

# Earth and Space Science

## TECHNICAL REPORTS: METHODS

10.1029/2021EA001757

### Special Section:

Machine learning for Solid Earth observation, modeling and understanding

### Key Points:

- The current state of the earthquake cycle of tectonic stress accumulation and release is unobservable with existing methods
- We show that readily observable small earthquake correlations can be used to nowcast the current state of the earthquake cycle
- Machine learning techniques indicate that signals corresponding to future large earthquakes can be detected in a correlation time series

### Correspondence to:

J. B. Rundle,  
[jbrundle@ucdavis.edu](mailto:jbrundle@ucdavis.edu)

### Citation:




Rundle, J. B., Donnellan, A., Fox, G., Crutchfield, J. P., & Granat, R. (2021). Nowcasting earthquakes: Imaging the earthquake cycle in California with machine learning. *Earth and Space Science*, 8, e2021EA001757. <https://doi.org/10.1029/2021EA001757>

Received 17 MAR 2021  
Accepted 2 AUG 2021

© 2021 The Authors.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial License](#), which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.

## Nowcasting Earthquakes: Imaging the Earthquake Cycle in California With Machine Learning

John B. Rundle<sup>1,2,3</sup> , Andrea Donnellan<sup>2</sup> , Geoffrey Fox<sup>4</sup>, James P. Crutchfield<sup>1</sup>, and Robert Granat<sup>5</sup> 

<sup>1</sup>University of California Davis, Davis, CA, USA, <sup>2</sup>Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA, <sup>3</sup>Santa Fe Institute, Santa Fe, NM, USA, <sup>4</sup>Indiana University, Bloomington, IN, USA, <sup>5</sup>City University of New York, New York, NY, USA

**Abstract** We propose a new machine learning-based method for nowcasting earthquakes to image the time-dependent earthquake cycle. The result is a timeseries that may correspond to the process of stress accumulation and release. The timeseries are constructed by using principal component analysis of regional seismicity. The patterns are found as eigenvectors of the cross-correlation matrix of a collection of seismicity timeseries in a coarse grained regional spatial grid (pattern recognition via unsupervised machine learning). The eigenvalues of this matrix represent the relative importance of the various eigenpatterns. Using the eigenvectors and eigenvalues, we compute the weighted correlation timeseries of the regional seismicity. This timeseries has the property that the weighted correlation generally decreases prior to major earthquakes in the region, and increases suddenly just after a major earthquake occurs. As in a previous paper (Rundle & Donnellan, 2020, <https://doi.org/10.1029/2020ea001097>), we find that this method produces a nowcasting timeseries that resembles the hypothesized regional stress accumulation and release process characterizing the earthquake cycle. We then address the problem of whether the timeseries contain information regarding future large earthquakes. For this, we compute a receiver operating characteristic and determine the decision thresholds for several future time periods of interest (optimization via supervised machine learning). We find that signals can be detected that can be used to characterize the information content of the timeseries. These signals may be useful in assessing present and near-future seismic hazards.

**Plain Language Summary** Major earthquakes on fault systems in a tectonically active region are thought to occur in approximately repetitive cycles as a result of the buildup and release of tectonic forces (stress). Nowcasting is a technique adopted from weather, finance, and other fields that use readily observable proxy data to represent the unobservable stress accumulation process of interest. This paper presents a method that computes a timeseries representing the weighted correlation of small earthquake activity in the California region from 1950 to 2020. Prior to major magnitude  $M > 7$  earthquakes, the timeseries trends toward lower values. Just after the earthquake occurs, the timeseries increases suddenly in association with the earthquake, before resuming its gradual trend toward lower values. Plotting the timeseries on an inverted scale, one sees a cyclic behavior that strongly resembles the hypothesized earthquake cycle. In principle, we can therefore use this timeseries for nowcasting, as a proxy for stress accumulation and release. Using methods of signal detection first developed for radar by the British in the 1940's, we find that the timeseries contain information about future large earthquakes that can be used for hazard assessment.

## 1. Introduction

Earthquake hazard analysis is hobbled by our inability to directly observe the accumulation and release of tectonic stress in regions of seismic activity (Scholz, 2019). As a result, research in this area has focused on several other lines of investigation. In forecasting, a major emphasis is now being placed on topologically realistic numerical simulations (Tullis et al., 2012).

Alternatively, recent research has developed the idea of earthquake nowcasting, which uses proxy variables to infer the current state of the earthquake cycle (Luginbuhl, Rundle, Hawkins, & Turcotte, 2018; Luginbuhl et al., 2018a, 2018b, 2019; Pasari & Mehta, 2018; Pasari, 2019, 2020; Pasari & Yogendra, 2020; Rundle et al.,

2016, 2018, 2019, 2021). In the nowcasting approach, one uses observations of small earthquake seismicity to estimate the conditional probability that a major earthquake might occur after the current number of small earthquakes has occurred, given that one has not occurred since the last major event.

A comprehensive review of the current state of earthquake nowcasting, forecasting, and prediction is given by Rundle and Donnellan (2020). Perez-Oregon et al. (2020) have also shown that nowcasting methods can be extended to forecasting methods as well. These methods have begun to be applied to India (Pasari, 2019), Japan (K. Nanjo, 2020; personal comm., 2020), and Greece (G. Chouliaras, personal comm. 2019). Fundamentally, nowcasting is based on the concept of natural time (Sarlis et al., 2018; Varotsos, Sarlis, & Skordas et al., 2020; Varotsos, Skordas, & Sarlis et al., 2020; Varotsos et al., 2001, 2002, 2011, 2013, 2014), which is unique in its characteristics (Pasari, 2019).

Beginning with the nowcasting idea, Perez-Oregon et al. (2020) have now shown that nowcasting models can be extended into forecasting models for two types of model systems, one being the slider block model of Olami et al. (1992), and the other being a system in which the events obey a log-normal distribution. These are toy models as described above but may be applicable to real data. The forecast methods are tested by means of the receiver operating characteristic method that we also describe below.

Recently, Rouet-LeDuc et al. (2017) have developed a timeseries prediction technique using machine learning for acoustic emissions from events in laboratory experiments on regular, nearly periodic stick-slip friction. They also applied a similar technique for Episodic Tremor and Slip events in the Pacific Northwest (Rouet-LeDuc et al., 2019), which are also relatively regular in time.

In a previous study, Rundle and Donnellan (2020) showed that a timeseries resembling the long-hypothesized earthquake cycle could be constructed from the analysis of bursts of small earthquakes that are clustered in space and time. Here we present an alternative method that yields similar results and is perhaps more robust. We identify the characteristic patterns of activity in a seismically active region, defined using principle component analysis (PCA; Tiampo, Rundle, Gross, et al., 2002; Tiampo, Rundle, McGinnis, & Klein, 2002).

To summarize our results: While the statistics of earthquake occurrence are known to be nearly Poisson in time and thus without memory of prior large events, there may nonetheless be unexamined signatures of high risk times that are defined by characteristics of the small earthquake distribution in time. It is this hypothesis that we examine here using a novel correlation timeseries. The timeseries so defined implies that the regional correlation of seismic activity generally decreases prior to major earthquakes in California. Just after the occurrence of a major earthquake, the correlation of seismic activity discontinuously increases. The result strongly resembles the expected earthquake cycle of stress accumulation and release and is similar to results published earlier from the analysis of bursts of small-magnitude seismic activity. We then applied a standard timeseries technique based on constructing a receiver operating characteristic (ROC) together with a Shannon Information metric (e.g., Rundle et al., 2019) to show that signals of future large earthquakes may be present. The method implies some level of signal detection of future large earthquakes, albeit with errors.

## 2. Method

The first step is to define a spatial coarse graining by assigning an array of grid boxes of a given latitude and longitude  $\Delta x$  (in degrees) to the area of interest. Each of these grid boxes (tiles or partitions), here chosen to be  $0.33^\circ$  both in latitude and longitude, is required to have a minimum number of small earthquakes over the entire time interval used. The reason for a minimum number is to avoid the problem of small or zero eigenvalues in the matrix inversion in the Principal Component Analysis (PCA) described below.

In the results example described in this paper, a minimum of 35 small earthquakes having magnitudes  $M > 3.29$  over the time period of 1950–2020 was used, or a rate of about 1 small earthquake every 2 years. Requiring more than 35 small earthquakes reduces the number of acceptable grid boxes, and can lead to poor spatial resolution. Variations of this value change result somewhat, but the general conclusions remain. This procedure produces a set of  $N_x$  “active” grid boxes.

We then extract data from the seismic catalog as follows. A catalog is a group of values that we can describe as the set  $\{t_i, M_i, \mathbf{z}_i\}$ , where  $i = 1, \dots, N_E$ , in which  $N_E$  is the number of earthquakes in the catalog. Here  $t_i$  is the origin time of the earthquake,  $M_i$  is the magnitude, and  $\mathbf{z}_i$  is the hypocentral location (latitude-longitude-depth). Note that  $\mathbf{z}_i$  is a container variable for an epicentral (horizontal) location  $\mathbf{x}_i$  and depth  $d_i$ .

We then digitize the catalog in time at increments  $\Delta t$ , and assign a given earthquake to a time interval  $(t_j - \Delta t, t_j)$ ,  $j = 1, \dots, J_T$  and to the grid box centered at  $\mathbf{x}_n$ ,  $n = 1, \dots, N_X$ . These assignments then yield a collection of time series  $\Phi(\mathbf{x}_n, t_j)$ , which for convenience we designate as  $\Phi(\mathbf{x}_n, t)$ . Thus we have a total of  $N_X$  time series, digitized at equidistant intervals  $\Delta t$ , extending over the long interval  $t_0, \dots, (t_0 + \Delta t J_T)$ . In other words,  $\Phi(\mathbf{x}_n, t_j)$  is the number of earthquakes in the grid box centered on  $\mathbf{x}_n$ , occurring between  $t_j - \Delta t$  and  $t_j$ .

The next step is to compute the (eigen) patterns. To do so, we use PCA of the correlation matrix. The correlation matrix involves centered, univariant time series  $\hat{\Phi}(\mathbf{x}_n, t)$ . Here  $\hat{\Phi}(\mathbf{x}_n, t)$  is obtained from the timeseries  $\Phi(\mathbf{x}_n, t)$ :

$$\mu_{n,t} = \left( \frac{1}{t - t_0} \right) \int_{t_0}^t \Phi(\mathbf{x}_n, t') dt', \quad (1)$$

$$\sigma_{n,t}^2 = \left( \frac{1}{t - t_0} \right) \int_{t_0}^t \left( \Phi(\mathbf{x}_n, t') - \mu_{n,t} \right)^2 dt', \quad (2)$$

$$\hat{\Phi}(\mathbf{x}_n, t) = \frac{\left( \Phi(\mathbf{x}_n, t) - \mu_{n,t} \right)}{\sigma_{n,t}}. \quad (3)$$

The correlation matrix element  $C_{nm}(t)$  is then given by:

$$C_{nm}(t) = \int_{t_0}^t \hat{\Phi}(\mathbf{x}_n, t') \hat{\Phi}(\mathbf{x}_m, t') dt'. \quad (4)$$

$C_{nm}(t)$  is then diagonalized to find its eigenvectors (eigenpatterns)  $\mathbf{e}_i(t)$ ,  $i = 1, \dots, N_X$ , and eigenvalues  $\lambda_i(t)$ . Because  $C_{nm}(t)$  is a positive definite, symmetric matrix of rank  $N_X$ , the eigenvalues  $\lambda_i(t)$  are real and positive.

The next step is to define a sliding window seismicity state vector  $\boldsymbol{\psi}(t)$ . The  $N_X$  components of  $\Phi(\mathbf{x}_n, t)$  are just the  $N_X$  time series  $\Phi(\mathbf{x}_n, t)$ , summed over a previous time interval  $\tau = S\Delta t$ . The  $n$ th component of  $\boldsymbol{\psi}(t)$  is then:

$$\psi_n(t) = \int_{t-\tau}^t \Phi(\mathbf{x}_n, t') dt'. \quad (5)$$

Since the  $\mathbf{e}_i(t)$  are orthonormal and complete, we can expand  $\boldsymbol{\psi}(t)$  in the eigenpatterns with expansion coefficients  $a_i(t)$ :

$$\boldsymbol{\psi}(t) = \sum_i a_i(t) \mathbf{e}_i(t), \quad (6)$$

we then compute the power spectrum of  $\boldsymbol{\psi}(t)$ , that is,  $a_i(t)^2$ . In computing  $\mathbf{e}_i(t)$ , we use only data for  $t' \leq t$ .

The weighted correlation of the seismicity at time  $t$  is then found as the dot product of the power spectrum with the vector of correlation eigenvalues. This dot product is then the weighted correlation value  $\chi(t)$  for the regional seismicity:

$$\chi(t) \equiv \langle \lambda(t) \rangle = \sum_i \lambda_i(t) a_i(t)^2, \quad (7)$$

$\chi(t)$  represents a timeseries containing (possibly) significant information content as we discuss below.

In computing (Equation 7), it is found that the number of time series with the required minimum number of events and therefore active grid boxes, generally increases with time. So in order to compute a continuous timeseries, uniformly valid for all times  $t$ , we adopt the normalization:

$$\sum_i \lambda_i(t) = 100, \quad (8)$$

$$\sum_i a_i(t)^2 = 1(L_2\text{norm}). \quad (9)$$

We plot  $\chi(t)$  as a function of time below, which we interpret as a nowcasting correlation timeseries.

### 3. Application

We apply this method to California as an illustration of the method. We begin by partitioning the region centered on Los Angeles (34.0522° latitude, 118.2437° west longitude), and within 5.0° (in latitude and longitude) of that point. We consider small earthquakes to be those having magnitudes  $M \geq 3.29$  from January 1, 1950 until the present December 31, 2020. Although the catalog is stated to be complete for  $M \geq 3.0$ , the use of a slightly larger minimum magnitude provides additional assurance.

For the coarse-grained time interval  $\Delta t$  as discussed above, we set  $\Delta t = 0.07692$  years, equal to 1/13 years or approximately 1 “lunar month,” equal to 4 “weeks” of length 1/52 years. The reason for this choice is that aftershocks with  $M \geq 3.29$  of large earthquakes in California tend to decay over the time scale of about a month (e.g., Rundle et al., 2021).

As is often the case in these machine learning methods (see, e.g., Rouet-Leduc et al., 2017, 2019), the state vector  $\psi(t)$  consists of a sliding window of length  $\tau = S\Delta t$ , that advances in time by the successive increment  $\Delta t$  on each time step. In other words, small earthquake activity is accumulated over the window length  $\tau$  and assigned to the time  $t$  at the end of the sliding window. As our sliding window, we set  $S = 13$ , thus  $\tau = 1$  year. The reason for this choice is that in the greater California region, large  $M > 6$  earthquakes tend to recur on average about once per year.

We downloaded the earthquake catalog from the USGS web site, collected and filtered the data to construct acceptable timeseries of small earthquakes. The size of the  $N$  coarse grained grid boxes was taken to be 0.33°. Requiring a minimum of 35 small earthquakes over the time period from January 1, 1950 to December 31, 2020, we find  $N_x = 100$  of the spatial grid boxes can be used. We then constructed the correlation matrix (Equation 1) and diagonalized it to find the eigenvalues and eigenvectors. As noted, when we computed the correlation matrix at time  $t$ , we used data only prior to that time.

As an example, we show in Figure 1 the four orthonormal eigenpatterns with the highest correlation values in the correlation matrix, computed for the entire time period January 1, 1950 to December 31, 2020. These eigenpatterns can clearly be recognized by their association with the four largest earthquakes in California during that time period.

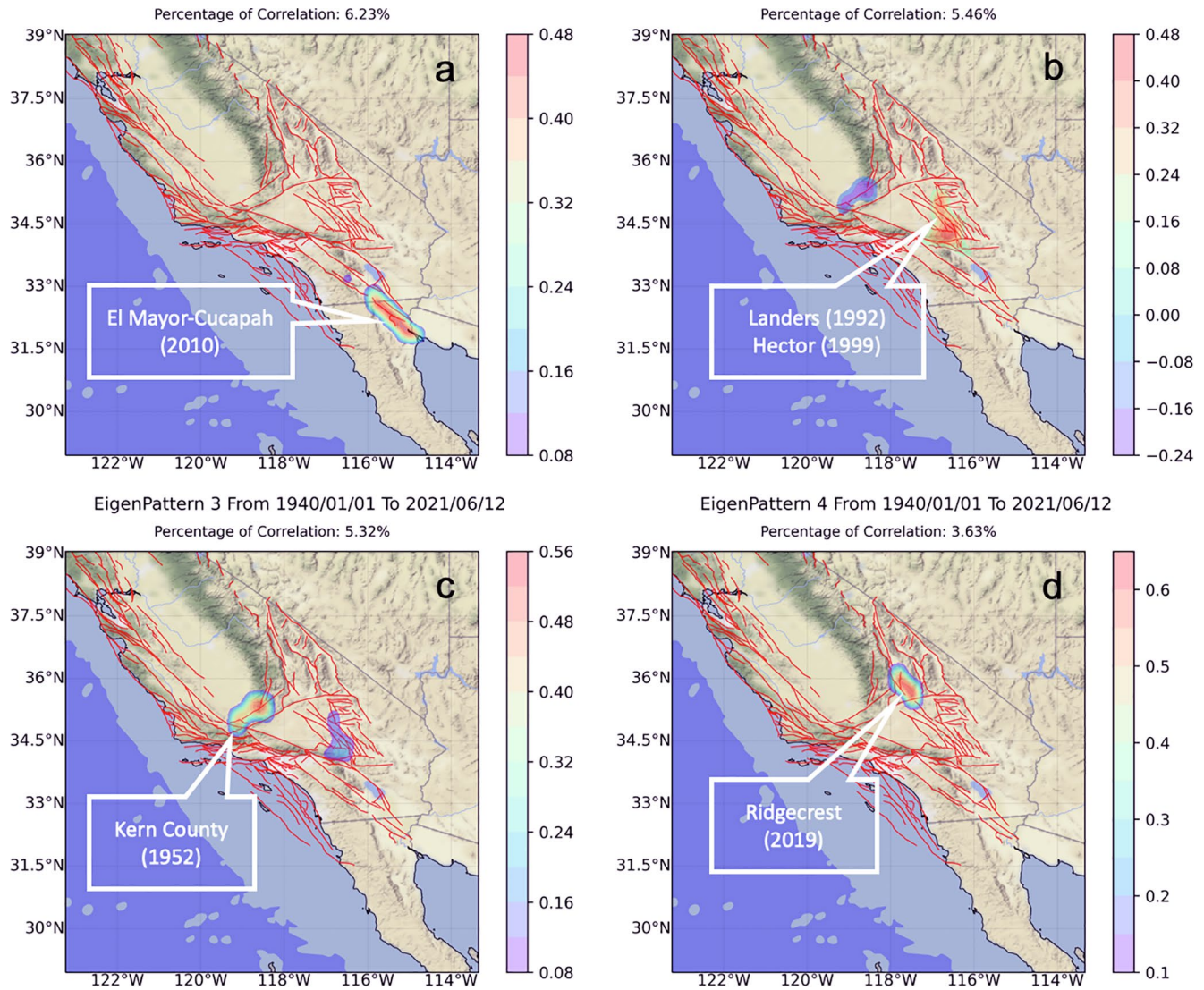
### 4. Signal Detection and Information Content

We now turn to investigate the information contained in the correlation timeseries  $\chi(t)$  that is, shown in Figure 2 from 1984 to December 31, 2020. More specifically, we are interested in determining whether the timeseries contain any information about future large earthquakes. This is basically a problem in signal detection in the presence of noise, which was considered in the 1940s in association with the advent of radar (Green & Swets, 1966; Joy et al., 2005). In that application, the problem was to determine whether an observed signal was actually a true radar return or a random fluctuation.

The researchers introduced the idea of a decision threshold, whereby if the signal had amplitude higher than the threshold, it was classified as a true return (true positive = TP). Of course, even if the signal was large enough, there was still the possibility that it was a random signal (false positive = FP). On the other hand, some returns might have had amplitudes less than the threshold, but still have been real returns (false negative = FN). Or alternatively, they might have been random fluctuations (true negative = TN).

We view our problem as lying in the domain of classification via unsupervised machine learning, sorting potential signals into the categories or classes of true positive (TP), true negative (TN), false positive (FP), and false negative (FN). The standard method (Green & Swets, 1966; Joy et al., 2005) is to construct a receiver operating curve (“ROC”) by plotting the true positive rate (TPR):





**Figure 1.** Top four eigenpatterns having the highest value of regional correlation eigenvalues, near the locations of the M7.2 2010 El Mayor Cucupah; M7.3 1992 Landers + M7.1 1999 Hector Mine; 1952 Kern County; and M7.1 2019 Ridgecrest earthquakes. Small earthquake activity at locations with (hot = reds/ cool = blues) color is *positively correlated* with activity at other (hot/cool) color locations and *anticorrelated* with activity at (cool/hot) color locations. (a) El Mayor Cucupah pattern (6.23% of total correlation). (b) Landers-Hector pattern (5.46% of total correlation). (c) Kern County pattern (5.32% of total correlation). (d) Ridgecrest pattern (3.63% of total correlation). *Note.* that Landers-Hector area appears to be anti-correlated with Kern County area.

$$TPR = \frac{TP}{TP + FN}, \quad (10)$$

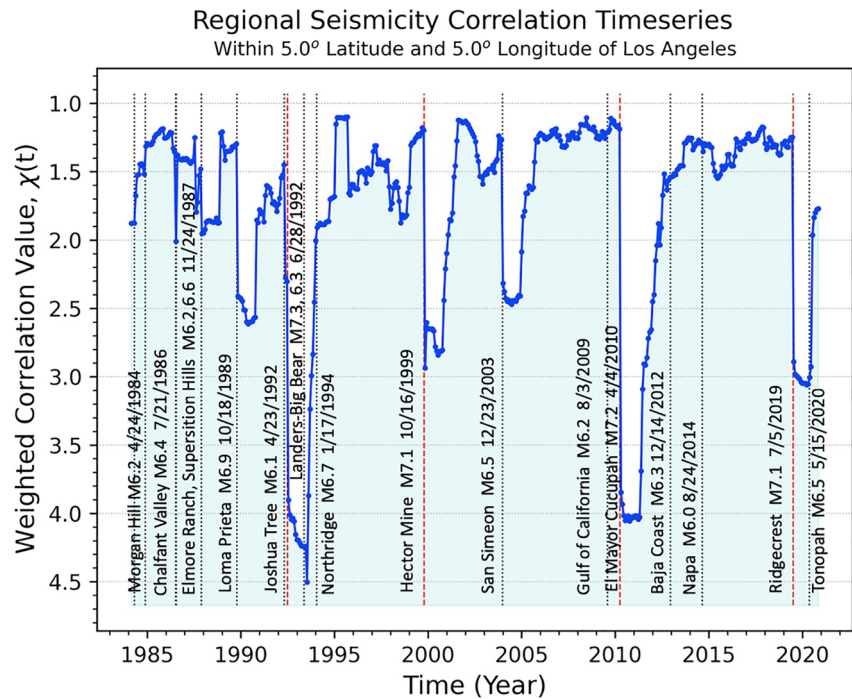
against the false positive rate (FPR), defined in terms of the *specificity* or true negative rate (TNR):

$$FPR = 1 - TNR = \frac{FP}{FP + TN}, \quad (11)$$

TPR is also called the *Recall* or *Hit Rate*, and FPR is also defined as 1 minus the *specificity* or the *False Alarm Rate*. As is well known, Recall measures how well the model performs at correctly predicting positive classes.

On the other hand, PPV or *Precision* measures how well the model performs when the prediction is positive:

$$PPV = \frac{TP}{TP + FP}. \quad (12)$$



**Figure 2.** Weighted correlation value  $\chi(t)$  as a function of time for the California region shown region in Figure 1, computed according to Equations 3 and 4. Vertical dashed red lines represent the four large earthquakes having magnitudes  $M > 7.0$  Vertical black dotted lines represent significant magnitude  $7.0 > M \geq 6.0$  earthquakes as listed in Rundle and Donnellan (2020).

Additionally, ACC or Accuracy measures the fraction of correct predictions, either TP or TN:

$$ACC = \frac{TP + TN}{TP + FN + FP + TN}. \quad (13)$$

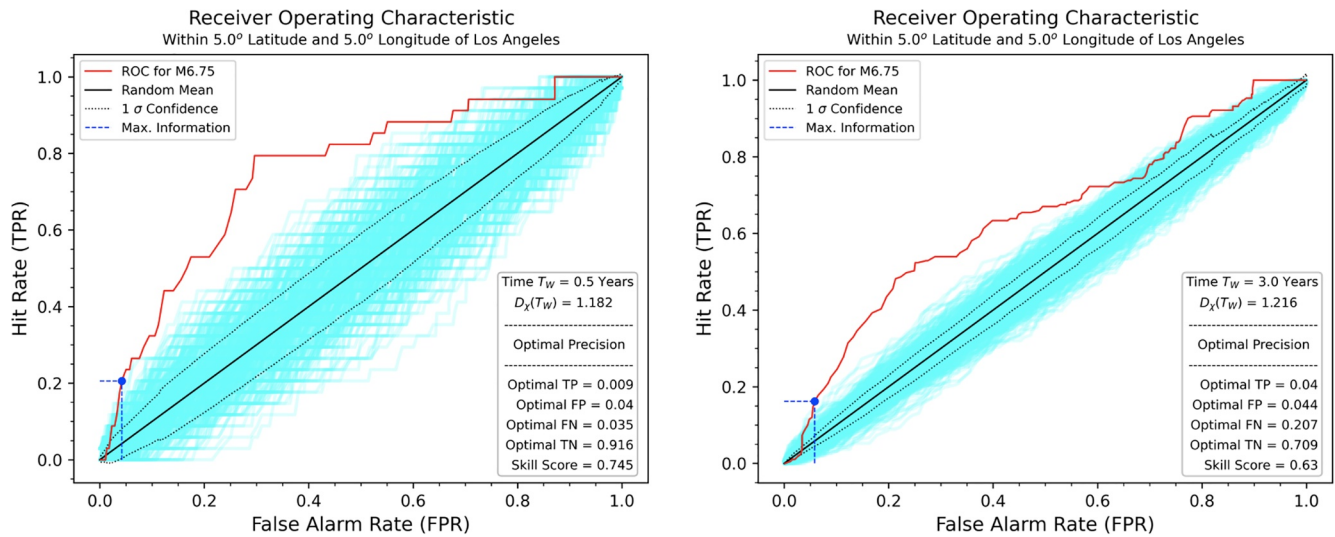
Inspection of the time series  $\chi(t)$  shown in Figure 2 indicates that the largest earthquakes having magnitude  $M \geq M_\lambda$  tend to occur when the correlation timeseries  $\chi(t)$  is near a small minimum value (the “floor”). To proceed, at each time  $t$  we define a future time window  $(t, t + T_w)$ , where  $T_w$  is the duration of the window. We then select an ensemble of decision thresholds  $D_\chi(T_w)$  to test  $\chi(t)$ . The decision thresholds sweep through all possible values to define the ensemble of values for TP, FP, FN, TN.

For each such decision threshold, we accumulate the following statistics. If the condition  $\chi(t) \leq D_\chi(T_w)$  exists, we take this as an indication (“prediction”) that a large earthquake having magnitude  $M \geq M_\lambda$  will occur during the future time window  $(t, t + T_w)$ . On the other hand, if  $\chi(t) > D_\chi(T_w)$ , the “prediction” is that no large earthquake will occur during the future time window. Thus:

1. If  $\chi(t) \leq D_\chi(T_w)$  (“**predicted**”: yes), and the future time window **does contain** at least 1 large earthquake  $M \geq M_\lambda$ , we increment TP  $\rightarrow$  TP + 1, that is., a true positive.
2. If  $\chi(t) \leq D_\chi(T_w)$  (“**predicted**”: yes), and the future time window **does not contain** at least 1 large earthquake  $M \geq M_\lambda$ , we increment FP  $\rightarrow$  FP + 1, that is., a false positive.
3. If  $\chi(t) > D_\chi(T_w)$  (“**predicted**”: no), and the future time window **does contain** at least 1 large earthquake  $M \geq M_\lambda$ , we increment FN  $\rightarrow$  FN + 1, that is., a false negative.
4. If  $\chi(t) > D_\chi(T_w)$  (“**predicted**”: no), and the future time window **does not contain** at least 1 large earthquake  $M \geq M_\lambda$ , we increment TN  $\rightarrow$  TN + 1, that is., a true negative.

Since these the quantities TP, FP, FN, TN only appear in ratios, in the results shown here, we list the quantities TP, FP, FN, TN as normalized by the sum TP + FP + FN + TN, for example:

$$TP \rightarrow \frac{TP}{TP + FP + FN + TN}, \quad (14)$$



**Figure 3.** Red curve is the receiver operating characteristic (ROC) diagram for the time series  $\chi(t)$  shown in Figure 2, obtained by systematically varying the decision threshold  $D_\chi(T_w)$  as described in the text for major earthquakes having magnitudes  $M \geq 6.75$ , and computing the true positive rate (hit rate or recall) TPR and plotting against the false positive rate (1—specificity) FPR. The black diagonal black line from the lower left to the upper right is the random predictor, TPR = FPR. To emphasize that the diagonal line does represent the ROC for a random predictor, we constructed 500 random timeseries by sampling from  $\chi(t)$  with replacement. These are represented by the mass of cyan colored lines in the figures. The  $1\sigma$  confidence level is indicated by the ellipsoidal dotted line enclosing the solid black random predictor line. The blue dashed vertical and horizontal lines represent the values of TPR and FPR obtained by optimizing the precision TP/(TP + FP) for the optimal value of  $D_\chi(T_w)$ . (a) ROC for  $T_w = 6$  months (b) ROC for  $T_w = 3$  years.

etc. Thus all the normalized quantities TP, FP, FN, TN listed here lie within the interval (0, 1).

We construct the ROC curves shown in Figure 3 for the time windows of duration  $T_w = 0.5$  years = 6 months and  $T_w = 3$  years. These are the red curves in the figures, which are the result of the decision threshold systematically ranging over all values, thus sweeping out the values of TPR and FPR.

As is well known (Green & Swets, 1966), a random predictor (no information) is represented by the condition:

$$\text{TPR} = \text{FPR}, \quad (15)$$

shown as the diagonal black line from (0, 0) to (1, 1). To emphasize that the diagonal line does indeed represent the ROC for a random predictor, we constructed 500 random timeseries by sampling from  $\chi(t)$  with replacement. These are represented by the mass of cyan colored lines in the figures. The  $1\sigma$  confidence level is indicated by the ellipsoidal dotted line enclosing the solid black random predictor line.

The fact that the red line lies substantially above the random predictions indicates that there are signals of large earthquakes contained in  $\chi(t)$ . In fact, the area under the red line corresponding to  $\chi(t)$  is a measure of skill, with values lying between (0,1).

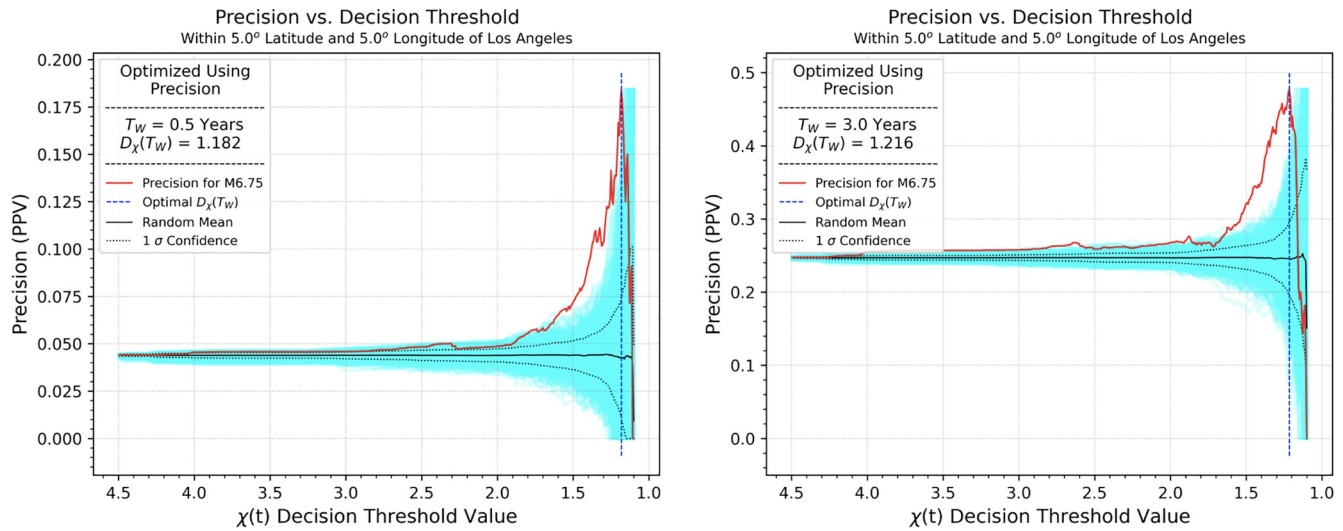
For the random predictor, (black diagonal line) the skill score = 0.5. However, for the 6 months  $T_w$ , the area under the red curve, the skill score = 0.74, indicating skill higher than random. For  $T_w = 3$  years, the skill score = 0.63, implying that skill degrades as  $T_w$  increases, a not unexpected result, although still better than random.

If we were to use the data in the ROC curve in a practical way, we would need to determine the optimal decision threshold, corresponding to an optimal point on the ROC curve for each value of  $T_w$  and  $M_\lambda$ . The possible presence of signals for large earthquakes motivates us to use Shannon information entropy  $I_s$  (Shannon, 1948) as a measure of the information content of points along the ROC curve:

$$I_s = p \log_2 p + (1 - p) \log_2 (1 - p), \quad (16)$$

where  $p$  is an appropriately chosen probability. Thus we are led to seek the value of decision threshold  $D_\chi(T_w)$  that optimizes  $I_s$  for a given value of  $T_w$ .





**Figure 4.** Similar to Figure 3, the red curve in Figure 4 shows the precision  $TP/(TP + FP)$  as a function of the decision threshold  $D_X(T_W)$  applied to the time series  $\chi(t)$  shown in Figure 2. The solid black line is the random predictor. To emphasize that the solid black line does indeed represent the precision for a random predictor, we constructed 500 random timeseries by sampling from  $\chi(t)$  with replacement. These are represented by the mass of cyan colored lines in the figures. The  $1\sigma$  confidence level is indicated by the dotted line enclosing the solid black random predictor line. The blue dashed vertical lines represent the values of the decision threshold obtained by optimizing the Shannon information entropy, using precision as the probability. (a)  $T_W = 6$  months (b)  $T_W = 3$  years.

As an example of this approach, we show in Figures 3 and 4, and Table 1, optimal values for TP, FP, FN, TN that arise from using Equation 16 and the probability measure of precision. In the figures, the optimal values are represented by the vertical dashed blue lines. As mentioned, Figure 3 shows the ROC curves for

**Table 1**  
Comparison of Optimal Data, Obtained by Optimizing the Shannon Information From Entropy of the Precision Variable, to Random Data

Statistic	Definition	Optimal precision		Random data ( $\mu_R \pm \sigma_R$ )		P - Value statistics	
		$T_W = 0.5$ Yrs	$T_W = 3.0$ Yrs	$T_W = 0.5$ Yrs	$T_W = 3.0$ Yrs	$T_W = 0.5$ Yrs	$T_W = 3.0$ Yrs
TP	True Positive	0.009	0.04	$0.002 \pm 0.001$	$0.02 \pm 0.005$	$P < 0.01$	$P < 0.01$
FP	False Positive	0.040	0.044	$0.045 \pm 0.008$	$0.059 \pm 0.009$	$P = 0.27$	$P < 0.05$
FN	False Negative	0.035	0.207	$0.042 \pm 0.001$	$0.227 \pm 0.005$	$P < 0.01$	$P < 0.01$
TN	True Negative	0.916	0.709	$0.911 \pm 0.008$	$0.694 \pm 0.009$	$P = 0.27$	$P < 0.05$
Hit Rate (TPR)	$TP/(TP+FN)$	0.205	0.162	$0.047 \pm 0.035$	$0.078 \pm 0.019$	$P < 0.01$	$P < 0.01$
Specificity (TNR)	$TN/(TN+FP)$	0.958	0.942	$0.953 \pm 0.008$	$0.921 \pm 0.013$	$P = 0.27$	$P < 0.1$
Precision (PPV)	$TP/(TP+FP)$	0.183	0.476	$0.044 \pm 0.032$	$0.244 \pm 0.051$	$P < 0.01$	$P < 0.01$
Accuracy	$TP+TN$	0.925	0.749	$0.913 \pm 0.008$	$0.712 \pm 0.01$	$P < 0.1$	$P < 0.01$
Skill Score	Area Under ROC	0.745	0.630	$0.5 \pm 0.247$	$0.5 \pm 0.113$	$P = 0.16$	$P = 0.12$

Note. Optimal value of decision threshold  $D_X(T_W)$  is found by this procedure. The null hypothesis is that our optimal precision data is generated by a random process. Thus, we compare our values to those generated by the random process and calculate a  $P$ -statistic based on  $Z$ -values using standard procedures.



the two time windows. Figure 4 is a plot of the precision as a function of the decision threshold  $D_\chi(T_w)$  for the two time windows  $T_w$ .

We also optimized the values of these quantities using hit rate (recall) and accuracy, but in general, found the results were not as good as using precision.

## 5. Statistical Tests of Significance

To test whether the information is contained in the time series  $\chi(t)$ , we take as our null hypothesis the idea that any information that may be apparent in  $\chi(t)$  is the result of a purely random process, and that  $\chi(t)$  might be a random time series. Definition of all quantities considered is given in Table 1, columns 1 and 2. Note that TP, FP, FN, TN have been normalized as in Equation 12.

Table 1 also contains the optimal values of the various quantities TP, FP, FN, TN, hit rate, precision, specificity, accuracy, and skill in columns 3 and 4. Columns 5 and 6 in Table 1 display the means and standard deviations for the random set of time series  $\{\chi_R(t)\}$  evaluated at the same particular decision thresholds  $D_\chi(T_w)$  defined previously by optimizing the precision of  $\chi(t)$ . Thus columns 5 and 6 contain the same quantities listed in columns 1 and 2, evaluated for a random predictor.

The random predictor was constructed by means of a bootstrap approach. The time series  $\chi(t)$  was repeatedly sampled randomly with replacement to construct 500 random time series that we can designate as the set of time series  $\{\chi_R(t)\}$ . These random time series are shown as the mass of green lines in Figures 3 and 4.

For all of the statistical quantities identified in column 1, we then compute the Z-statistic:

$$Z = \frac{S - \mu_R}{\sigma_R}, \quad (17)$$

where  $S$  is the statistical quantity (TP, FP, FN, TN, etc.) obtained by optimizing  $\chi(t)$ . The quantities  $\mu_R$  and  $\sigma_R$  are the means and standard deviations of the ensemble of random time series  $\{\chi_R(t)\}$ , also evaluated at the same optimized decision thresholds  $D_\chi(T_w)$ .

From the Z-statistics, we then calculate the  $P$ -values shown in columns 7 and 8 in Table 1. With few exceptions, it can be seen that for the most part,  $P < 0.05$ , a standard criterion for rejecting the null hypothesis at the 95% confidence level. In other words, the observed values of the quantities in column 1 listed in columns 3 and 4 are unlikely to be the result of a random process. There are several exceptions to this general finding for the shorter  $T_w = 6$  months, but for  $T_w = 3$  years, all quantities reject the null hypothesis at the 95% confidence level with the exception of skill score. This latter result may be a reflection of the nature of the earthquake Poisson statistics, which implies a lack of predictability (no memory) and therefore a lack of skill.

## 6. Discussion

Space-time correlations among earthquakes are a topic that has been discussed previously in the literature, but not in the sense discussed in this study. Here we discuss a few examples. Corral (2006) finds universal scaling behavior in earthquake catalogs. He finds that distances between succeeding earthquakes do not depend on the magnitude and are characterized by two jump values, 200 km for world wide behavior, and 15 km for southern California.

Michas et al. (2013) examine seismicity in the West Corinth (Greece) rift, using a data set that covers 2001–2008. Using detrended fluctuation analysis (DFA), they find evidence of fractal distributions in the non-stationary behavior of earthquake activity and show evidence of long-range correlations. Oncel and Wilson (2002) analyze the fractal attributes of seismotectonic variables such as magnitude-frequency, moment-magnitude, and moment-source relations. They relate the fractal properties (scaling exponents) of these distributions yield scaling relations among GR b-values, occurrence rates, and linear diameters of fault planes.

Finally, Mohan and Revathi (2011) quantify correlations between earthquake pairs using a metric developed by Baiesi and Paczuski (2004). They focus on catalogs from California, Japan, and the Himalayas using a network model, using recurrence lengths and times, the two important features in the model. They find a rooted tree structure growing from the largest earthquakes, as one might expect. They further find scaling exponents that illustrate a scaling regime similar to the Omori law describing a faster falloff for larger earthquakes.

With respect to the results of the present paper, we are led to the conclusion that there is evidently information content embedded within  $\chi(t)$ , and that there are optimal decision thresholds that can be determined by a procedure similar to that described above. From a practical perspective, one might imagine that these results might be used to identify signals for optimal threshold values. These could be in the form of “alerts” of future major earthquakes that are declared when  $\chi(t) \leq D_{\chi}(T_w)$  for pre-defined values of  $D_{\chi}(T_w)$ .

We also note that the time series shown in Figure 2 is very similar in character to the ensemble time series (Figure 6) shown in Rundle and Donnellan (2020) using a very different approach that considers seismic bursts. That paper showed that the mean horizontal size given by the radius of gyration  $R_G$  of clusters or bursts of small earthquakes was very large just after a major earthquake, then decreased systematically to a small relatively constant value just prior to the next large earthquake.

We note that for percolation clusters that are frequently studied in statistical physics problems (e.g., Stauffer & Aharony, 2018), the mean radius of gyration of percolation clusters is a measure of the correlation length. Therefore the results described here are consistent with the basic results obtained by Rundle and Donnellan (2020).

Furthermore, the feature shown in Figure 2 is very similar from a physical point of view to that exhibited before mainshocks by the variability of the order parameter of seismicity (introduced in natural time Varotsos et al., 2005) when plotting it versus the conventional time. See, for example, Figures 2 and 3 of the work by Varotsos et al. (2012), revealing that clear changes in the probability distribution of the order parameter of seismicity occur before the three mainshocks: Landers in 1992, Northridge in 1994, and Hector Mine in 1999.

Given the fact that the time series  $\chi(t)$  appears to contain some level of information about the hazard posed by future earthquakes, its use in nowcasting applications would seem to have promise. However, deterministic earthquake prediction using this method does not appear to be possible. Future investigations may allow further refinement and clarification of whatever information this and similar time series contain.

An important question is an extent to which the current method applies to other seismically active regions. A major factor is catalog completeness. For example, the USGS catalog that we use, while basically complete in California at the level that we use in our study,  $M > 3.29$  beginning in 1932, is not complete at that level in many other regions of the world.

A second issue is the time span of the catalog, which in California is about 90 years. In many other regions of the world, the data is not available in the USGS catalog prior to the advent of global seismographs, and in particular, not at the level that we used for this study. In addition, the region of California is particularly simple tectonically since there is one basically dominant fault system (San Andreas), and most importantly, the seismicity is essentially two-dimensional, having very little depth extent (about 15 km in most locations in California). Notwithstanding these limitations, we intend to explore the application of these correlation methods to other active seismic regions in future work.

## Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

## Data Availability Statement

The data used in this study were downloaded from the online earthquake catalog (<https://earthquake.usgs.gov/earthquakes/search/>) maintained by the US Geological Survey, accessed on December 31, 2020.

## Acknowledgments

The research of J. B. Rundle was supported in part by NASA grant (NNX17AI32G) to UC Davis (nowcasting) and in part by DOE grant (DOE DE-SC0017324) to UC Davis (data analysis). Portions of the research by A. Donnellan were carried out at the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration. The research by J. P. Crutchfield was supported by a DOE grant (DOE DE-SC0017324) to UC Davis. The authors thank colleagues including Donald Turcotte for helpful discussions and referees for careful reviews. This research was also supported in part by the Southern California Earthquake Center (Contribution No. 10929). SCEC is funded by NSF Cooperative Agreement EAR-1600087 and USGS Cooperative Agreement G17AC00047.

## References

- Baiesi, M., & Paczuski, M. (2004). Scale-free networks of earthquakes and aftershocks. *Physical Review E—Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 69, 066106. <https://doi.org/10.1103/physreve.69.066106>
- Corral, A. (2006). Universal earthquake-occurrence jumps, correlations with time, and anomalous diffusion. *Physical Review Letters*, 97, 178501. <https://doi.org/10.1103/physrevlett.97.178501>
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. John Wiley and Sons Inc.
- Joy, J. E., Penhoet, E. E., & Petitti, D. B., & Institute of Medicine (US) and National Research Council (US). (2005). *Committee on new approaches to early detection and diagnosis of breast cancer*. National Academies Press (US).
- Luginbuhl, M., Rundle, J. B., Hawkins, A., & Turcotte, D. L. (2018). Nowcasting earthquakes: A comparison of induced earthquakes in Oklahoma and at the Geysers, California. *Pure and Applied Geophysics*, 175(1), 49–65. <https://doi.org/10.1007/s00024-017-1678-8>
- Luginbuhl, M., Rundle, J. B., & Turcotte, D. L. (2018a). Natural time and nowcasting earthquakes: Are large global earthquakes temporally clustered? *Pure and Applied Geophysics*, 175, 661–670. <https://doi.org/10.1007/s00024-018-1778-0>
- Luginbuhl, M., Rundle, J. B., & Turcotte, D. L. (2018b). Natural time and nowcasting induced seismicity at the Groningen gas field in the Netherlands. *Geophysical Journal International*, 215, 753–759. <https://doi.org/10.1093/gji/ggy315>
- Luginbuhl, M., Rundle, J. B., & Turcotte, D. L. (2019). Natural time and nowcasting earthquakes: Are large global earthquakes temporally clustered? In C. Williams, Z. Peng, Y. Zhang, E. Fukuyama, T. Goebel, & M. Yoder (Eds.), *Earthquakes and multi-hazards around the Pacific Rim, Vol. II. PAGEOPH topical Volumes*. Springer.
- Michas, G., Vallianatos, F., & Sammonds, P. (2013). Non-extensivity and long-range correlations in the earthquake activity at the West Corinth rift (Greece). *Nonlinear Processes in Geophysics*, 20, 713–724. <https://doi.org/10.5194/npg-20-713-2013>
- Mohan, T. R. K., & Revathi, P. G. (2011). Earthquake correlations and networks: A comparative study. *Physical Review E*, 83, 046109. <https://doi.org/10.1103/physreve.83.046109>
- Nanjo, K. Z. (2020). Were changes in stress state responsible for the 2019 Ridgecrest, California, earthquakes? *Nature Communications*, 11. <https://doi.org/10.1038/s41467-020-16867-5>
- Olami, Z., Feder, H. J. S., & Christensen, K. (1992). Self organized criticality in a continuous non-conservative cellular automaton modeling earthquakes. *Physical Review Letters*, 68, 1244–1247. <https://doi.org/10.1103/physrevlett.68.1244>
- Oncel, A. O., & Wilson, T. H. (2002). Space-time correlations of seismotectonic parameters: Examples from Japan and from Turkey preceding the Izmit earthquake. *Bulletin of the Seismological Society of America*, 92, 339–349. <https://doi.org/10.1785/0120000844.1>
- Pasari, S. (2019). Nowcasting earthquakes in the Bay-of-Bengal region. *Pure and Applied Geophysics*, 23, 537–559. <https://doi.org/10.1007/s10950-019-09822-5>
- Pasari, S. (2020). Stochastic modeling of earthquake interevent counts (Natural Times) in Northwest Himalaya and adjoining regions. In S. Bhattacharyya, J. Kumar, & K. Ghoshal (Eds.), *Mathematical modeling and Computational Tools. Springer Proceedings in Mathematics & statistics* (Vol. 320, pp. 495–501). Springer. [https://doi.org/10.1007/978-981-15-3615-1\\_35](https://doi.org/10.1007/978-981-15-3615-1_35)
- Pasari, S., & Mehta, A. (2018). Nowcasting earthquakes in the northwest Himalaya and surrounding regions. *The International Society of Archival Photogrammetry and Remote Sensing Spatial Inf. Sci., XLII-5*, 855–859. <https://doi.org/10.5194/isprs-archives-xlii-5-855-2018>
- Pasari, S., & Yogendra, S. (2020). Contemporary earthquake hazards in the West-Northwest Himalaya: A statistical perspective through natural times. *Seismological Society of America*, 91.6, 3358–3369. <https://doi.org/10.1785/0220200104>
- Perez-Oregon, J., Jennifer-Brown, F. A., & Nicholas Vassiliou Sarlis, N. V. (2020). Nowcasting avalanches as earthquakes and the predictability of strong avalanches in the Olami-Feder-Christensen model. *Entropy*, 22, 1228. <https://doi.org/10.3390/e22111228>
- Rouet-Leduc, B., Hulbert, C., & Johnson, P. A. (2019). Continuous chatter of the Cascadia subduction zone revealed by machine learning. *Nature Geoscience*, 12, 75–79. <https://doi.org/10.1038/s41561-018-0274-6>
- Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C. J., & Johnson, P. A. (2017). Machine learning predicts laboratory earthquakes. *Geophysical Research Letters*, 44(18), 9276–9282. <https://doi.org/10.1002/2017gl074677>
- Rundle, J., Stein, S., Donnellan, A., Turcotte, D. L., Klein, W., & Saylor, C. (2021). The complex dynamics of earthquake fault systems: New approaches to forecasting and nowcasting of earthquakes. *Reports on Progress in Physics*, 84, 076801. <https://doi.org/10.1088/1361-6633/abf893>
- 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, e2020EA0010. <https://doi.org/10.1029/2020ea001097>
- Rundle, J. B., Donnellan, A., Grant Ludwig, L., Gong, G., Turcotte, D. L., & Luginbuhl, M. (2016). Nowcasting earthquakes. *Earth and Space Science*, 3, 480–486. <https://doi.org/10.1002/2016ea000185>
- 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, 456–472. <https://doi.org/10.1029/2018ea000464>
- Rundle, J. B., Luginbuhl, M., Giguere, A., & Turcotte, D. L. (2018). Natural time, nowcasting and the physics of earthquakes: Estimation of risk to global megacities. *Pure and Applied Geophysics*, 175, 647–660. <https://doi.org/10.1007/s00024-017-1720-x>
- Sarlis, N. V., Skordas, E. S., & Varotsos, P. A. (2018). A remarkable change of the entropy of seismicity in natural time under time reversal before the super-giant M9 Tohoku earthquake on 11 March 2011. *Europhysics Letters*, 124. <https://doi.org/10.1209/0295-5075/124/29001>
- Scholz, C. H. (2019). *The mechanics of earthquakes and faulting*. Cambridge University Press.
- Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27, 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>
- Stauffer, D., & Aharony, A. (2018). *Introduction to percolation theory*. CRC press.
- Tiampo, K. F., Rundle, J. B., Gross, S. J., McGinnis, S., & Klein, W. (2002). Eigenpatterns in southern California seismicity. *Journal of Geophysical Research*, 107(B12), 2354. <https://doi.org/10.1029/2001jb000562>
- Tiampo, K. F., Rundle, J. B., McGinnis, S. A., & Klein, W. (2002). Pattern dynamics and forecast methods in seismically active regions. *Pure and Applied Geophysics*, 159, 2429–2467. <https://doi.org/10.1007/s00024-002-8742-7>
- Tullis, T. E., Richards-Dinger, K., Barall, M., Dieterich, J. H., Field, E. H., Heien, E. M., et al. (2012). Generic earthquake simulator. *Seismological Research Letters*, 83, 959–963. <https://doi.org/10.1785/0220120093>
- Varotsos, P., Sarlis, N. V., & Skordas, E. S. (2001). Spatiotemporal complexity aspects on the interrelation between seismic electric signals and seismicity. *Practica of Athens Academy*, 76, 294–321.
- Varotsos, P., Sarlis, N. V., & Skordas, E. S. (2002). Long-range correlations in the electric signals that precede rupture. *Physical Review E—Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics*, 66, 011902. <https://doi.org/10.1103/physreve.66.011902>
- 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-Verlag.

- Varotsos, P., 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, 9192–9206. <https://doi.org/10.1002/2014ja020580>
- Varotsos, P., Sarlis, N. V., Skordas, E. S., & Lazaridou, M. S. (2013). Seismic electric signals: An additional fact showing their physical interconnection with seismicity. *Tectonophysics*, 589, 116–125. <https://doi.org/10.1016/j.tecto.2012.12.020>
- Varotsos, P. A., Sarlis, N. V., & Skordas, E. S. (2012). Remarkable changes in the distribution of the order parameter of seismicity before mainshocks. *EPL (Europhysics Letters)*, 100(3), 39002. <https://doi.org/10.1209/0295-5075/100/39002>
- Varotsos, P. A., Sarlis, N. V., & Skordas, E. S. (2020). Self-organized criticality and earthquake predictability: A long-standing question in the light of natural time analysis. *EPL (Europhysics Letters)*, 132, 29001. <https://doi.org/10.1209/0295-5075/132/29001>
- Varotsos, P. A., Sarlis, N. V., Tanaka, H. K., & Skordas, E. S. (2005). Similarity of fluctuations in correlated systems: The case of seismicity. *Physical Review E*, 72(4), 041103. <https://doi.org/10.1103/physreve.72.041103>
- 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. *Europhysics Letters*, 130. <https://doi.org/10.1209/0295-5075/130/29001>