*Geophys. J. Int.* (2000) **143,** 438–453

# Probabilistic forecasting of earthquakes

Yan Y. Kagan and David D. Jackson

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

Accepted 2000 June 29. Received 2000 June 22; in original form 2000 January 21

## SUMMARY

We present long-term and short-term forecasts for magnitude 5.8 and larger earthquakes. We discuss a method for optimizing both procedures and testing their forecasting effectiveness using the likelihood function. Our forecasts are expressed as the rate density (that is, the probability per unit area and time) anywhere on the Earth. Our forecasts are for scientific testing only; they are not to be construed as earthquake predictions or warnings, and they carry no official endorsement. For our long-term forecast we assume that the rate density is proportional to a smoothed version of past seismicity (using the Harvard CMT catalogue). This is in some ways antithetical to the seismic gap model, which assumes that recent earthquakes deter future ones. The estimated rate density depends linearly on the magnitude of past earthquakes and approximately on a negative power of the epicentral distance out to a few hundred kilometres. We assume no explicit time dependence, although the estimated rate density will vary slightly from day to day as earthquakes enter the catalogue. The forecast applies to the ensemble of earthquakes during the test period. It is not meant to predict any single earthquake, and no single earthquake or lack of one is adequate to evaluate such a hypothesis. We assume that 1 per cent of all earthquakes are *surprises*, assumed uniformly likely to occur in those areas with no earthquakes since 1977. We have made specific forecasts for the calendar year 1999 for the Northwest Pacific and Southwest Pacific regions, and we plan to expand the forecast to the whole Earth. We test the forecast against the earthquake catalogue using a likelihood test and present the results. Our short-term forecast, updated daily, makes explicit use of statistical models describing earthquake clustering. Like the long-term forecast, the short-term version is expressed as a rate density in location, magnitude and time. However, the short-term forecasts will change significantly from day to day in response to recent earthquakes. The forecast applies to main shocks, aftershocks, aftershocks of aftershocks, and main shocks preceded by foreshocks. However, there is no need to label each event, and the method is completely automatic. According to the model, nearly 10 per cent of moderately sized earthquakes will be followed by larger ones within a few weeks.

**Key words:** earthquake forecasting, Gutenberg–Richter relation, maximum seismic moment, seismicity kernel smoothing.

## 1 INTRODUCTION

Statistical studies (Kagan & Jackson 1991) show that earthquakes are clustered in both space and time and that this clustering is especially strong in the days and weeks following a strong earthquake. Here we address two different problems, short- and long-term forecasting, each involving different treatments of earthquake clustering. Because there is as yet no comprehensive model of earthquake occurrence, our forecasting procedures are derived from a variety of statistical, physical and intuitive (*ad hoc*) arguments. While the short-term forecast is based on a specific stochastic model of earthquake occurrence (Kagan 1991b), the long-term seismicity model is not yet developed to a stage where we could justify all our decisions. Thus the long-term forecast is essentially an empirical description of observed spatial clustering, and it has value only

 

to the degree that it can estimate well the probabilities of future earthquakes. In our long-term forecast we ignore temporal clustering and attempt to infer the degree of spatial clustering from the available earthquake history. We evaluate our long-term forecast annually, but in general the estimated rate density changes little from year to year, thus we assert that the forecast also serves well as an estimate of earthquake likelihood for periods of up to several decades.

Short-term forecasting considers temporal clustering of the sort that causes foreshock–main shock–aftershock sequences. In subduction regions these sequences may last for a few weeks or months (Kagan 1991b; Davis & Frohlich 1991), whereas in zones with a low tectonic deformation rate the sequences may last for centuries (Ebel *et al.* 2000). We treat short-term clustering as a perturbation to the long-term earthquake rate, and we update the short-term forecast daily.

Long-term forecasting may be appropriate when decisions such as insurance purchases, seismic upgrading or siting of permanent geophysical networks must be made using present information. For this problem we seek a probabilistic approach that will generally account for future clusters, even though we do not know the locations of specific events that will initiate these clusters. To solve this problem we estimate the probability of independent events and allow for dependent following events only in an 'average' sense. In our long-term procedure, there is no explicit time dependence. However, the estimated earthquake rate density depends on past seismicity, which will change slightly as earthquakes occur or as time elapses without earthquakes. The spatial clustering assumed in both the long-term and short-term forecasts is in some ways antithetical to the seismic gap model (e.g. McCann *et al.* 1979; Nishenko 1991; see also Kagan & Jackson 1995), which assumes that recent earthquakes deter future ones.

Our preliminary investigations indicate that following most moderate earthquakes (magnitude 6–7), the additional temporal increase of future seismic activity exceeds the time-independent (static) rate for a few weeks. Most of the events forecast by our short-term procedure would retrospectively be labelled as aftershocks. However, in real time we cannot reliably determine whether an event is larger or smaller than future ones. Furthermore, some dependent events will be larger than their predecessors, and would be normally labelled main shocks. Our procedure aims to forecast these events too.

Even our short-term probabilities generally do not rise high enough to satisfy the popular definition of earthquake prediction, yet they may reach levels that justify special attention by emergency personnel and by observational scientists. For instance, following a moderate earthquake, there is approximately a 10 per cent chance that a larger one will follow within a week (*cf.* Reasenberg 1999). Correlations between earthquakes also allow the probable focal mechanism to be estimated, allowing an early alert to possible tsunami generation.

Both our long-term and short-term forecasts are based on the most recent seismic record, updated within hours of significant earthquakes and published on the World-Wide Web.

As an exploratory step, in the beginning of 1999 we calculated long- and short-term hazard estimates for the northwest and southwest Pacific regions ($60.0^\circ \geq$ latitude $\geq 0.0^\circ$N, $170.0^\circ \geq$ longitude $\geq 110.0^\circ$E and $0.0^\circ \geq$ latitude $\geq 60.0^\circ$S, $170.0^\circ$W $\geq$ longitude $\geq 110.0^\circ$E, respectively). These regions account for 22 and 35 per cent of global seismicity, respectively. The general features of the forecast method are described in Jackson & Kagan (1999), hence we concentrate in this paper on technical details of procedure and on the preliminary results of the forecast testing. Earthquakes that occurred in 1999 in these regions are used for the testing.

## 2 LONG-TERM SEISMIC HAZARD ESTIMATES

We have developed a long-term forecast procedure based on earthquake data alone (Kagan & Jackson 1994, see also http://scec.ess.ucla.edu/ykagan.html). This procedure is based on the smoothing of past earthquake catalogues of seismic moment solutions. We assume that the smoothing function can be factored into kernel functions dependent on location, focal mechanism and earthquake size (*cf.* eq. 2 in Kagan & Jackson 1994 and eq. 1 in Jackson & Kagan 1999). That is, we let

$$\Xi(\mathbf{x}, m, \omega) = f(\mathbf{x}) f_{\mathrm{M}}(M) h(\omega), \tag{1}$$

where $\Xi$ is the rate of earthquakes per unit area, seismic moment and time, $\mathbf{x}$ is latitude and longitude, $M$ is the scalar seismic moment, $f$ is the spatial density function, $f_{\mathrm{M}}(M)$ is the moment distribution (see below) and $h(\omega)$ is the distribution of focal mechanisms, with $\omega$ a focal mechanism parametrization. In this work we use the focal mechanism function and its parametrization as described in Kagan & Jackson (1994). The functions $f_{\mathrm{M}}(M)$ and $h$ are normalized so that their integrals over moment and focal mechanism orientation, respectively, equal 1.0. Then, $f(\mathbf{x})$ is the rate density (number per unit time and unit area) of all earthquakes within the moment range of interest.

We consider the long-term earthquake rate $\Xi$ to be time-independent, although our estimates of it change with time, especially for catalogues of small time span. So far we have attempted to optimize only the spatial kernel of the earthquake forecast formula, to which the forecast is most sensitive.

### 2.1 Spatial dependence

The spatial kernel has units of earthquakes per unit area and time, and in our studies it applies to all earthquakes of moment $M = 10^{17.7}$ N m (magnitude 5.8) and greater. It is generally a function of the distance between earthquake epicentres or 'epicentroids', depending on which values are given in a particular catalogue. Here we define epicentroid as the vertical projection of the moment centroid onto the surface of the earth.

Many spatial smoothing techniques have been proposed to forecast future seismicity rates based on past earthquake data: cell averaging (Console 1998), smoothing kernels in the form of a disc (Kafka & Walcott 1998), a Gaussian function (Frankel 1995; Frankel *et al.* 1996) and power-law functions (Cao *et al.* 1996) have been proposed and tested.

Kafka & Walcott (1998) considered the problem of using the distributions of small earthquakes to predict the locations of future stronger events. They used disc-shaped kernel functions with uniform probability density over radii of 25, 50 and 100 km. Disc-shaped kernel functions have the disadvantage that forecasts made with them have spatial discontinuities, so the perceived success of the forecasts depends strongly on the radius of the discs. They can also lead to disputes over whether given earthquakes are within the discs.

Frankel (1995) and Frankel *et al.* (1996) proposed a model of seismic risk assessment for the conterminous United States (see also http://gldage.cr.usgs.gov/eq/). They used a Gaussian kernel to smooth seismic data. Cao *et al.* (1996) experimented with two types of smoothing kernel: a Gaussian one and the function $r^{-\lambda}$, where $r$ is the epicentral distance. They found that the latter kernel forecasts future earthquakes better. The value of $\lambda$ evaluated from the two-point epicentre correlation function in California is 1.2–1.4.

In a previous paper (Kagan & Jackson 1994) we used a spatial smoothing kernel proportional to inverse epicentroid distance $1/r$, except that it was truncated at short and very long distances. Since the integral of a power-law function over a plane diverges, the kernel function is truncated below a minimum distance, $R_{\min}$, and above a maximum distance, $R_{\max}$. The kernel is elongated along the fault plane, which is estimated from available focal mechanism solutions. An important feature of Kagan & Jackson's (1994) method is a jackknife-like procedure for testing the predictive power of the smoothing. In this procedure we optimize the parameters by choosing those values that best predict the second half of a catalogue, using a maximum likelihood criterion, from the first half. Based on these parameters we simulate thousands of earthquake catalogues to compare with a real second-half catalogue, and evaluate the likelihood function for both types of catalogues (Kagan & Jackson 1994). If the forecast is accurate, the likelihood value for the real catalogue should be in the middle range of the simulated values.

We have experimented with alternative kernels (Vere-Jones 1992; Woo 1996; Cao *et al.* 1996) that do not diverge at small epicentral distances but provide comparable effect. These kernels are more robust than power-law kernels, and they allow a more versatile smoothing of the available seismic history.

Here we employ kernels of the general form

$$f(r)=\mathrm{sgn}(\lambda-1)\left(\frac{\lambda-1}{\pi}\right)(r_s^{2(\lambda-1)})\,\frac{1}{(r^2+r_s^2)^{\lambda}} \quad \text{for} \quad \lambda\neq 1 \tag{2}$$

or

$$f(r)=\frac{1}{\pi}\left(\frac{1}{r^2+r_s^2}\right) \quad \text{for} \quad \lambda=1\,, \tag{3}$$

where $r$ is the epicentroid distance, $r_s$ is the scale parameter and $\lambda$ controls the degree of spatial smoothing. Preliminary calculations show that reasonable values of $\lambda$ lie in the range $0.5\le\lambda\le 1.5$. When $\lambda=0.5$, the kernel varies as the reciprocal of epicentroid distance for large $r$, as in our 1994 paper.

Integrated over two dimensions, the cumulative kernel functions are

$$F(R)=\mathrm{sgn}(\lambda-1)2(\lambda-1)r_s^{2(\lambda-1)}\int_0^R\frac{rdr}{(r^2+r_s^2)^{\lambda}}$$

$$=1-\left(1+\frac{R^2}{r_s^2}\right)^{1-\lambda} \quad \text{for} \quad \lambda\neq 1 \tag{4}$$

and

$$F(R)=2\int_0^R\frac{rdr}{r^2+r_s^2}=\log\left(1+\frac{R^2}{r_s^2}\right) \quad \text{for} \quad \lambda=1\,. \tag{5}$$

Although the kernel function converges as $r\to\infty$ for $\lambda>1$, the convergence is slow, so in these calculations we define the function $f(r)$ for $1.0\le\lambda\le 1.5$ up to the maximum distance $R_{\max}=1000$ km. The normalizing coefficients in eqs (2) and (3) have been modified accordingly. Several examples of new kernel functions are shown in Fig. 1, where they are compared with the kernel used in the previous work (Kagan & Jackson 1994) for the northwest Pacific. Our preliminary investigations have shown that two choices of the kernel parameters ($\lambda=1.1$, $r_s=15$ km and $\lambda=1.5$, $r_s=50$ km) are at least as effective in smoothing seismicity as the $1/r$ function used by Kagan & Jackson (1994).

In addition to the distance dependence, we multiply kernel (3) by an orientation function $D(\varphi)$ depending on the angle $\varphi$ between the fault plane of an earthquake and the direction to a map point (Jackson & Kagan 1999, p. 394):

$$D(\varphi)=1+\delta\cos^2(\varphi)\,. \tag{6}$$

The parameter $\delta$ controls the degree of azimuthal concentration (Kagan & Jackson 1994, Fig. 2).

The total spatial rate for a map point $j$ (or earthquake epicentre) at location $\mathbf{x}_j$ is calculated as the sum of contributions from all earthquakes closer than $R_{\max}$,

$$f(\mathbf{x}_j)=(1-C)\sum_i f(r_{ij})D(\varphi_{ij})+C\,, \tag{7}$$

where $C$ is a *surprise* coefficient, taking into account earthquakes in low-seismicity areas (see the next paragraph). The computation details are discussed in Kagan & Jackson (1994).

For some cells in NW and SW Pacific regions, no $M\ge 10^{17.7}$ N m (magnitude $\ge 5.8$) earthquakes occurred during the period 1977–1999 within $R_{\max}$ (1000 km) distance. Clearly, these zones are not immune from earthquakes forever. As in our earlier paper (Kagan & Jackson 1994), we assume that the background probability density is uniform over the whole region and integrates to 1 per cent of the total earthquake probability. In principle we could use a maximum likelihood procedure to estimate the surprise coefficient $C$. Here we simply adopted the above 1 per cent probability on an *ad hoc* basis. For zones in which earthquakes occurred during this period, the above probability has been augmented by the values calculated using kernel (3).

Functions $f(r)$ and $D$ in eqs (3) and (6) have a few adjustable parameters that need to be optimized for the best smoothing. We use the smoothed bootstrap (simulation) technique for choosing the optimal smoothing parameters (Silverman 1986, p. 145; Kagan & Jackson 1994, p. 13 691). The method involves subdividing the catalogue and using one subset (the 'learning' catalogue) to predict the other (the 'test' catalogue). As the learning subset we use the catalogue for 1977–1996, and as the test catalogue the data for 1997–1998. We then calculate the likelihood function (Kagan & Jackson 1994, eq. 6) and optimize it by modifying the values of the parameters in eqs (3) and (6).

Likelihood analysis reveals that there is a trade-off between the parameters $\lambda$ and $r_s$ such that for any reasonable value of $\lambda$, there is a corresponding value of $r_s$ with about the same value of maximum likelihood. Thus, we arbitrarily chose $\lambda=1.0$ and let the smoothing distance $r_s$ control the smoothing kernel. For the NW Pacific we found $r_s=15$ km and $\delta=100$ to be near the maximum likelihood values, whereas for the SW Pacific we used $r_s=5$ km and $\delta=25$ (Jackson & Kagan 1999).

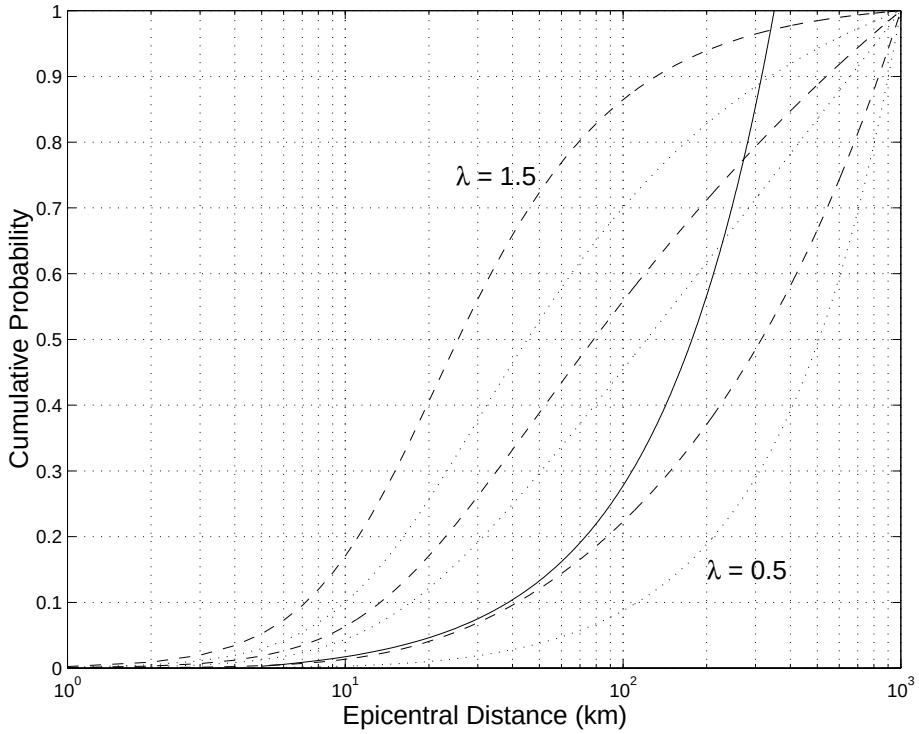


**Figure 1.** Examples of cumulative kernel graphs. Solid line: $1/r$ kernel (Kagan & Jackson 1994), $R_{\max}=350$ km. Dashed and dotted lines from right to left correspond to kernels (eqs 2–5) with $r_s=15$ km and $\lambda=0.5$, 0.75, 1.0, 1.1, 1.25 and 1.5.

### 2.2 Seismic moment distribution

We use the notation $M$ for the scalar seismic moment and $m$ for the moment magnitude,

$$m=\frac{2}{3}\,(\log_{10} M-9.05) \tag{8}$$

(Hanks & Kanamori 1979), where $M$ is measured in N m. The magnitude calculated from eq. (8) is used here only for illustration purposes; all pertinent computations are carried out with the moment $M$-values.

We assume that the occurrence rate $\phi(\mathbf{x}, M)$ of earthquakes at location $\mathbf{x}$ with moment $M$ may be written purely in terms of the marginal rates, i.e.

$$\phi(\mathbf{x}, M)=\phi_{\mathbf{x}}(\mathbf{x})f_{\mathrm{M}}(M)\,, \tag{9}$$

where $\mathbf{x}$ are the horizontal spatial coordinates and $\phi_{\mathbf{x}}(\mathbf{x})$ and $f_{\mathrm{M}}(M)$ are normalized probability densities of events in area and moment, respectively; the equation indicates that the spatial and moment distributions of earthquake clusters are independent. The distribution $f_{\mathrm{M}}(M)$ is defined by eq. (11) below. The methodology for the evaluation of the spatial distribution of earthquakes that corresponds to the long-term seismicity forecast is described in Kagan & Jackson (1994).

The seismic moments are modelled to follow the modified Gutenberg–Richter (MGR) relation with an exponential taper applied to the cumulative number of events with seismic moment larger than $M$ (D. Vere-Jones, personal communication, 1999; Vere-Jones *et al.* 2000):

$$F(M)=(M_{\mathrm{t}}/M)^{\beta}\exp\left[(M_{\mathrm{t}}-M)/M_{\mathrm{c}}\right] \quad \text{for} \quad M_{\mathrm{t}}\le M\le\infty\,, \tag{10}$$

where $M_{\mathrm{t}}$ is a catalogue completeness threshold (cut-off), taken here to be $M_{\mathrm{t}}=10^{17.7}$ N m ($m_{\mathrm{t}}\approx 5.8$), and $M_{\mathrm{c}}$ is the parameter that controls the distribution in the upper ranges of $M$ ('upper corner moment'). The corresponding probability density function is

$$f_{\mathrm{M}}(M)=\left[\frac{\beta}{M}+\frac{1}{M_{\mathrm{c}}}\right](M_{\mathrm{t}}/M)^{\beta}\exp\left[(M_{\mathrm{t}}-M)/M_{\mathrm{c}}\right]. \tag{11}$$

Previously we used the gamma distribution (Kagan 1991a, 1999; Main *et al.* 1999), where the exponential taper is applied to density, not to the cumulative function.

We apply the maximum likelihood method to estimate the parameters $M_{\mathrm{c}}$ and $\beta$ from catalogue data. The logarithm of the likelihood function ($\ell$) for $N$ observations of the seismic moment for the distribution (10, 11) is

$$\ell=N\beta\log(M_{\mathrm{t}})+\frac{1}{M_{\mathrm{c}}}\left(NM_{\mathrm{t}}-\sum_{i=1}^{N}M_i\right)-\beta\sum_{i=1}^{N}\log M_i$$
$$+\sum_{i=1}^{N}\log\left(\frac{\beta}{M_i}+\frac{1}{M_{\mathrm{c}}}\right). \tag{12}$$

A comparison of this expression with eq. (6) in Kagan (1991a) shows an advantage of the MGR relationship (11) over the gamma distribution: for the gamma distribution the normalizing coefficient involves the incomplete gamma function, which complicates the formula and its computation.

To determine the corner moment and $\beta$ we obtain maps of the log likelihood function for two parameters of the gamma and MGR distributions (10, 11). See Kagan (1991a, 1999) for a more complete description. We display these maps in Figs 2(a) and (b). The maximum likelihood estimates of the slope $\beta$ and the corner moment $M_{\mathrm{c}}$ are practically uncorrelated for the MGR distribution, whereas similar estimates for the gamma distribution show some positive correlation. Moreover, the log-

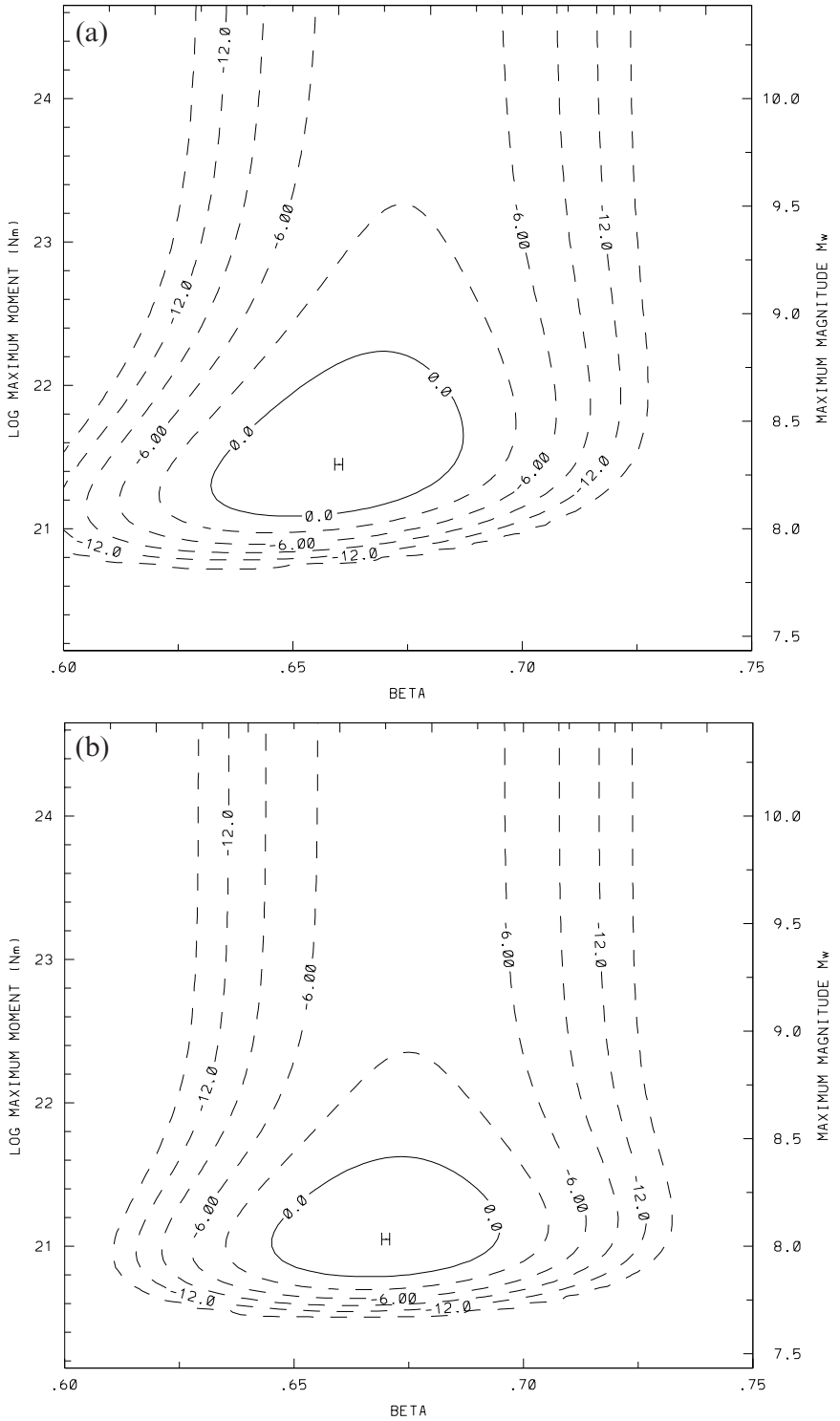


**Figure 2.** Log likelihood maps for distribution of scalar seismic moment of earthquakes. The Harvard catalogue time span is 1982 January 1–1999 July 1, the seismic moment cut-off is $10^{17.4}$ N m ($m_t - 5.6$) and the number of events is 4512. (a) Approximation by the gamma distribution; (b) approximation by the modified Gutenberg–Richter (MGR) distribution.

likelihood function for the MGR distribution is 0.16 higher than that for the gamma law. Although this difference for the log-likelihood is not statistically significant according to the Akaike information criterion (see Utsu 1999, p. 517 and his Table 2), it shows a marginally better fit of the MGR formula (10, 11) to the data.

In Fig. 3 we display the fit of the MGR and gamma distributions to the empirical distribution of shallow earthquake moments in the Harvard catalogue (Dziewonski *et al.* 2000). Both theoretical curves fit the data well. The standard errors of $\beta$ and the corner moment are smaller for the MGR distribution, in part because their estimates are less correlated.

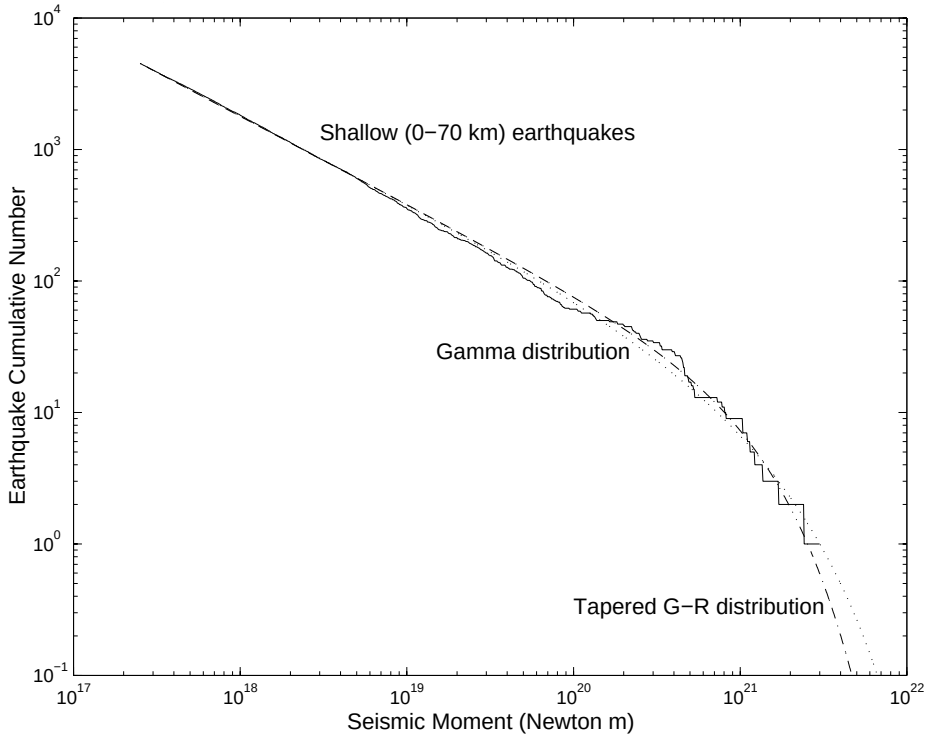


**Figure 3.** Earthquake cumulative number versus log of seismic moment for the global shallow earthquake distribution in the 1982 January 1–1999 July 1 Harvard catalogue. The curves show the numbers of events with moment larger than or equal to $M$, $M_t = 10^{17.4}$ N m ($m_t = 5.6$). We also show the approximation of curves by the Gutenberg–Richter law restricted at large magnitudes by an exponential taper (MGR) as well as by the gamma distribution (Kagan 1991a, 1999), where the exponential taper is applied to the probability density. The slope of the linear part of the modified Gutenberg–Richter curve corresponds to a $\beta$-value (eq. 10) of $0.669 \pm 0.013$ and the corner moment magnitude $m_c = 8.04 \pm 0.14$. For the gamma distribution, respective parameter values are $\beta = 0.660 \pm 0.015$ and $m_\chi = 8.32 \pm 0.19$.

An additional difference between these two seismic moment distributions is that the MGR distribution has a fatter tail. It follows that for the same value of $\beta$ and corner moment, the MGR distribution will predict a larger moment rate than the gamma distribution. Thus, if the corner moments for both distributions are estimated from the earthquake catalogues, and constrained to the same moment rates, then the estimated corner magnitude for the MGR distribution will be about 0.35 units smaller than that for the gamma distribution. Thus, the value of the corner moment $M_c = 10^{21.7}$–$10^{22.5}$ N m (magnitude 8.5–9.0) obtained for the gamma distribution in Kagan (1999) is equivalent to $M_c = 10^{21.1}$–$10^{22.0}$ N m ($m = 8.1$–$8.7$) for the MGR distribution used here.

### 2.3 Long-term seismicity forecast

Let us comment briefly on the long-term forecasting procedure, as it is implemented here, and its relation to the 'real' but unknown long-term earthquake occurrence rate density. Were earthquakes to follow the Poisson process (that is, show no clustering or quasi-periodic recurrence) the algorithm described above would yield a nearly optimal estimate of future seismicity. However, as we mentioned above, the major feature of earthquake occurrence is time–space clustering, both short and long term. If we use in our procedure catalogues that are longer than a typical aftershock sequence, the short-term clustering effects would be averaged out for all but the most productive sequences. Nevertheless, temporal clustering would still influence the forecast because clusters may last for decades or possibly centuries in areas of low seismicity (Kagan & Jackson 1994). We can expect our procedure to work well in the presence of such clustering because the rate of earthquakes in the catalogue, and the expected rate for the next few years, will be affected in about the same way. However, the rate density we estimate would be the present value of a temporally varying process, rather than the steady-state value. As an alternative model, not considered here, we may use geodetic, tectonic and geological information to infer an estimate for the long-term, steady-state earthquake rate (Jackson & Kagan 2000).

## 3 SHORT-TERM SEISMIC HAZARD ESTIMATES

In this study we extend the results of the long-term seismicity forecast to issue short-term forecasts of future earthquake activity. The short-term forecasts are based largely on the phenomenon of short-term earthquake clustering, the best-known manifestations of which are foreshock–main shock–aftershock sequences. Several authors (Kagan & Knopoff 1987; Reasenberg & Jones 1989; Maeda 1996; Console & Murru 1996; Utsu & Ogata 1997; Ogata 1988, 1998; Console 1998) described quantitative models of the clustering and use of foreshocks to predict main-shock probability. Although the general scientific algorithms that allow one to calculate short-term probabilities have been known for some time (see the references above), the application of these ideas to real-time forecasting was not feasible until recently. Several recent developments have made statistical short-term earthquake forecasting possible: the availability of earthquake solutions within a few hours, fast computers that allow the computation of earthquake

probability maps in a few minutes or even seconds, and the World-Wide Web, which allows fast transmission of the results to a wide audience.

### 3.1 Description of the forecasting procedure

Our technique for producing short-term hazard estimates is to establish a statistical model to fit the catalogue of earthquake times, locations and seismic moments, and subsequently to base forecasts on this model. While most of the components of the model have been tested (Kagan & Knopoff 1987; Kagan 1991b; Jackson & Kagan 1999), some require further exploration and may be modified as our research progresses.

The assumptions we make in constructing our initial model may be summarized as follows [see Kagan & Knopoff (1987) and Kagan (1991b) for more details; a similar model was proposed by Ogata (1998)]. Seismicity is approximated by a Poisson cluster process, in which clusters or sequences of earthquakes are statistically independent although individual earthquakes in the cluster are dependent events. The clusters are assumed to form a Poisson point process with a constant rate. The major assumption regarding the interrelationships between events within a cluster is that the propagation of an earthquake rupture is closely approximated by a stochastic space–time critical branching process. Under this assumption there is a sole trigger for any given dependent event. As shown below (Section 3.1.2), the space–time distribution of inter-related earthquake sources within a sequence is controlled by simple relations justified by analysing the available statistical data on seismicity.

#### *3.1.1 Earthquake clusters: independent events*

Usually the first event in a sequence is the largest one and it is called a main shock. Other dependent events are simply called aftershocks, although some of them are really aftershocks of aftershocks. If the first event in a sequence is smaller than subsequent shocks, it is called a foreshock. Retrospectively, it is relatively easy to subdivide an earthquake catalogue into foreshocks, main shocks and aftershocks. However, in real-time forecasting it is uncertain whether the most recent event registered by a network is a foreshock or a main shock. Although it is likely that subsequent dependent events will be smaller and called aftershocks, there is a significant chance that such earthquakes may be bigger (Kagan 1991b; Michael & Jones 1998).

#### *3.1.2 Dependent events*

Similar to the estimation of the long-term seismic hazard (Section 2), we assume that the distribution of dependent events *within* a cluster may be broken down into a product of its marginal distributions. That is, the conditional rate density of the $j$th shock dependent on the $i$th shock ($j>i$) with seismic moment $M_i$ is modelled as

$$\psi(\Delta t,\, \rho,\, M_j|M_i)=\psi_{\Delta t}(\Delta t)\psi_\rho(\rho)\psi_M(M_i)\phi_M(M_j)\,, \tag{13}$$

where $\Delta t=t_j-t_i$ and $\rho$ is the horizontal distance between the $i$th and $j$th centroids ($\rho=|\mathbf{x}_j-\mathbf{x}_i|$). $\psi_{\Delta t}$, $\psi_\rho$ and $\psi_\mu$ are the marginal temporal, spatial and moment densities and are detailed below. The total time-dependent rate density is a sum of the effects from all previous earthquakes,

$$\Psi(t_j,\, \mathbf{x}_j,\, M_j)=\sum_{i<j}\psi(\Delta t,\, \rho,\, M_j|M_i)\,. \tag{14}$$

The function $\psi$ decays rapidly with time and distance (see below). Thus, only neighbouring events substantially contribute to the sum (14), although the range of strong earthquakes is much longer in time and space than that of weak events.

The first three densities $\psi(x)$ in the right part of eq. (13) depend on $M_i$. We take a power law relation for the probability density of time intervals between earthquakes within a cluster,

$$\psi_{\Delta t}(\Delta t)=\theta t_M^{\theta}\Delta t^{-1-\theta}\,,\quad \Delta t\geq t_M\,, \tag{15}$$

which is similar to Omori's law. The parameter $\theta$ is an 'earthquake memory' factor, and $t_M$ is the coda duration time of an earthquake with seismic moment $M_i$. We assume for the coda duration

$$t_M=t_r M_i^{1/3}\,, \tag{16}$$

where $t_r$ is the coda duration time, taken here as $t_r=0.03125$ days, of an earthquake with the reference seismic moment $M_r=10^{18}$ N m, corresponding to $m_r\approx 6.0$ (Kagan 1991b). The probability of the next dependent shock occurring in the time interval ($t_1$, $t_2$), for $0<t_1<t_2$, given an event of a cluster occurring at time 0, can then be calculated simply as

$$\mathrm{Prob}(t_1<t<t_2)=t_M^{\theta}(t_1^{-\theta}-t_2^{-\theta})\,. \tag{17}$$

The non-normalized function $\psi_M(M_i)$ that corresponds to the number of dependent shocks generated on average by an earthquake with seismic moment $M_i$ is assumed to obey (Kagan 1991b)

$$\psi_M(M_i)=\mu(M_i/M_t)^{\zeta}\,. \tag{18}$$

We approximate by a Rayleigh distribution the probability density of the horizontal distance between two earthquake centroids in a cluster:

$$\psi_\rho(\rho)=\frac{\rho}{\sigma_\rho^2}\exp[-\rho^2/(2\sigma_\rho^2)]\,, \tag{19}$$

where $\sigma_\rho$ is the spatial standard deviation. The standard deviation is assumed to depend on the standard errors of centroid determination and on the seismic moment $M_i$ of the main event according to the relation (Kagan 1991b)

$$\sigma_\rho^2=\varepsilon_\rho^2+s_r^2(M_i/M_r)^{2/3}\,. \tag{20}$$

Here $\varepsilon_\rho$ is the standard error in centroid determination and $s_r$ is a characteristic size of a focal zone of an earthquake with some reference seismic moment $M_r$ (see eq. 16).

The density (19) is isotropic. There are several ways to improve the forecast of azimuth distribution of dependent events: (1) we can use the pattern of past seismicity to infer the scatter of future events, (2) we can use earthquake focal mechanisms to estimate the direction of the focal plane and project epicentres of dependent events along this plane (Kagan & Jackson 1994), and (3) we can use the aftershock pattern to infer the development of earthquake rupture (Ogata 1998). Only the first of these methods can be used practically in real-time forecasting. The focal mechanisms of temporary solutions, obtained in a few hours after an earthquake, are usually determined by the inversion of only a few seismograms, thus their accuracy is low

(Frohlich & Davis 1999). The aftershock pattern is known only after some time delay, and the preliminary epicentres also have relatively low accuracy. For these reasons we do not presently use focal mechanism information, nor do we try to infer the specifics of the aftershock distribution as suggested by Ogata (1998).

The spatial density function (19) needs to be normalized to take into account the global non-homogeneous seismicity distribution. We accomplish this by using instead of $\psi_\rho(\rho)$ the function

$$\psi_{\mathbf{u}}(\rho) = C_\phi^{-1}\psi_\rho(|\mathbf{u}-\mathbf{x}|)\phi_{\mathbf{x}}(\mathbf{x}), \tag{21}$$

where $\mathbf{u}$ denotes earthquake centroid coordinates, $|\mathbf{u}-\mathbf{x}|$ is the horizontal distance between points $\mathbf{u}$ and $\mathbf{x}$, $\phi(\mathbf{x})$ is the spatial density of earthquake clusters (see eq. 9) and $C_\phi$ is a spatial normalizing coefficient,

$$C_\phi(\mathbf{u}) = \int_{\mathbf{X}} \psi_\rho(|\mathbf{u}-\mathbf{x}|)\phi_{\mathbf{x}}(\mathbf{x})\, d\mathbf{x}, \tag{22}$$

where $\mathbf{X}$ is the total forecast area and $d\mathbf{x}$ is an area differential element.

The proposed short-term forecast is valid only for the 'next moment' after the end of a catalogue. To extend the forecast further in the time, we need to take into account the possibility that other earthquakes could occur in the time interval between the end of a catalogue and the forecast time. This can be carried out by the Monte-Carlo simulation procedure (Kagan 1973b): we simulate the first generation of dependent events, which in turn are fed into the algorithm to obtain the second generation, and so on. If the time horizon of the forecast is not long, only a few generations need to be computed for such simulations. For longer time intervals we may simply use our long-term forecast.

### 3.2 Examples of forecasts

Estimates of the parameters in the model (9–22) may yield important seismological information. The values of parameter estimates listed in Section 3.1.2 and obtained for the Harvard catalogue (Dziewonski *et al.* 2000) by the maximum likelihood optimization procedure (Kagan 1991b) are $\mu = 0.09$, $\theta = 0.21$, $s_r = 3.0$ km, $\varepsilon_\rho = 25.0$ km and $\zeta = 0.68$. For independent shocks (Section 3.1.1) the parameter values are (Kagan 1999) $\beta = 0.63$ and $M_c = 10^{22}$ N m ($m_c = 8.7$). With these values of parameters one can calculate from (18) that, for example, a $M = 10^{21}$ N m ($m = 8.0$) earthquake should have on average about 10 dependent events (aftershocks) of $M \geq 10^{18}$ N m ($m \geq 6.0$). However, according to the branching model, each of these earthquakes may generate its own offspring events, thus the total number of earthquakes in a sequence may be significantly larger.

Although the focal mechanisms of short-term future earthquakes can be predicted using the method described by Kagan & Jackson (1994), in this initial stage of our forecasting effort, we use the long-term estimate of the focal mechanism (see Table 1). The preliminary results suggest that within an earthquake sequence the focal mechanisms are very tightly clustered, with the Cauchy rotational distribution describing the scatter of the 3-D angle of rotation (Kagan 1992).

Table 1 displays an example of a forecast. We show a header and a few lines from the actual table for an area north of the Philippines. This area is selected because on 1999 February 6 a shallow earthquake with magnitude $m = 6$ occurred at 19.36°N, 120.97°E. While the earthquake was of moderate size, its influence on short-term probability is still felt. Near the centroid, the daily probability 5 days after the event is nine times higher than the long-term average. Although we display hazard estimates with three significant digits, the uncertainty in their determination may be of the order of tens of per cent. We supply more digits to facilitate trend calculations and comparison with other models. Estimation of uncertainties is clearly an important subject for future work.

In Fig. 4 and Table 2 we display another example, the 1992 sequence of earthquakes near the NE coast of Honshu island (Japan). The sequence starts with an earthquake on July 12, which slightly increases the daily seismicity forecast at the location of the sequence. Four days later the event is followed by another moderate sized earthquake, which, in retrospect, can be considered a foreshock of the $m7$ event occurring 2.4 days later. The later earthquake is in turn followed by a few aftershocks. Note, however, that during the sequence it appears uncertain whether or not a subsequent event will exceed the $m7$ earthquake in size. As shown in Fig. 4(f), the increase in seismicity attributed to the $m7$ earthquake is still high half a year later, while the increase in seismicity due to the smaller 1992 July 12 earthquake has essentially dissipated.

We calculate daily short-term hazard estimates for the NW and SW Pacific regions. The forecasts are computed from the preliminary Harvard catalogue. We use e-mail messages

**Table 1.** Example of long- and short-term forecasts, 1999 February 11, north of the Philippines.

| Latitude | Longitude | Long-term forecast | | | | | | Short-term forecast | |
|---|---|---|---|---|---|---|---|---|---|
| | | Probability $m \geq 5.8$ eq/day*km$^2$ | Focal mechanism *T*-axis Pl | *T*-axis Az | *P*-axis Pl | *P*-axis Az | Rotation angle degree | Probability $m \geq 5.8$ eq/day*km$^2$ time-dependent | Probability ratio time-dependent/ independent |
| 119.5 | 19.5 | 3.18E-09 | 31 | 208 | 10 | 304 | 64.8 | 1.79E-14 | 5.62E-06 |
| 120.0 | 19.5 | 5.23E-09 | 17 | 213 | 32 | 314 | 68.8 | 1.41E-10 | 2.71E-02 |
| 120.5 | 19.5 | 4.28E-08 | 7 | 93 | 75 | 335 | 21.4 | 2.12E-07 | 5.0 |
| 121.0 | 19.5 | 3.02E-08 | 69 | 135 | 21 | 302 | 28.2 | 2.84E-07 | 9.4 |
| 121.5 | 19.5 | 1.82E-08 | 77 | 106 | 13 | 296 | 40.9 | 6.14E-08 | 3.4 |
| 122.0 | 19.5 | 7.81E-09 | 60 | 32 | 3 | 297 | 48.4 | 1.13E-10 | 1.45E-02 |
| 122.5 | 19.5 | 4.15E-09 | 81 | 228 | 4 | 113 | 51.8 | 1.00E-12 | 2.41E-04 |
| 123.0 | 19.5 | 3.01E-09 | 78 | 251 | 9 | 110 | 50.3 | 7.70E-16 | 2.56E-07 |
| 123.5 | 19.5 | 2.43E-09 | 76 | 273 | 13 | 107 | 49.5 | 1.08E-20 | 4.43E-12 |

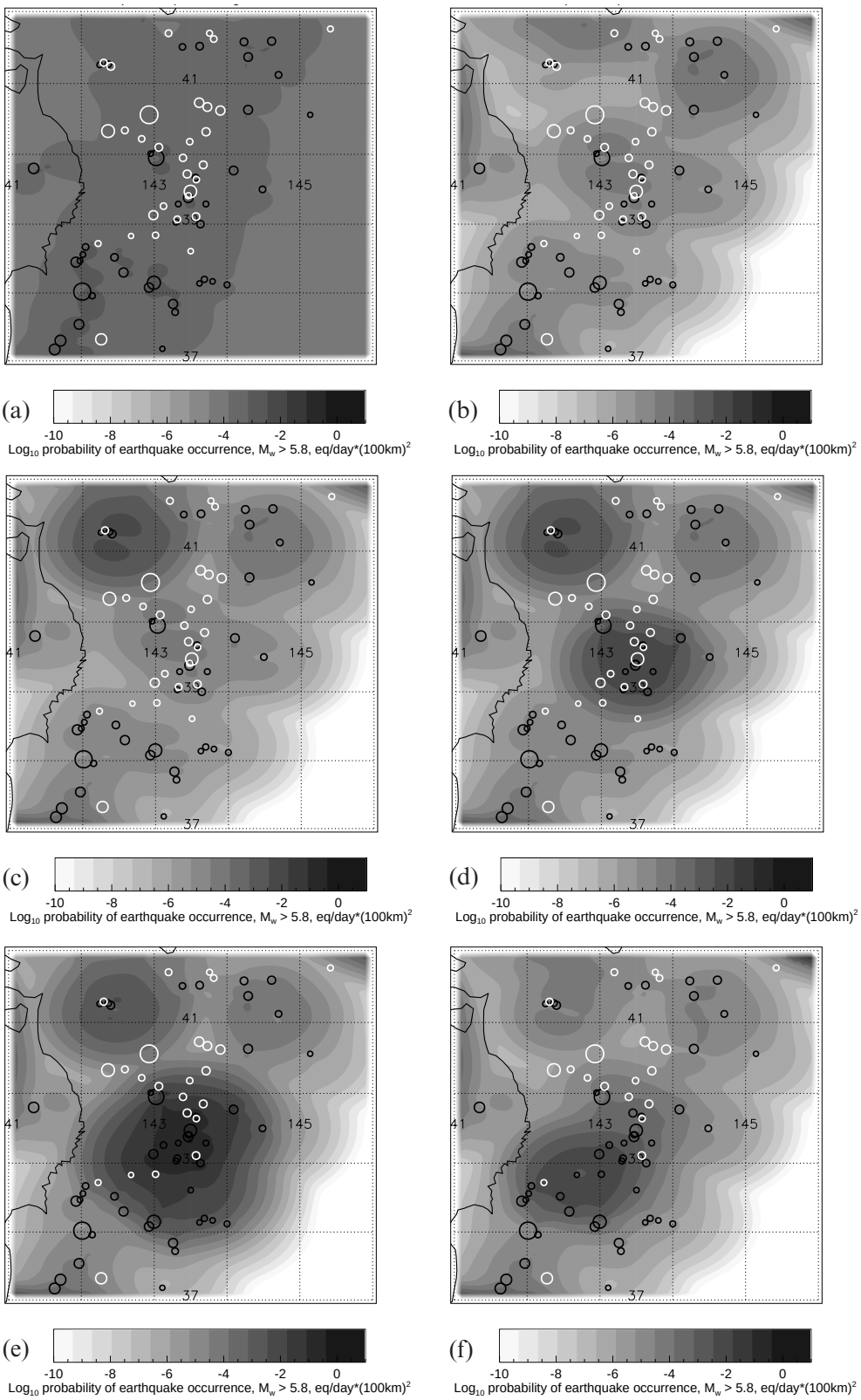


**Figure 4.** (a) Long-term seismicity forecast northwest of Honshu Island, Japan; latitude limits 37.0–42.0°N, longitude limits 141.0–146.0°E. Earthquakes before 1992 July 15 are shown in black, earthquakes after that are shown in white. The sizes of circles are proportional to earthquake magnitude. Greyscale tones show the long-term probability of earthquake occurrence calculated using the Harvard 1977–1992 July 15 catalogue. (b) Greyscale tones show the short-term probability of earthquake occurrence calculated using the Harvard 1977–1992 July 11 catalogue; earthquakes before 1992 July 11 are shown in black, earthquakes after that are shown in white. Time limits for the following maps of the short-term probability are as follows: (c) 1992 July 15; (d) 1992 July 16; (e) 1992 July 18; (f) 1993 January 1. Fig. 4 may be viewed in colour in the online version of the journal (www.blackwell-synergy.com).

**Table 2.** Earthquakes in the northeast Honshu (Japan) region in 1992.

| No | Date | | Time | | Latitude | Longitude | Depth | Moment |
|---|---|---|---|---|---|---|---|---|
| | Month | Day | Hour | Min | | | (km) | ($10^{18}$ Nm) |
| 41 | May | 7 | 6 | 23 | 41.12 N | 144.73 E | 15 | 1.15 |
| 42 | Jul | 12 | 11 | 8 | 41.24 N | 142.43 E | 60 | 2.19 |
| 43 | Jul | 16 | 0 | 0 | 39.41 N | 143.49 E | 19 | 0.85 |
| 44 | Jul | 18 | 8 | 37 | 39.47 N | 143.54 E | 15 | 27.54 |
| 45 | Jul | 18 | 10 | 20 | 39.13 N | 143.02 E | 15 | 4.47 |
| 46 | Jul | 18 | 12 | 2 | 38.61 N | 143.52 E | 15 | 0.62 |
| 47 | Jul | 18 | 13 | 56 | 39.26 N | 143.15 E | 29 | 1.07 |
| 48 | Jul | 18 | 21 | 19 | 39.07 N | 143.34 E | 20 | 1.17 |
| 49 | Jul | 25 | 2 | 53 | 38.84 N | 143.04 E | 18 | 0.93 |
| 50 | Jul | 29 | 4 | 30 | 39.72 N | 143.48 E | 15 | 2.34 |
| 51 | Dec | 31 | 7 | 26 | 38.83 N | 142.70 E | 16 | 0.51 |

sent by the Harvard team to update our catalogue for all earthquakes $M \geq 10^{17.7}$ N m ($m \geq 5.8$); the time delay between earthquake occurrence and update is of the order of a few hours for automatic solutions (Kawakatsu 1995) or up to 1 day for human analyst solutions. The updated catalogue is then used to estimate both long- and short-term daily probabilities of future earthquake occurrence. The short-term values are for 1 day periods from the current midnight to the subsequent midnight, Los Angeles time. These forecasts are stored in a table and displayed in two figures, which are accessible via the World-Wide Web at http://scec.ess.ucla.edu/ykagan.html (see 'FORECASTS FOR 1999–2000: TABLES AND FIGURES'). The tables (in ASCII form) list daily long-term and short-term probabilities of earthquakes on a $0.5 \times 0.5^{\circ}$ grid. The plots display long-term and short-term probabilities as colour PostScript maps of the regions.

The long-term probabilities (1) account for the rate of both independent and dependent events, whereas the short-term rates (14) are calculated only for dependent events. Thus, when displaying the sum of these rates, we should subtract from the long-term rate the averaged part of the seismic activity due to dependent events. For moment cut-off $M \geq 10^{17.7}$ N m, this part was estimated to be 0.2 (Kagan 1991b). Thus, we adjust the long-term rate in the plots of combined probability by multiplying it by 0.8:

$$\Lambda(t, \mathbf{x}, M, \omega) = 0.8\Xi(\mathbf{x}, M, \omega) + \Psi(t, \mathbf{x}, M)h(\omega), \qquad (23)$$

where $\Lambda$ is the combined time-dependent long- and short-term rate of earthquake occurrence. Examples of the combined rate for the hybrid model (23) at one point near the centre of the 1992 sequence (see Fig. 4) are shown in Figs 6 and 7 of Jackson & Kagan (1999).

The estimated value of 0.09 for $\mu$ in our preliminary analysis described below indicates that on average about 9 per cent of earthquakes would be followed by a stronger dependent event within a relatively short time (a few days or weeks). This fact explains why our statistical short-term forecast does not meet the popular definition of 'prediction' (Kagan 1997): only a few alarms would be followed closely by sufficiently large earthquakes. However, investigations by Kagan & Knopoff (1987), Reasenberg & Jones (1989), Michael & Jones (1998) and Reasenberg (1999) show that 20–40 per cent of earthquakes are preceded by events that are smaller and are identified later as statistical forerunners of future earthquakes. These sequences include foreshocks followed by main shocks, as well as strong aftershocks preceded by an increase in seismic activity of smaller events. This means that some inexpensive mitigation and scientific measures (Molchan & Kagan 1992) may be justified. Examples of such measures include putting emergency services on a higher-alert status, stepping up InSAR (Synthetic Aperture Radar Interferometry) and other geophysical measurements in the dangerous regions, and deploying temporary seismic stations.

## 4 TESTING FORECASTS

There are several methods for verifying a long-term forecast (Kagan & Jackson 1995). For example, we can (1) compare the observed and predicted numbers of events, and (2) test the spatial distribution of earthquakes against the forecast map. In this paper we discuss and carry out preliminary tests of forecast performance only for the long-term estimate of seismic hazard. At present, the Harvard catalogue for the second half of 1999 is only in preliminary form. Therefore, our test results may change (hopefully only slightly) when the catalogue becomes available in the final form in a few months.

The short-term forecast can be tested, in principle, by using similar likelihood methods (Kagan & Knopoff 1987; Kagan 1991b). However, when testing a short-term forecast we are faced with the problem of selecting the appropriate null hypothesis (Stark 1997). If we use a time-invariant null hypothesis (Kagan & Knopoff 1987), almost any algorithm that models aftershock sequences would enjoy a huge statistical advantage. Therefore, a high level of likelihood ratio would contain no relevant information. Moreover, our forecasts are now issued once per day, whereas a substantial part of the probability increase is concentrated in the first hours after a strong earthquake. Thus, we need, first, to develop the procedure so it can run continuously, and, second, to find a credible null hypothesis against which our short-term scheme can compete fairly. We have not yet found such a null hypothesis.

### 4.1 Distribution of earthquake numbers

To test whether earthquake numbers are within prescribed limits, we must know the number distribution. Fig. 5 displays the numbers of earthquakes ($m \geq 5.8$) in the NW Pacific for 22 yr of Harvard catalogue monitoring (Dziewonski *et al.* 2000). Even a cursory inspection of the plot suggests that the number distribution deviates from the average by more than

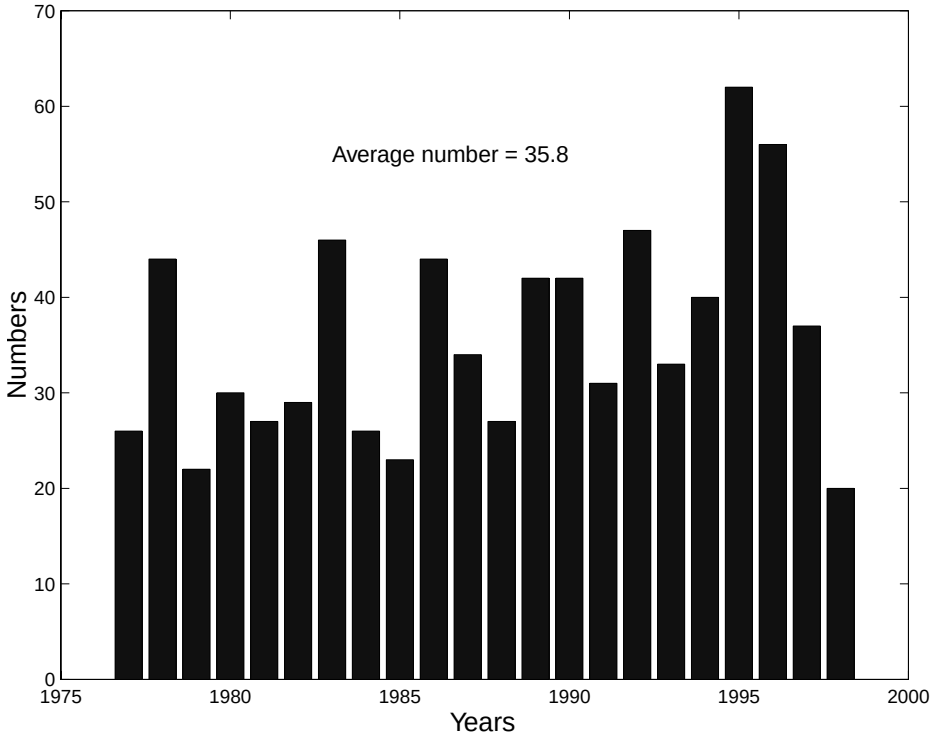


**Figure 5.** Yearly numbers of earthquakes $m \geq 5.8$ in the NW Pacific, 1977–1998.

the Poisson law allows. As Vere-Jones (1970) and Jackson & Kagan (1999) have shown, the statistical distribution of event numbers can be better approximated by the negative binomial law (Evans *et al.* 1993; Kotz & Johnson 1985, pp. 169–177).

Although declustered catalogues are closer to the Poisson process, the removal of aftershocks cannot be accomplished unambiguously. Therefore, we leave aftershocks in the catalogue and use the negative binomial law to approximate the earthquake numbers. The negative binomial function has two parameters, $\tau$ and $\upsilon$, that are adjusted to fit an observed distribution of earthquake numbers as described below. For positive $\tau$ the probability of $k$ events is

$$p(k) = \frac{\Gamma(\tau + k)}{\Gamma(\tau) k!} \upsilon^{\tau} (1-\upsilon)^{k}, \tag{24}$$

where $k = 0, 1, 2, \ldots$, $\Gamma$ is the gamma function and $0 \leq \upsilon \leq 1$. The parameter $\upsilon$ characterizes the degree of clustering or non-Poisson character of earthquake occurrence; if $\upsilon \to 1$ seismicity becomes close to the Poisson process.

If $\tau$ is a positive integer, eq. (24) is simplified and the probability of $k$ events is

$$p(k) = \binom{\tau + k - 1}{k} \upsilon^{\tau} (1-\upsilon)^{k}. \tag{25}$$

The expected number of earthquakes is

$$E(k) = \tau \frac{1-\upsilon}{\upsilon}; \tag{26}$$

its variance is

$$D(k) = \tau \frac{1-\upsilon}{\upsilon^2}. \tag{27}$$

We can see that the negative binomial distribution generally has a larger standard deviation than the Poisson law, for which the average is equal to the variance. We estimated $\upsilon$ and $\tau$ for various catalogues from the above equations, replacing $E(k)$ by the sample mean, and $D(k)$ by the sample variance for each catalogue (Kotz & Johnson 1985).

We applied various time and moment thresholds to calculate the yearly numbers of earthquakes in three catalogues: the Harvard catalogue, the CalTech/USGS catalogue (Hutton & Jones 1993 and references therein), and the global Pacheco & Sykes (1992) catalogue. In Fig. 6, we display the $\upsilon$-values obtained. From the diagram it is apparent that the estimates depend on the time span of a catalogue: shorter catalogues yield smaller values of $\upsilon$. For catalogues with a large earthquake threshold, the earthquake number distribution is described by a Poisson law.

Kagan (1973a,b) proposed a branching model of earthquake occurrence that predicted that, for large time–space intervals, earthquake numbers are distributed according to the negative binomial distribution. In this model, for the truncated Pareto distribution of seismic energy or seismic moment (equivalent to a Gutenberg–Richter relation with the density truncated at $M_{\max}$), the parameter $\upsilon$ should depend on $M_{\max}$,

$$\upsilon = \left( \frac{M}{M_{\max}} \right)^{\beta}, \tag{28}$$

where $\beta$ is the slope of the moment–frequency relation (10).

We approximate the dependence of $\upsilon$ on the moment threshold by the expression

$$\upsilon = \left( \frac{M}{M + M_c'} \right)^{\beta}, \tag{29}$$

where $M_c'$ is a corner moment, not necessarily equal to $M_c$ in the moment–frequency relation (see eq. 10). Expression (29) differs from eq. (28), since instead of the 'hard' limit for the

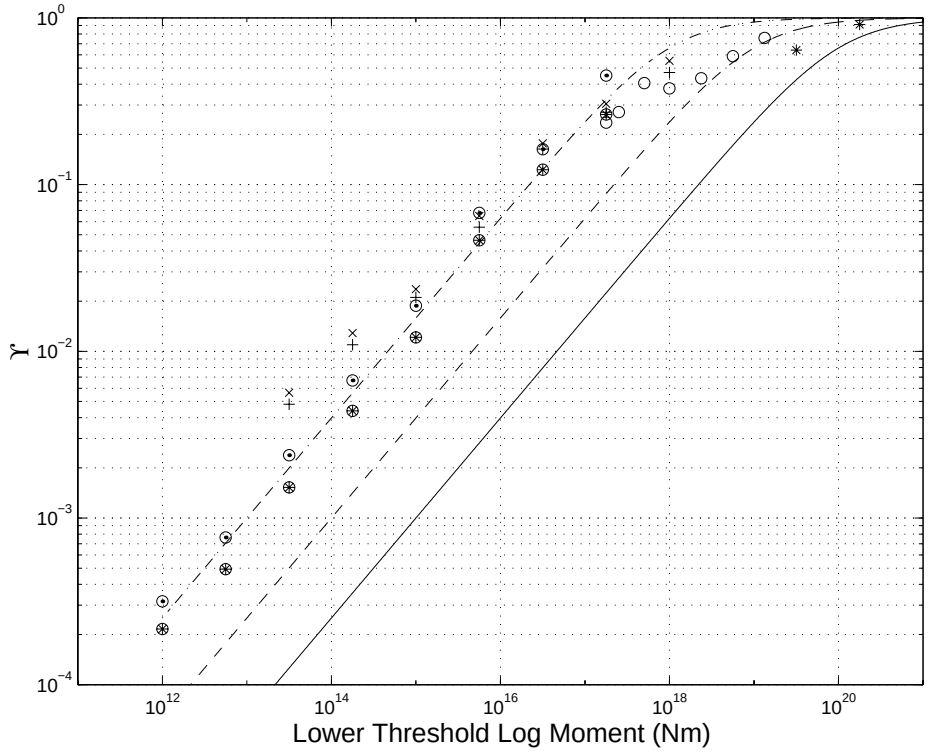


**Figure 6.** Parameter $\upsilon$ of the negative binomial distribution for several catalogues. Asterisks: Pacheco–Sykes, time span 1905–1989; circles: Harvard, time span 1977–1998, $M_t \geq 10^{17.7}$ N m; time span 1982–1998, $M_t \geq 10^{17.4}$ N m; time span 1987–1998, $M_t \geq 10^{17.25}$ N m. CalTech/USGS catalogue, 32.5°–36.5°N, 114.0°–122.0°W—multiplication signs: time span 1932–1998, local magnitude $m_L \geq 3.0$; plus signs: time span 1944–1998, $m_L \geq 3.0$; circles with dots: time span 1979–1998, $m_L \geq 2.0$; circles with asterisks: time span 1989–1998, $m_L \geq 2.0$. Three theoretical approximations are shown (eq. 29) for $\beta = 0.6$ and $M'_c = 10^{20}$ N m (solid line), $M'_c = 10^{19}$ N m (dashed line), $M'_c = 10^{18}$ N m (dash-dotted line).

Gutenberg–Richter relation in eq. (28), we are using a 'soft', exponential taper in (10, 11). The lines in Fig. 6 show eq. (29) with $M'_c = 10^{18}$, $10^{19}$, $10^{20}$ N m and $\beta = 0.6$. We conjecture that for sufficiently long and complete earthquake catalogues, the values of $M'_c$ would approach the value of $M_c = 10^{21}$ N m (Fig. 3).

It is clear from Fig. 6 that estimates for the parameters of the negative binomial distribution ($\tau$ and $\upsilon$) are unstable for small regions, thus we use an $\upsilon$-value based on the approximation (29): for $M = 10^{17.7}$ N m ($m \approx 5.8$), we take $\upsilon = 0.3$. The value of $\tau$ is calculated using the average yearly rate of events (26), i.e.

$$\tau = \frac{3}{7} E(k) . \tag{30}$$

From 1999 January 1 to 1999 December 31 there were 32 events with $m \geq 5.8$ registered in the NW Pacific area and 55 events in the SW Pacific area. Comparing these numbers with the annual rates of earthquakes averaged over 1977–1998 (Jackson & Kagan 1999), which are 35.8 and 57.7 for the NW and SW Pacific respectively, we expect that these numbers would pass the test. Indeed, in Fig. 7 we show two distributions that are used to approximate the earthquake numbers: the Poisson and the negative binomial. It is evident that the 1999 earthquake numbers are compatible with theoretical curves in both regions. However, based on the analysis of the number distributions in the Harvard catalogue in previous years (Jackson & Kagan 1999, their Fig. 3) and other results discussed earlier in this section, we expect the negative binomial distribution to be more appropriate, especially if large aftershock sequences occur in a region.

### 4.2 Testing spatial distribution of earthquake numbers

In Figs 8(a) and (b) we display long-term forecast maps computed in early 1999 for both Pacific regions. From visual inspection, the model predicts the spatial distribution of seismic activity in 1999 reasonably well. We have tested this forecast by a Monte-Carlo simulation (Kagan & Jackson 1994). The test involves a comparison of the forward prediction issued before the test period with a retrospective prediction that is optimized after 1999, when the earthquakes that occurred in 1999 are known. If these two forecasts differ within the 95 per cent confidence limit estimated by a simulation procedure, we consider the forward prediction as successful. In effect, instead of competing against a null hypothesis, which cannot be effectively defined for the spatial distribution of seismicity, we test our results against the 'perfect', ideal model, specified on the basis of a retroactive adjustment of the model parameters.

In Fig. 9 we display simulation distributions for prediction and earthquake data as shown in Fig. 8. We simulate earthquake locations with $r_s = 15.0$ km for the NW Pacific and $r_s = 5.0$ km for the SW Pacific (Jackson & Kagan 1999, their Table 1), each time calculating the likelihood function and comparing the function value with that obtained for the real catalogue in 1999. Whereas the choice of $r_s = 15$ km for the NW Pacific was close to optimal for the prediction of 1999 earthquakes, it is clear that for the SW Pacific $r_s \approx 2.5$ km is a more appropriate value. Thus, a forecast of all 1999 earthquakes, using the model with $r_s = 5.0$ km, would have failed at the 95 per cent confidence level. The reason is apparent from Fig. 8. Similar to the test for the NW Pacific described by

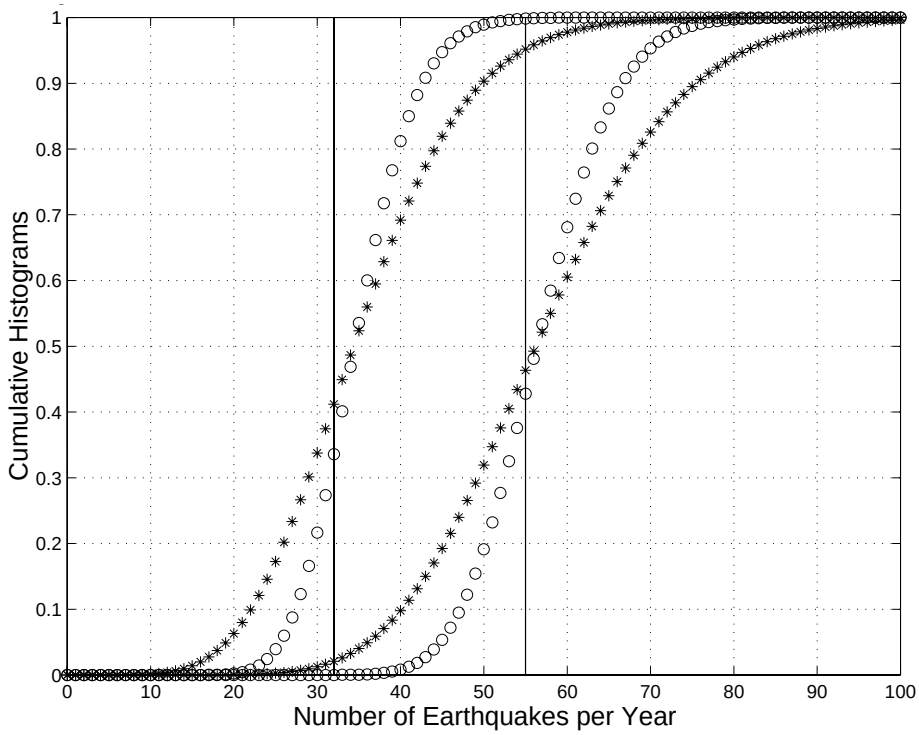


**Figure 7.** Testing the forecast for earthquake numbers. The left cluster of curves is for the NW Pacific; the right cluster is for the SW Pacific. Distribution of earthquake numbers—solid line: earthquakes registered during 1999; asterisks: approximation by the negative binomial distribution; circles: approximation by the Poisson distribution.

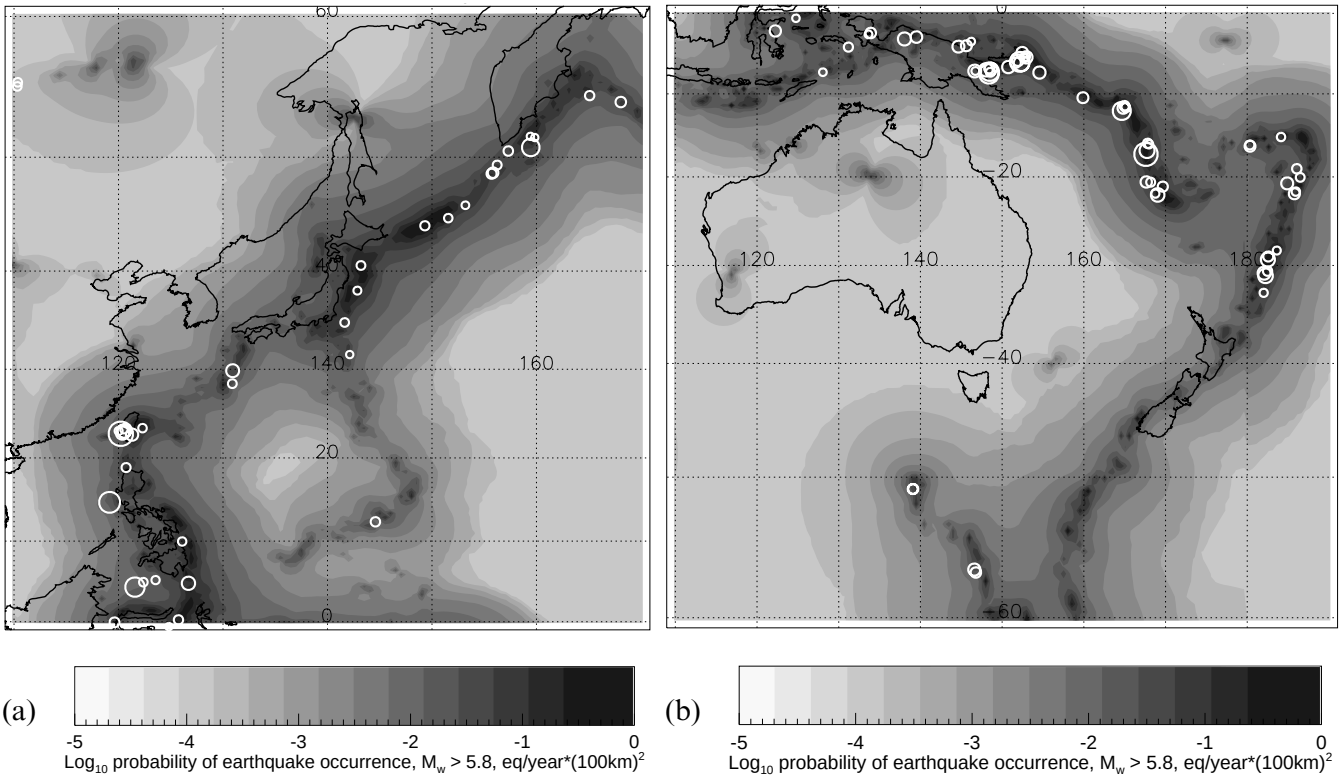


**Figure 8.** Greyscale tones show the probability of earthquake occurrence calculated using the Harvard 1977–1998 catalogue; earthquakes in 1999 are shown in white. (a) NW Pacific long-term seismicity forecast: latitude limits 0.25°S–60.25°N, longitude limits 109.75°E–170.25°E; 32 events from 1999 January 1 to 1999 December 31. (b) SW Pacific long-term seismicity forecast: latitude limits 0.25°N–60.25°S, longitude limits 109.75°E–169.75°W; 55 events from 1999 January 1 to 1999 December 31. Fig. 8 may be viewed in colour in the online version of the journal (www.blackwell-synergy.com).

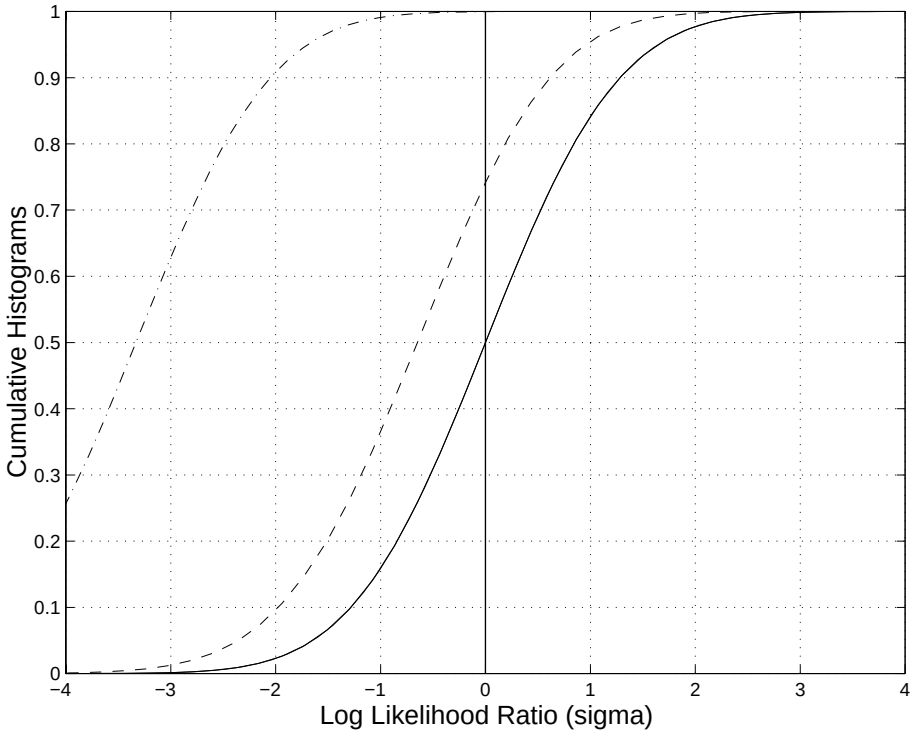


**Figure 9.** Testing the forecast in Fig. 8. We use the 1977–1998 earthquakes as a control set. The solid line is the best Gaussian curve, having the same standard deviation as simulations. The dashed curve corresponds to simulation distributions for the NW Pacific and the dash-dotted curve to the SW Pacific. Curves on the right of the Gaussian curve correspond to simulations worse than a real earthquake distribution; curves on the left correspond to simulations better than a real earthquake distribution.

Kagan (1997), earthquakes since 1998 have been much more strongly clustered than before. Even though these aftershocks occurred in already dangerous areas in Fig. 8 due to pre-1999 earthquakes, the model with $r_s = 5.0$ km fails to account for their strong clustering. As Kagan & Jackson (1994, p. 13 961) explain, the forecasting of aftershocks requires the selection of a smaller smoothing distance. There are two lessons to be learned here. First, like any forecast model, the model of Kagan & Jackson (1994) needs a quantitative procedure to account for aftershocks. Second, a visual inspection of a map such as Fig. 8 is not adequate: a quantitative hypothesis test is required.

In Fig. 10 we show the size distribution of 1999 earthquakes in both Pacific regions and their approximation by the MGR relation (10). Despite large scatter due to the relatively small number of events, it is obvious that in both cases earthquakes follow a similar distribution, and their distribution does not differ significantly from the theoretical curve.

## 5 DISCUSSION

Compared to other techniques for the estimation of time-independent and time-variable seismic hazard (Reasenberg & Jones 1989; NRC 1991; Frankel 1995; Frankel *et al.* 1996; Console & Murru 1996; Maeda 1996; Cao *et al.* 1996; Kafka & Walcott 1998; Console 1998; Michael & Jones 1998), the forecast methods discussed above have several advantages. Both of our models exist as computer algorithms, thus their implementation does not require additional human intervention. Moreover, they use only the information available at present, and no retrospective evaluation of seismic data is assumed. Therefore, these techniques can be used for an automatic computer-assisted evaluation of seismic hazard.

The long-term hazard calculations are based on the smoothed bootstrap optimization of the forecast procedure (Section 2). The parameters of the model are evaluated on the basis of success in the forecasting of seismic activity of a control data set (see also the discussion below eq. 7). Thus, our results are optimal in a certain well-defined sense, that is, the values of parameters in eq. (7) are evaluated as the best to predict earthquake distribution in the previous year for a particular region. In addition to the spatial distribution of future seismicity, we also report possible focal mechanisms of earthquakes and uncertainties of this estimate as the 'rotation angle' in Table 1 (see also Kagan & Jackson 1994).

For short-term forecasts we use an explicit model of earthquake clustering. We apply the maximum likelihood procedure to infer the optimal values of clustering model parameters. The method does not depend on the *post factum* identification of event type (foreshock, main shock or aftershock) and can thus function in a truly automatic fashion. It can be incorporated into real-time seismic networks to provide an almost instantaneous seismic hazard estimate.

The investigations described above may improve our understanding of the tectonic motions and stress accumulation causing earthquakes; the studies would result in a timely and effective hazard assessment. Because our earthquake probability assessments are rapid and up-to-date, they can contribute to improved decision making and mitigation of earthquake risks.

Both forecasts are well suited for statistical testing, which can be carried out in real time. Kagan & Knopoff (1987), Kagan & Jackson (1994, 1995) and Jackson (1996) demonstrated forecast verification techniques. Moreover, these earthquake probability calculations should provide a minimum threshold

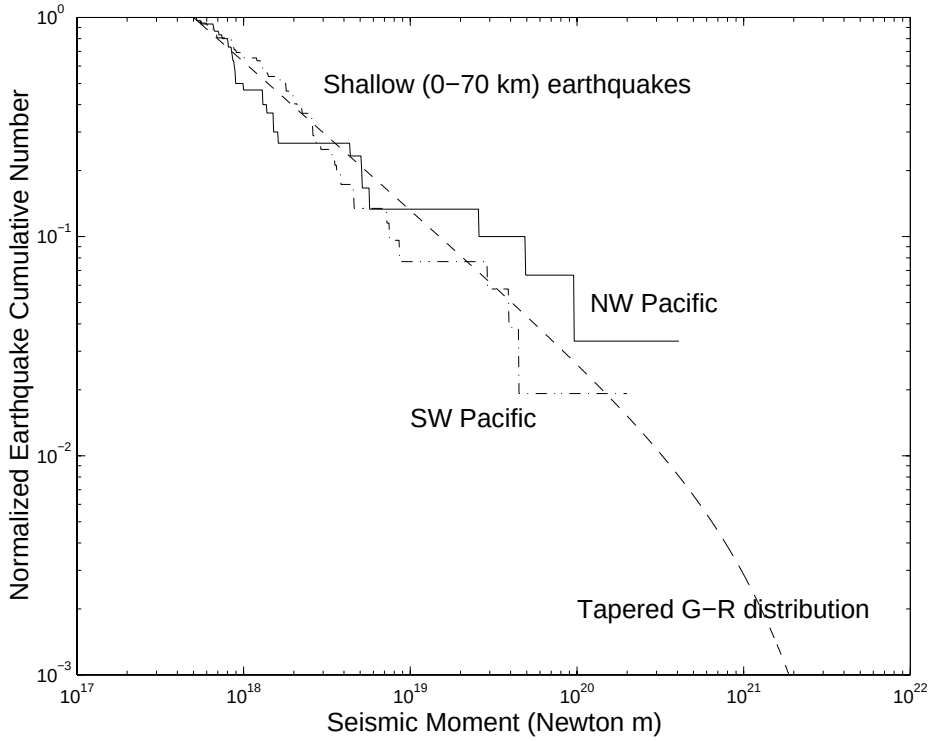


**Figure 10.** Earthquake cumulative number versus log of seismic moment for the shallow earthquake distribution in the 1999 January 1–1999 December 31 Harvard catalogue. The curves show the numbers of events with moment larger than or equal to $M$. We also show the approximation of curves by the tapered Gutenberg–Richter law (MGR).

(that is, a null hypothesis) for other prediction research and/or precursor claims.

This research has both scientific and practical results. Of scientific interest is our effort to characterize the statistical relationship between successive earthquakes in a quantitative way that will facilitate hypothesis testing. For example, it is widely believed that Coulomb stress increments from past earthquakes control aftershock behaviour, as well as the long-term occurrence of large earthquakes nearby (Stein *et al.* 1997; Toda *et al.* 1998). Case studies can provide important illustrative examples to support this model, but model verification as well as identifying the conditions under which the model works best and explaining exceptions require a thorough statistical treatment.

The forecasts will also have important practical benefits. Our quantitative predictive assessments may readily be adopted into rational mitigation strategies, which often require numerical hazard estimates. Our forecasts are in digital format, easily incorporated into Geographical Information Systems (GIS). By combining them with other GIS databases, the forecasts could be used to estimate probable loss of life, property damage and other consequences (HAZUS 1997).

Early alerts to probable strong earthquakes will allow emergency response personnel to anticipate and begin action for some potentially disastrous earthquakes. Since earthquake probability estimates tend to be modest, a chance of 'false alarms' will always be significant, thus the type of action justified will also be modest. Nevertheless, advanced planning of personnel schedules, intensive data collection, stockpiling of relief supplies and preparation for scientific monitoring can make earthquake response much more organized and effective.

Finally, we conclude that our long-term procedure forecasts earthquake probabilities in the NW Pacific reasonably well. The actual 1999 earthquake catalogue for the region is indistinguishable at the 95 per cent confidence level from synthetic ones generated from our forecast. To our knowledge, our NW Pacific forecast is the only one ever to pass such a test. For the SW Pacific, our forecast did not pass at the 95 per cent confidence level. There, the 1999 earthquakes were more concentrated in the high-rate regions, and less in the low-rate regions, than predicted by the forecast. We will update the forecast parameters, and time will tell whether the method will succeed there and in other places.

## NOTE ADDED IN PROOF

We obtained the Harvard catalogue for 1999 in final form too late to revise all pertinent computations. However, we updated Fig. 8 to show all of the predicted 1999 events. In the rest of the paper, we use only the first half of 1999 data from the final catalogue data; for the second half of the year we use preliminary solutions.

## ACKNOWLEDGMENTS

We appreciate partial support from the Southern California Earthquake Center (SCEC). SCEC is funded by NSF Cooperative Agreement EAR-8920136 and USGS Cooperative Agreements 14-08-0001-A0899 and 1434-HQ-97AG01718. The authors thank P. Bird and F. Schoenberg of UCLA, D. Vere-Jones of Wellington University and J. Ebel of Boston College for very useful discussions. We are grateful to D. Vere-Jones, A. F. Gangi and an anonymous referee for their reviews. Publication 516, SCEC.

## REFERENCES


Cao, T.Q., Petersen, M.D. & Reichle, M.S., 1996. Seismic hazard estimate from background seismicity in southern California, *Bull. seism. Soc. Am.,* **86,** 1372–1381.

Console, R., 1998. Computer algorithms for testing earthquake forecasting hypotheses, *Internal Rept,* Earthquake Research Institute, Tokyo.

Console, R. & Murru, M., 1996. Probability gain due to foreshocks following quiescence tested by synthetic catalogs, *Bull. seism. Soc. Am.,* **86,** 911–913.

Davis, S.D. & Frohlich, C., 1991. Single-link cluster analysis of earthquake aftershocks—decay laws and regional variations, *J. geophys. Res.,* **96,** 6335–6350.

Dziewonski, A.M., Ekström, G. & Maternovskaya, N.N., 2000. Centroid-moment tensor solutions for January-March 1999, *Phys. Earth planet. Inter.,* **118,** 1–11.

Ebel, J.E., Bonjer, K.-P. & Oncescu, M.C., 2000. Paleoseismicity: seismicity evidence for past large earthquakes, *Seism. Res. Lett.,* **71,** 283–294.

Evans, M., Hastings, N. & Peacock, B., 1993. *Statistical Distributions,* 2nd edn, Wiley, New York.

Frankel, A., 1995. Mapping seismic hazard in the central and eastern United States, *Seism. Res. Lett.,* **66**(4)**,** 8–21.

Frankel, A., Mueller, C., Barnhard, T., Perkins, D., Leyendecker, E.V., Dickman, N., Hanson, S. & Hopper, M., 1996. National seismic-hazard maps: documentation June 1996, *USGS Open-File Rept,* **96-532** (http://geohazards.cr.usgs.gov/eq/html/docmaps.shtml).

Frohlich, C. & Davis, S.D., 1999. How well constrained are well-constrained T, B, and P axes in moment tensor catalogs?, *J. geophys. Res.,* **104,** 4901–4910.

Hanks, T.C. & Kanamori, H., 1979. A moment magnitude scale, *J. geophys. Res.,* **84,** 2348–2350.

HAZUS Technical Manual, 1997. *Earthquake Loss Estimation Methodology: 'HAZUS' Technical Manual,* National Institute of Building Sciences, Federal Emergency Management Agency, Washington, DC.

Hutton, L.K. & Jones, L.M., 1993. Local magnitudes and apparent variations in seismicity rates in Southern California, *Bull. seism. Soc. Am.,* **83,** 313–329.

Jackson, D.D., 1996. Hypothesis testing and earthquake prediction, *Proc. Nat. Acad. Sci. USA,* **93,** 3772–3775.

Jackson, D.D. & Kagan, Y.Y., 1999. Testable earthquake forecasts for 1999, *Seism. Res. Lett.,* **70,** 393–403.

Jackson, D.D. & Kagan, Y.Y., 2000. Tectonic deformation and earthquake potential, *EOS, Trans. Am. geophys. Un.,* **81,** S407 (abstract).

Kafka, A.L. & Walcott, J.R., 1998. How well does the spatial distribution of smaller earthquakes forecast the location of larger earthquakes in the northeastern United States?, *Seism. Res. Lett.,* **69,** 428–440.

Kagan, Y.Y., 1973a. A probabilistic description of the seismic regime, *Izv. Acad. Sci. USSR, Phys. Solid Earth,* 213–219 (English translation).

Kagan, Y.Y., 1973b. Statistical methods in the study of the seismic process (with discussion), *Bull. Int. stat. Inst.,* **45**(3)**,** 437–453.

Kagan, Y.Y., 1991a. Seismic moment distribution, *Geophys. J. Int.,* **106,** 123–134.

Kagan, Y.Y., 1991b. Likelihood analysis of earthquake catalogues, *Geophys. J. Int.,* **106,** 135–148.

Kagan, Y.Y., 1992. Correlations of earthquake focal mechanisms, *Geophys. J. Int.,* **110,** 305–320.

Kagan, Y.Y., 1997. Are earthquakes predictable?, *Geophys. J. Int.,* **131,** 505–525.

Kagan, Y.Y., 1999. Universality of the seismic moment-frequency relation, *Pure appl. Geophys.,* **155,** 537–573.

Kagan, Y.Y. & Jackson, D.D., 1991. Long-term earthquake clustering, *Geophys. J. Int.,* **104,** 117–133.

Kagan, Y.Y. & Jackson, D.D., 1994. Long-term probabilistic forecasting of earthquakes, *J. geophys. Res.,* **99,** 13 685–13 700.

Kagan, Y.Y. & Jackson, D.D., 1995. New seismic gap hypothesis: five years after, *J. geophys. Res.,* **100,** 3943–3959.

Kagan, Y.Y. & Knopoff, L., 1987. Statistical short-term earthquake prediction, *Science,* **236,** 1563–1567.

Kawakatsu, H., 1995. Automated near-realtime CMT inversion, *Geophys. Res. Lett.,* **22,** 2569–2572.

Kotz, S. & Johnson, N.L., 1985. *Encyclopedia of Statistical Sciences,* Vol. 6, Wiley, New York.

Maeda, K., 1996. The use of foreshocks in probabilistic prediction along the Japan and Kuril trenches, *Bull. seism. Soc. Am.,* **86,** 242–254.

Main, I., Irving, D., Musson, R. & Reading, A., 1999. Constraints on the frequency–magnitude relation and the maximum magnitudes in the UK from observed seismicity and glacio-isostatic recovery rates, *Geophys. J. Int.,* **137,** 535–550.

McCann, W.R., Nishenko, S.P., Sykes, L.R. & Krause, J., 1979. Seismic gaps and plate tectonics: seismic potential for major boundaries, *Pure appl. Geophys.,* **117,** 1082–1147.

Michael, A.J. & Jones, L.M., 1998. Seismicity alert probabilities at Parkfield, California, revisited, *Bull. seism. Soc. Am.,* **88,** 117–130.

Molchan, G.M. & Kagan, Y.Y., 1992. Earthquake prediction and its optimization, *J. geophys. Res.,* **97,** 4823–4838.

Nishenko, S.P., 1991. Circum-Pacific seismic potential—1989–1999, *Pure appl. Geophys.,* **135,** 169–259.

NRC (National Research Council) Panel, 1991. *Real-time Earthquake Monitoring: Early Warning and Rapid Response,* National Academy Press, Washington, DC.

Ogata, Y., 1988. Statistical models for earthquake occurrence and residual analysis for point processes, *J. Am. stat. Assoc.,* **83,** 9–27.

Ogata, Y., 1998. Space-time point-process models for earthquake occurrences, *Ann. Inst. stat. Mech.,* **50,** 379–402.

Pacheco, J.F. & Sykes, L.R., 1992. Seismic moment catalog of large, shallow earthquakes, 1900–1989, *Bull. seism. Soc. Am.,* **82,** 1306–1349.

Reasenberg, P.A., 1999. Foreshock occurrence before large earthquakes, *J. geophys. Res.,* **104,** 4755–4768.

Reasenberg, P.A. & Jones, L.M., 1989. Earthquake hazard after a mainshock in California, *Science,* **243,** 1173–1176.

Silverman, B.W., 1986. *Density Estimation for Statistics and Data Analysis,* Chapman & Hall, London.

Stark, P.B., 1997. Earthquake prediction: the null hypothesis, *Geophys. J. Int.,* **131,** 495–499.

Stein, R.S., Barka, A.A. & Dieterich, J.H., 1997. Progressive failure on the North Anatolian fault since 1939 by earthquake stress triggering, *Geophys. J. Int.,* **128,** 594–604.

Toda, S., Stein, R.S., Reasenberg, P.A., Dieterich, J.H. & Yoshida, A., 1998. Stress transferred by the 1995 $M_w = 6.9$ Kobe, Japan, shock: effect on aftershocks and future earthquake probabilities, *J. geophys. Res.,* **103,** 24 543–24 565.

Utsu, T., 1999. Representation and analysis of the earthquake size distribution: a historical review and some new approaches, *Pure appl. Geophys.,* **155,** 509–535.

Utsu, T. & Ogata, Y., 1997. Statistical analysis of seismicity, in, *IASPEI Software Library,* Vol. 6, pp. 13–94, eds Healy, J.H., Keilis-Borok, V.I. & Lee, W.H.K., Int. Assoc. Seism. Phys. Earth's Inter. and Seism. Soc. Am., El Cerrito, CA.

Vere-Jones, D., 1970. Stochastic models for earthquake occurrence (with discussion), *J. R. stat. Soc.,* **B32,** 1–62.

Vere-Jones, D., 1992. Statistical methods for the description and display of earthquake catalogues, in *Statistics in the Environmental and Earth Sciences,* pp. 220–244, eds Walden, A.T. & Guttorp, P., Arnold, London.

Vere-Jones, D., Robinson, R. & Yang, W.Z., 2000. Remarks on the accelerated moment release model, *Geophys. J. Int.,* submitted.

Woo, G., 1996. Kernel estimation methods for seismic hazard area source modeling, *Bull. seism. Soc. Am.,* **86,** 353–362.