Bulletin of the Seismological Society of America, Vol. 101, No. 4, pp. 1618–1629, August 2011, doi: 10.1785/0120090164

Proximity to Past Earthquakes as a Least-Astonishing Hypothesis for Forecasting Locations of Future Earthquakes

by Alan L. Kafka and John E. Ebel

Abstract The cellular seismology (CS) method of Kafka (2002, 2007) is presented as a least-astonishing null hypothesis that serves as a useful standard of comparison for other, more complex, spatial forecast methods (i.e., methods that forecast the locations, but not the times, of earthquakes). Spatial forecast methods based on analyses of earthquakes in California, such as that of Ebel *et al.* (2007) and the pattern informatics (PI) method of Rundle *et al.* (2002, 2007) provide opportunities for comparing methods that incorporate information about rates of seismicity with a method (i.e., CS) that only assumes that future earthquakes will occur near epicenters of past earthquakes. The Ebel *et al.* (2007) five-year-forecast method (E07) maps the spatial distribution of rates of seismicity, and the PI method not only considers rates of seismicity but also incorporates temporal changes in local rates of seismicity as a measure of the potential for future earthquakes to occur at some location. Our comparison of success rates of the E07 method and the PI method with CS for earthquakes in California has yet to reveal any compelling evidence that inclusion of seismicity rates or temporal changes in local seismicity rates in a spatial forecast model improves the ability to forecast locations of earthquakes.

Introduction

If an earthquake forecast is to be considered successful, it should perform better than a reasonable least-astonishing null hypothesis—but what should that null hypothesis be? Here we propose that, for the case of spatial forecasts (i.e., forecasting the locations but not the times of earthquakes), the cellular seismology (CS) method of Kafka (2002, 2007) is an appropriate choice for a least-astonishing null hypothesis model.

A uniform spatial distribution of epicenters is an obvious first-order choice as a null hypothesis, and any forecast that is claimed to be successful should certainly be able to perform better than a uniform distribution. However, performing better than a uniform spatial distribution is not particularly instructive because (at least in plate boundary regions) it is clear that the spatial distribution of earthquakes is not uniform but rather exhibits a large degree of spatial clustering. We envision CS, which considers the region surrounding any past earthquake to be a potential source point of future earthquakes, as the next logical step beyond the uniform distribution hypothesis for the choice of a least-astonishing hypothesis for spatial earthquake forecasts. CS does assume that the spatial distribution of earthquakes is nonuniform but, just as we would want of a least-astonishing hypothesis, does not depend on any other preconceived notions (beyond the notion of proximity to past earthquakes) regarding the processes that are causing earthquakes in a given region. Other spatial forecasting methods typically depend on additional assumptions that, although they may be well based on theory and observation, render them more complicated than CS. Thus, we advocate for CS as an appropriate choice of a least-astonishing null hypothesis because it is the simplest clustered model of the spatial distribution of seismicity that we can envision.

To illustrate how success rates of other forecast methods can be tested against the CS null hypothesis, we evaluate two examples of spatial forecast methods that incorporate more information than CS. Specifically, two forecast methods based on analyses of earthquakes in California, the method of Ebel *et al.* (2007) and the pattern informatics (PI) method of Rundle *et al.* (2002, 2007), provide opportunities for comparing methods that incorporate information about rates of seismicity with a method (i.e., CS) that only assumes that future earthquakes will occur near epicenters of past earthquakes. The Ebel *et al.* (2007) five-year-forecast method (hereafter referred to as E07) maps the spatial distribution of rates of seismicity, and the PI method not only considers rates of seismicity but also incorporates changes in rates of seismicity as a measure of the potential for future earthquakes to occur at some location.

Cellular Seismology as a Least-Astonishing Null Hypothesis

Details of the CS method are described by Kafka (2007); here we summarize the essence of the method. Figure 1

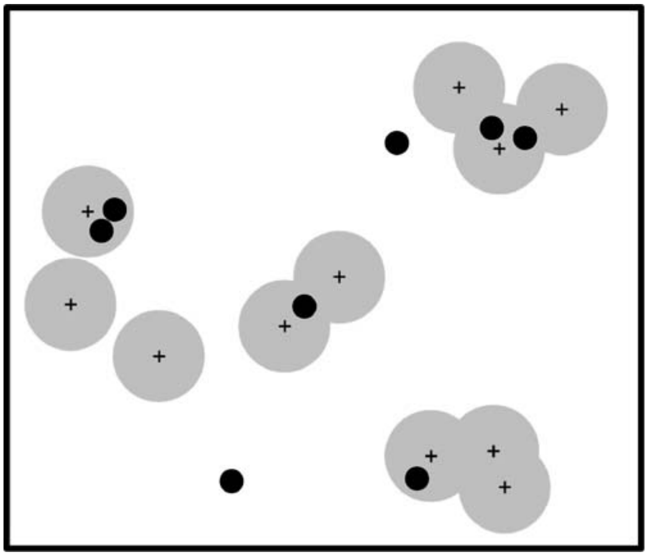

Figure 1. Illustration of the cellular seismology (CS) method, showing a map of a hypothetical region in which earthquakes are occurring. The before catalog epicenters are shown by plus signs, the CS zones are shown by the shaded zones surrounding those plus signs, and the after catalog epicenters are shown by the black-filled circles. For this hypothetical case, there are eight after-catalog earthquakes, and six of them occur within the chosen distance radius of at least one before-catalog earthquake; that is, there are 6/8 (75%) hits for this case.

shows a hypothetical region in which earthquakes are occurring, and for which the CS method (which is analogous to the configuration of a cellular phone system) is being applied. As illustrated for this hypothetical case, we construct circles of a given radius around each epicenter in a catalog of earthquakes that occurred before some specific date (the "before" catalog), and investigate the percentage of earthquakes that took place after that date (the "after" catalog) and were located within that radius of at least one previous earthquake. The radius value is varied so that the area covered by the interiors of the circles fills a given percentage of the map of the study area. If the epicenter of an after-catalog earthquake lies within the given radius of at least one past earthquake, we call that a hit, and the observed percentage of hits is investigated for a given combination of before and after catalogs.

Although CS is a very simple method of characterizing the relationship between past seismicity and locations of future earthquakes, in earlier analyses we tried more complex approaches (including Gaussian smoothing, following the method of Frankel, 1995) and found the results to be quite similar to what we obtained using this simpler CS method (e.g., Kafka and Levin, 2000). Thus, we have adopted the CS method as a simple and straightforward (time-independent) way of measuring the extent to which past seismicity delineates zones where future (i.e., later occurring) earthquakes will occur.

The Regional Earthquake Likelihood Models (RELM) study of California, well-documented in a special issue of *Seismological Research Letters* (*SRL*; e.g., Field, 2007; Schorlemmer, *et al.*, 2007), provides a basis for comparing CS with other candidate null hypotheses. RELM has evolved into several other entities, most notably the Collaboratory for the Study of Earthquake Predictability (see Data and Resources section). For the purpose of this study, however, we base the discussion of the candidate null hypotheses solely on the 2007 RELM issue of *SRL* because various forecast methods are presented there in a uniform format such that they can be easily compared and contrasted. In particular, the forecast methods in this issue of *SRL* are all presented for the same cell size and grid structure.

All of the forecast methods in the RELM issue of *SRL* that we can envision as potential candidates for null hypotheses involve some variation of additional assumptions beyond the simple assumption of proximity to past earthquakes upon which CS is based. For example, the method of Helmstetter *et al.* (2007), another potential candidate for a null hypothesis, involves: (1) declustering the catalog to remove large fluctuations of seismic activity that do not represent the long-term average, (2) estimating the spatial density of seismicity using a kernel to smooth the locations of the before catalog of earthquakes that is processed to form the forecast, and (3) choosing a kernel function. By comparison, CS does not require such assumptions.

For the E07/CS comparison described subsequently (see the section Comparison of Performance of the E07 Method with that of CS), the catalog is declustered of foreshocks and aftershocks but only because the catalog used by E07 was declustered and we wanted to compare the E07 and CS methods based on identical before catalogs. Declustering is not an inherent, or in any way fundamental, assumption of the CS method and is not used for the other comparisons described herein (see the sections Comparison of Performance of the E07 Method with that of CS and Comparison of Performance of the PI Method with that of CS). Thus, CS is not subject to any of the complications associated with declustering that some other methods need to address. CS assumes that any point on a map that is near a past earthquake (foreshock, mainshock, and/or aftershock) is a potential location of a future earthquake.

One other assumption needed for CS is the radius of the CS circles, but that radius is only chosen so that, in comparing with other methods, we assure that both forecast maps cover the same amount of map area. Thus, this assumption is more a part of the testing procedure than part of the theoretical underpinnings of CS. Given the simplicity of the assumptions underlying CS, we present it here as a choice of a least-astonishing hypothesis for forecasting locations of future earthquakes.

Comparison of Performance of the E07 Method with that of CS

The E07 method was developed as part of the RELM study and involves dividing the region into an array of cells

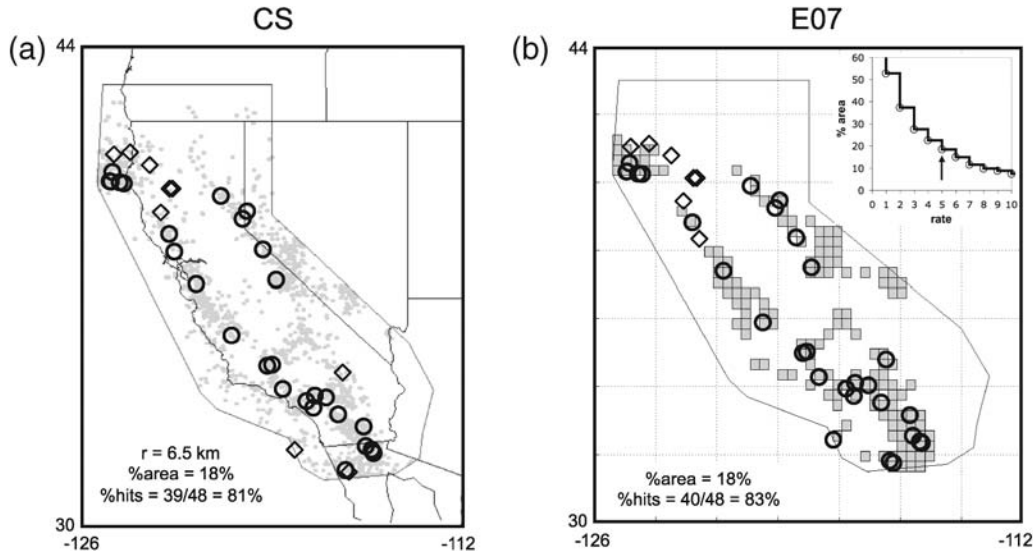

Figure 2. (a) Application of the CS method to earthquakes within the RELM study area polygon (Schorlemmer, *et al.*, 2007; polygon shown here surrounding California). The before catalog is the Advanced National Seismic System (ANSS) catalog of earthquakes with $M \geq$ 4.0 that occurred between 1932 and 2004 and fall within the RELM study area polygon. CS radius is 6.5 km, which corresponds to 18% of the map area (shaded zones within the RELM polygon). The after catalog (open circles for hits and open diamonds for misses) consists of ANSS earthquakes with $M \geq 4.5$ that occurred between 2005 and 2008. (b) Application of the E07 method using the same before and after catalogs as that of the CS analysis shown in (a), with open circles indicating hits and open diamonds indicating misses. Rate threshold for the case shown here (shaded cells) also corresponds to 18% of the area within the RELM polygon. The inset shows the percentage of the polygon map area as a function of rate threshold (number of observed earthquakes) for the cells. The arrow indicates that five earthquakes per cell corresponds to the 18% map area example shown here.

that are 0.3° on each side (Fig. 2) and then determining the rate of earthquake activity in each cell (based on a before catalog). In implementing this method for the RELM region, Ebel *et al.* (2007) chose to remove foreshocks and aftershocks from the before catalog (using the time–space windows of Gardner and Knopoff, 1974) to avoid the situation in which dependent events in cells near regions of large earthquakes dominate the map of the spatial distribution of the rates.

To compare the success rate of the E07 method with that of CS, we applied the CS and E07 methods to the same region and to the same before catalog as was analyzed by E07 (Fig. 2). The before catalog for this case consists of $M \geq 4.0$ earthquakes downloaded from the Advanced National Seismic System (ANSS) catalog that fall within the RELM study area and that occurred from the beginning of 1932 to the end of 2004 (see Data and Resources section). The after catalog for this analysis consists of ANSS earthquakes with $M \geq 4.5$ that occurred from the beginning of 2005 to the end of 2008 and includes 48 events (aftershocks were not removed from this after catalog).

When comparing two seismicity-based forecasts, it is important to consider not only the methods themselves but also the before catalogs that are used to test the forecasts. Rundle *et al.* (2007) noted that it is important to be careful that the before catalog chosen for testing the success rate of a forecast method is not chosen retrospectively so that it is clear that the catalog/forecast method combination is strictly prospective. For the E07/CS comparison, we deal with this issue by limiting the before catalog to being the same as that used by E07. In the section Comparison of Performance of the PI Method with that of CS (a case in which the two methods were developed for regions with significant spatial overlap), we conduct a more thorough analysis of the effect of the choice of a before catalog on the success rates of the different methods. In particular, for the PI/CS comparison, we explore the results of using the exact same before catalog as that used by Kafka (2002) (henceforth referred to as the K02 before catalog) in the development of the CS method, and we also explore the effects of varying the before catalog.

In the case of this E07/CS comparison, we decided that limiting the analysis to the before catalog used by Kafka (2002) was too restrictive for a proper comparison of the two forecast methods. The Kafka (2002) before catalog covers a significantly smaller part of California than the E07 study region, and we wanted to take this opportunity to test the applicability of the CS method on a larger spatial scale than if we were to limit the analysis to the smaller region covered by Kafka (2002). By using the same before catalog as that of E07 and applying the CS method to that before catalog, we are assured that we are testing differences in the method instead of differences in the before catalogs.

For the E07/CS analysis, we varied the size of the CS circles such that the interiors of the circles cover a given percentage of the map area within the RELM study area polygon, and we varied the E07 cutoff for the rates in the cells such that cells with a given cutoff rate or higher cover a given percentage of map area. The purpose of this approach is to convert the rate-based forecast of E07 into a binary (alarm-based) forecast like that of CS, but the approach does this by combining some of the E07 information regarding rates and associated probabilities within the cells.

Figure 2a shows the results of this comparison for a CS radius of 6.5 km, which corresponds to 18% of the map area (shaded zones) within the RELM polygon. In this case there are 39/48 (81%) hits (Fig. 2a). Application of the E07 method using the same before and after catalogs as that of the CS analysis is shown in Figure 2b for a rate threshold corresponding to the same 18% map area as that of Figure 2a, which results in 40/48 (83%) hits. Figure 3 shows the percentage of hits as a function of map area covered by the CS before catalog circles, as well as the percentage of hits for a given E07 rate cutoff as a function of map area covered by the E07 cells corresponding to that rate cutoff. For this analysis, the success rates are similar for the two methods (Fig. 3).

Statistical Significance of Differences in Performance of Forecast Models

To test the statistical significance of the previously described results of the E07/CS comparison, as well as that of other results presented subsequently herein, we use the following method. Consider the observed differences in performance of a given forecast method (FM1) versus some other forecast method (FM2). We model the forecast and testing process as a binomial experiment, with success defined as a hit and failure defined as a miss. We treat the FM1 success rate (π_0) as the null hypothesis, and the alternative hypothesis is that FM2 has a higher success rate than FM1.

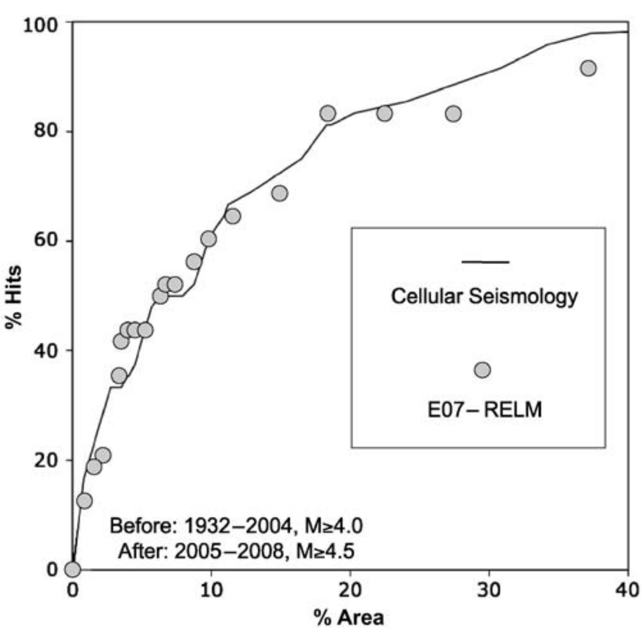

Figure 3. Comparison of performance (percentage of hits as a function of percentage of map area) of E07 versus CS methods of forecasting locations of earthquakes in the study area shown in the maps in Figure 2.

We want to know if an observed higher success rate of one forecast model above that of another is high enough that we can rule out the possibility that the higher success rate is only due to random variation. If π is the observed percentage of hits for FM2, then the p value for the significance test is the binomial probability of observing π or greater percentage of hits, given that π_0 is the true rate of success (i.e., given the null hypothesis).

Using this approach, we calculated p values for all cases shown in Figure 3 where the E07 method yielded a higher percentage of hits than the CS method, and we found that in none of those cases was the difference statistically significant at the 0.05 level. The p values for this E07/CS comparison ranged from 0.10 to 0.43, with a mean of 0.31 and a standard deviation of 0.15. We apply this same approach to test the statistical significance of the other forecast comparisons discussed in the sections Comparison of Performance of the E07 Method with that of CS and Comparison of Performance of the PI Method with that of CS.

Effect of Spatial Resolution on Comparison of Forecast Methods

Reviewers of this paper were concerned that a possible problem with the E07/CS comparison is that the E07 method is discretized but the CS method is not. The highest possible spatial resolution for the E07 method is the size of the 0.3° cells in Figure 2b, whereas CS has essentially infinite resolution (only limited by the resolution of the numerical calculations). We therefore test the effect of resolution on our results by modifying the CS method such that it is based on the same cell size as that of Figure 2b.

In order to make CS maps such as that shown in Figure 2a, a fine grid of discretized cells was used for the entire map area, with the cell size being 0.01° on a side. The method used for producing the CS maps for a given percentage of map area involves testing the centers of each of the cells to see if there are any before-catalog earthquakes within a given CS radius of that cell's central point, as illustrated in Figure 4. Percentage of map area is then calculated by counting the percentage of discretized cells with centers that are within that radius of a before-catalog earthquake (and also correcting for variation in cell area due to differences in the latitude of a cell's location). When the discretized cell size is small relative to the size of the CS radius, this approach provides an accurate estimate of the area covered by the CS circles.

We can investigate the extent to which differences in success rates are due to differences in resolution, as opposed to fundamental differences in the methods, by applying the same CS method but making the discretized cells larger; that is, using the same 0.3° cell size as that of the E07 method. There is, however, a technical difficulty inherent to this lowering of the resolution of the CS method. Consider the two situations shown in Figure 4. The CS radius in Figure 4a is less than half the diagonal across a cell, so there are

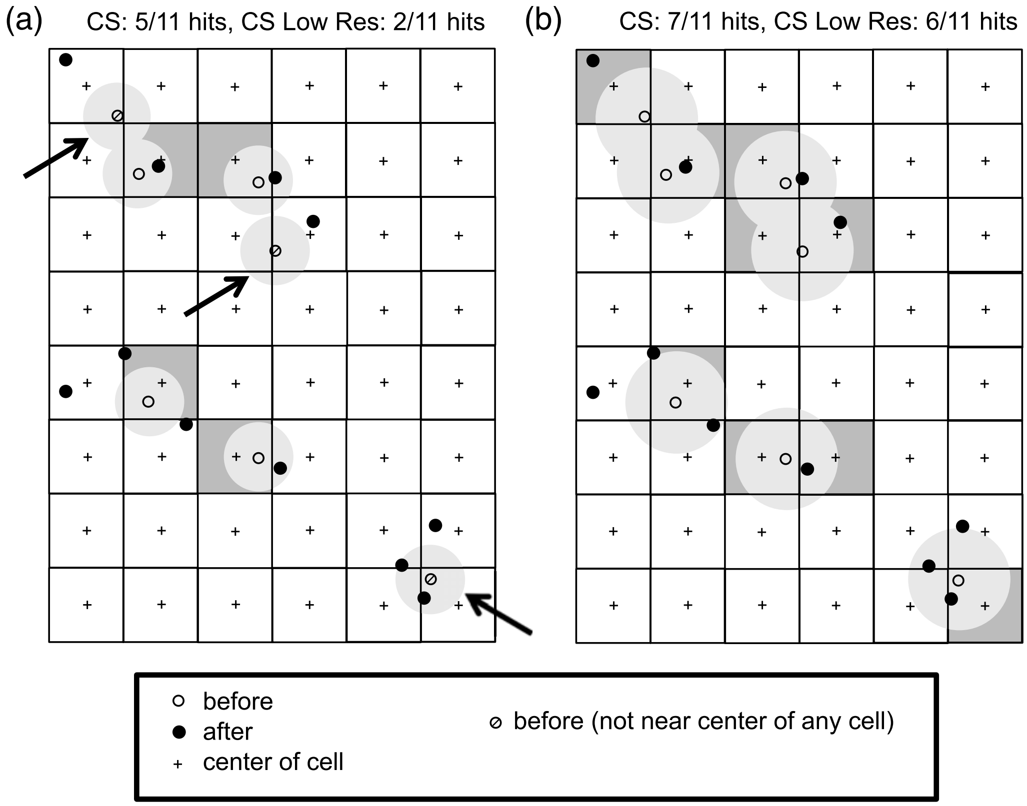

Figure 4. Lower-resolution modification of the CS method such that it can be based on cells of the same size as the grid-based methods discussed here. The method (identical to the CS method, except for the lower resolution) involves testing the centers of each of the cells (+ signs) to determine if there are any before-catalog earthquakes within a given CS radius of the center of that cell. If there is a before-catalog earthquake within that radius, the entire cell is identified (darker shading) as a region within which earthquakes are forecasted to occur. Arrows in (a) indicate before-catalog earthquakes that are too far from the centers of the cells to be counted as being within the CS radius of the center of any cell. When the circle radius is comparable to (or greater than) the size of the cells, as in (b), this method provides an accurate estimate of the map corresponding to the higher-resolution application of the CS method.

some before-catalog earthquake locations that are not near the center of any cell. For any choice of CS radius that is smaller than half the diagonal across a discretized cell, it is possible that there will be before-catalog earthquakes that are too far from the centers of the cells to be counted as being within that radius of the center of any cell (see Fig. 4a). The larger the CS radius relative to the cell size, the less this problem affects the low-resolution estimate of CS. Thus, for a given cell discretization size, there is a minimum radius such that the CS method will be assured of covering all possible locations of before-catalog earthquakes; and, the larger the size of the discretized cells, the more pronounced this effect will be.

For the case of cells 0.3° on a side, the minimum radius that assures that all map points are tested is 22 km (Fig. 5). Thus, for a CS radius of 22 km (or greater) this lower-resolution application of CS is not affected by the cell discretization size of 0.3° on a side. The 22-km CS radius covers 59% of the map area and yields 100% hits for this lower-resolution (0.3° cells) application of the CS method (Fig. 5a). For the CS method applied with the finer grid of cells, the 22-km radius covers 61% of the map area and also yields 100% hits (Fig. 5b). This is, however, hardly an interesting result.

In a separate E07/CS comparison, we chose the same 6.5 km CS radius as that of Figure 2a and applied it to a spatially discretized map with a cell size of 0.3° (Fig. 6b). This procedure results in 37/48 (77%) hits, which is lower than the number of hits obtained with very small discretized cells (39/48, or 81% hits; Fig. 6a). The difference in success rates between the two maps in Figure 6 is not statistically significant ($p = 0.31$). Thus, the effect of coarse discretization in this case is to slightly reduce the number of CS hits, from 81% to 77%. Furthermore, comparing the result of this lower-resolution application of CS (i.e., $37/48 =$ 77% hits, 18% map area) to the E07 result shown in Figure 2b (i.e., $40/48 = 83\%$ hits, 18% map area), we find that, although the E07 success rate is higher than that for the low-resolution CS, the difference is not statistically significant ($p = 0.20$).

The preceding analysis suggests that differences in the success rates of the CS versus the E07 method are not an artifact of the better resolution of the CS method. However, because of its limitations, we do not advocate the lower-resolution application of the CS method over the CS method with the finer grid of cells as the best null hypothesis.

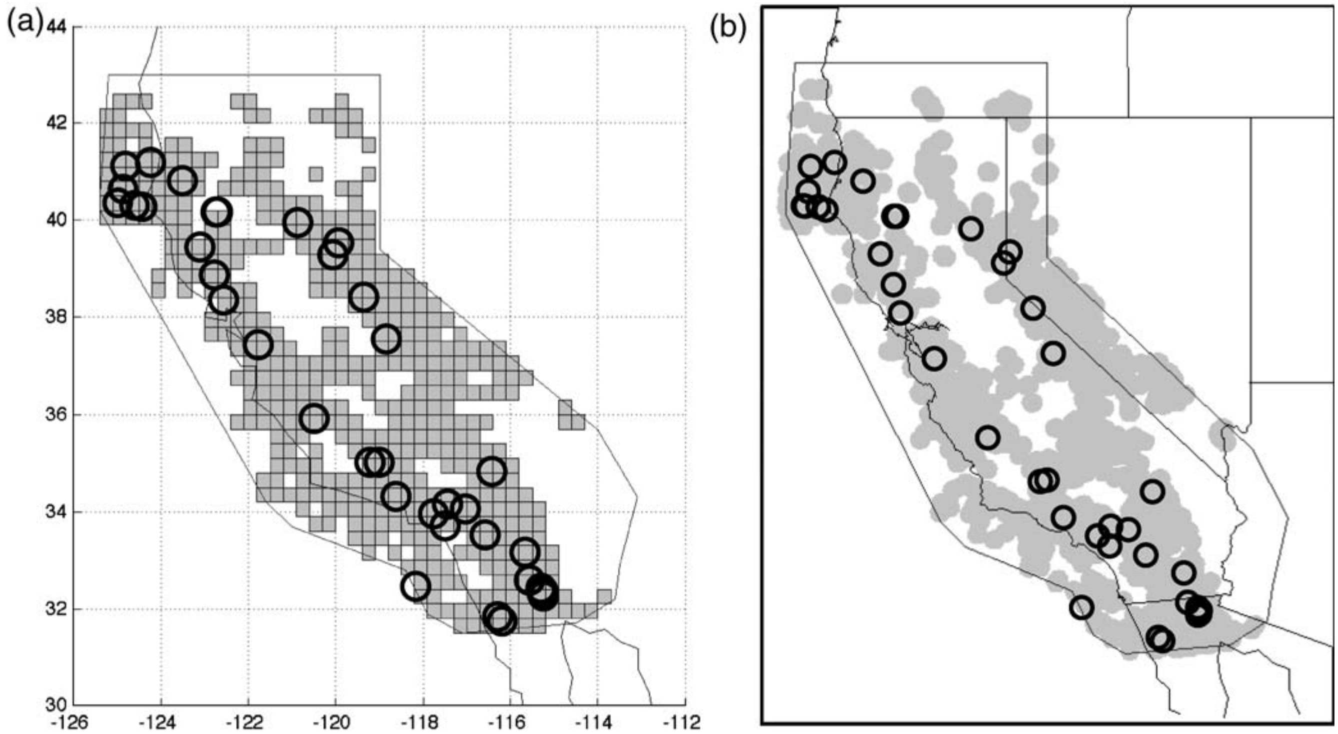

Figure 5. Effect of lowering resolution of the CS method so that it has the same resolution as that of E07 method. For this value of radius (22 km), the (a) lower-resolution result provides an accurate estimate of the CS map calculated at (b) higher resolution, enabling a direct comparison of the lower-resolution CS result with that of the E07 method. However, both applications of the CS method result in a large percentage of map area and 100% hits so are not very instructive results for the comparison of success rates of different methods. (a) radius = 22 km, area = 59%, 48/48 = 100% hits. (b) radius = 22 km, area = 61%, 48/48 = 100% hits.

Specifically, the lower-resolution application results in some of the before-catalog earthquakes not being counted in the CS calculations. In spite of this limitation, for the cases we have analyzed, we find that the differences in the success rates for the lower-resolution application of the CS method versus that of the E07 method are not statistically significant. Our earlier conclusion was that the information in the E07 method about rates of past activity does not add to the ability to forecast the locations of future earthquakes better than the information in the CS method (which contains only information about past

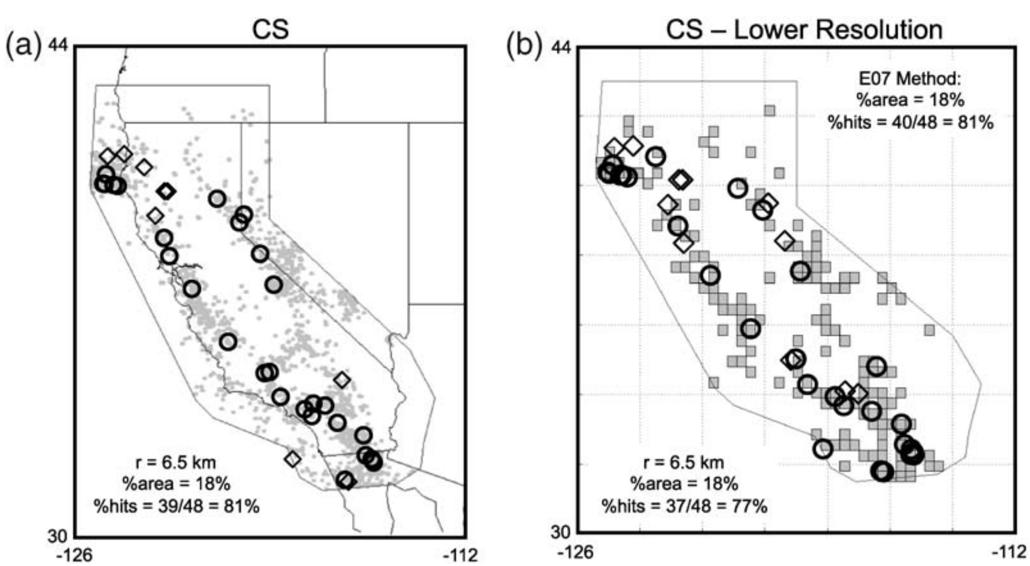

Figure 6. The effect of lowering the resolution of the CS method for the E07/CS comparison, with open circles indicating hits and open diamonds indicating misses. (a) Higher-resolution application of the CS method. (b) Lower-resolution application of the CS method.

earthquake locations). This conclusion does not appear to be an artifact of the resolution of the E07 method versus that of the CS method.

Comparison of Performance of the PI Method with that of CS

Having found evidence that inclusion of information about rates of activity does not add to the ability to forecast the locations of future earthquakes, we now proceed to investigate the extent to which incorporating changes in rates of seismicity in a forecast model affects the success rate of the model relative to that of CS. Here we compare the PI method to CS, based on the CS method exactly as it was published by Kafka (2002) and their before catalog (aftershocks were not removed from the K02 before catalog).

Because the K02 before catalog is the same as that used by Kafka and Levin (2000), it was chosen before the occurrence of any of the post-2000 earthquakes that are in the after catalog used in this analysis. Rundle *et al.* (2007) expressed concern that in our previous comparison of PI with CS (Kafka and Ebel, 2007), we used a different before catalog than that of the original CS method, which might have given us the opportunity to modify our forecast model to produce optimal results. By using the identical before catalog as that of Kafka (2002) and an after catalog consisting of events that occurred after that before catalog was chosen, we insure that this is a strictly prospective test and that no retrospective optimization is occurring here. Additional exploration of the effect of varying the before catalog for this analysis is discussed subsequently in this section.

In this comparison, we also limit the analysis to earthquakes (both before and after 2000) that lie within the map area common to that covered by both Kafka (2002) and Rundle *et al.* (2002); this study area is shown in Figure 7. The after catalog for the analysis shown in Figure 7 consists of earthquakes in the ANSS catalog with $M \geq 5.0$ that occurred from the beginning of 2000 to the end of 2008 and includes 16 events, including aftershocks (see Data and Resources section). We also conducted the same analysis for an after catalog for this same region that consisted of earthquakes in the National Earthquake Information Center (NEIC) catalog with $M \geq 5.0$ that occurred from the beginning of 2000 to the end of 2008 and includes 21 events, including aftershocks (see Table 1 and Data and Resources section).

The PI "hotspots" of Rundle *et al.* (2002) (Fig. 7b) cover 14% of the land in the study area; for that percentage of area, there are 14/16 (88%) hits. Using a CS circle radius that covers that same 14% area (6.3 km), there are also 14/16 (88%) hits for the CS method. Table 1 shows that, for the 14% forecast map area, the success rate is the same for the CS and PI methods. The 2003 San Simeon earthquake is not a hit for the case shown in Figure 7b but is very near the boundary of one of the PI cells; and, because of epicentral uncertainty, one might want to consider this earthquake to be a hit, changing the PI hit rate to 15/16 (94%). Given this 94% success rate for the PI method, one might argue that the PI method is an improvement over CS as a method of forecasting locations of future earthquakes. However, using the preceding formulation, the difference between 14/16 versus 15/16 hits is not statistically significant at the 0.05 level, because we find that the $p = 0.39$ if we assume 15/16 hits

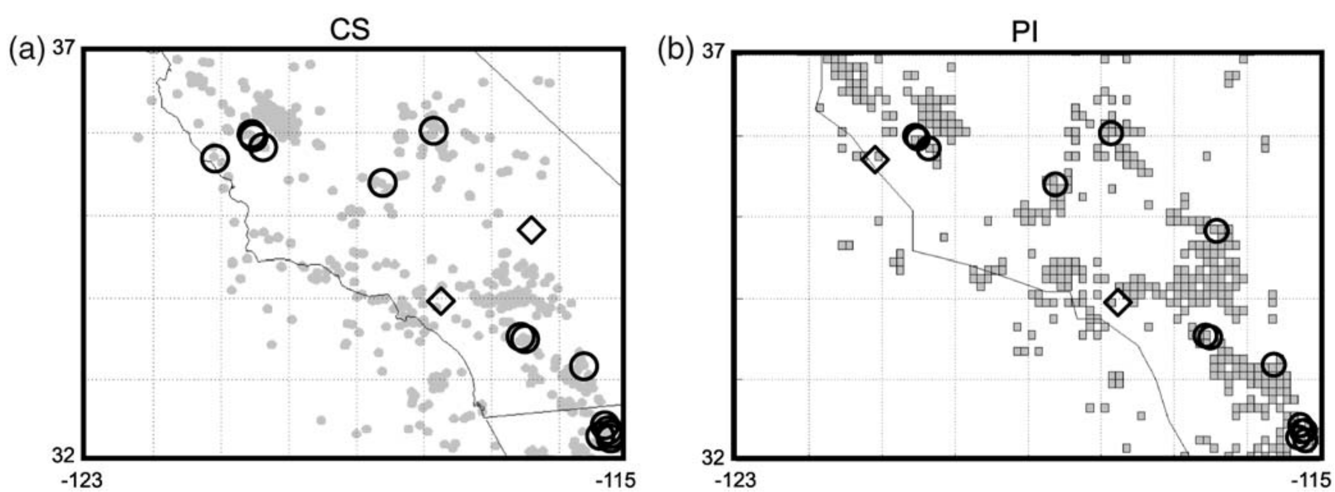

Figure 7. (a) Application of the CS method to earthquakes within the map area common to that covered by both Kafka (2002) and Rundle *et al.* (2002). The before catalog is the same as that used by Kafka (2002) (i.e., for 1984–1987) for $M \geq 3.0$ (aftershocks were not removed from this before catalog). The after catalog (open circles for hits and open diamonds for misses) consists of earthquakes in the ANSS catalog with $M \geq 5.0$ that occurred between 2000 and 2008 and fall within the study area shown here (aftershocks were not removed from this after catalog). This implementation of CS yields 14/16 (88%) hits for 14% map area (shaded zones). (b) PI forecast map of Rundle *et al.* (2002), along with the same after catalog as shown on the left (shaded cells are PI hotspots, which cover 14% of the map area). The before catalog consists of earthquakes with $M \geq 3.0$ from 1932 through 1999. For this after catalog, the PI method yields either 14/16 (88%) hits or 15/16 (94%) hits, depending on whether or not the 2003 San Simeon earthquake (which is on the boundary of one of the PI cells) is considered a hit or a miss.

Table 1
Comparison of Performance of Pattern Informatics (PI) versus Cellular Seismology (CS) Methods (14% Map Area)

After Catalog	% Hits (CS)	% Hits (PI)
ANSS	88%	88% (or 94%*)
NEIC	90%	81% (or 86%*)

*Result if the 2003 San Simeon, California, earthquake is counted as a hit.

for PI. Thus, this possibly higher success rate for PI is not statistically significant at 0.05 level.

The results of applying this same analysis to the NEIC after catalog are also shown in Table 1. For the CS method, the result is 19/21 (90%) hits; for the PI method, the result is either 17/21 (81%) hits or, if the San Simeon earthquake is included as a hit, 18/21 (86%) hits. Given the higher success rate for the CS method over for the PI method (for the same percentage of map area covered), one might argue that the CS method is an improvement over the PI method. However, using the preceding formulation, we again find that the difference is not statistically significant at the 0.05 level because $p = 0.21$ if we assume 17/21 hits for PI and $p = 0.41$ if we assume 18/21 hits for PI.

As in the case of the E07/CS comparison, we can explore the extent to which similarities in success rates are perhaps due to differences in resolution between the two different methods (as opposed to fundamental differences between the PI and CS methods). Using the same grid structure as shown in Figure 7b to compute the CS results (0.1° cells) and applying the CS method with that same lower resolution for the same 14% map area as in the PI forecast, the results for percentages of hits (Fig. 8) are identical to that of the higher-resolution CS analysis shown in Figure 7a. Also, Figure 9 shows that this lower-resolution CS map and the higher-resolution CS map calculated for the same 14% area are visually similar, illustrating that, for a 6.5-km CS radius, the lower-resolution map well approximates the higher-resolution CS map. This analysis suggests that neither the differences in resolution nor anything fundamental about the CS versus PI methods indicate that including changes in rates of seismicity improves the success rate of the forecast over that of the (least-astonishing) CS null hypothesis, in which only the locations of past earthquakes are used to predict of the locations of future earthquakes.

In the preceding analyses, we limited the before catalog to be strictly that of Kafka (2002), but the extent to which that choice of before catalog affects the results can be investigated by varying the before catalog. Table 2 shows the results of dividing the before catalog into ten subcatalogs of four years each, extending from 1999 back to 1963, and performing the same CS analysis to those subcatalogs as we did for the Kafka (2002) catalog. Figure 10 shows the CS analysis for the same before catalog as that used by Rundle *et al.* (2002) (1932–1999, $M \geq 3.0$). In all of the results shown in Table 2 and Figure 10, the CS radii are adjusted so that the CS circles cover the same 14% of land area as covered by the PI hotspots. Figure 11 shows the percentage of hits for the various subcatalogs. As the subcatalogs go back in time, the percentage of hits generally decreases from 94% (ANSS after catalog) and 86% (NEIC after catalog) for the 1996–1999 subcatalog to as low as 38% (ANSS) and 33% (NEIC) for the earlier subcatalogs. However, for the full 1932–1999 catalog, the percentages of hits are 81% (ANSS) and 86% (NEIC), which are again comparable to that of the PI method. None of the differences are statistically significant. Based on these results, it is difficult to discern if the decline

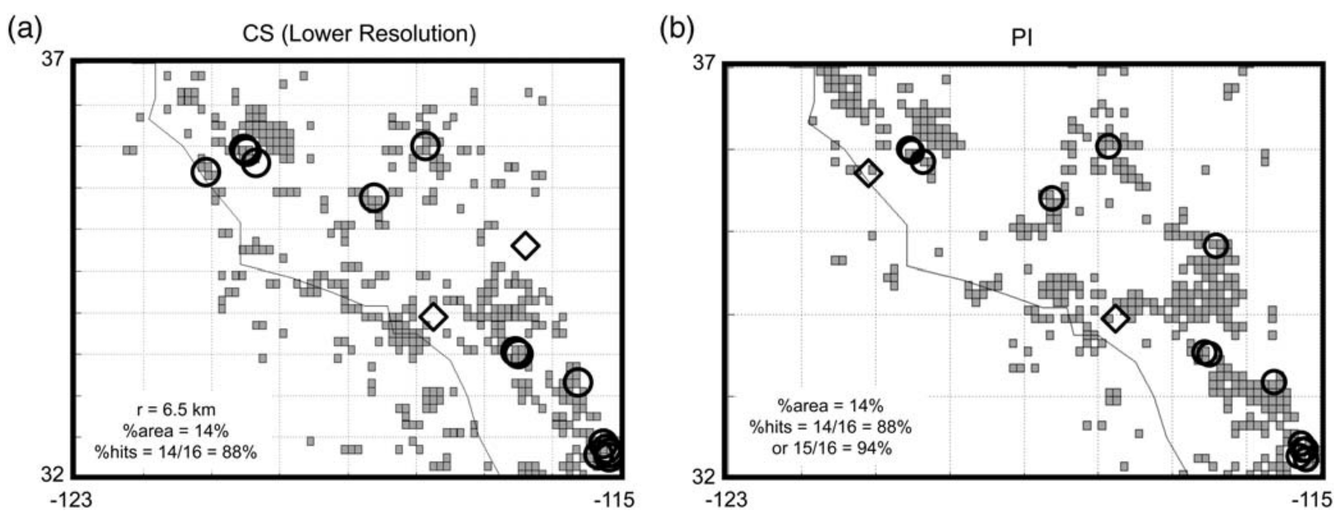

Figure 8. Comparison of (a) lower-resolution CS versus (b) PI forecasts for a 6.5-km CS radius (the same 14% map area as for the PI method). Lower-resolution application of the CS method yields 14/16 = 88% hits. PI method yields either 14/16 (88%) hits or 15/16 (94%) hits. The after catalog is shown by the open circles for hits and open diamonds for misses. The before and after catalogs are the same as in Figure 7.

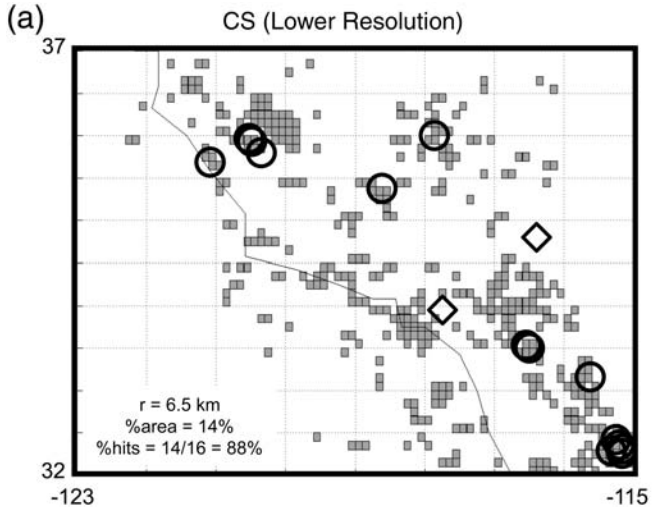

Figure 9. Comparison of lower-resolution CS map versus higher-resolution CS map for the PI forecast region. The after catalog is shown by the open circles for hits and open diamonds for misses; the before and after catalogs are the same as in Figure 7a.

in predictability when the CS method is applied to older subcatalogs is due to true differences in seismicity, catalog incompleteness, or both.

A more insidious complication in comparing forecast methods based on what is ostensibly the same before catalog needs to be noted here. The catalog downloaded from the Southern California Earthquake Center (SCEC) website for the Kafka (2002) study is very similar, but not identical, to the 1984–1987 catalog downloaded more recently from the SCEC website for this study. This is not surprising because earthquake catalogs sometimes change over time as new analyses revisit the locations and/or magnitudes of older events. Based on the Kafka (2002) catalog, there are 88% hits (ANSS) and 90% hits (NEIC) for the 14% map area, whereas, based on the 1984–1987 catalog downloaded for this study, there are 81% hits (ANSS) and 86% hits (NEIC) for the same 14% map area. These differences are not statistically significant, but they do show that uncertainties in earthquake catalogs can have some effect on success rates of forecasts for any particular choice of forecast model.

Table 2
Effect of Varying the Before Catalog of the CS Method Applied to the PI/CS Study Area (14% Map Area)

Before Catalog	% Hits (ANSS)	% Hits (NEIC)
1996–1999	94%	86%
1992–1995	75%	71%
1988–1991	81%	67%
1984–1987	81%	86%
1980–1983	94%	81%
1976–1979	88%	81%
1972–1975	50%	71%
1968–1971	38%	57%
1964–1967	44%	33%
1960–1963	56%	62%
1932–1999	81%	86%
K02 (1984–1987)	88%	90%

Cellular Seismology Forecasts for the RELM Study Region

In this section, we present CS forecasts using the same cell size and grid structure as that of the forecasts in the 2007 RELM *SRL* issue (Fig. 12). The cell size for these RELM forecasts is 0.1° on a side. Also shown in Figure 12 are forecasts based on the higher-resolution (0.01° on a side) application of CS. Although we prefer the higher-resolution application of the CS method, we consider it instructive to also cast the problem in the format of the cell size/grid structure of the

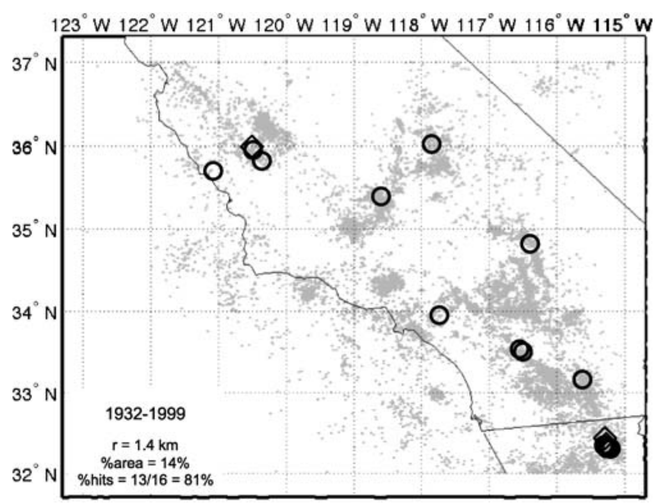

Figure 10. CS method applied to the PI/CS study area. Before catalog: 1932–1999, $M \geq 3.0$ (aftershocks included). After catalog: 2000–2008, $M \geq 5.0$ (aftershocks included); this is the same ANSS after catalog as that of Figure 7. CS radius = 1.4 km, 14% map area, 13/16 = 81% hits. Open circles indicate hits, and open diamonds indicate misses.

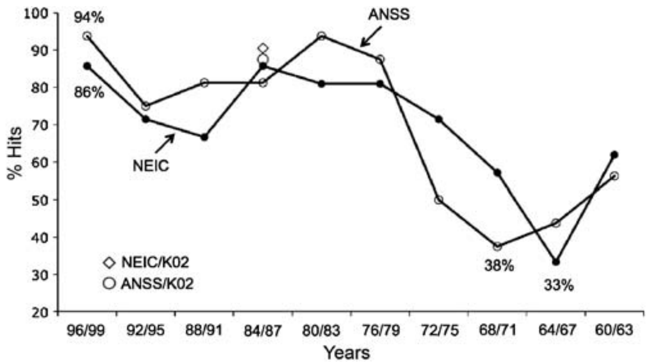

Figure 11. Percentage of hits as a function of the four-year before subcatalogs shown in Table 1, going back in time from the 1996–1999 subcatalog to 1960–1963 subcatalog. The larger open circle shows the result for the K02 before catalog (1984–1997) applied to the ANSS after catalog, and the open diamond shows the result for the K02 before catalog applied to the NEIC after catalog. Results for the K02 catalog are slightly different than for the 1984–1987 subcatalog downloaded for this study, due to different times that the data were downloaded from the ANSS and NEIC catalog search websites.

RELM forecasts because those forecasts are well documented and are compared and contrasted in a uniform manner in the RELM *SRL* issue. The forecasts presented in this section are provided here so that other researchers can compare our CS forecasts with published RELM models as new earthquakes occur within the RELM study region.

Using the cell size/grid structure of the RELM issue, forecasts are presented here for 13%, 18%, and 23% map areas (Fig. 12). These forecasts are based on a before catalog downloaded from the ANSS with $M \geq 4.0$ (1932–1998) and an after catalog, also from the ANSS, with $M \geq 5.0$ (1999–2008, including 47 events). Aftershocks were not removed from either the before or after catalog. For the lower-resolution (0.1° on a side) application of CS, with CS radii chosen to cover 13%, 18%, and 23% of the map area, the results are 31/47 (66%) hits, 42/47 (89%) hits, and 43/47 (91%) hits, respectively. For the higher-resolution application of CS, with CS radii again chosen to cover 13%, 18%, and 23% map area, the results are 32/47 (68%) hits, 38/47 (81%) hits, and 42/47 (89%) hits, respectively. In none of the cases shown in Figure 12 are the differences between the lower-resolution versus higher-resolution results statistically significant, again demonstrating that differences in resolution for the CS method do not appear to have an effect on the ability of the CS method to forecast locations of future earthquakes. Also shown in Figure 12 are confidence intervals for these forecasts. The confidence intervals are calculated for a binomial experiment model following the same procedure as that of Kafka (2007).

Discussion and Conclusions

None of the analyses discussed in our paper reveal any statistically significant evidence that the inclusion of rates and/or changes in rates in the forecast model improves the ability of the methods to forecast locations of future earthquakes beyond that of the least-astonishing CS hypothesis.

Although certainly not an exhaustive analysis of the questions addressed here, the results of this study do suggest the possibility that the ability to forecast locations of future earthquakes is not improved by including information about rates of seismicity and/or changes in rates of seismicity in binary (alarm-based) forecasts. Does this analysis, therefore, demonstrate that such rate-based information can be ruled out as possible indicators of locations of future earthquakes? We believe not. In this study we have only compared two examples of rate-dependent forecast models to the simpler CS model; both of those comparisons were in the same region, and the comparison involved converting the rate-based forecast models to binary (alarm-based) forecasts. It will require many more such hypothesis tests in many more regions (as well as other types of hypothesis tests) before it would be reasonable to conclude that proximity to past earthquakes is the only predictor of future earthquake locations.

It is intuitively reasonable to expect that patterns in rates of seismicity and/or changes in rates of seismicity at a given location are a reflection of the physical processes that eventually lead up to the occurrence of earthquakes. On the other hand, it is possible that, while patterns in rates of seismicity do reflect the physical processes leading up to an earthquake, any given pattern might be just as likely to indicate an impending large earthquake as to be an indicator of a lower probability of an impending earthquake. If that is indeed the case, then perhaps cellular seismology, simple as it is, may actually be a measure of all we can know about the future occurrence of earthquakes. Hard as it may be for seismologists (and the public?) to accept such a conclusion, the results of this study suggest that such a possibility needs to be considered.

In the meantime, we present cellular seismology as an appropriate choice for a standard of comparison for spatial forecast models. We propose that spatial forecast methods be tested against the cellular seismology method to ascertain whether or not they perform better than this least-astonishing hypothesis model.

Data and Resources

The Collaboratory for the Study of Earthquake Predictability (which may be accessed at http://www.cseptesting.org) is a virtual, distributed laboratory for the study of the predictability of earthquake rupture processes and how earthquake prediction experiments should be conducted and evaluated. Earthquake catalog data collected specifically for the analyses presented in this study were from the Advanced National Seismic System (ANSS) catalog (www.ncedc.org/anss), the National Earthquake Information Center (NEIC) catalog (http://earthquake.usgs.gov/regional/neic), and the Southern California Earthquake Center (SCEC) catalog (www.scec.org); all were last accessed in December

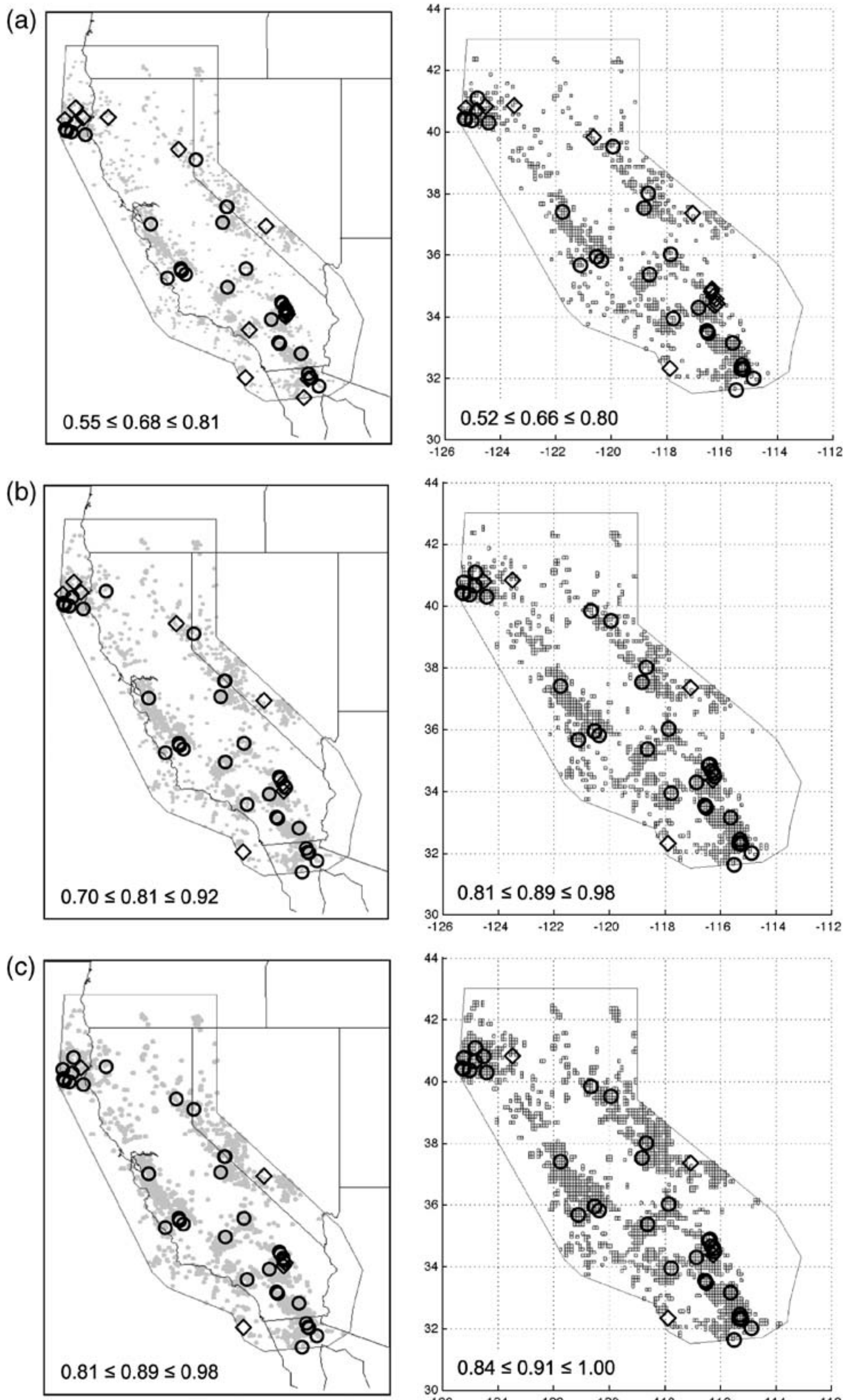

Figure 12. RELM CS forecasts for the high-resolution application of CS (left) and low-resolution application of CS (right). Open circles are hits, and open diamonds are misses. The before catalog is from ANSS with $M \geq 4.0$ (1932–1998), and the after catalog is also from ANSS with $M \geq 5.0$ (1999–2008). (a) 13% area; radius = 4.8 km, $32/47 = 68\%$ hits for the high-resolution case, and radius = 5.0 km, $31/47 = 66\%$ hits for the low-resolution case. (b) 18% area; radius = 6.2 km, $38/47 = 81\%$ hits for the high-resolution case, and radius = 6.6 km, $42/47 = 89\%$ hits for the low-resolution case. (c) 23% area; radius = 7.8 km, $42/47 = 89\%$ hits for the high-resolution case, and radius = 7.3 km, $43/47 = 91\%$ hits for the low-resolution case. Intervals in the lower left of the maps are 95% confidence intervals for forecasts.

2010. The before catalogs of the Kafka (2002) and Rundle *et al.* (2002) studies were both collected from the SCEC catalog.

Acknowledgments

Much of the research for this study was undertaken during a sabbatical leave granted to the first author (A. Kafka) by Boston College for the 2008–2009 academic year. We are grateful to our colleagues in the Boston College (BC) Department of Earth and Environmental Sciences, as well as the BC College of Arts and Sciences, for their support in providing the extended research and writing time necessary for us to complete this project. BSSA Editor A. Michael, reviewer V. Kossobokov, and anonymous reviewers provided helpful comments and critiques that led to improvements in this paper, in particular suggesting that the lower-resolution analysis would elucidate the extent to which differences in resolution, rather than more fundamental differences, might explain differences in success rates of the methods compared here.

Reference

Ebel, J. E., D. W. Chambers, A. L. Kafka, and J. A. Baglivo (2007). Non-Poissonian earthquake clustering and the Hidden Markov model as bases for earthquake forecasting in California, *Seismol. Res. Lett.* **78,** no. 1, 57–65.

Field, E. H. (2007). Overview of the Working Group for the Development of Regional Earthquake Likelihood Models, *Seismol. Res. Lett.* **78,** no. 1, 7–16.

Frankel, A. (1995). Mapping seismic hazard in the central and eastern United States, *Seismol. Res. Lett.* **66,** no. 4, 8–21.

Gardner, J. K., and L. Knopoff (1974). Is the sequence of earthquakes in southern California with aftershocks removed Poissonian?, *Bull. Seismol. Soc. Am.* **64,** 1363–1367.

Helmstetter, A., Y. Y. Kagan, and D. D. Jackson (2007). High-resolution time-independent grid-based forecast for $M \geq 5$ earthquakes in California, *Seismol. Res. Lett.* **78,** no. 1, 78–86.

Kafka, A. L. (2002). Statistical analysis of the hypothesis that seismicity delineates areas where future large earthquakes are likely to occur in the central and eastern United States, *Seismol. Res. Lett.* **73,** no. 6, 990–1001.

Kafka, A. L. (2007). Does seismicity delineate zones where future large earthquakes are likely to occur in intraplate environments?, Special Paper 425, in *Continental Intraplate Earthquakes: Science, Hazard, and Policy Issues*, S. Stein and S. Mazzotti (Editors), Geological Society of America, Boulder, Colorado 35–48, doi 10.1130/2007.2425(03).

Kafka, A. L., and J. E. Ebel (2007). Exaggerated claims about earthquake predictions, *Eos Trans. AGU* **88,** no. 1, 1, 6.

Kafka, A. L., and S. Z. Levin (2000). Does the spatial distribution of smaller earthquakes delineate areas where larger earthquakes are likely to occur?, *Bull. Seismol. Soc. Am.* **90,** no. 3, 724–738.

Rundle, J. B., K. F. Tiampo, W. Klein, and J. S. S. Martins (2002). Self-organization in leaky threshold systems: The influence of near-mean field dynamics and its implications for earthquakes, neurobiology, and forecasting, *Proc. Natl. Acad. Sci. Unit. States Am.* **99,** Suppl. 1, 2514–2521.

Rundle, J. B., K. F. Tiampo, and W. Klein (2007). Pattern informatics and cellular seismology: A comparison of methods, *Eos Trans. AGU* **88,** no. 24, 254.

Schorlemmer, D., M. C. Gerstenberger, S. Wiemer, D. D. Jackson, and D. A. Rhoades (2007). Earthquake likelihood model testing, *Seismol. Res. Lett.* **78,** no. 1, 17–27.

Weston Observatory
Department of Earth and Environmental Sciences
Boston College
381 Concord Rd.
Weston, Massachusetts 02493

Manuscript received 6 July 2009