Spatial and temporal analysis of earthquakes in Southern California based on K-means clustering and b-value analysis

Brian Zhu¹, Min Chen²

¹Department of Statistics and Data Science, Yale University, New Haven, CT 06520, USA
²Department of Computational Mathematics, Science, and Engineering, Michigan State University, East Lansing, MI 48824, USA

Motivation

- Analyzing earthquake clusters offers insight into seismic hazards of different geological and tectonic settings.
- Unsupervised clustering algorithms provide more objective clustering more efficiently than clustering done by human experts.

Objectives

- Improve spatial clustering of earthquakes by adding earthquake-fault distance constraints to a classical K-means algorithm.
- Analyze and evaluate spatial and temporal variations of the b-value of the earthquake clusters derived from the modified algorithm.

Methods

- K-means: widely used in earthquake clustering and seismic hazard analysis.
  - Silhouette score ($S$): measures how well data is clustered (averaged over all points $i$)
    \[
    S_i = \frac{b_i - a_i}{\max(a_i, b_i)}
    \]
  - Gutenberg-Richter Law:
    \[
    \log N = a - bM
    \]
    - $N$: earthquakes above magnitude $M$
    - $a$, $b$: constants
  - b-value: the value of $b$ in the G-R Law, roughly the ratio of small to large earthquakes
  - Magnitude of completeness ($M_c$): minimum magnitude where all earthquakes are detected by the local seismic network, estimated using the entire magnitude range (EMR) method

Results: Spatial variations of b-value

- Characterize the spatial distribution and magnitude of earthquake clusters.
- Notable spatial correlation with model crustal dilatation (extension or compression) and isotatic gravity (crustal density influence on gravity)
- Higher b-values in regions of crustal extension and in basin edges

Methods, cont.

- Added earthquake-fault distance constraints for major faults in Southern California (e.g. San Andreas, San Jacinto, Elsinore) to K-means.
- Computed silhouette scores for combinations of faults and weights to find the best one.
- Derived a maximum likelihood estimate of the b-value from the G-R Law:
  \[
  \hat{b} = \log e \frac{M_c - M}{M_c}
  \]
  - $M_c$: magnitude of completeness
  - $M$: mean magnitude of events above $M_c$

Results: Temporal variations of b-value

- Data Set
  - July 4, 2019 – July 3, 2019
  - 53,000 earthquake locations
  - Retrieved from Southern California Seismic Network
  - Ridgecrest Earthquakes
    - July 4, 2019: $M_c$ 6.4
    - July 5, 2019: $M_c$ 7.1
  - Largest earthquakes in the region over the last 10 years
  - Some distant clusters have similar time series: suggests interactions and distant stress transfer between faults

Data Set

- 900,000 earthquake locations
- Locations obtained from earthquake template matching
- Maps of various physical properties: Hauksson (2011)¹
- Notable spatial correlation with model crustal dilatation (extension or compression) and isotatic gravity (crustal density influence on gravity)

Conclusions

- The modified K-means algorithm produced better spatial clustering of earthquakes.
- The spatial and temporal b-value analysis of earthquake clusters suggests that the physical properties of the crust control the earthquake spatial distribution and that the interactions between different fault systems can affect the occurrence of large crustal earthquakes.

References

5. Modified K-means: $S = 0.572697$
6. Classical K-means: $S = 0.561586$

Acknowledgements

For help in preparing this paper, we acknowledge support from NSF Grant OAC-1560168, the Institute for Cyber-Enabled Research, and the Advanced Computational Research Experience for Students at Michigan State University.