California Earthquake Forecasts Based on Smoothed Seismicity: Model Choices

by Qi Wang, David D. Jackson, and Yan Y. Kagan

Abstract  We constructed 5- and 10-yr smoothed-seismicity forecasts of moderate- to-large California earthquakes, and we examined the importance of several assumptions and choices. To do this, we divided the available catalog into learning and testing periods and optimized parameters to best predict earthquakes in the testing period. Fourteen different 5-yr testing periods were considered, in which the number of earthquakes varies from 18 to 63. We then compared the likelihood gain per target earthquake for the various choices. In this study, we assumed that the spatial, temporal, and magnitude distributions were independent of one another, so that the joint probability distribution could be factored into those three components. We compared several disjoint test periods of the same length to determine the variability of the likelihood gain. The variability is large enough to mask the effects of some modeling choices. Stochastic declustering of the learning catalog produced a significantly better forecast, and representing larger earthquakes by their rupture surfaces provided a slightly better result, all other choices being equal. Inclusion of historical earthquakes and the use of an anisotropic smoothing kernel based on focal mechanisms failed to improve the forecast consistently. We chose a lower threshold magnitude of 4.7 for our learning catalog so that our results could be compared in the future to other forecasts relying on shorter catalogs with a smaller magnitude threshold.

Online Material: Probability gain based on different conditions and earthquake rate forecast map.

Introduction

Our purpose is to explore the validity of assumptions built into various choices made in earthquake forecast models and to present improved forecasts based on results of that exploration. We adopt the smoothed-seismicity model, in which the rate of future earthquakes is estimated from past earthquakes only. Kagan and Jackson (1994) introduced the basic method, and it has been widely used since (e.g., Helmstetter et al., 2006, 2007; Kagan et al., 2007; Kagan and Jackson, 2010; Werner et al., 2010). Kagan et al. (2007) developed a forecast for southern California using historical and instrumental seismicity, extended sources approximating faulting geometry for large events, and anisotropic smoothing kernels based on focal mechanisms. Helmstetter et al. (2007) constructed a forecast for all of California, obtaining high spatial resolution by using recent instrumentally recorded earthquakes as small an M 2.0. Kagan and Jackson (2010) provided a forecast for California and Nevada using isotropic kernels centered on the epicenters of instrumentally recorded earthquakes in both the Preliminary Determinations of Epicenters (PDE) and Advanced National Seismic System (ANSS) catalogs. Werner et al. (2010) improved the program proposed by Helmstetter et al. (2007) and provided a revised long-term forecast together with a short-term forecast for California. All of the forecast models mentioned and many others have been or are being submitted to an organized project for prospective testing and comparison of such forecasts. The testing program began as Regional Earthquake Likelihood Models (RELM; Field, 2007; Schorlemmer and Gerstenberger, 2007; Schorlemmer et al., 2007; Schorlemmer et al., 2010), and it has since been subsumed under the broader name Collaboratory for the Study of Earthquake Predictability (CSEP; see Data and Resources section and Jordan, 2006).

In our smoothed-seismicity method, we estimate the geographically varying rate density of earthquakes by superposing spatial smoothing kernels centered at the epicenters of cataloged earthquakes. We thus assume that the future earthquake rate density will be proportional to a smoothed version of past seismicity. While time-dependence can be included, we assume here that earthquakes rates are independent of time over the length of the test period. Kagan et al. (2007) presented a 5-yr forecast of southern California earthquakes with M 5 or larger based on these assumptions. In this work,
we hypothesize that each earthquake in the catalog provides equal information, regardless of its magnitude, about future earthquakes. We then evaluate the validity of this hypothesis based on the results.

The modeling choices are as follows: (1) Should we include triggered events (generally aftershocks) in formulating the forecast? (2) Should we include historical earthquake data or older instrumental data? (3) Should we represent larger earthquakes as multiple fault patches? (4) Should we use anisotropic spatial kernels to represent focal mechanisms and the likely orientation of nearby faults?

The answers to these four questions help us to understand better the earthquake process and to construct a more effective forecast. In the following subsections, we discuss those questions in more detail.

Triggered Events

Earthquakes obviously cluster in space and time, and they can be divided into two categories: apparently spontaneous events and those probably triggered by previous earthquakes in the catalog. The spontaneous ones may be stimulated by earthquakes not in the catalog (Sornette and Werner, 2005; Wang et al., 2010b) or by external tectonic forces, but their times are effectively random for our purposes. As Wang et al. (2010a) stated, attempts to relate earthquake occurrence to external phenomena are confounded by the numerous triggered events.

There are reasons for or against declustering based on assumptions about the interaction of earthquakes with ambient stress. Declustering might be appropriate if spontaneous earthquakes are simply witnesses responding to a long-lived stress field without changing it much. Then triggered events, caused by short-lived effects of previous earthquakes, can give excessive weight to the locations of their spontaneous triggers. An alternative theory is that every earthquake is a perpetrator, causing a substantial persistent change in the stress field. In that case, every earthquake of a given magnitude might have equal effect, regardless of whether it was spontaneous or triggered.

Some forecasters have not declustered, and others have. Kagan et al. (2007) and Kagan et al. (2010) used all events, treating spontaneous and triggered events equally. Helmstetter et al. (2006, 2007) and Werner et al. (2010) used the Reasenberg (1985) method to decluster their learning catalog. As shown in Wang et al. (2010a), different declustering methods can give different results, and the Reasenberg (1985) method misses some apparently triggered events in the California catalog. In addition, the Reasenberg (1985) method is deterministic, and the choice of parameters in it is quite subjective.

Here, we compare forecast results for models with and without declustering. In the model without declustering, we treat all earthquakes above the magnitude threshold as equal, whether or not they appear to be triggered. In the model with declustering, we used stochastic declustering based on the

epidemic type aftershock sequence (ETAS) model of Ogata (1988, 1998) and Zhuang et al. (2005). In this method, for each event, a probability that the event was triggered by previous events in the catalog is estimated. The complement of that probability is the independence, or the probability that each event was spontaneous. Rather than removing triggered events from the declustered catalog, we multiply each smoothing kernel by the independence of the associated earthquake, thus down-weighting earthquakes that were probably triggered. In this paper, spontaneous catalog refers to a catalog has been weighted by spontaneous probability, and full catalog means all the events are equally weighted.

Historical and Early Instrumentally Recorded Earthquakes

The quality of earthquake recording in California has improved markedly with time. Relatively consistent reporting of large earthquakes began in about 1850, when California’s population increased rapidly, partly in response to the California gold rush. We thus treat 1850 as the beginning of the historical earthquake catalog. Instrumental recording of earthquakes began in earnest with the installation of seismic networks in the early 1930s; we consider 1932 to be the beginning of the early instrumental catalog. We treat 1981, about when digital computers had a profound effect, as the beginning of the modern instrumental catalog. Clearly, these dates are somewhat arbitrary, and we use them simply to divide the catalog into substantially different parts. The choice of starting date for the learning catalog obviously involves a trade-off between accuracy and the need to sample the earthquake process completely. The catalog is clearly not complete at our lower threshold of 4.7 during the historic period, but that period included the three largest earthquakes (1857, 1872, and 1906). The early instrumental period has a substantially more complete catalog but still not at $M \geq 4.7$. That period included the largest instrumentally recorded event, in 1952. The catalog is complete almost everywhere in California at $M \geq 4.7$ after 1981, but that period misses the four largest-known events. Kagan and Jackson (1994) and Kagan et al. (2007) suggested that the optimum time span of the learning catalog should increase as the time span of the test window. Yesterday’s events may provide a good forecast of today’s, but a poor forecast for the next few decades. What is the appropriate learning period for 5- and 10-yr forecasts? We approach this question by constructing forecasts using learning periods starting in 1850, 1932, and 1981 and comparing results for nonoverlapping test periods.

Multiple Fault Patches

Large earthquakes are not accurately represented by point sources due to their large rupture dimensions. Collections of subevents represented as rectangular slip patches, so-called extended sources, are essential to represent the stress perturbations and ground motion from large earthquakes. Kagan et al. (2007) and Wang et al. (2009) used the
extended sources to represent some large earthquakes in California. The importance of the rupture surface representation depends on whether earthquakes in the catalog are acting as witnesses only or as perpetrators. If they are perpetrators, then clearly the full rupture extent is important. For the witness model, the answer is not so clear. The epicenter might best reveal the persistent stress field, but the extent of the rupture might also depend on that stress field. To address this question, we have constructed and tested forecast models that treat earthquakes as point sources at their epicenters and, alternately, as extended rupture sources.

Anisotropic Spatial Kernels

Kagan and Jackson (1994), Jackson and Kagan (1999), and Kagan et al. (2007) employed anisotropic spatial kernels to represent focal mechanism information. They found that in subduction zones, the focal mechanisms of past earthquakes provide a reasonably accurate indication of the direction of the fault planes and the local stress directions controlling past and future earthquakes. They devised an anisotropic form of the spatial kernel, described in the following section, that utilized the focal mechanism advantageously in the forecast. Kagan and Jackson (2010), Helmstetter et al. (2006, 2007), and Werner et al. (2010) did not consider focal mechanisms in their forecast. We address this question by comparing pairs of forecasts, one pair employing a spatial kernel as used by Jackson and Kagan (1999) and the other constrained to be isotropic.

Data and Method

We used both epicenters and extended sources from the catalog by Wang et al. (2009) for both learning and test catalogs. The epicenter catalog covers from 1800 to 2007 with a minimum magnitude of 4.7 and maximum depth of 30 km. For this paper, we extended the catalog to the end of 2009 using the same methods. The point-source catalog was compiled from several previous catalogs by choosing the most reliable location and magnitude of each earthquake. This catalog is more complete than previous catalogs and includes more accurate magnitudes and hypocenter locations. Earthquakes less than $M_{\text{L}}$ 6.5 are treated as point sources. Some earthquakes larger than $M_{\text{L}}$ 6.5 are replaced by extended sources, which are ensembles of smaller events with equivalent total moment, distributed along the rupture surface (Kagan et al., 2007).

The test region employed in the RELM project (Schorlemmer and Gerstenberger, 2007, fig. 2) was used in this paper, and all forecasting and testing were performed in $0.1' \times 0.1'$ cells. The density of future seismicity in each cell was estimated by superposing the spatial kernels, centered on earthquakes in the learning catalog, and weighted by their independence as appropriate. The isotropic kernel (Kagan and Jackson, 2010) has the form

$$f(r) = A \times \frac{1}{r^2 + d^2},$$  \hspace{1cm} (1)$$

while the anisotropic one (Kagan et al., 2007) has the form

$$f(r) = A \times \frac{1}{r^2 + d^2} \times [1 + \delta \cos^2(\theta)],$$  \hspace{1cm} (2)$$

where $r$ is the epicentral distance, $d$ is the smoothing distance, $A$ is a normalizing factor, $\theta$ represents the orientation of the map point relative to the fault-plan azimuth for each event, and $\delta$ is the coefficient of azimuthal concentration. The normalization makes the sum of earthquake rates over the whole test area, after summing over all sources in the learning catalog, equal to the reciprocal of the test interval. Thus, the superposed kernels form a spatial probability density function for a single earthquake during the test period. This probability density is assumed constant over the test period. Magnitudes matter only in defining the learning and test catalogs; once in, earthquakes of all magnitudes are treated identically.

Model Optimization and Probability Gain

We optimize the parameters, $d$ in kernel (1) and $d$ and $\delta$ in kernel (2), by choosing those which best forecast earthquakes in the test catalog from those in the learning catalog according to a likelihood test. We calculate the likelihood and probability gain in the same way as Kagan and Knopoff (1977), Helmstetter et al. (2007), and Werner et al. (2010). The normalized spatial density is

$$\mu^*(i_x, i_y) = \frac{\mu(i_x, i_y) \times N}{\sum_{i_x} \sum_{i_y} \mu(i_x, i_y)},$$  \hspace{1cm} (3)$$

where $N$ is the total number of target events in the testing part and $\mu(i_x, i_y)$ is the nonnormalized spatial density in each cell.

We assume that the observed earthquakes in each cell are independent and follow a Poisson distribution with mean $\mu^*(i_x, i_y)$, so the log likelihood of the model is given by

$$\log(L) = \sum_{i_x} \sum_{i_y} \log(p[\mu(i_x, i_y), n(i_x, i_y)]),$$  \hspace{1cm} (4)$$

where $n(i_x, i_y)$ is the number of observed events during the test interval in the cell $(i_x, i_y)$, and the probability $p[\mu(i_x, i_y), n(i_x, i_y)]$ is given by the Poisson distribution

$$p[\mu(i_x, i_y), n(i_x, i_y)] = [\mu(i_x, i_y)]^{n(i_x, i_y)} \frac{\exp[-\mu(i_x, i_y)]}{n(i_x, i_y)!}. \hspace{1cm} (5)$$

In order to evaluate our forecast performance, the spatially uniform density model is used as the reference model. The probability gain per earthquake of our model relative to the reference model is given by

$$G = \exp\left(\frac{\log(L) - \log(L_0)}{N}\right),$$  \hspace{1cm} (6)$$
where \( \log(L_0) \) is the log likelihood of the uniform in the space model.

**Results**

In this paper, we used all earthquakes larger than \( M \geq 4.7 \) in the learning catalog to forecast events larger than \( M \geq 5.0 \) in the test catalog. We used three different start times for the learning catalogs: recent catalogs begin in 1981, instrumental catalogs begin in 1932, and full catalogs begin in 1850. Helmstetter et al. (2007) and Werner et al. (2010) made relatively successful forecasts using just the recent catalog. We used two different testing period durations of 5 yr and 10 yr. In each case, the end of the learning period was the start of the test period. All results are shown in Tables 1 through 4.

**Modeling Choices**

*Do Triggered Events Improve the Forecast?* Table 1 provides the probability gain per earthquake for 5-yr pseudoprospective tests over disjoint test periods. Table 2 shows similar results for 10-yr pseudoprospective tests. From Tables 1 and 2, we find that in most cases (42 out of 47 pseudoprospective tests) forecasts using just the recent catalog are significantly better than those using all events. Thus, inclusion of triggered events in the learning catalog degrades the forecast. Henceforth, we consider only those forecasts based on spontaneous learning catalogs, in which each event is weighted by its independence probability.

*Does Historic Seismicity Improve the Forecast?* For 5-yr tests, no. The probability gains of models 1, 2, and 3 in Table 3 show the different forecasting performance, measured as likelihood gain per event, using the recent, instrumental, and full catalogs. The probability gain for forecasts using the recent catalog is the highest, and the gain for the full catalog is the lowest, for the 5-yr test period beginning in 2005. The historical seismicity does not improve the 5-yr forecast. However, the results are somewhat different for other test intervals. Table 1 shows results for several disjoint 5-yr test periods. The full catalog provided the best forecast twice, the instrumental catalog beat the others once, and the most recent catalog won twice. From these ambiguous results, we cannot conclude that longer catalogs provide better forecasts.

For 10-yr tests, the answer to the question asked in the section head is no again. Table 2 and Table 3 show the results when all events are used in the learning part. Again, we did several different forecasting experiments with different test intervals. The full catalog did the best twice, the instrumental catalog beat the others four times, and the recent seismicity was always worst. For both 5-yr and 10-yr forecasts, the variability resulting from different choices of test interval obscured any clear conclusion about the optimum learning interval.

**Do Multiple Fault Patches for Larger Earthquakes Improve the Forecast?** In Table 3, we show that for both 5-yr and 10-yr tests, forecasts based on extended sources do a bit better than those based on point sources. We tried different testing intervals and found that these results do not depend on the specific testing interval.

**Do Anisotropic Smoothing Kernels Improve the Forecast?** Table 3 shows that forecasts employing anisotropic kernels (labeled A in column 2 of Table 3) generally had slightly higher likelihood-per-target-event than similar models using isotropic kernels (labeled S in column 2 of Table 3). For some anisotropic models, the optimal asymmetry, \( \delta \), was close to 0; these models were not included in the table.

### Table 1

<table>
<thead>
<tr>
<th>End of the Testing Catalog</th>
<th>1981 Full</th>
<th>1981 Spontaneous</th>
<th>1932 Full</th>
<th>1932 Spontaneous</th>
<th>1850 Full</th>
<th>1850 Spontaneous</th>
</tr>
</thead>
<tbody>
<tr>
<td>2009</td>
<td>21 3.57</td>
<td>5.35</td>
<td>3.36</td>
<td>4.48</td>
<td>3.35</td>
<td>4.37</td>
</tr>
<tr>
<td>2004</td>
<td>21 2.20</td>
<td>2.99</td>
<td>2.97</td>
<td>3.57</td>
<td>2.98</td>
<td>3.46</td>
</tr>
<tr>
<td>1999</td>
<td>23 3.40</td>
<td>3.75</td>
<td>2.70</td>
<td>2.79</td>
<td>2.43</td>
<td>2.33</td>
</tr>
<tr>
<td>1994</td>
<td>58 1.78</td>
<td>2.29</td>
<td>2.20</td>
<td>2.40</td>
<td>2.27</td>
<td>2.57</td>
</tr>
<tr>
<td>1989</td>
<td>39 2.58</td>
<td>2.71</td>
<td>3.33</td>
<td>3.72</td>
<td>3.69</td>
<td>3.98</td>
</tr>
<tr>
<td>1984</td>
<td>56 N/A</td>
<td>N/A</td>
<td>2.50</td>
<td>3.24</td>
<td>2.55</td>
<td>3.01</td>
</tr>
<tr>
<td>1979</td>
<td>34 N/A</td>
<td>N/A</td>
<td>1.81</td>
<td>2.21</td>
<td>1.90</td>
<td>2.20</td>
</tr>
<tr>
<td>1974</td>
<td>18 N/A</td>
<td>N/A</td>
<td>2.08</td>
<td>2.45</td>
<td>2.20</td>
<td>2.60</td>
</tr>
<tr>
<td>1969</td>
<td>33 N/A</td>
<td>N/A</td>
<td>1.77</td>
<td>2.13</td>
<td>1.75</td>
<td>1.90</td>
</tr>
<tr>
<td>1964</td>
<td>29 N/A</td>
<td>N/A</td>
<td>2.15</td>
<td>2.44</td>
<td>2.17</td>
<td>2.15</td>
</tr>
<tr>
<td>1959</td>
<td>47 N/A</td>
<td>N/A</td>
<td>2.96</td>
<td>2.87</td>
<td>2.51</td>
<td>2.12</td>
</tr>
<tr>
<td>1954</td>
<td>63 N/A</td>
<td>N/A</td>
<td>1.51</td>
<td>1.61</td>
<td>1.53</td>
<td>1.64</td>
</tr>
<tr>
<td>1949</td>
<td>31 N/A</td>
<td>N/A</td>
<td>1.68</td>
<td>1.87</td>
<td>1.47</td>
<td>1.50</td>
</tr>
<tr>
<td>1944</td>
<td>42 N/A</td>
<td>N/A</td>
<td>2.08</td>
<td>2.280</td>
<td>2.280</td>
<td>2.281</td>
</tr>
</tbody>
</table>

* \( N \) is the total number of events in the “test part.” Full catalog weights all recorded events in the learning catalog equally; spontaneous catalog weights each according to the probability that it was not triggered. In each row, the target events are the same for each forecast, and the probability gain is highlighted for the most successful forecast. Please note that 1981, 1932, and 1850 are the beginning years of the “learning catalog.”
because they are redundant with those employing isotropic kernels.

However, an additional penalty could be considered for more complex models, and anisotropic models (equation 2) have one more free parameter than isotropic ones. We used the Akaike information criterion (AIC; Akaike, 1974) to compare models. The statistic is

$$\text{AIC} = 2K - 2 \log(L),$$

where $K$ is the number of free parameters in the model, and $L$ is the maximized value of its likelihood function. The model with the smaller AIC value fits the data better. Allowing for the extra free parameter, we found no systematic advantage for the anisotropic smoothing kernel for 5-yr or 10-yr tests, regardless of the length of the learning interval or the choice of point-source or extended-source catalogs.

### What Determines the Likelihood Scores?

The log of the likelihood score (equation 4) involves a sum over all cells, which can be divided into those with earthquakes and those without. Alternatively, the log-likelihood score can be decomposed to show the contribution of each earthquake, and the combined contributions of all empty cells. Table 4 shows the results for each of the 5-yr forecast models 1–7 in Table 3, evaluated for the test period beginning in 2005. All of the values in Table 4 are negative because they represent logarithms of probabilities smaller than one. Thus, the log likelihood can be viewed as a cost; the positive values of likelihood gain per event in Tables 1

Table 2

<table>
<thead>
<tr>
<th>End of the Testing Catalog</th>
<th>1981</th>
<th>1992</th>
<th>1850</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Full</td>
<td>Spontaneous</td>
<td>Full</td>
</tr>
<tr>
<td>2009</td>
<td>42 2.62</td>
<td>3.75</td>
<td>3.36</td>
</tr>
<tr>
<td>1999</td>
<td>81 2.02</td>
<td>2.51</td>
<td>2.27</td>
</tr>
<tr>
<td>1989</td>
<td>95 N/A</td>
<td>N/A</td>
<td>2.71</td>
</tr>
<tr>
<td>1979</td>
<td>52 N/A</td>
<td>N/A</td>
<td>1.86</td>
</tr>
<tr>
<td>1969</td>
<td>62 N/A</td>
<td>N/A</td>
<td>1.80</td>
</tr>
<tr>
<td>1959</td>
<td>110 N/A</td>
<td>N/A</td>
<td>1.72</td>
</tr>
<tr>
<td>1949</td>
<td>73 N/A</td>
<td>N/A</td>
<td>1.69</td>
</tr>
</tbody>
</table>

*The method of obtaining this information is the same as the method mentioned in the note below Table 1.

Table 3

<table>
<thead>
<tr>
<th>Model Number</th>
<th>M</th>
<th>Learning Catalog</th>
<th>Testing Catalog</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$T_{\text{begin}}$</td>
<td>$T_{\text{end}}$</td>
<td>$N$</td>
</tr>
<tr>
<td>1</td>
<td>S</td>
<td>1981</td>
<td>2004</td>
<td>160.47</td>
</tr>
<tr>
<td>2</td>
<td>S</td>
<td>1932</td>
<td>2004</td>
<td>518.09</td>
</tr>
<tr>
<td>3</td>
<td>S</td>
<td>1850</td>
<td>2004</td>
<td>693.52</td>
</tr>
<tr>
<td>4</td>
<td>E</td>
<td>1981</td>
<td>2004</td>
<td>160.47</td>
</tr>
<tr>
<td>5</td>
<td>E</td>
<td>1932</td>
<td>2004</td>
<td>518.09</td>
</tr>
<tr>
<td>6</td>
<td>E</td>
<td>1850</td>
<td>2004</td>
<td>693.52</td>
</tr>
<tr>
<td>7</td>
<td>A</td>
<td>1932</td>
<td>2004</td>
<td>518.09</td>
</tr>
<tr>
<td>8</td>
<td>S</td>
<td>1981</td>
<td>1999</td>
<td>135.20</td>
</tr>
<tr>
<td>9</td>
<td>S</td>
<td>1932</td>
<td>1999</td>
<td>492.82</td>
</tr>
<tr>
<td>10</td>
<td>S</td>
<td>1850</td>
<td>1999</td>
<td>668.25</td>
</tr>
<tr>
<td>11</td>
<td>E</td>
<td>1981</td>
<td>1999</td>
<td>135.20</td>
</tr>
<tr>
<td>12</td>
<td>E</td>
<td>1932</td>
<td>1999</td>
<td>492.82</td>
</tr>
<tr>
<td>13</td>
<td>E</td>
<td>1850</td>
<td>1999</td>
<td>668.25</td>
</tr>
</tbody>
</table>

*The second column, $M$, indicates the methods used in smoothing seismicity: $S$ means isotropic smoothing kernel (equation 1), and point sources are used; $E$ means extended sources represented by isotropic kernels; $A$ means anisotropic smoothing kernels (equation 2), and point sources are used. $T_{\text{begin}}$, $T_{\text{end}}$, and $N$ are the beginning time, ending time, and the total number of events in “learning part” or “testing part.” $d$ and $\delta$ are adjusted to maximize log likelihood ($\log(L)$). $G$ is the probability gain per earthquake based on equation (6). Please note that all $M \geq 4.7$ events are included in “learning part,” and all $M \geq 5.0$ events are included in “testing part.” The log likelihood of the uniform space model $\log(L_o)$ is $-155.00$ for the 5-yr forecast and $-296.50$ for the 10-yr forecast. For some anisotropic models, the asymmetry was negligible (optimal $\delta$ close to 0). These are not included in the table, because they are redundant with models employing isotropic kernels.
Earthquake Number Distribution

We have so far discussed the spatial distribution of the earthquakes. The earthquake number distribution is another important issue. In our testing, we assumed that earthquakes during the test period are independent of one another, implicitly obeying the widely used Poisson distribution (e.g., Cornell, 1968). However, earthquake catalogs, including ours (Wang et al., 2009), show significant departures from Poisson behavior even after declustering. Kagan (2010) discussed various statistical distributions of earthquake numbers and found that the negative binomial distribution (NBD) fits well the annual earthquake number for the CalTech (CT) catalogue, 1932–2001, \( M \geq 5.0 \). Kagan (1996), Jackson and Kagan (1999), and Kagan and Jackson (2000) have used the NBD to describe the earthquake numbers in other catalogs. In addition, Werner et al. (2010) found that the AIC favors the NBD over the Poisson distribution to fit the ANSS catalog of magnitude 1932–2008, \( M \geq 4.95 \). The discrete Negative binomial probability mass function is defined as

\[
f(k) = \frac{\Gamma(r + k)}{\Gamma(r) k!} (1 - p)^r p^k, \tag{8}\]

where \( \Gamma \) is the gamma function, \( r > 0 \), and \( p \in (0, 1) \).

The mean of the NBD is \( \frac{\tau}{(1 - p)} \) and the variance is \( \frac{\tau p}{(1 - p)^2} \).

Based on Felzer (2008), the catalog is approximately complete for \( M \geq 5.0 \) earthquakes after 1957. Figure 1 shows that the NBD fits our earthquake catalog from 1957 better than the Poisson distribution. For these reasons, we propose to use the NBD in a consistency test (the “N-test” of Schorlemmer et al., 2007). There were 332 \( M \geq 5.0 \) and 99 \( M \geq 5.5 \) events from 1960 to 2009 in California. Therefore, we expect that in the future, the number of \( M \geq 5.0 \) should follow the NBD with a mean of 33.2 events and standard deviation of 14.2 events per five years, and \( M \geq 5.5 \) events should follow the same distribution with a mean of 9.90 and standard deviation of 7.65 events per five years.

New Grid-Based Forecast for Earthquakes in California

Here, we provide a 5-yr forecast for \( M \geq 5.0 \) earthquakes and a 10-yr forecast for \( M \geq 5.5 \) earthquakes in California. In these forecasts, we used only spontaneous events after 1932 and the isotropic smoothing kernel (equation 1). We used multiple point sources to represent large events. We optimized the \( d \) value in smoothing kernel (1): \( d \) is 2.1 km for \( M \geq 5.0 \) and 5.0 km for \( M \geq 5.5 \). We believe these values reflect the size of the rupture area and the consequent location uncertainty of earthquakes in those

*All forecasts use spontaneous learning catalogs.
magnitude ranges. Figure S1, available as an electronic supplement to this paper, shows the 5-yr forecast, with probability density per unit area represented by the color code. A similar figure for the 10-yr forecast is not shown. Because it is based on the same learning catalog and uses very slightly wider kernels, the 10-yr forecast map is nearly indistinguishable from Figure S1, available as an electronic supplement to this paper.

### Discussion and Conclusions

We also tested whether our results depend on the minimum magnitude choices in both learning and testing catalogs. Tables S1 and S2, available as an electronic supplement to this paper, provide the probability gain per earthquake for 5-yr pseudoprospective tests over disjoint test periods when the minimum magnitudes in the testing catalog are $M \geq 4.5$ and $M \geq 5.5$, respectively, and the minimum magnitude in the learning catalog is $M = 4.7$. Tables S3–S6, available as an electronic supplement to this paper, show similar results when the minimum magnitudes in the learning catalog are $M = 5.0$, $M = 4.5$, $M = 4.0$, and $M = 3.5$, respectively, and the minimum magnitude in the testing catalog is $M = 5.0$. From Tables S1–S6, available as an electronic supplement to this paper, using spontaneous earthquakes only in the learning catalog provides a better forecast, and inclusion of historical earthquakes fails to improve the forecast consistently, both of which are consistent with the conclusions in this paper.

We also performed conditional $L$ test (Werner et al., 2010) to test whether our 5-yr forecast is consistent with the observed data. We simulated 1000 earthquake catalogs based on the forecast probability and got the log-likelihood distribution of simulated catalogs. Then, we calculated the percentile of observed log likelihoods in the distribution. We found that all forecasts are consistent with the observations at 95% confidence level.

We gave the same weight to large earthquakes and small events in the learning catalogs because there is no clear rule to assign magnitude-dependent weights to different events and because the speculation of a specific rule might be dangerous. As shown in Table 3, the improvement from using extended sources is not significant, and it depends on this assumption. If higher weights were given to larger events, the effect of using extended sources might be more significant. The possible gain from giving large events higher weight will be investigated in our future research.

Helmstetter et al. (2007) and Werner et al. (2010) used earthquakes with magnitude as small as $M = 2.0$ in their learning catalog. We used $M = 4.7$ as the minimum magnitude threshold for three reasons. First, we wanted to explore the forecasting power of longer learning intervals. Second, we used the stochastic declustering method to classify the triggered events, and our preliminary results suggested that triggered events, if not declustered, degrade the forecast performance. Third, we wanted to draw a clear distinction between our forecasts and those of Helmstetter et al. (2007) and Werner et al. (2010) so that the effects of small events can be clearly determined by CSEP in fully prospective tests.

For moderate and large earthquakes in California, we assume that the earthquake rate is stationary in time, and we estimate the rate of future earthquakes directly from the mean past rate. This assumption is also used by Kagan et al. (2007), Helmstetter et al. (2006, 2007), and Werner et al. (2010). The stationarity could be strongly influenced by short-term clusters of large events, but because these extreme quakes are rare, the stationarity assumption holds true most of the time. Wang, Schoenberg, et al. (2010) simulated earthquakes using an ETAS model and found that the total number of quakes reaches a stationary level when the time window is long enough. In addition, Wang et al. (2010a) showed that there is no strong evidence to reject the hypothesis that the spontaneous events in California occur at a constant rate from 1932 to 2007. One way to capture clustering effects better is to give more weight to the recent quakes compared to older ones. However, the effect is limited because we have shown that recent quakes do not always have more forecasting power, and giving different weights increases the complexity of the model. Another possible way to capture clustering effects is to consider stochastic process models such as ETAS. These models are widely used in short-term forecasting, but their effectiveness in the long term is yet to be fully explored. We take that as a challenge for the future. Kagan et al. (2010, ch. 6.2), describe a number of problems with ETAS models.

The magnitude completeness level cannot be ignored in seismicity studies. We adapted completeness information from Felzer (2008) in estimating the expected number of
quakes in the future. However, the historical earthquakes are not complete for the $M \geq 4.7$ level from 1850. We did not try to add those missed or unobserved historical events in our learning catalogs. We could estimate their total number from the magnitude-frequency distribution, but we do not know where they occurred. Missing earthquakes in the historical and early instrumental parts of the catalog would cause us to underestimate future seismicity where these missed events occurred. On the other hand, many earthquakes in the early part of the catalog occurred in locations that are now seismically quiet (as around the 1857 event). Thus, the long-term forecast for such locations may not be seriously biased. Also, if many of these missed events were triggered, their influence would be small. One way to compensate for catalog incompleteness is to give more weight to events where the duration of complete recording is short. This assumes that the spatial distribution of missed events matches that of observed ones. As a test, we weighted earthquakes by the reciprocal of the magnitude-dependent completeness duration estimated by Felzer (2008). The results are shown in Table S7, available as an electronic supplement to this paper, and are consistent with what we have concluded in our paper.

In our model, we assume that earthquakes are independent and that their rate is constant in time. These assumptions describe a Poisson process, and we have evaluated likelihood scores assuming a Poisson distribution. However, we found that the numbers of events in disjoint time intervals are fit better by an NBD than by a Poisson (Fig. 1). Clearly, there is an inconsistency. We believe it is minor because the spatial distribution is assumed independent of the temporal one. Furthermore, the likelihood score is still useful, even with the incorrect Poisson assumption, just as the chi-square distribution is a useful misfit criterion even for non-Gaussian data. In addition, the number of earthquakes in every year or 5-yr period is not independent due to the triggering. This influence of triggering is of interest for future studies.

We did not explore the uncertainty caused by errors of location and magnitude in our earthquake catalog. We will study this important effect in the future.

Two earthquakes larger than 6.0 occurred in our forecast region since 2010. The first is the $M$ 6.5 offshore northern California earthquake of 10 January 2010 which occurred in one of the areas with the top 5% highest forecasted probability. The other one is the $M$ 7.2 Sierra El Mayor earthquake of 4 April 2010, which occurred in one of the areas with the top 6% highest forecasted probability.

Data and Resources

The catalog of earthquakes in California comes from Wang et al. (2009). Some plots were made using the Generic Mapping Tools version 4.3.1 (www.soest.hawaii.edu/gmt, last accessed April 2010). To access the CSEP web site, go to http://www.cseptesting.org/node/55 (last accessed May 2010).

Acknowledgments

We thank Marco Mucciarelli and an anonymous reviewer for very constructive reviews. We appreciate support from the National Science Foundation through grants EAR-0711515, EAR-0409890, EAR-0944218, and from the Southern California Earthquake Center (SCEC). SCEC is funded by NSF Cooperative Agreement EAR-0529922 and USGS Cooperative Agreement 07HQAG0008. The SCEC contribution number for this paper is 1450.

References


Department of Earth and Space Sciences
University of California
595 Charles E. Young Dr.
Los Angeles, California
qiwang@ucla.edu

Manuscript received 13 May 2010