A Testable Five-Year Forecast of Moderate and Large Earthquakes in Southern California Based on Smoothed Seismicity

Yan Y. Kagan and David D. Jackson
Department of Earth and Space Sciences, University of California-Los Angeles

Yufang Rong
AIR-Worldwide Corporation

INTRODUCTION

We present a five-year forecast of southern California earthquakes with magnitudes 5.0 and greater. The forecast uses earthquake data only, with no explicit use of tectonic, geologic, or geodetic information. The forecast is based on observed regularity of earthquake occurrence rather than on any physical model. The earthquake rate density (probability per unit area, time, and magnitude) is assumed constant in time. We estimate it as the sum of contributions from all magnitude 5 and larger earthquakes in our catalog, which for large quakes extends from 1800 to the present. The contribution from each quake is inversely proportional to epicentral distance and directly dependent on the magnitude. We use the same model to estimate the probable focal mechanisms of future earthquakes, using a weighted sum of the moment tensors of previous quakes.

We assume that the rate density is proportional to a smoothed version of past seismicity, using a recently published (Kagan *et al.* 2006) catalog of California earthquakes. That catalog includes all known earthquakes larger than magnitude 7.5 since 1800 and smaller earthquakes as the lower threshold of completeness has decreased with time. We treat earthquakes as point sources, except that large earthquakes ($M \geq 6.5$) are represented by multiple rectangular dislocation patches (see figure 2 in Kagan *et al.* 2006). For the forecast that is the subject of this paper, we represent each of the patches by a point source at its center.

The estimated rate density depends linearly on the magnitude of past earthquakes and inversely as epicentral distance out to a few hundred kilometers. We assume that 2% of all earthquakes are *surprises*, assumed uniformly likely in those areas with no earthquakes. Foreshocks and aftershocks are treated as any other earthquake.

The method is further described in Kagan and Jackson (1994). We have used the same method to forecast earthquakes for the northwest and southwest Pacific regions (Jackson and Kagan 1999; Kagan and Jackson 2000), shown on our Web site, http://scec.ess.ucla.edu/~ykagan/predictions_index.html. We have also applied this technique since 1992 (Jackson *et al.* 1995; Kagan *et al.* 2003) in an experimental long-term forecast for southern California. The latest version is on the Web at http://moho.ess.ucla.edu/~kagan/s_cal_haztbl.tmp.

The model contains just a few adjustable parameters, which we normally determine so that a forecast based on the first half of a catalog best fits the second half. At any given time we can revise the parameter estimates and the forecast. For the California testing experiment, we will fix all of the parameters and suspend revisions for the five-year duration of the test.

Our forecast differs from some other forecasts presented in this issue (Bird and Liu 2007, this issue; Helmstetter *et al.* 2007, this issue; Shen *et al.* 2007, this issue) and elsewhere in several important respects.

1. By using the longest possible time span for our "learning" catalog, we have tried to construct a model that will be valid over several decades. Some other models are based on smaller earthquakes over shorter periods; these may not capture effects that are only evident in the long term.
2. We use an extended source representation for large earthquakes: We replace each epicenter point by a series of rectangular dislocations that cover the whole rupture area. Therefore, we forecast non-negligible earthquake rates for the whole 1857 Fort Tejon rupture extent, though some of these areas currently exhibit almost no seismic activity.
3. We use only moderate and large ($\mathbf{M} \geq 5.0$) earthquakes in our forecast. Thus we largely avoid the complications of smaller earthquakes caused by volcanism, geothermal activity, or fault creep.
4. Our forecasting algorithm predicts not only the rate, size distribution, and location of future earthquakes, as almost all other forecast programs do, but also focal mechanisms of these events with an indication of forecast uncertainty.

SMOOTHING METHOD

We assume that the rate-density function can be factored into three functions depending independently on location, magnitude, and time. That is, we let

$$\Lambda(\theta, \phi, m, t) = f(\theta, \phi) \times g(m) \times h(t), \qquad (1)$$

where Λ is the rate of earthquakes per unit area, magnitude, and time; θ is latitude, ϕ is longitude, m is magnitude, t is time, f is the spatial density function, g is the normalized magnitude distribution, and h is the time function. The functions f and g are normalized so that their integrals over space equal 1.0. Then $h(t)$ is the rate (number per unit time) of all earthquakes within the area and magnitude range of interest. Foreshocks and aftershocks are treated like any other earthquakes. In this work we take $h(t)$ to be constant, *i.e.*, our forecast is time-independent. However, the same separation of variables implied by equation (1) is also a useful approximation in time-dependent models.

The spatial density function f is a weighted sum of smoothing kernels, each centered at the epicenter of a previous ith quake

$$f(\theta,\phi) = \sum_i f_i(\theta_i,\phi_i) + s, \qquad (2)$$

where $s = 0.02$ is a small constant to allow for "surprises" far from past earthquakes. Our spatial smoothing kernels have the form

$$f_i(r_i) = A \times (m_i - 5.0) \times \frac{1}{r_i} \times \left[1 + \delta \cos^2(\psi_i)\right] + s, \qquad (3)$$

for $r_i < 200$ km and 0 otherwise. Here A is a normalization constant assuring that the integral of f over space is unity, r_i is epicentral distance, d is a smoothing distance, δ is a parameter indicating the degree of azimuthal concentration, and ψ_i measures the orientation of the map point relative to the fault-plane azimuth for an event in a catalog (Kagan and Jackson 1994). The parameter δ above controls the degree of azimuthal concentration in a direction related to the earthquake's focal mechanism (Kagan and Jackson 1994, figure 2). We take δ to be 100 in our calculations. The shape of the spatial smoothing kernel is somewhat arbitrary, except that it should decrease with distance, remain above zero out to several hundred kilometers, and be integrable over space. For the function above, we had to truncate it at some distance (we chose 200 km) to make the kernel function integrable. We chose the kernel's form after experimenting with several alternatives.

We assume that the earthquake size distribution follows the tapered Gutenberg-Richter (G-R) relation (Gutenberg and Richter 1944; Kagan 2002; Bird and Kagan 2004). The parameters of the relation are assumed to be constant over southern California: the slope (b-value) $b = 0.95$ and the corner moment magnitude $M_C = 8.0$. These values are suggested by Bird and Kagan (2004) for continental transform fault boundaries.

The smoothing kernel has just a few adjustable parameters: a normalizing factor, a smoothing distance, and an anisotropy factor. Even the normalizing factor is barely adjustable, because we always set it equal to the average earthquake rate of the learning catalog. Normally we determine the smoothing distance and anisotropy factor by optimizing the forecast of the second half of the available catalog from the first half. We could not do that for southern California because the catalog includes too few earthquakes and because the catalog quality changes with time. Thus, we adopted the values from our analysis of global seis-micity (Jackson and Kagan 1999; Kagan and Jackson 2000). Although these global values have been determined for point sources, and in the California forecast we use extended representation of any earthquake with $M \geq 6.5$, our tests (Kagan *et al.* 2003) show that this smoothing yields a reasonable estimate of future seismicity.

RESULTS

Figure 1 shows the rate density of earthquakes of magnitude 5 forecasted by the smoothed seismicity model. We assume that larger earthquakes will have a rate consistent with a tapered Gutenberg-Richter magnitude distribution with $b = 0.95$ and corner magnitude 8.0. Because we assume a uniform magnitude distribution, a rate density map for magnitude 7 and above, for example, would look the same as figure 1, if the color scale were shifted accordingly.

We extended our long-term seismicity forecast to include focal mechanisms as a predicted quantity (see figure 2), with a statement of their uncertainty. These focal mechanism forecasts have implications for seismic hazard because ground shaking depends on faulting style. The focal mechanisms of earthquakes depend on the distribution of fault orientation, so our mechanism studies provide useful data for future models based on stress interactions. FORTRAN programs used in creating forecasts described above are available at our Web site: http://scec.ess.ucla.edu/~ykagan/programs_cal_index.html.

As an experiment we tested a "pseudo-prospective" forecast in southern California. We defined the probabilities using earthquakes before the beginning of 1993 and tested against later earthquakes, obtaining diagrams similar to figure 1. The smoothed seismicity model predicted that 90% of the earthquakes should lie in the "hottest" 58% of the area covered, while in fact all events after 1993 lay in the hottest 41%. Thus we smoothed too much; that is, a smaller value of the smoothing distance would have concentrated the probability in a smaller zone where the test earthquakes eventually occurred. Optimizing the parameters with southern California data would probably result in a sharper forecast. We also used the model to construct random synthetic earthquake catalogs that include focal mechanisms. The likelihood scores of the synthetic catalogs were close to those of the real catalog. Because the likelihood score is a measure of the compatibility of the model and the data, the similarity mentioned above suggests that the earthquake data fit the model as well as could be expected.

The forecasting methodology we employ in this paper can be relatively easily extended to northern California and other seismic regions. What is needed for such an extension is an earthquake catalog in which large earthquakes have their rupture pattern documented.

DISCUSSION

In our forecast we've made several assumptions worth mentioning here. Perhaps the most important is that the earthquake rate is stationary in time, implicit in estimating the future earth-

Southern California Smoothed Seismicity Forecast: 1850-2005

Log_{10} Earthquake Rate $M_w > 5.0$, eq/year*(100km)^2
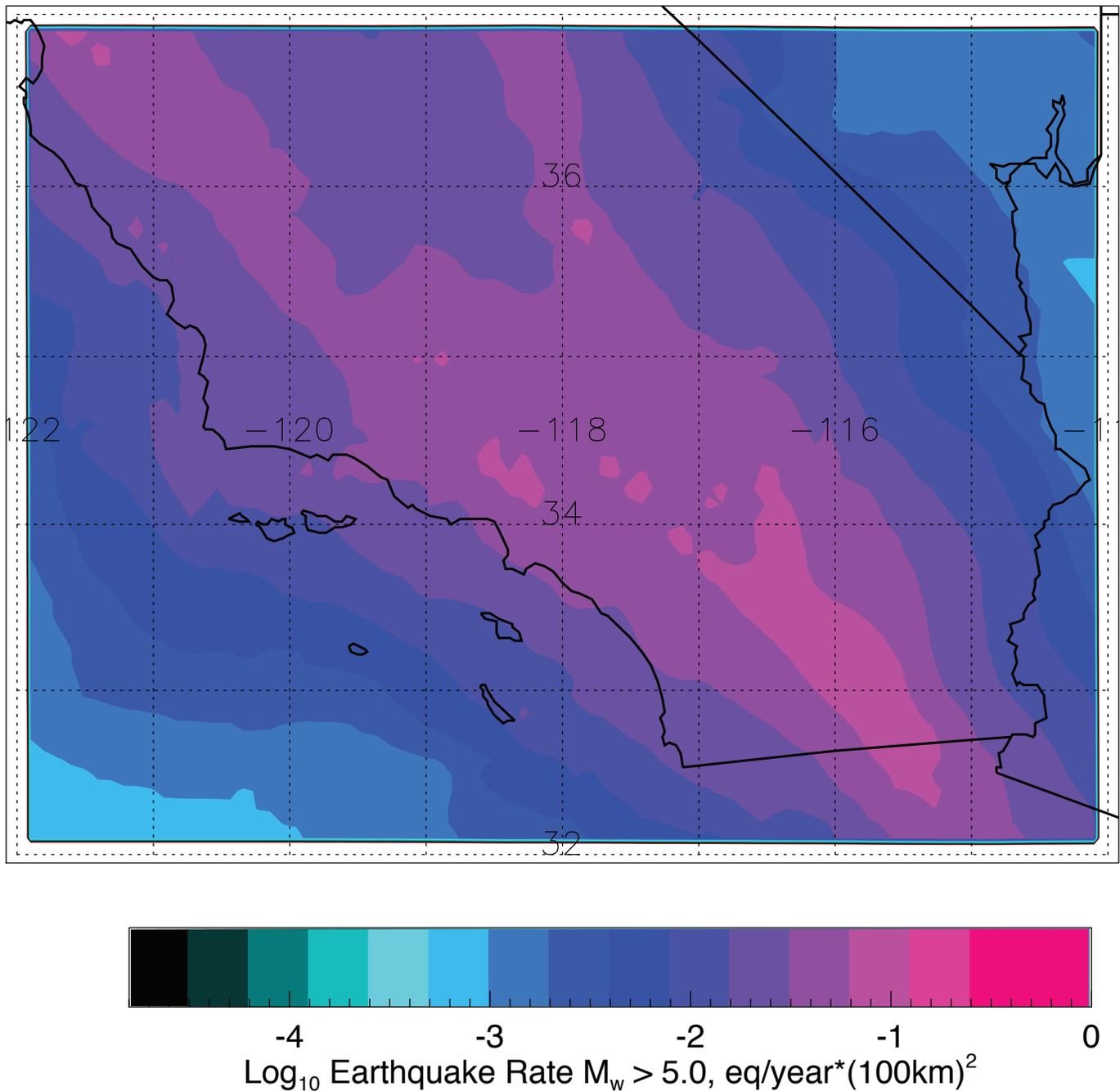
▲ **Figure 1.** Earthquake potential based on smoothed seismicity. Earthquakes from the RELM catalog since 1850 are used. Earthquake occurrence is modeled by a time-independent (Poisson) process. Color scale tones show the long-term probability of earthquake occurrence. URL http://moho.ess.ucla.edu/~kagan/s_calif_fps1850-2005.ps.

quake rate directly from the past rate. Our studies of earthquake catalogs (for instance, Kagan and Jackson 1991) have revealed many exceptions to the assumed stationarity, but they are fairly forgiving. Most of the departures from stationarity result from clusters of events following fairly large earthquakes. Such large earthquakes can alter the seismicity rather strongly, but since they are rare, the stationarity assumption works most of the time. We also assume that the relative (that is, normalized) mag-

nitude distribution is independent of location. The probability of a magnitude 7 earthquake, for example, varies from place to place only because the multiplicative regional density function $f(\theta, \phi)$ does. Bird and Kagan (2004) found that this assumption was consistent with observed seismicity in several different continental strike-slip regions, but there are likely to be exceptions in southern California. Bird and Kagan studied tectonic events, but earthquakes in volcanic and geothermal zones in
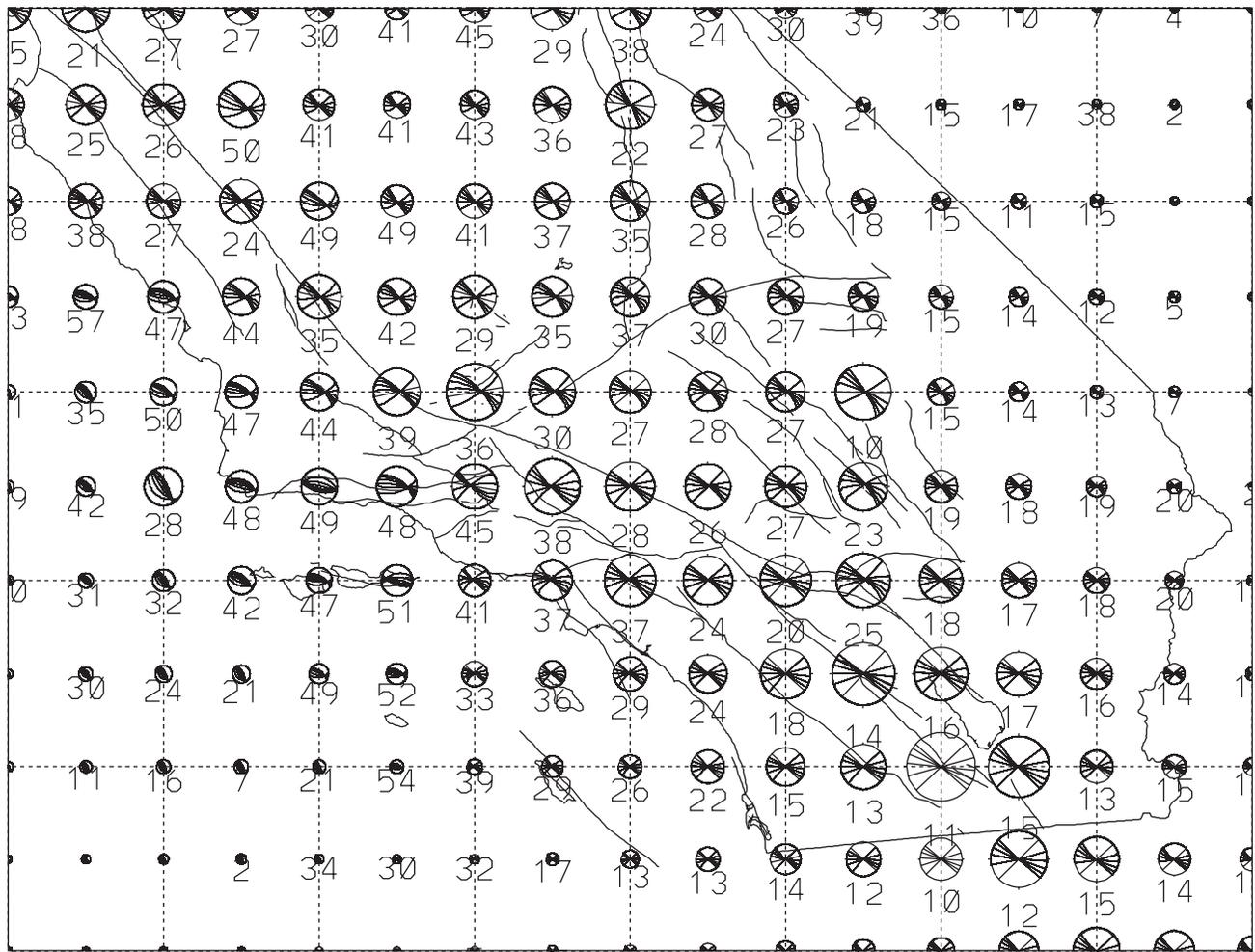
▲ **Figure 2.** Long-term forecast diagrams of earthquake focal mechanisms in southern California. Lower hemisphere diagrams of focal spheres are shown. Size of the focal mechanism diagram is proportional to forecasted rate of occurrence (see figure 1). Stripes in beachballs are concentrated toward the assumed earthquake fault plane. The numbers below the diagrams of earthquake focal mechanisms correspond to a standard deviation of a weighted 3-D rotation angle. We first calculate the average seismic moment tensor and then compute the rotation of earthquake focal mechanisms with regard to the average double-couple source. Therefore the average rotation angle shows degree of tectonic complexity. Points without beachball diagrams denote places for which data are inadequate to forecast focal mechanism. The plot is displayed at URL http://moho.ess.ucla.edu/~kagan/s_cal_fps.ps.

southern California might have a different magnitude distribution. For these earthquakes the corner magnitude is generally smaller than that for tectonic earthquakes, the productivity (*a*-value) is generally higher, and the *b*-value may be different from 0.95 (Kagan 2002, 538–39). Earthquakes in creeping zones, like those on the San Andreas fault north of Parkfield, and those due to filling of water reservoirs, may also have very different magnitude distributions. Clearly such earthquakes need to be treated differently from those associated with global tectonic deformation.

Earthquakes in volcanic, geothermal, and creeping zones have different magnitude distributions from other zones, possibly resulting from different physical processes. The difference between these zones and the rest of California appear most extreme for small-magnitude earthquakes. If these zones have more normal magnitude distributions for moderate and large

quakes, as we assume, then our forecast should be reasonable there. Forecasts based on small quakes could overestimate the rate of larger events in those zones unless the magnitude distribution is properly adjusted. Problems of volcanic, geothermal, and creeping zones may be more serious in forecasts based on events below magnitude 5.

By assuming time-independent probabilities, we miss the opportunity to exploit the clustering behavior so evident in most earthquake catalogs. We capture some of the effects of clustering in the sense that the average seismicity has generally been high where major clusters have occurred. We could capture more of the clustering effect by giving more weight to more recent earthquakes, but we have not done so in this model. We have developed a clustering model for short-term forecasting that gives much stronger weight to recent events (*e.g.*, Jackson

and Kagan 1999; Kagan and Jackson 2000; Kagan *et al.* 2003; Helmstetter *et al.* 2006).

Our forecasting approach is in some ways antithetical to the seismic gap model, which assumes that recent earthquakes deter future ones (Kagan and Jackson 1995; Rong *et al.* 2003; Jackson and Kagan 2006). Moreover, we treat main shocks, foreshocks, and aftershocks equally, without trying to distinguish between them. In so doing we assume that they all have the same magnitude distribution, which is consistent with the results of our own investigations (see the citations above).

CONCLUSIONS

We present a five-year forecast of southern California earthquakes with magnitudes 5.0 and greater. The earthquake rate density (probability per unit area, time, magnitude, and earthquake focal mechanism) is assumed constant in time (*i.e.*, the forecast is time-independent or long-term) and is calculated as the sum of contributions from all magnitude 5 and larger earthquakes in our catalog extending from 1800 to 2005. We represent smaller earthquakes by point sources and each larger one ($M \geq 6.5$) by a series of rectangular dislocations covering the whole rupture area. For recent earthquakes the aftershock zone provides a reasonable measure of the rupture zone. For historic earthquakes we had to use published results of geologic slip and geodetic displacement studies to delineate the rupture zone. ◄

ACKNOWLEDGMENTS

REFERENCES

Bird, P., and Y. Y. Kagan (2004). Plate-tectonic analysis of shallow seismicity: Apparent boundary width, beta, corner magnitude, coupled lithosphere thickness, and coupling in seven tectonic settings. *Bulletin of the Seismological Society of America* **94** (6), 2,380–2,399 (with electronic supplement).

Bird, P., and Z. Liu (2007). Seismic hazard inferred from tectonics: California. *Seismological Research Letters* **78** (1), 37–48.

Gutenberg, B., and C. F. Richter (1944). Frequency of earthquakes in California. *Bulletin of the Seismological Society of America* **34**, 185–188.

Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2006). Comparison of short-term and long-term earthquake forecast models for southern California. *Bulletin of the Seismological Society of America* **96** (1), 90–106.

Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2007). High-resolution time-independent grid-based forecast for *m* ≥ 5 earthquakes in California. *Seismological Research Letters* **78** (1), 78–86.

Jackson, D. D., and Y. Y. Kagan (1999). Testable earthquake forecasts for 1999. *Seismological Research Letters* **70** (4), 393–403.

Jackson, D. D., and Y. Y. Kagan (2006). The 2004 Parkfield earthquake, the 1985 prediction, and characteristic earthquakes: Lessons for the future. *Bulletin of the Seismological Society of America* **96** (4B), S397–S409.

Jackson, D. D., K. Aki, C. A. Cornell, J. H. Dieterich, T. L. Henyey, M. Mahdyiar, D. Schwartz, and S. N. Ward (1995). Seismic hazards in southern California: Probable earthquakes, 1994–2024. *Bulletin of the Seismological Society of America* **85**, 379–439.

Kagan, Y. Y. (2002). Seismic moment distribution revisited: I. Statistical results. *Geophysical Journal International* **148**, 520–541.

Kagan, Y. Y., and D. D. Jackson (1991). Long-term earthquake clustering. *Geophysical Journal International* **104**, 117–133.

Kagan, Y. Y., and D. D. Jackson (1994). Long-term probabilistic forecasting of earthquakes. *Journal of Geophysical Research* **99**, 13,685–13,700.

Kagan, Y. Y., and D. D. Jackson (1995). New seismic gap hypothesis: Five years after. *Journal of Geophysical Research* **100**, 3,943–3,959.

Kagan, Y. Y., and D. D. Jackson (2000). Probabilistic forecasting of earthquakes. *Geophysical Journal International* **143**, 438–453.

Kagan, Y. Y., D. D. Jackson, and Y. F. Rong (2006). A new catalog of southern California earthquakes, 1800–2005. *Seismological Research Letters* **77** (1), 30–38.

Kagan, Y. Y., Y. F. Rong, and D. D. Jackson (2003). Probabilistic forecasting of seismicity. In *Earthquake Science and Seismic Risk Reduction*, ed. F. Mulargia and R. J. Geller, 185–200. Dordrecht: Kluwer Academic Publishing.

Rong, Y.-F., D. D. Jackson and Y. Y. Kagan (2003). Seismic gaps and earthquakes. *Journal of Geophysical Research* **108** (B10), 2471, **ESE-6**, pp. 1–14, doi:10.1029/2002JB002334.

Shen, Z.-K., D. D. Jackson, and Y. Y. Kagan (2007). Implications of geodetic strain rate for future earthquakes, with a five-year forecast of M5 earthquakes in southern California. *Seismological Research Letters* **78** (1), 116–120.

Department of Earth and Space Sciences
University of California-Los Angeles
Los Angeles, California 90095-1567 USA
ykagan@ucla.edu
(Y.Y.K.)
djackson@ucla.edu
(D.D.J.)

AIR-Worldwide Corporation
Boston, Massachusetts 02116 USA
yrong@air-worldwide.com
(Y.R.)