Assimilating Multicycle Rupture Simulations into Probabilistic Forecasting Models

Luis Vazquez[1]; Thomas H. Jordan[1]

[1]University of Southern California

USC University of Southern California

SCEC AN NSF+USGS CENTER

Abstract

We combine long earthquake catalogs (~ 10^6 yr) from the multi-cycle Rate-State Quake Simulator (RSQSim) of Dieterich & Richards-Dinger (2010) with the time-independent Uniform California Earthquake Rupture Forecast Version 3 (UCERF3), of Field et al. (2014). RSQSim ruptures were mapped into the UCERF3 rupture set. Our Bayesian approach uses the UCERF3-TI model as the prior distribution of earthquake rates, which we update using an RSQSim catalog. We model the catalog as a time-independent Poisson process and adopt a multi-variate gamma distribution as the conjugate prior. We assess the efficacy of the updating schemes by logarithmic scoring of the mean forecasts against independent RSQSim catalogs.

Mapping RSQSim into UCERF3

The RSQSim catalog was split into two subsets of 500 kyr each, one for training and the other one for testing.

We define an X% **rupture criterion** to map RSQSim into UCERF3, i.e., if the RSQSim rupture area of a given event within an UCERF3 subfault section is at least X% of such section, the mapped RSQSim area takes the value of the UCERF3 subfault area.

Our goal is to **minimize the moment bias between RSQSim and UCERF3**, i.e., keep the seismic moments in both sets as similar as possible for the same ruptures. Figure 1 compares the magnitude (moment) bias for rupture criteria of 20% and 50%, we conclude that 50% is better than 20% due to the smaller bias and because the number of mapped events for 50% is greater than for 20%. Thus, RSQSim was mapped into the UCERF3 event set using **a 50% rupture criterion.**

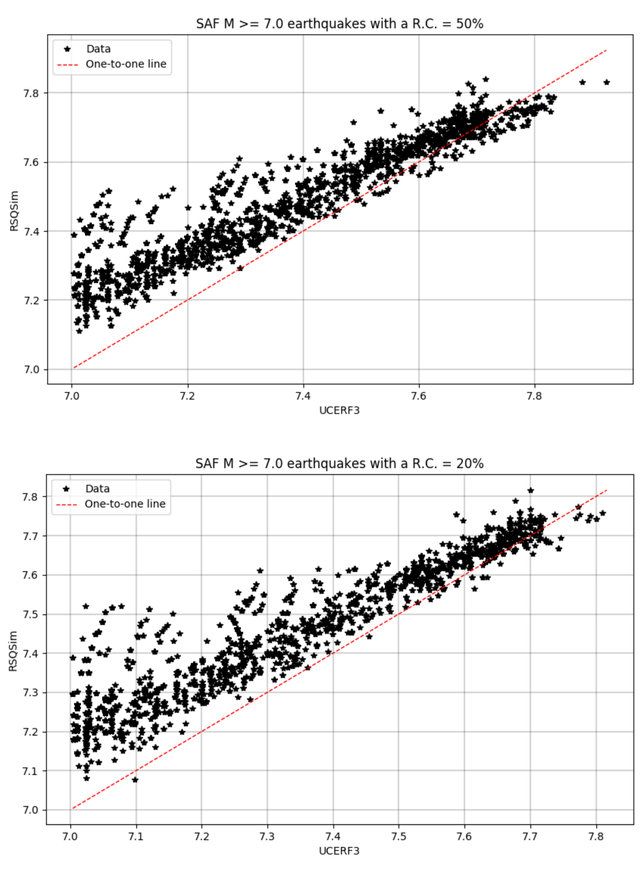

Figure 1: RS and U3 magnitude scatterplot. A rupture criterion of 50% and 20% was analyzed in the upper and lower subplot, respectively.

We further reduce these subsets by only analyzing $\boldsymbol{M \geq 7.0}$ **ruptures in the San Andreas Fault system (SAF)**. Events that were not mapped by RSQsim whose UCERF3 rate is smaller than 1×10^{-6} were excluded, yielding a total of $M = 1415$.

Bayesian calibration

We assume that UCERF3 rates obey a **multivariate Poisson-Gamma model** whose posterior PDF is defined in equation (1).

$$p(n_m|\alpha_m,\beta_m) = \int_0^\infty f(n_m|\lambda_m)g(\lambda_m|\alpha_m,\beta_m)d\lambda_m \qquad (1),$$

where $f(n_m|\lambda_m) = \frac{\lambda_m^{n_m} e^{-\lambda_m}}{n_m!}$ (likelihood function),

and $g(\lambda_m|\alpha_m,\beta_m) = \frac{\beta_m^{\alpha_m}}{\Gamma(\alpha_m)}\lambda_m^{a_m-1}e^{-\beta_m\lambda_m}$ (conjugate prior PDF),

n_m is the RSQSim count of the m-th rupture in the training set and λ_m is its rate.

The Poisson-Gamma model yields the expected value:

$$\bar{\lambda}_m = E[\lambda_m|n_m] = \frac{n_m + aM\tilde{\lambda}_m}{T + aM} \qquad (2).$$

The parameter $\alpha_0 = aM$ is called the pseudocount. The estimate $\bar{\lambda}_m$ will lie between the RSQSim rate $n_m/T(a \to 0)$, and the UCERF3 rate $\tilde{\lambda}_m$ $(a \to \infty)$.

Log-likelihood scoring

We assess the calibrated model against the test catalog by computing the negative-oriented logarithmic score of $f(n_m|\lambda_m)$ defined in (3). Here, n'_m is taken from the testing set.

$$LS = -\ln\left(\prod_{m=1}^{M}\frac{\lambda_m^{n'_m}e^{-\lambda_m}}{n'_m!}\right) = -\sum_{m=1}^{M}[n'_m\ln(\lambda_m) - \ln(n'_m!) - \lambda_m] \qquad (3).$$

The MLE is the one that minimizes (3):

$$\hat{a} = \underset{a}{\operatorname{argmin}}\, LS(a) \qquad (4),$$

where $LS(a) = -\ln\big(f(\boldsymbol{n}'|\boldsymbol{\lambda}(a))\big)$.

Finally, the skill score of our forecast was computed:

$$LSS = \frac{LS(posterior\ forecast) - LS(prior\ forecast)}{LS(optimal\ forecast) - LS(prior\ forecast)} \qquad (5).$$

Optimization of a

We compute $\bar{\lambda}_m$ in (2) with different values of a ranging from 0 to 1×10^9, we then obtain LS (3). Figure 2 presents LS as a function of a and shows that the negative-oriented logarithmic score is minimized at $\boldsymbol{\hat{a} = 8}$. Furthermore, the larger the value of a is, the closer the logarithmic score gets to the prior score meaning that $\bar{\lambda}_m$ collapses to $\tilde{\lambda}_m$.

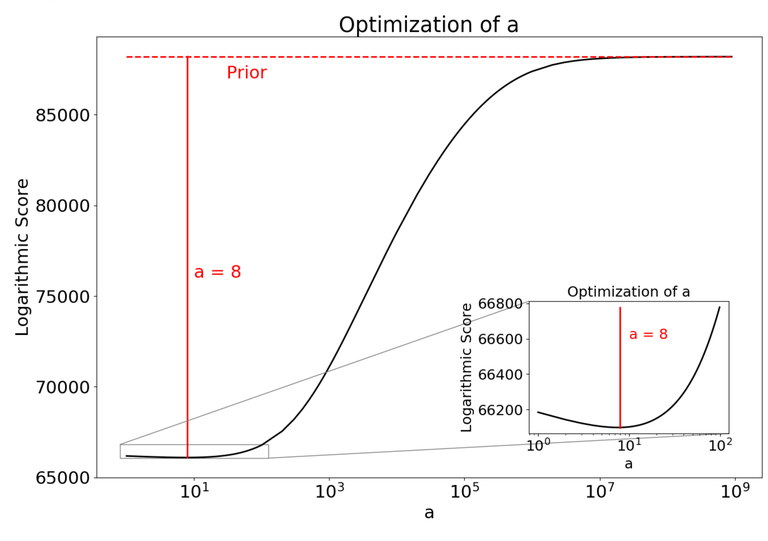

Figure 2. Optimization of a using a 500 kyr RSQSim catalog.

Bayesian calibrated rates

$\bar{\lambda}_m$ was computed with $a = 8$ for $m = 1,2,\ldots,1415$ and these results are shown in figure 3. Event rates above and below the one-to-one line were increased and decreased, respectively, after the Bayesian calibration.

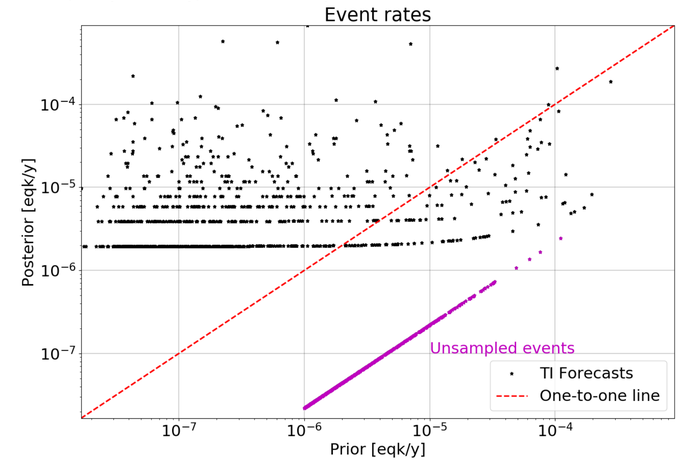

Figure 3. Event rates. UCERF3 event rates (prior event rates) were updated by RSQSim data to obtain the posterior event rates.

We then compute the participation rates for all the parent sections of the SAF. We find all events that have at least two fault subsections of the X parent section, and then we sum all those event rates to obtain the participation rate of the X parent section (figure 4). **The most frequent parent sections to rupture in a $\boldsymbol{M \geq 7.0}$ earthquake, are Santa Cruz and North Coast according to the prior and posterior forecast, respectively**.

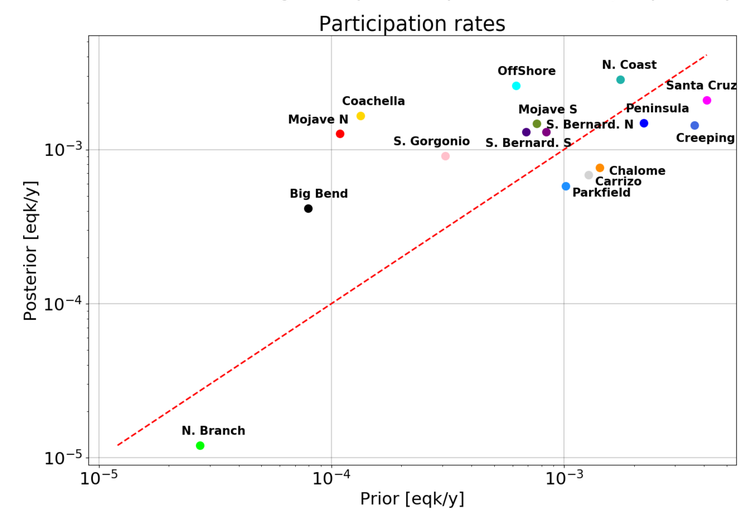

Figure 4. SAF parent section participation rates.

Finally, the logarithmic scores were computed to obtain the skill score in (5).

- $\boldsymbol{LS(prior\ forecast) = 88201}$
- $\boldsymbol{LS(posterior\ forecast) = 66098}$
- $\boldsymbol{LS(optimal\ forecast) = 65928}$
- **LSS = 0.9924**

These results show that the posterior is as good as the optimal forecast, $LSS \approx 1.0$.

Final remarks

- ✓ RSQSim was mapped into the UCERF3 with a **rupture criterion of 50%**, yielding a total of 1415 $\boldsymbol{M \geq 7.0}$ **UCERF3 events in the SAF**.
- ✓ The best mean time-independent forecast is obtained with $\boldsymbol{a = 8}$.
- ✓ The **skill score of this forecast is 0.9924** which means that the posterior is as good as the optimal forecast.
- ✓ RSQSim-calibrated rates for $\boldsymbol{M \geq 7.0}$ earthquakes on SAF parent sections differ from UCERF3-TI rates by factors of 2 to 10. In particular, these rates show order of magnitude increases for the Mojave N. and Coachella sections.
- ✓ The **North Branch Mill Creek** section is the **least participating** parent section in both forecasts.

References

- Dieterich, J. H., & Richards-Dinger, K. B. (2010). Earthquake recurrence in simulated fault systems. In *Seismogenesis and Earthquake Forecasting: The Frank Evison Volume II* (pp. 233-250). Springer, Basel.
- Field, E. H., Arrowsmith, R. J., Biasi, G. P., Bird, P., Dawson, T. E., Felzer, K. R., ... & Michael, A. J. (2014). Uniform California earthquake rupture forecast, version 3 (UCERF3)—The time-independent model. *Bulletin of the Seismological Society of America*, *104*(3), 1122-1180.