Whole Earth high-resolution earthquake forecasts

Yan Y. Kagan and David D. Jackson

Department of Earth and Space Sciences, University of California, Los Angeles, CA 90095-1567, USA. E-mail: ykagan@ucla.edu

Accepted 2012 April 23. Received 2012 April 23; in original form 2011 December 11

SUMMARY
Since 1977 we have developed statistical short- and long-term earthquake forecasts to predict earthquake rate per unit area, time and magnitude. The forecasts are based on smoothed maps of past seismicity and assume spatial and temporal clustering. Our new program forecasts earthquakes on a 0.1° grid for a global region 90N–90S latitude. We use the PDE catalogue that reports many smaller quakes (M ≥ 5.0). For the long-term forecast we test two types of smoothing kernels based on the power law and on the spherical Fisher distribution. The effective width of these kernels is much larger than the cell size, thus the map discretization effects are insignificant. We employ adaptive kernel smoothing which improves our forecast both in seismically quiet and active areas. Our forecasts can be tested within a relatively short-time period since smaller events occur with greater frequency. The forecast efficiency can be measured by likelihood scores expressed as the average probability gains per earthquake compared to spatially or temporally uniform Poisson distribution. Another method uses the error diagram to display the forecasted point density and the point events. Our short-term forecasts also assume temporal clustering described by a variant of Omori’s law. Like the long-term forecast, the short-term version is expressed as a rate density in location, magnitude and time. Any forecast with a given lower magnitude threshold can be recalculated, using the tapered Gutenberg–Richter relation, to larger earthquakes with the maximum (corner) magnitude determined for appropriate tectonic zones.

Key words: Probabilistic forecasting; Probability distributions; Earthquake interaction, forecasting, and prediction; Seismicity and tectonics, Statistical seismology; Dynamics: seismotectonics.

1 INTRODUCTION
We have developed a time-independent (long-term) and time-dependent (short-term) earthquake forecast by using several earthquake catalogues (Kagan & Knopoff 1977, 1987; Kagan & Jackson 1994; Jackson & Kagan 1999). The importance of earthquake forecasting for seismic hazard and risk estimation and the difficulty of resolving basic differences in forecast models have motivated an international effort to report and test earthquake forecasts. That effort is organized by the Collaboratory for Study of Earthquake Predictability (CSEP; Schorlemmer & Gerstenberger 2007; Schorlemmer et al. 2010; Marzocchi & Zechar 2011).

Our purpose is to adapt a clustering model that we used to make testable forecasts over large regions of the western Pacific (Kagan & Knopoff 1977, 1987; Kagan & Jackson 1994; Jackson & Kagan 1999; Kagan & Jackson 2000) to include long- and short-term regional and worldwide forecasts in areas designated as natural laboratories by CSEP. Our earlier effort (Kagan & Jackson 2011) forecasts seismicity only within 75N to 75S latitude with low-resolution 0.5° cells. To get the test running as quickly as possible, we adopted arbitrary parameter values, similar to those that Kagan & Jackson (2000) used. Later we calculated optimized forecast parameters (Kagan & Jackson 2011, fig. 11) which were implemented here in our new high-resolution forecast.

We had to solve two problems to produce the high-resolution (0.1° cells) whole Earth forecasts: (1) Close to the poles we needed to use the exact spherical distance formula (e.g., Bullen 1979, eq. 5, p. 155) which requires about twice the computation time. (2) The size of the grid increased by a factor of 30 compared to our quasiglobal 0.5° forecast. Therefore, executing the forecast programs on the computers we employed previously (0.6 Ghz Alpha machines) required several days of the CPU time. On faster computers the time can be reduced by a factor of 10 at least.

Our new program forecasts earthquakes on a 0.1° grid for the global region 90N–90S latitude. We used the PDE earthquake catalogue (Preliminary determinations of epicentres 2011) that reports many smaller events (M ≥ 5.0) and a maximum cell size around the equator of about 11 km. The resolution of these forecasts is probably the maximum desirable: the earthquake depth distribution peaks around 10 km, and the PDE catalogue location uncertainty is about the same (Kagan et al. 2010).

Many local forecasts are now produced within the CSEP infrastructure (Schorlemmer et al. 2010). For these forecasts the
maximum resolution is usually 0.1°; therefore, localized forecasts can be extracted from ours (as in Kagan and Jackson 2010) and compared to those predictions. Moreover, a combined forecast employing local and global data can be calculated.

In this work we rigorously test long-term forecasts. The advantage of worldwide forecast testing is that many earthquakes will be available. Because these quakes occur over the globe, there are many practically independent sequences of events; thus seismicity is not dominated by only a few large earthquakes and their aftershocks, as happens in local and regional earthquake catalogues. Therefore, the testing results are more reliable and reproducible. Another problem we avoid is that earthquakes outside of a regional catalogue area may trigger earthquakes inside it (Wang et al. 2011).

2 METHODS AND DATA

For our long-term model we factor the rate density into spatial, magnitude and number distributions, assuming their independence from one another. We estimate the spatial distribution using all earthquakes above the assumed completeness threshold in the catalogue. For the magnitude distribution, we assume a tapered Gutenberg-Richter (TGR) distribution (Bird & Kagan 2004; Kagan et al. 2010). Any forecast with a given lower magnitude threshold applies as well to larger earthquakes with the maximum (corner) magnitude determined for appropriate tectonic zones. The map of tectonic zones shown in Kagan et al. (2010, fig. 1) assigns any 0.1° cell to a specific tectonic category. Then we can recalculate any forecast rate using the TGR distribution with the appropriate corner magnitude value (ibid., Table 1).

The PDE catalogue is distributed in preliminary form within a few minutes after each earthquake and in the final form with a few months latency. The catalogue reports earthquake size using several magnitude scales, providing the body-wave ($m_b$) and surface-wave ($M_S$) magnitudes. The moment magnitude ($m_w$) estimate has been added recently. Depending on the time period and region, the completeness threshold is on the order 4.5 to 4.7 (Kagan 2003; Romashkova 2009). To construct our smoothed seismicity maps, we need a single standard magnitude estimate for each earthquake. Like Kagan & Jackson (2011), in this work we use the maximum magnitude among those shown as a standard magnitude for each event.

3 LONG-TERM RATE DENSITY ESTIMATES

3.1 Smoothing kernel selection

Our previous regional and global forecasts (Kagan & Jackson 1994; Jackson & Kagan 1999; Kagan & Jackson 2000, 2011) were based on fixed kernel smoothing. We selected a fixed kernel with the degree of spatial smoothing controlled by the function which is asymptotic to a power law at distances much larger than $r_c$,

$$f(r) = \frac{1}{\pi} \times \frac{1}{r^2 + r_c^2},$$

where $r$ is epicentre distance, $r_c$ is the scale parameter of about 7.5 km and $r \leq 1000$ km (Kagan & Jackson 2011).

Unfortunately, the 1000 km distance limit causes sharp discontinuities in the smoothed maps as can be seen in Fig. 4 around the Hawaii islands. These discontinuities can be avoided if we use a kernel with the whole Earth support. However, in this case the choice of available kernels is significantly restricted. If we employ a power-law kernel like (1), its normalization on a sphere involves the application of cumbersome hyper-geometric functions. Practically, the best simple expression for a spherical surface kernel is the Fisher distribution (Fisher 1953, p. 296; Watson 1983, p. 14–15; Fisher et al. 1987, eqs 4.19–4.23; Mardia & Jupp 2000, eq. 9.3.4).

These authors propose expressions for the spherical Fisher distribution in a general form. For our purpose we assume that the distribution centre is at a pole of a sphere and the distribution has a rotational symmetry. Then the probability density function (PDF) of the spherical Fisher distribution is

$$f(\rho, \eta) = \frac{\kappa}{4\pi \sin \kappa} \exp (\kappa \cos \rho \times \sin(\rho) \times \phi(\eta)) = \frac{\kappa}{2\pi(e^\kappa - e^{-\kappa})} \exp (\kappa \cos \rho \times \sin(\rho) \times \phi(\eta)), \quad (2)$$

where $\eta$ is an azimuthal angle, $\phi(\eta)$ is angular azimuthal distribution density, $\rho = r/R$ is the epicentre distance in radians, $R$ is the Earth radius and $\kappa$ is a scale parameter.

It is more convenient to consider the Fisher distribution as depending only on distance, that is, we take $\phi(\eta) = 1/(2\pi)$. For the distance distribution only $1/(2\pi)$ term in (2) can be omitted. Then, the cumulative spherical Fisher distribution function is

$$F(\rho) = \frac{\exp [\kappa (\cos \rho - 1)] - 1}{e^{\kappa^2} - 1}. \quad (3)$$

For $\kappa > 100$, these equations can be simplified

$$f(\rho) \approx \kappa \exp [\kappa (\cos \rho - 1)] = \kappa \exp [-\kappa (\sin^2(\rho/2))]. \quad (4)$$

Since for small distance values $\sin(\rho) \approx \rho$, the above equation suggests that the probability density decays like a Gaussian function. In our California forecasts we applied the Gaussian as well as the power-law kernel smoothing distributions (Werner et al. 2011) and found that they yield similar results.

To forecast California seismicity, we also employed an adaptive kernel smoothing (Helmstetter et al. 2006, 2007; Werner et al. 2011) where the smoothing distance associated with an earthquake is proportional to the horizontal distance between the event and its $n^{th}$ closest neighbour (Silverman 1986, Ch. 2.6). Zhuang (2011) applied such variable kernels in the forecasts for the region around Japan. The number of the neighbours, $n$, is an adjustable parameter, estimated by optimizing the forecasts.

However, adaptive smoothing based on the distance between the epicentre and its $n^{th}$ neighbour has a certain disadvantage: if the $(n - 1)$th and $n^{th}$ events are separated by a large distance, the kernel width would jump from one value to another. Thus, we propose a method which in effect smooths the differences in the kernel width.

To carry out adaptive smoothing based on the Fisher distribution, we follow the advice of Silverman (1986, ch. 5.3): we first create an initial weight value estimate for earthquake epicentre location ($\chi_i$) by using eq. (4)

$$\chi_i = \sum_{j=1}^{n} f(\rho_{ij}), \quad (5)$$

where $n$ is the number of earthquakes and $\rho_{ij}$ is the distance between two epicentres.

Then local bandwidth factors ($\Lambda_i$) are defined

$$\Lambda_i(\rho) = \kappa (\chi_i/\chi_\theta)^{-\alpha}, \quad (6)$$
where $x_g$ is the geometric mean of $x_i$.

$$\log x_g = \frac{1}{n} \sum_{i=1}^{n} \log(x_i).$$

(7)

The $\kappa$-values in eqs (5) and (6) could be different, but as Silverman (1986, ch. 5.3) suggests and we tested (see later), the parameter value in the initial estimate does not significantly influence the final result. Since the larger $\kappa$-values correspond to a narrower kernel (see eq. 4), in actual computations we use an inverse of the $\Lambda_1(\rho)$ function in (6). The forecast density at any point $\vec{r}$ is then estimated by

$$\mu(\vec{r}) = \sum_{i=1}^{n} \Lambda_i(\vec{r} - \vec{r}_i).$$

(8)

In Fig. 1 we display two kernel examples: the densities for the power-law and for the spherical Fisher distribution. The density maximum for the Fisher law can be calculated by equating the derivative of its PDF (2) to zero. For large $\kappa$ we obtain

$$\cos \rho - \kappa \sin^2 \rho \approx 0.$$  

(9)

Since for large $\kappa$ (cos $\rho$) ↑ 1, the distance for the maximum is

$$\rho_m \approx \arcsin \sqrt{1/\kappa}.$$  

(10)

To demonstrate the difference between the kernels, Fig. 2 shows several examples of the cumulative distribution for various kernels: the Fisher distribution with the change of the $\kappa$ parameter and two variants of the power-law distribution—linearly normalized for a distance of 20,000 km and normalized on a sphere. The latter kernel normalization was adjusted so that the distributions coincide for smaller distances.

According to eq. (6), the two parameters $\kappa$ and $\alpha$ control the choice of kernel bandwidth; $\kappa$ is estimated iteratively from a starting value. But as described earlier, the final estimate depends very weakly on the starting one. The full spatial model includes one more parameter, $\epsilon$: a background rate to cover ‘surprise’ earthquakes in places far from any event locations in the catalogue.

### 3.2 Testing long-term forecasts

Kagan (2009) proposed a new method for testing the performance of a spatial point process forecast. It is based on a log-likelihood score for predicted point density and the information gain for events that actually occurred in the test period. The method largely avoids simulation and allows us to calculate the information score for each event as well as the standard error of each forecast. As the number of predicted events increases, the score distribution approaches a Gaussian distribution. To display the forecasted point density and the point events, we use the Error Diagram (ED) (Molchan 1990, 1997, 2010; Molchan & Kagan 1992; Zaliapin & Molchan 2004; Zechar & Jordan 2008, 2010). Rong & Jackson (2002) describe an event concentration diagram which contains the same information in a slightly different format. The general idea is to specify rate densities $\lambda_i$, calculated at points on a grid, using a hypothesis to be tested against a null hypothesis with rate densities $\xi_i$. Here $i$ are grid point numbers, and $N$ is the total number of grid points. It is also useful to construct non-overlapping cells about each grid point, integrate the rate densities in each cell, and normalize to 1.0 to get cell probabilities. We can then pick either the test or null hypothesis as the reference, and information scores relative to it can be computed from $\nu_i$, $\tau_i$, and $n_i$, the number of earthquakes in any cell (Kagan 2009; see also eqs (11-13) below).

To construct the error diagram, we sort the cells in decreasing order of reference probability and compute a cumulative distribution function (CDF) for each hypothesis. A theoretical error diagram is a plot of the complementary CDF (CCDF = 1 − CDF) for each hypothesis as a function of the CDF for the reference. The horizontal axis represents a kind of ‘water level,’ determined by a threshold.
Figure 2. Cumulative kernel functions: Solid and dashed diagonal lines correspond to power-law kernels (eq. 1) with \( r_s = 7.5 \text{ km} \); dashed line integrated over plane surface, solid line integrated over spherical surface. Sigmoidal solid lines from right to left correspond to the Fisher distribution kernels (eq. 3) with \( \kappa = 10, 100, 1000, 10000 \) and 100 000.

cell probability; what’s plotted is the probability that an event would be forecast by the reference model. The vertical axis represents the probability that the earthquake would not be forecast according to the model in question. If the reference CDF is \( p \), then by definition its error curve is \( 1 - p \). A forecast with a curve that lies below that of the reference curve predicts better, and will have a higher information score.

Empirical error curves and information scores can also be computed for earthquake catalogues by assigning epicentres to cells, again sorting the cells in decreasing order of reference probability, and plotting the empirical CCDF versus the CDF for the reference hypothesis.

We calculate the theoretical forecast information score, measured in the bits of information per event, as

\[
i_0 = \sum_{i=1}^{N} \nu_i \log_2 \left( \frac{\nu_i}{\tau_i} \right),
\]

where the reference \( \tau_i \) is the corresponding probability for a spatially uniform probability density. Under that assumption, \( \tau_i \) is proportional to cell area (Kagan 2009, eq. 14). \( I_0 \) contains no forecast earthquake data; rather, it measures the degree to which the forecast rate itself is concentrated relative to the null hypothesis of uniform rate density. The equation above can be easily generalized to measure the information score relative to any other reference hypothesis.

Using the forecasted rate values \( \lambda_i \) for cell centres in which earthquakes occurred, we compute the empirical information score

\[
i_1 = \frac{1}{n} \sum_{i=1}^{n} \log_2 \frac{\lambda_i}{\bar{\xi}},
\]

where \( n \) is the earthquake number during a forecast period and \( \bar{\xi} \) is a similar rate for the event occurrence according to the Poisson process with a uniform rate over a region (Kagan 2009, eq. 7). \( I_1 \) measures the degree to which the earthquakes in the test period are concentrated in cells where the forecast rate is high. If \( \lambda_i \) truly represents the event rate for each cell, then within the limit of infinitely many earthquakes \( I_1 \) converges to \( I_0 \).

As another option, instead of (12) we compute the empirical information score for the actual epicentre locations \( (\lambda_k) \)

\[
i_2 = \frac{1}{n} \sum_{k=1}^{n} \log_2 \frac{\lambda_k}{\bar{\xi}}.
\]

As Fig. 3 demonstrates, the values of \( I_1 \) and \( I_2 \) may also be significantly different (Kagan & Jackson 2011) unless the cell size is much smaller than the smoothing distance \( \rho \).

Similarly to (12) we also calculate the empirical score \( I'_1 \) for earthquakes which occurred in the training period. These values are significantly larger than those calculated for the forecast intervals. Therefore, as expected, our forecast ‘predicts’ better the locations of past earthquakes than those of future (target) events (see Kagan 2009, fig. 10 and its explanations).

Fig. 3 demonstrates how the fixed Fisher kernel was optimized using earthquake history from 1969–2005 to forecast earthquake occurrence in 2006–2010. The red curve displays the theoretical gain of a forecast (specificity): the scores increase for the narrower kernels.

However, to select the most internally consistent kernel, we should compare these scores with those obtained for the test period (2006–2010). The latter scores can be calculated for the cell centres where the earthquakes occur \( (I_j) \) or for actual epicentre locations \( (I_j) \). The intersection of \( I_0 \) and \( I_1 \) or \( I_2 \) scores suggests the best choice of kernel, because of the convergence property \( (I_1 \rightarrow I_0) \) mentioned earlier.
To test whether the initial estimate does not significantly influence the score values, as Silverman suggested (1986, ch. 5.3, see also eq. 7), we produced two forecasts. In one the initial estimate for earthquake epicentre locations (4) was executed using \( \kappa = 10000 \) and in the second \( \kappa \) was taken \( 100000 \). In both cases the actual forecast \( \kappa \) is selected \( 100000 \). The resulting scores are \( I_0 = 5.538; 5.420, I_1 = 3.865; 3.859, I_2 = 4.145; 4.145, I'_1 = 4.940; 4.972 \), respectively for both forecasts: differences connected with the change in the initial \( \kappa \) value are negligible.

In Fig. 4 we show the global long-term forecast made with the PDE catalogue using \( M \geq 5.0 \) earthquakes from 1969 to the present. The fixed power-law kernel (1) with \( r_s = 7.5 \) km is used on the \( 0.1^\circ \) grid. Fig. 5 shows the worldwide long-term forecast made using an adaptive Fisher distribution kernel. Comparing this with Fig. 4 it is obvious that the width of seismicity peaks is reduced at subduction zones. This is due to narrower kernels at concentrations of earthquake epicentres. Conversely, seismicity contours for the low activity regions are smoother for the adaptive kernels than for the fixed ones.

Optimizing the three parameters \( \kappa, \alpha \) and \( \epsilon \) is computationally challenging, so we’ve taken some shortcuts. We used a low-resolution grid (0.5 by 0.5 \( ^\circ \)) to estimate \( \kappa \) and \( \alpha \), and we made only a partial search for the maximum likelihood values of all parameters. The model used in Fig. 5 is based on that approximate optimization. We regard it as a respectable model worthy of comparison to others, but the values could probably be improved by further optimization.

4 TESTS AND FORECASTS OPTIMIZATION

Fig. 6 shows two theoretical and two empirical error curves relevant to the fixed power-law forecast. To spread the curves on the graph,

© 2012 The Authors, *GJI*, 190, 677–686
Geophysical Journal International © 2012 RAS
Figure 5. Adaptive Fisher kernel: long-term earthquake rates based on smoothed seismicity from the PDE catalogue 1969 to 2011 April. Adaptive smoothing kernel based on the Fisher spherical distribution (eqs 6–8) is used. Values of parameters are: $\kappa = 100,000$, $\alpha = 0.5$ and $\epsilon = 0.003$. Earthquake occurrence is modelled by a time-independent Poisson process.

we use as a reference the forecast based on the power-law kernel (eq. 1) with $r_s = 7.5$ km and the earthquake catalogue from 1969 to 2005. This display is equivalent to calculating the information scores by using forecast $\lambda_i$ as a reference density

$$I = \frac{1}{n} \sum_{j=1}^{n} v_j \log_2 \frac{\zeta_i}{\lambda_i},$$

(14)

where $\zeta_i$ is a rate density for all of the other point distributions (see Kagan 2009, fig. 10; Kagan & Jackson 2011, fig. 7 and their explanations).

The information scores reported below are still calculated with respect to the null hypothesis of uniform spatial rate density. The red line describes the reference hypothesis, so it satisfies $(\nu = 1 - \tau)$. The black line, corresponding to the null hypothesis, lies above the reference curve, indicating that its performance is worse than that of the power-law kernel. Also shown are the two empirical distributions, one based on the learning catalogue (1969–2005) and the other on the test catalogue (2006–2010).

Information scores for the curves shown in the plot are as follows: for the forecast (red curve) $I_0 = 3.13$; for earthquakes in the 2006–2010 test period (blue curve), $I_1 = 3.85$; for the backward forecast (magenta) $I'_1 = 4.46$. The probability gain $G$ can be calculated as $G = 2^I$ for all the information rate values. The backward forecast does best as expected, because the test and learning data are the same (Kagan 2009).

In principle, we could adjust our forecast so that $I_0 = I_1$, because for a correct forecast the empirical information score should converge to the theoretical one. In this case, the empirical score exceeds the theoretical one, indicating that the earthquakes are more concentrated than the forecast predicts. As shown in Fig. 4, this occurs when the kernel width is too large. Decreasing the value of $r_s$ would bring closer agreement. However, the sample size for $I_1$ is fairly small, so the agreement between $I_0$ and $I_1$ is close enough.

The major part of the numerical difference between different forecasts in Fig. 6 generally corresponds to the curves’ behaviour in the upper-left part of the plot; the higher $I$ values are associated with curves closer to the vertical axis (Kagan 2009). This is observed if only the upper-left corner of a diagram is plotted.
Figure 7. Error diagram $(\tau, \nu)$ for global long-term seismicity ($M \geq 5.0$) forecast relative to that of adaptive kernel forecast. Same format as in Fig. 6 is used. Solid black line, theoretical curve for spatially uniform rate density; red line, theoretical results for reference forecast based on adaptive Fisher kernel (eq. 2) with $\kappa = 10\,000$ and $\alpha = 0.5$; blue line, empirical results for test period 2006–2010; magenta line, empirical results for retrospective application to training period, 1969–2005.

Figure 8. Partial error diagram for fixed power-law kernel. Same as Fig. 6, except that reference forecast is the spatially uniform rate density model, horizontal axis is plotted on linear scale, and only the right half of the diagram is plotted.

In Fig. 6 the complete curves are shown, so this feature is not readily observable.

Fig. 7 displays error diagrams for adaptive kernel smoothing. The scores are higher than in the former case. The diagram shows the adaptive Fisher kernels more effectively approximate earthquake productivity than the fixed power-law kernel does in Fig. 6: $I_0 = 4.26$, $I_1 = 4.04$ and $I_1' = 4.57$.

In Figs 8 and 9 we show the same error diagrams in a different, more regular format (Kagan 2009, fig. 5; Kagan & Jackson 2011, fig. 6) to better display the approximation in low seismicity regions.
The adaptive kernel smoothing forecast is closer to the test period earthquake distribution. The scores for Fig. 8 are the same as for Fig. 6, and Fig. 9 has scores identical to Fig. 7.

Our forecasts do not predict well in regions where the previous earthquake rate was low. We have investigated whether our predictions can be improved by using the strain-based forecast (Bird et al. 2010a,b). We normalized the seismicity and strain-based spatial densities, and tested a linear or geometric combination that preserves the normalization. Preliminary results show that a 50/50 linear combination works best.

The tests indicated above are ‘pseudo-prospective’, in that except for $I'_1$, the tests are performed on data not used in the forecast. They provide a good template for truly prospective tests.

5 SHORT-TERM RATE DENSITY ESTIMATES

The short-term forecast in this work was carried out according to the technique described by Kagan & Jackson (2000, section 3.1.2) and by Kagan & Jackson (2011, section 3.3.3). Fig. 10 shows the calculated short-term forecast. Since our forecasts are calculated daily, they are only valid for the ‘next moment’ after the end of a catalogue (Kagan & Jackson 2000, 2011). To extend the forecast’s horizon in time, we must note the possibility that other earthquakes might occur in the time between the end of a catalogue and the forecast time. Accounting for this can be managed by the Monte-Carlo simulation of multiple generations of dependent events (Kagan 1973; Kagan & Knopoff 1977; Zhuang 2011).

The short-term forecast’s efficiency can be measured by its average probability gain per earthquake compared to the temporally uniform Poisson distribution (Kagan et al. 2010). For the short-term forecast, the gain $G$ is about 3.7 for the PDE catalogue relative to the temporally random but spatially localized null hypothesis (see table 4 in Kagan et al. 2010).

Figure 9. Partial error diagram for adaptive Fisher kernel. Same as Fig. 7, except that reference forecast is the spatially uniform rate density model, horizontal axis is plotted on linear scale, and only the right half of the diagram is plotted.

Figure 10. Earthquake short-term potential on 0.1° grid based on smoothed seismicity from the PDE catalogue since 1969. Earthquake occurrence is modelled by a branching temporal process controlled by Omori’s law type dependence.
Forecasting and testing in this way does not provide information for rapid decisions, but it is an effective prospective operation as long as the algorithms and parameters are fixed in advance. Short-term probabilities computed by these programs can be used in the operational earthquake forecasting discussed by Jordan & Jones (2010), van Stiphout et al. (2010) and Jordan et al. (2011).

6 DISCUSSION

The earthquake forecast technique described here continues and in a certain sense completes the work we have pursued since 1977 (see Section 1). Our question has been: how can we use the available seismicity record to optimally evaluate future seismic risk? Because of the fundamentally random character of earthquakes, such a forecast must be statistical.

Several problems must be solved to make such a forecast quantitatively efficient and applicable to the local and global environment. Relevant properties of earthquake occurrence need to be thoroughly investigated, including classical statistical earthquake distributions. How do Omori’s law and the Gutenberg-Richter relation apply in various regions? We also need to consider in a quantitative, objective manner the spatial distribution of seismicity (Kagan 2007). As with any application of measurement results to practical problems, we must have quantitative estimates for the measurement uncertainties and data completeness (Kagan 2003).

Our results (Bird & Kagan 2004, and references therein; Kagan et al. 2010) indicate that whereas the $b$-value of the Gutenberg-Richter relation has a universal value 0.93–0.98, the maximum or corner magnitude changes significantly for various tectonic zones, from a high of $M_{9.6}$ for subduction zones to a low of $M_{5.9}$ for oceanic spreading ridge normal faulting earthquakes. Moreover, maximum earthquake size was shown to be the same, at least statistically, for all the studied subduction zones (Kagan 1997). These maximum size determinations, combined with observed very large ($M \geq 9.0$) events in the other subduction zones before 1997 warn us that such an earthquake could occur in any major subduction zone, including the Sumatra and Tohoku (Japan) regions.

The Bird & Kagan (2004) regionalization is based, among other factors, on the earthquake focal mechanism. However, the PDE catalogue does not have this information for many earthquakes. Kagan et al. (2010) subdivide global seismicity into five tectonic zones without using focal mechanism information. These zones are displayed in their Fig. 1. The global zone assignment table with 0.1° grid is available at http://bemlar.ism.ac.jp/wiki/index.php/Bird%27s_Zones. The corner magnitudes for these zones are shown in table 1 by Kagan et al. (2010). These results allow us to extend our forecasts to evaluate the rate of the strongest earthquakes in each tectonic zone by extrapolating the tapered Gutenberg-Richter distribution (Bird & Kagan 2004; Kagan et al. 2010).

Our tests of forecast efficiency, described in Sections 3.2 and 4, are conducted mostly for $M_5$ to $M_6$ earthquakes. These earthquakes are less interesting from a practical point of view, because they usually cause little damage. Preliminary tests for the forecast of higher magnitude earthquakes have shown that, although $M \geq 7.5$ are mostly concentrated in trenches, the forecast skill for these events is similar to that for medium size earthquakes.

Our investigation of earthquake time behaviour (Kagan 2011) suggests that the Omori’s law parameters are also universal: even strong earthquakes are clustered in time instead of being quasi-periodic. We assume that in addition to earthquakes resulting from clustering, spontaneous earthquakes, approximately Poissonian, occur at a rate proportional to the regional strain rate. Thus, as clustered earthquakes decay in frequency according to the Omori law, the total rate becomes asymptotically Poissonian. Therefore, for tectonic plate interiors where the strain rate is low, only Omori’s law decay for historical and palaeo-earthquakes is usually observed. No transition to the Poisson rate can be unambiguously seen.

England & Jackson (2011) indicate that most earthquake related casualties occur not at tectonic plate boundaries such as subduction zones but in the continental interiors far from the boundaries. This presents a challenge for seismicity-based earthquake forecasts: if earthquakes are rare, our technique cannot give a good estimate of future seismic hazard with temporally limited earthquake catalogues. However, a forecast based on including the global strain rate data (Bird et al. 2010a,b) can significantly improve the seismicity-based forecasts, especially in regions where spatial geodesy and geologic data allow us to evaluate the strain rate accurately. Using the calibration provided by Bird & Kagan (2004) and Kagan et al. (2010), Bird et al. (2010a) converted the global strain rate to an earthquake hazard map.

For most of the active continents (Alpine-Himalayan belt) we have enough earthquake occurrence data and tectonic rate estimates to evaluate the seismic risk, albeit with some difficulty. However, for plate interiors where the fault deformation rates are less than 1 mm yr$^{-1}$, it is still a problem. In these regions only rough estimates of long-term earthquake rates can be obtained (Kagan 2011). However, by using our results, extrapolation of seismic activity in aftershock zones can be obtained.

As we indicated before, the forecast method described in this paper is a concluding step in the development of seismicity-based earthquake forecasts. Though a significant effort would be needed to obtain optimally chosen parameters for the smoothing kernel and to organize details of catalogue preparation and processing, we feel that the main scientific problems have been addressed. The forecast covers the whole surface of the Earth. The resolution of the forecast is unlikely to be significantly improved with the available data, and the adaptive feature of the smoothing kernel produces a nearly optimal adjusted forecast map. However, as discussed in Section 5, if a temporarily extended forecast is needed, then future seismic activity needs to be simulated. That may involve significant computation and would require greater resources than we have available.

7 DATA AND RESOURCES


ACKNOWLEDGMENTS

We are grateful to Peter Bird and Paul Davis of UCLA, and Stefan Hiemer of ETH Zurich for useful feedback and suggestions. Reviews by Warner Marzocchi and an anonymous reviewer have been helpful in revising the manuscript. We thank Kathleen Jackson who edited the manuscript. The authors appreciate support from the National Science Foundation through grants EAR-0711515, EAR-0944218 and EAR-1045876, as well as from the Southern California Earthquake Center (SCEC). SCEC is funded by NSF Cooperative Agreement EAR-052992 and USGS Cooperative Agreement 07HQAG0008. Publication 1634, SCEC.
REFERENCES


