Nowcasting Earthquakes:
Imaging the Earthquake Cycle in California with Machine Learning

by

John B Rundle

Andrea Donnellan

Geoffrey Fox

James P. Crutchfield

Robert Granat

Abstract

We propose a new machine learning-based method for nowcasting earthquakes to image the time-dependent earthquake cycle. The result is a timeseries which may correspond to the process of stress accumulation and release. The timeseries is 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 then compute the weighted correlation timeseries (WCT) 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 and Donnellan, 2020), 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 contains 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 hazard.

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 uses 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 1990-2020. 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 towards 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 contains information about future large earthquakes that can be used for hazard assessment.

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 in a correlation time series corresponding to future large earthquakes can be detected.

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. (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 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 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 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 (Tiampo et al, 2000a,b).

<u>To summarize our results:</u> The timeseries so defined implies 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. 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.

Method

The first step is to define a spatial coarse graining by assigning an array of grid boxes of given latitude and longitude $\Delta \boldsymbol{x}$ (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.

For example, in the results example described in this paper, a minimum of 35 small earthquakes having magnitudes M>3.29 over the time period of 1960-2020 was required, or a rate of about 1 small earthquake every year. Variations of this value changes results 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 $\{t_i, M_i, \boldsymbol{z}_i\}$, where i = 1,..., 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 $\boldsymbol{z}_i$ is hypocentral location (latitude-longitude-depth). Note that $\boldsymbol{z}_i$ is a container variable for an epicentral (horizontal) location $\boldsymbol{x}_i$ and depth d_i.

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

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

$$\mu_{n,t} \;=\; (\frac{1}{t - t_0}) \int_{t_0}^{t} \Phi(\boldsymbol{x}_n, t')\, \mathrm{d}t' \tag{1}$$

$$\sigma_{n,t}^2 \;=\; (\frac{1}{t - t_0}) \int_{t_0}^{t} \left(\Phi(\boldsymbol{x}_n, t') \;-\; \mu_{n,t}\right)^2 \mathrm{d}t' \tag{2}$$

$$\widehat{\Phi}(\boldsymbol{x}_n, t) \;=\; (\Phi(\boldsymbol{x}_n, t) \;-\; \mu_{n,t}) /\, \sigma_{n,t} \tag{3}$$

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

$$C_{nm}(t) \;=\; \int_{t_0}^{t} \widehat{\Phi}(\boldsymbol{x}_n, t')\widehat{\Phi}(\boldsymbol{x}_m, t')\, \mathrm{d}t' \tag{4}$$

$C_{nm}(t)$ is then diagonalized to find its eigenvectors (eigenpatterns) $\boldsymbol{e}_i(t)$, i = 1,...,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 $\Phi(\boldsymbol{x}_n, t)$ are just the N_X time series $\Phi(\boldsymbol{x}_n, t)$, summed over a previous time interval $\tau = S\Delta t$. The n^{th} component of $\boldsymbol{\psi}(t)$ is then:

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

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

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

we then compute the power spectrum of $\boldsymbol{\psi}(t)$, i.e., $a_i(t)^2$. In computing $\boldsymbol{e}_i(t)$, 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 vector with the vector of correlation eigenvalues. This dot product is then the weighted correlation value $\chi(t)$ for the regional seismicity:

$$\chi(t) \equiv \langle \lambda(\mathrm{t}) \rangle = \sum_i \lambda_i(t)\, a_i(t)^2 \tag{7}$$

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

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, we adopt the normalization:

$$\sum_i \lambda_i(t) = 100 \tag{8}$$

$$\textstyle\sum_i a_i(t)^2 = 1 \ \ (\mathrm{L_2\ norm}) \tag{9}$$

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

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^o latitude, 118.2437^o west longitude) , and within 5.0^o (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 present 12/31/2020. For the time interval Δt as discussed above, we set Δt = 0.07692 year, equal to 1/13 year or approximately 1 "lunar month", equal to 4 "weeks" of length 1/52 year.

As is often the case in these machine learning methods (see, e.g., Rouet-Leduc et al., 2017, 2019), the state vector $\boldsymbol{\psi}(t)$ consists of a sliding window of length τ = SΔt, 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 τ = 1 year.

We downloaded the earthquake catalog from the USGS web site, collected and filtered the data to construct acceptable timeseries of small earthquakes. Size of the N coarse grained grid boxes was taken to be 0.33^o. 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.

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 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.

Signal Detection and Information Content

We now turn to investigating the information contained in the correlation timeseries $\chi(t)$ that is shown in Figure 2 from 1984 to 12/31/2020. More specifically, we are interested in determining whether the timeseries contains 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*).

We wish to determine whether signals of future large earthquakes can be detected by analysis of $\chi(t)$. 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*):

$$TPR = TP/(TP + FN) \tag{5}$$

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

$$FPR = 1 - TNR = FP/(FP + TN) \tag{6}$$

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:

$$PPV = TP/(TP + FP) \tag{7}$$

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

$$ACC = (TP + TN)/(TP + FN + FP + TN) \tag{8}$$

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:

- 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 $TP \rightarrow TP+1$. i.e, a true positive.
- 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 $FP \rightarrow FP+1$. i.e, a false positive.
- 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 $FN \rightarrow FN+1$. i.e, a false negative.
- 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 $TN \rightarrow TN+1$. 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.:

$$TP \rightarrow TP/(TP + FP + FN + TN) \tag{9}$$

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

$$TPR = FPR \tag{10}$$

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 σ 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 month 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_λ,. 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:

$$I_S = p \log_2 p + (1-p)\log_2(1-p) \tag{11}$$

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 3 and 4, and Table 1, optimal values for *TP, FP, FN, TN* that arise from using the equation (11) 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 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.

Statistical Tests of Significance

To test whether 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 (7).

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:

$$Z = \frac{S - \mu_R}{\sigma_R} \tag{10}$$

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 $\{\chi_R(t)\}$, 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 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 $\chi(t)$, 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 and 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).

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. Future investigations may allow further refinement and clarification of whatever information this and similar time series contain.

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 catalog[1] 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/

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.

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.

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).

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)

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).

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, 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., Giguere, A., Turcotte, D.L., Crutchfield, J.P. and Donnellan, A. Global seismic nowcasting with Shannon information entropy. *Earth and Space Science*, 6, 456-472 (2019).

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, John 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

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).

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).

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).

Figure 1. Top four eigenpatterns having the highest value of regional correlation eigenvalues, near the locations of the 2010 El Mayor Cucupah; 1992 Landers + 1999 Hector Mine; 1952 Kern County; and 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(6.22% of total correlation). b) Landers-Hector pattern (5.5% of total correlation). c) Kern County pattern (5.3% of total correlation). d) Ridgecrest pattern (3.63% 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.

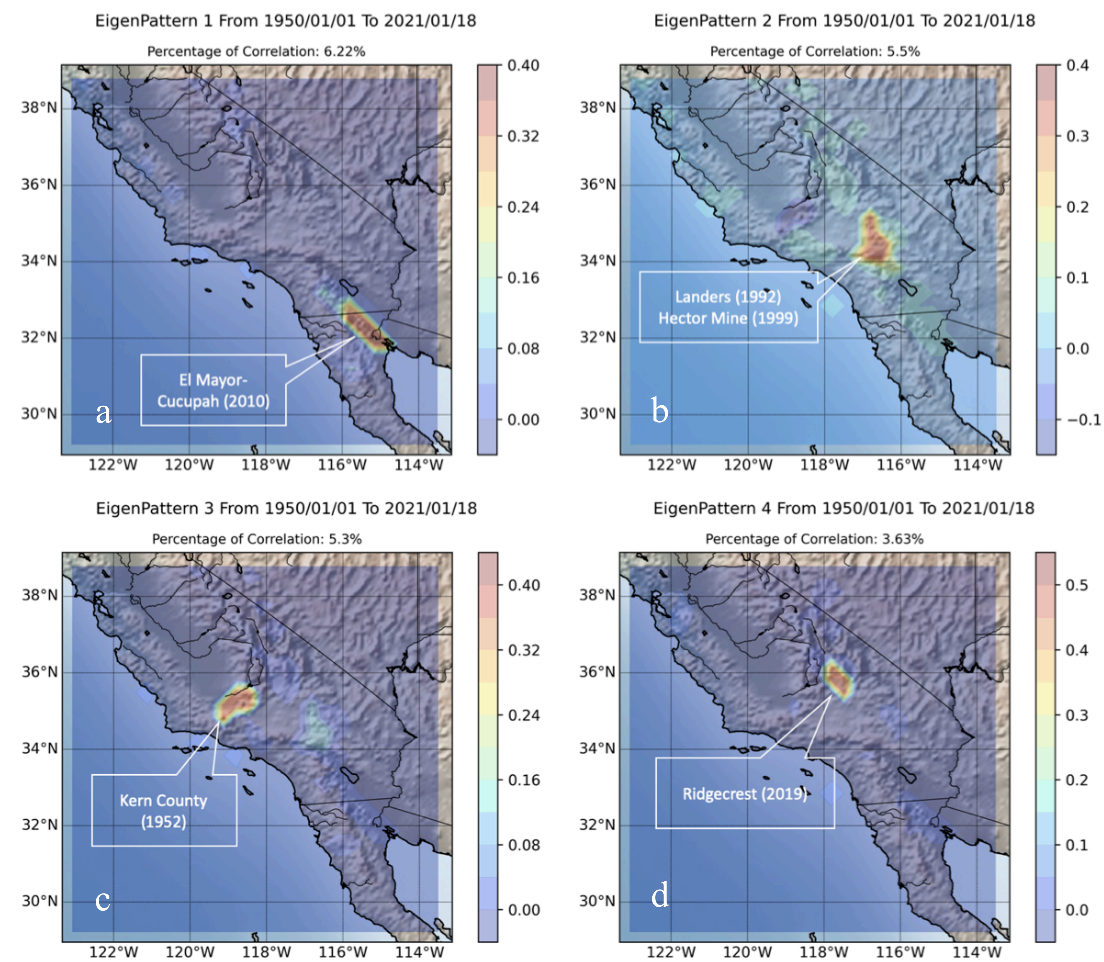

Figure 2. Weighted correlation value $\chi(t)$ as a function of time for the California region shown as the shaded regions in Figure 1. Computed according to equations (3)-(4). Vertical dashed red lines represent the four large earthquakes having magnitudes M>7.0 Vertical black dotted lines represent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

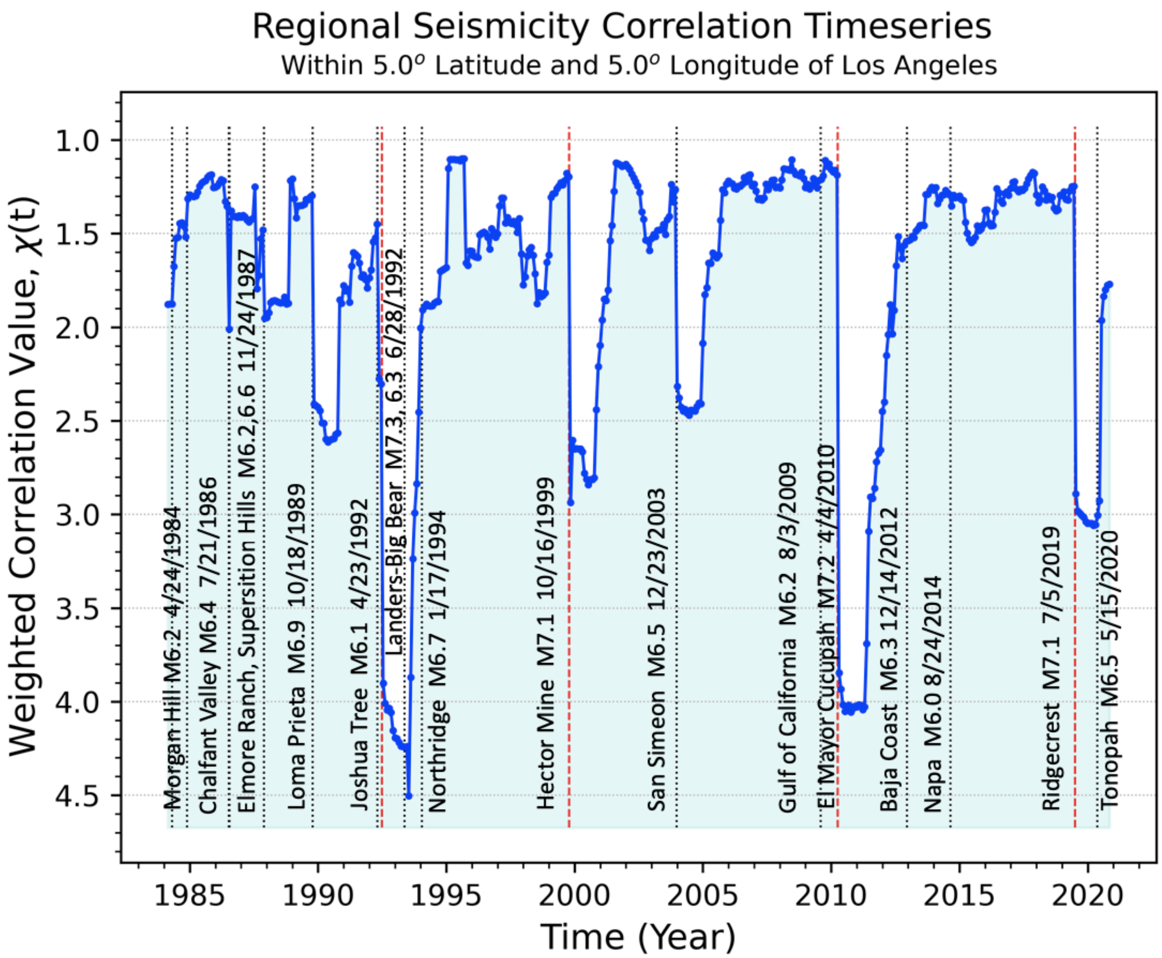

Figure 3. Red curve is the Receiver Operating Characteristic (ROC) diagram for the time series $\chi(t)$ shown in Figure 2, obtained by systematically varying the decision threshold $D_\chi(T_W)$ as described in the text for major earthquakes having magnitudes M≥6.75, and computing the true positive rate (hit rate or recall) TPR and plotting against the false positive rate (1 - specificity) FPR. The black diagonal black line from 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 $\chi(t)$ with replacement. These are represented by the mass of cyan colored lines in the figures. The 1 σ 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.

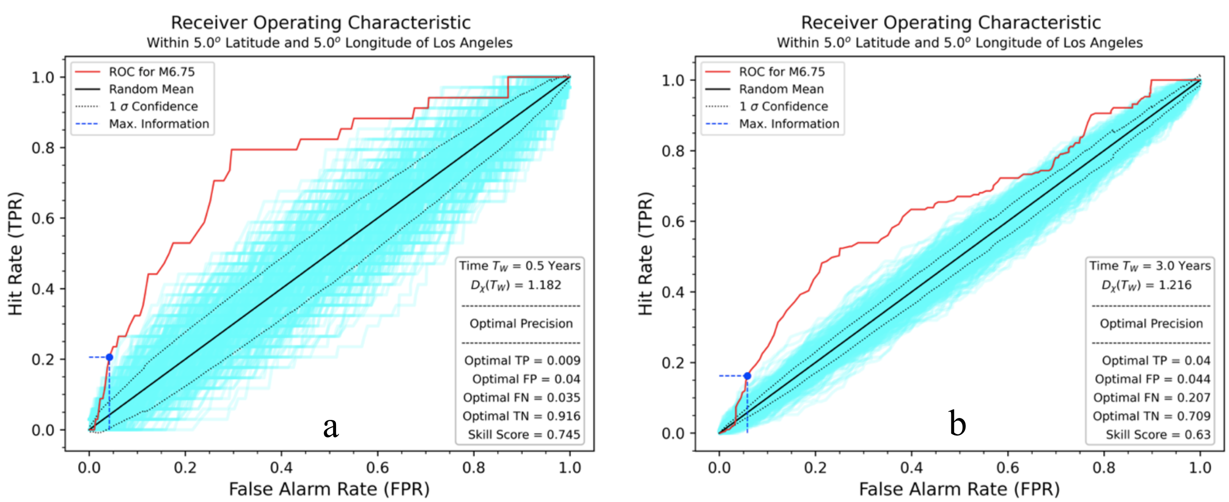

Figure 4. Similar to Figure 3, the red curve in Figure 4 shows the precision $TP/(TP + FP)$ as a function of the decision threshold $D_\chi(T_W)$ applied to the time series $\chi(t)$ shown in Figure 2. The solid black black line is the random predictor. To emphasize that the solid black line does indeed 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 σ 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. a) T_W = 6 months. b) T_W = 3 years.

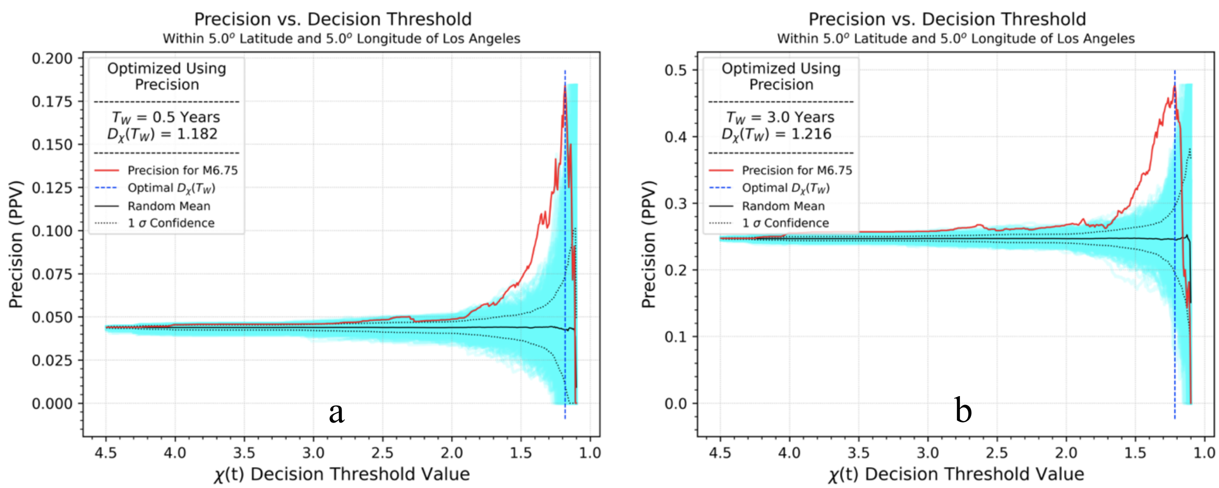

Table 1. Comparison of optimal data, obtained by optimizing the Shannon information from entropy of the precision variable, to random data. Optimal value of decision threshold $D_{\chi}(T_W)$ is found by this procedure. The null hypothesis is that our optimal precision data is generated by a random process. Thus we compare our values to those generated by the random process and calculate a *P*-statistic based on Z-values using standard procedures.

		Optimal Precision		Random Data ($\mu_R \pm \sigma_R$)		*P* - Value Statistics	
Statistic	Definition	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 (ACC)	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