Nowcasting earthquakes

J. B. Rundle[1,2,3,4], D. L. Turcotte[3], A. Donnellan[5], L. Grant Ludwig[6], M. Luginbuhl[1], and G. Gong[4]

[1]Department of Physics, University of California, Davis, California, USA, [2]Santa Fe Institute, Santa Fe, New Mexico, USA, [3]Department of Earth and Planetary Science, University of California, Davis, California, USA, [4]Open Hazards Group, Davis, California, USA, [5]Science Division, Jet Propulsion Laboratory, Pasadena, California, USA, [6]Program in Public Health, University of California, Irvine, California, USA

Abstract Nowcasting is a term originating from economics and finance. It refers to the process of determining the uncertain state of the economy or markets at the current time by indirect means. We apply this idea to seismically active regions, where the goal is to determine the current state of the fault system and its current level of progress through the earthquake cycle. In our implementation of this idea, we use the global catalog of earthquakes, using "small" earthquakes to determine the level of hazard from "large" earthquakes in the region. Our method does not involve any model other than the idea of an earthquake cycle. Rather, we define a specific region and a specific large earthquake magnitude of interest, ensuring that we have enough data to span at least ~20 or more large earthquake cycles in the region. We then compute the earthquake potential score (EPS) which is defined as the cumulative probability distribution $P(n < n(t))$ for the current count $n(t)$ for the small earthquakes in the region. From the count of small earthquakes since the last large earthquake, we determine the value of EPS = $P(n < n(t))$. EPS is therefore the current level of hazard and assigns a number between 0% and 100% to every region so defined, thus providing a unique measure. Physically, the EPS corresponds to an estimate of the level of progress through the earthquake cycle in the defined region at the current time.

1. Introduction

Given a seismically active local region, a problem of interest is to determine how much stress and strain has accumulated since the last major earthquake. In other words, we would like to determine the level of progress of the region through the "earthquake cycle." This earthquake cycle problem is complicated because the absolute stress and strain since the last major earthquake cannot be determined from direct observations at all locations of interest. However, we describe a new method, "nowcasting," that may provide an answer to this problem. Note that the term nowcasting arises in economics, finance, and political analysis (see for example: https://en.wikipedia.org/wiki/Nowcasting_(economics)). It refers to the determination of the uncertain current state of a system via indirect means.

In earthquake fault systems, earthquakes are observed to recur within a seismically active region [*Scholz*, 1990]. Determining the progression of the region through its earthquake "cycle" is important for estimating the level of seismic hazard. In the past, this determination of the state of a regional fault system has focused on trying to estimate the state of stress in the earth, its relation to the failure strength of the active faults in a region, and the rate of accumulation of tectonic stress [*Scholz*, 1990]. Determining the values of these parameters would allow researchers to estimate the proximity to failure of the faults in the region. This would be an answer to the question of "how far along is the region in the earthquake cycle?" However, such a program of research is very difficult, because it is not possible to determine the state of stress and strain in general. As a result, the current state of the fault system remains uncertain and elusive.

We propose to use the idea of nowcasting to answer the question of the current hazard state of the region. More specifically, we count the number of small earthquakes since the last large earthquake in a defined region to estimate the current hazard level in the region. Event counts as a measure of "time," rather than the clock time, is known as "natural" time [*Varotsos et al.*, 2002, 2005, 2011; *Holliday et al.*, 2006]. We will show that the use of natural time has at least two advantages when applied to earthquake seismicity:

1. It is not necessary to decluster the aftershocks. The natural time count is uniformly valid when aftershocks dominate, when background seismicity dominates, and when both contribute.

2. Natural time statistics are independent of the level of the seismicity as long as the *b* value is approximately constant. In computing nowcasts, the concept of natural time, counts of small earthquakes, is used as a measure of the accumulation of stress and strain between large earthquakes in a defined geographic region.

Nowcasting, which describes the present state of a system, is distinct from the idea of forecasting, which looks forward in time [*Holliday et al.*, 2005, 2016; *Field*, 2007; *Rundle et al.*, 2012]. Forecasting is the calculation of probabilities for the future. Nowcasting is the calculation of the current state of the system. Nowcasting can be used as a basis for forecasting if a method is used to project the current state into future states. In fact, nowcasting should be a prerequisite to forecasting, the estimation of the future state of the system. The current state must be known, at least approximately, before the future state can be accurately estimated.

2. Procedure

Our goal is to assign a number to a defined small geographic region at a given time, to characterize the current potential for a "large" earthquake to occur in the region subject to recurring large earthquakes. We define an "earthquake cycle" as the interval between two large earthquakes in a seismically active region. We denote the magnitude of the "large" earthquake as M_λ. "Large" in this sense means having the potential to cause damage or injuries if it were to occur close by. We compute the potential by using "small" earthquakes, whose magnitude is denoted by M_σ. The large earthquake magnitude is selected to ensure there are enough earthquake cycles to provide reasonable statistics. The small earthquake magnitude is typically set by the catalog completeness level but can be also viewed as a free parameter within a certain range.

It should be noted at the outset that one of the problems associated with this method is due to the fact that earthquake catalogs are not necessarily complete at the desired small-earthquake magnitude level. Small events observed during aftershock sequences in particular may not be completely recorded due to instrumental issues. We hope that this issue will resolve over time as earthquake detection and recording is more fully automated. Another problem is that the "large" earthquakes of interest are necessarily limited in magnitude to events for which "many" cycles of activity have occurred in the large region. As a result, this constraint probably limits the magnitude of the large earthquakes to about $M_\lambda \leq 7.5$.

The basis for our nowcast is the use of "natural" time. Natural time is simply the count of small earthquakes, in our case the number *n* of small earthquakes since the last large earthquake. The statistical distribution of *n* will be an essential part of our analysis. There are several advantages to using natural time. The first is that it is not necessary to separate the aftershocks ("decluster") from the background seismicity. A second is that only the natural interevent count statistics are used rather than the seismicity rate, which involves calendar time as well.

The Gutenberg-Richter magnitude-frequency relation [*Scholz*, 1990] can be used to show that the number of small earthquakes having magnitudes larger than M_σ but less than magnitude M_λ is on average a known value *N*. We can calculate *N* using the Gutenberg-Richter law for the average number N_{avg} of earthquakes greater than *M*:

$$N_{avg} = 10^a 10^{-bM} \tag{1}$$

where the *b* value is typically a number near 1, and the value of *a* is set by the level of seismicity in the region [*Scholz*, 1990].

If we denote by $N_\sigma = 10^a 10^{-bM_\sigma}$ the average number of earthquakes having magnitude larger than M_σ and by $N_\lambda = 10^a 10^{-bM_\lambda}$ the average number of earthquakes having magnitude larger than M_λ, we find:

$$N = \frac{N_\sigma - N_\lambda}{N_\lambda} = 10^{b(M_\lambda - M_\sigma)} - 1 \tag{2}$$

The number of small earthquakes is often used as a measure of the passage of *natural time* for the earthquake system. Notice that *N* does not depend on *a*.
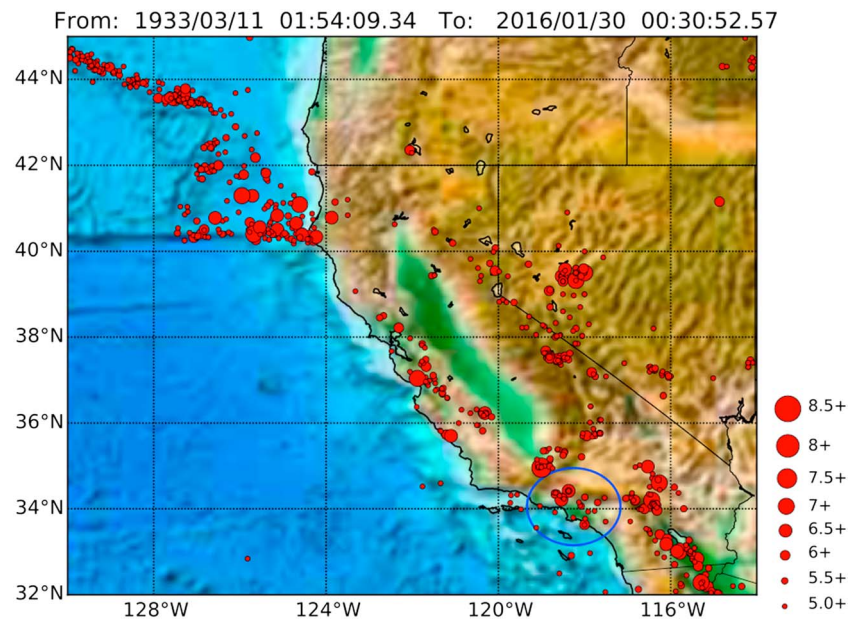
Figure 1. The region shown in this figure (California-Nevada region, latitudes 32.0°N to 43.0°N, longitudes 130.0°W to 114.0°W) will be used to illustrate the nowcast method. We also consider the Los Angeles region at the center of the blue circle with a radius of 100 km. Also shown are $M_\lambda \geq 5$ for the period 11 March 1933 to 30 January 2016.

The average number N of small earthquakes between large earthquakes will be called the mean *interevent count*. Mathematically, we can express the progress through the regional earthquake cycle, or *earthquake potential* for a large earthquake to occur having magnitude larger than M_λ, by computing the cumulative distribution function (CDF) of small earthquakes of magnitude larger than M_σ but less than M_λ: $M_\sigma \leq M \leq M_\lambda$.

To compute the CDF, we tabulate the number of small earthquakes for each large earthquake cycle from an appropriate database such as the Advanced National Seismic System (ANSS) earthquake catalog [ANSS]. We then use this to define the probability density function (PDF) and the CDF by standard methods [e.g., *Bevington and Robinson*, 2003].

Once we have the CDF, we can then use the current count of small earthquakes $n(t)$ at time t to compute the current value of the CDF, $P\{n \leq n(t)\}$. Note that t is the (calendar) time since the last large earthquake, and $n(t)$ is the number of small earthquakes since the last large earthquake. This value is defined to be the *earthquake potential score* (EPS) at time t:

$$\text{EPS} \equiv P\{n \leq n(t)\} \tag{3}$$

We interpret the EPS as the potential for the occurrence of the next large earthquake having magnitude larger than M_λ within the defined geographic region. By construction, EPS will increase monotonically with time since the last large earthquake. It will reset to EPS $= P\{n \leq n(t = 0)\} = 0$ immediately after the next large earthquake, and then again begin to increase monotonically until the next large earthquake occurs. We note that the nowcasting method does not, a priori, involve a model; it is only a process of tabulating and interpreting data. However, to the extent that the results have meaning, it is a transparent way of estimating the progress of a region through the seismic cycle of large earthquakes. Since our analysis is independent of the rate of seismic activity as shown in equation (2), we can apply it either to the large geographic region used to obtain statistics or to any smaller region within the larger region.

3. Examples

3.1. California-Nevada Region

To illustrate the method, we begin with nowcasts of the earthquake potential in the California-Nevada region illustrated in Figure 1. We obtain our earthquake statistics for this entire region. We also obtain the present
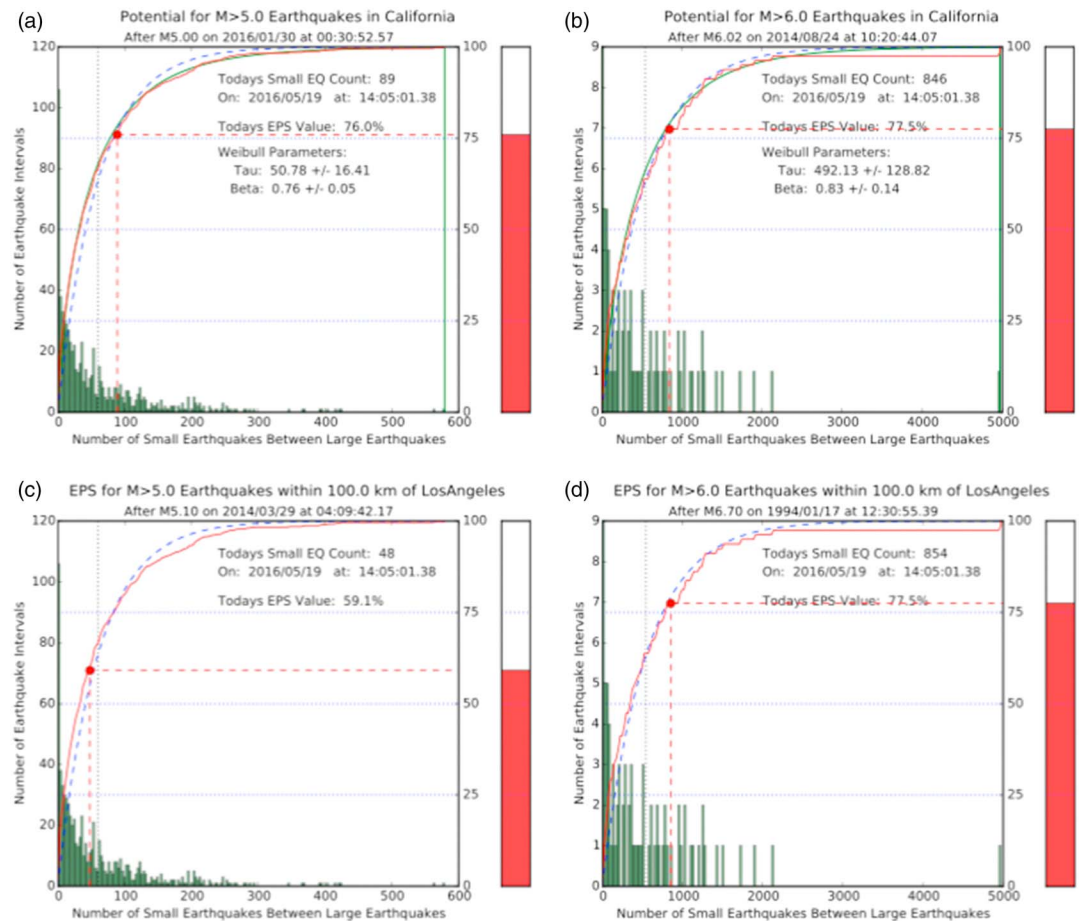
Figure 2. Earthquake potential score (EPS) (a and b) for the California-Nevada region and (c and d) for the region within 100 km radius of Los Angeles. Results are for $M_\lambda \geq 5$ (Figures 2a and 2ac) and for $M_\lambda \geq 6$ (Figures 2b and 2d). The green bars are the histograms of the number of small earthquakes ($M_\sigma \geq 3.0$) between the large earthquakes. The red curves are the CDFs $P\{n \leq n(t)\}$ of the histogram values. The dashed blue curves are the Poisson distributions having the same mean as the red curves. The green curve is the best fitting Weibull statistical distribution. The red circles correspond to the number of small earthquakes since the last large earthquake. The mean number of small earthquakes ($M_\sigma \geq 3.0$) between large earthquakes is shown by the vertical dotted line.

EPS values for earthquakes with magnitudes greater than $M_\lambda = 5$ and for earthquakes with magnitudes greater than $M_\lambda = 6$ in the entire region. We then turn to a region within a circle of 100 km radius about the center of Los Angeles. We again obtain the present EPS values for earthquakes in this region greater than $M_\lambda = 5$ and $M_\lambda = 6$.

We note that this value of 100 km, which we adopt as a standard, is arbitrary. For $M_\lambda = 6$ earthquakes, we note that such an event would produce a peak ground acceleration in the range of Mercalli Intensity IV, which can be felt and may even damage older buildings. In any event, the value of the radius used should be noted if other standards are developed.

Since 1933, there have been 721 earthquakes in the California-Nevada region (between latitudes 32°N and 43°N and between longitudes 130°W and 114°W) having $M_\lambda \geq 5$, of which 81 earthquakes had magnitudes $M_\lambda \geq 6$. Thus, there have been 720 earthquake-to-earthquake intervals between $M_\lambda \geq 5$ events and 80 such earthquake-to-earthquake intervals between $M_\lambda \geq 6$ events. In this region, we use a completeness magnitude of $M_\sigma = 3.0$.

The results for the computation of the earthquake potential score (EPS) for $M_\lambda \geq 5$ in the California-Nevada region are shown in Figure 2a. The vertical bars are a histogram of the number of small earthquakes ($M_\sigma \geq 3.0$) n in the 720 interevents between the $M_\lambda \geq 5$ earthquakes. The mean number (mean natural time)

is $N = 60$ small earthquakes, corresponding to $b = 0.89$. These values are used to obtain the corresponding CDF $P\{n \leq n(t)\}$ shown in the red curve. The dashed blue curve is the Poisson distribution having the same mean as the CDF $P\{n \leq n(t)\}$. As we discuss below, statistical tests of the data-derived CDF clearly indicate that the data-derived CDF is not a Poisson distribution. The red dot marks the position of the current small earthquake count. There have been $n = 89$ earthquakes in the region having magnitudes $M_\sigma \geq 3.0$ since the last $M_\lambda \geq 5$ event, which was on 30 January 2016. The red "thermometer" on the right side is a visual representation of the current cumulative probability (CDF) which defines the earthquake potential score EPS = 76.0%. The green curve underlying the red data-derived curve is the best fitting Weibull distribution [*Weibull*, 1951], also as described below.

The results for the computation of the earthquake potential score (EPS) for the $M_\lambda \geq 6$ earthquakes in the California-Nevada region are shown in Figure 2b. The vertical bars are a histogram of the number of small earthquakes ($M_\sigma \geq 3.0$) n in the 80 intervents between the $M_\lambda \geq 6$ earthquakes. The mean number (mean natural time) is $N = 544$ small earthquakes corresponding to $b = 0.91$. The corresponding CDF $P\{n \leq n(t)\}$ and Poisson distributions are shown. The red dot marks the position of the current small earthquake count n. There have been $n = 846$ earthquakes in the region having magnitudes $M_\sigma \geq 3.0$ since the last $M_\lambda \geq 6$ event, which was the $M6.04$ Napa earthquake on 24 August 2014. The red "thermometer" on the right side is a visual representation of the current cumulative probability (EPS), which is EPS = 77.5%. The green curve underlying the red data-derived curve is the best fitting Weibull distribution.

3.2. Cities

As a second illustration of nowcasting, we consider a local value of EPS appropriate for estimating the current state of hazard for individual cities. For these, we use a circle of radius 100 km around the geographic center of the city. Within each circle, there must be at least one large earthquake having magnitude M_λ or greater during the time period considered. Assuming that the statistics of the large region in which the city is embedded also describe the statistics within the circular city region, we can then use the regional CDF $P\{n \leq n(t)\}$. We then count the number of small earthquakes within the city circle since the last large earthquake $M_\sigma \leq M \leq M_\lambda$. Using the regional CDF, we can then compute the value of EPS for the city.

As a concrete example, we consider a 100 km radius circle around Los Angeles, California, and compute the EPS value for earthquakes having $M_\lambda \geq 5$. Figure 2c is a plot of the EPS score for the Los Angeles circle for events having $M_\lambda \geq 5$. The most recent event was the $M5.1$ La Habra earthquake on 29 March 2014. From Figure 2a, we see that the current EPS score as of 25 April 2016 for the area within 100 km radius of Los Angeles is 59.1. This means that this region has progressed 59.1% toward the next $M_\lambda \geq 5$ earthquake in the circular region. There have been 32 such $M_\lambda \geq 5$ earthquakes within the 100 km circle since 1933 (83 years).

It is also of interest to repeat this calculation for $M_\lambda \geq 6$ earthquakes within the 100 km circular region, using the 81 $M_\lambda \geq 6$ earthquakes that have occurred in the large region shown in Figure 1. Doing so, we find the results of Figure 2d. Here we see that the circular region around Los Angeles has an EPS score of 77.5, meaning that the small region has progressed 77.5% of the way to the next $M_\lambda \geq 6$ earthquake since the last such event, which was the $M6.7$ Northridge earthquake of 17 January 1994. The next-previous $M_\lambda \geq 6$ earthquake in the circular region was the $M6.6$ San Fernando earthquake of 9 February 1971, 23 years earlier. The next-previous $M_\lambda \geq 6$ earthquake was the $M6.4$ Long Beach earthquake of 11 March 1933, some 38 years earlier than the San Fernando event. Thus, the two interval times on record for such $M_\lambda \geq 6$ events in the 100 km circle around Los Angeles are 23 years and 38 years.

We can also compute the EPS score for major global cities located in seismically active areas using the ANSS Composite Catalog (ANSS). As parameters, we again adopt a circle radius of 100 km. We choose $M_\lambda \geq 6$ for the large earthquake magnitude. We also choose $M_\sigma = 4.5$ for the small earthquake magnitude, an estimate of the catalog completeness level for global events. The results for five megacities having populations of more than 1 million persons is shown in Table 1. We note that in the 100 km circles around these cities, many of the earthquakes that might affect them are hundreds of kilometers deep in the respective subduction zones, so no significant damage would result.

Table 1. EPS Values for Five Global Megacities for Earthquakes as of 23 May 2016

City	Region	Population (Millions)	Number of Regional Earthquake Cycles	EPS (Highest = 100)
Santiago, Chile (33.45°S, 70.66°E)	57°S to 18°S, 83°W to 56°W	5	313	86.3
Manila, Philippines (14.60°N, 120.98°E)	5°N to 20°N, 116°E to 129°E	1.6	196	74.0
Tokyo, Japan (35.69°N, 139.68°E)	25°N to 50°N, 125°E to 155°E	39	659	68.7
Taipei, Taiwan (25.o3°N, 121.63°E)	19°N to 27°N, 117°E to 124°E	2.6	95	47.4
Jakarta, Indonesia (6.17°S, 106.82°E)	12.25°S to 6.25°N, 90°E to 115°E	32	235	38.3

4. Sensitivity Testing

4.1. Statistical Distribution

As shown, for example, in Figure 2a, with 721 earthquake cycles, the data-derived statistical distributions are similar to a Poisson distribution with the same mean. For example, to determine whether the data-derived CDF in Figure 2a was in fact Poisson, we applied a standard Kolmogorov-Smirnoff test [*Stephens*, 1974]. We find an extremely small *p* value of 2.2×10^{-8}, which is the proportion of trials that the data-derived distribution would be indistinguishable from the Poisson. Therefore, we conclude that in fact the data-derived CDF is not Poisson.

We then fit the data-derived CDF to a Weibull statistical distribution having the same mean number of small earthquakes between large earthquakes (thus corresponding to the same *b* value), for both $M_\lambda \geq 5$ and $M_\lambda \geq 6$ from the California-Nevada region. These are shown as the green curves in Figures 2a and 2b. Weibull statistics have two parameters, a scale parameter τ and an exponent β. For the $M_\lambda \geq 5$ data, we found $\tau = 50.78 \pm 16.41$ and $\beta = 0.76 \pm 0.05$. For the $M_\lambda \geq 6$ data, we found $\tau = 492.13 \pm 128.82$ and $\beta = 0.83 \pm 0.14$.

4.2. Parameter Variations

In computing the EPS scores for M_λ earthquakes within a defined geographic region, we use a small earthquake magnitude threshold M_σ that we associate with the stated completeness magnitude for the catalog. In addition, we also use a large geographic region to build the data-derived CDF. So it is of interest to vary both these parameters to determine the effect on the EPS values.

Table 2 shows examples of nowcasts for the 100 km circle around Los Angeles. We vary the small earthquake magnitude M_σ from 3.0 to 3.5, and change the large earthquake region, for $M_\lambda \geq 5$ and $M_\lambda \geq 6$ from the California-Nevada region, as of 27 April 2016. The new region includes both California and Nevada also, with less of Northern California and Oregon and more of Baja California. Latitudes and longitudes are given in the table.

It can be seen from the data in Table 2 that the EPS values for $M_\lambda \geq 5$ are generally consistent across all parameter values and range from 46.6 to 59.1. The same is generally true for $M_\lambda \geq 6$; these EPS values range from 66.0 to 77.5. For $M_\lambda \geq 5$ we find that the EPS score is then $EPS_5 = 54.5 \pm 4.6$. For $M_\lambda \geq 6$ we find $EPS_6 = 72.7 \pm 4.1$.

For comparison, we have also computed the value of EPS within the 100 km Los Angeles circle just prior to the 29 March 2014 *M*5.1 La Habra earthquake and just prior to the 17 January 1994 *M*6.7 Northridge earthquake. These values were computed using the California-Nevada region and using a value $M_\sigma = 3.0$.

Table 2. Nowcasts for Los Angeles and California-Nevada Region

M_σ	EPS Values for $M_\lambda \geq 5$	EPS Values for $M_\lambda \geq 6$
Los Angeles		
3.0	59.1	77.5
3.25	59.1	77.5
3.5	49.2	73.8
California-Nevada Region[a]		
3.0	54.3	66.0
3.25	52.6	69.8
3.5	46.6	71.7

[a]Latitude: 32.0°N–45.0°N; longitude: 130.0°W–114.0°W.

For the La Habra earthquake, we found a value EPS$_5$ = 82.0. At the time of the event, 109 small earthquakes had occurred since the last $M_\lambda \geq 5$ earthquake had occurred within the Los Angeles circle on 29 July 2008, a M5.4 event. For the Northridge earthquake, we found a value EPS$_6$ = 81.0 just before the event. There had been 863 small earthquakes since the 9 February 1971 M6.6 San Fernando earthquake.

Since the nowcasting method describes only the present state of a system, there is not necessarily an implication for future activity, given an observed value of EPS. To judge future activity, the rate of accumulation of small earthquakes must be known. The reason for this is that while the minimum magnitude is $\boldsymbol{M_\lambda}$, the maximum possible magnitude of the future large earthquake is undetermined. According to the Gutenberg-Richter relation, we would generally expect more small earthquakes during an earthquake cycle as the magnitude of the terminating large earthquake increases.

5. Discussion and Summary

We have introduced a new method for characterizing the current state of the earthquake cycle, the nowcasting method, a concept borrowed from economics and finance. In this method, the idea is to more accurately characterize the current state of the earthquake cycle within a defined geographic area, for a defined large earthquake magnitude. Once the current state of the fault system in the defined area is better understood, it should be possible to better characterize the future state of the region and the calculation of forecast probabilities.

In our nowcasting method, we construct the cumulative probability distribution function (CDF) for the number of small earthquakes between large earthquakes during a sequence of earthquake cycles. These cycles occur in a large region around the area of interest. An advantage of this method is that it offers a systematic means of ranking locations as to their current exposure to the earthquake hazard. We calculate an earthquake potential score (EPS) that is found from determining the number of small earthquakes since the last large earthquake, then using the CDF found from the regional earthquake cycles. Specifically, if $n(t)$ is the number of small earthquakes since the last large earthquake, the EPS score is defined to be: EPS = $P\{n \leq n(t)\}$, where P is the CDF of small earthquakes occurring between large earthquakes.

We emphasize again that the nowcasting method yields no information about the future. For example, an EPS value of 50 in a location with few small earthquakes may reside in that state for many years. On the other hand, an EPS value of 50 in a very active location would see a rapid increase in EPS over a very short time.

We applied these ideas primarily to a 100 km circular region around Los Angeles, California, using data from the region from latitude 32°N to 45°N and from longitude 130°W to 114°W. We also showed that the methods are general and can be used to easily compute EPS scores for any global city in a seismically active region. We conclude that our method offers a fast and easily reproducible method to estimate the current level of progress through the earthquake cycle for any local region subject to seismic hazard worldwide.

References

Bevington, P. R., and D. K. Robinson (2003), *Data Reduction and Error Analysis in the Physical Sciences*, McGraw-Hill, Boston, Mass.

Field, E. H. (2007), Overview of the working group for the development of Regional Earthquake Likelihood Models (RELM), *Seismol. Res. Lett.*, *78*, 7–16.

Holliday, J. R., K. Z. Nanjo, K. F. Tiampo, J. B. Rundle, and D. L. Turcotte (2005), Earthquake forecasting and its verification, *Nonlin. Proc. Geophys.*, *12*, 965–977.

Holliday, J. R., J. B. Rundle, D. L. Turcotte, W. Klein, K. F. Tiampo, and A. Donnellan (2006), Using earthquake intensities to forecast earthquake occurrence times, *Phys. Rev. Lett.*, *97*, 238501.

Holliday, J. R., W. R. Graves, J. B. Rundle, and D. L. Turcotte (2016), Computing earthquake probabilities on global scales, *Pure Appl. Geophys.*, *173*, 739–748, doi:10.1007/s00024-014-0951-3.

Rundle, J. B., J. R. Holliday, W. R. Graves, D. L. Turcotte, K. F. Tiampo, and W. Klein (2012), Probabilities for large events in driven threshold systems, *Phys. Rev. E, 86*, 021106.

Scholz, C. H. (1990), *The Mechanics of Earthquakes and Faulting*, Cambridge Univ. Press, Cambridge, UK.

Stephens, M. A. (1974), EDF statistics for goodness of fit and some comparisons, *J. Am. Stat. Assoc.*, *69*(347), 730–737.

Varotsos, P. A., N. V. Sarlis, and E. S. Skordas (2002), Long-range correlations in the electric signals that precede rupture, *Phys. Rev. E, 66*, 011902.

Varotsos, P. A., N. V. Sarlis, H. K. Tanaka, and E. S. Skordas (2005), Some properties of the entropy in natural time, *Phys. Rev. E, 71*, 032102.

Varotsos, P. A., N. V. Sarlis, and E. S. Skordas (2011), *Natural Time Analysis: The New View of Time*, Springer, Berlin.

Weibull, W. (1951), A statistical distribution function of wide applicability, *J. Appl. Mech.*, *18*, 293–297.