Likelihood analysis of earthquake focal mechanism distributions

Yan Y. Kagan and David D. Jackson

Department of Earth, Planetary, and Space Sciences (EPSS), Geology Bldg, 595 Charles E. Young Dr, University of California, Los Angeles (UCLA), Los Angeles, CA 90095-1567, USA. E-mail: kagan@moho.ess.ucla.edu

Accepted 2015 February 19. Received 2015 February 18; in original form 2014 September 5

SUMMARY

In our paper published earlier we discussed forecasts of earthquake focal mechanism and ways to test the forecast efficiency. Several verification methods were proposed, but they were based on ad hoc, empirical assumptions, thus their performance is questionable. We apply a conventional likelihood method to measure the skill of earthquake focal mechanism orientation forecasts. The advantage of such an approach is that earthquake rate prediction can be adequately combined with focal mechanism forecast, if both are based on the likelihood scores, resulting in a general forecast optimization. We measure the difference between two double-couple sources as the minimum rotation angle that transforms one into the other. We measure the uncertainty of a focal mechanism forecast (the variability), and the difference between observed and forecasted orientations (the prediction error), in terms of these minimum rotation angles. To calculate the likelihood score we need to compare actual forecasts or occurrences of predicted events with the null hypothesis that the mechanism’s 3-D orientation is random (or equally probable). For 3-D rotation the random rotation angle distribution is not uniform. To better understand the resulting complexities, we calculate the information (likelihood) score for two theoretical rotational distributions (Cauchy and von Mises-Fisher), which are used to approximate earthquake source orientation pattern. We then calculate the likelihood score for earthquake source forecasts and for their validation by future seismicity data. Several issues need to be explored when analyzing observational results: their dependence on forecast and data resolution, internal dependence of scores on forecasted angle and random variability of likelihood scores. Here, we propose a simple tentative solution but extensive theoretical and statistical analysis is needed.

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

1 INTRODUCTION

Kagan & Jackson (2014, hereinafter abbreviated as KJ2014) discussed two problems: forecasting earthquake focal mechanisms and evaluating a forecast skill. The first problem was initially addressed by Kagan & Jackson (1994), but no attempt to verify the forecast has been carried out until now. KJ2014 proposed several verification methods, but the techniques were based on ad hoc, empirical assumptions, thus their performance was not clear. In this work we apply a more conventional likelihood method to measure the skill of forecast.

The likelihood estimate for the focal mechanism prediction compares actual forecasts or later occurrences of predicted events with the null hypothesis that the mechanism’s orientation is equally probable or random. It is similar to our forecast testing for long- or short-term earthquake rate predictions (Kagan & Jackson 1994, 2011), where we use a Poisson process uniform in space or time as the null hypothesis. For 3-D rotation the random probability distribution function (PDF) for the rotation angle (Φ) is not uniform in the relevant parameters. Kendall & Moran (1963, chapters 4.25–4.30), for example, show that rotating an object around an axis selected uniformly from all possible axes in space and using a uniformly distributed angle yields no truly (or uniform) random 3-D rotation. With the earthquake source orientation distribution, the situation is even more complex because of the double-couple (DC) source symmetry; though the PDF analytical expression is known: see Kagan (1990, eq. 3.1) or Table 3 for orthorhombic symmetry by Grimmer (1979). See also eqs (3)–(6) below.

Kagan & Jackson (1994, 2011) forecast earthquake rate by smoothing the past seismicity record with a spatial kernel. They optimize the kernel by searching for the best prediction of future seismicity based on past earthquakes (see also Molchan 2012). The assumption is that the true forecast model belongs to a general class of parametric models which is used to approximate seismicity.
Present we adjust only the width of the smoothing kernel and its functional form. However, several other kernel parameters (such as its directivity, magnitude dependence, etc.) can be optimized as well (see, for instance, Kagan & Jackson 2011).

Likelihood scores can be calculated to measure the forecast skill (Vere-Jones 1998; Daley & Vere-Jones 2004; Harte & Vere-Jones 2005; Kagan & Jackson 2011). In the optimized forecast Kagan (2009, his table 3) calculated several likelihood scores ($I_0$, $I_1$, $I_2$; and $I_3$). We compare the forecasted earthquake spatial rate density with an ignorant model that has only a single Poisson global earthquake rate. The score $I_0$ measures the potential efficiency of a forecast or its ‘specificity’ (Bird & Kreemer 2015). Scores from $I_1$ to $I_3$ measure the forecast ‘success’ (ibid.), that is how well the future events were predicted by the forecast model. Most of these scores have similar values, their difference is due to the forecast map resolution and other known and controlled factors. This feature is explained by our spatial smoothing forecast procedure which yields a relatively homogeneous seismicity map. The forecast is generally based on the probability density estimation technique (Silverman 1986): future earthquakes are concentrated close to the location of the past events, thus our general forecast model is obtained by smoothing the past earthquake locations to infer the spatial distribution of future events.

A major problem with the focal mechanism forecast is that we lack a model for earthquake source pattern similar to that for earthquake rate. Thus, our forecast distribution contains many relatively sharp steps, and if integrated, they produce the $I_4$ score (Kagan 2009, table 3), which yields a consistent estimate that converges to the true value only when the sample size tends to infinity (Kagan 2007; Molchan 2010). Therefore, we use various approaches to determine the properties of the $I_4$ score for the earthquake catalogues under investigation. We hope that this new measure of focal mechanism forecast skill will be useful in earthquake prediction efforts.

2 Statistical Focal Mechanism Forecasts

KJ2014 have constructed (see its Fig. 1) daily worldwide long- and short-term earthquake forecasts based on the Global Centroid-Moment-Tensor (GCMT) catalogue (Ekström et al. 2012, and its references). The GCMT (1977–2012) completeness threshold corresponds to scalar moment $M_0 = 10^{17.7}$ Nm, or moment magnitude $m_w = 5.8$; only earthquakes above the threshold are used in the analysis (KJ2014). The forecasts specify the earthquake rate per unit area, time and magnitude with the spatial resolution to 0.1 degree and the latitude range from pole to pole. The focal mechanism orientation is also computed in the forecast. We limit this study to the latitude bandwidth [75°S–75°N] to reduce computation time. Moreover, almost all earthquakes in the GCMT catalogue are concentrated there (KJ2014, fig. 2).

In KJ2014 paper, we analysed the focal mechanism orientation distribution in the GCMT catalogue. Though the GCMT provides a general description of the earthquake source as a symmetric deviatoric moment tensor, in our studies we used only the DC solution. The orientation of the DC source ($\hat{O}$) can be expressed by several largely equivalent methods: seismic moment tensor, fault-plane solutions, principal axes ($T$, $P$, $B$-axes) and quaternions (Kagan 1991). Because of the DC symmetry, four rotations with an angle less than or equal to 180° are needed to transform one DC into another. We select the smallest rotation angle ($\Phi$) as the representative (ibid.).

As described in Kagan & Jackson (1994) and in KJ2014 (see its eqs 2–5), we estimate the focal mechanism orientations of future earthquakes using a distance-dependent weighted combination of focal mechanisms of past earthquakes. For seismicity rate we optimize the parameters of the combination (e.g. width of the smoothing kernel) as follows: we divide the available GCMT earthquake catalogue into a learning or training set (here 1977–2007 inclusive) and a test or validation set (2008–2012), and select those parameters that best predict the earthquake rate in the test set using only data from the learning set. Because of difficulties of measuring focal mechanism forecast efficiency for earthquake source we did not optimize our prediction. As we explained in Section 1, our aim in KJ2014 paper and in this work is to develop methods of evaluating the forecast skill and employ it in future work.

For shallow (depth 0–70 km) earthquakes the forecasted focal mechanism can be calculated at any point in seismically active areas on Earth; for efficiency we can calculate at the centres of cells spaced uniformly in latitude and longitude. Our tasks here are to estimate the uncertainty of the forecast and evaluate its predictive performance. For both tasks, we employ the minimum rotation angle $\Phi$ between pairs of orientations as described in the previous paragraph. The uncertainty estimate is based on the variability of the data that go into the forecast,

$$\Phi = \frac{1}{N} \sum_{i=1}^{N} \psi (O_i, \hat{O}_i),$$

where $O_i$ is the predicted focal mechanism orientation at a cell centre (as shown in eqs 2–5 of KJ2014), $\hat{O}_i$ is the orientation of each event in the learning set within 1000 km of the cell centre, the operator $\psi (O_i, \hat{O}_i) = \Phi_i$ determines a minimum rotation angle between two DC orientations, and $N$ is the number of events for this cell in the training set (1977–2007). We need to compute the variability only at test points, that is the centres of cells containing an earthquake in the test set. Note that at any location $\Phi$ depends
only on the information in the learning set and will be small if the focal mechanisms near the test point are very similar. The angle \( \Phi_1 \) is non-negative, and its empirical distribution is asymmetrical and peaked near zero (see Fig. 1 below). We employed cell sizes of 0.5 and 0.1 degrees, and as shown in Table 4 below the choice made little difference.

As a measure of prediction error we use the minimum disorientation angle \( \Phi_2 \) between the predicted and observed focal mechanisms for each event in the test set 2008–2012.

\[
\Phi_2 = \Psi(O_b, O_f),
\]

(2)

where \( O_b \) is the earthquake orientation in a forecasted cell, and \( O_f \) is the orientation of an earthquake which occurred in this cell during 2008–2012.

In Fig. 1 we display a scatterplot of two angles \( \Phi_1 \) and \( \Phi_2 \). To investigate the interdependence of two angles, we subdivide the plot of 1069 \( \Phi_1 \) angles (i.e. the forecast uncertainty) into 10 subsets with an increasing angle and calculate the quantiles of \( \Phi_2 \) distribution in each of these subsets. Generally, the smaller the value of \( \Phi_1 \), or the smaller \( \Phi_2 \) and their variation, the better the forecast. The ideal forecast should have \( \Phi_1 = \Phi_2 = 0 \).

The gridded focal mechanism forecast is displayed in fig. 4 of KJ2014. About 42 percent of the cells in this forecast have variables \( \Phi_1 \) and \( \Phi_2 \) equal to zero. These cell centres do not have any earthquake epicentroid within 1000 km distance and they are shown as empty focal mechanism cells in the above diagram. (We refer to the surface point overlying the (hypo)centroid as the epicentroid). To avoid future ‘surprises’, in our earthquake rate forecasts we assume that 1 percent of all earthquakes occur uniformly over the Globe (Jackson & Kagan 1999; Kagan & Jackson 2011). We did not make such or similar assumptions for the focal mechanism forecast, since it is not known what default value needs to be adopted for these cells (KJ2014).

However, almost all 2008–2012 earthquakes occurred in the places within 1000 km of 1977–2007 epicentroids, thus their angles could be evaluated. Only three events out of 1069 are outside of the 1000 km limit. In Fig. 1 these zero values of \( \Phi_1 \) and \( \Phi_2 \) are all plotted at the point [0.0, 0.0] and thus are not separately visible. In our likelihood studies below we sometimes use only 1066 events for which both angles can be measured.

In Table 1 we list the properties of both angles distribution shown in Fig. 1. The average \( \Phi_2 \) value increases steadily with the increase of \( \Phi_1 \), though the standard deviation is generally stable, thus the coefficient of variation also decreases for later subsets. In fig. 7 by KJ2014 this interdependence of two angles was characterized by regression lines.

3 Rotation Angle Distributions

Kagan (2013, section 5) considers three theoretical statistical distributions for the DC source orientation:

(1) The uniform random rotation, which corresponds to crystallographic orthorhombic symmetry for a general DC source. This distribution is defined for the orientation angle range 0°–120° for the minimum rotation angle \( \Phi \) (see Section 2). This random distribution is important as a potential null hypothesis for comparing it with observational and theoretical rotational distributions. The PDF is

\[
f(\Phi) = (4/\pi)(3 \sin \Phi + 2 \cos \Phi - 2) \quad \text{for} \quad 90^\circ \leq \Phi \leq \Phi_5; \tag{3}
\]

and

\[
f(\Phi) = (4/\pi) \left[ 3 \sin \Phi + 2 \cos \Phi - 2 - (6/\pi) \left( 2 \sin \Phi \arccos \left( \frac{1 + \cos \Phi}{-2 \cos \Phi} \right) \right)^{1/2} - (1 - \cos \Phi) \arccos \left( \frac{1 + \cos \Phi}{-2 \cos \Phi} \right) \right] \quad \text{for} \quad \Phi \leq \Phi_5 \leq 120.0^\circ, \tag{4}
\]

where

\[
\Phi_5 = 2 \arccos (3^{-1/2}) = \arcsin \left( -\frac{1}{3} \right) \approx 109.47^\circ. \tag{5}
\]

(2) Two non-uniform theoretical statistical distributions: the rotational Cauchy law (Kagan 1992, 2013) and the rotational von Mises-Fisher (VMF) distribution (Kagan 2000; Mardia & Jupp 2000, pp. 289–292; Morawiec 2004, pp. 88–89).

Two Cauchy distributions need to be distinguished: the regular Cauchy law in an infinite 3-D Euclidean space and the rotational Cauchy distribution for rotation in the 3-D space (Kagan 1990, 2013). The Cauchy law in the 3-D Euclidean space has a power-law tail for large distances. This Cauchy distribution is especially important for representing earthquake geometry, since theoretical arguments and simulations (Kagan 1990) show that the stress tensor in a medium with uniformly distributed random defects follows this distribution. The regular Cauchy law is a stable distribution (Kagan 2013).

The distribution of rotation angles between neighbouring earthquakes was found out to be well approximated by the rotational Cauchy law (Kagan 1992, 2013). The rotational Cauchy is defined on the 3-D hypersphere of a normalized quaternion (Kagan 1990).

| No. | \( \Phi_1 \) | \( \Phi_2 \) | \( \sigma_\Phi \) | \( C_v \) | Coef. | Score |
|-----|-------------|-------------|----------------|--------|-------|-------|
| 1   | 0.2         | 18.0        | 23.3           | 1.29   | 106   | 6.56  |
| 2   | 6.4         | 17.3        | 21.6           | 1.25   | 106   | 6.47  |
| 3   | 9.5         | 20.8        | 25.6           | 1.23   | 107   | 5.93  |
| 4   | 12.1        | 20.7        | 24.6           | 1.19   | 106   | 5.85  |
| 5   | 15.1        | 30.6        | 30.9           | 1.01   | 107   | 4.32  |
| 6   | 18.8        | 34.5        | 29.7           | 0.86   | 107   | 3.35  |
| 7   | 23.5        | 31.7        | 24.0           | 0.76   | 106   | 3.42  |
| 8   | 29.4        | 40.4        | 28.3           | 0.70   | 107   | 2.36  |
| 9   | 36.4        | 50.4        | 28.9           | 0.57   | 106   | 1.45  |
| 10  | 47.2        | 58.4        | 24.4           | 0.42   | 108   | 0.77  |
The PDF of the rotation angle $\Phi$ for the rotational Cauchy distribution can be written as

$$f(\Phi) = \frac{2}{\pi} \left[ \frac{\kappa A^2 (1 + A^2)}{(\kappa^2 + A^2)^2} \right]$$

$$= \frac{4\kappa [1 - \cos(\Phi)]}{\pi [1 + \kappa^2 + (\kappa^2 - 1) \cos(\Phi)]^2}, \quad \text{for } 0^\circ \leq \Phi \leq 180^\circ,$$

(7)

where $A = \tan(\Phi/2)$. The scale parameter $\kappa$ of the Cauchy distribution represents the degree of incoherence or complexity in a set of earthquake focal mechanisms.

The VMF distribution is a Gaussian-shaped function defined on the 3-D hypersphere of a normalized quaternion. It is concentrated near the zero angle $\Phi$. This distribution can be implemented to model random errors in determining focal mechanisms.

The VMF distribution is obtained by generating a 3-D normally distributed random variable $u$ ($u_1, u_2, u_3$) with a zero mean and the standard deviation $\sigma_u = (\sigma_{u_1}, \sigma_{u_2}, \sigma_{u_3})$ and then calculating the unit quaternion

$$q_0 = 1/\sqrt{1 + u_1^2 + u_2^2 + u_3^2},$$

$$q_i = u_i/\sqrt{1 + u_1^2 + u_2^2 + u_3^2}, \quad \text{for } i = 1, 2, 3.$$  

(8)

The 3-D rotation angle is calculated

$$\Phi = 2 \arccos(q_0).$$

(9)

Since components of the vector $u$ are normally distributed, the sum $(u_1^2 + u_2^2 + u_3^2)$ in eq. (8) follows the Maxwell distribution. The Maxwell equation describes the distribution of the vector length in three dimensions, if the vector components have a Gaussian distribution with a zero mean and a standard error $\sigma_u$. For small $\sigma_u$ the distribution of angle $\Phi$ follows the Maxwell law (Kagan 2013). The PDF and cumulative distribution function (CDF) for the VMF law can be taken from Kagan (2013, eqs 20–21).

In above equations these Cauchy and VMF laws are theoretically defined for orientation angle range $0^\circ$–$180^\circ$. However, because of the orthorhombic symmetry of the general DC source (Kagan 2013), its maximum disorientation cannot exceed $120^\circ$. Since we cannot obtain an analytical representation of these distributions for the DC source, we use simulation (Kagan 1992) to derive these distributions for the range $0^\circ$–$120^\circ$.

Fig. 2 displays the DC random cumulative rotation distribution (integrated from eqs 3 to 6), as well as several Cauchy distributions. The Cauchy distribution (7) has only one parameter ($\kappa$), a smaller $\kappa$-value corresponds to the rotation angle $\Phi$ concentrated closer to zero. The Cauchy laws CDF for a DC source are taken from Kagan (1992, see fig. 3a). These theoretical laws can be compared to the cumulative $\Phi_1$ and $\Phi_2$ angle distributions. The Cauchy law approximates the $\Phi_1$ curve reasonably well for $\kappa = 0.075$ up to about $20^\circ$. Even for larger angles the observation curve is close to the Cauchy law line. A similar effect is observed for the $\Phi_2$ curve.

In Fig. 3 the focal mechanism angles are compared to the cumulative VMF distribution for a DC source (see Kagan 1992, fig. 3c). As expected (see Kagan 2013) the fit of the VMF law to observation curves is not as good as for the Cauchy distribution, which provides a much better approximation (Fig. 2).

Table 2 shows the values of the information scores, $I$ (Kagan 2009) for both theoretical distributions, Cauchy and VMF. The information score in this case can be calculated as

$$I = \frac{1}{n} \sum_{i=1}^{n} \lambda_i \log \lambda_i,$$

(10)

that is as the sum of the cells for ensemble average. Above $\lambda_i$ is the normalized frequency in an $i$-interval. As can be seen from
Table 2. Information scores in bits for theoretical distributions (see Section 3). $\kappa$ is the parameter of the rotational Cauchy distribution (see eq. 7 and Fig. 2), $\sigma_i$, ditto for the von Mises-Fisher (VMF) rotational distribution (see eq. 8 and Fig. 3).

| # | Parameter | Cauchy ($\kappa$) | VMF ($\sigma$) |
|---|-----------|-------------------|----------------|
| 1 | 0.025     | 7.48              | 0.50          | 8.15         |
| 2 | 0.050     | 4.86              | 0.10          | 5.21         |
| 3 | 0.075     | 3.49              | 0.20          | 4.44         |
| 4 | 0.100     | 2.60              | 0.30          | 1.03         |
| 5 | 0.200     | 0.95              | 0.40          | 0.30         |
| 6 | 0.500     | 0.05              | 0.50          | 0.03         |

Figs 2 and 3, the curves that are close to the random rotation line (eqs 3–6) have the score value closer to zero.

4 ERROR DIAGRAMS

Figs 2–3, as well as Fig. 4, show cumulative distributions of rotation angles in a concentration diagram format which is similar to the error diagram (ED) display (Molchan & Kagan 1992; Kagan 2009). Characterizing prediction performance is a major challenge for EDs 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 forecast efficiency or skill value.

Several functionals have been proposed as a measure of ED forecast efficiency: the minimax strategy (Molchan 1991; Molchan & Kagan 1992; Kossobokov 2006; Molchan & Keilis-Borok 2008), the sum of errors ($v + \tau$) (ibid., see also eq. 11 below), the area skill score (Zechar & Jordan 2008, 2010), 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 behaviour have the same ‘sum of errors’ value.

Moreover, as Kagan (2009, fig. 10) suggests, the fluctuations of the synthetic curves also indicate that some strategies proposed to measure the performance of a prediction algorithm by considering the ED, like a sum of errors ($v + \tau$) or minimax errors (see above) are biased for a small number of forecasted events.

The original ED plots assume that the null hypothesis is a descending diagonal connecting square corners. It corresponds to the uniform Poisson distribution of the points in a region, that is a random guess forecast strategy or unskilled forecast. However, as we explained earlier (see Section 1 and eqs 3–6), in the case of earthquake source orientation the random rotation is a non-linear function of the angle. Qualitatively, one can judge the difference between a skillful forecast and a random guess by looking at the distance between these curves. However, our goal is to provide a quantitative score.

As we explained in Section 1, an efficient way to measure the forecast skill is by evaluating the likelihood information scores. There are considerable problems in computing the information score for the rotation angles $\Phi_1$ and $\Phi_2$. As Kagan (2009) discusses, we have a reasonable model for the earthquake spatial rate forecast. We evaluate the smoothing kernel parameters using some optimizing criteria (Kagan & Jackson 2011). The optimized theoretical earthquake spatial density model allows us to evaluate several scores $I_0$, $I_1$, $I_2$ and $I_3$ (Kagan 2009, table 3) by using the direct likelihood computation as well as by integrating the EDs.

However, no such general model is yet available for the distributions of angles $\Phi_1$ and $\Phi_2$. Since the number of observations is relatively small, the concentration diagrams or cumulative distributions for these angles contains step-like jumps. In such a case, we can calculate the score that Kagan (2009) called $I_1$, which can be computed from the ED curve in absence of a probabilistic model. The $I_1$ estimate is biased upwards for small samples (ibid.), though for large number of events $I_1$ approaches the theoretical score value (Kagan 2007).

If the number of segments $n_i$ in cumulative curves (cf. Figs 2–4) is less or equal to the number of events $n$, then by applying generalized eq. 14 in Kagan (2009) to the cumulative curve of earthquakes, we evaluate the information score (compare also eq. 11 by Kagan 2007)

$$I_1 = \frac{n_i + 1}{n} \sum_{j=1}^{n_i} \Delta v_j \log_2 \left[ -\frac{\Delta v_j}{\Delta \tau_j} \right]$$

$$= \sum_{j=1}^{n_i} (v_j - v_{j+1}) \times \log_2 \left[ -\frac{v_j - v_{j+1}}{\tau_j - \tau_{j+1}} \right], \quad (11)$$

where $\tau$ is the cumulative fraction of the alarm time, $v$ is the cumulative fraction of failures to predict, and $\log_2$ is used to obtain the score measured in the Shannon bits of information (Kagan 2009). However, in some considerations (see Table 3) the segment (cell) number was selected greater than $n$. For such a case we calculate

$$I_1 = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \Delta v_j \log_2 \left[ -\Delta v_j \sum_{k=1}^{n_i} \frac{\tau_k}{\tau_{k+1}} \right], \quad (12)$$

$k_i$ is the cell number corresponding to the $i$th event.

In Table 3 several estimates of the $I_1$ score are shown for both angles ($\Phi_1$ and $\Phi_2$) and for two choices of smoothing kernel width ($r_1$). We subdivide the angle range ($0°–120°$) into various numbers of grid cells to see how it influences the $I_1$-value. The number of cells with non-zero number of events is relatively small for large cell size, but for a finer subdivision it approaches the total number.

Figure 4. GCMT catalogue, 2008–2012, $r_1 = 2.5$ km, latitude range [75°S–75°N], earthquake number $n = 1066$. Curves are cumulative distributions of the observed rotation angle $\Phi_1$ at earthquake centroids, for 10 subsets with an increasing angle of $\Phi_1$ (see Fig. 1 and Table 1). Left green solid curve is for the rotational Cauchy distribution (7) with $\kappa = 0.025$. Right green solid line is for the random rotation.
of angle measurements (1066). The score value also approaches an upper limit for a finer subdivision. The results do not appear to depend on the smoothing kernel \(r_s\)-value.

The final \(I_s\)-values can be reasonably well forecast by comparing their approximation by the Cauchy distribution in Fig. 2 with the appropriate score values in Table 2. For example, the major part of the \(\Phi_1\) curve is between Cauchy curves \(\kappa = 0.05\) and \(\kappa = 0.1\), and the score value is also between the respective values in Table 2.

Table 4 shows the values of the information scores for the 2008–2012 GCMT catalogue in a format similar to table 3 by Kagan (2009). We vary several parameters to investigate the dependence of scores. For example, the scores change with the grid modification, partly because almost all the events are located in separate cells for a higher resolution forecast. However, since the score \(I_2\) is calculated for the actual earthquake locations in the test period, their value does not depend on the cell size, as expected.

The influence of the smoothing kernel width \(r_s\) on the score appears to be insignificant. An additional study is necessary to optimize our forecast by changing \(r_s\); unfortunately the needed computations are very extensive.

Fig. 4 displays cumulative \(\Phi_2\) distributions for 10 subsets shown in Fig. 1 and Table 1. The distributions move from left to right; the distribution for the first subset is close to the Cauchy \(\kappa = 0.025\) law curve, whereas the last distribution is close to the random curve. This observation is confirmed by \(\Phi_1\) distribution in Fig. 1: the maximum value of the angle is 75.07°. It is only slightly less than 75.16°—the average angle for the random rotation of DC source (Grimmer 1979, his table 5; Kagan 1992). The score values shown in the last column of Table 1 verify this pattern: for the initial subsets, \(I_s\)-value is close to \(\kappa = 0.25\) Cauchy distribution (see Table 2), whereas for the 10th subset, \(I_s\)-value approaches zero.

### Table 3.

| # | Subdivision | \(\xi\) | \(\Phi_1\) | \(\Phi_2\) | \(\xi\) | \(\Phi_1\) | \(\Phi_2\) |
|---|-------------|--------|--------|--------|--------|--------|--------|
| 1 | 120         | 70     | 4.18   | 108    | 3.49   | 69     | 4.04   |
| 2 | 1200        | 448    | 4.41   | 538    | 3.74   | 452    | 4.28   |
| 3 | 120 000     | 955    | 4.74   | 975    | 4.03   | 957    | 4.59   |
| 4 | 120 000     | 1058   | 4.94   | 1053   | 4.18   | 1049   | 4.76   |
| 5 | 1 200 000   | 1064   | 4.97   | 1065   | 4.24   | 1050   | 4.77   |

### Table 4.

| # | 2008–2012 [75°S–75°N] |
|---|------------------------|
| Grid | 0.1° | 0.5° | 0.5° |
| \(r_s\) | 2.5 km | 2.5 km | 6.0 km |
| \(\xi\) | 1025 | 758 | 758 |

### Figure 5.

Figure 5. GCMT catalogue, 2004–2006, North-west (NW) Pacific, \(r_s = 15\) km, earthquake number \(n = 108\). Relation between two scores \(I_2\) and \(I_1\) in bits, shown by circles, for first 10 simulated catalogues, displayed in fig. 10 by Kagan (2009). We calculate two regression lines approximating the interdependence of the \(I_1\) and \(I_2\) scores, the linear and quadratic curves. To make expressions clearer in regression curves we use \(I_1 = 2.4\) as a reference point. The coefficient of correlation \(\rho\) between the angles is high 0.9, indicating that the \(I_2\) estimate can be well evaluated by \(I_1\) regression. Results of both regressions are written at the top of the plot: \(\sigma\), standard error, \(\epsilon_{\text{max}}\), maximum difference. A cross sign inside a square shows \(I_1\) and \(I_2\) estimates for the NW Pacific region (lower) and the South-west Pacific region (upper) from Table 3 by Kagan (2009). These two points are for illustration only, they are not used in regression calculations.

### 5 DISCUSSION

The likelihood approach for focal mechanism orientation pattern interpretation allows for earthquake rate prediction (Kagan & Jackson 2011) to be adequately combined with the focal mechanism forecast. This should result in a general earthquake forecast optimization.

As Kagan (2009) observed, the \(I_1\) score estimates are biased and have a higher random variation compared to the other information scores. In order to understand the properties of the \(I_1\) score, we study the correlation between \(I_1\) and other scores. To investigate the relation between the two scores, \(I_1\) and \(I_2\), in Fig. 5 we calculated these scores for 10 simulated catalogues shown in fig. 10 (Kagan 2009) for the North-west (NW) Pacific region. The range of \(I_1\) values (0.8) can be compared with the standard deviation for \(I_0\) (\(\sigma_{\text{std}} = 0.215\)), shown in table 3 by Kagan (2009). The scores \(I_0\) and \(I_1\) are optimized to have close values. As the diagram demonstrates, \(I_2\) is usually larger than \(I_1\), but their correlation coefficient is high,
thus one can estimate the $I_1$ score using regression as shown in Fig. 5.

As mentioned in Section 1, a significant effort needs to be extended to incorporate the methods developed in this and previous publication (KJ2014) to forecast earthquake focal mechanisms. A similar investigation needs to be carried out to fully optimize the global earthquake rate forecast (Kagan & Jackson 2011). However, these efforts will be mostly of technical nature, the major forecast scientific issues are addressed in this and abovementioned papers.

6 CONCLUSIONS

(1) We apply a likelihood method to measure the skill of an earthquake focal mechanism forecast. The advantage of such an approach is that the likelihood scores for the earthquake rate prediction can quantitatively be combined with the focal mechanism forecast, resulting in a general forecast optimization.

(2) We compare actual forecasts or occurrences of event source properties with the null hypothesis that the mechanism’s 3-D orientation is random.

(3) We calculate the information (likelihood) score for two rotational distributions (Cauchy and VMF) which are used to approximate a source orientation pattern.

(4) We calculate the likelihood score for earthquake source forecasts based on the GCMT catalogue and their validation by future seismicity data. We explored the dependence of the results on data resolution, internal dependence of scores on forecasted angle and a random variability of likelihood scores.

ACKNOWLEDGEMENTS

The authors appreciate support from the National Science Foundation through grants EAR-0711515, EAR-0944218 and EAR-1045876, as well as from the Southern California Earthquake Center (SCEC). SCEC is funded by NSF Cooperative Agreement EAR-0529922 and USGS Cooperative Agreement 07HQAG0008. The comments by two anonymous reviewers as well as by the associate editor Xiaofei Chen have significantly improved the presentation. Publication 1982, SCEC.

REFERENCES

Bird, P. & Kreemer, C., 2015. Revised tectonic forecast of global shallow seismicity based on version 2.1 of the Global Strain Rate Map, Bull. Seism. Soc. Am., 105(1), 152–166.

Daley, D.J. & Vere-Jones, D., 2004. Scoring probability forecasts for point processes: the entropy score and information gain, J. Applied Probab., 41A, 297–312.

Ekström, G., Nettles, M. & Dziewonski, A.M., 2012. The global CMT project 2004–2010: centroid-moment tensors for 13,017 earthquakes, Phys. Earth planet. Inter., 200–201, 1–9.

Grimmer, H., 1979. Distribution of disorientation angles if all relative orientations of neighboring grains are equally probable, Scripta Metall., 13(2), 161–164.

Harte, D. & Vere-Jones, D., 2005. The entropy score and its uses in earthquake forecasting, Pure appl. Geophys., 162(6–7), 1229–1253.

Jackson, D.D. & Kagan, Y.Y., 1999. Testable earthquake forecasts for 1999, Seism. Res. Lett., 70(4), 393–403.

Kagan, Y.Y., 1990. Random stress and earthquake statistics: spatial dependence, Geophys. J. Int., 102(3), 573–583.

Kagan, Y.Y., 1991. 3-D rotation of double-couple earthquake sources, Geophys. J. Int., 106(3), 709–716.

Kagan, Y.Y., 1992. Correlations of earthquake focal mechanisms, Geophys. J. Int., 110(2), 305–320.

Kagan, Y.Y., 2000. Temporal correlations of earthquake focal mechanisms, Geophys. J. Int., 143(3), 881–897.

Kagan, Y.Y., 2007. On earthquake predictability measurement: information score and error diagram, Pure appl. Geophys., 164(10), 1947–1962.

Kagan, Y.Y., 2009. Testing long-term earthquake forecasts: likelihood methods and error diagrams, Geophys. J. Int., 177(2), 532–542.

Kagan, Y.Y., 2013. Double-couple earthquake source: symmetry and rotation, Geophys. J. Int., 194(2), 1167–1179.

Kagan, Y.Y. & Jackson, D.D., 1994. Long-term probabilistic forecasting of earthquakes, J. geophys. Res., 99(B7), 13 685–13 700.

Kagan, Y.Y. & Jackson, D.D., 2006. Comment on ‘Testing earthquake prediction methods: “the West Pacific short-term forecast of earthquakes with magnitude MwHRV ≥ 5.8”’ by V.G. Kossobokov, Tectonophysics, 413(1–2), 33–38.

Kagan, Y.Y. & Jackson, D.D., 2011. Global earthquake forecasts, Geophys. J. Int., 184(2), 759–776.

Kagan, Y.Y. & Jackson, D.D., 2014. Statistical earthquake focal mechanism forecasts, Geophys. J. Int., 197(1), 620–629 (KJ2014).

Kendall, M.G. & Moran, P.A.P., 1963. Geometrical Probabilities, Hafner, 125 pp.

Kossobokov, V.G., 2006. Testing earthquake prediction methods: “the West Pacific short-term forecast of earthquakes with magnitude MwHRV ≥ 5.8”, Tectonophysics, 413(1–2), 25–31.

Mardia, K.V. & Jupp, P.E., 2000. Directional Statistics, Wiley, 429 pp.

Molchan, G.M., 1991. Structure of optimal strategies in earthquake prediction, Tectonophysics, 193(4), 267–276.

Molchan, G., 2010. Space-time earthquake prediction: the optimal diagrams, Pure appl. Geophys., 167(8/9), 907–917.

Molchan, G., 2012. On the testing of seismicity models, Acta Geophysica, 60(3), 624–637.

Molchan, G.M. & Kagan, Y.Y., 1992. Earthquake prediction and its optimization, J. geophys. Res., 97(B4), 4823–4838.

Molchan, G. & Keilis-Borok, V., 2008. Earthquake prediction: probabilistic aspect, Geophys. J. Int., 173(3), 1012–1017.

Morawiec, A., 2004. Orientations and Rotations: Computations in Crystallographic Textures, Springer-Verlag, pp. 200.

Silverman, B.W., 1986. Density Estimation for Statistics and Data Analysis, Chapman and Hall, 175 pp.

Vere-Jones, D., 1998. Probabilities and information gain for earthquake forecasting, Comput. Seismol., 30, 248–263.

Zechar, J.D. & Jordan, T.H., 2008. Testing alarm-based earthquake predictions, Geophys. J. Int., 172(2), 715–724.

Zechar, J.D. & Jordan, T.H., 2010. The area skill score statistic for evaluating earthquake predictability experiments, Pure appl. Geophys., 167(8/9), 893–906.