Ⓔ

Turing-Style Tests for UCERF3 Synthetic Catalogs

by Morgan T. Page and Nicholas J. van der Elst

Abstract   Epidemic-type aftershock sequence (ETAS) catalogs generated from the third Uniform California Earthquake Rupture Forecast (UCERF3) model are unique in that they are the first to combine a complex, fault-based long-term forecast with short-term earthquake clustering statistics. We present Turing-style tests to examine whether these synthetic catalogs can successfully imitate observed earthquake behavior in California. We find that UCERF3-ETAS is more spatially diffuse than the observed historic catalog in California and that it lacks quiet periods that are present in the real catalog. Although mean aftershock productivity of the observed catalog is matched closely by UCERF3-ETAS, the real catalog has more intersequence productivity variability and small mainshocks have more foreshocks. In sum, we find that UCERF3-ETAS differs from the observed catalog in ways that are foreseeable from its modeling simplifications. The tests we present here can be used on any model that produces suites of synthetic catalogs; as such, in addition to providing avenues for future improvements to the model, they could be incorporated into testing platforms such as Collaboratory for the Study of Earthquake Predictability (CSEP).

Electronic Supplement: Five 1000-yr third Uniform California Earthquake Rupture Forecast epidemic-type aftershock sequence (UCERF3-ETAS) synthetic catalogs, and figures showing aftershock decay rates for *M* 5.5–6.5 mainshocks, mean foreshock and aftershock productivity as a function of differential magnitude, and results from the clustering analysis.

Introduction

Traditionally, earthquake probability models have largely been in two separate time domains. Probabilistic hazard analysis employing fault-based information and elastic rebound has operated on time scales of decades to centuries, whereas empirical statistical models incorporating clustering due to aftershocks have operated on the scale of days to years. Bringing these different models into alignment requires bridging the "medium-term forecasting gap" (Jordan, 2012) that separates them as well as reconciling inherent statistical and spatial differences.

The third Uniform California Earthquake Rupture Forecast (UCERF3) model is the first unified earthquake forecasting model to include elastic rebound effects, fault-based ruptures, and clustering driven by Omori–Utsu statistics (Omori, 1895; Utsu, 1972). The base of the model is the time-independent model (UCERF3-TI) (Field *et al.*, 2014), which uses an inversion method (Page *et al.*, 2014) to relax segmentation and solve for the rates of hundreds of thousands of potential ruptures on an interconnected network of faults. To these fault-based ruptures, off-fault seismicity is added from smoothed historical seismicity data. The next progression of the model, the long-term time-dependent model (UCERF3-TD) (Field *et al.*, 2015), includes elastic rebound effects related to the time of the last event on major faults. Finally, short-term time dependence is added to this model in the form of epidemic-type aftershock sequence (ETAS) modeling (Ogata, 1988) to make UCERF3-ETAS (Field *et al.*, 2017).

UCERF3-ETAS differs from conventional ETAS due to the explicit modeling of faults. In particular, although conventional ETAS employs the Gutenberg–Richter (G-R) magnitude distribution (Gutenberg and Richter, 1944) at the smallest spatial scales, the magnitude distributions in UCERF3 are more variable. UCERF3 is constrained to follow G-R scaling in large regions (northern California and southern California), but in smaller regions it deviates from this scaling. On average, the magnitude distributions on the major modeled faults are characteristic (Wesnousky *et al.*, 1983; Schwartz and Coppersmith, 1984), meaning that the rates of $M \gtrsim 6.5$ earthquakes are higher than a G-R extrapolation from small earthquake rates would predict. However, there is considerable variation—in addition to faults with varying degrees of characteristic magnitude behavior, some

faults have magnitude distributions that are approximately G-R and still others are somewhat anticharacteristic.

The variable magnitude distributions in the UCERF3 model have a substantial effect on the triggering probabilities for large earthquakes. For potential foreshocks that occur in close proximity to characteristic faults, the probability of triggering a large earthquake is elevated relative to the foreshock probabilities in conventional ETAS; similarly, the model predicts that earthquakes occurring near anticharacteristic faults have a lower probability of triggering a large ($M \gtrsim 6.5$) earthquake. This effect is in agreement with a similar effect seen in the Agnew and Jones (1991) model that is also due to the adoption of characteristic magnitude distributions (Michael, 2012). A fault with a sufficiently characteristic magnitude distribution, could, in theory, cause a runaway sequence of aftershocks for parameters that would be subcritical in conventional ETAS; however, in UCERF3-ETAS this effect is mitigated by elastic rebound, which effectively removes many large earthquakes from the pool of potential aftershocks following rupture of a fault.

By modeling faults, elastic rebound, and short-term clustering, UCERF3-ETAS marries previously disparate types of earthquake forecasting and bridges the medium-term forecasting gap. Given both the complexity and novelty of the model, testing is required to ensure that results are consistent with observed seismicity. To this end, we devise a series of seismological Turing tests. The original Turing test was proposed by mathematician Alan Turing to sidestep what he saw as the poorly defined question of whether or not machines could truly think. Instead, he devised a game to determine whether a machine could successfully imitate the language behavior of a human (Turing, 1950). In our seismological Turing tests, we ask whether synthetic catalogs produced by UCERF3-ETAS can successfully imitate the statistical behavior seen in the real California earthquake catalog. Although the observed California catalog is quite limited in duration, we have many long synthetic catalogs. We can thus construct many synthetic-catalog snapshots that match the duration of the real catalog. These snapshots can be used to construct error bounds for various metrics; UCERF3-ETAS fails the Turing test if the observed data are outside of those bounds.

We compare the statistics of one hundred 1000-yr synthetic UCERF3-ETAS catalogs to the observed catalog in California for a 28-yr period from 1984 to 2011 (Felzer, 2013a). These tests are fully retrospective, because we compare UCERF3-ETAS to an observational catalog that was available at the time of its construction. Both observed and synthetic catalogs contain earthquakes down to M 2.5. UCERF3-ETAS catalogs are generated with direct Omori parameters from Hardebeck (2013). All synthetic catalogs start in 2012 and include aftershocks from the UCERF3 historical catalog (Felzer, 2013a). Rates of spontaneous background events in synthetic catalogs, which in the ETAS model are a proxy for aftershocks of earthquakes outside of the spatial and temporal bounds of the model, vary with time.

Early in the UCERF3-ETAS catalogs, these spontaneous events account for 30% of all events; after 1000 yrs, they account for 20%. Thus, it is possible for the catalogs to have different statistical behavior as they evolve. Also, because all synthetic catalogs include aftershocks from the same historical events, we expect the catalogs to be more similar at the start of 1000-yr simulations than at the end. For these reasons, in many of our tests we compare the 28-yr observed catalog to 28-yr snapshots taken from both the beginning and end of 1000-yr synthetic catalogs.

Seismological Turing Tests

Seismicity Rate

The mean seismicity rate in UCERF3 is 8.3 $M \geq 5$ earthquakes per year. By design, this is a bit higher than the historical (1850–2011) seismicity rate in California of 7.9 $M \geq 5$ earthquakes per year, and higher still than the recent (1984–2011) seismicity rate of 5.8 $M \geq 5$ earthquakes per year (Felzer, 2013b). Although there are several logic-tree branches with different seismicity rates in UCERF3, more weight was given to high-seismicity rates in the model because the true long-term seismicity rate in California is likely to be higher than the historical rates due to the lack of $M \geq 8$ events in the post-1850 era (Felzer, 2013d). The difference in mean seismicity rate is evident in Figure 1, which shows the range of small earthquake rates for 10-day windows in both the real and synthetic UCERF3-ETAS catalogs. However, even when the rates are normalized by the mean rate (Fig. 1c), the distributions are still different. The real California catalog from 1984 to 2012 (Felzer, 2013a) has more quiet periods than synthetic catalogs. This difference is significant given the scatter seen in shorter UCERF3-ETAS catalogs as well (Fig. 1c). This may be because of the spontaneous background events in UCERF3-ETAS, which account for ∼30% of the events early in synthetic catalogs and 20% of events 1000 yrs into a synthetic catalog (the fraction of events that are spontaneous diminishes as the catalogs are run for longer, because there are more previous events producing aftershocks). These spontaneous events are Poissonian in time, but in reality many of them represent orphaned aftershocks of parent earthquakes that occur prior to the start, outside the spatial bounds of the catalog, or below the minimum magnitude used (van der Elst, 2017). Thus, they may be more temporally clustered than is assumed in the modeling. At the high end of seismicity rates, UCERF3-ETAS and the observed catalog agree (Fig. 1c, inset).

The spatial distribution of small ($M \geq 2.5$) earthquakes for real and synthetic catalogs is shown in Figure 2. UCERF3-ETAS catalogs are more spatially diffuse than the real catalog. The spontaneous background events from the UCERF-ETAS catalog are shown in Figure 2c, and they are also more diffuse than the real catalog, which shows that the source of the discrepancy is not due (or only due) to the distance kernel used for aftershock triggering. Small earthquake
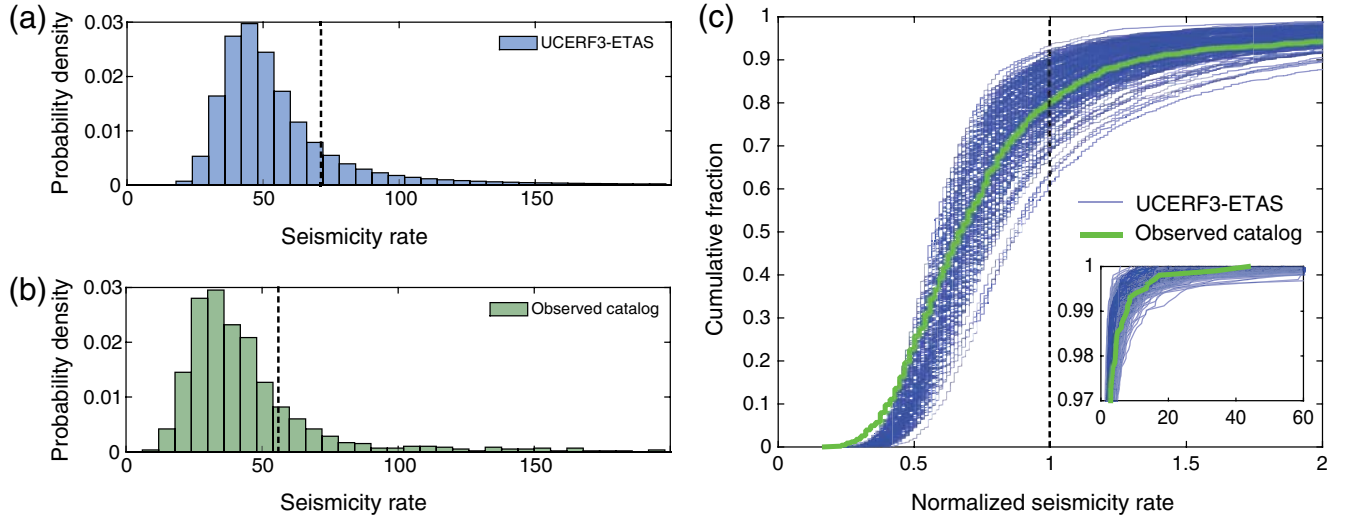
Figure 1. Seismicity rate ($M \geq 2.5$) for 10-day windows for (a) one hundred 1000-yr third Uniform California Earthquake Rupture Forecast epidemic-type aftershock sequence (UCERF3-ETAS) catalogs and (b) the real California catalog. UCERF3-ETAS catalogs lack quiet periods present in the real catalog. Vertical dashed lines show the mean seismicity rate. (c) Two hundred 28-yr snapshots of UCERF3-ETAS catalogs (half of the snapshots used are from the beginning of 1000-yr synthetic catalogs and half are from the end) have fewer quiet periods than the observed catalog, even when the mean rate difference is removed by normalization. The inset shows that for the high end of the seismicity-rate distribution, the observed catalog falls within the spread defined by the UCERF3-ETAS catalogs. The color version of this figure is available only in the electronic edition.

rates in UCERF3 are from the smoothed seismicity model (Felzer, 2013c), which is based on the adaptive algorithm of Helmstetter *et al.* (2007). This adaptive smoothing technique performed the best in a 5-yr prospective test of 17 forecast models in California (Zechar *et al.*, 2013). Although this smoothing technique produces a mean forecast that does well in testing, it does not produce synthetic catalogs with the spatial behavior of the real catalog. Ideally, individual synthetic catalogs would have similar spatial statistics as the real catalog, whereas the mean of many synthetic catalogs would be diffuse enough to perform well as a probability map. This could be accomplished by applying a stochastic curdling algorithm to individual catalogs to make them less diffuse. The idea would be to model the observed spatial clustering statistics in individual catalogs, whereas appropriately capturing the uncertainty as to where future seismicity clusters might be with intercatalog variability.

We quantify the difference between the synthetic and observed spatial distributions by calculating the 2D interevent distance (ignoring depth) between each pair of earthquakes in each catalog, respectively, following work by Kagan (2007). The number of pairs within a given interevent distance R approximately follows a power-law distribution

$$N_2(R) \propto R^\delta, \qquad (1)$$

as shown in Figure 2f. We estimate the power-law exponent δ for $M \geq 3$ events between interevent distances of 1 and 100 km. For the observed catalog, we fit $\delta = 1.16$; for twenty 28-yr synthetic catalogs, δ ranges between 1.31 and 1.47. The lower slope in the observed California catalog is consistent with the tighter spatial clustering seen in the map-based plots (Fig. 2a–d).

Magnitude Distribution

The yearly magnitude distribution for UCERF3-ETAS is compared with the observed catalog in Figure 3a. In this figure, the higher mean seismicity rate in UCERF3-ETAS is visible, as well as more severe magnitude rounding. The observed catalog has a surplus in the rate of $M \geq 6.4$ events relative to UCERF3-ETAS, but this is not statistically significant given the small numbers of these larger earthquakes.

Figure 3b shows b-value estimates versus minimum magnitude for each catalog, which allows us to focus on the magnitude probability distribution rather than differences in rate. The b-value estimates are computed using maximum-likelihood (Aki, 1965) and are corrected for magnitude rounding. The estimates assume no cutoff in maximum magnitude. This has the effect of inflating the b-value estimate for higher magnitudes, as shown in the b-value estimation for a pure G-R distribution with a magnitude truncation of $M_{max} = 8.4$ (thin line, Fig. 3b). The UCERF3-ETAS catalog has a slower fall-off above M 8 than this example. Also visible are rising b-values relative to $b = 1$ at intermediate magnitudes (M 4–5.7); these are consistent with a probability drop relative to G-R scaling at around M 7. The b-value estimates between the synthetic and observed catalogs are consistent at 95% confidence for the M 5.65 bin and above.

The observed California catalog has b-values consistent with 1.0 for most minimum magnitudes; however, there are deviations for several bins at small magnitudes. Although these deviations are statistically significant and preclude a single b-value at 95% confidence, we caution that the catalog is not a perfect representation of earthquakes that occur. M 2.5 is a reasonable completeness threshold for this time period in most
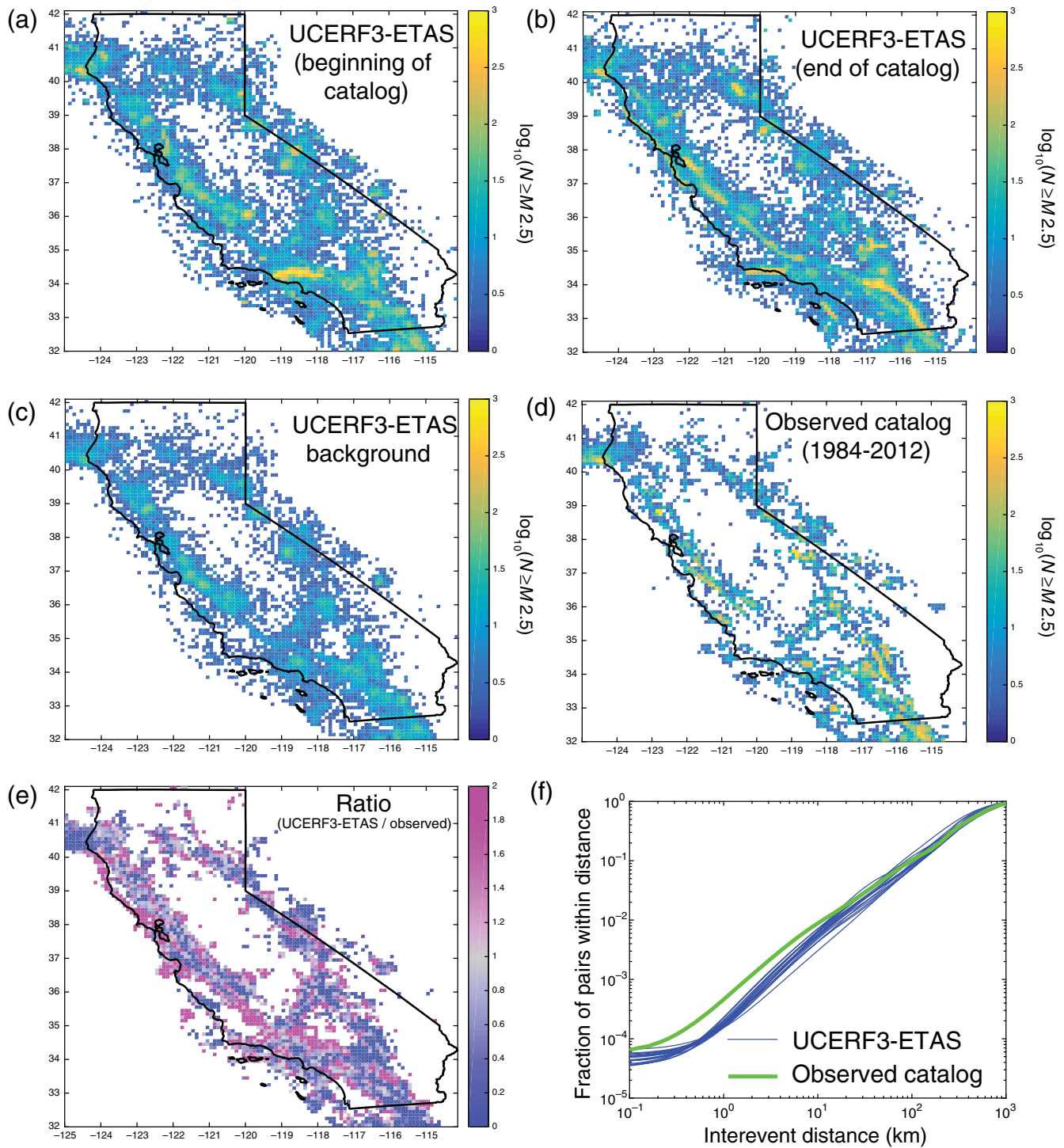
Figure 2. 28-yr snapshots of UCERF3-ETAS at (a) the start and (b) the end of synthetic 1000-yr catalogs. (c) The background (spontaneous) events from the catalog shown in (a). (d) The real catalog (1984–2012) is considerably less diffuse than the synthetic UCERF3-ETAS catalogs, and also less diffuse than the synthetic background events. (e) Event-rate ratio between one 1000-yr UCERF3-ETAS catalog and the observed catalog, for bins with a mean rate of at least 5 $M \geq 2.5$ earthquakes in either the synthetic or observed catalog in a 28-yr period (note that the color scale is clipped). (f) The fraction of $M \geq 3$ earthquake pairs within a given interevent distance, for 20 UCERF3-ETAS 28-yr snapshots (at the beginning of synthetic catalogs) and the observed catalog. The color version of this figure is available only in the electronic edition.

of the state (Felzer, 2013b), but may not be adequate immediately following large earthquakes. In addition, there is catalog nonuniformity in this time period in the definition of local magnitude (Tormann *et al.*, 2010; Uhrhammer *et al.*, 2011).

UCERF3-ETAS also has b-value nonuniformity at small magnitudes. At $M \geq 2.5$, the b-value estimate is 0.98. It continues to rise for higher minimum magnitudes, reaching $b = 1$ at $M_{\min} = 3.3$. UCERF3 was intended to strictly
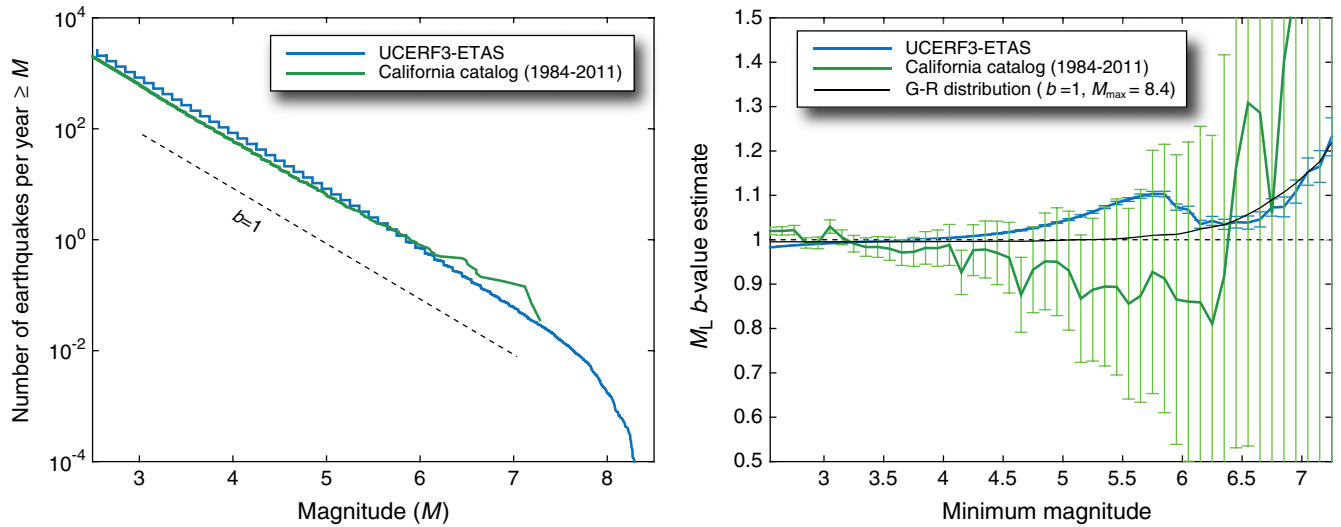
Figure 3. (Left) The cumulative magnitude–frequency distribution for one hundred 1000-yr UCERF3-ETAS and the recent observed California catalog. (Right) Maximum-likelihood *b*-value estimates (Aki, 1965) for UCERF3-ETAS and the observed catalog, corrected for magnitude rounding, with 95% confidence bounds. *b*-value estimates for a Gutenberg–Richter distribution with an upper magnitude cutoff at *M* 8.4 are shown for comparison. The color version of this figure is available only in the electronic edition.

match a *b*-value of 1.0 at low magnitudes; the discrepancy is not due to either magnitude rounding or the rolloff at high ($M \geq 8$) magnitudes. Rather, it represents a bug that will be fixed in future updates of the model.

Depth

By construction, the UCERF3-ETAS model matches the mean observed depth distribution observed in California. However, this depth distribution is (intended to be) uniformly applied throughout the state. Thus, the UCERF3-ETAS model does not have spatial variations in depth that are observed in the real catalog (Fig. 4). There is, however, a small amount of spatial heterogeneity in mean event depth visible in the UCERF3-ETAS catalogs. Earthquakes along the major

modeled faults in UCERF-ETAS have a slightly smaller mean depth. This may occur because modeled faults have a depth limited by their lower seismogenic depth, which for 80% of faults is between 9.0 and 16.1 km. Even small earthquakes are affected by this modeling assumption because most aftershocks occur near to their parents; thus, aftershocks of large fault-based earthquakes will be affected by this lower seismogenic depth. Off-fault spontaneous seismicity, however, has depths drawn from a continuous probability distribution extending to 24 km.

Foreshock and Aftershock Productivity

In contrast to conventional ETAS, in UCERF3-ETAS, after a large on-fault earthquake, probabilities of subsequent large earthquakes on the fault that just ruptured are reduced by elastic rebound. Another difference in UCERF3-ETAS comes from local magnitude distributions that may differ from G-R scaling. In the case of characteristic magnitude distributions, this enhances the probability of triggering large ($M \geq 6.5$) earthquakes following a potential foreshock.

We measure aftershock rates by fitting sequences to the modified Omori Law, assuming the aftershock rate λ scales as
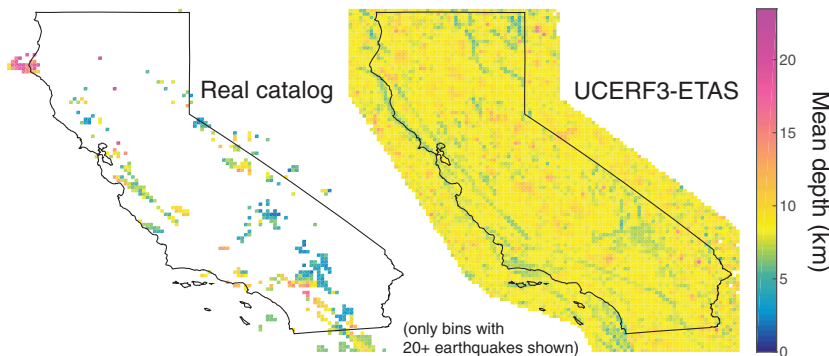

Figure 4. Mean depths for bins with more than 20 earthquakes for UCERF3-ETAS (one hundred 1000-yr catalogs) and the real catalog. By construction, the UCERF3-ETAS model does not have depth variation that is present in the real catalog. However, mean depths are not constant in UCERF3-ETAS as visible differences can be seen in the mean depth of on-fault earthquakes. The color version of this figure is available only in the electronic edition.
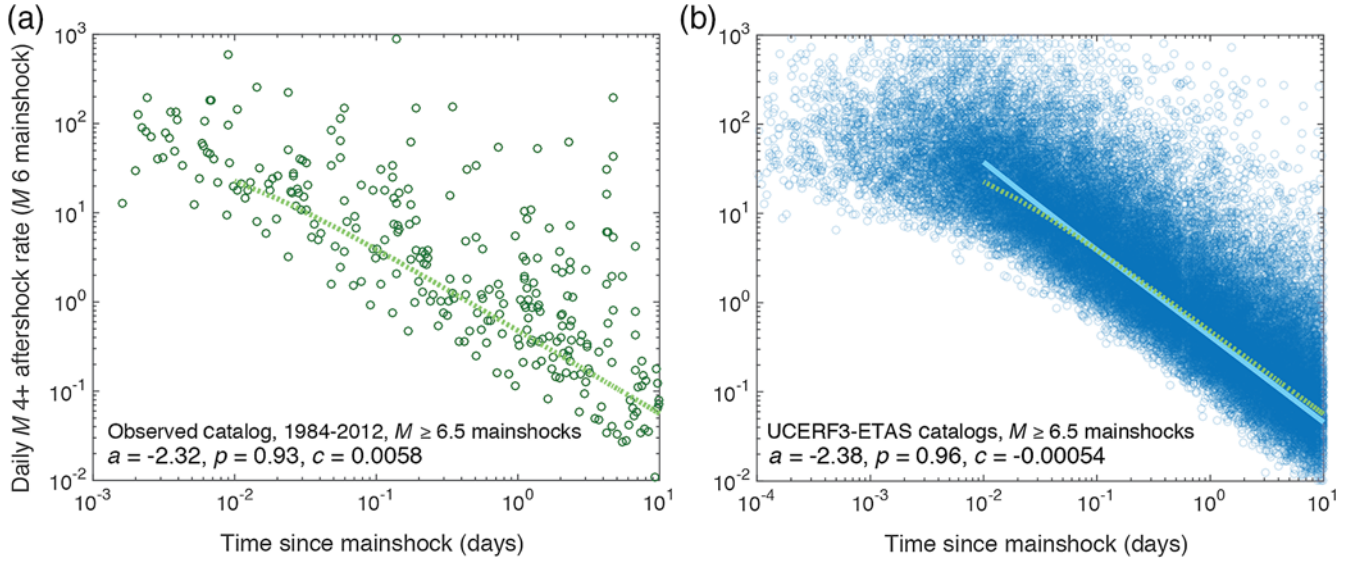
$$\lambda(t) = 10^{a+b(M_{\mathrm{main}}-M_{\mathrm{min}})}(t + c)^{-p}, \quad (2)$$

in which t is the time elapsed from the mainshock, M_{main} is the mainshock magnitude, M_{min} is the minimum magnitude of aftershocks under consideration, a, c, and

Figure 5. Interevent rate for subsequent aftershock pairs for $M \geq 6.5$ mainshocks, for (a) the observed catalog and (b) five 1000-yr synthetic UCERF3-ETAS catalogs. Aftershock rates are scaled to equivalent rates for a single M 6 mainshock, assuming productivity scales as $10^{M_{\mathrm{main}}}$, in which M_{main} is the mainshock magnitude. A maximum-likelihood fit to the temporal decay of the aftershock rate is shown for the real catalog (dotted line, shown in both panels for comparison with synthetic) and for the synthetic catalogs (solid line, b). The color version of this figure is available only in the electronic edition.

p are constants, and the G-R b-value is assumed to be 1 (Omori, 1895; Utsu, 1961; Reasenberg and Jones, 1989). The a-, p-, and c-values here are indirect Omori parameters and differ from the direct ETAS parameters that are applied to each generation of aftershocks in the modeling (Hardebeck, 2013). We stack all sequences relative to time since the mainshock and solve for a, c, and p by maximizing the log likelihood, which is given by

$$\log(L) = \sum_{i=1}^{N} \log(\lambda(t_i)) - \int_0^T \lambda(t)dt, \qquad (3)$$

in which N is the number of aftershocks that occur within T days of their mainshocks and t_i is the occurrence time of the ith aftershock.

On average, UCERF3-ETAS aftershock productivity and decay of $M \geq 4$ aftershocks following $M \geq 6.5$ mainshocks (as defined by the exclusion criteria of Page *et al.*, 2016, which includes aftershocks within $T = 10$ days of mainshocks) are very similar to the observed catalog (Fig. 5). The Omori c-value, which is a time constant that regularizes the rate function at small times after the mainshock, is smaller in UCERF3-ETAS—in fact, it is slightly negative (which is unphysical) in the maximum-likelihood fit (the direct Omori c-value used in UCERF3-ETAS is quite small: 1.78×10^{-5} days). The UCERF3-ETAS model currently does not make any attempt to model short-term aftershock incompleteness that is present in the observed catalog following moderate and large earthquakes. This detection bias in the observed catalog inflates the observed c-value, so this difference is (at least, partly) observational rather than physi-

cal. Fits of $M \geq 3$ aftershocks of M 5.5–6.5 mainshocks, shown in Ⓔ Figure S1 (available in the electronic supplement to this article) show a similar trend to the $M \geq 6.5$ mainshock data. Omori a-values and p-values are slightly higher and lower, respectively, in the observed catalog relative to synthetic catalogs. Ⓔ Figure S1c shows that the higher observed productivity at 10 days postmainshock is at the high end of 200 synthetic UCERF3-ETAS catalogs the same length as the observed catalog. This high-rate observed productivity is borderline statistically significant ($p = 0.03$) given the range of fits seen in short synthetic catalogs.

Although on average aftershock productivity of UCERF3-ETAS is close to what is observed in California, more stark differences emerge on a sequence-by-sequence basis. The observed California catalog contains more intersequence variability in aftershock productivity—both more quiet sequences and more productive sequences—than UCERF3-ETAS catalogs, as shown in Figure 6. This difference is evident after the first day, and is therefore not due to incompleteness problems in the observed catalog. UCERF3-ETAS does not include sequence-specific parameters—rather, the direct Omori a-, p-, and c-values used are constant from sequence to sequence, although stochasticity of the ETAS algorithm leads to some intersequence variability in productivity. For large ($M \geq 6.5$) mainshocks, UCERF3-ETAS results only diverge from the limited observational data for quiet sequences, which, again, are more numerous in the observed data. For the most productive sequences of $M \geq 6.5$ mainshocks, the observed results fall within the range given by short (28-yr) UCERF3-ETAS catalogs. It may be there is not enough data in a short catalog to see a difference in productivity for mainshocks as large as
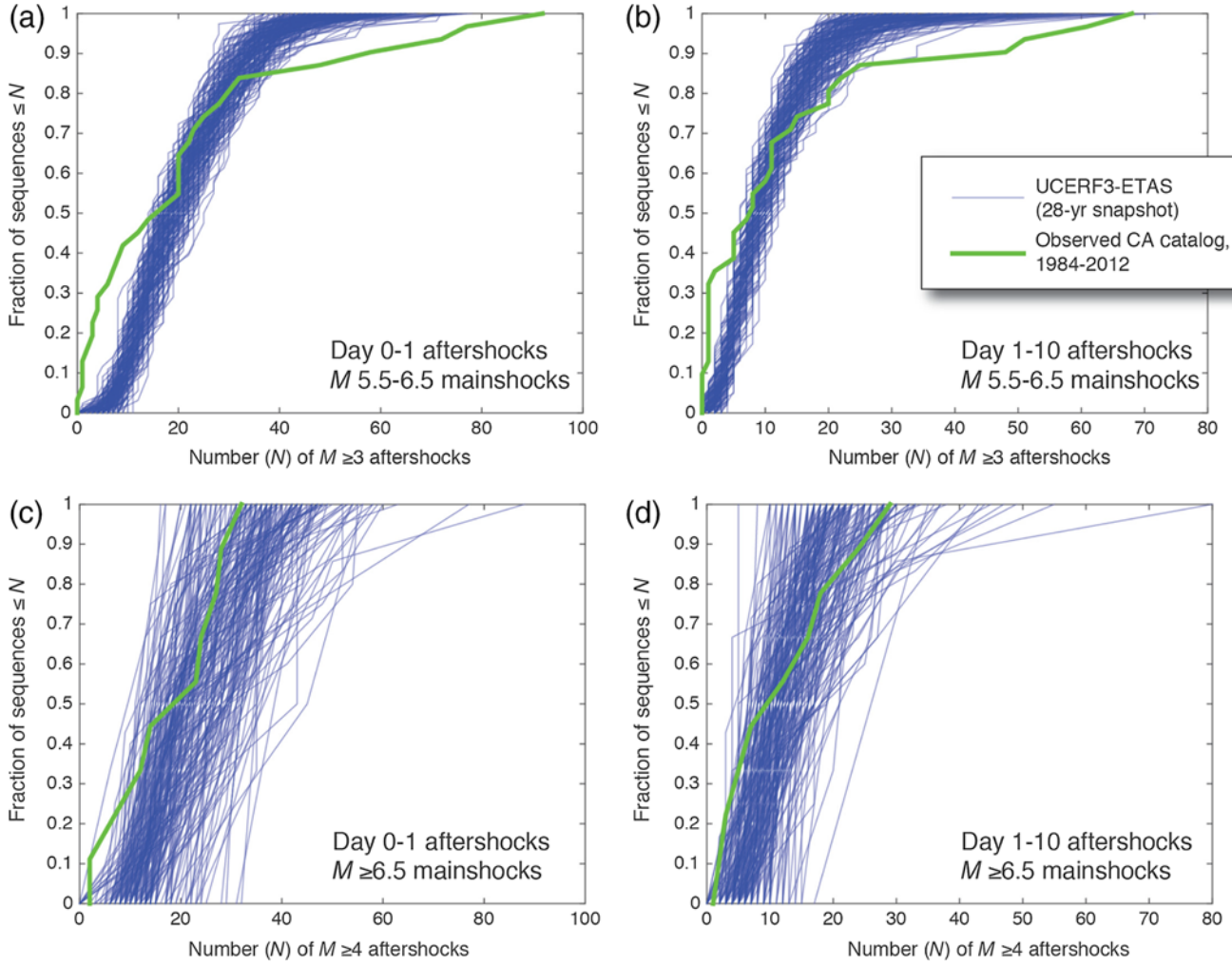
Figure 6. Number of $M \geq 3$ aftershocks per M 5.5–6.5 mainshock, plotted cumulatively, for observed and synthetic catalogs in (a) the first day following the mainshock, and (b) 1–10 days following the mainshock. Number of $M \geq 4$ aftershocks of $M \geq 6.5$ mainshocks are shown for (c) the first day following the mainshock and (d) 1–10 days following the mainshock. Each line in the plot represents the distribution of the number of aftershocks per sequence for one 28-yr catalog. For synthetic catalogs, we show two hundred 28-yr snapshots, half from the beginning of 1000-yr synthetic catalogs and half from the end. The color version of this figure is available only in the electronic edition.

M 6.5, or it may be that enhanced productivity differences due to large aftershocks drawn from characteristic magnitude distributions in UCERF3-ETAS bring the results into alignment for larger mainshocks.

We also compare foreshock and aftershock productivity as a function of differential magnitude, following Shearer (2012). Here, we again use the Page *et al.* (2016) exclusion criteria, but apply it to foreshocks as well as aftershocks. In the case of foreshocks, we include events 10 days prior to mainshocks, and as above, mainshocks are defined as the largest event in the sequence. Differential magnitude is defined as the difference between the aftershock or foreshock magnitude and the mainshock magnitude. Observed foreshock and aftershock productivity agrees with the range produced by UCERF3-ETAS catalogs the same length as the observed catalog (Fig. 7), with the exception of foreshocks of M 4.5–5.5 mainshocks, for which the foreshock produc-

tivity of the observed catalog is higher. Brodsky (2011) first pointed out that the observed catalog has a factor of 2–3 more foreshocks than is predicted by scale-free ETAS models. This result was verified by Shearer (2012), who observed that the effect was independent of differential magnitude, as we reproduce here. We find, however, that this effect disappears for higher mainshock magnitudes, for which both the foreshock and aftershock curves at different mainshock magnitudes are approximately coincident (Ⓔ Fig. S2).

Clustering Analysis

To compare spatiotemporal patterns in the UCERF3-ETAS model, we employ a time–space–magnitude clustering analysis (Baiesi and Paczuski, 2004; Zaliapin *et al.*, 2008). In this approach, the pairwise distance between events is defined as
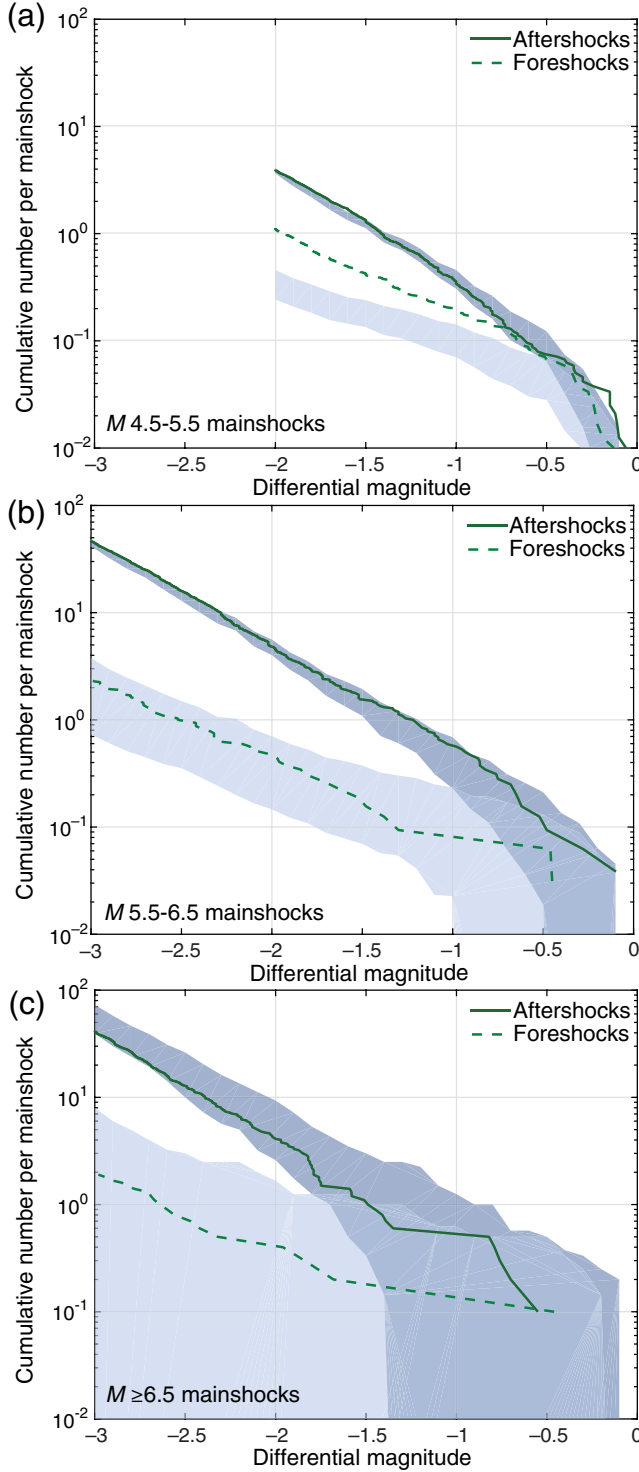
Figure 7. Foreshock and aftershock productivity for (a) M 4.5–5.5, (b) M 5.5–6.5, and (c) $M \geq 6.5$ mainshocks, as a function of differential magnitude (the difference between the foreshock or aftershock magnitude and mainshock magnitude). Solid and dashed lines show results for aftershocks and foreshocks, respectively, from the observed California catalog. Shaded regions show the 95% confidence bounds derived independently for each differential magnitude value from two hundred 28-yr snapshots of UCERF3-ETAS catalogs (half of the snapshots used are from the beginning of 1000-yr synthetic catalogs and half are from the end). The color version of this figure is available only in the electronic edition.

$$\eta_{ij} = \begin{cases} c\tau_{ij}r_{ij}^d 10^{-b(m_i-M_{\min})}, & \text{if } \tau_{ij} > 0 \\ \infty, & \text{if } \tau_{ij} \leq 0 \end{cases}, \quad (4)$$

in which τ_{ij} is the time separation between events i and j, r_{ij} is the epicentral distance, m_i is the magnitude of the ith event, $M_{\min}$ is the minimum catalog magnitude, and c and d are constants.

The nearest-neighbor distance for event j is the minimum η_{ij} over all earthquakes i. This pairwise distance is roughly equivalent to the number of earthquakes expected between event i and j—conceptually similar to the transformed time in the ETAS model (Ogata, 1992), with the addition of a spatial component. The nearest-neighbor distances are used to establish single-link networks of events that constitute clusters.

To visualize the spatiotemporal clustering, we decompose the nearest-neighbor distance into a spatiotemporal component:

$$T_{ij} = \tau_{ij}10^{-0.5bm_i} \quad (5)$$

$$R_{ij} = r_{ij}^d 10^{-0.5bm_i}, \quad (6)$$

in which we ignored the constant terms (setting $c = 1$ and $M_{\min} = 0$). Plotting T_{ij} versus R_{ij} allows us to visually separate highly correlated pairs (aftershocks) from uncorrelated background earthquakes (Zaliapin *et al.*, 2008). We use $b = 1$ and $d = 1.6$, the values used in Zaliapin *et al.* (2008).

The UCERF3-ETAS model treats large and small earthquakes differently, with the fault-based machinery kicking in only for earthquakes above a certain size ($M \sim 6.5$). We here plot only pairs for which the first event i is larger than M 6.5, and separate the population of second events j into four magnitude slices (M 3–4, 4–5, 5–6, ≥ 6). We contour the point density using an adaptive Gaussian kernel (Botev *et al.*, 2010). For the UCERF3-ETAS results, we stack the kernel density estimates from 28-yr time periods for 100 simulated catalogs.

The observed and simulated catalogs appear generally consistent (Figs. 8 and 9), except that the data are much sparser for the (single) observed catalog than for the stack of UCERF3-ETAS catalogs. (The diffusion of the contours at higher magnitudes is at least partly due to the reduced data density. For plots of individual 28-yr UCERF3-ETAS catalogs, see Ⓔ Fig. S4.) The other major difference is the relative lack of very short rescaled times in the observed catalog **M** 3–4 aftershocks, due to early-time catalog incompleteness.

We also examine the distribution of total cluster size and the number of generations within a cluster, called the branching depth (Zaliapin and Ben-Zion, 2013). A cluster is defined as an unbroken chain of nearest-neighbor links. For each earthquake i, we find all the subsequent events J for which event i is the nearest neighbor. Then, for each event j in J, we find all the subsequent events for which events j are the
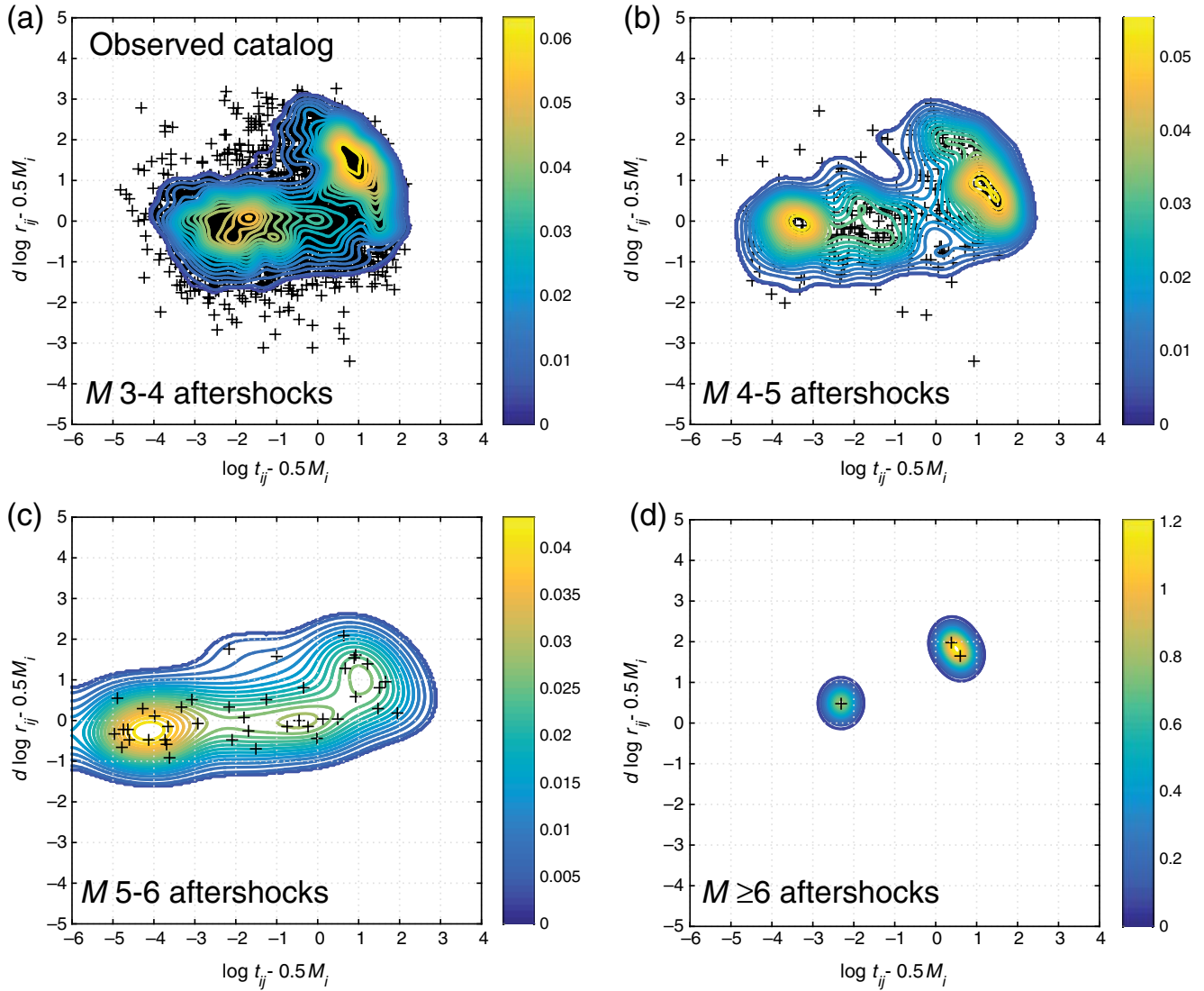
Figure 8. Distribution of pairwise nearest-neighbor distances (equations 5 and 6) in the observed catalog. Contours are probability density. The *x* axis is the normalized interevent time and *y* axis is the normalized epicentral separation. The distribution is bimodal, with the population divided into likely triggered earthquakes (smaller distances) and likely background earthquakes (larger distances). Parent earthquakes (at the origin) are *M* 6.5 and greater. Different panels correspond to different magnitude ranges of the daughter events. Compare with Figure 9. The color version of this figure is available only in the electronic edition.

nearest neighbor, and so on. This differs from the forward-looking pairwise links described above in that each event *i* can branch to multiple (or zero) events at later times. The cluster terminates when no subsequent events can be found for which an event in the cluster is the nearest neighbor.

The distribution of cluster sizes in the UCERF3-ETAS catalogs compares well with the observed catalog cluster sizes. Figure 10a shows the rate of clusters of different sizes in each catalog. The rate of small clusters is somewhat higher in the UCERF3-ETAS catalogs, consistent with the somewhat higher long-term rate used in the model. The branching depth is rather larger in the observed catalog than in the UCERF3-ETAS catalog (Fig. 10b). This implies that the branching process in the observed catalog is somewhat more robust than modeled by the simulations, with a higher ratio of

triggered to background earthquakes. This is again consistent with the proportion of background earthquakes in synthetic catalogs being slightly overestimated due to the inclusion of orphaned aftershocks.

Discussion and Conclusions

Many of the differences we find between the observed catalog statistics and UCERF3-ETAS are to be expected given UCERF3 modeling simplifications. Spatial variability in the depth profile of earthquakes, for example, is not included in UCERF3-ETAS; neither are direct aftershock productivity variations. Other differences, such as the higher rate of foreshocks for *M* 4.5–5.5 mainshocks, are known (and poorly
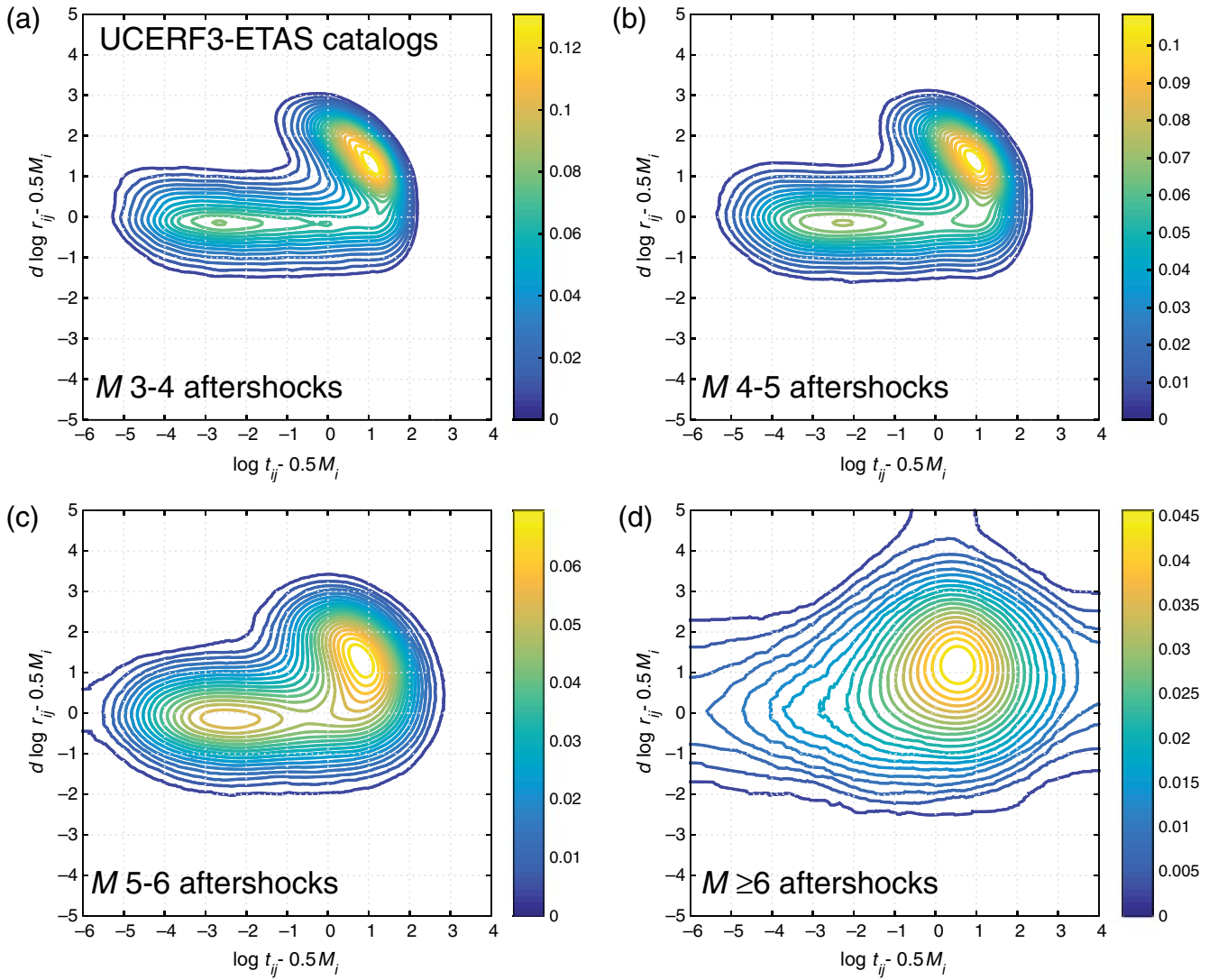
Figure 9.   Stacked distribution of nearest-neighbor distances (equations 5 and 6) in the first 28 yrs of 100 synthetic UCERF3-ETAS catalogs. Only the contours of probability are shown, for clarity. Compare with Figure 8. Stacked results from the last 28 yrs of one hundred 1000-yr simulated catalogs are shown in Ⓔ Figure S3 (available in the electronic supplement to this article). Those results are very similar. The color version of this figure is available only in the electronic edition.

understood) differences between any self-similar triggering model such as ETAS and observations in California.

UCERF3-ETAS lacks quiet periods that are present in the observed catalog. This means that the spontaneous events, which are ∼1/3 of all events early in the synthetic catalogs, are in fact not well modeled as Poissonian. The time dependence in UCERF3-ETAS comes from aftershock triggering—so either not enough of the UCERF3-ETAS earthquakes are modeled as aftershocks or there are other un-modeled time-dependent processes in the observed catalog (e.g., swarms) that are significant. Both of these deficiencies are known modeling simplifications—background events in ETAS are, at least partly, a proxy for aftershocks outside of the time, magnitude, and spatial window modeled—and changes in background rate, which occur during swarms, are not modeled.

The cluster-size analysis, which shows that UCERF3-ETAS has a smaller branching depth than the observed cata-log, also suggests that the background rate used in the ETAS modeling is too high. The background rate in the simulations is a function of the direct ETAS parameters (Hardebeck, 2013), which for UCERF3 define a branching ratio (Sornette and Helmstetter, 2002) of 0.67 for 28-yr time periods and 1.07 at infinite times. To be consistent with a deeper branch-ing depth, these parameters should be more critical at short times. Recent work analyzing the effect of orphaned after-shocks on ETAS parameter estimations has found that when contributions from earthquakes prior to the start of the cata-log are accounted for, the branching ratio for the California catalog rises from a naive estimate of 0.65 to at least 0.78, and possibly up to 1 depending on the assumed duration over which Omori's law applies (van der Elst, 2017). At long
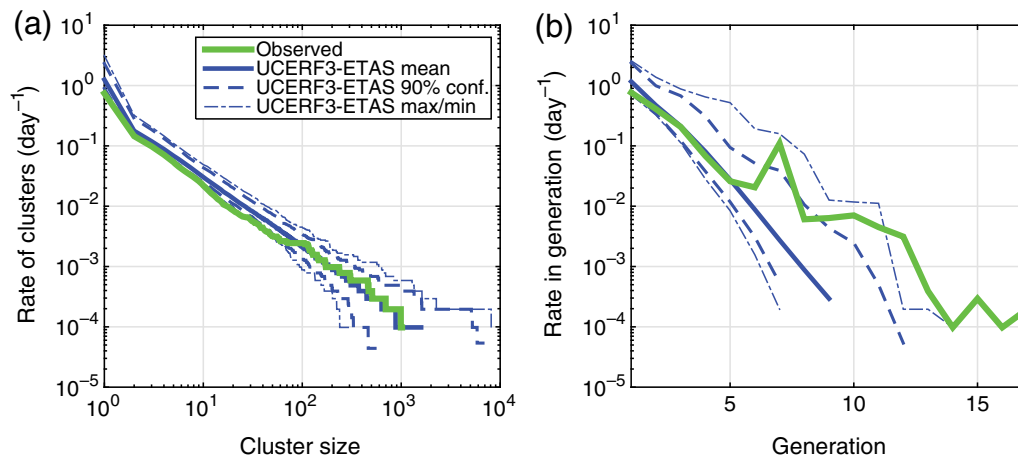
Figure 10. Clustering analysis of the UCERF3-ETAS catalog. (a) The rate of clusters as a function of cluster size. Dashed and dashed–dotted lines are, respectively, the 90% confidence range and minimum–maximum range of observations in one hundred 1000-yr UCERF3-ETAS synthetic catalogs. (b) Rate of events as a function of the generation number (branching depth) within the cluster. The color version of this figure is available only in the electronic edition.

times, the UCERF3-ETAS parameters should be less critical (physical, subcritical values for the branching ratio are less than 1). Setting the branching ratio to 1 at infinite times, which in effect assumes that there is no true background, is consistent with California data (van der Elst, 2017) and would help constrain the parameter estimation.

The adaptively smoothed background model for off-fault seismicity in UCERF3-ETAS produces catalogs that are more spatially diffuse than the observed catalog. This suggests that the level of smoothing necessary for a mean forecast to perform well in short-term tests smooths too much seismicity into aseismic areas when applied to individual synthetic catalogs. One aggravating problem here may be that the smoothing kernel is isotropic and does not account for the geometry of the fault network. Besides the potential problem of overestimating earthquake rates in off-fault areas, the spatial distribution of small earthquakes in UCERF3 is important scientifically, because it directly affects how characteristic the model is. On average, the faults in UCERF3 have a characteristic magnitude distribution (i.e., large earthquakes have a higher frequency than a G-R extrapolation from the small earthquake rate would imply) and the surrounding areas are anticharacteristic. This is because small earthquake rates, determined from smoothed seismicity, are more spatially diffuse than large earthquakes, a large proportion of which must be placed on mapped faults to match geologic and geodetic slip rates. If small earthquakes were more clustered spatially near the major faults in the model, magnitude distributions would be less characteristic, which has significant implications for earthquake foreshock rates near faults (Michael, 2012).

In our opinion, the most pressing difference between the observed catalog and UCERF3-ETAS is the underestimation of intersequence aftershock productivity in UCERF3-ETAS. There are existing forecasting models in California that include sequence-specific aftershock productivity (e.g., the short-term earthquake probability [STEP] model of Gerstenberger *et al.*, 2005). As a modeling simplification, UCERF3-ETAS uses a single set of direct Omori parameters for all earthquakes. The observed catalog has more variability than this. When UCERF3-ETAS is run prospectively following a large earthquake—and this is the most likely mode for it to be run because it may be too computationally demanding to run continuously—it will underestimate the range of likely numbers of aftershocks. As aftershocks are recorded and included in the model, secondary triggering in the ETAS model does make it somewhat adaptable—for example, sequences with large numbers of early aftershocks will have more later aftershocks as well triggered by those early aftershocks. However, this adaptability is somewhat limited because $\sim 1/2$ of late aftershocks are primary (i.e., triggered directly by the mainshock).

In this article, we focused on a specific class of tests, seismological Turing tests, that put observed and synthetic catalogs through identical statistical machinery to look for differences in spatial, temporal, and size behavior. Other types of tests are also possible—for example, one could mask tests to just look at known fault locations, for example, to focus on where UCERF3-ETAS is likely to differ from other forecasting models. Prospective tests at the Collaboratory for the Study of Earthquake Predictability (CSEP; Jordan, 2006) would also quantify how well UCERF3-ETAS performs relative to other forecasting methods in California. Currently, CSEP tests use mean forecast rates and assume Poissonian variability in rate, which limits their applicability for models like UCERF3-ETAS. Tests of the style presented here, however, can be used on any model that generates suites of synthetic catalogs in a region. Data limitations in any of these tests do make it difficult to test UCERF3-ETAS and its assumptions for the largest ($M \geq 6.5$) earthquakes, which is precisely where we expect the model to be dramatically different than simpler models that lack explicitly mod-

eled faults. Thus, in future work we will compare UCERF3 with global data of the biggest earthquakes, specifically to look for evidence of the effect of modeled characteristic magnitude distributions on foreshock/aftershock statistics.

Data and Resources

The third Uniform California Earthquake Rupture Forecast epidemic-type aftershock sequence (UCERF3-ETAS) catalogs were generated using the UCERF3 code, which is available at https://github.com/OpenSHA (last accessed October 2017). The first 5 of the one hundred 1000-yr UCERF3-ETAS catalogs are included in an Ⓔ electronic supplement to this article. The observed California catalog used in this work was developed in the UCERF3 project (Felzer, 2013a) and is available at https://pubs.usgs.gov/of/2013/1165/ (last accessed August 2015). Code used to sort earthquakes into mainshock/aftershock sequences can be found at https://github.com/mtpage/Aftershocks (last accessed February 2016).

Acknowledgments

References

Agnew, D. C., and L. M. Jones (1991). Prediction probabilities from foreshocks, *J. Geophys. Res.* **96,** 11,959–11,971.

Aki, K. (1965). Maximum likelihood estimate of b in the formula $\log N = a - bM$ and its confidence limits, *Bull. Earthq. Res. Inst.* **43,** 237–239.

Baiesi, M., and M. Paczuski (2004). Scale-free networks of earthquakes and aftershocks, *Phys. Rev. E* **69,** 066106, doi: 10.1103/PhysRevE.69.066106.

Botev, Z. I., J. F. Grotowski, and D. P. Kroese (2010). Kernel density estimation via diffusion, *Ann. Stat.* **38,** 2916–2957, doi: 10.1214/10-AOS799.

Brodsky, E. E. (2011). The spatial density of foreshocks, *Geophys. Res. Lett.* **38,** doi: 10.1029/2011GL047253.

Felzer, K. R. (2013a). Appendix K, The UCERF3 earthquake catalog. Uniform California Earthquake Rupture Forecast, Version 3 (UCERF3), *U.S. Geol. Surv. Open-File Rept. 1165,* 5 pp.

Felzer, K. R. (2013b). Appendix L, Estimate of the seismicity rate and magnitude-frequency distribution of earthquakes in California from 1850 to 2011. Uniform California Earthquake Rupture Forecast, Version 3 (UCERF3), *U.S. Geol. Surv. Open-File Rept. 1165,* 13 pp.

Felzer, K. R. (2013c). Appendix M, Adaptive smoothed seismicity model. Uniform California Earthquake Rupture Forecast, Version 3 (UCERF3), *U.S. Geol. Surv. Open-File Rept. 1165,* 12 pp.

Felzer, K. R. (2013d). Appendix Q, The empirical model. Uniform California Earthquake Rupture Forecast, Version 3 (UCERF3), *U.S. Geol. Surv. Open-File Rept. 1165,* 6 pp.

Field, E. H., R. J. Arrowsmith, G. P. Biasi, P. Bird, T. E. Dawson, K. R. Felzer, D. D. Jackson, K. M. Johnson, T. H. Jordan, C. Madden, *et al.* (2014). Uniform California Earthquake Rupture Forecast, version 3 (UCERF3)—The time-independent model, *Bull. Seismol. Soc. Am.* **104,** 1122–1180, doi: 10.1785/0120130164.

Field, E. H., G. P. Biasi, P. Bird, T. E. Dawson, K. R. Felzer, D. D. Jackson, K. M. Johnson, T. H. Jordan, C. Madden, A. J. Michael, *et al.* (2015). Long-term time-dependent probabilities for the Third Uniform California Earthquake Rupture Forecast (UCERF3), *Bull. Seismol. Soc. Am.* **105,** 511–543, doi: 10.1785/0120140093.

Field, E. H., K. R. Milner, J. L. Hardebeck, M. T. Page, N. van der Elst, T. H. Jordan, A. J. Michael, B. E. Shaw, and M. J. Werner (2017). A spatiotemporal clustering model for the Third Uniform California Earthquake Rupture Forecast (UCERF3-ETAS): Toward an operational earthquake forecast, *Bull. Seismol. Soc. Am.* **107,** no. 3, 1049, doi: 10.1785/0120160173.

Gerstenberger, M. C., S. Wiemer, L. M. Jones, and P. A. Reasenberg (2005). Real-time forecasts of tomorrow's earthquakes in California, *Nature* **435,** 328–331.

Gutenberg, B., and C. F. Richter (1944). Frequency of earthquakes in California, *Bull. Seismol. Soc. Am.* **4,** 185–188.

Hardebeck, J. L. (2013). Appendix S, Constraining epidemic type aftershock sequence (ETAS) parameters from the Uniform California Earthquake Rupture Forecast, version 3 catalog and validating the ETAS model for magnitude 6.5 or greater earthquakes. Uniform California Earthquake Rupture Forecast, Version 3 (UCERF3), *U.S. Geol. Surv. Open-File Rept. 1165,* 24 pp.

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,** 78–86, doi: 10.1785/gssrl.78.1.78.

Jordan, T. H. (2006). Earthquake predictability, brick by brick, *Seismol. Res. Lett.* **77,** 3–6, doi: 10.1785/gssrl.77.1.3.

Jordan, T. H. (2012). Earthquake forecasting as a system-science problem, *AGU Fall Meeting Abstracts*, San Francisco, California, 3–7 December.

Kagan, Y. Y. (2007). Earthquake spatial distribution: The correlation dimension, *Geophys. J. Int.* **168,** 1175–1194, doi: 10.1111/j.1365-246X.2006.03251.x.

Michael, A. J. (2012). Fundamental questions of earthquake statistics, source behavior, and the estimation of earthquake probabilities from possible foreshocks, *Bull. Seismol. Soc. Am.* **102,** 2547–2562, doi: 10.1785/0120090184.

Ogata, Y. (1988). Statistical models for earthquake occurrence and residual analysis for point processes, *J. Am. Stat. Assoc.* **83,** 9–27.

Ogata, Y. (1992). Detection of precursory relative quiescence before great earthquakes through a statistical model, *J. Geophys. Res.* **97,** 19,845–19,871, doi: 10.1029/92JB00708.

Omori, F. (1895). On after-shocks of earthquakes, *J. Coll. Sci. Imp. Univ. Tokyo* **7,** 111–200.

Page, M. T., E. H. Field, K. R. Milner, and P. M. Powers (2014). The UCERF3 grand inversion: Solving for the long-term rate of ruptures in a fault system, *Bull. Seismol. Soc. Am.* **104,** 1181–1204, doi: 10.1785/0120130180.

Page, M. T., N. van der Elst, J. Hardebeck, K. Felzer, and A. J. Michael (2016). Three ingredients for improved global aftershock forecasts: Tectonic region, time-dependent catalog incompleteness, and intersequence variability, *Bull. Seismol. Soc. Am.* **106,** no. 5, 2290, doi: 10.1785/0120160073.

Reasenberg, P. A., and L. M. Jones (1989). Earthquake hazard after a mainshock in California, *Science* **243,** 1173–1176.

Schwartz, D. P., and K. J. Coppersmith (1984). Fault behavior and characteristic earthquakes: Examples from the Wasatch and San Andreas fault zones, *J. Geophys. Res.* **89,** 5681–5698.

Shearer, P. M. (2012). Self-similar earthquake triggering, Båth's law, and foreshock/aftershock magnitudes: Simulations, theory, and results

for southern California, *J. Geophys. Res.* **117,** no. B6, doi: 10.1029/2011JB008957.

Sornette, D., and A. Helmstetter (2002). Occurrence of finite-time singularities in epidemic models of rupture, earthquakes, and starquakes, *Phys. Rev. Lett.* **89,** 158501, doi: 10.1103/PhysRevLett.89.158501.

Tormann, T., S. Wiemer, and E. Hauksson (2010). Changes of reporting rates in the southern California earthquake catalog, introduced by a new definition of M_L, *Bull. Seismol. Soc. Am.* **100,** 1733–1742, doi: 10.1785/0120090124.

Turing, A. M. (1950). Computing machinery and intelligence, *Mind* **59,** 433–460.

Uhrhammer, R. A., M. Hellweg, K. Hutton, P. Lombard, A. W. Walters, E. Hauksson, and D. Oppenheimer (2011). California integrated seismic network (CISN) local magnitude determination in California and vicinity, *Bull. Seismol. Soc. Am.* **101,** 2685–2693, doi: 10.1785/0120100106.

Utsu, T. (1961). A statistical study on the occurrence of aftershocks, *Geophys. Mag.* **30,** 521–605.

Utsu, T. (1972). Aftershocks and earthquake statistics (3): Analyses of the distribution of earthquakes in magnitude, time and space with special consideration to clustering characteristics of earthquake occurrence (1), *J. Facul. Sci. Hokkaido Univ.* **3,** 379–441.

van der Elst, N. J. (2017). Accounting for orphaned aftershocks in the earthquake background rate, *Geophys. J. Int.* **211,** 1108–1118, doi: 10.1093/gji/ggx329.

Wesnousky, S. G., C. H. Scholz, K. Shimazaki, and T. Matsuda (1983). Earthquake frequency distribution and the mechanics of faulting, *J. Geophys. Res.* **88,** 9331–9340.

Zaliapin, I., and Y. Ben-Zion (2013). Earthquake clusters in southern California I: Identification and stability, *J. Geophys. Res.* **118,** 2847–2864, doi: 10.1002/jgrb.50179.

Zaliapin, I., A. Gabrielov, V. Keilis-Borok, and H. Wong (2008). Clustering analysis of seismicity and aftershock identification, *Phys. Rev. Lett.* **101,** 018501, doi: 10.1103/PhysRevLett.101.018501.

Zechar, J. D., D. Schorlemmer, M. J. Werner, M. C. Gerstenberger, D. A. Rhoades, and T. H. Jordan (2013). Regional earthquake likelihood models I: First-order results, *Bull. Seismol. Soc. Am.* **103,** 787–798, doi: 10.1785/0120120186.

U.S. Geological Survey
525 South Wilson Avenue
Pasadena, California 91106
pagem@caltech.edu
nvanderelst@usgs.gov