1960 to 2023. I also ensured that key features such as date, location, and magnitude were standardized for consistency and calculated seismic energy using established formulas.

The second step involved identifying and implementing the several analysis components. This included creating charts and spreadsheets of analyzed ETAS data and USGS data to identify any discrepancies. I then enhanced the data further through feature engineering, temporal binning (e.g., two-week intervals), and spatial grid partitioning through latitude and longitude. Finally, I conducted a cell-grid-based analysis on ETAS and USGS data which incorporated a custom Long Short-Term Memory (LSTM) model that was built using TensorFlow.

The front end of the application was a Python Dash web application that served to display the research works. This framework was chosen due to its native compatibility with Plotly and the ability to visualize animated charts on the web. Additionally, the website also includes some notable observations and key insights derived from the visualizations. You can access the website and explore it publicly at the following link: https://capstone-2024-website.onrender.com/.

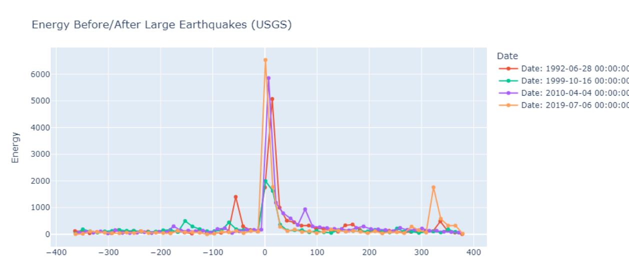

Figure 1: Energy Before/After Large Earthquakes (USGS)

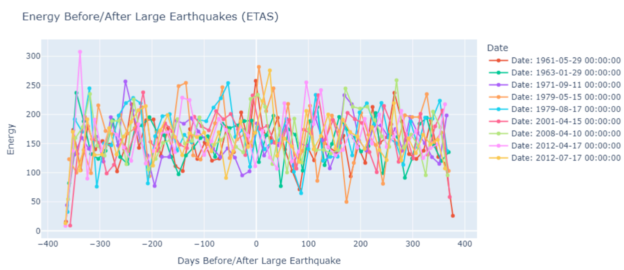

Figure 2: Energy Before/After Large Earthquakes (ETAS)

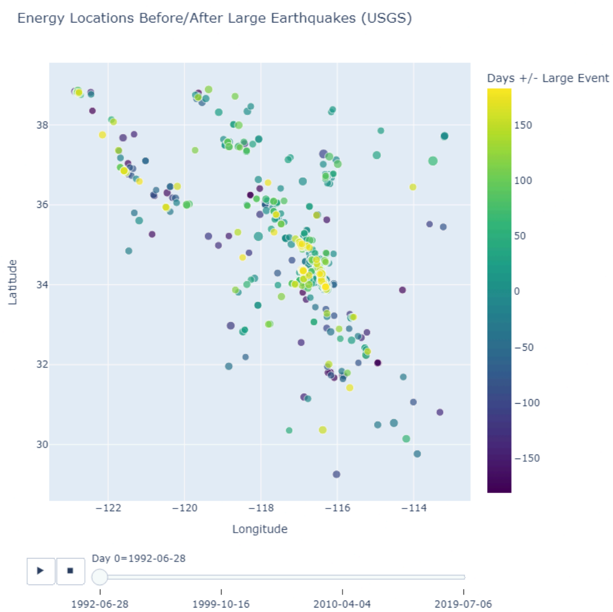

Figure 3: Energy Locations Before/After Large Earthquakes (USGS)

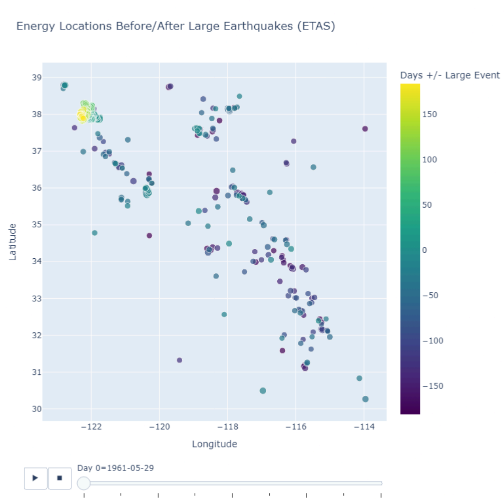

Figure 4: Energy Locations Before/After Large Earthquakes (ETAS)

4. RESULTS

The results of the analysis showed several discrepancies between ETAS and USGS data. Notably, while both ETAS and USGS data exhibit clustering of earthquakes around the day of major events, ETAS tends to display higher average energy both before and after large earthquakes, whereas USGS data records a pronounced spike in energy at the time of the main event itself. This spike is orders of magnitudes higher than energy calculations from generated ETAS data. In addition, USGS data also demonstrates sustained aftershock activity up to 100 days following a large earthquake, a feature not adequately captured by the ETAS model, which produces more methodological and short-lived clustering. The locations of these aftershocks are also more spread out in recorded USGS data, however, those events might be unrelated to the initial mainshock.

Additional observations have also been made following the implementation of the LSTM for cell-grid analysis. For example, the ETAS model sometimes centralizes energies within a single grid or a group of grids that are next to each other following a period of high energy. This behavior is not seen in recorded USGS data which shows that Earthquake energies are not centralized near places of high energy and rather show earthquake activity in other locations away from the place of centralized energy.

5. CONCLUSION

My capstone research project conducted a detailed comparison between the ETAS simulation model and historical USGS earthquake data. By identifying where ETAS diverges from observed data from USGS, my research offers valuable insights for improving the ETAS model and advancing the accuracy of earthquake hazard assessment in California.

By openly sharing the research findings through an interactive website and making the code available on GitHub, this project not only helps advance the development of the ETAS model but also fosters transparency and collaboration within the scientific community. This foundation encourages further research and additional refinement to seismic prediction models, supporting the ongoing pursuit of more accurate earthquake forecasting and risk mitigation strategies.

6. FUTURE WORK

Building on the insights gained from comparing ETAS simulations with USGS earthquake data, I propose several directions for future research. One major avenue is the development of cell-based models to analyze and predict seismic energy trends at a regional level. By dividing California into spatial grid cells and calculating the relationships between earthquake energies within these cells, it would help determine whether the seismic activity in one cell is influenced by neighboring cells.

Another key area is investigating the connection between earthquake clustering and known fault lines. Understanding how fault structures relate to the multiplicity and magnitude of clustered events could improve the accuracy of predictive models. Additionally, leveraging these relationships, future work should aim to develop models capable of forecasting spikes in earthquake energies, as well as identifying early indicators of heightened seismic risk.

7. ACKNOWLEDGMENTS

I would like to express my gratitude to my capstone advisor, Professor Geoffrey C. Fox, who provided guidance and support throughout the course of the project. I would also like to thank the members of the AI for Science 2023 Summer Independent study members and the members of the 2024 Fall Independent study whose work set the groundwork for my capstone project. Finally, I would like to thank Professor John B. Rundle and his team at UC Davis who provided me with their ETAS model for testing.

REFERENCES

[1] G. C. Fox, J. B. Rundle, A. Donnellan, and B. Feng, "Earthquake nowcasting with deep learning," Geohazards, vol. 3, no. 2, pp. 199–226, 2022.

[2] J. Zhuang, "Second-order residual analysis of spatiotemporal point processes and applications in model evaluation," Journal of the Royal Statistical Society Series B: Statistical Methodology, vol. 68, no. 4, pp. 635–653, 2006.

[3] G.C. Fox, M.K. Islam, V. Kamalakrishnan, J. Zheng,, K. Ma, Z. Russell, "Surrogate Modeling for Efficient Earthquake Prediction," doi: 1693166726736

[4] V. Kamalakrishnan, D Jani, P. Patel, "Earthquake Historical Data Analysis". https://thev123.github.io/EarthquakeHistoricalDataAnalysis/intro.html