Results of the Regional Earthquake Likelihood Models (RELM) test of earthquake forecasts in California

Ya-Ting Lee[a,b], Donald L. Turcotte[a,1], James R. Holliday[c], Michael K. Sachs[c], John B. Rundle[a,c,d], Chien-Chih Chen[b], and Kristy F. Tiampo[e]

[a]Geology Department, University of California, Davis, CA 95616; [c]Physics Department, University of California, Davis, CA 95616; [b]Graduate Institute of Geophysics, National Central University, Jhongli, Taiwan 320, Republic of China; [e]Department of Earth Sciences, University of Western Ontario, London, ON, Canada N6A 5B7; and [d]Santa Fe Institute, Santa Fe, NM 87501

Contributed by Donald L. Turcotte, August 19, 2011 (sent for review January 5, 2011)

The Regional Earthquake Likelihood Models (RELM) test of earthquake forecasts in California was the first competitive evaluation of forecasts of future earthquake occurrence. Participants submitted expected probabilities of occurrence of $M \geq 4.95$ earthquakes in $0.1° \times 0.1°$ cells for the period 1 January 1, 2006, to December 31, 2010. Probabilities were submitted for 7,682 cells in California and adjacent regions. During this period, 31 $M \geq 4.95$ earthquakes occurred in the test region. These earthquakes occurred in 22 test cells. This seismic activity was dominated by earthquakes associated with the $M = 7.2$, April 4, 2010, El Mayor–Cucapah earthquake in northern Mexico. This earthquake occurred in the test region, and 16 of the other 30 earthquakes in the test region could be associated with it. Nine complete forecasts were submitted by six participants. In this paper, we present the forecasts in a way that allows the reader to evaluate which forecast is the most "successful" in terms of the locations of future earthquakes. We conclude that the RELM test was a success and suggest ways in which the results can be used to improve future forecasts.

earthquake forecasting | forecast verification | earthquake clustering

Reliable short-term earthquake prediction does not appear to be possible at this time. This was confirmed by the failure to observe any precursory phenomena prior to the 2004 Parkfield earthquake (1). However, earthquakes do not occur randomly in space and time. Large earthquakes occur preferentially in regions where small earthquakes occur. Earthquakes on active faults occur quasiperiodically in time.

Earthquakes obey several scaling laws. One example is Gutenberg–Richter frequency-magnitude scaling (2). The cumulative number of earthquakes, N_c, with magnitudes greater than M in a region over a specified period of time is well approximated by the relation

$$\log N_c = a - bM, \qquad [1]$$

where b is a near universal constant in the range $0.8 < b < 1.1$ and a is a measure of the level of seismicity. Small earthquakes can be used to determine a, and Eq. **1** can be used to forecast the probability of occurrence of larger earthquakes.

An alternative approach to quantifying earthquake hazard is to specify the recurrence statistics of earthquakes on mapped faults. Geodetic observations can be used to determine rates of strain accumulation, and paleoseismic studies can be used to determine the occurrence of past earthquakes. A problem with this approach is that many damaging earthquakes do not occur on mapped faults.

A *pattern informatics* (PI) approach to earthquake forecasting has been proposed (3–5). In forecasting $M \geq 5$ earthquakes, a region is divided into a grid of $0.1° \times 0.1°$ subregions. The rates of seismicity in the subregions are studied to quantify anomalous behavior. Precursory changes that include either increases or decreases in seismicity are identified during a prescribed time interval. If changes exceed a prescribed threshold, hot spots are defined. The forecast is that future $M \geq 5$ earthquakes will occur in the hot-spot regions in a 10-y time window. Therefore, this is an alarm-based forecast. Utilizing the PI method, a forecast of California hot spots valid for the period 2000–2010 was given (3); 16 of the 18 earthquakes that occurred during the period 2000–2005 occurred in these hot-spot regions (6).

Another alternative forecasting technique is the relative intensity (RI) approach. The RI forecast is based on the direct extrapolation of the rate of occurrence of small earthquakes using Eq. **1**. Comparisons of these approaches have come to different conclusions regarding their validity (6, 7). These comparisons emphasize the difficulties in evaluating the performance of seismicity forecasts.

Extensive studies of earthquake hazards in California have been carried out (8). These studies quantified the relative risk of earthquakes in various parts of the state and specifically are used to set earthquake insurance premiums. Because extrapolations of past seismicity to establish risk play an important role, the working group for Regional Earthquake Likelihood Models (RELM) was established (9). Research groups were encouraged to submit forecasts of future earthquakes in California. The submissions were required by January 1, 2006, and the test period extended from January 1, 2006, until December 31, 2010.

The test region extended somewhat beyond the boundaries of the state as shown in Fig. 1. Earthquakes with magnitudes greater than $M = 4.95$ were to be forecast. Probabilities of occurrence of test earthquakes were required for 7,682 spatial cells with $0.1° \times 0.1°$ dimensions. These conditions for the RELM test were identical to those used for the PI forecast (3). However the RELM test was not a threshold (hot-spot) test. Participants were expected to submit a continuous range of earthquake probabilities for the 7,682 cells. Details of the RELM test are given in *Data and Methods*.

Results and Discussion

During the test period January 1, 2006, to December 31, 2010, there were 31 earthquakes with $M \geq 4.95$ in the test region (Table 1). The locations of these earthquakes are given in Figs. 1–4. These earthquakes occurred in 22 forecast cells. Association of earthquakes with cells is illustrated in Fig. 3. Further details regarding the test earthquakes are given in *Data and Methods*.

In this paper we consider only the forecasts of whether a test earthquake was expected to occur in the cells in which earthquakes actually occurred. These probabilities λ_{in} are given in Table 2 and are the probabilities that a $M \geq 4.95$ will occur in cell i during the test period. The probability λ_{in} is normalized so that the sum of the probabilities over all cells is 22, the number of cells

Author contributions: D.L.T., J.B.R., C.-C.C., and K.F.T. designed research; Y.-T.L., J.R.H., and M.K.S. performed research; Y.-T.L. and J.R.H. analyzed data; and D.L.T. and M.K.S. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence should be addressed. E-mail: dlturcotte@ucdavis.edu.

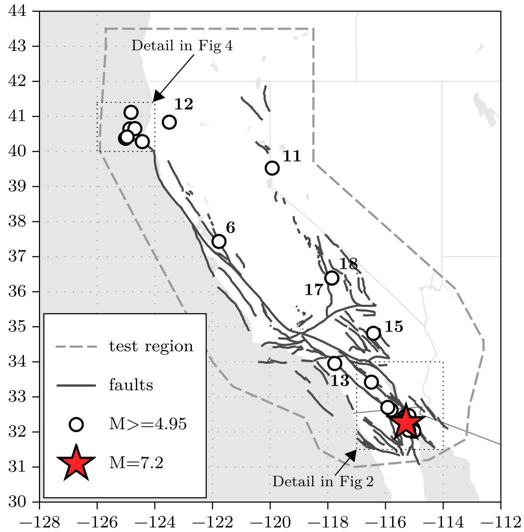

Fig. 1. Map of the test region, the coast of California, major faults, and the 31 earthquakes with $M \geq 4.95$ that occurred in the test region. The earthquakes are given in Table 1. Also shown are the square regions where large-scale maps are given in Figs. 2 to 4.

in which earthquakes actually occurred. A perfect forecast would have $\lambda_{in} = 1$ in each of these cells and $\lambda_{in} = 0$ in all other cells. Seven submissions of probabilities are given in Table 2. The details of the way in which the submitted probabilities λ_{im} were used to obtain the normalized probabilities λ_{in} are given in the RELM subsection of *Data and Methods*. Further details of the submitted forecasts are given in The Forecasts subsection of *Data and Methods*. It is also of interest to compare the submitted forecast probabilities with random (no skill) values. This has been given in Eq. **5** and is $\lambda_{inr} = 2.86 \times 10^{-3}$.

Table 1. Times of occurrence, locations, and magnitudes of the 31 earthquakes in the test region with $M \geq 4.95$ from January 1, 2006, until December 31, 2010

No.	Event time (universal time)		Lat.	Long.	M
1	2006/05/24	04:20:26.01	32.3067	−115.2278	5.37
2	2006/07/19	11:41:43.46	40.2807	−124.4332	5.00
3	2007/02/26	12:19:54.48	40.6428	−124.8662	5.40
4	2007/05/09	07:50:03.83	40.3745	−125.0162	5.20
5	2007/06/25	02:32:24.62	41.1155	−124.8245	5.00
6	2007/10/31	03:04:54.81	37.4337	−121.7743	5.45
7	2008/02/09	07:12:04.55	32.3595	−115.2773	5.10
8	2008/02/11	18:29:30.53	32.3272	−115.2568	5.10
9	2008/02/12	04:32:39.24	32.4475	−115.3175	4.97
10	2008/02/19	22:41:29.66	32.4325	−115.3130	5.01
11	2008/04/26	06:40:10.60	39.5253	−119.9289	5.00
12	2008/04/30	03:03:06.90	40.8358	−123.4968	5.40
13	2008/07/29	18:42:15.71	33.9530	−117.7613	5.39
14	2008/11/20	19:23:00.19	32.3288	−115.3318	4.98
15	2008/12/06	04:18:42.85	34.8133	−116.4188	5.06
16	2009/09/19	22:55:17.84	32.3707	−115.2612	5.08
17	2009/10/01	10:01:24.67	36.3878	−117.8587	5.00
18	2009/10/03	01:16:00.31	36.3910	−117.8608	5.19
19	2009/12/30	18:48:57.33	32.4640	−115.1892	5.80
20	2010/01/10	00:27:39.32	40.6520	−124.6925	6.50
21	2010/02/04	20:20:21.97	40.4123	−124.9613	5.88
22	**2010/04/04**	**22:40:42.15**	**32.2587**	**−115.2872**	**7.20**
23	2010/04/04	22:50:17.08	32.0972	−115.0467	5.51
24	2010/04/04	23:15:14.24	32.3000	−115.2595	5.43
25	2010/04/04	23:25:06.95	32.2462	−115.2978	5.38
26	2010/04/05	00:07:09.07	32.0180	−115.0172	5.32
27	2010/04/05	03:15:24.46	32.6282	−115.8062	4.97
28	2010/04/08	16:44:25.92	32.2198	−115.2760	5.29
29	2010/06/15	04:26:58.48	32.7002	−115.9213	5.72
30	2010/07/07	23:53:33.53	33.4205	−116.4887	5.43
31	2010/09/14	10:52:18.00	32.0485	−115.1982	4.96

The $M = 7.2$ El Mayor–Cucapah earthquake is in bold.

There are a variety of ways in which cell forecasts can be scored relative to each other. Three of these are given in Table 3. The three are as follows:

1. The number of submitted forecasts $N_{\lambda\,max}$ that had the highest cell probabilities λ_{in} for the 22 cells in which earthquakes occurred. By this method of scoring the Holliday et al. forecast was the best with $N_{\lambda\,max} = 8$; the second best was the Wiemer and Schorlemmer forecast with $N_{\lambda\,max} = 6$.
2. The mean forecast cell probabilities $\bar{\lambda}_{in}$ for all 22 cells in which earthquakes occurred. By this method of scoring the Helmstetter et al. forecast was the best with $\bar{\lambda}_{in} = 2.84 \times 10^{-2}$; the second best was the Wiemer and Schorlemmer forecast with $\bar{\lambda}_{in} = 2.66 \times 10^{-2}$. It is of interest to note that these values were about a factor of 10 better than the random (no skill) forecast $\lambda_{inr} = 2.86 \times 10^{-3}$.
3. Likelihood (L_{test}) test results for the 22 cells in which earthquakes occurred. By this measure the Helmstetter et al. forecast was the best with $L = -114$ with the Holliday et al. and Ebel et al. forecasts the next best with $L = -123$.

It is clear that different accepted methods of scoring rank the forecasts differently. Further discussion of the scoring is given in *Data and Methods*.

An important question is, what have we learned from the RELM test? In terms of seismic hazard mitigation and insurance premiums, forecasts of the locations of future large earthquakes are required. As an evaluation of alternative forecast methods, the RELM test has clearly been a success. The best forecasts are about an order of magnitude better than a random forecast.

In summary, we enumerate the project successes and limitations as follows:

1. Forecasts of earthquakes with $M \geq 5$ can be successfully evaluated in a relatively short time period (5 y) in a seismically active area.
2. Submission of forecast rates for $0.1° \times 0.1°$ cells is appropriate.
3. Forecast rates should be made for the cumulative number of earthquakes expected to exceed some specified minimum magnitude in each spatial cell.
4. In evaluating cell probabilities, unless explicitly testing for multiple earthquakes, the occurrence of multiple earthquakes in a cell should not be considered.
5. Forecasts of the numbers and locations of earthquakes are basically independent following the approach suggested in this paper.
6. Forecasts of the locations of earthquakes are independent of whether the forecasts are made with or without aftershocks.
7. There is no optimum approach to the scoring of results. Some scoring methods emphasize successes and others penalize low forecast rates. Alternative scoring approaches should be used.
8. Results should also be evaluated using an alarm-based approach. This can be done utilizing RELM continuum forecasts. The developed scoring techniques used for weather (specifically tornadoes) can then be applied. Results can be scored using relative operating characteristic (10), or similar diagrams.

Data and Methods

RELM. The Working Group on California Earthquake Probabilities was established to evaluate the potential for large earthquakes in California, and studies were published in 1988, 1990, 1995, 2003, and 2007 (8). These studies have concentrated on the probabilities of earthquake occurrence on mapped faults in California. In order to aid these assessments, the Southern California Earthquake Center formed the working group for RELM in 2000 (9). Research groups were encouraged to submit

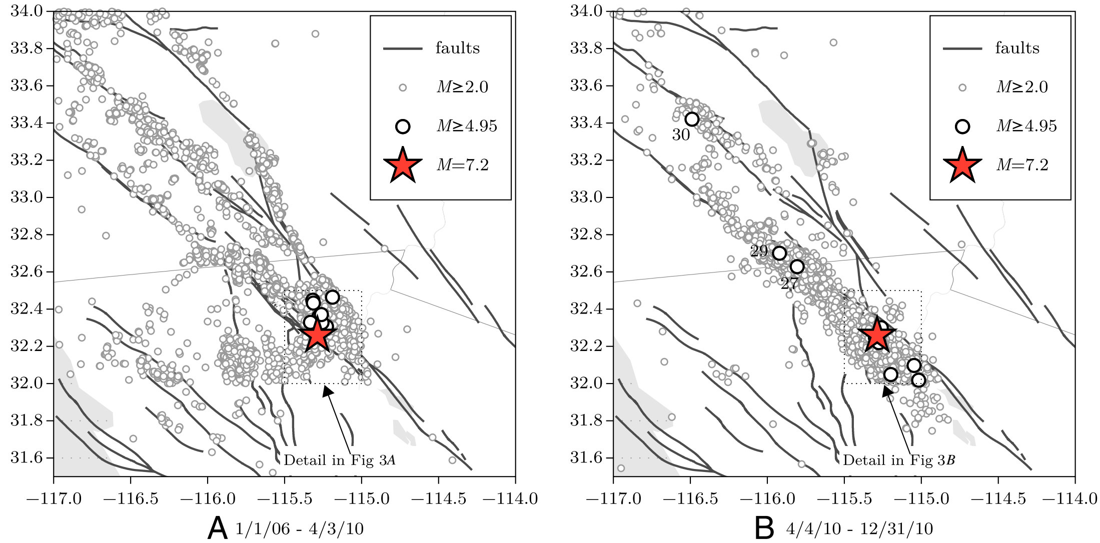

Fig. 2. Map of the southeast region around the epicenter of the $M = 7.2$ El Mayor–Cucapah earthquake that occurred on April 4, 2010 (event #22 in Table 1, shown as a star). (*A*) Earthquakes during the period January 1, 2006, through April 3, 2010. (*B*) Earthquakes during the period April 4, 2010, through December 31, 2010 (includes aftershocks). Included are the test earthquakes given in Table 1 as well as background earthquakes with $M \geq 2.0$. More details in the square region are given in the larger-scale maps in Fig. 3.

forecasts of future earthquakes in California. At the end of the test period, the forecasts would be compared with the actual earthquakes that occurred.

The ground rules for the RELM test were as follows:

The test region to be studied was the state of California; however, the selected region extended somewhat beyond the boundaries of the state as shown in Fig. 1.

The objective was to forecast the largest earthquakes for which a reasonable number could be expected to occur in a reasonable time period. A 5-y time period for the test was selected extending from January 1, 2006, to December 31, 2010. Earthquakes with $M \geq 5$ were to be forecast. This magnitude cutoff was chosen because at least 20 $M \geq 5$ earthquakes could be expected. For $M \geq 6$, only about 2 would be expected so the 5-y period would be much too short. The applicable magnitudes were taken from the Advanced National Seismic System online catalog http://www.ncedc.org/anss/anss-detail.html.

Participants were required to submit the expected probabilities of occurrence of earthquakes for the test region. In order to do this, the test region was subdivided into 7,682 spatial cells with dimensions $0.1° \times 0.1°$ (approximately 10 km × 10 km). These spatial cells were further divided into 41 magnitude bins: $4.95 \leq M < 5.05$, $5.05 \leq M < 5.15$, $5.15 \leq M < 5.25, \ldots,$ $8.85 \leq M < 8.95$, $8.95 \leq M < \infty$. Participants were required to specify the probability of occurrence, λ_{im}, for each spatial-magnitude bin i for the 5-y test period. In this paper, we sum over the magnitude bins in each spatial cell to give the forecast probability of occurrence λ_i of an $M > 4.95$ earthquake in cell i during the test period

$$\lambda_i = \sum_{m=4.95}^{\infty} \lambda_{\mathrm{im}}. \qquad [2]$$

The sum of the λ_i over all cells is the total number of earthquakes N_e forecast to occur during the test period

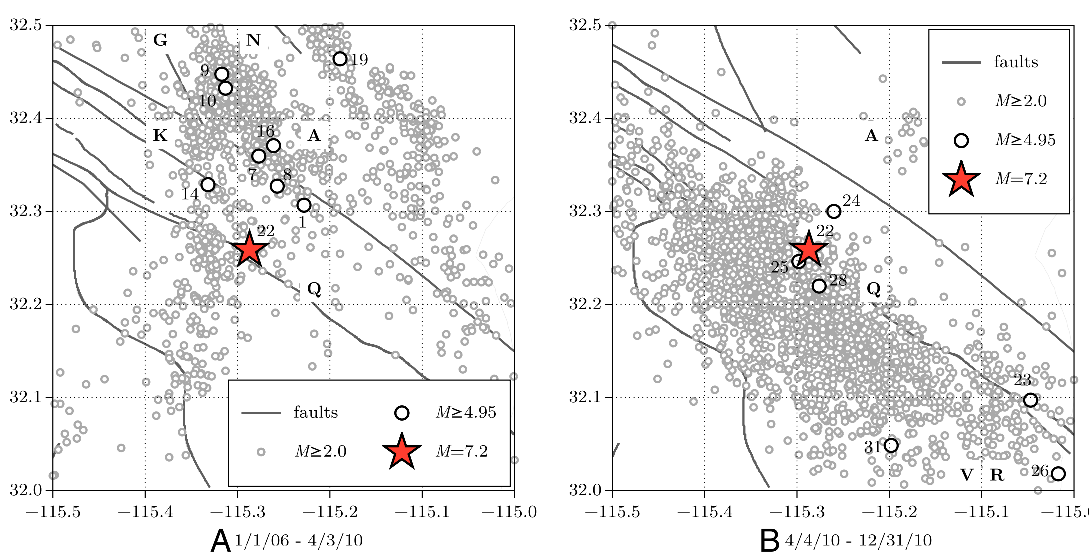

Fig. 3. Map of the region in the immediate vicinity of the epicenter of the $M = 7.2$ El Mayor–Cucapah earthquake. (*A*) Earthquakes during the period January 1, 2006, through April 3, 2010. (*B*) Earthquakes during the period April 4, 2010, through December 13, 2010. Included are the test earthquakes given in Table 1, as well as background earthquakes with $M \geq 2.0$. The association of lettered $0.1° \times 0.1°$ cells in which earthquakes occurred with the numbered earthquakes is illustrated.

Table 2. Normalized probabilities of occurrence λ_{in} of an earthquake with $M \geq 4.95$ for the 22 cells in which earthquakes occurred during the test period

Cell ID	EQ ID	B and L	Ebel	Helm.	Holl.	W-C	W-G	W and S
-A-	1,7,8,16,24	1.99e-2	2.20e-2	1.17e-1	3.32e-2	1.87e-2	1.28e-2	**1.24e-1**
-B-	2	1.41e-2	3.40e-2	**7.20e-2**	3.32e-2	1.08e-3	1.86e-3	4.99e-2
-C-	3	7.40e-3	6.59e-3	7.41e-3	**3.32e-2**	8.93e-4	1.54e-3	7.91e-3
-D-	4	3.54e-2	3.29e-2	**6.97e-2**	3.32e-2	9.50e-4	1.64e-3	3.59e-2
-E-	5	**7.23e-3**	1.10e-3	2.29e-3	9.72e-5	9.25e-4	1.59e-3	1.58e-7
-F-	6	9.37e-3	2.85e-2	3.07e-2	3.32e-2	5.29e-3	8.12e-3	**4.55e-2**
-G-	9,10	9.11e-3	5.49e-3	2.55e-2	**3.32e-2**	2.25e-2	1.27e-2	2.38e-2
-H-	11	3.42e-4	**5.49e-3**	9.15e-4	1.62e-4	3.77e-4	6.49e-4	2.06e-4
-I-	12	2.14e-3	1.10e-3	3.65e-3	2.05e-4	1.14e-3	1.96e-3	**9.89e-3**
-J-	13	1.68e-3	8.78e-3	1.11e-2	**3.32e-2**	8.11e-3	5.12e-3	1.13e-2
-K-	14	3.12e-2	2.20e-2	3.30e-2	3.32e-2	1.93e-2	1.17e-2	**5.90e-2**
-L-	15	2.07e-3	5.49e-3	**6.93e-3**	3.32e-3	4.80e-3	5.45e-3	2.64e-3
-M-	17,18	1.74e-3	2.20e-3	5.78e-3	**3.32e-2**	3.88e-3	4.61e-3	5.38e-4
-N-	19	**5.83e-2**	6.59e-3	1.49e-2	3.32e-2	1.65e-2	1.23e-2	7.44e-3
-O-	20	1.25e-2	1.43e-2	9.45e-3	**3.32e-2**	9.30e-4	1.60e-3	1.62e-2
-P-	21	6.48e-3	3.29e-2	2.71e-2	**3.32e-2**	9.03e-4	1.55e-3	7.46e-3
-Q-	22,25,28	2.88e-2	2.20e-2	2.84e-2	3.32e-2	1.66e-2	1.30e-2	**5.23e-2**
-R-	23,26	**3.06e-2**	1.54e-2	1.43e-2	1.73e-4	1.78e-2	1.38e-2	1.58e-2
-S-	27	2.13e-2	5.49e-3	1.26e-2	**3.32e-2**	9.55e-3	7.93e-3	1.19e-2
-T-	29	1.83e-2	1.32e-2	2.43e-2	3.32e-2	6.35e-3	3.90e-3	**4.99e-2**
-U-	30	1.26e-2	3.07e-2	**1.03e-1**	3.32e-3	1.61e-2	5.47e-3	5.16e-2
-V-	31	6.76e-3	1.54e-2	5.55e-3	**3.32e-2**	1.54e-2	1.43e-2	2.64e-3

The association of cell id's (A–V) with the earthquake id's (1–31) from Table 1 is illustrated in Fig. 1. Seven submitted forecasts are given: (1) Bird and Liu (B and L), (2) Ebel et al. (Ebel), (3) Helmstetter et al. (Helm.), (4) Holliday et al. (Holl.), (5) Ward combined (W-C), (6) Ward geodetic (W-G), and (7) Wiemer and Schorlemmer (W and S). The highest (best) probabilities are in bold.

$$N_e = \sum_{i=1}^{7682} \lambda_i, \qquad [3]$$

where N_c is the total number of cells. The total number of forecast earthquakes, N_e, directly influences the distribution of individual cell probabilities, λ_i: Doubling N_e would double each λ_i, therefore increasing the likelihood of a successful forecast. In order to overcome this problem, we rescale each forecast to take into account the actual number of cells in which earthquakes occurred during the test period N_{ce}. The normalized cell probabilities, λ_{in}, are defined by the relation

$$\lambda_{in} = \frac{N_{ce}}{N_e} \lambda_i. \qquad [4]$$

The forecast values of λ_{in} are a direct measure of the success of a forecast in locating future earthquakes.

Participants could submit forecasts that included all earthquakes in the test region as well as forecasts that excluded aftershocks. Because of our rescaling approach, we eliminate the difference between these two types of forecast. This is desirable because–as we will show–it is difficult to define which earthquakes are aftershocks. The normalized rates λ_{in} are equal for the two forecasts with and without aftershocks.

The Earthquakes. During the test period January 1, 2006, to December 31, 2010, there were $N_e = 31$ earthquakes in the test region with $M \geq 4.95$. The times of occurrence, locations, and magnitudes of these earthquakes are given in Table 1. The locations of the test earthquakes are also shown in Figs. 1–4. The earthquakes are identified by the event numbers given in Table 1.

The major earthquake that occurred during the test period was the $M = 7.2$ El Mayor–Cucapah earthquake on April 4, 2010 (event #22 in Table 1). This earthquake was on the plate boundary between the North American and Pacific plates. The epicenter was about 50 km south of the Mexico–United States border, and the aftershocks indicate a rupture zone with a length of about 75 km. Both the epicenter and the aftershock sequence are illustrated in Fig. 2.

We first discuss the test earthquakes in the region of the El Mayor–Cucapah earthquake. The earthquakes within a 0.5°× 0.5° region centered on the epicenter are illustrated in Fig. 3. The El Mayor earthquake and the test earthquakes that occurred later, April 4, 2010, to December 31, 2010, are given in Fig. 3*B*. Events 23, 24, 25, 26, 28, and 31 are certainly aftershocks. The El Mayor earthquake and the test earthquakes that occurred earlier, January 1, 2006, to April 3, 2010, are given in Fig. 3*A*. Events 1, 7, 8, 9, 10, 14, 16, and 19 constitute a precursory swarm of eight test earthquakes in this region in the magnitude range 4.97 to 5.80,

Table 3. Comparisons of the forecasts

	$\lvert N_{\lambda \max} \rvert$	$\lvert \bar{\lambda}_{in} \rvert$	L_{test}
Bird and Liu	3	1.53e-2	−126
Ebel et al.	1	1.51e-2	−123
Helmstetter et al.	4	**2.84e-2**	**−114**
Holliday et al.	**8**	2.45e-2	−123
Ward combined	0	8.55e-3	−141
Ward geodetic	0	6.53e-3	−141
Wiemer and Schor.	6	2.66e-2	−129

Column 1: The number of maximum cell probabilities $N_{\lambda \max}$. Column 2: The mean cell probabilities forecast $\bar{\lambda}_{in}$. Column 3: The maximum likelihood scores. The best scores in each category are in bold.

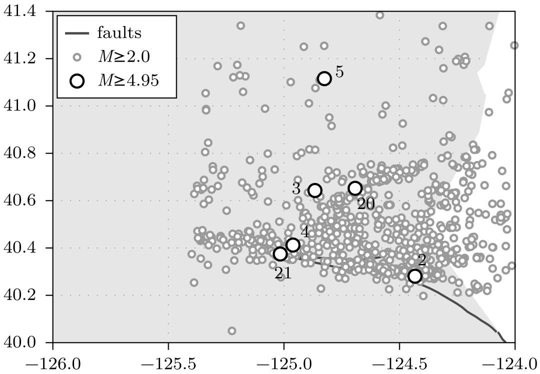

Fig. 4. Map of the northwest region near Cape Mendocino. Test earthquakes given in Table 1 are shown as well as background earthquakes with $M \geq 2.0$.

including four in the 10 d period between February 9 and February 19, 2008 (events 7–10). These events are located some 5 to 20 km north of the subsequent epicenter of the El Mayor–Cucapah earthquake and lie outside the primary aftershock region of that event, as illustrated in Fig. 3*A*. This swarm of earthquakes certainly cannot be considered foreshocks, due to their relatively small magnitudes and early occurrence, but may represent a seismic activation.

The locations of the earthquakes given in Table 1 identify the $0.1° \times 0.1°$ cells in which the earthquakes occurred. These cells are illustrated in Fig. 3. Cells in which earthquakes occurred are identified by capital letters. Earthquakes in Fig. 3*A* occurred in cells A, G, N, K, and Q. Earthquakes in Fig. 3*B* occurred in cells A, Q, R, and V. The association of earthquake event numbers with cell letters is given in Table 2. The occurrence of five test earthquakes in cell A is not surprising because this is the Cerra Prieto geothermal area that is recognized as having a high level of seismic activity.

We next turn to the somewhat larger region ($3.0° \times 2.5°$) illustrated in Fig. 2. The El Mayor earthquake and the test earthquakes that occurred later, April 4, 2010, to December 31, 2010, are given in Fig. 2*B*. The aftershock region of the El Mayor–Cucapah earthquake is clearly illustrated, and events 27 and 29 are certainly aftershocks. Event 30 may or may not be an aftershock. The El Mayor earthquake and the test earthquakes that occurred earlier, January 1, 2006, to April 3, 2010, are given in Fig. 3*A*. No test earthquakes occurred outside the smaller region considered in Fig. 3*A*.

We next consider the $2° \times 1.4°$ region adjacent to Cape Medocino, illustrated in Fig. 4. Six test earthquakes occurred in this region (events 2, 3, 4, 5, 20, and 21) in the magnitude range 5.0 to 6.5. This is a region of high seismicity, and this concentration of events is expected. Event 21 may or may not be an aftershock of event 20.

There were seven test earthquakes that occurred outside of the regions considered above. These are illustrated in Fig. 1, and their magnitudes ranged from 5.0 to 5.45. The pair of earthquakes #17 and #18 is interesting. It is very likely that the $M = 5.0$ earthquake on October 1, 2009, was a foreshock of the $M = 5.19$ earthquake on October 3, 2009.

The Forecasts. The submitted forecasts have been discussed in some detail (9, 11). The 19 forecasts submitted by eight groups are available on the RELM web site (http://relm.cseptesting.org/). In order to have a common basis for comparison, we consider only forecasts that cover the entire test region. Nine forecasts were submitted that gave forecast probabilities, λ_{im}, for $M \geq 4.95$ earthquakes in 0.1 magnitude bins during the 5-y test period for all $N_c = 7{,}682$ $0.1° \times 0.1°$ cells. We then converted the forecast binned probabilities λ_{im} to cumulative probabilities λ_i that an earthquake with $M \geq 4.95$ would occur in cell i during the test period using Eq. **2**. Taking the actual number of cells in which earthquakes occurred to be $N_{ce} = 22$ and the total number of earthquakes forecast in each submission N_e using Eq. **3**, we obtained the normalized forecast probabilities λ_{in} using Eq. **4**. The normalized forecast probabilities λ_{in} for each of the seven submissions are given in Table 2 for the $N_{ce} = 22$ cells in which an earthquake occurred. A perfect normalized forecast in which only the 22 cells were forecast to have earthquakes would have $\lambda_{\mathrm{in}} = 1$ in each of the 22 cells. A random normalized forecast in which all $N_c = 7{,}682$ cells were given equal probabilities would have

$$\lambda_{\mathrm{inr}} = \frac{N_{ce}}{N_c} = \frac{22}{7682} = 2.86 \times 10^{-3}. \qquad \textbf{[5]}$$

The submitted forecast probabilities in Table 2 have a wide range of values from $\lambda_{\mathrm{in}} = 1.58 \times 10^{-7}$ to $\lambda_{\mathrm{in}} = 1.24 \times 10^{-1}$.

The submitted forecasts are based on a variety of approaches. The Bird and Liu forecast (12) was based on a kinematic model of neotectonics. The Ebel et al. forecast (13) was based on the average rate of $M \geq 5$ earthquakes in $3° \times 3°$ cells for the period 1932 to 2004. The Helmstetter et al. forecast (14) was based on the extrapolation of past seismicity. The Holliday et al. forecast (15) was based on the extrapolation of past seismicity using a modification of the PI technique. Ward (16) submitted two forecasts that cover the entire test region. The first was a geodetic forecast based on Global Positioning System velocities for the test region. The second was a composite forecast based on seismic and geological datasets in addition to the geodetic data. The Wiemer and Schorlemmer forecast (17) was based on the asperity-based likelihood model (ALM).

We now discuss the Holliday et al. (15) forecast in somewhat greater detail. The basis of this RELM forecast followed the format introduced in the PI forecast methodology (3, 5). The magnitude range $M \geq 5$ and the cell dimensions $0.1° \times 0.1°$ were the same. However, the PI method was alarm based. Earthquakes were forecast to either occur or not occur in specified regions (hot spots) in a specified time period. In the PI-based RELM forecast, all hot-spot cells are given equal probabilities of an earthquake. For the normalized values in Table 2, $\lambda_{\mathrm{in}} = 3.32 \times 10^{-2}$. Instead of being alarm based, the RELM test was based on probabilities of occurrence of an earthquake in each cell in the test region. This required a continuous assessment of risk rather than a binary, alarm-based assessment. To do this, the Holliday et al. forecast introduced a uniform probability of occurrence for hot-spot regions and added smaller probabilities for non-hot-spot regions based on the RI of seismicity in the region.

Forecast Evaluations. Because the forecasts are for specific $0.1° \times 0.1°$ cells, it is necessary to consider how to handle the forecasts when more than one earthquake occurs in a cell. In our analysis a cell in which more than one earthquake occurred is treated the same as a cell in which only one earthquake occurred. For the test earthquakes given in Table 1, events 1, 7, 8, 16, and 24 occurred in the same cell, and similarly for events 9 and 10, events 17 and 18, events 22, 25, and 28, and events 23 and 26. This multiplicity is shown in Table 2. Thus, we will consider forecasts made for 22 cells.

The results given in Table 2 can be used to compare the forecast probabilities for each of the cells in which earthquakes occurred. The highest probabilities are shown in bold. Clearly there are many ways in which to evaluate the results of the forecasts. There is a trade-off between good forecasts with large λ_{in} and poor forecasts with small λ_{in}. We first consider the forecasts that had the highest forecast probabilities. The Holliday et al. forecast had the largest λ_{in} for 8 of the 22 cells in which (target) earthquakes occurred. The Wiemer and Schorlemmer forecast had 6 of the largest λ_{in}. Helmstetter et al. had 4 of the largest λ_{in}. Finally, the Bird and Liu forecast had 3 of the largest λ_{in}. These values are also given in Table 3. The range of the highest normalized cell probabilities ranged from $\lambda_{\mathrm{in}} = 2.29 \times 10^{-2}$ for event 1 to $\lambda_{\mathrm{in}} = 1.05 \times 10^{-3}$ for event 11.

It is also of interest to compare the mean cell forecast probabilities for the 22 cells in which earthquakes occurred. These values $\bar{\lambda}_{\mathrm{in}}$ are given in Table 3. The Helmstetter et al. forecast had the highest $\bar{\lambda}_{\mathrm{in}} = 2.84 \times 10^{-2}$, the Wiemer and Schorlemmer forcast had $\bar{\lambda}_{\mathrm{in}} = 2.66 \times 10^{-2}$, and the Holliday et al. forecast had $\bar{\lambda}_{\mathrm{in}} = 2.45 \times 10^{-2}$. The Helmstetter et al. forecast did the best in an average sense but did relatively poorly in providing the best cell forecasts. It should be noted that the best average forecast $\bar{\lambda}_{\mathrm{in}} = 2.84 \times 10^{-2}$ is one order of magnitude better than the random (no skill) forecast $\lambda_{\mathrm{inr}} = 2.86 \times 10^{-3}$.

A complex series of statistical tests based on maximum likelihood was proposed (18, 19) to simultaneously evaluate both N_e

and λ_i for each forecast. This approach was utilized to evaluate the forecasts after the first 2½ y of the 5-y test period (11, 20). In this paper, we carry out a direct evaluation of the forecasts for the entire 5–y period. In Table 3 we give the likelihood (L_{test}) test results for the forecast probabilities given in Table 2. The best score is the least negative so that the Helmstetter et al. forecast has the best score.

As noted above, the Holliday et al. forecast is primarily a threshold (hot spot) forecast. The PI method was used to determine the cells in which earthquakes were most likely to occur (hot spots). In the normalized cell forecasts given in Table 2, these cells had forecast probabilities $\lambda_{\text{in}} = 3.32 \times 10^{-2}$ and consisted of 8.3% of the total area of the test region (637 of the 7,682 cells). Of the 22 cells in which earthquakes occurred, 17 occurred in hot-spot cells. In 8 of the 17 cells, the normalized forecast cell probabilities given by the Holliday et al. forecast were the highest.

ACKNOWLEDGMENTS. Y.T.L. is grateful for research support from both the National Science Council and the Institute of Geophysics (National Central University). J.R.H. and J.B.R. have been supported by National Aeronautics and Space Administration Grant NNXO8AF69G.

1. Bakun WH, et al. (2005) Implications for prediction and hazard assessment from the 2004 Parkfield earthquake. *Nature* 437:969–974.
2. Gutenberg B, Richter CF (1954) *Seismicity of the Earth and Associated Phenomena* (Princeton Univ Press, Princeton, NJ).
3. Rundle JB, Tiampo KF, Klein W, Martins JSS (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 USA* 99(Suppl. 1):2514–2521.
4. Rundle JB, Turcotte DL, Shcherbakov R, Klein W, Sammis C (2003) Statistical physics approach to understanding the multiscale dynamics of earthquake fault systems. *Rev Geophys* 41(4):1019.
5. Tiampo KF, Rundle JB, McGinnis S, Klein W (2002) Pattern dynamics and forecast methods in seismically active regions. *Pure Appl Geophys* 159:2429–2467.
6. Holliday JR, Nanjo KZ, Tiampo KF, Rundle JB, Turcotte DL (2005) Earthquake forecasting and its verification. *Nonlin Process Geophys* 12:965–977.
7. Zechar JD, Jordan TH (2008) Testing alarm-based earthquake predictions. *Geophys J Int* 172:715–724.
8. Field EH (2007) A summary of previous working groups on California earthquake probabilities. *Seis Soc Am Bull* 97:1033–1053.
9. Field EH (2007) Overview of the working group for the development of regional earthquake likelihood models (RELM). *Seis Res Lett* 78:7–16.
10. Jolliffe IT, Stephenson DB (2003) *Forecast Verification* (John Wiley, Chichester, UK).
11. Schorlemmer D, et al. (2010) First results of the regional earthquake likelihood models experiment. *Pure Appl Geophys* 167:859–876.
12. Bird P, Liu Z (2007) Seismic hazard inferred from tectonics: California. *Seis Res Lett* 78:37–48.
13. Ebel JE, Chambers DW, Kafka AL, Baglivo JA (2007) Non-Poissonian earthquake clustering and the hidden Markov model as bases for earthquake forecasting in California. *Seis Res Lett* 78:57–65.
14. Helmstetter A, Kagan YY, Jackson DD (2007) High-resolution time-independent grid-based for $m \geq 5$ ecast for earthquakes in California. *Seis Res Lett* 78:78–86.
15. Holliday JR, et al. (2007) A RELM earthquake forecast based on pattern informatics. *Seis Res Lett* 78:87–93.
16. Ward SN (2007) Methods for evaluating earthquake potential and likelihood in and around California. *Seis Res Lett* 78:87–93.
17. Wiemer S, Schorlemmer D (2007) ALM: An asperity-based likelihood model for California. *Seis Res Lett* 78:134–140.
18. Schorlemmer D, Gerstenberger MC, Wiemer S, Jackson DD, Rhoades DA (2007) Earthquake likelihood model testing. *Seis Res Lett* 78:17–29.
19. Schorlemmer D, Gerstenberger MC (2007) RELM testing center. *Seismol Res Lett* 78:31–36.
20. Zechar JD, Gierstenberger MC, Rhodes DA (2010) Likelihood based tests for evaluating space-rate-magnitude earthquake forecasts. *Seis Soc Am Bull* 100:1184–1195.