Rate-and-State California Earthquake Forecasts: Resolving Stress Singularities

Anne Strader, David D. Jackson.
GfZ-Potsdam, University of California, Los Angeles

Abstract

In previous studies, we confirmed an association between static Coulomb stress change (ΔCFF) and earthquake location in southern California, which indicated potential in prospective earthquake forecasting.

We prospectively tested a time-dependent alarm-based rate-and-state stress evolution-based forecast against a smoothed seismicity model for aftershocks following the M7.1 Hector Mine earthquake.

We derive empirical earthquake probability distributions from quantifying ΔCFF or conditional intensity, and observing the fraction of cells containing at least one earthquake for each magnitude.

In order to include areas with little to no seismicity in our forecast evaluation, we modified the stress inversion method proposed by Dieterich (2000) to use the rate of cumulative Poissonian probability as observed seismicity, instead of the rate of the number of earthquakes, when calculating stress evolution.

To determine how the impact of magnitude completeness and the effectiveness of an inverse rate-and-state earthquake forecast, we test how the ΔCFF and resulting empirical cell activation probability distributions vary for four lower magnitude thresholds: Mw 2.0, 2.5, 3.0 and 3.5.

To test the impact of magnitude/ΔCFF outside of the study area, we vary the window in which earthquakes are included in likelihood tests, while the area within which earthquakes are included as stress sources remains constant.

At the 95% confidence interval, both the rate-and-state ΔCFF-based and smoothed seismicity-based forecasts on the test set, due to underestimating the number of events following a magnitude threshold.

The stress inversion concentrated stress perturbations in areas that were much smaller than the actual aftershock distribution, and failed to represent the stress distribution that would have resulted from the forward model.

We defined the southern California study area similarly to the area used by Deng and Sykes (1997), when determining ΔCFF-based cell activation probabilities, we considered the entire study area (area 1).

When conducting likelihood tests, we defined three additional areas (Figure 1), outside of which earthquakes were not considered in forecast evaluation.

During the learning period, defined as the 180-day time interval just before the 16/10/1999 M7.1 Hector Mine earthquake, we used ΔCFF variations over 30-day time windows to calculate the probability of each cell containing at least one earthquake (cell “activation”) during the subsequent 30-day time windows.

We defined the target period as the 180-time interval following the Hector Mine earthquake. Stress steps were calculated at the beginning of each of the 30-day time windows.

Four lower magnitude thresholds were defined when calculating stress steps over 30-day intervals during the learning and target periods: Mw 2.0, 2.5, 3.0 and 3.5.

We used the “number, ” or “N”-test to determine whether the observed number of filled zones comes weighted by the empirical probability from either hypothesis.

Alternate hypothesis: ΔCFF-based earthquake forecasts significantly improve upon seismicity-based forecasts.

Figure 4: The southern California study area used in the Deng and Sykes (1997) study, displaying the locations of major faults and the Hector Mine earthquake epicenter. Four areas used to evaluate the rate and state and smoothed seismicity forecasting are outlined (area 1-4). The earthquake location bin size was 0.1˚x0.1˚ cell containing stress steps and subsequent cell activation probability distributions.

Seismicity Rate Distributions/Probabilities

Assuming observed seismicity rates define stress evolution throughout the study area, earthquake distributions may be inverted to obtain ΔCFF evolution (Figure 2a) from seismic stress theory (Dieterich, 2000).

where ΔCFF is the Coulomb stress step. A is a dimensionless, constitutive parameter, is the effective normal stress, 5, and 5r are the total and tectonic stress rates, respectively, N1 and N2 are the numbers of earthquakes during time intervals t1 and t2, respectively.

We divided the resulting ΔCFF steps and smoothed seismicity distribution (Figure 3a) during the learning period into quantiles and calculated the average number of cells within each quantile containing at least one earthquake, from which we derived empirical cell activation probability distributions applicable to the target period (Figure 2b).

At the 95% confidence interval, both the rate-and-state ΔCFF-based and smoothed seismicity-based forecasts fail to reliably forecast the areas and times of aftershocks following the Hector Mine earthquake.

Using a two-tailed test, we defined the p-value as the percentage of synthetic seismicity distribu- tions that are weighted by the empirical probability from either hypothesis.

Alternate hypothesis: ΔCFF-based earthquake forecasts significantly improve upon seismicity-based forecasts.

We define our null and alternate hypotheses as follows:

- Null hypothesis: Earthquake forecasts based on rate-and-state ΔCFF evolution do not indica- te future earthquake locations and times more effectively than seismicity-based forecasts.

- Alternate hypothesis: ΔCFF-based earthquake forecasts significantly improve upon seismicity-based forecasts.

Likelihood Tests

For each spatiotemporal cell, we calculate the probability that the cell will contain at least one earthquake, based on empirical seismicity and ΔCFF or smoothed seismicity distributions.

- We define null and alternate hypotheses as follows:

  - Null hypothesis: Earthquake forecasts based on rate-and-state ΔCFF evolution do not indica- te future earthquake locations and times more effectively than seismicity-based forecasts.
  - Alternate hypothesis: ΔCFF-based earthquake forecasts significantly improve upon seismicity-based forecasts.

We define our null and alternate hypotheses as follows:

- Null hypothesis: Earthquake forecasts based on rate-and-state ΔCFF evolution do not indica- te future earthquake locations and times more effectively than seismicity-based forecasts.
  - Alternate hypothesis: ΔCFF-based earthquake forecasts significantly improve upon seismicity-based forecasts.

Number of Filled Zones (N) Test

We used the “number, ” or “N”-test to determine whether the observed number of filled zones during the target period was sufficiently similar to what each forecast would construct.

- For both ΔCFF and smoothed seismicity distributions, we generated 1000 synthetic seismicity distributions, with each cell assigned as containing zero or at least one earthquake, with the outcome weighted by the empirical probability from either hypothesis.

- Using a two-tailed test, we defined the p-value as the percentage of synthetic seismicity distribu- tions with more earthquakes than the observed seismicity distribution.

If p ≤ 0.025 or p ≥ 0.975, we reject both the hypotheses for over- or underestimating seismicity, respectively.

- We reject both hypotheses at the 95% confidence interval, due to underestimating the Hector Mine earthquake (Figure 6).

Spatial Likelihood (S) Test

The spatial likelihood test, or “S”-test, evaluates observed seismicity locations relative to what each hypothesis would construct. We base the test on the probability of a specific set of zones containing at least one earthquake.

We generated 1000 synthetic seismicity distributions from each ΔCFF- and seismicity-based forecasts.

- For both the observed and synthetic seismicity distributions, we calculated the cumulative log likelihood, or “S” statistic.

where L is the log likelihood, N is the total number of cells in the study area, is the specific spatiotemporal cell in the probability cell containing at least one earthquake, and it is equal to zero or one, indicating whether we observed at least one earthquake in the cell during the target period (16/10/1999-12/31/1999).

We defined a test statistic  or as the percentage of synthetic log likelihood values exceeding the ob- served log likelihood.

To account for all potential bias from earthquake clustering, we chose a two-tailed test, rejecting either hypothesis at the 95% confidence interval of p ≤ 0.025 or p ≥ 0.975.

We reject both hypotheses at the 95% confidence interval, as both fail to indicate the full extent of aftershock locations over time (Figure 5).

Conclusions/Future Work

- At the 95% confidence interval, both the rate-and-state ΔCFF forecast and ETAS forecast fail to reliably forecast the areas and times of aftershocks following the Hector Mine earthquake.

- Although the ΔCFF-based forecast indicates earthquake locations during the target period more effectively than ETAS, the stress steps obtained from the rate-and-state stress inversion are confined to smaller locations than what would be physically realis- tic from a forward stress model.

- The rate-and-state stress inversion method effectively avoids stress singularities near faults that we observed from the forward model. However, the resulting stress steps underestimated aftershock areas.

- Varying the lower magnitude threshold from which we calculated ΔCFF yielded no significant change in the rate-and-state-based forecast’s effectiveness. Likewise, the stress field outside of the forecast evaluation windows did not significantly contribute to the resulting cell activation probability distribution’s effectiveness compared to the smoothed seismic model.

- Although stress singularities from the forward-rate-and-state model result in overestimating of expected seismicity, the forward model more effectively defines stress evolution over time.

- Instead of representing stress heterogeneities through observed seismicity patterns, optimizing fault section dislocations may both indicate future seismicity and delineate aftershock zones, as seen in previous studies.

Figure 5: S-test results when varying the lower magnitude threshold from the rate-and-state seismicity model (bottom row). Both the rate- and state- and smoothed seismicity-based forecasts are rejected at the 95% confidence interval, due to underestimating the number of acti- ve cells. In contrast to the smoothed seismicity model, the ΔCFF model shows several numbers of active cells for smoothed seismicity-detected probabilities. The black line shows how synthetic seismicity distributions compare to the observed results.