(what citation format are we using? We are free to choose but need to pick one): APA

Depth influences magnitudes:

<https://www.nature.com/articles/365045a0>

citation for our datasets :

* Waldhauser, F. and D.P. Schaff, Large-scale relocation of two decades of Northern California seismicity using cross-correlation and double-difference methods, J. Geophys. Res.,113, B08311, [doi:10.1029/2007JB005479, 2008](http://dx.doi.org/10.1029/2007JB005479).
* Waldhauser, F., Near-real-time double-difference event location using long-term seismic archives, with application to Northern California, Bull. Seism. Soc. Am., 99, 2736-2848, [doi:10.1785/0120080294, 2009](http://dx.doi.org/10.1785/0120080294).

Papers that uses our dataset:

* A Comparison of Bayesian Hierarchical Space-Time Models for Earthquake Data By Bent NATVIG and Ingunn Fride TVETE, <http://webdoc.sub.gwdg.de/ebook/serien/e/uio_statistical_rr/07-04.pdf>
  + The focus of this paper is: how large the maximal earthquake for a given time period and grid cell going
* Gordon J. Ross; Bayesian Estimation of the ETAS Model for Earthquake Occurrences. *Bulletin of the Seismological Society of America* 2021;; 111 (3): 1473–1480. doi: <https://doi.org/10.1785/0120200198>
  + This uses our dataset, and uses ETAS model to do prediction.

Suggested ways for our modeling (for numbers of earthquake happening) (by month? We can also filter and focus on stronger earthquakes like M4.0+ instead of M3.0+)

For the first two we can go with the easier theme: comparing it with frequentist time-series models. We can say that since we are analyzing them in monthly basis, so we can try to treat the major earthquakes are the same as foreshock and aftershocks, and see what the results are, versus the ETAS model described below

1. Log-random walk model

This approach acknowledges the temporal clustering and periods of quiescence characteristic of seismic activity.

Key points to support this basic modelling:

* Ensures positivity of the rate parameter
* allows for smooth transitions between time periods
* captures both gradual shifts and more abrupt changes in seismicity rates

1. Hierarchical Model (spatial modeling)

Then place prior on hyperparameters

1. ETAS model (done by several papers already and very difficult, also most of them used our dataset. It considers a large earthquake (an independent Poisson process), and aftershocks it excites (another Poisson process))

* Ogata, Yosihiko. "Statistical models for earthquake occurrences and residual analysis for point processes." *Journal of the American Statistical association* 83.401 (1988): 9-27.
  + This describes the ETAS model
* Gordon J. Ross; Bayesian Estimation of the ETAS Model for Earthquake Occurrences. *Bulletin of the Seismological Society of America* 2021;; 111 (3): 1473–1480. doi: <https://doi.org/10.1785/0120200198>
  + This uses our dataset, and uses ETAS model to do prediction.

Bayesian Hierarchical Modeling for Count Data

<https://www.taylorfrancis.com/books/mono/10.1201/b16018/bayesian-data-analysis-david-dunson-donald-rubin-john-carlin-andrew-gelman-hal-stern-aki-vehtari>

 Hierarchical models for grouped/structured data (e.g., per-region effects uru\_rur​)

 Use of negative binomial distribution to handle overdispersion in count data

 Half-Cauchy priors for scale parameters

Bayesian Hierarchical Modeling of Earthquake Data

<https://link.springer.com/article/10.1007/s11009-006-9008-0>

 Introduced Bayesian models of earthquake counts over time and space

 Modeled overdispersed earthquake counts, incorporating spatial clustering and time-varying features

How is our Bayesian model different:

**1. Inclusion of Time-Lagged Magnitude (LagMag)**

* **model includes LagMag**, i.e., average magnitude in the *previous month*, as a covariate.
* **Most literature** (e.g., Clements et al., 2011) uses only current-time covariates (like Depth, Lat, RMS).

**2. Modeling Overdispersion via Covariates**

* 1. **Region Defined by Spatial Grid (Lat/Lon Binning)**

While our model structure is inspired by the hierarchical negative binomial framework used in Clements et al. (2011) and Ogata (1998), we extended the modeling approach in several ways:

* First, we incorporated temporal dependence by including lagged average magnitude as a covariate.
* Second, we modeled dispersion as a function of sensor reliability indicators (Nst, RMS, Clo), rather than treating it as a global or region-specific scalar.
* Third, we constructed spatial regions dynamically using binned latitude and longitude, enabling a generalizable pipeline.
* Lastly, our Bayesian workflow adheres closely to modern best practices (Gelman et al., 2013; STAT 447), including prior and posterior predictive checks and safe posterior simulations.