

Nowcasting Earthquakes by Visualizing the Earthquake Cycle with Machine Learning: A Comparison of Two Methods

by

John B Rundle^{1,2,3}, Andrea Donnellan², Geoffrey Fox⁴ and James P. Crutchfield¹

¹University of California, Davis, CA

²Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA

³Santa Fe Institute, Santa Fe, NM

⁴Indiana University, Bloomington, IN

Abstract

The earthquake cycle of stress accumulation and release is associated with the elastic rebound hypothesis proposed by H.F. Reid following the M7.9 San Francisco earthquake of 1906. However, observing details of the actual values of time- and space-dependent tectonic stress is not possible at the present time. In two previous papers, we have proposed methods to image the earthquake cycle in California by means of proxy variables. These variables are based on correlations in patterns of small earthquakes that occur nearly continuously in time. The purpose of the present paper is to compare these two methods by evaluating their information content using decision thresholds and Receiver Operating Characteristic methods together with Shannon information entropy. Using seismic data from 1950 to present in California, we find that both methods provide nearly equivalent information on the rise and fall of earthquake correlations associated with major earthquakes in the region. We conclude that the resulting timeseries can be viewed as proxies for the cycle of stress accumulation and release associated with major tectonic activity.

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 (Rundle et al., 2016, 2018, 2019, 2020; Pasari and Mehta, 2018; Pasari, 2019, 2020; Pasari and Sharma, 2020; Luginbuhl et al. 2019; 2020). 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 et al. (2021). 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 has been based on the concept of natural time (Varotsos et al., 2001; 2002; 2011, 2013; 2014; 2020a,b; Sarlis et al., 2018). 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 Rundle and Jackson (1977) and Olami-Feder Christensen (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 paper, Rundle and Donnellan (2020) showed that a timeseries resembling the long-hypothesized earthquake cycle could be constructed from the time dependence of horizontal radius of gyration $R_G(t)$ of bursts of small earthquakes that are clustered in space and time. In fact, the quantity $R_G(t)$ is often used in the calculation of correlation length for models in statistical mechanics (Rundle and Donnellan, 2020). In a subsequent paper,

Rundle et al. (2021) have developed an alternate method based on constructing a correlation time series $\chi(t)$ of small earthquakes in the seismically active region of California. Both of these methods use patterns of small earthquakes, and the correlations among them, to define proxy timeseries that have many of the characteristics expected of the cycle of tectonic stress accumulation and release.

To summarize our results: The two timeseries so defined imply that regional correlation of seismic activity generally decreases prior to major earthquakes in California. Just after occurrence a major earthquake, correlation of seismic activity discontinuously increases, as does the quantity $R_c(t)$. Both of the resulting timeseries strongly resemble the expected earthquake cycle of stress accumulation and release. We then applied standard timeseries methods based on constructing Receiver Operating Characteristic (ROC) diagrams 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, in both $R_c(t)$ and $\chi(t)$.

Methods

Summary of the Radius of Gyration $R_c(t)$ Method. We begin with our definition of a seismic burst, or cluster of small events. Our definition of a seismic burst is the occurrence of an unusual sequence of generally small earthquakes closely clustered in space and time (e.e., Hill and Prejean, 2007; Peresan and Gentili, 2018; Zaliapin and Ben-Zion, 2016a,b).

We define two general types of bursts, Type I and Type II:

- We define a Type I seismic burst as a mainshock-aftershock sequence, in which the initiating event has the largest magnitude in the sequence, and is typically followed by a power-law Omori decay of occurrence of smaller events (Omori, 1900; Scholz, 2019).
- A Type II seismic burst is defined as a sequence of similar magnitude events in which the largest magnitude event is not the initiating event, and in which there is not typically a subsequent power-law decay.

The earthquakes defining the bursts are small, usually of magnitudes characterizing the catalog completeness level. For the Southern California region, we consider small earthquakes of magnitudes $M \geq 3.3$. This magnitude threshold was chosen as a value high enough to ensure completeness of the catalog data used. The catalog containing these events is downloaded from the US Geological Survey earthquake search database¹. The method we describe proceeds in 5 stages, and only a summary is provided here. For a more detailed discussion, we refer readers to Rundle and Donnellan (2020).

The first stage consists of an automated definition and classification of seismicity into candidates of seismic bursts. The second stage involves automated rejection of outliers. The third stage selects the members of the ensemble of accepted bursts which will then be

displayed as a time series. The fourth stage applies an exponential moving average to the bursts to construct the burst time series. The fifth stage involves optimization of the ensemble of possible bursts with a simple cost function. We note that both classification/clustering and optimization/regression are well known components of new ideas in machine learning, along with other ideas in deep learning and decision tree analysis².

We begin by coarse-graining time in the catalog into units of single days, and consider an elementary burst to be a day on which there are 2 or more small earthquakes of magnitude $M \geq 3.3$ within the region of interest, which for this study is the 600 km radius Southern California region. Note that, over the last 10 years, the rate of occurrence of these small earthquakes have been about 0.75 such earthquakes per day in that geographic region.

In Stage I, classification, the daily seismic catalog is searched to find bursts consisting of connected sequences of days in which 2 or more $M \geq 3.3$ events occur without any intervening days of fewer events. For each such set of days, we also, as a rule, include the preceding day to allow for any foreshock events.

This stage yields many hundreds of candidate bursts. This process will of necessity yield bursts that include purely random, uncorrelated events. To remove these, the bursts are then filtered in the following two ways.

In Stage II of the method, rejection of outliers, we detect and remove small earthquakes that may be random outliers. We begin by computing the spatial centroid, or center-of-mass, of each burst. In this calculation, all events having $M \geq 3.3$ are treated as a particle or unit of mass, each of equal computational weight.

We now compute the horizontal distance or radius (" R_i ") of each small event from the centroid, then the median distance ("MedianR") is calculated from the set $\{R_i\}$. A factor F_{CL} is defined and applied to each of the candidate bursts. Using all the accepted small events in the burst, the burst radius-of-gyration R_G is computed about the burst centroid. R_G is the square root of the mean square radius of the small events in the burst. These filtered bursts now define the ensemble of accepted clusters. Radius of gyration is a parameter used to study fracture mechanics (e.g. Kuchеров and Ryvkin, 2014; Sayers and Calvez, 2010).

In Stage III of the method, the collection of bursts is filtered according to their mass ratio or density ρ , which we define as the ratio of the cluster mass μ to the radius-of-gyration R_G , $\rho = \mu / R_G$. Mass is defined as the number of small events in the cluster or burst.

To implement this filter, we define a filter or threshold value corresponding to a particular value of mass ratio ρ . Each burst is tested, with the criterion for *acceptance* being that the density is greater than the threshold value. With this condition, we accept only high-density clusters, which are typically the most compact clusters. Clusters that are accepted by this criterion correspond to long wavelength fluctuations in the time series, so this condition represents a low-pass filter (Rundle and Donnellan, 2020).

In Stage IV, an exponential moving average² (EMA) was applied to the filtered burst time series data. The choice to be made with this method is the value of N , the number of averaging steps. For our purposes, a 1-year averaging interval was adopted for the temporal resolution, corresponding to an average of $N \sim 23$ bursts per year.

In Stage V of the approach, the collection or ensemble of the bursts was optimized and combined into a single time series using a simple cost function. The result of this stage is an ensemble in which the largest earthquakes of $M \geq 7$ occur at approximately the same value of R_G for each event. The strategy involves defining a cost function that seeks to optimize the value of radius of gyration R_G for the largest earthquakes $M \geq 7$, just before they occur. The cost function that used requires that the radius of gyration of these large earthquakes just prior to failure be a relatively uniform value. This would allow a crude nowcast or forecast of when a large such earthquake might occur in the future.

In general, it was found that there is a recharge period where average $R_G(t)$ decreases prior to each magnitude $M \geq 7$, followed by a sudden discharge where $R_G(t)$ increases in average due to the large aftershock bursts following the mainshock. Between these large mainshocks, it can be seen that lesser magnitude earthquakes result in lesser but similar effects.

For the methods described here, we downloaded the earthquake catalog from the USGS web site, collected and filtered the data to construct acceptable timeseries of small earthquakes.¹ The largest of these earthquakes since 1994 are given in Table 1. As examples of bursts that were used in computing $R_G(t)$, we show in Figure 1 four bursts of small earthquakes in California. The top two represent moderate sized bursts, the bottom two represent aftershocks of major earthquakes $M \geq 6.0$

Summary of the Correlation Timeseries $\chi(t)$ Method. This method begins by defining a spatial coarse graining, assigning an array of grid boxes of given latitude and longitude $\Delta \mathbf{x}$ (in degrees) to the area of interest. Each of these grid boxes (tiles or partitions) is required to have a minimum number of small earthquakes over the entire time interval used. This procedure produces a set of N_X "active" grid boxes.

We then extract data from the seismic catalog, which is the set of values $\{t_i, M_i, \mathbf{z}_i\}$, 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 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 .

The catalog is then digitized in time at increments Δt . A given earthquake is then assigned 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 Δt , extending over the interval $t_0, \dots, (t_0 + \Delta t J_T)$. In 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. Principal Component Analysis (PCA) is used to analyze the correlation matrix $C_{nm}(t)$, which involves centered, univariant time series $\hat{\Phi}(\mathbf{x}_n, t)$. $\hat{\Phi}(\mathbf{x}_n, t)$ is obtained from the timeseries $\Phi(\mathbf{x}_n, t)$ by removing the mean and normalizing to unit variance (Rundle et al., 2021). $C_{nm}(t)$ 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:

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

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

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

In computing $\mathbf{e}_i(t)$, only data for $t' \leq t$ are used.

The weighted correlation of the seismicity at time t is then found as the dot product of the power spectrum $a_i(t)^2$ 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 Weighted Correlation Timeseries (WCT) containing (possibly) significant information content.

In computing (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 , both $\lambda(t)$ and $a_i(t)^2$ were normalized to standard values of, respectively, 100 and 1 (Rundle et al., 2021).

To compute the timeseries $\chi(t)$, we use construct a state vector $\boldsymbol{\psi}(t)$. As is often the case in these machine learning methods (see, e.g., Rouet-Leduc et al., 2017, 2019), $\boldsymbol{\psi}(t)$ consists

of a sliding window of length $\tau = S\Delta t$, that advances in time by the successive increment Δt on each time step. In other words, small earthquake activity is accumulated over the window length τ 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.

At each time t , $\psi(t)$ is expanded in the eigenpatterns and the coefficients of expansion $a_i(t)$ are computed. The current eigenvalues $\lambda_i(t)$ are then used as in equation (7) to compute the value of $\chi(t)$ at that time step. We plot $\chi(t)$ as a function of time as shown below, which we interpret as a nowcasting correlation timeseries.

Comparison of The Two Methods.

We have applied both methods to California, and summarize previous results here, together with some new analysis. We begin by focusing on 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 1/1/1950 until 12/31/2020. For the time interval Δt as discussed above, we set $\Delta t = 0.07692$ year, equal to 1/13 year or approximately 1 "lunar month", equal to 4 "weeks" of length 1/52 year.

For the method used to compute the WCT $\chi(t)$, 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 1/1/1950 to 12/31/2020, we find $N_x = 100$ of the spatial grid boxes can be used. We then constructed the correlation matrix (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.

In Figure 2, we show four orthonormal eigenpatterns with the high correlation values in the correlation matrix, computed for the entire time period 1/1/1950 to 12/31/2020. These eigenpatterns can clearly be recognized by their association with the four largest earthquakes in California during that time period. Again for reference, Table 1 lists the large earthquakes having magnitude $M \geq 6$ from 1984 - present.

Timeseries. For both timeseries methods, we compute the timeseries, $R_G(t)$ and $\chi(t)$, by the methods described above. These are shown in Figure 3, over the time periods from 1984-present, where the red dashed vertical lines represent earthquakes having magnitudes larger than $M \geq 6.75$, and the black dotted vertical lines represent earthquakes having magnitudes $6.0 \leq M < 6.75$. The green dashdot line ("decision threshold" $D\chi(T_w)$) will be discussed in the following.

Note that the time series values on the vertical axis are plotted on an inverted scale, so that the smallest values of $R_G(t)$ and $\chi(t)$ are at the top of the figure, and the largest values at the bottom. As a result, the two timeseries resemble the hypothesized cycle of regional

stress accumulation and release that is thought to be associated with the elastic rebound hypothesis of H.F. Reid (Scholz, 2019). These time series might then be viewed as a proxy for the regional tectonic stress cycle.

Since $R_G(t)$ is often used in statistical mechanics as a measure of correlation length, Figure 3 implies that just after a major earthquake occurs, correlation length suddenly increases. Then, over the subsequent period of time, correlations gradually decrease, eventually reaching a low value prior to the next major earthquake. This cycle of sudden correlation increase at the time of a major earthquake, is presumably associated with a sudden decrease in regional tectonic stress. The gradual decrease of correlation following a major event, is then apparently associated with the preparation and stress buildup leading to the next large earthquake. In any case, the two time series resemble the expected behavior of the cycle of regional tectonic stress in seismically active areas.

Time Series Prediction. Given a time series such as those in Figure 3, there are a number of machine learning models and techniques that have been developed to predict future behavior of the time series, given a period of training of the parameters in the models. One of these that we briefly describe is the one-step walk-ahead method based on the random forest algorithm².

In this method, the data set of values $V \in \{V_i\}, i = 1, \dots, N_i$, are partitioned into a training set and a test set, typically 25%-75% training, with the remainder in the test set. The parameters of model are then adjusted, and the result is checked by application to the test set. The idea is that the model parameters are assigned using a sliding window of N_P time series values, resulting in a the prediction for the next (unknown) value of the time series. In the examples shown below, we use $N_P = 13$ monthly values of the time series to predict the next value.

We show time series prediction of the time series for $R_G(t)$ and $\chi(t)$ in Figures 4 and 5, respectively. The left panel in both figures is shown for the entire time period since 1984, with the original blue curve representing the time series, and the red superposed curve representing the time series prediction. The right panel of each figure is a zoom closeup of the last ~ 2.5 years of the left side panel, encompassing the time period of the M7.1, July 5, 2019 Ridgecrest earthquake.

Superficially, from the left panels, both figures show that the red prediction curve appears to predict the next time series values reasonably well. However, a closer look represented by the right panel indicates that the prediction algorithm does not predict the onset of the large earthquakes, but in fact shows a delayed response, delayed by one time step.

Signal Detection and Information Content

Decision Thresholds. We now turn to investigating the information contained in the correlation timeseries $R_G(t)$ and $\chi(t)$ that is shown in Figure 3 from 1984 to 12/31/2020. More specifically, we are interested in determining whether those 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 1940's in association with the advent of radar (Green and 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, where 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 threshold, but still have been real returns (false negative = FN). Or alternatively, they might have been random fluctuations (true negative = TN).

The problem is to determine whether signals of future large earthquakes can be detected by analysis of $R_G(t)$ and $\chi(t)$. 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 and Swets, 1966; Joy et al., 2005) is then to construct a Receiver Operating Curve ("ROC") by plotting the true positive rate (TPR):

$$TPR = TP/(TP + FN) \quad (5)$$

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

$$FPR = 1 - TNR = FP/(FP + TN) \quad (6)$$

TPR is also called the *Recall* or *Hit Rate*, and FPR is also defined as 1-*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:

$$PPV = TP/(TP + FP) \quad (7)$$

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

$$ACC = (TP + TN)/(TP + FN + FP + TN) \quad (8)$$

Inspection of the time series $R_G(t)$ and $\chi(t)$ shown in Figure 3 indicates that the largest earthquakes having magnitude $M \geq M_\lambda$ tend to occur when the correlation timeseries $R_G(t)$ and $\chi(t)$ are 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 $R_G(t)$ and $\chi(t)$. 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 $\{R_G(t), \chi(t)\} \leq D_\chi(T_W)$ 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 $\{R_G(t), \chi(t)\} > D_\chi(T_W)$, the "prediction" is that no large earthquake will occur during the future time window. Thus:

- If $\{R_G(t), \chi(t)\} \leq D_\chi(T_W)$ ("**predicted**": **yes**), and the future time window **does contain** at least 1 large earthquake $M \geq M_\lambda$, we increment $TP \rightarrow TP+1$. i.e, a true positive.
- If $\{R_G(t), \chi(t)\} \leq D_\chi(T_W)$ ("**predicted**": **yes**), and the future time window **does not contain** at least 1 large earthquake $M \geq M_\lambda$, we increment $FP \rightarrow FP+1$. i.e, a false positive.
- If $\{R_G(t), \chi(t)\} > D_\chi(T_W)$ ("**predicted**": **no**), and the future time window **does contain** at least 1 large earthquake $M \geq M_\lambda$, we increment $FN \rightarrow FN+1$. i.e, a false negative.
- If $\{R_G(t), \chi(t)\} > D_\chi(T_W)$ ("**predicted**": **no**), and the future time window **does not contain** at least 1 large earthquake $M \geq M_\lambda$, we increment $TN \rightarrow TN+1$. i.e, 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 $TP+FP+FN+TN$, e.g.:

$$TP \rightarrow TP/(TP + FP + FN + TN) \quad (9)$$

etc. Thus all the normalized quantities TP, FP, FN, TN listed here lie within the interval $[0,1]$.

Receiver Operating Characteristic. Figure 6 shows a comparison of the Receiver Operating Characteristics (ROC) diagrams for $R_G(t)$ and $\chi(t)$, obtained by plotting TPR against FPR . The plot for $R_G(t)$ is at left, and that for $\chi(t)$ is at right. These diagrams are computed for the time window $T_W = 3$ years. The red curve is the ROC curve in both panels.

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

$$TPR = FPR \quad (10)$$

On the ROC plots for $R_G(t)$ and $\chi(t)$, this line is 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 computed 500 random timeseries by sampling from $R_G(t)$ and $\chi(t)$ with replacement (bootstrap method). These are represented by the mass of cyan-colored lines in the figures. The 1σ confidence level is indicated by the ellipsoidal dotted line enclosing the solid black random predictor line.

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

For the random predictor, (black diagonal line) the skill score = 0.5. It can be seen that the area under the red curve in both figures evidently has more area line beneath it, indicating skill better than random. Values of the skill score for both $R_G(t)$ and $\chi(t)$ are shown in in the figure and in Table 2.

Optimal Decision Thresholds. 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 the large earthquakes to be nowcasted. The possible presence of signals for large earthquakes motivates us to use Shannon information entropy I_S (Shannon, 1948) as a measure of information content of points along the ROC curve:

$$I_S = -p \log_2 p - (1 - p) \log_2 (1 - p) \quad (11)$$

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 .

As an example of this approach, we show in Figures 6, and Table 2, the optimal values for TP , FP , FN , TN that arise from using the equation (11) and the probability measure of *precision*. In Figure 6, the optimal values are represented by the vertical dashed blue lines. Figure 7 is a plot of the precision as a function of the decision threshold $D_\chi(T_W)$ for the same time window $T_W = 3$ years, corresponding to the ROC plots of Figure 6.

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.

Statistical Tests of Significance

To test whether information is contained in the time series $R_G(t)$ and $\chi(t)$, we take as our null hypothesis the idea that any information that may be apparent in $R_G(t)$ and $\chi(t)$ is the result of a purely random process, and that $R_G(t)$ and $\chi(t)$ might be a random time series. Definition of all quantities considered is given in Table 2, columns 1 and 2. Note that TP , FP , FN , TN have been normalized as in equation (7).

Table 2 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 for $R_G(t)$ and $\chi(t)$. Columns 5 and 6 in Table 2 display the means and standard deviations for the random set of time series $\{R_{G,R}(t), \chi_R(t)\}$ evaluated at the same particular decision thresholds $D_\chi(T_W)$ defined previously by optimizing the precision of $R_G(t)$ and $\chi(t)$. Thus columns 5 and 6 contain the same statistical quantities listed in columns 1 and 2, evaluated for a random predictor.

As stated, the random predictor was constructed by means of a bootstrap approach. The time series $R_G(t)$ and $\chi(t)$ were repeatedly sampled randomly with replacement to construct 500 random time series that we can designate as the set of time series $\{R_{G,R}(t), \chi_R(t)\}$.

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

$$Z = \frac{S - \mu_R}{\sigma_R} \quad (10)$$

where S is the statistical quantity (TP , FP , FN , TN , etc.) obtained by optimizing $\chi(t)$. The quantities μ_R and σ_R are the means and standard deviations of the ensemble of random time series $\{R_{G,R}(t), \chi_R(t)\}$, 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 2. 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 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.

Discussion

We are led to the conclusion that there is evidently information content embedded within $R_G(t)$ and $\chi(t)$, 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 $\{R_G(t), \chi(t)\} \leq D_\chi(T_W)$ for pre-defined values of $D_\chi(T_W)$.

From examination of Table 2, columns 7 and 8, it would appear that the time series $R_G(t)$ performs somewhat better than $\chi(t)$. It should be noted that the time series $R_G(t)$ has been optimized using machine methods, while $\chi(t)$ has not, which may account for the difference. This is a subject for future work.

Given the fact that the time series $R_G(t)$ and $\chi(t)$ appear to contain some level of information about the hazard posed by future earthquakes, its use in nowcasting applications would seem to have promise. Future investigations may allow further refinement and clarification of whatever information this and similar time series contain.

The elastic rebound theory of earthquakes (Richter, 1958) proposes that tectonic stresses build up, recharge or increase, in a region following a large earthquake until another large earthquake occurs and stress discharges or decreases. At that point the stresses are suddenly reduced, and a new cycle of stress recharge and discharge begins. By presenting our results in the manner shown in Figures 3-6, the similarity with the elastic rebound theory can be seen.

Our present results contribute to the development of seismic nowcasting methods that we have discussed earlier (Rundle et al., 2016; 2018; 2019a,b). In previous methods, elastic rebound is introduced as a constraint, by counting small earthquakes since the last large earthquake. In contrast using this method, elastic rebound emerges naturally, in that it follows directly from time-dependent properties of the bursts. Another difference is that the seismic nowcasting method produces a cumulative distribution function, or equivalently a survivor distribution of future large earthquake activity. By contrast, the present method computes an observable property of the region with a clear physical meaning.

Other studies have shown that large earthquakes tend to occur in relatively small regions where small earthquake activity has been the greatest for a number of years (Rundle et al., 2003; Tiampo et al., 2002a,b,c; Holliday et al., 2006a,b; 2007; 2008). This is essentially a consequence of the universal applicability of the Gutenberg-Richter relation (Rundle et al., 2016; 2018). The RELM earthquake forecasting test suggests that this approach may be fruitful (Holliday et al., 2007; Lee et al., 2011)A

A strategy to anticipate major earthquakes that combines methods such as those proposed by (e.g., Rundle et al., 2003) to estimate candidates for spatial locations of future events, combined with the ensemble time series methods discussed here, might be useful to consider. A question that remains is the applicability of the methods described here to other seismically active regions, which in turn depends on the completeness of the seismic catalog over a wide range of magnitudes. A major advantage of the Southern California region is that the catalog is complete to small magnitudes, a property that is not generally the case

elsewhere. Future work will be directed at answering this question as it will be important to test the method in other seismically active areas.

Acknowledgements

The research of JBR 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 Andrea 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 JPC was supported by DOE grant (DOE DE-SC0017324) to UC Davis. None of the authors have identified financial conflicts of interest. We thank colleagues including Donald Turcotte for helpful discussions. The data used in this paper was downloaded from the online earthquake catalog¹ maintained by the US Geological Survey, accessed on December, 31, 2020. 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 & USGS Cooperative Agreement G17AC00047.

Notes

[1] <https://earthquake.usgs.gov/earthquakes/search/>

[2] <https://machinelearningmastery.com/random-forest-for-time-series-forecasting/>

References

- Green, David M.; Swets, John A. (1966). Signal detection theory and psychophysics. New York, NY: John Wiley and Sons Inc. ISBN 978-0-471-32420-1.
- Hill DP and Prejean SG, Dynamic triggering, in Earthquake seismology, vol. 4, G. Schubert, H. Kanamori, Eds. *Treatise on Geophysics*, Schubert, G (Editor-in Chief) (Elsevier, Amsterdam), pp. 257–291 (2007).
- Holliday, JR, Turcotte, DL, and Rundle, JB, Self-similar branching of aftershock sequences, *Physica* **A387**, 933-943 (2008).
- Holliday, JR, Chen, CC, Tiampo, KF, Rundle, JB, Turcotte, DL, and Donnellan, A, A RELM earthquake forecast based on Pattern Informatics, *Seism. Res. Lett.*, 78, 87-93 (2007).
- Holliday, JR, Rundle, JB, Tiampo, KF, and Donnellan, A systematic procedural and sensitivity analysis of pattern informatics method for forecasting large ($M > 5$) earthquake events in southern California, *Pure Appl. Geophys.*, 10.1007/s00024-006-0131-1 v. 163, No. 11-12, 2433-2454 (2006a).
- Holliday, JR, Rundle, JB, Turcotte, DL, Klein, W. and Tiampo, KF, Space-time correlation and clustering of major earthquakes, *Phys. Rev. Lett.*, 97, 238501 (2006b)

- Joy JE, Penhoet EE, Petitti DB, eds, Institute of Medicine (US) and National Research Council (US) Committee on New Approaches to Early Detection and Diagnosis of Breast Cancer. Washington (DC): National Academies Press (US); 2005.
- Kuchеров, L. and Ryvkin, M., 2014. Fracture toughness of open-cell Kelvin foam. *International Journal of Solids and Structures*, 51(2), pp.440-448.
- Lee, YT, Turcotte, DL, Holliday, JR, Sachs, MK, Rundle, JB, Chen, CC, and Tiampo, KF, Results of the Regional Earthquake Likelihood Models (RELM) test of earthquake forecasts in California, *Proc. Nat. Acad. Sci USA*, **108**, 16533-16538 (2011) DOI: 10.1073/pnas.1113481108
- Luginbuhl, M., Rundle, J.B. and Turcotte, D.L. Natural time and nowcasting earthquakes: are large global earthquakes temporally clustered? *Pure Appl. Geophys.*, 175, 661-670 (2018b).
- Luginbuhl, M., Rundle, J.B. and Turcotte, D.L. Natural time and nowcasting induced seismicity at the Groningen gas field in the Netherlands. *Geophys. J. Int.*, 215, 753-759 (2018c).
- Luginbuhl, M., Rundle, J.B. and Turcotte, D.L. Natural Time and Nowcasting Earthquakes: Are Large Global Earthquakes Temporally Clustered? In: Williams, C., Peng, Z., Zhang, Y., Fukuyama, E., Goebel, T., Yoder, M. (eds) *Earthquakes and Multi-hazards Around the Pacific Rim*, Vol. II. PAGEOPH Topical Volumes. Birkhäuser, Cham. (2019).
- Luginbuhl, M., Rundle, J.B., Hawkins, A. and Turcotte, D.L. Nowcasting earthquakes: a comparison of induced earthquakes in Oklahoma and at the Geysers, California. *Pure Appl. Geophys.*, 175(1), 49-65 (2018a).
- Nanjo, K. Z. "Were changes in stress state responsible for the 2019 Ridgecrest, California, earthquakes?." *Nature communications* 11 (2020).
- Olami, Z., Feder, H.J.S. and Christensen, K. Self Organized Criticality in a Continuous Non-Conservative Cellular Automaton Modeling Earthquakes. *Phys. Rev. Lett.*, 68, 1244-1247 (1992).
- Omori, F., 1894. On the aftershocks of earthquakes: *Journal of the College of Science*, Imperial University of Tokyo.
- Pasari, S. Nowcasting earthquakes in the Bay-of-Bengal region. *Pure Appl. Geophys.* 23, 537-559 (2019).
- Pasari, S. Stochastic Modeling of Earthquake Interevent Counts (Natural Times) in Northwest Himalaya and Adjoining Regions. In: Bhattacharyya, S., Kumar, J. and Ghoshal, K. *Mathematical Modeling and Computational Tools, Springer Proceedings in Mathematics & Statistics*, 320, 495-501, Springer, Singapore (2020).

- Pasari, S., and Mehta, A. 2018. Nowcasting earthquakes in the northwest Himalaya and surrounding regions. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, XLII-5, 855–859 (2018).
- Pasari, S., and Sharma, Y. Contemporary Earthquake Hazards in the West-Northwest Himalaya: A Statistical perspective through Natural Times. *Seismol. Res. Lett.* (in print)
- Pasari, S., Nowcasting Earthquakes in the Bay of Bengal Region, *Pure Appl. Geophys.* (2019) 176: 1417. <https://doi.org/10.1007/s00024-018-2037-0>
- Peresan, A., & Gentili, S., 2018. Seismic clusters analysis in Northeastern Italy by the nearest-neighbor approach. *Physics of the Earth and Planetary Interiors*, 274, 87-104.
- Perez-Oregon, Jennifer, Fernando Angulo-Brown, and Nicholas Vassiliou Sarlis. "Nowcasting Avalanches as Earthquakes and the Predictability of Strong Avalanches in the Olami-Feder-Christensen Model." *Entropy* 22.11 (2020): 1228. (2020).
- Richter, C.F, *Elementary Seismology*, W.H Freeman (1958)
- Rouet-Leduc, B., Hulbert, C. and Johnson, P.A. Continuous chatter of the Cascadia subduction zone revealed by machine learning. *Nature Geoscience*, 12, 75-79 (2019).f
- Rouet-Leduc, B., Hulbert, C., Lubbers, N., Barros, K., Humphreys, C.J. and Johnson, P.A. Machine Learning Predicts Laboratory Earthquakes. *Geophys. Res. Lett.*, 44(18) (2017).
- Rundle, John B., and David D. Jackson, Numerical simulation of earthquake sequences, *Bulletin of the Seismological Society of America* 67.5 (1977): 1363-1377.
- Rundle, J. B., and Andrea Donnellan, 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 (2020): e2020EA0010
- Rundle, J.B., Donnellan, A., Fox, GCF, Crutchfield, JP, and Granat, R., Nowcasting earthquakes: Imaging the earthquake cycle in California with machine learning, submitted to *Earth and Space Science*, (2021).
- Rundle, J.B., Donnellan, A., Grant Ludwig, L, Gong, G., Turcotte, D.L. and Luginbuhl, M. Nowcasting earthquakes. *Earth and Space Science*, 3, 480-486 (2016).
- Rundle, J.B., Luginbuhl, M., Giguere, A., and Turcotte, D.L. Natural time, nowcasting and the physics of earthquakes: Estimation of risk to global megacities. *Pure Appl. Geophys.*, 175, 647-660 (2018).
- Rundle, J.B., Luginbuhl, M., Khapikova, P. et al. Nowcasting Great Global Earthquake and Tsunami Sources, *Pure Appl. Geophys.* (2019b) doi:10.1007/s00024-018-2039-y
- Rundle, J.B., Stein, S., Donnellan, A., Turcotte, D.L. Klein, W., and Saylor, C., The complex dynamics of earthquake fault systems: New approaches to forecasting and nowcasting of earthquakes, *Reports on Progress in Physics*, invited (2021)

- Rundle, JB, Giguere, A, Turcotte, DL, Crutchfield, JP, and Donnellan, A, Global seismic nowcasting with Shannon information entropy, *Earth and Space Science*, 6, 456-472 (2019a)
- Rundle, JB, Luginbuhl, M, Giguere, A and Turcotte, DL, Natural time, nowcasting and the physics of earthquakes: Estimation of risk to global megacities, *Pure Appl. Geophys.*, 175, 647-660 (2018)
- Rundle, JB, Tiampo, KF, Klein, W and Martins, JSS, Self-organization in leaky threshold systems: The influence of near mean field dynamics and its implications for earthquakes, neurobiology and forecasting, *Proc. Nat. Acad. Sci. USA*, 99, Supplement 1, 2514-2521, (2002)
- Rundle, JB, Turcotte, DL, Donnellan, A., Grant Ludwig, L, Luginbuhl, M., and Gong, G., Nowcasting earthquakes, *Earth and Space Science*, 3, 480-486 doi:10.1002/2016EA000185 (2016)
- Rundle, JB, Turcotte, DL, Sammis, C, Klein, W, and Shcherbakov, R., Statistical physics approach to understanding the multiscale dynamics of earthquake fault systems (invited), *Rev. Geophys. Space Phys.*, **41**(4), DOI 10.1029/2003RG000135 (2003).
- Sarlis, N.V., Skordas, E.S. and Varotsos, P.A. 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. *EPL*, 124 (2018).
- Sayers, C.M. and Calvez, J.L., 2010. Characterization of microseismic data in gas shales using the radius of gyration tensor. In SEG Technical Program Expanded Abstracts 2010 (pp. 2080-2084). Society of Exploration Geophysicists.
- Scholz, C.H, *The Mechanics of Earthquakes and Faulting* (2019). Cambridge University Press.
- Shannon, C.E. A Mathematical Theory of Communication, *Bell System Technical Journal*, 27, 379-423 & 623-656 (1948).
- Stauffer, D., and Aharony, A. *Introduction to percolation theory*. CRC press, 2018.
- Tiampo, K.F., Rundle, J.B., Gross, S.J., McGinnis, S. and Klein, W. Eigenpatterns in southern California seismicity. *J. Geophys. Res.*, 107(B12), 2354 (2002a).
- Tiampo, K.F., Rundle, J.B., McGinnis, S.A. and Klein, W. Pattern dynamics and forecast methods in seismically active regions. *Pure Appl. Geophys.*, 159, 2429-2467 (2002b).
- Tiampo, KF, Rundle, JB, McGinnis, S, Gross, S. and Klein, W, Mean field threshold systems and phase dynamics: An application to earthquake fault systems, *Europhys. Lett.*, 60, 481-487, (2002c).
- Tullis, T.E., Richards-Dinger, K., Barall, M., Dieterich, J.H., Field, E.H., Heien, E.M., Kellogg, L.H., Pollitz, F.F., Rundle, J.B., Sachs, M.K., Turcotte, D.L., Ward, S.N. and Yikilmaz, M.B. Generic Earthquake Simulator. *Seismol. Res. Lett.*, 83, 959-963 (2012).

- Varotsos, P. A., N. V. Sarlis, and E. S. Skordas, Self-organized criticality and earthquake predictability: A long-standing question in the light of natural time analysis, *EPL (Europhysics Letters)* 132.2 (2020b): 29001.
- Varotsos, P., Sarlis, N.V. and Skordas, E.S. Long-range correlations in the electric signals that precede rupture. *Phys. Rev. E*, 66 (2002).
- Varotsos, P., Sarlis, N.V. and Skordas, E.S. *Natural Time Analysis: The new view of time. Precursory Seismic Electric Signals, Earthquakes and other Complex Time-Series*. Springer-Verlag, Berlin Heidelberg (2011).
- Varotsos, P., Sarlis, N.V. and Skordas, E.S. Spatiotemporal complexity aspects on the interrelation between Seismic Electric Signals and seismicity, *Practica of Athens Academy*, 76, 294-321 (2001).
- Varotsos, P., Sarlis, N.V. and Skordas, E.S. Study of the temporal correlations in the magnitude time series before major earthquakes in Japan. *J. Geophys. Res. Space Phys.*, 119, 9192-9206 (2014).
- Varotsos, P., Sarlis, N.V., Skordas, E.S. and Lazaridou, M.S. Seismic Electric Signals: An additional fact showing their physical interconnection with seismicity. *Tectonophys.*, 589, 116-125 (2013).
- Varotsos, P.A., Skordas, E.S. and Sarlis, N.V. Fluctuations of the entropy change under time reversal: Further investigations on identifying the occurrence time of an impending major earthquake. *EPL*, 130 (2020a).
- Zaliapin, I. and Y. Ben-Zion, A global classification and characterization of earthquake clusters. *Geophys. J. Intl.*, 207 (1): 608-634, (2016b)
- Zaliapin, I. and Y. Ben-Zion, Discriminating characteristics of tectonic and human-induced seismicity. *Bull. Seismol. Soc. Am.*, 106(3), 846-859 (2016a)

Figure and Table Captions

Figure 1. Examples of moderate bursts having $M \geq 3.29$. Symbol color: cooler colors represent earlier events in the burst, hotter colors later events. Symbol size represents magnitudes. In each figure, the right side is a chart of magnitudes of the sequence of events in the burst. a) and b) represent bursts in which the largest earthquake is not the first. c) and d) represent bursts associated with large earthquakes having $M \geq 6$ and their aftershocks. In addition, c) illustrates events associated with the M6.2 Elmore Ranch and M6.6 Superstition Hills earthquakes of November 11, 1987. d) illustrates events associated with the M6.9 earthquake of October 17, 1989.

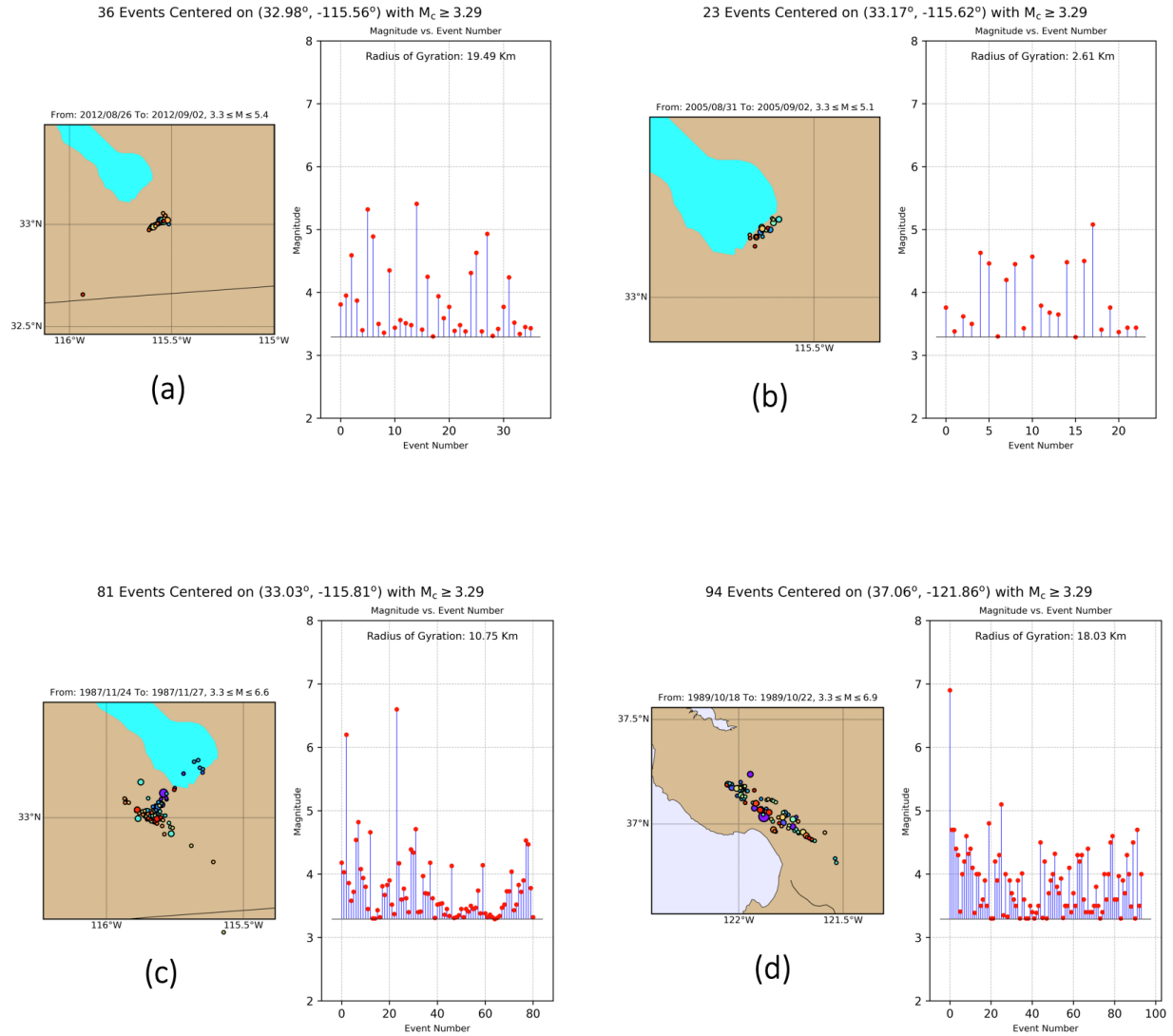


Figure 2. Four prominent eigenpatterns having high regional correlation eigenvalues, near the locations of a) the April 4, 2010 M7.2 El Mayor Cucupah; b) the June 28, 1992 M7.3 Landers + October 16, M 7.1 1999 Hector Mine events; c) July 4, 2019 M7.1 Ridgecrest earthquake; and the d) October 17, 1989 M6.9 Loma Prieta earthquake. 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(6.2% of total correlation). b) Landers-Hector pattern (5.5% of total correlation). c) Ridgecrest pattern (3.6% of total correlation). d) Loma Prieta pattern (2.7% of total correlation). Region of computation is shown as the shaded regions in the figures, within 5° latitude and 5° longitude of Los Angeles, CA. Data used to compute the patterns spanned the interval from 1/1/1950 to 12/31/2020.

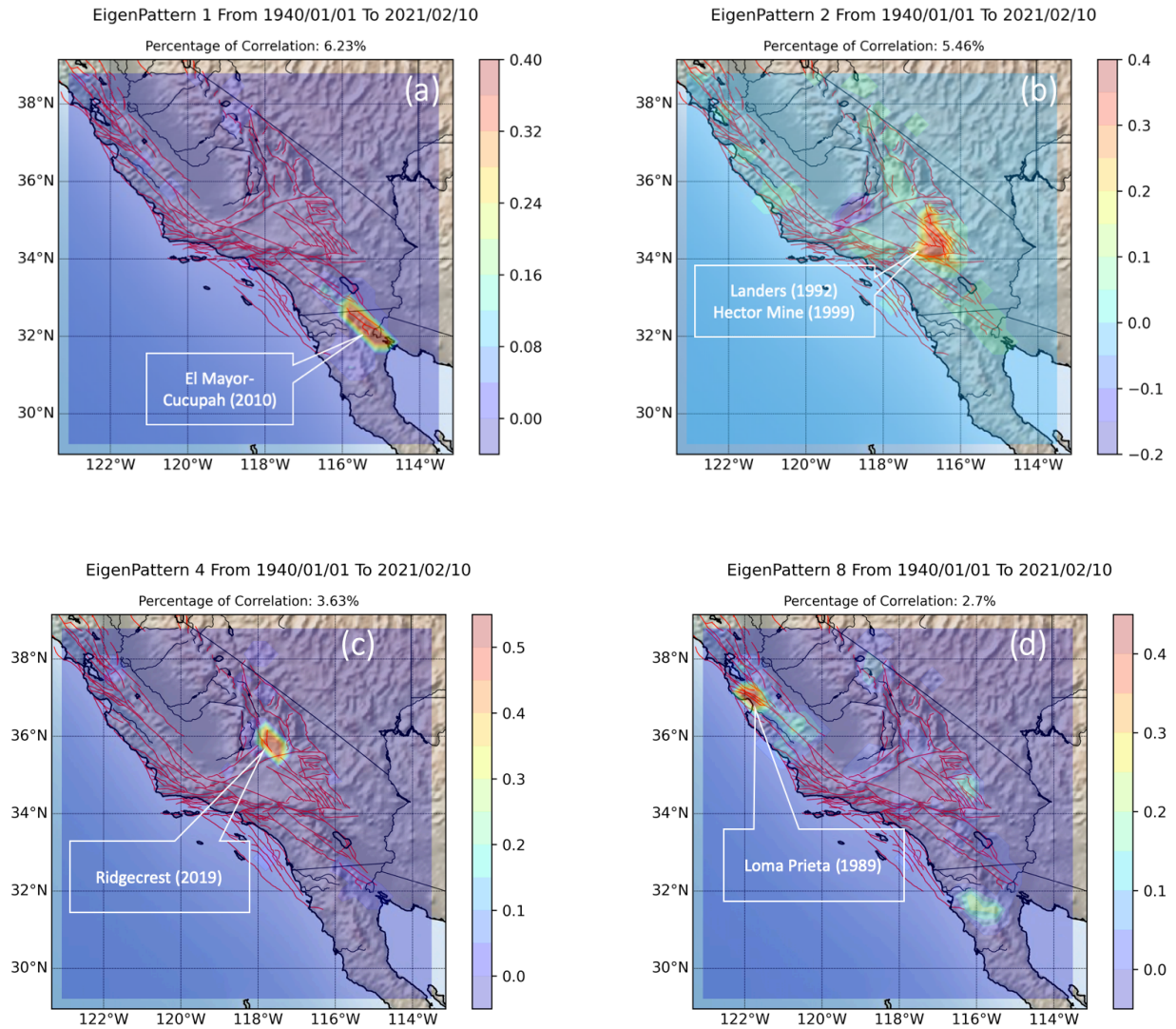


Figure 3. Comparison of two time series, a) $R_G(t)$ and b) $\chi(t)$ from 1984 through 2020. $R_G(t)$ is calculated as a filtered optimized ensemble average of radii of gyration of small earthquake bursts as a function of time. Note that the average radius of gyration of a cluster of events is often taken to be a measure of the correlation length in statistical mechanics. $\chi(t)$ is the weighted correlation time series computed from principal component analysis of the gridded timeseries of small earthquake events. In both figures, vertical red dashed lines represent large earthquakes $M \geq 6.9$, vertical dotted lines represent earthquakes $6.9 > M \geq 6.0$. Note that the vertical scale is inverted so that small values of $R_G(t)$ and b) $\chi(t)$ are at the top of the figure, so values increase towards the bottom. The horizontal green dash dot line in each figure represents the decision threshold $D_\chi(T_W)$ for the time window $T_W = 3$ years as discussed in the text. For $R_G(t)$ time series at the left, $D_\chi(T_W) = 4.15$. For $\chi(t)$ time series at right, $D_\chi(T_W) = 1.216$.

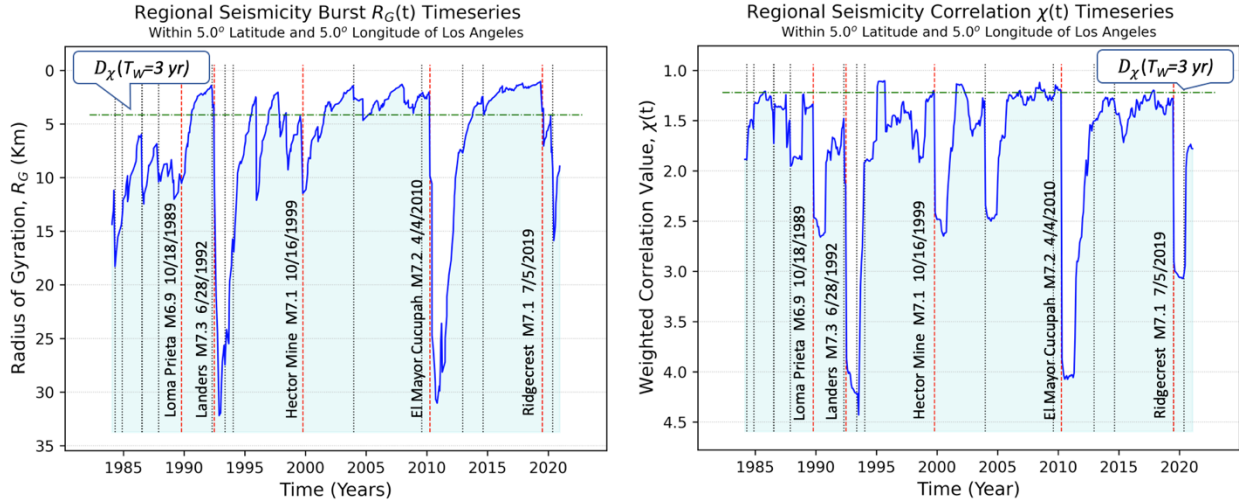


Figure 4. Time series forecasting for $R_G(t)$ time series by a 1 step walk-forward algorithm using the random forest method adapted from the scikit-learn library of machine learning methods [1]. In this application, we use 13 features in the feature vector, representing 1 year of prior data to forecast the next time series value in the future. Blue curve with dots at left panel is the same curve as shown in Figure 3 left. The red curve overlying the blue curve is the time series forecast. The left panel seems to suggest that the method has promise at forecasting future values of the curve. However, the right panel is a zoomed version of the same curve for 2018 to 2020, showing that the forecast does not correctly anticipate or forecast the large M7.1 July 5, 2019 Ridgecrest earthquake. The same is true for the other large earthquakes upon closer examination.

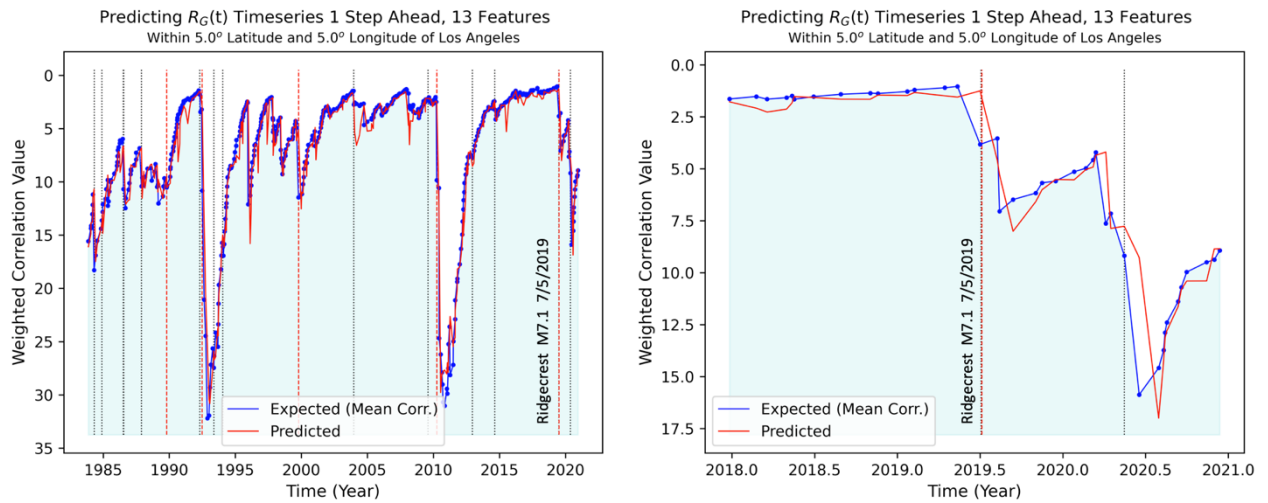


Figure 5. Similar to Figure 4, Figure 5 represents a time series forecast for $\chi(t)$ time series by a 1 step walk-forward algorithm using the random forest method adapted from the scikit-learn library of machine learning methods [2]. In this application, we also use 13 features in the feature vector, representing 1 year of prior data to forecast the next time series value in the future. Blue curve with dots at left panel is the same curve as shown in Figure 3 right. The red curve overlying the blue curve is the time series forecast. The left panel seems to suggest that the method has promise at forecasting future values of the curve. However, the right panel is a zoomed version of the same curve for 2018 to 2020, showing that the forecast does not correctly anticipate or forecast the large M7.1 July 5, 2019 Ridgecrest earthquake. The same is true for the other large earthquakes upon closer examination.

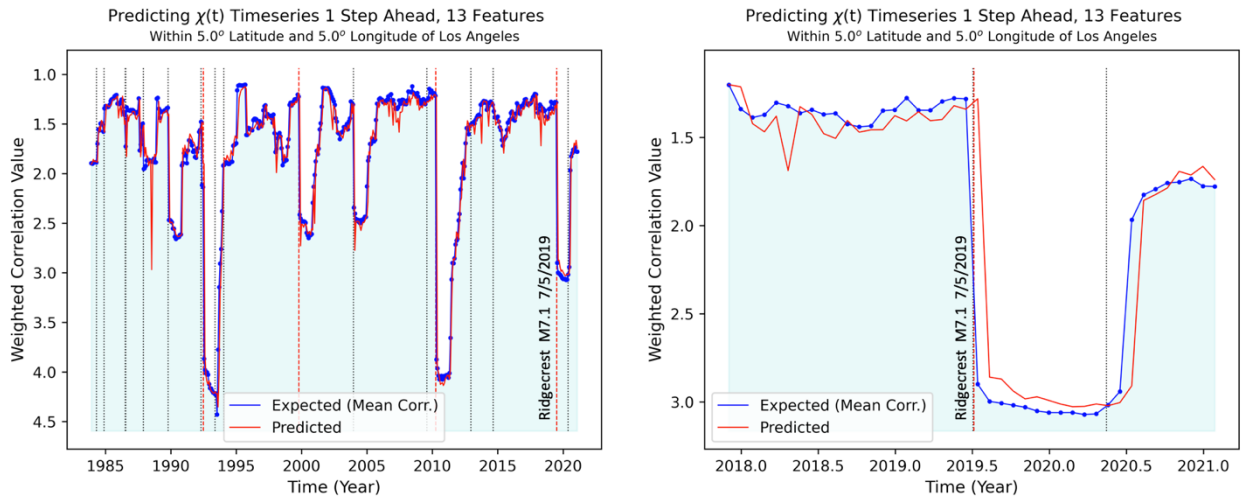


Figure 6. Receiver Operating Characteristic (ROC) diagram for two time series, a) $R_G(t)$ and b) $\chi(t)$ using data from 1960 through 2020. Red curve is the Receiver Operating Characteristic (ROC) diagrams for the time series a) $R_G(t)$ at left panel and b) $\chi(t)$ at right panel, shown in Figure 3, 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 line from lower left to upper right is the random predictor, $TPR = FPR$. To emphasize that the diagonal line does indeed represent the ROC for a random predictor, we constructed 500 random timeseries by sampling from $R_G(t)$ and $\chi(t)$ panels, respectively, with replacement. These are represented by the mass of cyan colored lines in the figures. The 1σ 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 $D_\chi(T_W)$ for the ROC using the time window of $T_W = 3$ years.

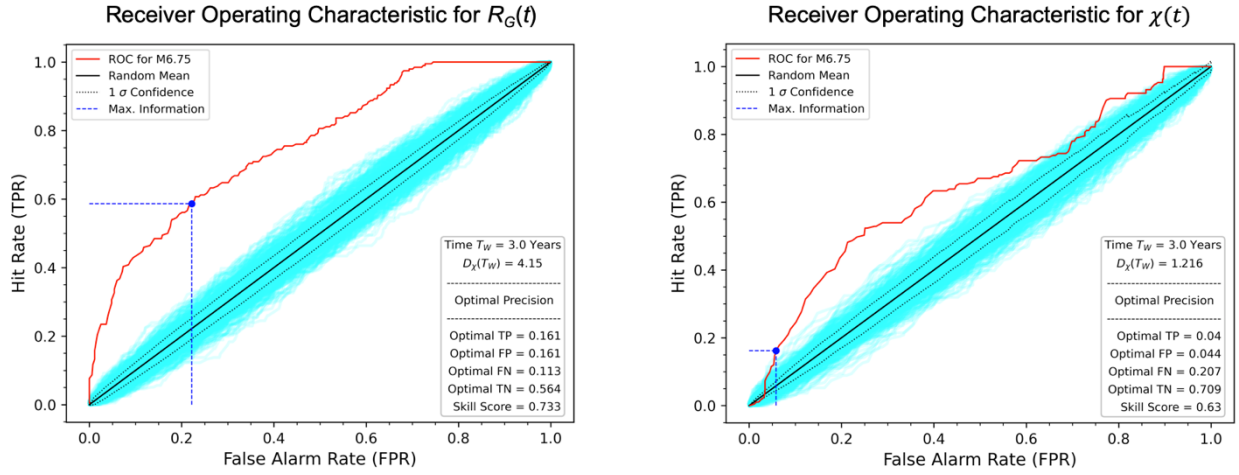


Figure 7. Similar to Figure 6, the red curve in Figure 7 shows the precision $TP/(TP + FP)$ as a function of the decision threshold $D_\chi(T_W)$ applied to the time series a) $R_G(t)$ at left panel and b) $\chi(t)$ at right panel, as shown in Figure 3. The solid black line is the random predictor. To emphasize that the solid black line does indeed represent the precision for a random predictor, we constructed 500 random timeseries by sampling from $R_G(t)$ and $\chi(t)$, from the left and right panels, respectively, with replacement. These are represented by the mass of cyan colored lines in the figures. The 1σ confidence level is indicated by the ellipsoidal dotted line enclosing the solid black random predictor line. The blue dashed vertical lines represent the optimal values of decision threshold obtained by optimizing the Shannon information entropy, using precision as the probability for a time window $T_W = 3$ years.

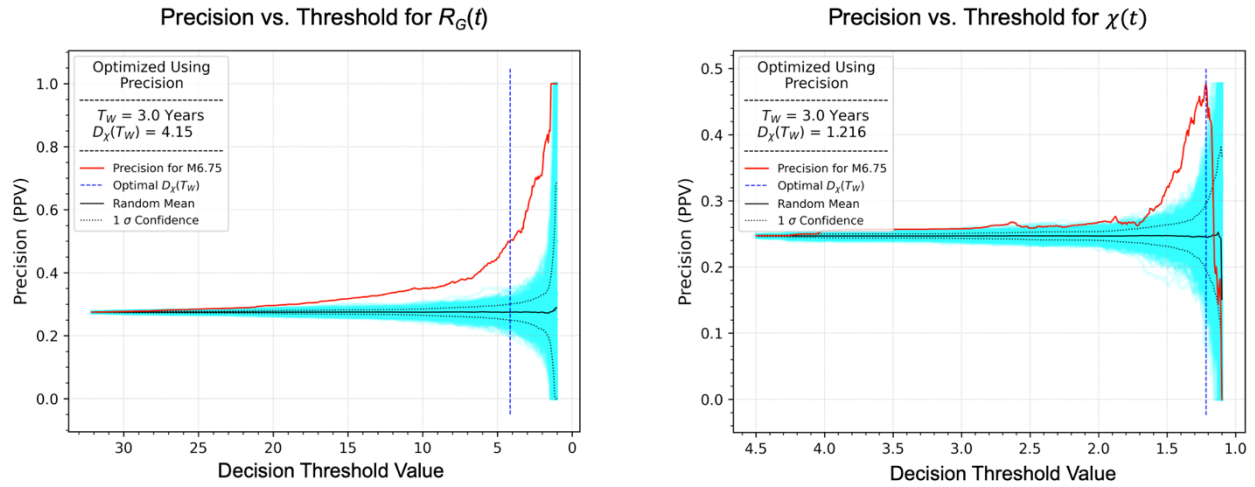


Table 1. Large earthquakes in the Los Angeles region between January 1, 1984 and December 31, 2020. These correspond to the vertical lines in Figures 3-5.

Table 1. Large Earthquakes near Los Angles from 1/1/1984 to 12/21/2019

Date (Z)	Time (Z)	Magnitude	Location
4/24/84	21:15:18.760	6.2	Morgan Hill
11/23/84	18:08:25.360	6.1	Round Valley
7/8/86	09:20:44.560	6	Morongo Valley
7/21/86	14:42:26.000	6.4	Chalfant Valley
11/24/87	01:54:14.660	6.2	Elmore Ranch
11/24/87	13:15:56.710	6.6	Superstition Hills
10/18/89	00:04:15.190	6.9	Loma Prieta
4/23/92	04:50:23.230	6.1	Joshua Tree
6/28/92	11:57:34.130	7.3	Landers
6/28/92	15:05:30.730	6.3	Big Bear
5/17/93	23:20:50.250	6.1	Big Pine
1/17/94	12:30:55.390	6.7	Northridge
10/16/99	09:46:44.460	7.1	Hector Mine
12/22/03	19:15:56.240	6.5	San Simeon
8/3/09	18:40:50.100	6.2	Gulf of California
4/4/10	22:40:42.360	7.2	El Major Cucupah
12/14/12	10:36:01.590	6.3	Baja Coast
8/24/14	10:20:44.070	6.02	Napa
7/4/19	17:33:49.000	6.4	Ridgecrest
7/6/19	03:19:53.040	7.1	Ridgecrest
5/15/20	11:03:27.176	6.5	Tonopah, NV

Table 2. Comparison of optimal data, obtained by optimizing the Shannon information from entropy of the precision variable, to random data for a time window $T_w = 3$ years for both $R_G(t)$ and $s\chi(t)$ time series. Optimal values of decision threshold $D_\chi(T_w)$ are 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 one-sided P -statistic based on Z-values using standard procedures.

Statistic	Definition	Optimal Precision		Random Data		P- Value Statistics	
		$R_G(t)$	$\chi(t)$	$R_G(t)$	$\chi(t)$	$R_G(t)$	$\chi(t)$
TP	True Positive	0.161	0.04	0.087 ± 0.009	0.02 ± 0.005	$P \ll 0.01$	$P \ll 0.01$
FP	False Positive	0.161	0.044	0.228 ± 0.015	0.059 ± 0.009	$P \ll 0.01$	$P < 0.05$
FN	False Negative	0.113	0.207	0.188 ± 0.009	0.227 ± 0.005	$P \ll 0.01$	$P \ll 0.01$
TN	True Negative	0.564	0.709	0.498 ± 0.015	0.694 ± 0.009	$P \ll 0.01$	$P < 0.05$
Hit Rate (TPR)	$TP/(TP+FN)$	0.587	0.162	0.316 ± 0.032	0.078 ± 0.019	$P \ll 0.01$	$P \ll 0.01$
Specificity (TNR)	$TN/(TN+FP)$	0.778	0.942	0.686 ± 0.021	0.921 ± 0.013	$P \ll 0.01$	$P < 0.1$
Precision (PPV)	$TP/(TP+FP)$	0.5	0.476	0.276 ± 0.023	0.244 ± 0.051	$P \ll 0.01$	$P \ll 0.01$
Accuracy	$TP+TN$	0.726	0.749	0.584 ± 0.017	0.712 ± 0.01	$P \ll 0.01$	$P \ll 0.01$
Skill Score	Area Under ROC	0.733	0.630	0.5 ± 0.185	0.5 ± 0.113	$P = 0.10$	$P = 0.12$