

# UCLA

## UCLA Previously Published Works

### Title

Testing long-term earthquake forecasts: likelihood methods and error diagrams

### Permalink

<https://escholarship.org/uc/item/6v37q066>

### Journal

Geophysical Journal International, 177(2)

### Author

Kagan, Yan Y

### Publication Date

2009-04-01

### DOI

10.1111/j.1365-246X.2008.04064.x

Peer reviewed

# Testing long-term earthquake forecasts: likelihood methods and error diagrams

Yan Y. Kagan<sup>1\*</sup>

<sup>1</sup> Department of Earth and Space Sciences  
University of California, Los Angeles, California, USA

December 10, 2008

## Abstract

We propose a new method to test the performance of a spatial point process forecast 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 use and allows us to calculate the information score for each event or set of events as well as the standard error of each forecast. As the number of predicted events increases, the score distribution approaches the Gaussian law. The degree of its similarity to the Gaussian distribution can be measured by the computed coefficients of skewness and kurtosis. To display the forecasted point density and the point events, we

---

\*Yan Y. Kagan, Department of Earth and Space Sciences, University of California, Los Angeles, California, 90095-1567, USA; (e-mail: ykagan@ucla.edu)

use an event concentration diagram or a variant of the *Error Diagram* (ED). We find forward relation between the error diagram curve and the information score as well as inverse relation for one simple model of point spatial fields. We again show that the error diagram is more informative than the likelihood ratio.

We demonstrate the application of the method by using our long-term forecast of seismicity in two western Pacific regions. We compare the ED for these regions with simplified diagrams based on two-segment approximations. Since the earthquakes in these regions are concentrated in narrow subduction belts, using the forecast density as a template or baseline for the ED is a more convenient display technique. We also show, using simulated event occurrence, that some proposed criteria for measuring forecast effectiveness at EDs would be strongly biased for a small event number.

KEY WORDS: Probabilistic forecasting; Spatial analysis; Fractals and multifractals; Probability distributions; Earthquake interaction, forecasting, and prediction; Statistical seismology.

# 1 Introduction

This paper continues our analysis of stochastic point process forecast verification (Kagan 2007b). There (*ibid.*) we had discussed two interrelated methods for measuring the effectiveness of earthquake prediction algorithms: the information score based on the likelihood ratio (Kagan 1991) and the “Error Diagram” (ED). These methods have been applied (Kagan 2007b) to 1-D temporal renewal stochastic processes, but only for very long processes with the number of events approaching infinity.

In this work we extend our analysis by

- 1) discussing spatial (not temporal) random processes (fields);
- 2) considering forecast testing if the number of events is relatively small;
- 3) applying newly developed techniques to long-term earthquake forecasts.

Two issues are related to the problem of testing point process forecasts:

- 1) Spatial random point fields density evaluation and its prediction is a mature discipline with many publications. Baddeley *et al.* (2005), Baddeley (2007), Daley & Vere-Jones (2003, 2008) provide reviews. As we explain below, the earthquake forecasting problem is different in many respects from regular density evaluation and requires special treatment. However, some results of this paper can be applied to test the forecast of a random spatial pattern.
- 2) Well-developed application methods exist in weather and climate pre-

diction; their reviews have been recently published by Jolliffe & Stephenson (2003), Palmer & Hagedorn (2006), DelSole & Tippett (2007). These prediction methods consider continuous processes and fields; however, with necessary modifications some of these methods can be used for stochastic point processes.

Our main focus here is on the two most widely used approaches to assessing earthquake prediction methods (Zaliapin & Molchan 2004). Both approaches evaluate how a prediction method reveals new information about impending earthquake activity. The first approach starts by estimating the expected spatio-temporal distribution of seismicity and uses the classical likelihood paradigm to evaluate predictive power. Accordingly, it uses the nomenclature of statistical estimation. The second one applies the results by Molchan (1990, 1997; see also Molchan & Keilis-Borok 2008 and Molchan, 2008) who proposed error diagrams for measuring prediction efficiency. The EDs plot the normalized rate of failures-to-predict ( $\nu$ ) versus the normalized time of alarms ( $\tau$ ). The ED can be considered as a time-dependent analog of the Neyman-Pearson lemma on making a decision: should we expect an earthquake within a given spatio-temporal region? Consequently, it uses the language of hypothesis testing.

Starting with Molchan's (1990) paper, previous EDs were almost exclusively time-dependent. We apply the ED to time-independent 2-D spatial earthquake distributions. In some respects, the earthquake spatial pattern is

more difficult to analyze than the temporal distribution. In the latter case, we have a reasonable null model (the uniform in time Poisson process) which can be compared to any test model. In the spatial case, the simple model of uniformly distributed seismicity can hardly serve as an initial approximation; even large earthquakes (which often can be well-approximated by a Poisson temporal process) are strongly clustered in space. This property of seismicity is caused by the fractal nature of earthquake spatial distribution (Kagan 2007a). Although in our forecast (Kagan & Jackson 2000) we use a projection of earthquake centroids on the Earth surface which smoothes their spatial distribution (Kagan 2007a), the spatial distribution still preserves a self-similar fractal pattern with large parts of the Earth practically aseismic.

Diagrams similar to EDs have been used previously to describe spatial distribution of seismicity: Rong & Jackson (2002, their Fig. 3); Kagan *et al.*, (2003, their Fig. 5.3); Helmstetter *et al.*, (2007, their Fig. 4); and Shen *et al.*, (2007, their Figs. 1B, 2B) created spatial “Concentration Diagrams” to characterize the agreement (or lack thereof) between the predicted seismicity distribution and future earthquakes. These diagrams plot the fraction of the event success rate (equivalent to  $1 - \nu$ ) versus the normalized area ( $\tau$ ), sorted by probability density. The sorting is largely analogous to water-level threshold analysis (Zechar & Jordan 2008). These concentration diagrams can easily be converted to EDs by adding an ascending diagonal and then reflecting the plot in the line ordinate ( $\nu = 1/2$ ).

The error diagram is related to the *Relative Operating Characteristic* (ROC) (Swets 1973; Mason 2003, pp. 66-76), used in signal detection and weather prediction efforts. In the ROC diagrams the success rate of an event prediction is compared against the false alarm rate.

In principle, ED diagrams can be treated like ROC plots where cells (bins) with events are considered as success and empty cells as false alarms. However, this interpretation encounters difficulties when the cells are not infinitesimally small, so some may contain more than one event. Even for very small cells due to the fractal nature of earthquake spatial distribution (Kagan, 2007a) several events may be located in a single cell. However, in our forecast we use a projection of earthquake centroids on a 2-D Earth surface; this transformation smoothes out earthquake spatial pattern, thus for a sufficiently dense grid the probability of more than one event falling into the same bin can be kept low. Moreover, as we show below, for a point process on a sphere it is difficult to define cells of equal size. A usual sphere subdivision yields cells of unequal size that are larger at the equator and smaller towards the poles. Below we discuss the techniques for overcoming these difficulties.

Characterizing prediction performance is a major challenge for ED analysis. Since prediction results are represented by a function (curve), it is important to find a simple one-parameter criterion (a functional) that briefly expresses the efficiency value. Several functionals have been proposed as

a measure of ED forecast efficiency: the *minimax* strategy (Molchan 1997; Molchan & Keilis-Borok 2008), the *sum of errors* ( $\nu + \tau$ ) (*ibid.*), the *area skill score* (Zechar & Jordan 2008), etc. Each of these criteria has some advantages or disadvantages. For example, Kagan & Jackson (2006, Section 5) in their discussion of Kossobokov’s (2006) paper, show that two ED trajectories with a very different behavior have the same ‘sum of errors’ value.

In this work, as well as in our previous paper (Kagan, 2007b), we advocate the use of the log-likelihood to characterize the ED performance. The advantage of the likelihood score is that it is well-known in statistics and its properties are discussed in many statistical treatises. Moreover, as Kagan & Jackson (2006, Section 5) argue, the likelihood score yields a better measure of forecast performance in terms of possible earthquake warning strategies and their cost.

## 2 Long-term earthquake forecasts

Kagan & Jackson (1994, 2000) present long-term and short-term earthquake forecasts in several regions using the CMT catalog (Ekström *et al.* 2005; <http://www.globalcmt.org/>). The forecasted earthquake rate is calculated as a spatially smoothed earthquake location distribution. The spatial kernel used in the smoothing has units of earthquakes per unit area and time. In our studies it applies to all shallow (depth less or equal 70 km) earthquakes with moment  $M = 10^{17.7}$  Nm (magnitude 5.8) and greater. The kernel

is elongated along the fault-plane, which is estimated from available focal mechanism solutions.

To take into account the impact of earthquakes outside the boundaries of the region, we allow events up to 1000 km outside the region to contribute to the rate density inside the forecast region. The additional rate density from outside events is on average balanced by a contribution ‘leakage’ from many ‘insider’ earthquakes close to the boundaries.

An important feature of Kagan & Jackson’s (1994) method is a procedure similar to the jack-knife (Silverman 1986) for testing the predictive power of the smoothing. It optimizes the kernel parameters choosing those values which best predict the second half of a catalogue, using a maximum likelihood criterion, from the first half. We argue that because the seismicity pattern exhibits a long-term clustering (Kagan & Jackson 1991), such a procedure is better suited than the standard density estimation techniques to predicting the future earthquake rate. We also assume on an *ad hoc* basis (Kagan & Jackson 1994, 2000) that the background rate density is uniform over the whole region and integrates to 1% of the total earthquake rate

$$(\text{Background rate}) = \epsilon \times (\text{Total rate}), \quad (1)$$

with  $\epsilon = 0.01$ . A quantitative determination of the  $\epsilon$ -value would require an extensive study not of a regional but the global spatial earthquake distribution, a task for future work.

Kagan (2007a) shows that the fractal dimension of earthquake hypocenters,  $\delta$ , strongly depends on the earthquake catalog time interval. For temporally short catalogs  $\delta$  is close to zero and approaches the asymptotic value  $\delta \approx 2.3$  for catalogs of decades length. In a more intuitive setting, this result signifies that in short time intervals, hypocenters are concentrated in a few point clouds. With increased time, seismicity spreads over a fault system or seismic belt length, eventually occupying a set with the dimension in excess of the 2-D plane. Therefore, if one uses a set of earthquake epicenters in a relatively short catalog to predict the future seismicity rate, the optimal forecast kernel should spread beyond the presently available event points, i.e., to be smoother than the standard density estimators (Silverman 1986) would suggest.

The forecasts are expressed as the rate density (that is, the probability per unit area and time). They are updated every day and posted for two western Pacific regions at [http://scec.ess.ucla.edu/~ykagan/predictions\\_index.html](http://scec.ess.ucla.edu/~ykagan/predictions_index.html) (see FORECAST TEST FOR 2004-2006:). Table 1 displays a slightly modified small extract of the forecast tables available at the Web site. The values of earthquake rate densities (column 3) or cell rates (column 5) are calculated at the end of year 2003 and they are used as a forecast for 2004-2006 earthquakes. In this Table we sorted (ordered) entries by the values of earthquake forecast densities (column 3).

In Fig. 1 we display the long-term forecast map computed for the northwest (NW) Pacific region using the CMT catalog for 1977-2003. Shallow earthquakes in 2004-2006 are shown in white. Similar maps for NW and southwest (SW) Pacific are shown, for instance, in Figs. 8a,b by Kagan & Jackson (2000). The NW Pacific region boundaries are: latitude limits from  $0.25^{\circ}\text{S}$  to  $60.25^{\circ}\text{N}$ , longitude limits from  $109.75^{\circ}\text{E}$  to  $170.25^{\circ}\text{E}$ . The SW Pacific boundaries are: latitude limits from  $0.25^{\circ}\text{N}$  to  $60.25^{\circ}\text{S}$ , longitude limits from  $109.75^{\circ}\text{E}$  to  $169.75^{\circ}\text{W}$ .

In this work we use the same values for a smoothing kernel as in Kagan & Jackson (2000, see their Eq. 3): the spatial scale parameter,  $r_s$  is 15 km and 2.5 km for the NW and SW Pacific, respectively. The azimuthal concentration factor (Eq. 6, *ibid.*) is 100 and 25, respectively. On visual inspection, the model predicts the spatial distribution of seismic activity reasonably well. We tested this forecast by a Monte-Carlo simulation (Kagan & Jackson 1994).

In Fig. 2 we show the forecasted earthquake density with 10 sets of synthetic catalogs, each having 108 events. Earthquakes are assumed to occur at the centers of grid cells with the rates defined by the forecast. We normalize the cell rates as shown in Table 1 (column 5) and simulate a random number uniformly distributed in the interval  $[0, 1]$ . The random number corresponding to a particular segment of the cumulative normalized rate curve defines the cell where an event occurs. We obtain one synthetic catalog by repeating this procedure  $n$  times.

Some of the grid points are occupied by more than one event. Some of the simulated points occur in areas of low seismicity (compare Fig. 2 with Fig. 1). As mentioned above, this feature of the forecast is used to prevent surprises, i.e., an occurrence of earthquakes in zones where no nearby events happened in 1977-2003.

Table 2 summarizes annual earthquake rates for both western Pacific regions. Because events outside the region's boundaries have influence, the rates calculated through the smoothing procedure and evaluated by a direct method (dividing the earthquake numbers by time interval) are close but do not coincide. The difference between the predicted ( $v_0$ ) and observed ( $v_i$ ) numbers is not statistically significant; for a Poisson process the standard error is  $\sqrt{v}$ . Kagan & Jackson (2000) argue that the earthquake number distribution follows a negative-binomial law which has a higher variance than the Poisson distribution with the same mean.

### 3 Log-likelihood

The standard method of statistical analysis for a stochastic point process is a likelihood function computation. For an inhomogeneous Poisson process in which  $n$  points are observed  $(x_1, \dots, x_n)$  in a region  $A$ , the log-likelihood can be written as (Daley & Vere-Jones 2003, Eq. 7.1.2)

$$\log L(x_1, \dots, x_n) = \sum_{i=1}^n \log \lambda(x_i) - \int_A \lambda(x) dx, \quad (2)$$

where  $\lambda(x_i)$  is the process rate (density) at a point  $x_i$ .

The log-likelihood of an inhomogeneous Poisson process is normally compared to a similar log-likelihood,  $L_0$ , calculated for a Poisson process with constant intensity ( $\xi$ ) to obtain the log-likelihood ratio (Daley & Vere-Jones 2003, Ch. 7; Schorlemmer *et al.* 2007)

$$\log(L/L_0) = \sum_{i=1}^n \log \lambda(x_i/\xi) - \int_A [\lambda(x) - \xi] dx. \quad (3)$$

In our calculations we normalize both rates ( $\lambda, \xi$ ) by the observed event number  $n$ , hence the integral term in (3) is zero.

Kagan & Knopoff (1977, see also Vere-Jones 1998) suggested measuring the performance of the earthquake prediction algorithm by first evaluating the likelihood ratio to test how well a model approximates an earthquake occurrence. In particular, they estimated the information score,  $\hat{I}$ , per one event by

$$\hat{I} = \frac{\ell - \ell_0}{n} = \frac{1}{n} \sum_{i=1}^n \log_2 \frac{\lambda_i}{\xi}, \quad (4)$$

where  $\ell - \ell_0$  is the log-likelihood ratio,  $n$  is the number of earthquakes in a catalog,  $\log_2$  is used to obtain the score measured in the Shannon bits of information,  $\lambda_i$  is the rate of earthquake occurrence according to a stochastic model, conditioned by the past:

$$\lambda_i = \text{Prob} \{ \text{an event in } (t_i, t_i + \Delta) \mid I(t_i) \}, \quad (5)$$

where  $I(t_i)$  is the past history of the process up to the moment  $t_i$ , and  $\xi$  is a similar rate for the event occurrence according to the Poisson process with

a uniform rate over a region. The interval  $\Delta$  is infinitesimally small so that the probability and the intensity (rate) are equivalent.

The Poisson process rate can be calculated by normalizing the seismicity level in the forecast regions. Several rates, such as shown in Table 2, can be used in the normalization. To make our results comparable to the forecast rate density, we use  $\nu_0$  values

$$\xi = \frac{\pi \nu_0}{180.0 \times [\sin(\theta_u) - \sin(\theta_l)] (\phi_u - \phi_l) \times 111.111^2 \times 365.25}, \quad (6)$$

where  $\nu_0$  is the annual rate of earthquakes in each region in 1977-2003 (Table 2),  $\theta_u$  and  $\theta_l$  are the upper and lower latitudes, respectively,  $\phi_u$  and  $\phi_l$  ditto for longitudes. For the NW-Pacific region  $\xi_{NW} = 2.6289 \times 10^{-9}$  eq/(day  $\times$  km<sup>2</sup>); for the SW-Pacific  $\xi_{SW} = 3.3479 \times 10^{-9}$  eq/(day  $\times$  km<sup>2</sup>). Below we use  $\xi$  without a subscript, since it is usually clear which Pacific region is discussed.

Several methods can be used in calculating the information score for a set of forecasted events. Using the forecasted rate values ( $\lambda_i$  for cell centers in which earthquakes occurred) we compute

$$I_1 = \frac{1}{n_j} \sum_{i=1}^{n_j} \log_2 \frac{\lambda_i}{\xi}, \quad (7)$$

where  $n_j$  is the earthquake number in two Pacific regions during 2004-2006 ( $j = 1, 2$ , for NW or SW, respectively, see Table 2). Below we use  $n$  without a subscript, since it is usually clear which Pacific region is discussed.

In Eq. 7 and in derivations below, we assume that earthquakes in the cells are identically distributed independent (i.i.d.) events. The assumed independence may be challenged by the clustered nature of earthquake occurrence of which foreshock-mainshock-aftershock sequences are the most clear example (Kagan & Knopoff 1977; Kagan 1991). However, given the high magnitude (5.8) threshold for the CMT catalog, the clustering is less pronounced. The dependent events on average constitute only about 20% of the total seismic activity (Kagan & Jackson 2000, Eq. 23). Thus, we expect that earthquake statistical inter-dependence would have relatively small influence. A more complete investigation of this problem will be done in our future work.

As another option, instead of (7) we compute the information score for the actual epicenter (centroid) locations ( $\lambda_k$ )

$$I_2 = \frac{1}{n} \sum_{k=1}^n \log_2 \frac{\lambda_k}{\xi}. \quad (8)$$

As we see from Table 3 below, the values of  $I_1$  and  $I_2$  may be significantly different. The score  $I_1$  depends on the cell grid subdivision and therefore is a less reliable indicator of the forecast effectiveness. On the other hand, its calculation can be accomplished using an available forecast table such as Table 1, whereas  $I_2$  requires new extensive computation.

The scores  $I_1$  and  $I_2$  show statistical behaviour of earthquakes that occurred after the forecast was made. In our examples 2004-2006 events are compared to the forecast based on a smoothed seismicity of 1977-2003. We

compare these forecasts with a score estimate for an extrapolated earthquake record based on the forecast. One way to calculate such a record is to create synthetic earthquake catalogs (Kagan & Jackson 1994).

In simulated catalogs we generate multiple ( $\mathcal{N} = 10,000$ ) sets of  $n$  events (Table 2) and calculate the rate for cell centers as the earthquake location (see Fig. 2)

$$I_3 = \frac{1}{n} \sum_{l=1}^n \log_2 \frac{\lambda_l}{\xi}. \quad (9)$$

and

$$\langle I_3 \rangle = \frac{1}{\mathcal{N}} \sum_{\ell=1}^{\mathcal{N}} (I_3)_\ell. \quad (10)$$

Similar to  $I_1$  calculation, this method has an advantage that we do not need to compute the rate densities again (as for  $I_2$ ), but instead we use the previously computed forecast tables (as shown in Table 1) to evaluate the scores. The score  $\langle I_3 \rangle$  is equivalent to (the opposite of) an entropy measure (Daley and Vere-Jones, 2003), since the summation (sum on the earthquakes, Eq. 10) is similar to

$$I = \frac{1}{n} \sum_{l=1}^n \lambda_l \log \lambda_l, \quad (11)$$

(sum on the cells) on ensemble average.

The score  $\langle I_3 \rangle$  is a measure of potential prediction effectiveness as compared to a Poisson process with a uniform rate over a region. Although the score is constructed using a previous earthquake record, its usefulness for forecast purposes needs to be established – it may be argued, for example,

that earthquakes ‘avoid’ the locations of past recent strong events (as in the seismic gap model, see Rong *et al.* 2003). Therefore, we calculate the score difference  $I_3 - I_2$  or  $I_3 - I_1$  to characterize the forecast properties. In short, if the difference is close to zero, it means that future earthquakes in a statistical sense follow the predictions. Kagan & Jackson (1994, see their Fig. 7 and its discussion) explain the test in more detail.

In Fig. 3 we display the log-likelihood function distribution differences for the simulation as shown in Fig. 2. We simulate  $n$  earthquake locations according to 1977-2003 forecasts for each region. For each synthetic catalog we calculate the log-likelihood function and subtract the log-likelihood function value obtained for the real catalogue in 2004-2006. Thus, we display the histogram of  $I_3 - I_2$  (Eqs. 8 and 9).

## 4 Error diagrams

To display and test the long-term forecast efficiency numerically, we calculate the concentration diagram. To make these diagrams, we divide the region into small cells (0.5 by 0.5 degrees for west Pacific regions) and estimate the theoretical forecast rate of earthquakes above the magnitude threshold for each cell. We then count the events that actually occurred in each cell, sort the cells in the decreasing order of the theoretical rate, and compute the cumulative values of forecast and the observed earthquake rates (see

Table 1). Similar plots have been used in several of our papers (Kagan *et al.* 2003; Helmstetter *et al.* 2007; Shen *et al.* 2007).

In effect, these concentration diagrams are equivalent to the error diagrams (EDs) proposed by Molchan (1990, 2003) and Molchan & Kagan (1992). The error diagram evaluates how well a prediction program performs. For any prediction algorithm, the diagram plots the cumulative fraction of the alarm time,  $\tau$ , versus the cumulative fraction of failures to predict,  $\nu$ . But in our case we use the normalized spatial area, not time, as the horizontal axis.

#### 4.1 Relation between the error diagram and information score

We illustrate the ED by a sketch in Fig. 4. For the spatial point distribution, this example is easier to construct and explain than for temporal renewal processes (Kagan 2007b). In the plot we show a theoretical pattern for two ED diagrams. The square's diagonal corresponds to the uniform Poisson distributions of the points in a region, i.e., a random guess forecast strategy or *unskilled* forecast. As a test example, we assume that the region consists of three sub-areas, their normalized surfaces  $\tau_i$  is 0.1, 0.5, and 0.4 of the total, and the normalized number of events  $\nu_i$  is 0.4, 0.5, and 0.1, in each zone respectively. Contrary to the temporal one-dimensional point process models (as discussed in Kagan, 2007b), for a spatial field these sub-areas do

not need to be contiguous, each of them can be subdivided in any number of patches. It is important that coordinates of all patches be known. The points in these zones are distributed according to the Poisson spatial process with the density  $\nu_i/\tau_i$ . Then, the information score for such a point distribution can be calculated as (see Eq. 4)

$$\begin{aligned} I &= \sum_{i=1}^3 \nu_i \log_2 \frac{\nu_i}{\tau_i} = 0.4 \log_2 4.0 + 0.5 \log_2 1.0 + 0.1 \log_2 0.25 \\ &= 0.8 - 0.2 = 0.6. \end{aligned} \quad (12)$$

For the normalized point Poisson distribution in the ED, the point density is unity. Hence its contribution to the information rate (12) is zero.

The information score can be calculated for continuous theoretical concave curves in an error diagram (Kagan 2007b; Molchan, 2008, Eq. 20)

$$I = \int_0^1 \log_2 \left( -\frac{\partial \nu}{\partial \tau} \right) d\nu. \quad (13)$$

If the ED consists of several linear segments (as in Fig. 4), then (13) converts to a sum

$$I_0 = \sum_{i=1}^N \nu_i \log_2 \left( \frac{\nu_i}{\tau_i} \right), \quad (14)$$

where  $i$  are cell numbers,  $N$  is the total number of grid points, and  $\nu_i$  and  $\tau_i$  are the normalized rates of occurrence and cell area:

$$\nu_i = \frac{R_i}{\sum_{i=1}^N R_i} \quad \text{and} \quad \tau_i = \frac{S_i}{\sum_{i=1}^N S_i}, \quad (15)$$

see Table 1. The above formulae show that both ED components are normalized, i.e.,

$$\sum_{i=1}^N \nu_i = 1, \quad \text{and} \quad \sum_{i=1}^N \tau_i = 1. \quad (16)$$

In our ED plots, we use as abscissa and ordinate of the diagrams the cumulative fraction of alarm space and failures to predict

$$\tau = \sum_i \tau_i, \quad \text{and} \quad \nu = \sum_i \nu_i. \quad (17)$$

When calculations of  $\tau$  are made for a spherical surface (as in Figs. 1-2), the  $\tau_i$  steps are usually unequal in size, unless a special effort is made to partition a sphere into equal-area cells (see more in Kagan & Jackson 1998). This cell inequality complicates the calculation.

Figs. 5 and 6 show the EDs for both Pacific regions. The red curves are for the forecast, based on 1977-2003 seismicity, and the blue curves are for the earthquakes which occurred in these regions from 2004-2006. Both sets of curves are calculated using the forecast tables like those in the example (Table 1). In principle, the calculations such as in (15) can be made with unordered cells. The density ordering in Table 1 and Figs. 5, 6 is performed to create the ED diagrams.

The score values  $I_0$  (14) in Table 3 are calculated using the distributions shown by the red curves in Figs. 5, 6. The  $I_0$  values for NW- and SW-Pacific indicate that the forecast yields an information score higher than 2-3 bits per event compared to a homogeneous Poisson process. To obtain the average

probability gain  $G$  or the predictive ratio (Kagan and Knopoff, 1977, p. 101) we calculate

$$G = 2^{I_0}. \quad (18)$$

Zechar & Jordan (2008, Eq. 9) derived a similar formula for the average probability gain of a forecast; their equation can be converted into our (13) taking into account their different normalization.

Eq. 18 means that on average the probability gain ( $G$ ) is a factor of 5 to 10 ( $2^{2.36}$  to  $2^{3.38}$ ) when using the long-term forecast compared to a random guess. Of course, these  $I_0$  values do not fully describe the forecast advantage. The boundaries of both regions have already been selected to contain the maximum number of earthquakes in relatively small areas. If we extend any of the regions toward the seismically quiet areas, the information score would significantly increase. The proper measure of long-term forecast performance would extend the forecast method globally, i.e., over the whole Earth surface. Limited numerical experiments suggest that depending on the degree of smoothing, the value of  $\epsilon$  (Eq. 1), and other factors, the  $G$ -value for world-wide seismicity varies from about 10 to 25.

The above values of the probability gain,  $G$ , can be compared with similar calculations by Rhoades & Evison (2005, 2006), Console *et al.* (2006), and Rhoades (2007). These authors calculated *the information rate per earthquake* for a model of smoothed seismicity (PPE), similar to our long-term

model. The PPE model was compared to a stationary and spatially uniform Poisson (SUP) model. The probability gain, computed using the information rate, for New Zealand, Japan, Greece, and California is about 4.5, 1.6, 1.6, and 3.4, respectively. These relatively small gain values are caused by the authors' choice of the regions that include only seismically active areas (see *ibid.*). Helmstetter *et al.* (2007, Table 1) obtained for different long-term seismicity predictive models in California the  $G$ -values ranging from 1.2 to 4.8.

The ED curves for earthquakes in Figs. 5, 6 are similar to the forecast earthquake curves. The computation of the likelihood scores (7, 8) shows that the NW earthquakes have a better score than the forecast, whereas SW events display the opposite behavior (see also Fig. 3). The scores using the actual centroid position ( $I_2$ ) are larger than those for the cell centers ( $I_1$ ), an anticipated feature. Similarly, Table 3 shows that the average scores for synthetics ( $\langle I_3 \rangle$ ) are very close to those of  $I_0$ , which is understandable, since the simulation runs are extensive (see Eqs. 9, 10).

Fig. 7 shows the frequency curves for the log-likelihood function of both western Pacific regions. We display  $\log_2$  of the normalized rate (see column 5 of Table 1) against the normalized cumulative area of the cells (column 4). Curves for both regions exhibit high values of the rate ( $R_i$ ) concentrated in a relatively small fraction of area. For example, the rate density exceeding the Poisson rate ( $\xi$ , see Eq. 6) is observed only at less than 18% of the

total area. Low values at the right-hand end of the diagram correspond to the assumed uniform background rate density (Eq. 1, see also Figs. 1-2). We also display two curves which are calculated as if the log of forecasted normalized earthquake density was distributed according to the normal law with the standard deviations ( $\sigma$ ) as shown in Table 3.

The curves shown in Fig. 7 are dependent on the log-likelihood function for one forecasted event on the fraction of total area  $\tau$ . The log-likelihood distribution for both regions is highly non-Gaussian: very few cells with low density would be expected for the Gaussian law, whereas in our forecast a significant part of the area has the density equal to the  $\epsilon$ -value (see Figs. 1-2). When we test the prediction for several events, the distribution seems to be close to the Gaussian distribution (Fig. 3). The tests of the forecast effectiveness are much simpler if the distribution is Gaussian or sufficiently close to it.

To measure the difference between the actual log-likelihood distribution and the Gaussian one, we calculate the higher order moments for the error curve ( $I_0$  of Eq. 14 corresponds to the first moment  $\mu_1$ )

$$\mu_k = \sum_{i=1}^N \nu_i \left[ \log_2 \left( \frac{\nu_i}{\tau_i} \right) - I_0 \right]^k, \quad (19)$$

where  $k = 2, 3, 4, \dots$

The standard deviation of the log-likelihood for the set of  $n$  events is

$$\sigma_n = \sqrt{\mu_2/n}. \quad (20)$$

The coefficient of skewness (Abramowitz and Stegun, 1972, p. 928) is

$$\eta = \mu_3/\mu_2^{3/2}, \quad (21)$$

and coefficient of kurtosis (*ibid.*) is

$$\psi = \mu_4/\mu_2^2 - 3. \quad (22)$$

These coefficients ( $\eta$  and  $\psi$ ) characterize how the likelihood curve differs from the Gaussian distribution; for the latter law both coefficients should be zero. The *Central Limit Theorem* states that the distribution of a sum of a large number of i.i.d. events with finite second moment (variance) should approach the Gaussian law. If the event number is small, we would need to find an efficient way to numerically approximate the distribution of the sum of i.i.d. random variables.

In Table 3 both coefficients are large for one event likelihood curve (see also Fig. 7), but for the set of  $n$  events they are small: the distribution is close to the Gaussian law as demonstrated in Fig. 3. The difference between the score values  $I_0$  to  $I_2$  is less than the standard error value (see Table 3). Thus both forecasts can be considered statistically successful.

The difference

$$I' = I_0 - I_1 \quad \text{or} \quad I'' = I_0 - I_2, \quad (23)$$

shows the predictive efficiency of a forecast, i.e., whether on average earthquakes in 2004-2006 occurred at the sites listed in the prediction table (see

an example in Table 1). For this particular time interval, both forecasts are sufficiently good. However, as other examples (Kagan & Jackson 2000, Fig. 9; Kagan *et al.* 2003, Fig. 5.2) demonstrate, this is not always the case. The values of differences (negative for the NW-Pacific and positive for the SW-Pacific) correspond to those simulations in Fig. 3, where we display the distribution of the difference  $I_3 - I_2$ .

By applying (14) to the blue curve of earthquakes in 2004-2006 in Figs. 5, 6 we evaluate the information score

$$I_4 = \frac{1}{n} \sum_{i=1}^n \nu_i \log_2 \left[ \frac{\nu_i}{\tau_i} \right], \quad (24)$$

(see Table 3). The value of  $I_4$  is obviously significantly larger than all the other estimates of the score. Earthquake simulations provide an explanation for this feature (see Fig. 10 below).

## 4.2 Two-segment error diagrams and information score

Similarly to Fig. 5 in Kagan (2007b), in Fig. 8 we display an approximation of the ED for the NW-Pacific by several two line segment diagrams with the same value of the information score,  $I_0$ . These ED segments correspond to two sub-areas with different point densities. As we discussed above (Eq. 12), for a spatial stochastic field these sub-areas may be subdivided into any number of patches with known coordinates.

For the assumed information score  $I$ , the contact point of two segments is defined by the equation (corrected Eq. 22 by Kagan 2007b)

$$D_1 \left[ \frac{\nu}{\nu - 1 - D_1} \right]^\nu = -2^I. \quad (25)$$

By solving this equation for any value of the first segment slope  $D_1$  (non-positive by definition), one obtains the  $\nu$ -value for the contact point of two linear segments,

$$\tau = (1 - \nu)/D_1. \quad (26)$$

The first of these curves has the second segment coinciding with the abscissa axis. This means that one can obtain the same information score by concentrating all the points in the  $2^{-I_0} = 0.194$  ‘active’ part of the region. However, though the  $I$ -value for such a pattern would be 2.36 bits, all points would have the same value of the probability gain. Hence, for such a likelihood value distribution, the variance and higher-order moments would be zero: very different from the actual probability gain pattern (Table 1).

In this two-segment curve we presume that there are no events at the non-active part of the region, thus the  $\epsilon$ -value (1) is zero. Such an arrangement is possible only if the model is assumed to be correct, because if even one event occurs in such a subregion, the value of the likelihood score would immediately be  $-\infty$ . This is why in the real-life situation, where an earthquake occurrence model is not known exactly, we need to adopt a non-zero  $\epsilon$ -value (see Eq. 1).

If we modify the two-segment model to distribute the events with different non-zero densities over the both sub-areas, the variance and the other moments would be also non-zero. In Fig. 9 we show the dependence of the lower-order moments for the likelihood score on the  $D_1$  slope. We calculate the moments for each two-segment model, using a modification of (19). For example

$$\sigma = \sqrt{(1 - \nu) \left[ \log_2 \left( \frac{1 - \nu}{\tau} \right) - I_0 \right]^2 + \nu \left[ \log_2 \left( \frac{\nu}{1 - \tau} \right) - I_0 \right]^2}, \quad (27)$$

where  $\nu$  and  $\tau$  are defined in (25-26) and  $I_0$  for this case can be calculated as

$$I_0 = (1 - \nu) \log_2 \left( \frac{1 - \nu}{\tau} \right) + \nu \log_2 \left( \frac{\nu}{1 - \tau} \right). \quad (28)$$

The higher-order moments (19), as well as the skewness (21) and the kurtosis (22) are computed similarly.

For  $D_1 = -2 \times 2^I$  (dashed magenta line, fifth curve from the bottom) the 2nd, 3rd, and 4th moments correspond roughly to the moments of the forecasted densities. Thus, such a two-segment model would reasonably well approximate the actual event distribution.

The contact coordinates of two segments for this curve are:  $\nu_5 = 0.1732$  and  $\tau_5 = 0.0803$ . Therefore, the point pattern having apparently the same lower-order moments as the actual earthquake forecast would have about 83% of points concentrated in 8% of the area, i.e., the point density will be 10.3 times higher than the uniform Poisson rate. The rest of the events

would be distributed in 92% of the area and have the rate of 0.19 compared to the uniform Poisson distribution. As we mention in Section 2, in our Pacific forecast 0.01 part of the total earthquake rate is spread over the entire region (see Eq. 1 and Figs. 1-2).

The equality of two scores, one based on the ED ( $I_0$ ) and other on the likelihood ratio ( $\langle I_3 \rangle$ ), shown in Table 3, demonstrates again that the ED is a more complete description of the point process. One can obtain the information score from the ED, but not *vice versa*: as we see from Fig. 8 even for two-segment diagrams there is by their construction an infinite number of ED curves having the same score.

### 4.3 Information score for 1977-2003 CMT and PDE catalogs

ED displays in Figs. 5, 6 are inconvenient since the most interesting parts of the curves are concentrated near  $\nu$ - and  $\tau$ -axes. The reason for this feature is that seismicity is concentrated in relatively narrow seismic belts having a fractal spatial earthquake distribution. Now we focus on how other curves deviate from the forecasted (red) one. To make these deviations show more prominently, we need to display the curves in a relative abscissa format, using the 1977-2003 forecast density as a template or baseline for the likelihood score calculation.

Fig. 10 shows several curves in a new format; in effect we convert the red curve in Fig. 5 to the diagonal. This is equivalent to calculating the

information scores by using  $\lambda_i$  as a reference density

$$I_m = \frac{1}{n} \sum_{j=1}^n \nu_j \log_2 \frac{\zeta_j}{\lambda_j}, \quad (29)$$

where  $\zeta_i$  is a rate density for all the other point distributions. In effect we displace all the curves horizontally by the abscissa difference between the forecast line and the unskilled forecast (the descending diagonal of the error diagram, *cf.* Fig. 5).

Fig. 10 shows the difference between the forecast curve (red) and the earthquake curve (blue) better than Fig. 5. Fig. 10 also displays the curve for the 1977-2003 CMT catalog. The numbers of events in the cell areas are shown in Table 1, column 8. Also shown is the curve for the PDE catalog (U.S. Geological Survey 2008) for 1968-2006. We obtain  $I_1 = 3.5991$  bits/event for the 1977-2003 CMT catalog and  $I_1 = 2.9789$  bits for the PDE. These values are significantly larger than those forecasted for 2004-2006. Therefore, our forecast predicts better locations of past earthquakes than those of future events. Why this paradox? In our forecast we use a broader smoothing kernel to capture the spread of seismicity with time (Section 2). Kagan & Jackson (1994, p. 13,696) performed two optimizations for the NW Pacific region: one by using the standard leave-one-out method (Silverman 1986) the other by subdividing the catalog into two parts. The standard density estimation method yielded the optimum value of the maximum distance of a smoothing kernel  $R_{max} \simeq 180$  km as compared to the

value  $R_{max} \simeq 350$  km for predicting the second half of the catalog based on the first half, i.e., the leave-one-out procedure greatly under-smoothed the predicted future spatial distribution of earthquakes. A similar explanation is apparently valid for the PDE score value. Helmstetter *et al.* (2007, Table 1) obtained  $G = 7.1$  (significantly higher than the  $G$ -values for predictive algorithms) when the same data were used to build the long-term seismicity model and to test it (see Section 4.1).

In Fig. 10 we also show several curves for the simulated earthquakes. These curves explain why the  $I_4$ -value (24) is significantly larger than the other measures of the information score. The reason is twofold. First, the number of events in the 3-year interval is relatively small and the curves often fluctuate around the expected value (the red curve). These fluctuations increase the sum value in (24). The curves are often below the red forecast line, which would usually cause the score value to increase. Second, the ED curve should be concave (Molchan 1997; 2003).  $I_4$ -values, listed in Table 3, are calculated with the original curves shown in Figs. 5, 6 which have many convex sections. If we make a lower envelope of the curve points, this would decrease the  $I_4$ -value. However, our numerical experiments show that the decrease is not significant enough to bring the value sufficiently close to the  $I_0$  score.

The fluctuations of the synthetic curves also suggest that some strategies proposed to measure the performance of a prediction algorithm by consider-

ing the ED, like a sum of errors ( $\nu + \tau$ ) or minimax errors (Molchan 1991; Molchan & Kagan 1992; Kossobokov 2006; Molchan & Keilis-Borok 2008) are biased for a small number of forecasted events. For western Pacific regions the number of predicted events ( $n_j$ ) is relatively large; in many other applications of the ED (*ibid.*) this number is less than 10.

In Fig. 10 the forecast distribution curve is used as the template or the reference model (Zechar & Jordan 2008). Thus, we can measure the difference between this line and the other curves using many standard statistical techniques, like the Kolmogorov-Smirnov test, the Cramer-von Mises test, etc., (Stephens 1974) to infer whether these distributions are statistically different.

## 5 Discussion

Several information scores are displayed in Table 3. Although these scores appear different, the difference is caused either by the small event number or a small number of simulations. The following limits can be easily conjectured

$$I_0 = \lim_{N \rightarrow \infty} \langle I_3 \rangle, \quad (30)$$

(see Eq. 10). In Table 3 the difference between these two scores is small due to the large number of simulations. Similarly,

$$I = \lim_{|S_i| \rightarrow 0} I_0, \quad \text{or} \quad I = \lim_{N \rightarrow \infty} I_0, \quad (31)$$

(*cf.* Eq. 14). Also

$$I_1 = \lim_{|S_i| \rightarrow 0} I_2, \quad (32)$$

(see Eqs. 7, 8).

In addition, if the model of the long-term forecast is correct, then

$$I_1 = \lim_{n \rightarrow \infty} I, \quad \text{and} \quad I_4 = \lim_{n \rightarrow \infty} I, \quad (33)$$

(see Eqs. 7, 24).

As Table 3 demonstrates, the convergence of the  $I_4$  score, shown in (33) is slow, even as tens or hundreds of events have been forecasted in our examples. In many considerations and applications of error diagrams (see Section 1), behavior of the curves is considered only for a few observed events, thus random fluctuations of the observed distribution *vs.* the theoretical distribution should be much larger.

In this paper we wanted to extend statistical analysis of the stochastic point processes on line (usually time) to multidimensional space. In particular, we wished to find the relation between two widely used statistical measures of prediction efficiency: likelihood scores and error diagrams. The equations derived here can be easily transformed to describe quantitative connection between the information scores and concentration diagrams (Section 1).

Summarizing our results, we list the following major points:

- 1. As with temporal stochastic processes (Kagan 2007b), we find forward

and inverse relations between the information score and the error diagram curve for point spatial fields. The error diagram represents a much more complete picture of the stochastic point process than does likelihood analysis.

- 2. Since we are using a Poisson process to represent the long-term spatial point pattern, the resulting models are easier to visualize and calculate. However, the assumption of earthquake statistical independence and its influence on the information score value both need to be investigated.
- 3. We extend our analysis for relatively small samples of events and show that for such samples we should modify some of the testing criteria proposed for error diagrams.
- 4. We show that the forecasting blueprint for estimating future earthquake point density differs from standard methods of statistical density evaluation. Nevertheless, the connection between the likelihood score and error diagrams described above can be used in many density estimation problems.
- 5. We show that for testing the long-term forecast, it is sufficient to process the forecast table to obtain the error diagram and most information scores. Thus, the simulation which was used in previous work, and which requires significant computational resources, can be avoided in most cases (Rhoades 2008).
- 6. In this paper we mostly considered relations between the theoretical EDs and likelihood scores (Eqs. 12 and 13) under the assumptions that the

number of events is large and the theoretical model of the stochastic point process is true. Thus, the derived relations are mathematical expectations. Although some theoretical analysis has been previously published for the case when the above assumptions have been violated (see references in Section 1), a rigorous consideration of statistical behavior for both measures of prediction efficiency would require an effort well beyond the bounds of the present study.

## **Acknowledgments**

I appreciate partial support from the National Science Foundation through grants EAR 04-09890, EAR-0711515, and DMS-0306526, as well as from the Southern California Earthquake Center (SCEC). SCEC is funded by NSF Cooperative Agreement EAR-0106924 and USGS Cooperative Agreement 02HQAG0008. I thank David D. Jackson of UCLA and David Rhoades of GNS Science, New Zealand for very useful discussions. Comments by Warner Marzocchi, by two anonymous reviewers and by the Associate Editor Massimo Cocco have been very helpful in revising the manuscript. I am grateful to Kathleen Jackson for significant improvements in the text. Publication 1252, SCEC.

## References

- [1] Abramowitz, M. & Stegun, I. A., (Eds.), 1972. *Handbook of Mathematical Functions*, Dover, NY, pp. 1046.
- [2] Baddeley, A., 2007. Spatial point processes and their applications, *Lecture Notes in Mathematics*, Springer-Verlag, **1892**, pp. 1-75.
- [3] Baddeley, A., Turner, R., Møller, J. & Hazelton, M., 2005. Residual analysis for spatial point processes, *J. Royal Stat. Soc.*, **B67**(5), 617-651.
- [4] Console, R., Rhoades, D.A., Murru, M., Evison, F.F., Papadimitriou, E.E., and Karakostas, V.G., 2006. Comparative performance of time-invariant, long-range and short-range forecasting models on the earthquake catalogue of Greece, *J. geophys. Res.*, **111**, B09304, doi:10.1029/2005JB004113.
- [5] Daley, D. J. & Vere-Jones, D., 2003. *An Introduction to the Theory of Point Processes*, Springer-Verlag, New York, 2-nd ed., Vol. 1, pp. 469.
- [6] Daley, D. J. & Vere-Jones, D., 2008. *An Introduction to the Theory of Point Processes*, Springer-Verlag, New York, 2-nd ed., Vol. 2, pp. 573.
- [7] DelSole, T. & M. K. Tippett, 2007. Predictability: Recent insights from information theory, *Rev. Geophys.*, **45**, RG4002, doi:10.1029/2006RG000202.

- [8] Ekström, G., A. M. Dziewonski, N. N. Maternovskaya & M. Nettles, 2005. Global seismicity of 2003: Centroid-moment-tensor solutions for 1087 earthquakes, *Phys. Earth planet. Inter.*, **148**(2-4), 327-351.
- [9] Helmstetter, A., Y. Y. Kagan & D. D. Jackson, 2007. High-resolution time-independent grid-based forecast for  $M \geq 5$  earthquakes in California, *Seism. Res. Lett.*, **78**(1), 78-86.
- [10] Jolliffe, I. T. & D. B. Stephenson, Eds., 2003. *Forecast Verification: a Practitioner's Guide in Atmospheric Science*, J. Wiley, Chichester, England, 240 pp.
- [11] Kagan, Y. Y., 1991. Likelihood analysis of earthquake catalogues, *Geophys. J. Int.*, **106**(1), 135-148.
- [12] Kagan, Y. Y., 2007a. Earthquake spatial distribution: the correlation dimension, *Geophys. J. Int.*, **168**(3), 1175-1194.
- [13] Kagan, Y. Y., 2007b. On earthquake predictability measurement: information score and error diagram, *Pure appl. Geoph.*, **164**(10), 1947-1962. DOI: 10.1007/s00024-007-0260-1
- [14] Kagan, Y. Y. & D. D. Jackson, 1991. Long-term earthquake clustering, *Geophys. J. Int.*, **104**(1), 117-133.
- [15] Kagan, Y. Y. & Jackson, D. D., 1994. Long-term probabilistic forecasting of earthquakes, *J. geophys. Res.*, **99**(B7), 13,685-13,700.

- [16] Kagan, Y. Y. & D. D. Jackson, 1998. Spatial aftershock distribution: Effect of normal stress, *J. geophys. Res.*, **103**(B10), 24,453-24,467.
- [17] Kagan, Y. Y. & D. D. Jackson, 2000. Probabilistic forecasting of earthquakes, *Geophys. J. Int.*, **143**(2), 438-453.
- [18] Kagan, Y. Y. & D. D. Jackson, 2006. Comment on ‘Testing earthquake prediction methods: “The West Pacific short-term forecast of earthquakes with magnitude  $M_{wHRV} \geq 5.8$ ”’ by V. G. Kossobokov, *Tectonophysics*, **413**(1-2), 33-38, 10.1016/j.tecto.2005.10.007.
- [19] Kagan, Y. & L. Knopoff 1977. Earthquake risk prediction as a stochastic process, *Phys. Earth planet. Inter.*, **14**(2), 97-108.
- [20] Kagan, Y. Y., Y. F. Rong & D. D. Jackson, 2003. Probabilistic forecasting of seismicity, Chapter 5.2 in “*EARTHQUAKE SCIENCE AND SEISMIC RISK REDUCTION*”, NATO Science Series IV: Earth and Environmental Sciences, vol. 32, Eds. F. Mulargia & R. J. Geller, pp. 185-200, Kluwer, Dordrecht, The Netherlands.
- [21] Kossobokov, V. G., 2006. Testing earthquake prediction methods: “The West Pacific short-term forecast of earthquakes with magnitude  $M_{wHRV} \geq 5.8$ ”, *Tectonophysics*, **413**(1-2), 25-31.

- [22] Mason, I. B., 2003. Binary events, in *Forecast Verification: a Practitioner's Guide in Atmospheric Science*, I. T. Jolliffe & D. B. Stephenson, Eds., pp. 37-76, J. Wiley, Chichester, England.
- [23] Molchan, G. M., 1990. Strategies in strong earthquake prediction, *Phys. Earth planet. Inter.*, **61**(1-2), 84-98.
- [24] Molchan, G. M., 1991. Structure of optimal strategies in earthquake prediction, *Tectonophysics*, **193**(4), 267-276.
- [25] Molchan, G. M., 1997. Earthquake prediction as a decision-making problem, *Pure appl. Geophys.*, **149**(1), 233-247.
- [26] Molchan, G. M., 2003. Earthquake prediction strategies: A theoretical analysis, In: Keilis-Borok, V. I. & A. A. Soloviev, (Eds) *Nonlinear Dynamics of the Lithosphere and Earthquake Prediction*, Springer, Heidelberg, 208-237.
- [27] Molchan, G., 2008. Space-Time Earthquake Prediction: the Error Diagrams, <http://arxiv.org/abs/0807.2533v1>.
- [28] Molchan, G. M. & Y. Y. Kagan, 1992. Earthquake prediction and its optimization, *J. geophys. Res.*, **97**(B4), 4823-4838.
- [29] Molchan, G. & V. Keilis-Borok, 2008. Earthquake Prediction: Probabilistic Aspect, *Geophys. J. Int.*, **173**(3), 1012-1017, doi: 10.1111/j.1365-246X.2008.03785.x

- [30] Palmer, T. N. & R. Hagedorn, 2006. *Predictability of Weather and Climate*, 702 pp., Cambridge Univ. Press, New York.
- [31] Rhoades, D. A., 2007. Application of the EEPAS Model to Forecasting Earthquakes of Moderate Magnitude in Southern California, *Seism. Res. Lett.*, **78**(1), 110-115.
- [32] Rhoades, D., 2008. Is there a less computationally intensive way to test forecasting models? <http://us.cseptest.org/WGTMeeting05022008?action=AttachFile&do=get&target=rhoades.pdf>
- [33] Rhoades, D. A., & F. F. Evison, 2005. Test of the EEPAS forecasting model on the Japan earthquake catalogue, *Pure appl. Geophys.*, **162**(6-7), 271-1290.
- [34] Rhoades, D. A., & F. F. Evison, 2006. The EEPAS forecasting model and the probability of moderate-to-large earthquakes in central Japan, *Tectonophysics*, **417**(1-2), 119-130.
- [35] Rong, Y.-F. & Jackson, D. D., 2002. Earthquake potential in and around China: Estimated from past earthquakes, *Geophys. Res. Lett.*, **29**(16): art. no. 1780, DOI: 10.1029/2002GL015297.

- [36] Rong, Y.-F., D. D. Jackson and Y. Y. Kagan, 2003. Seismic gaps and earthquakes, *J. geophys. Res.*, **108**(B10), 2471, **ESE-6**, pp. 1-14, doi:10.1029/2002JB002334.
- [37] Schorlemmer, D., M. C. Gerstenberger, S. Wiemer, D. D. Jackson & D. A. Rhoades, 2007. Earthquake likelihood model testing, *Seism. Res. Lett.*, **78**(1), 17-29.
- [38] Shen, Z.-K., D. D. Jackson & Y. Y. Kagan, 2007. Implications of geodetic strain rate for future earthquakes, with a five-year forecast of M5 earthquakes in southern California, *Seism. Res. Lett.*, **78**(1), 116-120.
- [39] Silverman, B. W. 1986. *Density Estimation for Statistics and Data Analysis*, London, Chapman and Hall, 175 pp.
- [40] Stephens, M. A. 1974. EDF statistics for goodness of fit and some comparisons, *J. Amer. Statist. Assoc. (JASA)*, **69**, 730-737.
- [41] Swets, J. A. 1973. The Relative Operating Characteristic in Psychology, *Science*, **182**(4116), 990-1000.
- [42] U.S. Geological Survey, *Preliminary determination of epicenters (PDE)*, 2008. U.S. Dep. of Inter., Natl. Earthquake Inf. Cent., <http://neic.usgs.gov/neis/epic/epic.html>
- [43] Vere-Jones, D., 1998. Probabilities and information gain for earthquake forecasting, *Computational Seismology*, **30**, Geos, Moscow, 248-263.

- [44] Zaliapin, I. & Molchan, G., 2004. Tossing the Earth: How to Reliably Test Earthquake Prediction Methods, *Eos Trans. AGU*, **85**(47), Fall Meet. Suppl., Abstract S23A-0302.
- [45] Zechar, J. D. & T. H. Jordan, 2008. Testing alarm-based earthquake predictions, *Geophys. J. Int.*, **172**(2), 715-724, doi:10.1111/j.1365-246X.2007.03676.x

Lat. $\theta$	Long. $\phi$	EQ Rate ( $\lambda_i$ ) EQ/(day $\times$ km <sup>2</sup> )	1° $\times$ 1° ( $S_i$ ) km <sup>2</sup>	Cell Rate ( $R_i$ ) EQ/[yr $\times$ (.5°) <sup>2</sup> ]	EQs 04-06	Probab. Gain ( $G_i$ )	EQs 77-03
1	2	3	4	5	6	7	8
7.0	127.0	1.7909E-07	12254	0.2003970	0	68.3700	6
40.5	143.0	1.6302E-07	9388	0.1397471	0	62.2331	1
45.5	151.5	1.5134E-07	8653	0.1195765	0	57.7738	5
24.0	122.0	1.4759E-07	11278	0.1519959	1	56.3445	10
44.5	150.0	1.4496E-07	8805	0.1165462	0	55.3376	8
44.5	149.5	1.4252E-07	8805	0.1145830	0	54.4055	9
12.5	125.5	1.4152E-07	12053	0.1557541	0	54.0252	6
44.0	148.5	1.4150E-07	8881	0.1147490	0	54.0181	8
....	....	....	....	....	....	....	....
48.5	128.0	2.6198E-11	8180	0.0000196	0	0.0100	0
48.5	127.5	2.6189E-11	8180	0.0000196	0	0.0100	0
48.5	127.0	2.6183E-11	8180	0.0000196	0	0.0100	0

Table 1: Beginning and end of earthquake rate forecast table for NW-Pacific based on CMT catalog (Ekström *et al.* 2005) for 1977-2003 and ordered by descending rate density (column 3,  $\lambda$ ). Cells are  $0.5^\circ \times 0.5^\circ$ , they form a  $121 \times 121$  grid, EQ – earthquake(s). The probability gains ( $G_i$ ) for displayed cells are ratio of the rate density ( $\lambda_i$ ) to the Poisson rate density equal to  $2.6289 \times 10^{-9}$  eq/(day  $\times$  km<sup>2</sup>) (see Eq. 6).

	Time Interval	Pacific Regions	
		NW	SW
		Annual Rate	
$v_0$	77-03 Forecast	35.7159	60.7509
$n_1$	77-03 EQs	968	1591
$v_1$	77-03 EQs	35.8546	58.9304
$n_2$	04-06 EQs	108	170
$v_2$	04-06 EQs	35.9918	56.6537

Table 2: Annual earthquake rate ( $v_i$ ) estimates. Actual rate calculations are made with time intervals measured in days (9861 days in 1977-2003 and 1096 days in 2004-2006). For display convenience, we convert the daily rates into annual rates by multiplying them by 365.25. EQs – earthquakes.

#	Info Score	Pacific Regions ( $n_j$ )	
		NW (108)	SW (170)
1	$I_0$	2.3645	3.3772
2	$I_1$	2.3675	3.0506
3	$I_2$	2.4067	3.2777
4	$\langle I_3 \rangle$	2.3609	3.3768
5	$I_4$	3.0970	3.9823
6	$I_0 - I_1$	-0.0030	0.3266
7	$I_0 - I_2$	-0.0422	0.1002
8	$\sigma$	2.2102	2.9720
9	$\eta$	-0.6574	-0.2264
10	$\psi$	0.3685	-0.3078
11	$\sigma_n$	0.2151	0.2296
12	$\eta_n$	-0.0196	0.0078
13	$\psi_n$	-0.0369	0.0441

Table 3: Information scores for one event in west Pacific regions, the standard error ( $\sigma$ ) and coefficients of skewness ( $\eta$ ) and kurtosis ( $\psi$ ) (20-22). The numbers in parentheses are event counts in 2004-2006 for each region,  $n_j$ . Variables  $\sigma_n$ ,  $\eta_n$ , and  $\psi_n$  are for the set of  $n_j$  events.

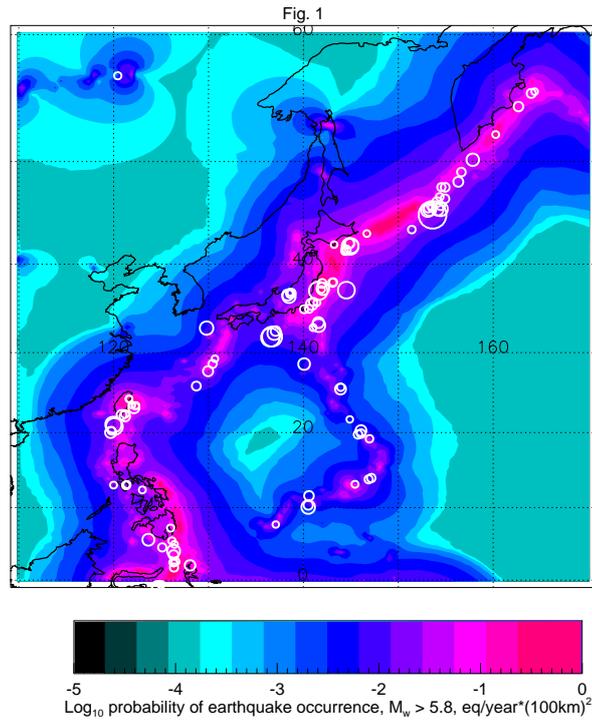


Figure 1: NW-Pacific long-term seismicity forecast. The forecast is calculated at  $121 \times 121$  grid. Colour tones show the rate density of shallow (depth less or equal to 70 km) earthquake occurrence calculated using the CMT 1977-2003 catalog; 108 earthquakes for 2004-2006 are shown in white. The uniform background rate density ( $\epsilon = 0.01$ , see Eq. 1) can be observed at northwest and southeast corners of the map as greyish-green areas.

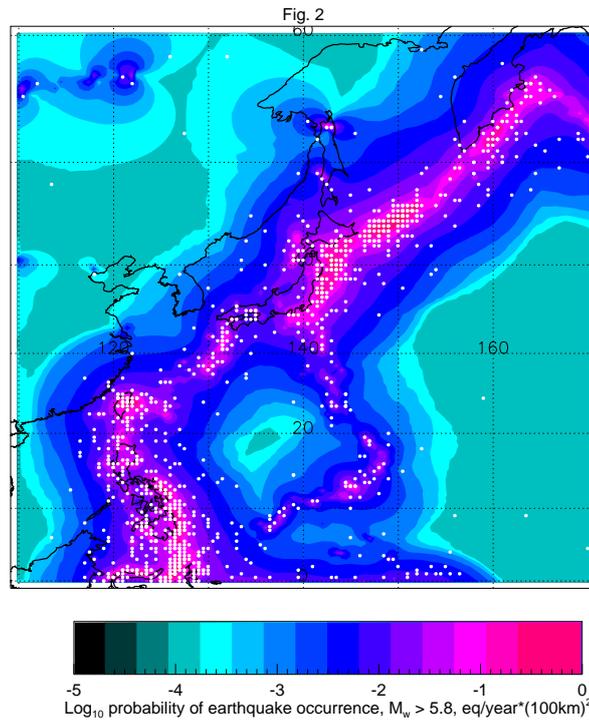


Figure 2: NW-Pacific long-term seismicity forecast. Colour tones show the rate density of earthquake occurrence calculated using the CMT 1977-2003 catalog; 1080 simulated earthquakes for 2004-2006 are shown in white.

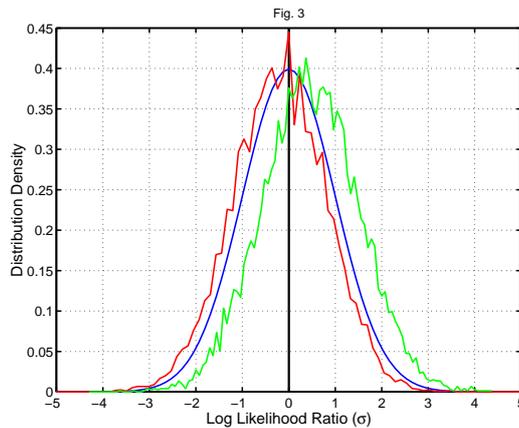


Figure 3: Histograms of the log-likelihood function differences for 2004-2006 simulated earthquakes (see Fig. 2). The functions are normalized to have a unit standard deviation. We simulate 10,000 sets of 108 events for the NW-Pacific and of 170 events for the SW-Pacific. The blue line is the Gaussian curve with a zero mean and unit standard deviation. Red curve corresponds to simulation distributions for NW-Pacific; green curve to SW-Pacific. Curves on the right from the Gaussian curve correspond to simulations that are worse than a real earthquake distribution; curves on the left correspond to simulations that are better than a real earthquake distribution.

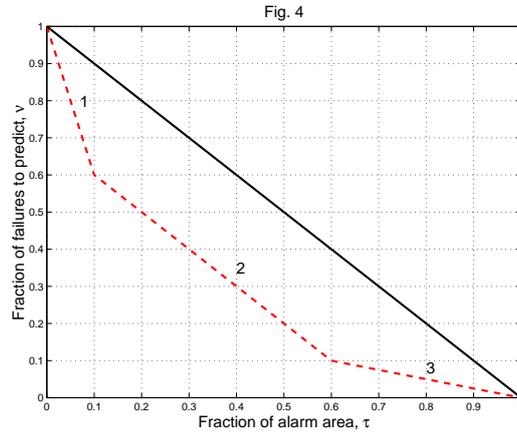


Figure 4: Error diagram  $(\tau, \nu)$  schematic example (see Eq. 17). Two ED plots are shown: 1) for the uniform Poisson process, corresponding to the diagonal of the square, and 2) for a region consisting of three sub-areas with different occurrence rates in each (see text).

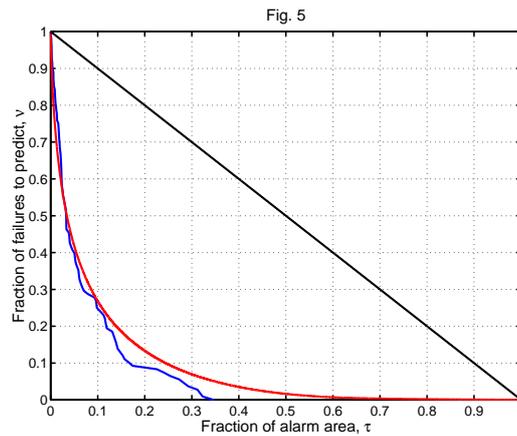


Figure 5: Error diagram  $(\tau, \nu)$  for NW-Pacific long-term seismicity forecast. The forecast is calculated at  $121 \times 121$  grid. Solid black line – the strategy of random guess, red line – the ordered density for long-term forecast, blue line – earthquakes in 2004-2006.

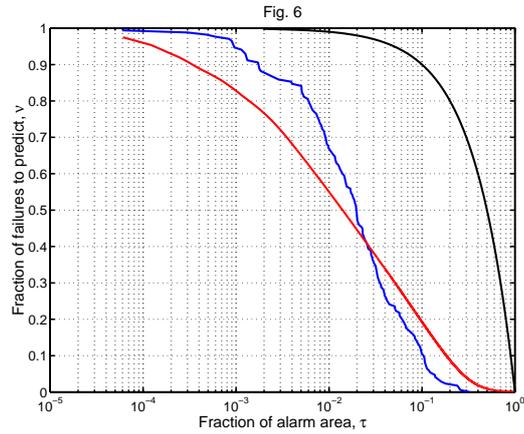


Figure 6: Error diagram  $(\tau, \nu)$  for SW-Pacific long-term seismicity forecast. The forecast is calculated at  $121 \times 161$  grid. Solid black line – the strategy of random guess, red line – the ordered density for long-term forecast, blue line – earthquakes in 2004-2006.

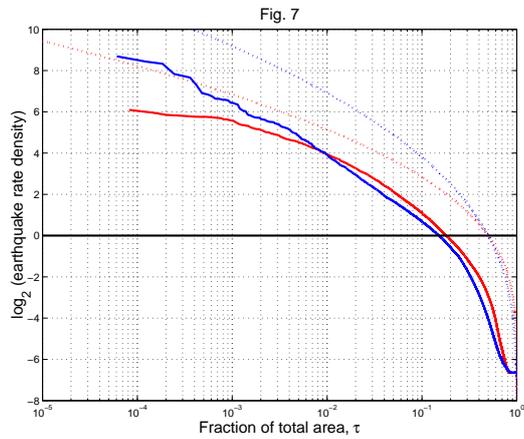


Figure 7: Likelihood function dependence on the fraction of the total area,  $\tau$ , for west Pacific long-term seismicity forecasts: solid red curve – NW-Pacific, blue curve – SW-Pacific (see boundaries in Section 2). Dotted lines are for the Gaussian distribution of log forecast rates: red – NW-Pacific, blue – SW-Pacific. Thick black line shows the log-likelihood distribution if all cells are assumed to have a uniform Poisson rate.

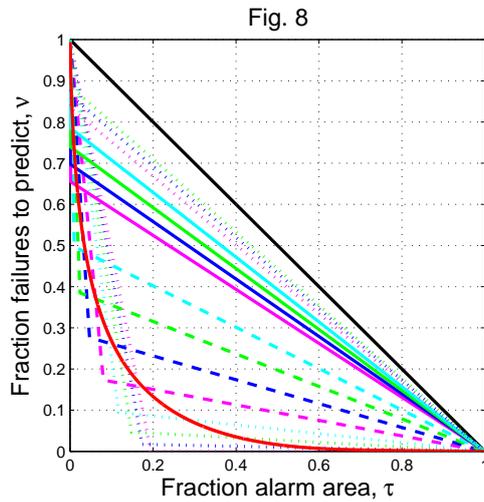


Figure 8: Error diagram  $(\tau, \nu)$  for NW-Pacific long-term seismicity forecast, approximated by two-segment distributions. The solid thick black straight line corresponds to a random guess, the thick red solid line is for the NW forecast. Thin two-segment solid lines are for the curves with the information score  $I_0 = 2.3645$  bits. The slope  $D_1$  for the right-hand first segment is  $D_1 = -2^{I_0}$ . For the next first segments slopes are  $D_1 \times 1.1$ ,  $D_1 \times 1.25$ ,  $D_1 \times 1.5$ ,  $D_1 \times 2.0$ ,  $D_1 \times 3.0$ ,  $D_1 \times 5.0$ ,  $D_1 \times 10$ ,  $D_1 \times 50$ ,  $D_1 \times 100$ ,  $D_1 \times 250$ ,  $D_1 \times 1000$ ,  $D_1 \times 10,000$ ,  $D_1 \times 100,000$ , and  $D_1 \times 1000,000$ .

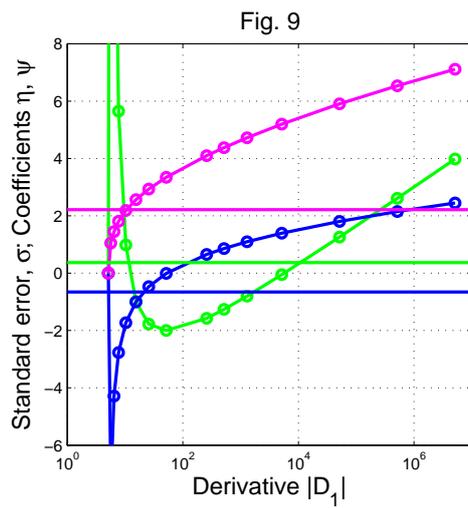


Figure 9: Dependence on slope  $|D_1|$  of standard deviation (magenta), coefficients of skewness (blue) and kurtosis (green) for two-segment curves in Fig. 8. Horizontal lines are these variables for the red curve in the cited plot.

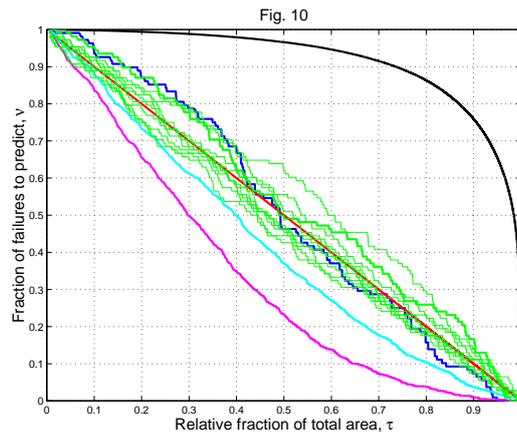


Figure 10: Error diagram  $(\tau, \nu)$  for NW-Pacific long-term seismicity forecast. Solid black line – the strategy of random guess. The solid thick red diagonal line is a curve for the NW forecast. Blue line is earthquake distribution from the CMT catalog in 2004-2006 (forecast); magenta line corresponds to earthquake distribution from the CMT catalog in 1977-2003; cyan line is earthquake distribution from the PDE catalog in 1968-2006. Thin green lines are ten simulations, displayed in Fig. 2, the first realization is shown by a thick green line.