Title

Nowcasting Earthquakes With Stochastic Simulations: Information Entropy of Earthquake Catalogs

Journal

Earth and Space Science, 11(6)

ISSN

2333-5084

Authors

Rundle, John B

Baughman, Ian

Zhang, Tianjian

Publication Date

2024-06-01

DOI

10.1029/2023ea003367

AGU ADVANCING EARTH AND SPACE SCIENCES

Earth and Space Science

METHOD
10.1029/2023EA003367

Key Points:

- Earthquake nowcasting tracks the change in the potential for large earthquakes, using information contained in seismic catalogs
- We analyze the information contained in the space-time clustering that is observed in earthquake seismicity
- We find that "aftershock de-clustering" of catalogs removes information about future large earthquakes that the nowcasting method uses

Correspondence to:
J. B. Rundle,
jbrundle@ucdavis.edu

Citation:
Rundle, J. B., Baughman, I., & Zhang, T. (2024). Nowcasting earthquakes with stochastic simulations: Information entropy of earthquake catalogs. *Earth and Space Science*, *11*, e2023EA003367. https://doi.org/10.1029/2023EA003367

Received 20 OCT 2023
Accepted 17 MAY 2024

Author Contributions:
Conceptualization: Ian Baughman
Data curation: John B. Rundle
Formal analysis: John B. Rundle
Methodology: John B. Rundle
Software: John B. Rundle, Tianjian Zhang
Validation: Ian Baughman
Writing – original draft: John B. Rundle, Ian Baughman
Writing – review & editing: Ian Baughman, Tianjian Zhang

Nowcasting Earthquakes With Stochastic Simulations: Information Entropy of Earthquake Catalogs

John B. Rundle[1,2,3], Ian Baughman[1], and Tianjian Zhang[1]

[1]Department of Physics and Astronomy, University of California, Davis, CA, USA, [2]Department of Earth and Planetary Sciences, University of California, Davis, CA, USA, [3]Santa Fe Institute, Santa Fe, NM, USA

Abstract Earthquake nowcasting has been proposed as a means of tracking the change in large earthquake potential in a seismically active area. The method was developed using observable seismic data, in which probabilities of future large earthquakes can be computed using Receiver Operating Characteristic methods. Furthermore, analysis of the Shannon information content of the earthquake catalogs has been used to show that there is information contained in the catalogs, and that it can vary in time. So an important question remains, where does the information originate? In this paper, we examine this question using stochastic simulations of earthquake catalogs. Our catalog simulations are computed using an Earthquake Rescaled Aftershock Seismicity ("ERAS") stochastic model. This model is similar in many ways to other stochastic seismicity simulations, but has the advantage that the model has only 2 free parameters to be set, one for the aftershock (Omori-Utsu) time decay, and one for the aftershock spatial migration away from the epicenter. Generating a simulation catalog and fitting the two parameters to the observed catalog such as California takes only a few minutes of wall clock time. While clustering can arise from random, Poisson statistics, we show that significant information in the simulation catalogs arises from the "non-Poisson" power-law aftershock clustering, implying that the practice of de-clustering observed catalogs may remove information that would otherwise be useful in forecasting and nowcasting. We also show that the nowcasting method provides similar results with the ERAS model as it does with observed seismicity.

Plain Language Summary Earthquake nowcasting was proposed as a means of tracking the change in the potential for large earthquakes in a seismically active area, using the record of small earthquakes. The method was developed using observed seismic data, in which probabilities of future large earthquakes can be computed using machine learning methods that were originally developed with the advent of radar in the 1940s. These methods are now being used in the development of machine learning and artificial intelligence models in a variety of applications. In recent times, methods to simulate earthquakes using the observed statistical laws of earthquake seismicity have been developed. One of the advantages of these stochastic models is that it can be used to analyze the various assumptions that are inherent in the analysis of seismic catalogs of earthquakes. In this paper, we analyze the importance of the space-time clustering that is often observed in earthquake seismicity. We find that the clustering is the origin of information that makes the earthquake nowcasting methods possible. We also find that a common practice of "aftershock de-clustering", often used in the analysis of these catalogs, removes information about future large earthquakes.

1. Introduction

Earthquake nowcasting (Rundle et al., 2019, 2021a, 2021b, 2021c, 2022a, 2023; Rundle and Donnellan, 2020; Pasari, 2019, 2020, 2022; Pasari and Mehta, 2018; Chouliaras et al., 2023) is a relatively new method that uses elements of machine learning to track the current state of the potential for large earthquakes, as well as the recent past and near future. Nowcasting is based on the Receiver Operating Characteristic (ROC) method that was developed with the invention of radar in 1941 relating to the observation of "signals" associated with reflections, or "events" (https://en.wikipedia.org/wiki/Receiver_operating_characteristic).

The fact that one can compute a probability for large earthquakes (Rundle et al., 2023) implies that there is information (Shannon, 1948) contained within seismic catalogs. An important question then is:

> From where does this information arise? What property of the catalogs is associated with this information content? This is the question that we seek to answer in the current paper.

While random clustering can occur naturally as a result of Poisson statistics (Gross & Rundle, 1998), the hypothesis that we consider here is that the information arises from the "non-Poisson", scale-invariant power-law clustering of aftershocks. This non-Poisson clustering typically arises following large earthquakes, or as swarms or "bursts." More specifically, we find that information is associated with the relative quiescence that follows aftershock activity, meaning that removing aftershocks obscures the boundary between an "active" phase versus a "quiescent" phase.

We are currently developing methods (Rundle, 2023) for earthquake nowcasting that make use of recent developments in science transformers, which are the basis for Large Language Models (LLMs) such as ChatGPT (e.g., Chang et al., 2023; Fox et al., 2022; Yang et al., 2023). In these science transformers, models are trained on a large corpus of data, allowing the transformer to build a predictive model based on the concept of query, key and vector inputs to an encoder-decoder architecture combined with the concept of dot-product attention (i.e., Vaswani et al., 2017). LLMs are trained on large collections of facts, internet sites, documents, computer codes, images, and other data. We adapt this science transformer architecture to analyze time series for predictive purposes.

In the earthquake application, we do not generally have access to the volume of earthquake catalog data that we require, since these catalogs only extend back a few decades in time with reliable data. As a result, we must turn to other means of training the transformer models, specifically, by the use of extensive simulations of seismic activity. For this reason, we turn initially to stochastic earthquake simulations which are physics-informed models based on the observational laws of earthquake seismicity. Identifying the origin of the information contained in the catalogs is important for the evaluation of the utility of the nowcasting methods.

2. Testing Nowcast Models With Stochastic Simulations

To test our hypothesis of information origin, we turn to stochastic earthquake simulations in which the temporal clustering of seismicity can be easily varied to investigate its effects. Examples of these stochastic models include the Epidemic Type Aftershock Sequences (ETAS) formalism. These ETAS models are based on using the fundamental observational laws of earthquake seismicity to build the rate for non-homogeneous Poisson models that can be used to generate space-time seismicity catalogs (Hardebeck, 2013; Helmstetter & Sornette, 2003; Lombardi, 2015; Mancini et al., 2021; Mancini & Marzocchi, 2023; Seif et al., 2017; Veen & Schoenberg, 2008; Zhuang, 2011; Zhuang et al., 2012).

The observational laws of interest are the Gutenberg-Richter (GR) magnitude-frequency relation; the Omori-Utsu (OU) aftershock time decay law; an analogous Omori-Utsu law for spatial aftershock migration away from the epicenter; and the Bath's law that governs the magnitude of the largest aftershock relative to the mainshock magnitude.

By themselves, these observed laws involve the scaling exponents b (GR); p and a parameter c_p (OU-time); and q together with a corresponding parameter c_q (OU-space). In addition, there are additional parameters and relations that are assumed, for example, an "earthquake productivity" relation having 2 parameters (K and α), and a background seismicity rate $\eta(\boldsymbol{x},\boldsymbol{y})$.

The typical process of fitting all these parameters to a catalog is usually carried out by maximizing a log-likelihood value involving the difference between the parameterized rate, and the rates determined from the catalog. Because all these parameters are involved in the overall log-likelihood equation, they are necessarily correlated, so that changing the value of one parameter necessarily involves changes in the values of other parameters. This makes the fitting process time-consuming and subject to correlation errors.

In this paper, we develop a different and simpler stochastic simulation model, an Earthquake Rescaled Aftershock Seismicity ("ERAS") model, which involves only 2 free, novel, independent, and uncorrelated parameters, one for time, one for space. The first of these parameters, f, determines the Omori-Utsu aftershock time decay relation with exponent p. The second parameter, g, governs the Omori-Utsu aftershock spatial migration scaling with scaling exponent q. These two parameters are determined by fitting the (p, q) values observed from stacked aftershock seismicity. The model is constructed using the standard statistical relations of magnitude-frequency scaling, aftershock scaling in time and space, and Bath's law. This process is similar to the ETAS models, but without the additional 4–5 parameters that are used in the ETAS models. We then apply the nowcast method to these ERAS models.

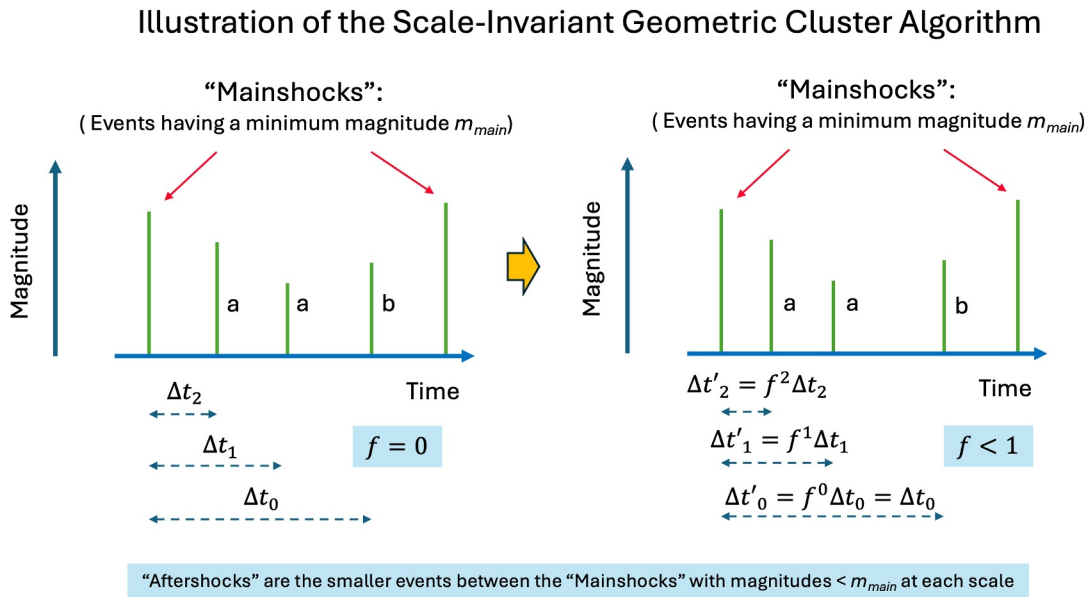

Figure 1. Schematic illustration of the re-clustering method described in the text. Given two earthquake "mainshocks" of some magnitude m_{main}, the smaller magnitude events between those mainshocks are identified. Using Bath's law, the subset of those smaller events that represent "aftershocks" are identified. Of the 3 smaller events shown, the first 2 were labeled as "aftershocks" ("a"), and the last event is labeled as "background" ("b"). These 2 "aftershocks" then have their times of occurrence altered to be earlier, and closer to the first mainshock. The time of the "background" event is unchanged. The "aftershocks" are moved closer to the earlier mainshock in time, using a geometric rescaling factor f as shown in the figure.

3. Stochastic Earthquake Rescaled Aftershock Seismicity (ERAS) Model

3.1. General Approach

Although the ERAS model has been formulated for a full time-space simulation, we are concerned in this paper only with the time dependence, so for the moment, we focus only on f and defer detailed discussion of g to a future paper, although we briefly summarize the process below.

We build the simulated catalog in 3 steps. We first generate simulated catalogs having random (Poisson) inter-event earthquake times to find a candidate catalog that has a b value within the observed margin of fitting error. Magnitudes of these events are drawn from a Gutenberg-Richter distribution. Next, since a (fractal) power law implies a geometric scaling or recursion, for example, similar to the von Koch curve and other fractal curves (e.g., Turcotte, 1997), we apply a geometric clustering algorithm in time using the single parameter f. The basic method is illustrated in Figure 1. We then search for a value of f that produces a match to the observed p value within the margin of fitting error for an ensemble of stacked mainshocks.

For the spatial (fractal) power law scaling, we assume that aftershocks migrate away from the mainshock epicenter by a random walk, which is known to have scaling (power law) distribution properties. We then analyze the observational catalog to determine the ratio of random walk step sizes in latitude/longitude, and introduce a multiplicative scaling parameter g. Finally we search for a value of g that produces a fit to the observed value of q for the simulated random walk within the observational error.

An important point to note is that the three steps are done sequentially, and that the two novel parameters f and g are independent and uncorrelated. The model is therefore characterized by only the 2 uncorrelated free parameters that are determined from the observed scaling exponents.

As a practical matter, it typically takes only a few minutes of wall clock time to find acceptable models using a Monte Carlo or grid search algorithm that simultaneously fits the b-value, the p-value and the q-value of the observed California catalog within the observational error.

3.2. Initial Model Set-Up and Event Classification

Given a USGS catalog with completeness magnitude μ, we create a catalog with the same exact number of events, having completeness magnitude μ, and with events drawn from a Gutenberg-Richter magnitude-frequency distribution with the observed b-value that is determined from fitting the catalog.

We also use a value for Bath's law $\Delta m_B = 1.1$, a typical value for earthquake sequences determined from many previous works, which a number of recent studies show in the range 1.0–1.2 (Guglielmi et al., 2022; Holliday et al., 2008; Shcherbakov et al., 2004, 2005). We do not consider either the b-value or the Δm_B value to be "free parameters", since their definition and method of evaluation are well-known.

Initially, the time between these events Δt is set using a Poisson (exponential distribution) for inter-event times Δt:

$$\Delta t = -\mathrm{Log}(1 - \zeta) \tag{1}$$

where ζ is a random deviate drawn from a uniform distribution. Once the catalog has been defined, the times are proportionately rescaled (expanded or compressed) to lie between an observed catalog start date (e.g., such as 1/1/1990) and the catalog final date (e.g., such as the present 3/1/2024).

Next, the Bath's number Δm_B, which is the typical difference between the mainshock magnitude and the largest aftershock magnitude, is used to estimate the number of aftershocks from a mainshock. Given a mainshock magnitude m, we label the next N_a events in time as "aftershocks" ("a"), where N_a is given by:

$$N_a = 10^{b(m - \mu - \Delta m_B)} \tag{2}$$

Events not labeled as aftershocks are labeled as "background" events ("b"). Background events having aftershocks are considered to be "mainshocks."

An implication of the above classification is that only background events with magnitudes $m \geq \mu + \Delta m_B$ can have aftershocks.

3.3. Rescaling Time Intervals for Aftershock Power Law Scaling: Parameter f

We begin by noting that the existence of the Omori law implies, like other fractal power laws, a geometric scale invariance. The next step is to identify cycles of "mainshock" events in order of descending magnitude. Finally, a geometric rescaling of the time intervals between the "aftershocks" and the "mainshocks" is applied in each cycle using a scale-invariant algorithm. The process is illustrated in Figure 1 and described in the following steps.

1. Assume that the largest magnitude event in the simulation is m_L. Then the largest magnitude aftershock is expected to have magnitude $m_a = m_L - \Delta m_B$. As described above, the smallest mainshock that can have aftershocks is $m_{min} = \mu + \Delta m_B$.
2. We store the small magnitude background ("b") events having magnitudes $m < m_{min}$ in a "random background" file.
3. Beginning with $m_a = m_L - \Delta m_B$, and progressing down in magnitude at intervals of 0.1 magnitude unit to m_{min}, we consider the intervals between these events as a set of m_{main}- "mainshock" cycles.
4. The rescaling factor $f < 1$ is shown in Figure 1. It is applied to the N_a aftershock-labeled ("a") events that have magnitudes $m < m_{main}$ that occur between successive m_{main} events.
5. The rescaling algorithm is then applied recursively. Beginning with earthquake magnitudes $m_{main} \geq m_a$, we repeat this procedure at the level $m_{main} = m_a$–0.1, in other words at a level 0.1 magnitude unit lower. This recursive procedure continues to be applied down smallest mainshock magnitude $m_{min} = \mu + \Delta m_B$.
6. The adjusted time intervals are found to be proportionately smaller than the original time intervals as a result of this geometric rescaling algorithm. Their time ordering remains unchanged. But as a result, the original time interval 1/1/1990—present will have changed, and we again need to linearly expand all times to restore that total time period for the entire catalog.
7. Following step 7, we remove all the small magnitude background events with $m < \mu + \Delta m_B$ which cannot by definition have aftershocks (events labeled with "b"), and replace these with randomized (in time) events from

the file of small events from step 2, with random times assigned between 1/1/1990 - present. This produces random background events in the catalog.

8. The optimal geometric rescaling factor f is finally found by fitting the aftershock decay times to the p-value from the observed catalog, by repeating the above process until a satisfactory value of f is obtained. An example is shown in the following.

4. Examples: Application of the Rescaling Algorithm

4.1. Compression of Aftershock Delay Time Intervals

The algorithm described in 1–8 above compresses the random delay time intervals of labeled random aftershocks to produce a scale-invariant set of aftershock delay times with an Omori power law p-value exponent.

Suppose there are 500 aftershocks of a large main shock, and further suppose that $f = 0.999$. The delay time, or time interval between the first aftershock following the mainshock and the mainshock, is compressed to a value 0.999^500 = 0.60637 => 61% of the original delay time interval. Continuing on, the delay time between the 100th aftershock and the mainshock is compressed to a value 0.999^400 = 0.67019 => 67% of the original random delay time.

The final aftershock has an adjusted time interval that is compressed to a value 0.999^1 => 99.9% of the original, or barely compressed at all. Note that the ordering of aftershock delay time intervals is unchanged by this rescaling, so that the 100th aftershock remains the 100th aftershock after the rescaling algorithm is applied.

Since this time compression process is repeated recursively, a given delay time will be recursively compressed to progressively smaller values. Fundamentally, this is a manifestation of the scale-invariant process that is associated with the Omori p-value scaling.

4.2. Aftershocks Can Produce Daughter Aftershocks

Suppose there are 100 initially computed aftershocks of a mainshock, and aftershock number 10 is found to be large enough to have 20 daughter aftershocks. In this case we assume that these daughter aftershocks do not extend the aftershock number, which remains set at 100. On the other hand, suppose that aftershock number 90 is large enough to have 20 daughter aftershocks. In this case, we assume that the final 10 of the initially computed aftershocks represent the first 10 of the 20 daughter aftershocks. We then extend the total number of aftershocks by 10 to include the final 10 of the daughter aftershocks. The total number of aftershocks of the mainshocks has now increased to 110, rather than the original value of 100. These additional 10 aftershocks are then labeled "a" as well.

With this mechanism, aftershocks can themselves produce daughter aftershocks. And with the recursive clustering algorithm discussed above, delay time intervals of the daughter aftershocks with respect to the parent aftershock are themselves rescaled in time to be scale invariant with respect to the parent aftershock.

5. Fitting the Model to the Observed Catalog

As described earlier, we construct long catalogs of stochastic earthquake simulations for use in training transformer models for earthquake nowcasting. As part of this program, we demonstrate now that the ERAS model parameters can be selected so as to fit the statistical earthquake scaling laws within the observational error for the much shorter, observed catalog. These scaling laws are, specifically, the Gutenberg-Richter b-value, the Omori-Utsu (OU) aftershock time scaling law with scaling power p, and the analogous OU aftershock space scaling law with scaling power q.

We proceed by the following steps:

- We start by generating a series of sample catalogs using random selection of earthquake magnitudes from a Gutenberg-Richter magnitude-frequency distribution as discussed above, selecting the same number of events in the observed catalog.
- We then use a Monte Carlo or grid search to select the value of f that leads to an acceptable fit to observed the OU p value, using stacked aftershock data.

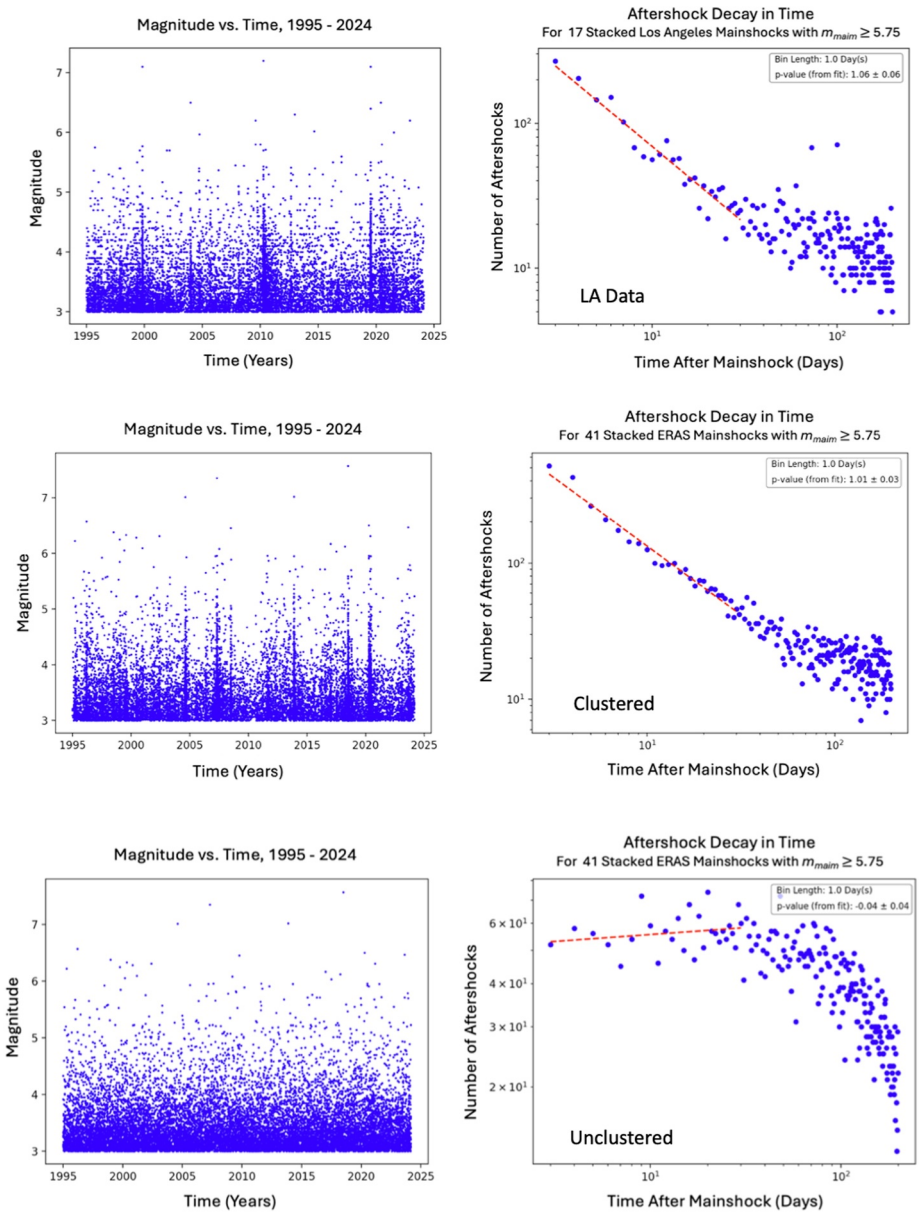

Figure 2. Data for California earthquakes (real + simulated) having magnitude $m \geq 3.0$ (left) magnitude versus time for 1995 - present (3/1/2024) (right) Omori-Utsu decay for mainshocks having magnitudes $m \geq 5.75$. Top: California data centered on Los Angeles in a $10° \times 10°$ latitude-longitude region. Middle: earthquake rescaled aftershock seismicity (ERAS) clustered simulation data. Bottom: ERAS un-clustered simulation data.

- Although we have not discussed in detail the spatial parameter g, we proceed in a similar way, by using a Monte Carlo or grid search to fit the OU q value.

Figure 2 shows 3 catalogs (left) and the corresponding stacked OU aftershock data for decay times (right) using 1 Day time bins. Top shows the observed catalog data, middle shows data from shows data from a clustered model ($f < 1$), and bottom shows data from an un-clustered model ($f = 1$). For the observed data and the clustered model, it can be seen that the $p \sim 1$ from the 29 stacked mainshocks having $m > 5.75$.

Many studies for California and elsewhere have documented p values for individual earthquakes that vary significantly, depending on the earthquake, the time interval over which the data are fit, the magnitude range, and to some extent how the data are binned (e.g., Hardebeck et al., 2019; Kisslinger & Jones, 1991; Reasenberg & Jones, 1989; Shcherbakov et al., 2005, 2006). Fits quoted in these papers for individual California earthquakes

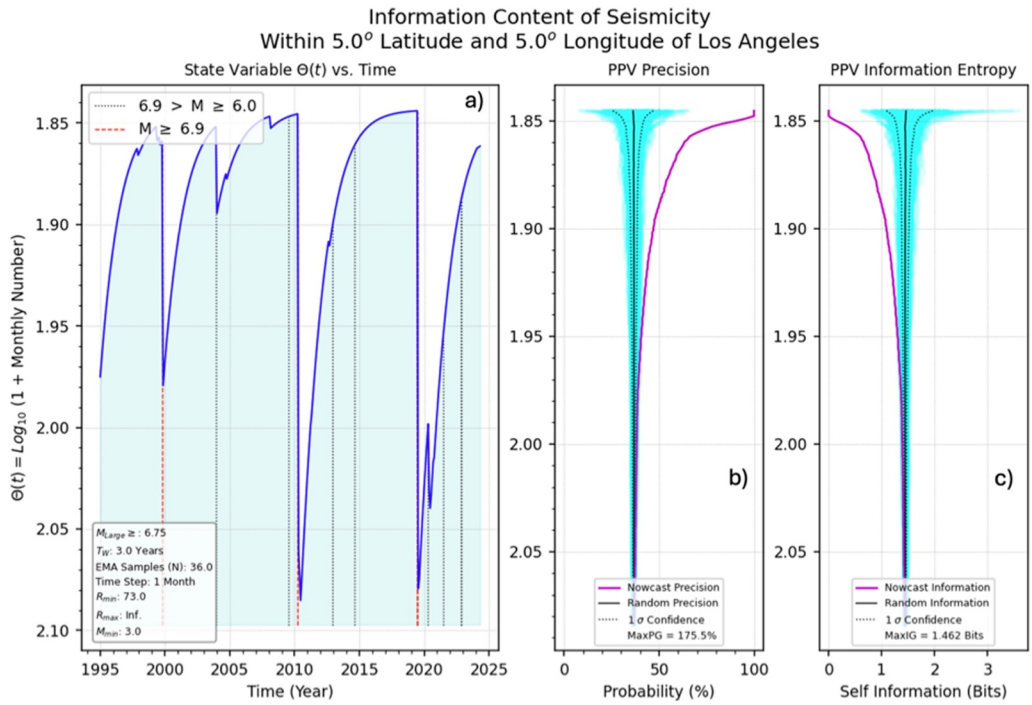

Figure 3. (a) Optimized earthquake potential state state variable for the California data as a function of time, using the same optimized filter parameters as in Rundle et al. (2022). Data were taken in the region of a 10° × 10° latitude-longitude box centered on Los Angeles, with the nowcast curve shown for 1/1/1995–3/1/1/2024. (b) Positive Predictive Value, PPV or Precision. Magenta curve is the PPV for the nowcast curve (state variable) shown in (a), where the vertical axis is the threshold T_H. The cyan lines represent the PPV for 50 random time series. Mean of the time series is the solid black line, and 1σ confidence is shown as the dashed lines. (c). Magenta curve is the corresponding self information I, Equation 6, on the horizontal axis as a function of the threshold value T_H on the vertical axis. The cyan curves are the self-information for the ensemble of 50 random time series, with mean (solid black line) and 1σ confidence as the dashed lines.

range from p values of 0.7–1.8 (Kisslinger & Jones, 1991), with a sample mean of 1.11 ± 0.25. Reasenberg and Jones (1989) find a value $p = 1.08$. Hardebeck et al. find a value $p = 1.03$, close to values in the SCSN and NCSN regions that they use for a variety of ETAS simulations.

For the observed data in Figure 2, we find $p = 1.06 \pm 0.06$ for 1 day bins, consistent with these other studies. Fitting our simulation data to this value, we found $f = 0.9987$, yielding a value $p = 1.01 \pm 0.03$ using 1 Day bins, compared to the observed p-value $p = 1.06 \pm 0.06$, the simulated data fitting clearly within the observational error. From analysis of other catalogs that fit the observed p-value data, we have found that values for f typically cluster around $f = 0.999$.

Note that the ERAS catalog was generated using an "input" b-value of $b = 0.95$, the value for the observed catalog. The "output" b-value found for the generated ERAS catalog was the same, $b = 0.95 \pm 0.01$. Furthermore, it can be seen that the p-value shown for the un-clustered catalog in Figure 5, $p = -0.04 \pm 0.04$, is not significantly different from $p = 0$, indicating no power law aftershock clustering.

6. Nowcasting Methods

For nowcasting we proceeded by computing the current large earthquake potential state (EPS), which is defined as the exponential moving average (EMA) of the inverted monthly rate of small earthquakes. We then defined a "signal", which is the current value of the EPS above or below a selected threshold, and an "event", which is the occurrence or non-occurrence of a large earthquake within a selected future time window T_W. In (Rundle et al., 2022a, 2023), the value of T_W is typically 1–3 years.

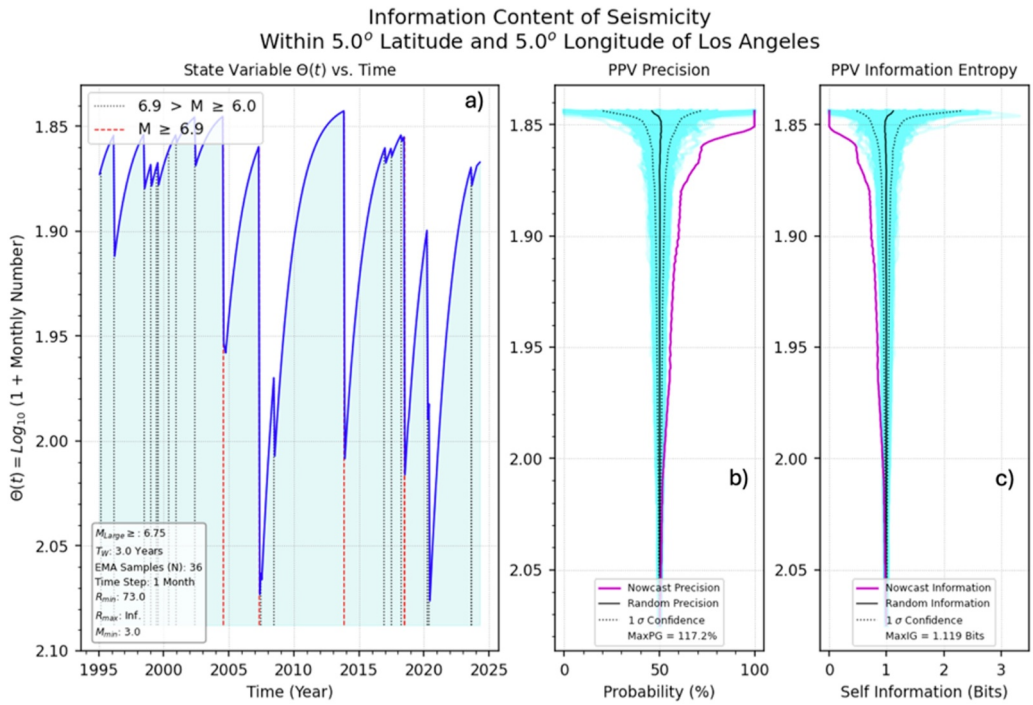

Figure 4. (a) Optimized state variable for the earthquake rescaled aftershock seismicity *clustered* simulation data as a function of time, using the same optimized filter parameters as in Rundle et al. (2022), and with model parameters fit to the California data. (b) Positive Predictive Value, PPV or Precision. Magenta curve is the PPV for the nowcast curve (state variable) shown in (a), where the vertical axis is the threshold T_H. The cyan lines represent the PPV for 50 random time series. Mean of the time series is the solid black line, and 1σ confidence is shown as the dashed lines. (c). Magenta curve is the corresponding self information I, Equation 6, on the horizontal axis as a function of the threshold value T_H on the vertical axis. The cyan curves are the self-information for the ensemble of 50 random time series, with mean (solid black line) and 1σ confidence as the dashed lines.

To implement the ROC method, we establish an arbitrary threshold value for the EPS curve, and categorize the "signals" and "events" into 4 categories. These are:

- If the "signal "is above the threshold, and a large earthquake *does occur* within the following T_W years ("event"), the "signal" is a True Positive (TP).
- If the "signal "is above the threshold, and a large earthquake *does not occur* within the following T_W years ("event"), the "signal" is a False Positive (FP).
- If the "signal "is below the threshold, and a large earthquake *does occur* within the following T_W years ("event"), the "signal" is a False Negative (FN).
- If the "signal "is below the threshold, and a large earthquake *does not occur* within the following T_W years ("event"), the "signal" is a True Negative (TN).

The initial threshold value is established at the lowest value of the EPS curve, all points on the timeseries curve are evaluated, and the confusion matrix is built for that value of the threshold. The threshold is then incrementally increased, and a new confusion matrix is built for that new threshold. This process continues until the maximum value of the EPS curve is reached. Typically one chooses several hundred threshold values, each resulting in its characteristic confusion matrix. From these confusion matrices, the ROC curve is then constructed by plotting the TP Rate (TPR), against the FP Rate (FPR), where:

$$\text{TPR} = \frac{TP}{TP + FN}, \text{FPR} = \frac{FP}{FP + TN} \tag{3}$$

From these quantities, we can also compute the Positive Predictive Value, or Precision (PPV), which is defined as:

10.1029/2023EA003367

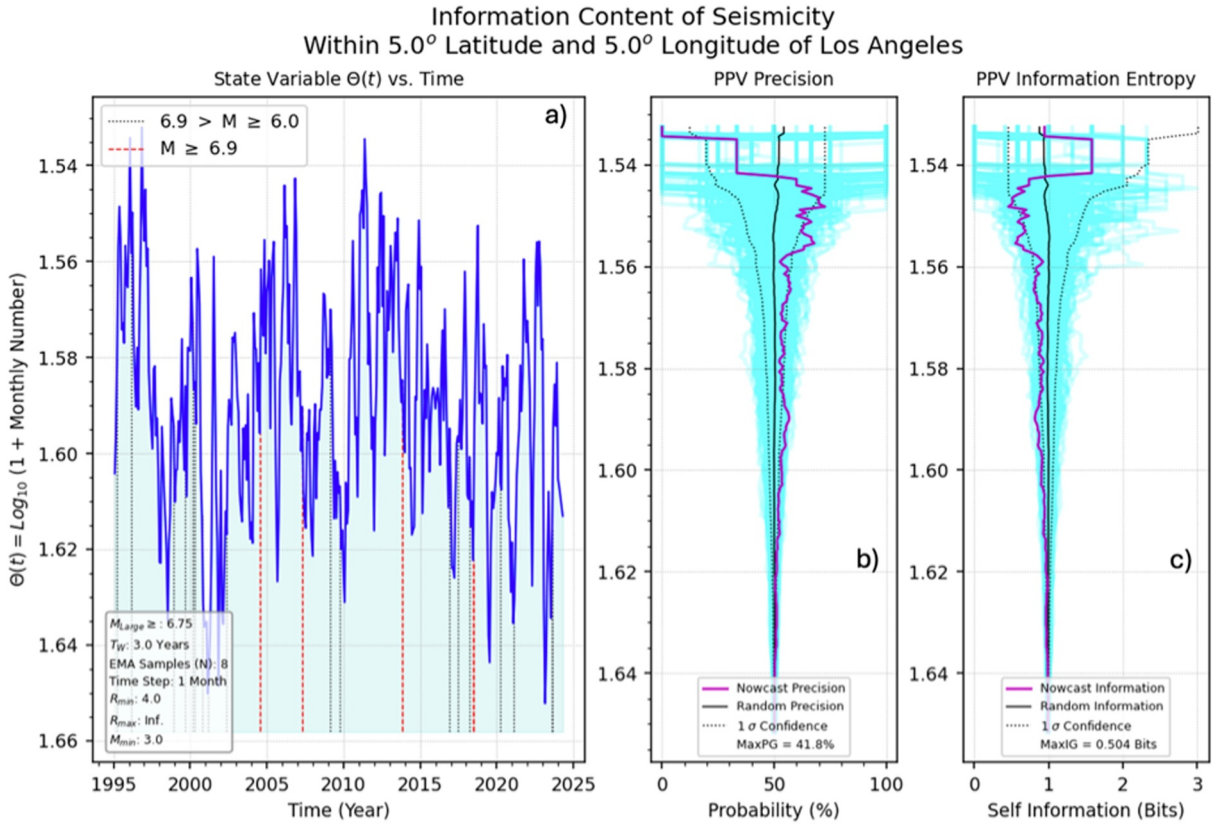

Figure 5. (a) Optimized state variable for the *unclustered* earthquake rescaled aftershock seismicity simulation data as a function of time, with no filter applied. (b) Positive Predictive Value, PPV or Precision. Magenta curve is the PPV for the nowcast curve (state variable) shown in (a), where the vertical axis is the threshold T_H. The cyan lines represent the PPV for 50 random time series. Mean of the time series is the solid black line, and 1σ confidence is shown as the dashed lines. (c). Magenta curve is the corresponding self information I, Equation 6, on the horizontal axis as a function of the threshold value T_H on the vertical axis. The cyan curves are the self-information for the ensemble of 50 random time series, with mean (solid black line) and 1σ confidence as the dashed lines.

$$\text{PPV} = \frac{TP}{TP + FP} \tag{4}$$

The precision can be regarded as the probability of a future large earthquake occurring within the following T_W years if the EPS curve is at, or above, a reference value on the EPS curve.

7. Nowcasting With California Earthquakes

In previous papers, we have described the nowcasting model, which tracks the current state of the complex dynamical earthquake system in space and time (e.g., Rundle et al., 2022). Briefly, the method uses data in the seismically active region around Los Angeles. It then applies an exponential moving average (EMA) to the monthly rate of small earthquakes in the region, using a number of weights $N_{EMA} = 36$ months, in this case. Finally, a correction to optimize the ROC skill of the nowcast was applied, partially to account as well for observational catalog incompleteness. In that paper, we used earthquakes having $M \geq 3.3$ since 1960.

After applying the EMA and small earthquake correction, the result is a smoothed version of the monthly rate of small earthquakes, which resembles an "upside-down" version of the earthquake cycle of stress accumulation and release, as has been pointed out in Rundle et al. (2022). Inverting this smoothed seismicity curve, we obtain the nowcast timeseries curve, examples of which are shown in the figures, in which seismic activation is at the bottom of the diagram (many earthquakes such as aftershocks), and quiescence is at the top. It can clearly be seen that quiescence is the precursor to large earthquakes. Results for this process applied to both observed California and simulated seismicity are shown in Figures 3–5 for the period 1/1/1995–3/1/2024.

8. Shannon Information Entropy of Catalogs

To analyze the information content of the earthquake catalogs, we consider Shannon information entropy. As discussed in Rundle et al. (2023), we use the Positive Predictive Value (PPV) of for large earthquakes ($m \geq 6.75$) within a future time window of duration $T_W = 3$ years as our probability value. The self-information I_{self} is then defined by:

$$I_{self} = -\mathrm{Log}_2 p(\omega) \tag{5}$$

where $p(\omega)$ = PPV, and ω is the Earthquake Potential Score (ω = EPS = value of the nowcast curve).

In Figures 3–5, the nowcast is shown in panel (a), the PPV in panel (b), and the Shannon self-information in panel (c). Figure 3 is the observed California seismicity, Figure 4 is the clustered simulated seismicity, and Figure 5 is the unclustered simulated seismicity.

Both the precision and the information of the data are represented by the solid line. The 50 cyan curves in each of the panels (b) and (c) are computed by randomizing the nowcast curve using a bootstrap procedure (random sampling with replacement), then computing the precision and self-information. Mean and standard deviation of the cyan curves are indicated by the dashed and dotted lines.

Also shown in Figure 3 for the California seismicity in panels b) and c) is the maximum probability gain, MaxPG = 175.5%, and the maximum information gain, MaxIG = 1.462 bits. We define probability gain as:

$$\mathrm{MaxPG} \equiv \max\left(\left(\mathrm{PPV}(\Theta)/<\mathrm{PPV}_R(\Theta)>\right) - 1\right) \tag{6}$$

where PPV(Θ) is precision, and $<\mathrm{PPV}_R(\Theta)>$ is the mean precision of the random PPV curves.

We define information gain as:

$$\mathrm{MaxIG} \equiv \max(<I_R(\Theta)> - I(\Theta)) \tag{7}$$

where $I(\Theta)$ is information entropy, and $<I_R(\Theta)>$ is mean information entropy. Both PPV and I are functions of the Earthquake Potential State (EPS ≡ Θ).

For the clustered simulation in Figure 4, MaxPG = 117.2%, and the maximum information gain, MaxIG = 1.119 bits. For the unclustered simulation in Figure 5, MaxPG = 41.8%, and the maximum information gain, MaxIG = 0.504 bits. For Figure 5, these values are essentially within the margin of error defined by the 50 cyan randomized nowcast curves.

In summary, it is clear from Figure 3 that the "precursor" to large earthquakes is anomalous or relative quiescence, which is the reason earthquakes are so hard to anticipate. Just as clearly, if one "de-clusters" the catalogs, as is common practice in many papers analyzing seismicity (e.g., Rundle et al., 2021a), the chance of detecting this anomalous quiescence will be reduced, thus implying that information has been removed by the de-clustering process.

9. Discussion

We emphasize that in this paper, the main point of our modified ERAS model is not to present a method to optimally fit California data, but rather to examine the effect of non-Poisson clustering in the catalog. In all models, Poisson (random, non-scale-invariant, non-power law) clustering continues to exist. The precursor to the large earthquakes in Figures 3 and 4 is clearly quiescence.

We note that many other authors have discussed the importance of seismic quiescence, which clearly can only be identified in comparison to activation. These include Kanamori (1981); Wiemer and Wyss (1994); Wyss and Habermann (1988); Chouliaras (2009); Katsumata (2011); Ben-Zion and Zaliapin (2020); Varotsos et al. (2011, 2014, 2020); Zaliapin and Ben-Zion (2022).

Comparing the results from the two ERAS catalogs (Figures 4 and 5), it can be seen that both precision and information content increase as clustering increases. This implies that de-clustering earthquake catalogs, as is

often done today, removes information. The information clearly lies in the anomalous quiescence that follows the clustered events, so that removing the clustered events obscures the quiescence "precursor."

In conclusion, our results suggest that earthquake catalogs should not be de-clustered prior to analysis, so that information is not removed. As noted in the introduction, we plan to use ERAS models together with AI-enhanced models to explore the utility of AI and deep learning in the problem of anticipating earthquake hazard in future publications.

Data Availability Statement

Python code that can be used to reproduce the simulation results of this paper can be found at Rundle (2024). Data for this paper was downloaded from the USGS earthquake ComCat catalog for California, and are available there. Python code at Rundle (2022) can be used to download and model these data for analysis using the methods of Rundle et al. (2022). Data for this paper was downloaded from the USGS earthquake catalog for California, and are freely available there. The Python code mentioned above can be used to download these data for analysis.

Acknowledgments
Research by JBR and IB was supported in part under DoE grant DE- SC0017324 to the University of California, Davis. JBR would also like to acknowledge generous support from a gift to UC Davis by John Labrecque. The authors would also like to acknowledge a helpful review by an anonymous referee.

References

Ben-Zion, Y., & Zaliapin, I. (2020). Localization and coalescence of seismicity before large earthquakes. *Geophysical Journal International*, *223*(1), 561–583. https://doi.org/10.1093/gji/ggaa315

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., et al. (2023). A survey on evaluation of large language models. *ACM Transactions on Intelligent Systems and Technology*, *15*(3), 1–45. https://doi.org/10.1145/3641289

Chouliaras, G. (2009). Seismicity anomalies prior to the 13 December 2008, M s= 5.7 earthquake in Central Greece. *Natural Hazards and Earth System Sciences*, *9*(2), 501–506. https://doi.org/10.5194/nhess-9-501-2009

Chouliaras, G., Skordas, E. S., & Sarlis, N. V. (2023). Earthquake nowcasting: Retrospective testing in Greece. *Entropy*, *25*(2), 379. https://doi.org/10.3390/e25020379

Fox, G. C., Rundle, J. B., Donnellan, A., & Feng, B. (2022). Earthquake nowcasting with deep learning. *Geohazards*, *3*(2), 199–226. https://doi.org/10.3390/geohazards3020011

Gross, S., & Rundle, J. (1998). A systematic test of time-to-failure analysis. *Geophysical Journal International*, *133*(1), 57–64. https://doi.org/10.1046/j.1365-246x.1998.1331469.x

Guglielmi, A. V., Zotov, O. D., Zavyalov, A. D., & Klain, B. I. (2022). On the fundamental laws of earthquake physics. *Journal of Volcanology and Seismology*, *16*(2), 143–149. https://doi.org/10.1134/s0742046322020026

Hardebeck, J. L. (2013). *Appendix S: Constraining epidemic type aftershock sequence (ETAS) parameters from the Uniform California Earthquake Rupture Forecast, Version 3 catalog and validating the ETAS model for magnitude 6.5 or greater earthquakes*. US Geol. Surv. Open-File Rept. 2013-1165-S, and California Geol. Surv. Special Rept. 228-S.

Hardebeck, J. L., Llenos, A. L., Michael, A. J., Page, M. T., & Van Der Elst, N. (2019). Updated California aftershock parameters. *Seismological Research Letters*, *90*(1), 262–270. https://doi.org/10.1785/0220180240

Helmstetter, A., & Sornette, D. (2003). Foreshocks explained by cascades of triggered seismicity. *Journal of Geophysical Research*, *108*(B10). https://doi.org/10.1029/2003jb002409

Holliday, J. R., Turcotte, D. L., & Rundle, J. B. (2008). A review of earthquake statistics: Fault and seismicity-based models, ETAS and BASS. *Earth Sciences and Mathematics*, *1*, 1003–1024. https://doi.org/10.1007/978-3-7643-8907-9_2

Kanamori, H. (1981). The nature of seismicity patterns before large earthquakes. In D. W. Simpson & P. G. Richards (Eds.), *Earthquake prediction, an international review* (Vol. 4, pp. 1–19). American Geophysical Union, Maurice Ewing Series.

Katsumata, K. (2011). A long-term seismic quiescence started 23 years before the 2011 off the Pacific coast of Tohoku Earthquake (M= 9.0). *Earth Planets and Space*, *63*(7), 709–712. https://doi.org/10.5047/eps.2011.06.033

Kisslinger, C., & Jones, L. M. (1991). Properties of aftershock sequences in Southern California. *Journal of Geophysical Research*, *96*(B7), 11947–11958. https://doi.org/10.1029/91jb01200

Lombardi, A. M. (2015). Estimation of the parameters of ETAS models by simulated annealing. *Scientific Reports*, *5*(1), 8417. https://doi.org/10.1038/srep08417

Mancini, S., & Marzocchi, W. (2023). SimplETAS: A benchmark earthquake forecasting model suitable for operational purposes and seismic hazard analysis. *Seismological Research Letters*, *XX*, 1–12. https://doi.org/10.1785/0220230199

Mancini, S., Werner, M. J., Segou, M., & Bapti, B. (2021). Probabilistic forecasting of hydraulic fracturing-induced seismicity using an injection-rate driven etas model. *Seismological Research Letters*, *92*(6), 3471–3481. https://doi.org/10.1785/0220200454

Pasari, S. (2019). Nowcasting earthquakes in the bay of Bengal region. *Pure and Applied Geophysics*, *176*(4), 1417–1432. https://doi.org/10.1007/s00024-018-2037-0

Pasari, S. (2020). Stochastic modeling of earthquake interevent counts (Natural Times) in Northwest Himalaya and adjoining regions. In *International conference on applied and computational mathematics* (pp. 495–501). Springer.

Pasari, S. (2022). Nowcasting earthquakes in Iran: A quantitative analysis of earthquake hazards through natural times. *Journal of African Earth Sciences*, *198*, 104821. https://doi.org/10.1016/j.jafrearsci.2022.104821

Pasari, S., & Mehta, A. (2018). Nowcasting earthquakes in the northwest Himalaya and surrounding regions. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *42*, 855–859. https://doi.org/10.5194/isprs-archives-xlii-5-855-2018

Reasenberg, P. A., & Jones, L. M. (1989). Earthquake hazard after a mainshock in California. *Science*, *243*(4895), 1173–1176. https://doi.org/10.1126/science.243.4895.1173

Rundle, J. (2022). jbrundle/OptimizingEarthquakeNowcasts: Optimal earthquake nowcasting (V1.0) [Software]. *Zenodo*. https://doi.org/10.5281/zenodo.7186635

Rundle, J. (2024). Earthquake_Rescaled_Aftershock_Seismicity_ERAS [Software]. *Zenodo*. https://doi.org/10.5281/zenodo.10810321

Rundle, J. B. (2023). Nowcasting earthquakes with QuakeGPT, an AI-enhanced earthquake generative pretrained transformer. *Authorea Preprints*. AGU23 https://essopenarchive.org/doi/full/10.22541/essoar.170034944.42344050

Rundle, J. B., & Donnellan, A. (2020). Nowcasting earthquakes in Southern California with machine learning: Bursts, swarms, and aftershocks may be related to levels of regional tectonic stress. *Earth and Space Science*, *7*(9), e2020EA0010. https://doi.org/10.1029/2020ea001097

Rundle, J. B., Stein, S., Donnellan, A., Turcotte, D. L., Klein, W., & Saylor, C. (2021a). The complex dynamics of earthquake fault systems: New approaches to forecasting and nowcasting of earthquakes. *Reports on Progress in Physics*, *84*(7), 076801. https://doi.org/10.1088/1361-6633/abf893

Rundle, J. B., Donnellan, A., Fox, G., & Crutchfield, J. P. (2021b). Nowcasting earthquakes by visualizing the earthquake cycle with machine learning: A comparison of two methods. *Surveys in Geophysics*, *43*(2), 1–19. https://doi.org/10.1007/s10712-021-09655-3

Rundle, J. B., Donnellan, A., Fox, G., Crutchfield, J. P., & Granat, R. (2021c). Nowcasting earthquakes: Imaging the earthquake cycle in California with machine learning. *Earth and Space Science*, *8*(12), e2021EA001757. https://doi.org/10.1029/2021ea001757

Rundle, J. B., Donnellan, A., Fox, G., Ludwig, L. G., & Crutchfield, J. (2023). Does the catalog of California earthquakes, with aftershocks included, contain information about future large earthquakes? *Earth and Space Science*, *10*(2), e2022EA002521. https://doi.org/10.1029/2022ea002521

Rundle, J. B., Giguere, A., Turcotte, D. L., Crutchfield, J. P., & Donnellan, A. (2019). Global seismic nowcasting with Shannon information entropy. *Earth and Space Science*, *6*(1), 191–197. https://doi.org/10.1029/2018ea000464

Rundle, J. B., Yazbeck, J., Donnellan, A., Fox, G., Ludwig, L. G., Heflin, M., & Crutchfield, J. (2022). Optimizing earthquake nowcasting with machine learning: The role of strain hardening in the earthquake cycle. *Earth and Space Science*, *9*(11), e2022EA002343. https://doi.org/10.1029/2022ea002343

Seif, S., Mignan, A., Zechar, J. D., Werner, M. J., & Wiemer, S. (2017). Estimating ETAS: The effects of truncation, missing data, and model assumptions. *Journal of Geophysical Research: Solid Earth*, *122*(1), 449–469. https://doi.org/10.1002/2016jb012809

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell system technical journal*, *27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shcherbakov, R., Turcotte, D. L., & Rundle, J. B. (2004). A generalized Omori's law for earthquake aftershock decay. *Geophysical Research Letters*, *31*(11), L11613. https://doi.org/10.1029/2004gl019808

Shcherbakov, R., Turcotte, D. L., & Rundle, J. B. (2005). Aftershock statistics. *Pure and Applied Geophysics*, *162*(6–7), 1051–1076. https://doi.org/10.1007/s00024-004-2661-8

Shcherbakov, R., Van Aalsburg, J., Rundle, J. B., & Turcotte, D. L. (2006). Correlations in aftershock and seismicity patterns. *Tectonophysics*, *413*(1–2), 53–62. https://doi.org/10.1016/j.tecto.2005.10.009

Turcotte, D. L. (1997). *Fractals and chaos in geology and geophysics*. Cambridge university press.

Varotsos, P., Sarlis, N. V., & Skordas, E. S. (2011). *Natural time analysis: The new view of time: Precursory seismic electric signals, earthquakes and other complex time series*. Springer Science and Business Media.

Varotsos, P. A., Sarlis, N. V., & Skordas, E. S. (2014). Study of the temporal correlations in the magnitude time series before major earthquakes in Japan. *Journal of Geophysical Research: Space Physics*, *119*(11), 9192–9206. https://doi.org/10.1002/2014ja020580

Varotsos, P. A., Skordas, E. S., & Sarlis, N. V. (2020). Fluctuations of the entropy change under time reversal: Further investigations on identifying the occurrence time of an impending major earthquake. *EPL*, *130*(2), 29001. https://doi.org/10.1209/0295-5075/130/29001

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. *Advances in Neural Information Processing Systems*, *30*.

Veen, A., & Schoenberg, F. P. (2008). Estimation of space-time branching process models in seismology using an EM-type algorithm. *Journal of the American Statistical Association*, *103*(482), 614–624. https://doi.org/10.1198/016214508000000148

Wiemer, S., & Wyss, M. (1994). Seismic quiescence before the landers (M= 7.5) and big bear (M= 6.5) 1992 earthquakes. *Bulletin of the Seismological Society of America*, *84*(3), 900–916.

Wyss, M., & Habermann, R. E. (1988). Precursory seismic quiescence. *Pure and Applied Geophysics*, *126*(2), 319–332. https://doi.org/10.1007/bf00879001

Yang, J., Jin, H., Tang, R., Han, X., Feng, Q., Jiang, H., et al. (2023). Harnessing the power of llms in practice: A survey on chatgpt and beyond. *ACM Transactions on Knowledge Discovery from Data*, *18*(6), 1–32. https://doi.org/10.1145/3649506

Zaliapin, I., & Ben-Zion, Y. (2022). Perspectives on clustering and declustering of earthquakes. *Seismological*, *93*(1), 386–401. Society of America. https://doi.org/10.1785/0220210127

Zhuang, J. (2011). Next-day earthquake forecasts for the Japan region generated by the ETAS model. *Earth Planets and Space*, *63*(3), 207–216. https://doi.org/10.5047/eps.2010.12.010

Zhuang, J., Harte, D., Werner, M. J., Hainzl, S., & Zhou, S. (2012). Basic models of seismicity: Temporal models. *Community Online Resource for Statistical Seismicity Analysis*. https://doi.org/10.5078/corssa-79905851