Has the global risk of big earthquakes recently increased?

Peter M. Shearer
Scripps Institution of Oceanography, U.C. San Diego

Philip B. Stark
Department of Statistics, U.C. Berkeley

30 September 2011

Abstract

No.

1 Introduction

The above-average rate of earthquakes of magnitude 8 and larger in recent years (e.g., ref. 1) has prompted speculation that the underlying rate of earthquake activity has changed (2–5), that is, that the observed apparent rate fluctuation is larger than would be expected for a homogeneous random process. Similarly, the recent 2011 Tohoku-Oki, Japan, $M_{9.0}$ earthquake, together with the 2004 $M_{9.0}$ Sumatra-Andaman earthquake and the 2010 $M_{8.8}$ Maule, Chile, earthquake, have fueled concern that these giant quakes may not have been independent events (see discussion in ref. 6). Temporal earthquake clustering, including aftershock sequences, is well known at local and regional scales. However, whether earthquake catalogs, with aftershocks removed, follow a temporal Poisson process is a longstanding area of research in seismology (7–11).

True earthquake rate changes at global scales would have important implications for assessments of seismic hazard and our understanding of how faults interact. Here we ask whether the recent elevation in large earthquake activity is statistically significant and the larger question of whether the locally declustered global catalog is Poissonian. Using cataloged events from 1900 to 2011, we address this question three ways: (1) Plotting earthquake activity versus time to identify apparent anomalies in present and past rates of large earthquakes; (2) Performing Monte Carlo tests to estimate the probability of specific observed anomalies if seismicity were Poissonian with the observed average occurrence rate; and (3) Testing whether the locally declustered catalog is statistically distinguishable from a realization of a homogenous Poisson process, using three statistical tests. Our main conclusion is that the observed fluctuations in the rate of large earthquakes ($M \geq 8$) are not surprising if global seismicity follows a Poisson process with constant expected rate. Moreover, the recent rates of smaller earthquakes ($7 \leq M \leq 8$) are near their historic norms and it is difficult to devise a physical mechanism that would increase the rate of the largest
earthquakes but not the rate of smaller earthquakes. This suggests that the risk of large
earthquakes in regions far from recent enhanced activity is no higher today than it has been
in the past.

2 Catalog and local declustering method

We use moment magnitudes ($M_w$) and times from the PAGER-CAT catalog (12) for 1900 to
6/30/2008 seismicity, and the PDE and PDE-W catalogs (available from the USGS/NEIC
website, http://earthquake.usgs.gov/earthquakes) from 7/1/2008 to 8/13/2011. We con-
sider only $M \geq 7$ events to reduce catalog completeness issues that may dominate results at
smaller magnitudes. Non-uniformity in earthquake magnitude assignments is a serious issue
in estimating earthquake rate changes (13, 14). PAGER-CAT attempts to use consistent mo-
ment magnitude estimates (see http://earthquake.usgs.gov/research/data/pager/PAGER
CAT Sup.pdf), but the catalog might still have artificial changes in rate related to magnitude estimation.

Separating triggered aftershock seismicity from ‘background’ seismicity is not trivial and
a variety of methods have been proposed. Because our focus here is on the global scale, we
adopt the conservative and simple approach of removing events for which preceding larger
earthquakes occur within 3 years and 1000 km. This removes many events that might
not traditionally be classified as aftershocks. For example, we remove both the March
2005 $M$ 8.6 and September 2007 $M$ 8.5 Sumatra earthquakes, retaining only the December
2004 $M$ 9.0 Sumatra-Andaman earthquake. We do this because we only want to consider
whether distant events are correlated, such as the 2010 $M$ 8.8 Chile and 2011 $M$ 9.0 Japan
earthquakes, not whether regional-scale clustering may exist. Thus, it is important to
decluster events at distances of less than about 1000 km because they are not independent
from a global perspective. We explore the effects of changing these declustering criteria
below.

3 Magnitude versus time

Figure 1 shows earthquake magnitudes versus time and smoothed yearly rates of $M \geq 8$,$
M \geq 7.5$, and $M \geq 7$ activity. As expected, there are many more small earthquakes than
large earthquakes, consistent with the Gutenberg-Richter power law. There are only 16
earthquakes of $M \geq 8.5$, which limits the power of statistical tests of whether these giant
events cluster in time. The eye is a poor judge of randomness and tends to find patterns in
random sequences (e.g., ref. 15). Thus, the simple appearance of clustering should not be
considered convincing evidence for non-randomness.

Nonetheless, past researchers have pointed to several possibly anomalous features that
are visible in this plot. First, there were a number of very large $M \geq 8.5$ earthquakes between
1950 and 1965. Second, there was a dearth of such large earthquakes in the 38 years from
1966 and 2003. Finally, since 2004 there has been an elevated rate of $M \geq 8$ earthquakes:
The five-year running average is at a record high, although there have been rates nearly
as high in the past. These anomalies are evident only for the largest earthquakes and are much weaker or absent for smaller earthquakes. This implies that if the large earthquake clustering is caused by a physical mechanism, the mechanism must affect $M \geq 8$ earthquakes without changing the rate of smaller events. This is inconsistent with the triggering behavior implied by aftershock sequences, which are observed to have Gutenberg-Richter magnitude-frequency relationships reflecting a preponderance of smaller events (e.g., ref. 16).

4 Monte Carlo tests

How statistically significant are these anomalies for the large earthquakes? Assessing this is complicated by the fact that virtually every realization of a random process will have features that appear anomalous. If the statistical test is chosen after looking at the data, the true significance level or $p$-value can be substantially larger than the nominal value computed as if the test had been chosen before collecting the data. Nonetheless, it is illuminating to find the probability of observed anomalies under various assumptions about the underlying process. For instance, one might assume that seismicity follows a homogeneous Poisson
process with expected rate equal to the observed rate in the catalog, and generate a series of
random catalogs. The fraction of these catalogs that have anomalies like those observed—or
even more “extreme”—is an estimate of the $p$-value of the hypothesis that seismicity satisfies
the assumptions in the simulation, which include conditioning on the estimated rate.

This approach was used by Bufe and Perkins (2) to assess the 1950 to 1965 peak in great
earthquake activity. They estimated that there was only a 4% chance that the three $M \geq 9$
earthsquakes observed through 2001 would occur within an 11.4 year period for a 100 year
catalog, and that there was only a 0.2% chance that seven of the nine $M \geq 8.6$ earthsquares
observed through 2001 would occur within any 14.5 year period. We have not recalculated
these probabilities, but note that they are likely underestimates for at least two reasons.
First, the rate of $M \geq 9$ and $M \geq 8.6$ earthquakes that Bufe and Perkins used in their
calculations is almost certainly lower than the true long-term rate, as activity since 2001
has shown. Second, they appear to have selected details of their statistical tests, such as
the magnitude thresholds, to maximize the apparent anomaly. As mentioned above, this
causes the $p$-value or significance level to appear to be smaller than it really is, taking into
account the post hoc selection.

Bufe and Perkins (2) also considered the gap in $M \geq 8.4$ earthquakes between 1966 and
2001 and estimated that a 36-year gap would occur in only 0.5% of random catalogs of 18
$M \geq 8.4$ earthquakes. We show below that this gap is perhaps the most anomalous feature
in the global catalog. However, the 0.5% value is misleadingly low because: (1) Bufe and
Perkins included among the 18 events a July 23, 2005 $M 8.5$ earthquake that is likely an
aftershock of a nearby $M 8.6$ earthquake occurring 14 days earlier, and (2) Bufe and Perkins
apparently selected the $M \geq 8.4$ cutoff to maximize the apparent anomaly (there were three
$M 8.3$ earthquakes between 1966 and 2001).

We use Monte Carlo simulations to estimate the probability both of the recent elevated
large earthquake activity and of the gap that preceded it, under the null hypothesis that
seismicity follows a Poisson process that generates exactly as many events as were in fact
observed. That is, under the null hypothesis, the number of events is given and the times
of these events are independent, identically distributed (iid) random variables, all with a
uniform distribution on the interval in days $[0, 40,767]$. Our estimates are based on 100,000
random catalogs simulated from that joint distribution. The estimated probabilities are the
fractions of those 100,000 catalogs that have the apparent anomaly at issue, for instance,
the fraction that have at least a given number of events within an interval of a specified
length.

Nine of the 75 (after local declustering) $M \geq 8$ earthquakes occurred in the 2,269-day
Tohoku-Oki earthquake. Under the null hypothesis, there is about an 85% chance that at
least nine of 75 events would occur within 2,269 days of each other: The recent elevated rate
of large earthquakes is hardly surprising even if regionally declustered seismicity follows a
homogeneous Poisson process.

Three of 16 $M \geq 8.5$ declustered earthquakes occur during the 2,266 days between the
12/26/2004 M 9.0 Sumatra earthquake and the Tohoku-Oki earthquake. Under the null hypothesis, there is about a 97% chance that at least three of 16 events will occur within 2,266 days of each other. Even if we cherry-pick the lower magnitude threshold to be 8.8 (the size of the 2/27/2010 Maule earthquake), so that three of six \( M \geq 8.8 \) events occur in a 2,266-day interval, this has a 14% chance under the null hypothesis that regionally declustered seismicity is Poisson.

The lack of \( M \geq 8.5 \) events in the \( \sim 40 \) years between 2/4/1965 and 12/26/2004 is more anomalous than the recent elevated rate. Under the null hypothesis, the probability that 16 events in a 111-year interval would contain such a long gap is only about 1.3%. However, this feature was selected in retrospect, and it is essentially always possible to find a feature of any specific realization of a random process that appears improbable—that only a small fraction of random realizations would have. Hence, this gap is hardly evidence that the underlying process is nonuniform.

5 Tests of the Poisson hypothesis

A more general question is whether the earthquake catalog, after removing regional clustering, is statistically distinguishable from a realization of a homogenous Poisson process. We consider three tests, all of which condition on the number of events after declustering, so that the times of those events are iid uniform random variables under the null hypothesis. For additional tests, see ref. 17.

The first test compares the empirical distribution of the times with the uniform distribution. For each magnitude threshold, we determine the times of (locally declustered) events of that magnitude or above. We then perform a Kolmogorov-Smirnov (KS) test (e.g., ref. 18) of the hypothesis that those times are a sample of iid uniform random variables, estimating the \( p \)-value by simulation. The second and third tests use the chi-square statistic, but in different ways. Both partition the observation period into equal-length windows (we used \( N_w = 100 \) windows). These tests are more complicated than the KS test and require ad hoc choices, such as the lengths of the windows; the \( p \)-values depend on those choices.

The first chi-square test (the Poisson dispersion test) uses the fact that the conditional joint distribution of the number of events in different windows, given the total number of events, is multinomial with equal category probabilities. The test statistic for this test is proportional to the variance of the counts across windows. We estimated the \( p \)-value by simulating 100,000 catalogs with the same number of events that the declustered catalogs contained, and iid uniformly distributed event times.

The second chi-square test (the multinomial chi-square test) assesses how well the numbers of windows with various numbers of events agree with the numbers expected for homogenous Poisson seismicity. To perform the multinomial chi-square test, we calculated the average rate \( \lambda \) of events per period. On the assumption that seismicity follows a Poisson distribution with expected rate per period equal to the average rate \( \lambda \), let \( K^- \) denote the smallest integer such that the expected number of periods with no more than \( K^- \) events is
at least 5, and let $K^+$ denote the largest integer such that the expected number of periods with at least $K$ events is at least 5:\footnote{The threshold 5 was chosen because many textbooks state that the chi-square approximation to the null distribution of the chi-square statistic holds when the expected number of counts in every bin is at least 5. However, this method of selecting the categories for the chi-square test makes the number and nature of the categories depend on the observed data through the empirical rate of seismicity. Conditioning on the data in this way alters the null distribution of the test statistic, which we take into account by using simulation instead of relying on the chi-square distribution to find the $p$-value.}

$$K^- \equiv \min \left\{ k : N_w e^{-\lambda} \sum_{j=0}^{k-1} \lambda^j / j! \geq 5 \right\},$$  

\hspace{1cm} (1)

$$K^+ \equiv \max \left\{ k : N_w \left( 1 - e^{-\lambda} \sum_{j=0}^{K^+-1} \lambda^j / j! \right) \geq 5 \right\},$$  

\hspace{1cm} (2)

Define

$$E_k \equiv \begin{cases} N_w e^{-\lambda} \sum_{j=0}^{K^-} \lambda^j / j!, & k = K^- , \\ N_w e^{-\lambda} \lambda^k / k!, & k = K^- + 1, \ldots, K^+ - 1 , \\ N_w (1 - e^{-\lambda} \sum_{j=0}^{K^-} \lambda^j / j!), & k = K^+ . \\ \end{cases}$$  

\hspace{1cm} (3)

Let $X_{K^-}$ denote the number of periods that contain $K^-$ or fewer events. For $k = K^- + 1, \ldots, K^+ - 1$, let $X_k$ denote the number of periods that contain $k$ events. And let $X_{K^+}$ denote the number of periods that contain $K^+$ or more events. The test statistic is

$$\chi^2 \equiv \sum_{k = K^-}^{K^+} (X_k - E_k)^2 / E_k.$$  

\hspace{1cm} (4)

The values of $K^-$ and $K^+$ were 1 and 7 for the 330 $M \geq 7.5$ events and 0 and 2 for the 75 $M \geq 8.0$ events.

The Kolmogorov-Smirnov test and the Poisson dispersion test are sensitive to whether the rate varies with time, that is, to clustering. In contrast, the multinomial chi-square test is more sensitive to whether the distribution of the number of events in each window differs from the distribution expected if declustered seismicity were Poisson. For instance, if event times were equispaced, neither the KS test nor the Poisson dispersion test would reject the Poisson hypothesis, but the multinomial chi-square test would, given enough data.

Table 1 gives the results of all three tests. For the catalogs with aftershocks removed, the $p$-values range from 10\% to 94\%. If we remove all smaller events within 3 years and 1000 km, regardless of whether they are foreshocks or aftershocks, the $p$-values range from 34.3\% to 95.1\%. For both magnitude thresholds and all three tests, the observed distribution of declustered event times is consistent with the hypothesis that declustered event times follow a homogeneous Poisson process.

These results agree with those of Michael (19), who concluded that the global $M \geq 7.5$ and $M \geq 8$ catalogs (1900–2011), after aftershock removal, are well-described by a Poisson process. As mentioned above, it is essentially always possible in retrospect to find some feature of a data set that would be prospectively unlikely under the null hypothesis, so it
Table 1: Estimated \( p \)-values for the hypothesis that times of events in the original and regionally declustered catalogs are independent, identically distributed uniform random variables, for several hypothesis tests. Column 1: magnitude threshold. Column 2: events removed by declustering: “AS” means aftershocks were removed, “AS, FS” means both aftershocks and foreshocks were removed, and “none” means there was no declustering. Column 3: events in the catalog with magnitude greater than or equal to the threshold, after any declustering. Column 4: \( p \)-value for the Kolmogorov-Smirnov test. Column 5: \( p \)-value for the Poisson dispersion test using 100 windows. Column 6: \( p \)-value for the multinomial chi-square test using 100 windows. Each estimated \( p \)-value is the percentile of the observed value of the corresponding test statistic among 100,000 values of that statistic obtained in simulations in which event times are independent, identically distributed uniform random variables. The standard error of the estimated \( p \)-values is on the order of 0.16%.

<table>
<thead>
<tr>
<th>magnitude threshold</th>
<th>removed</th>
<th>events</th>
<th>( p )-value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>KS PD MC</td>
</tr>
<tr>
<td>7.5</td>
<td>none</td>
<td>444</td>
<td>22.9% 24.1% 62.0%</td>
</tr>
<tr>
<td></td>
<td>AS</td>
<td>330</td>
<td>94.0% 88.8% 10.0%</td>
</tr>
<tr>
<td></td>
<td>AS, FS</td>
<td>268</td>
<td>82.3% 95.1% 56.3%</td>
</tr>
<tr>
<td>8.0</td>
<td>none</td>
<td>82</td>
<td>33.8% 79.1% 25.7%</td>
</tr>
<tr>
<td></td>
<td>AS</td>
<td>75</td>
<td>60.3% 89.4% 22.3%</td>
</tr>
<tr>
<td></td>
<td>AS, FS</td>
<td>72</td>
<td>49.0% 89.8% 34.4%</td>
</tr>
</tbody>
</table>

is essentially always possible—after looking at the data—to find or contrive some test that formally rejects the null hypothesis (in this case, for instance, a test based on the longest gap between \( M \geq 8.5 \) events). But the formal significance level or \( p \)-value would not be meaningful, because it does not take into account the “data snooping” involved in selecting the test.

6 Sensitivity to declustering parameters

We have shown that the global catalog of large earthquakes, with aftershocks or foreshocks and aftershocks removed as described above, is statistically indistinguishable from a homogeneous Poisson process. However, it is well known that earthquakes cluster in local and regional catalogs: There are swarms, foreshocks, and aftershocks. Thus, we should expect the \( p \)-value of the Poisson hypothesis to be lower for the original catalogs than for the declustered catalogs. This is the case: As table 1 shows, the \( p \)-values are generally rather smaller for the original catalogs than for the declustered catalogs (except for the multinomial chi-square test, which, as mentioned, measures something other than clustering). But the \( p \)-values are still not small: Even without declustering, the null hypothesis that event times follow a homogeneous Poisson process would not be rejected by any of these tests.

Since the original catalog is consistent with the Poisson hypothesis, our conclusions clearly do not depend strongly on details of the declustering method. For example, milder cutoffs of 400 days and 333 km give KS \( p \)-values of 56.5% and 35.9% for the \( M \geq 8 \) and \( M \geq 7.5 \) declustered catalogs (AS removed), respectively. The estimated \( p \)-values depend
on the choice of declustering parameters, but our main conclusions do not depend upon these details.

7 Discussion

Global clustering of large earthquakes is not statistically significant: The data are statistically consistent with the hypothesis that these events arise from a homogeneous Poisson process. However, it is possible that rate changes are at least partially responsible for the surplus of large earthquakes during 1950–1965 and 2004–2011 and for the intervening gap in activity. The long-term average rate of large earthquakes is uncertain. McCaffrey (20) argues based on global subduction zone properties that the expected rate of $M_9$ earthquakes may be only 1–3 per century, implying that the five $M_9$ earthquakes observed since 1900 exceed the expected number. Given the low rate of large earthquakes, there will not be enough data to place tight constraints on the long-term average rate and possible rate changes for many years.

The stability of earthquake magnitude estimates is also critical. Individual magnitudes are typically uncertain to about 0.1 units. For a Gutenberg-Richter $b$-value of one, a systematic increase in magnitude of 0.1 would increase the apparent rate of earthquakes by 25%. It is not our purpose here to revisit discussion of possible systematic changes in catalog magnitude assignments (e.g., 13, 14, 21) but simply to note that magnitude estimates matter for evaluating possible rate changes. For example, Engdahl and Villaseñor (21) wrote “Moreover, it was impossible to match the seismicity rates of the historical period to those of the modern period without making a reduction in the older magnitudes by about 0.2 units. However, final resolution of this problem is presently beyond the scope of this study so that, for example, the apparent higher seismicity rate during the 1940–1960 period (Fig. 3b) will remain problematic.” The uncertainty in catalog magnitude stability introduces hard-to-quantify errors in assessing possible long-term rate changes.

Since empirical earthquake data do not settle the question definitively, the case for global rate changes depends on the plausibility of physical mechanisms that might cause such changes. One possibility is that large earthquakes trigger other large earthquakes, like the triggering observed in local and regional aftershock sequences. Although static stress changes resulting from fault rupture decay rapidly with distance, dynamic triggering is often observed at much larger distances than the traditional aftershock zone (e.g., ref. 22). Surface waves generated by large earthquakes have been observed to trigger small ($M < 5$) earthquakes at global distances (e.g., 23, 24). However, Parsons and Velasco (25) find that the last 30 years of seismicity show no short-term (< 1 day) surface-wave triggering of larger ($M > 5$) earthquakes at distances beyond 600 to 1000 km. They conclude that the regional hazard of larger earthquakes is increased following a mainshock but that the global hazard is not.

Thus, although earthquake-to-earthquake triggering has been hypothesized to cause apparent global clustering of large earthquakes (5), such triggering of large earthquakes, if it
exists, would need to behave differently from ordinary aftershock sequences, which typically follow both the Gutenberg-Richter distribution of event sizes and Omori’s Law for decay of frequency with time. As discussed above, during periods with an above-average rate of large earthquakes, there has been no corresponding increase in the rate of smaller earthquakes. As just noted, Parsons and Velasco (25) show that there is no peak in large earthquake activity just after the surface waves from large earthquakes pass. And the above-average rates of large earthquake activity from 1950–1965 and 2004–2011 do not start with the flurry of activity typical in aftershock sequences. The largest and second-largest recorded earthquakes since 1900, the $M_{9.6}$ 1960 Chile and the $M_{9.2}$ 1964 Alaska events, were near the end of a period of enhanced activity, not near its beginning, and were followed by a notable gap in large earthquakes.

These observations imply that global earthquake clustering, if it has a physical explanation, is more analogous to seismic swarms than to mainshock–aftershock sequences. Swarms are spatially compact clusters of events that occur for a limited time and typically do not begin with their largest event. They are difficult to explain with standard earthquake-to-earthquake triggering models, and are often ascribed to slow slip or fluid flow near the swarm site (e.g., 26, 27); i.e., there is an underlying physical driving mechanism. But no physical mechanism has been proposed to explain possible global seismicity swarms that is physically plausible and that would not be detected in other observations. Although global cycles of earthquake energy release have been hypothesized (2, 28), there is as yet no observational evidence to support these ideas, other than apparent changes in seismicity rate, which, as we have shown here and Michaels (19) has shown, are not statistically significant.

Another hypothesis is that stress diffused through postseismic relaxation of the asthenosphere triggers events on a global scale. Pollitz et al. (29) suggested that four great subduction zone earthquakes from 1952 to 1964 in the Kurils-Aleutians-Alaska arc caused a stress pulse that reached California in 1985 and may have increased seismicity rates there. However, the predicted stress changes at large distances are small compared to those usually thought to trigger earthquakes (e.g., ref. 30). An additional difficulty with this explanation for global clustering of large earthquakes is that a stress pulse would travel too slowly to cause the observed global grouping of large earthquakes. The 2010 $M_{8.8}$ Chile earthquake is too far away from the 2004 $M_{9}$ Sumatra earthquake for stress diffusion to be a factor.

Our conclusion that the global risk of large earthquakes currently is not elevated is based both on the lack of statistical evidence that regionally declustered seismicity is temporally heterogeneous on a global scale and on the implausibility of physical mechanisms proposed to explain global clustering. The estimated global rate of very large ($M > 9$) earthquakes is still very uncertain because only five such events have occurred since 1900. The recent elevated rate of large earthquakes has increased estimates of large earthquake risk: The empirical rate of such events is higher than before. However, there is no evidence that the rate of the underlying process has changed. In other words, there is no evidence that the risk has changed, but our estimates of the risk have changed.

While there is little evidence that the global earthquake risk has changed in areas far
from recent activity, the current risk of large earthquakes is certainly above its long-term average in regions like Sumatra, Chile, and Japan, which have recently experienced large earthquakes. Finally, of course, even if the risk has not increased recently, that does not mean that the ongoing risk is small or should be ignored.

8 Acknowledgements

This research was supported by the USGS/NEHRP program and by the Southern California Earthquake Center.

9 References