Heterogeneity of Aftershock Productivity Along the Mainshock Ruptures and Its Advantage in Improving Short-Term Aftershock Forecast

Yicun Guo*, Jiancang Zhuang†, and Huai Zhang‡

*Key Laboratory of Computational Geodynamics, University of Chinese Academy of Sciences, Beijing, China; †Institute of Statistical Mathematics, Tokyo, Japan

Abstract This study introduces an improved hypocentral version of the space-time Epidemic-Type Aftershock Sequence (ETAS) model, entitled as the 3D-finite source (3D-FS) ETAS model, and applies it to the analysis of the Southern California earthquake catalog. By stochastic reconstruction, we are able to reconstruct the patterns of aftershock productivity density along the mainshock ruptures. Detailed analysis of the productivity patterns reveals that: (1) Directly triggered aftershocks make up 21% to 41% of all earthquakes within the mainshock rupture areas, and show significant spatial heterogeneity and temporal migrations; (2) Major aftershocks tend to locate in low productivity areas, at the edges of clusters formed by small aftershocks; (3) Large slip areas are depleted of aftershocks, over 60% of all productivity distributes in areas with slip less than 0.3 times of the maximum slip, and the trajectory of the productivity pattern on the fault plane demonstrates apparent compensation to coseismic slip. We relate the difference in triggering abilities of four mainshocks to the heat flow in corresponding regions. Simulation results suggest that the 3D-FS ETAS model has apparent advantages of improving the performance of short-term aftershock forecast. Moreover, the later aftershocks are more correlated with the locations of subsequent events than earlier aftershocks, suggesting that the migration of aftershocks is important for mitigating aftershock hazard.

Plain Language Summary How many aftershocks can be produced at each part along the rupture of a large earthquake is crucial for us to understand how aftershocks are generated and how the strain energy in the crust relaxes afterward. It is difficult to know whether an earthquake is an aftershock triggered by some previous earthquake or a background earthquake that has no ancestry since earthquake clusters not only overlap with each other but also with background seismicity in space and time. We proposed a new earthquake clustering model to disentangle such overlapping and quantify family trees among the earthquakes. Our study on the earthquake data in Southern California shows that aftershocks tend to occur at low slip areas in the mainshock rupture and that large aftershocks occur farther away near the boundary of such concentration of aftershocks. Moreover, this model improves the short-term aftershock forecasting.

1. Introduction

Aftershocks following large earthquakes form significant clusters near the mainshock ruptures. Large aftershocks can cause notable damage comparable with the mainshock. Detailed studies on aftershocks provide information of fault geometries (Ross, Trugman, et al., 2019; Yue et al., 2017), the migration of aftershocks (Kato et al., 2016; Peng & Zhao, 2009), and the distribution of strong aftershocks (Stallone & Marzocchi, 2018; van der Elst & Shaw, 2015). Resolving the spatiotemporal patterns of aftershocks is important for understanding the underlying physical mechanisms of earthquake triggering and the postseismic relaxation process (Enescu et al., 2007; Tajima & Kanamori, 1985).

Most of the aftershocks occur along the mainshock ruptures, which is generally interpreted as the response to static or dynamic stress changes imparted by the mainshock (Freed, 2005). However, it is difficult to explain the near-fault aftershocks through quantitative calculations of stress changes because the limited resolution of mainshock slip or unknown distribution of the receiver faults (Steacy et al., 2004). Therefore, many studies about the relationship between aftershocks and mainshocks are established by visual comparisons between aftershock locations and mainshock slip distribution (Das & Henry, 2003; Ebel &
Chambers, 2016; Mendoza & Hartzell, 1988; Wetzler et al., 2018). Specifically, Woessner et al. (2006) did systematic statistical tests to quantify the correlation between aftershock locations and the distribution of coseismic slip and stress drop of mainshocks, and found that aftershocks correlate with regions of low coseismic slip and of increased shear stress. These studies reveal that aftershocks are preferentially located in relatively low mainshock slip areas, near the periphery of large slip asperities, suggesting that the stress level in large slip areas has reduced to below the failure criticality. Moreover, van der Elst and Shaw (2015) found that there is a spatial dependency of aftershock magnitudes, with larger aftershocks occurring farther away than smaller ones from the centroid of cluster formed by early aftershocks.

The previous studies help us qualitatively understand the nucleation and onset of aftershocks within the mainshock ruptures. However, we still need more detailed quantitative studies by using sophisticated models which can objectively describe earthquake triggering to verify this relationship, and further to incorporate it in the forecast of short-term aftershocks. The key issue in the quantitative studies of the relationship between aftershocks and mainshocks is that the exclusion of high-order aftershocks, that is, we should exclude those aftershocks triggered by aftershocks and take the left aftershocks as directly triggered by the mainshocks. Comparisons should be made between these “directly triggered aftershocks” and mainshock slip distributions. Hence, using statistical models which well describe the clustering features of earthquakes to detect the aftershock heterogeneity is feasible. Marsan and Helmstetter (2017) investigated the fluctuations of aftershock productivity in Southern California, and related the results to the stress drop variability of mainshocks. Hainzl et al. (2019) found a linear relationship between the aftershock productivity and the coupling coefficient in the Chile subduction zone.

In the past decades, the Epidemic-Type Aftershock Sequence (ETAS) model has been proved to be an effective tool in quantifying earthquake triggering (Console & Murru, 2001; Ogata, 1988, 1998), estimating background seismicity (Console et al., 2003; Zhuang et al., 2002, 2005), extracting seismicity anomalies (Reverso et al., 2015; Nishikawa & Ide, 2017), and forecasting short-term aftershocks (Tsuruoka et al., 2012; Zhuang, 2011). In the space-time ETAS model, the distribution of aftershocks is assumed to follow the empirical power laws: The Omori-Utsu law in time (Omori, 1894; Utsu, 1969) and an inverse power law in space. In addition, the Gutenberg-Richter law (Gutenberg & Richter, 1956) for determining magnitude-frequency distribution of earthquakes is combined with the ETAS model in simulations. To improve the biased estimation of direct aftershocks of large earthquakes, Guo et al. (2015) proposed the 2D finite source (2D-FS) ETAS model which incorporates rupture geometries of mainshocks. This model enhances the productivity of mainshocks and successfully reproduces the anisotropic distribution of aftershocks in the mainshock ruptures (Zhuang et al., 2018). Furthermore, Guo et al. (2019) proposed the 3D version of the finite source (3D-FS) ETAS model, and validated it through the Kumamoto earthquake sequence.

This study aims to apply the 3D-FS ETAS model to the Southern California catalog, to reconstruct the 3D patterns of aftershock productivity within the mainshock ruptures through stochastic reconstruction method (Zhuang et al., 2004, 2005), and to carry out quantitative comparisons between aftershock productivity and mainshock slip distributions. Also, we expect to get high resolution patterns of long-term background seismicity in depth from the 3D model, which facilitates the probabilistic seismic hazard analysis. In the last section, a simulation experiment is done to test to what extent the heterogeneity of aftershock productivity can improve the performance of aftershock forecasts.

2. Methodology

The conditional intensity function of the 3D-finite source ETAS model can be written in the form:

\[
\lambda(t,x,y,z \mid H_t) = \mu(x,y,z) + \sum_{i \in \mathcal{J}} \kappa(M_i)g(t-t_i)f_{3D}(x,y,z; S, M_i),
\]

(1)

where \(H_t\) representing historical events before the fitting period or events beyond the study region. In the above equation, \(\mu(x,y,z)\) is the background seismicity rate estimated at location \((x,y,z)\), \(\kappa(M_i)\) denotes the aftershock productivity of an event with magnitude \(M_i\):

\[
\kappa(M_i) = Ae^{\alpha(M_i-M_c)},
\]

(2)
where $M_c$ is the cut-off magnitude. The probability density function (PDF) in time $g(t)$ is assumed to follow the empirical Omori-Utsu law in which aftershocks decay according to

$$g(t) = \frac{p-1}{c} \left(1 + \frac{t}{c}\right)^{-p}.$$  \hfill (3)

The spatial PDF $f_{3D}(x, y, z; S_i, M_i)$ is further decoupled into two terms: The epicenter PDF $f_{2D}(x, y; S_i, M_i)$ and depth PDF $h(z; S_i)$:

$$f_{3D}(x, y, z; S_i, M_i) = f_{2D}(x, y; S_i, M_i) h(z; S_i),$$  \hfill (4)

where

$$f_{2D}(x, y; S_i, M_i) = \frac{q-1}{\pi D'^2 e^{\gamma(M_i - M_c)}} \left(1 + \left(\frac{x - x_i}{D}\right)^2 + \left(\frac{y - y_i}{D}\right)^2\right)^{-q},$$  \hfill (5)

and

$$h(z; S_i) = \frac{\left(\frac{z}{Z}\right)^{\frac{q}{2}} \left(1 - \frac{z}{Z}\right)^{\left(\frac{1}{2} - \frac{q}{2}\right)}}{ZB \left(\eta \left(\frac{z}{Z}\right) + 1, \eta \left(1 - \frac{z}{Z}\right) + 1\right)}.$$  \hfill (6)

Note that we assume the focal depth to follow the Beta distribution, and thus it is easy to satisfy the fact that earthquakes are distributed in a limited seismogenic layer by defining the maximum depth $Z$ in Equation 6.

Equations 5 and 6 give the spatial response kernel of point-source events, whose scaling effect in the epicenter is exponentially proportional to the magnitude. For simplicity, the scaling effect in the depth is neglected. That is, different events share the same aftershock distribution in depth, although their magnitudes might be different.

For the large earthquakes with ruptures extending to tens of kilometers or more, it is necessary to incorporate the rupture geometries into the model. In the 3D-FS ETAS model (Guo et al., 2019), the spatial kernel of mainshocks is written as

$$f_{3D}(x, y, z; S_i, M_i) = \int \int [f_{2D}(x - u, y - v, z; w) \tau_i(u, v, w) du dv dw] \int \int \tau_j(u, v, w) du dv dw,$$  \hfill (7)

where $\tau(u, v, w)$ is the productivity of any location $(u, v, w)$ on source $S_i$, and the integral kernel is

$$f'_{3D}(x - u, y - v, z; w) = \frac{q-1}{\pi D'^2} \left(1 + \left(\frac{x - u}{D}\right)^2 + \left(\frac{y - v}{D}\right)^2\right)^{-q} h(z, w).$$  \hfill (8)

Note that we introduce the new parameter $D'$ to account for the scaling effect of finite sources. Therefore, there are in total 9 parameters ($\theta = (A, \alpha, c, p, D, D', q, \eta, \eta)$) to be estimated in the 3D-FS ETAS model.

To extract background earthquakes from the catalog, the ETAS model introduces the probabilistic description of triggering relationship between mainshocks and aftershocks by using the following definitions: $\varphi$, the probability of event $j$ to be a background earthquake, and $\rho$, the probability of event $j$ triggered by previous event $i$, as follows

$$\varphi_j = \frac{\mu(x_j, y_j, z_j \mid \mathcal{H}_j)}{\lambda(t_j, x_j, y_j, z_j \mid \mathcal{H}_j)},$$  \hfill (9)
Suppose that $S_i$ is partitioned into many patches, say $S_{n_i}, i = 1, 2, ..., n_i$. Similarly, the probability of event $j$ triggered by patch $i$ on finite source $S_{n_i}$ i.e., $\rho_{ij}$ can be defined as:

$$\rho_{ij} = \left\{ \begin{array}{ll} \frac{\int \int \int f_{3D}(x_j-u, y_j-v, z_j; w)\tau(u, v, w)du dv dw}{\lambda(t_j, x_j, y_j, z_j) H_j} & t_j > t_i; \\
0 & \text{otherwise.} \end{array} \right. \tag{11}$$

Note that the probability of event $j$ to be a triggered event $\rho_j = \sum_{i=1}^{n} \rho_{ij}$, and $\varphi_j + \rho_j = 1$. In the case of $j = 1$, $\rho_0 = 0$, $\varphi_1 = 1$.

Equations 9 and 10 give the criterion of how to discriminate the directly or indirectly triggered aftershocks. Given a mainshock $i$, we can realize the process of its direct offspring by keeping each event $j, j = i + 1, ..., N_i$, with probability $\rho_{ij}$. The stochastic declustering method proposed by (Zhuang et al., 2002) is based on such an idea and can classify the events in a catalog into different family trees. The inherent uncertainty of such an approach is accounted for through the seeds of random numbers used in the selection and quantified by repeating the selection for many times.

Guo et al. (2017) introduce an iterative algorithm for the 2D-FS ETAS model, which can simultaneously estimate the model parameters, background seismicity, and aftershock productivity. In this study, we modify this algorithm for the 3D-FS ETAS model. Please refer to the appendix for the detailed steps of estimating the parameter vector $\Theta$, background seismicity $\mu(x, y, z)$, and aftershock productivity $\tau(u, v, w)$.

The forecasting ability of the 3D-FS ETAS model can be tested through simulations, as done by Guo et al. (2018) for the 3D-PS (point source) ETAS model. The general simulation algorithm of the ETAS model can be found in Zhuang and Touati (2015). The forecasting performance of different models is evaluated through the information gain score (Vere-Jones, 1998), defined by $G = \sum X_k \log \frac{\tilde{P}_k}{\tilde{P}_{ok}} + (1 - X_k)\log\frac{1 - \tilde{P}_k}{1 - \tilde{P}_{ok}}$, where $\tilde{P}_k$ and $\tilde{P}_{ok}$ represent the probabilities that at least one event occurs in cell $k$ from the test and reference models, respectively. $X_k = 1$ or 0 denotes cell $k$ is or is not occupied by events, respectively. In this study, we use the 2D-PS ETAS model as the reference model.

## 3. Data Set and Computational Settings

The catalog data is downloaded from the California Earthquake Data Center (SCEDC, https://service.scedc.caltech.edu/ftp/catalogs/SCEC_DC/, last accessed: December 17, 2019) for the period from 1980 to 2019. The ruptures of four mainshocks are taken into account: the 1992 Landers earthquake, the 1994 Northridge earthquake, the 1999 Hector Mine earthquake, and the 2014 Ridgecrest earthquake. The focal mechanisms of four mainshocks are shown in Figure 1, referred to the Global Centroid Moment Tensor solution (Ekström et al., 2012). A brief introduction of four mainshocks is given below:

1. The 1992 $M_{w}7.3$ Landers earthquake is the largest mainshock investigated in this study. It is dominated by a right-lateral strike-slip mechanism (strike/dip/rake: 341°/70°/−172°) with coseismic rupture extended to 70 km, and its aftershock zone extends to over 100 km along the ruptured fault (Zeng & Anderson, 2000; Hauksson et al., 1993).
2. The 1994 $M_{w}6.7$ Northridge earthquake occurs on a buried thrust fault beneath the San Fernando Valley, with strike/dip/rake being 130°/53°/111° according to the GCMT solution. The mainshock hypocenter located at about 19 km, and its aftershock sequence extended to a depth of 23 km into the lower crust (Hauksson et al., 1995).
3. The 1999 Hector Mine $M_w$7.1 earthquake occurs on a right-lateral strike-slip fault (strike/dip/rake: $336^\circ/80^\circ/174^\circ$), only about 20 km away from the Landers earthquake. Simulation results of regional viscoelastic stress changes after the Landers earthquake showed that the occurrence of the Hector Mine earthquake might be promoted by the Landers earthquake (Freed & Lin, 2001).

4. The 2019 Ridgecrest earthquake sequence is initiated with a $M_w$6.4 left-lateral strike-slip foreshock, and the $M_w$7.1 mainshock occurs about 34 h later. According to the GCMT solutions, the foreshock and mainshock occur on a left-lateral fault (strike/dip/rake: $227^\circ/86^\circ/3^\circ$) and a right-lateral fault (strike/dip/rake: $321^\circ/81^\circ/180^\circ$), respectively, presenting cross-fault ruptures. The surface rupture of the mainshock extends to about 46 km with peak slip values of about 5 meters (Barnhart et al., 2019).

The source boundaries of the four mainshocks and the Ridgecrest foreshock are constrained by early aftershocks (aftershocks occurred within 1 day following the mainshocks, shown in Figure 1). The vertices of the source polygons are determined relatively arbitrarily by following the principle that the on-fault aftershocks should locate inside the mainshock source. The maximum depths for the Landers, Northridge, Hector Mine, and the Ridgecrest mainshocks are set to 17, 25, 16, and 20 km, respectively, according to the depth distributions of aftershocks. The source of Ridgecrest foreshock has the same maximum depth as the Ridgecrest mainshock. We divide the source volumes into $0.01^\circ \times 0.01^\circ \times 1$ km grids, with each grid triggering aftershocks isotropically as Equation A4. The magnitude threshold $M_c$ is set to 3.0 to include more events in the analysis, although there are some small events missing immediately following the mainshocks (Figure 2).
4. Results

The fitting results of the 3D-PS and 3D-FS ETAS models are shown in Table 1. The most significant change is in the parameter pair $\alpha$ and $A$, which control the aftershock productivity term in Equation 2. The value of $\alpha$ is usually underestimated in the 3D-PS ETAS model, and $A$ is overestimated as a trade-off effect (Guo et al., 2017, 2019; Hainzl et al., 2008). A smaller $\alpha$ value means less difference between large and small earthquakes in producing aftershocks, and thus the productivity of large earthquakes is underestimated. We attribute this biased estimation to the ignorance of rupture geometries of mainshocks in the 3D-PS ETAS model. Using the 3D-FS ETAS model, the $\alpha$-value increases from 1.02 to 1.70, and $A$ decreases from 0.58 to 0.25 in Table 1. For an earthquake of magnitude 7.0, its expected number of aftershock productivity increases from 34.3 to 224.5 in the 3D-FS ETAS model, suggesting that the incorporation of rupture geometries significantly enhances the productivity of large earthquakes, and an $\alpha$ value of 1.70 is closer to the value obtained from the temporal ETAS model (Hainzl et al., 2008).

![Figure 2. Magnitude-serial number plots for mainshocks: (a) Landers, (b) Northridge, (c) Hector Mine, and (d) Ridgecrest. The mainshock is marked as the red pentagram in each plot. The top horizontal axis represents the time relative to the mainshock with negative numbers being days before and positive numbers being minutes after the mainshock.](image)

| Model       | Unit | $A$   | $\alpha$ | c    | $D^2$ | $q$   | $\gamma$ | $D^2$ | $\eta$ | log $L$ |
|-------------|------|-------|----------|------|-------|-------|----------|-------|--------|---------|
| 3D-point    |      | 0.58  | 1.02     | 0.46 | 1.10  | 1.48  | 1.57     | 1.07  | –      | 53.8    | 2063.3  |
|             |      | (0.02)| (0.02)   | (0.04)| (0.01)| (0.11)| (0.02)   | (0.06)| –      | (2.2)   | –       |
| 3D-finite   |      | 0.25  | 1.70     | 0.72 | 1.11  | 0.95  | 1.69     | 1.38  | 1.72   | 65.2    | 2916.8  |
|             |      | (0.01)| (0.02)   | (0.07)| (0.01)| (0.08)| (0.02)   | (0.06)| (0.21) | (3.2)   | –       |
Although investigating the spatial variation of the model parameters is beyond the scope of this study, it is worth mentioning that the parameter variations could be very significant in the study region (Nandan et al., 2017), which could result in the difference in model fitting. Therefore, the results shown in Table 1 only represent the average model fitting for the whole study region.

### 4.1. Aftershock Productivity

Figures 3–6 present the density distributions of aftershock productivity for four mainshocks obtained from the 3D-FS ETAS model. As we can see from these results, the aftershock productivity has strong heterogeneity in space. For the Landers earthquake in Figure 3, high density values mostly exist in shallow depth...
Generally, the areas of high density values indicate that aftershocks are highly clustered in these areas. These productivity asperities formed by high density areas have several local centers and extend to different depth ranges. To the north part, high productivity exists from 0 to 10 km, while to the middle and south parts, the asperities are mainly distributed in the range of 2–6 km. The total productivity obtained from the 2D-FS ETAS model is 336.2, as shown in Figure 3j, consisting of about 20.5% of all events in the source region after the mainshock.

In Figure 4, areas of relatively high productivity the Northridge earthquake are mainly from 2 to 15 km, showing significant migrations with depth. From the shallow part to the middle part, the center of productivity asperity seems to rotate in the clockwise order, with location changing from northeast to southeast of the mainshock source. The high productivity in the layers shallower than 15 km reflects the complex and densely distributed sub-faults in the hanging wall of the rupture fault (Hauksson et al., 1995). A small portion of aftershock productivity distributes between 16 and 20 km layers, which is rare in Southern California region, indicating a complicated rupture process of the mainshock and coseismic ruptures may extend...
The productivity patterns of the Hector Mine earthquake are shown in Figure 5. High productivity values mostly distribute in layers less than 10 km, with a significant decaying trend in depth: The total productivity takes maximum value of 42.7 at 0–1 km-layer, and decay rapidly as it goes deeper. At 10–11 km-layer, The productivity decrease to only 4.4 (Figure 5f), less than 2% of the total productivity. There are two aftershock asperities within the source region, the middle one takes high values from 0 to 10 km, while the one in the north end of the source emerges at 4 km and diminishes around 10 km. The total productivity from the 2D-FS model is 247.8 (Figure 5i), which suggests that nearly 40% of all 617 earthquakes are directly triggered by the mainshock.

The aftershock productivity distributions of the latest Ridgecrest earthquake are presented in Figure 6. The areas with high productivity values form two aftershock asperities from 2 to 10 km, one in the north end to as deep as 20 km. The total productivity in Figure 4n is 150.5, about 28.6% of all events are classified as directly triggered aftershocks.

Figure 5. The aftershock productivity density patterns of the Hector Mine earthquake. The illustration is the same as Figure 3.
and the other in the south part of the mainshock source. About 33.6% of all aftershocks (293 out of 871) are produced by the Ridgecrest earthquake, as the number of total productivity in Figure 6j shows. Most strong aftershocks (M4.5+ in Figures 6c–6h, m4.8+ in Figure 6k) are located in the north end of the source, suggesting that the triggering potentials of the mainshock are highly concentrated in this area. Less than 1% of the total productivity occurs in layers deeper than 15 km, which is consistent with the Landers and Hector Mine earthquakes.

It is interesting to investigate the relationship between the Ridgecrest foreshock and mainshock since one of the powers of the ETAS model is to quantify the triggering relationship between two earthquakes (Jia et al., 2014). The total productivity of the Mw 6.4 foreshock is 86.9, about 29.7% of the mainshock productivity. The number of events between the foreshock and the mainshock is 138, of which about 67 are directly triggered by the foreshock, making of 48.7% of all events. Moreover, we inspect the probability of the Ridgecrest mainshock triggered by previous foreshocks. We find that the mainshock can be recognized as directly triggered by the Mw 6.4 foreshock at about 10%, and nearly 70% chance of being directly triggered by a closer

![Figure 6](image-url)

Figure 6. The aftershock productivity density patterns of the Ridgecrest earthquake. The illustration is the same as Figure 3.
M5.4 event (2.6 km), which occurred 0.68 days before the mainshock. After the mainshock, only a small proportion of aftershocks are classified as aftershocks directly triggered by the foreshock.

4.2. Distribution of Large Aftershocks

It has been reported in some literatures that large aftershocks tend to be located at the edges other than the centers of areas where aftershocks are highly clustered (van der Elst & Shaw, 2015; Stallone & Marzocchi, 2018). This finding holds true for the 2016 Kumamoto earthquake sequence in Guo et al. (2019), which demonstrate that most of M5.0+ aftershocks following the mainshock nucleate at the edges of aftershock asperities formed by high productivity density areas. To investigate the distribution of strong aftershocks, we mark the epicenters of M4.5+ events in the aftershock productivity density patterns (Figures 3–6). It is easy to see that most of the aftershocks are located at the boundaries of the asperities with high density values, and very few of them are located at the centers.

To further test the significance of the spatial dependency of aftershock magnitudes, we plot the ROC (receiver operating characteristic) curves in Figure 7. In each panel of Figure 7, the diagonal line denotes the cumulative number of M3.0+ aftershocks, which follows exactly the ratio of 1 to 1. Significant deviations are found for the curves of M4.0+ and M4.5+ aftershocks of the Landers, Northridge, and Hector Mine.
earthquakes: Over 80% of M4.5+ aftershocks are distributed in areas with cumulative productivity less than 60% of the total productivity, and more than half of them are located in areas with productivity less than 30% of the total. This suggests that larger aftershocks of these three mainshocks tend to be located in areas with relatively low to moderate productivity values.

However, the results of the Ridgecrest mainshock in Figure 7d show that more M4.5+ aftershocks occur in areas with relatively high productivity, making it an exception of the four mainshocks. As the red curve in Figure 7d shows, the M4.5+ aftershocks accumulates slowly in areas with low productivity, and then increases rapidly by nearly 40% in the interval between 0.6 and 0.8 times of the total productivity. We attribute this discrepancy to the dense distribution of large aftershocks in the northwestern region. As shown in Figure 6, most of the M4.5+ aftershocks are located in the northwestern end of the source, where the coseismic rupture is terminated (Liu et al., 2019; Ross, Idini, et al., 2019). Among these aftershocks, two largest ~ M5.5 aftershocks are located in the north end near 35.9°N, and they have 62 and 8% to be directly triggered by the mainshock, respectively.

Regional fault geometries revealed by the focal mechanisms of aftershocks in Wang and Zhan (2020) suggest that the faulting in the northwestern region of Ridgecrest earthquake is dominated by a series of splay faults and antithetic faults. Since this region is close to the northern termination of the mainshock rupture, the stress changes imparted by the mainshock on local seismogenic faults are substantial. The densely distributed sub-faults with complex geometries make this region vulnerable to stress perturbations, and 19 M4.5+ aftershocks and numerous small aftershocks indicate that the strain energy has been released in the postseismic period. Therefore, the reason why the large and small aftershocks of the Ridgecrest earthquake are more overlapped than aftershocks of the other mainshocks is because of complex regional faulting in the northwestern end of the source. Nevertheless, the larger aftershocks are located in the outer side of the productivity asperity at the local scale as shown in Figure 6, which is consistent with the finding in van der Elst and Shaw (2015) and Stallone and Marzocchi (2018).

4.3. Temporal Migration of Aftershocks

Because of the probability description of earthquake triggering in the ETAS model, we are able to disentangle aftershock productivity of the mainshock from the aftershock sequence within different time scales. The temporal migrations of mainshock productivity and all aftershocks (including high order aftershocks) within 0.1, 0.1–1, 1–10, and 10–100 days following the four mainshocks, are presented in Figures 8 and 9, respectively.

For the Landers earthquake, direct aftershocks (Figure 8a1–a4) exist in both ends of the source in 1 day, and then extend to the whole source region. In contrast, high order aftershocks (Figure 9a1–a4) distribute more broadly within the mainshock source. An aftershock asperity mainly composed of high order aftershocks is found in the south end, lasting from 0.1 to 100 days. Meanwhile, many high order aftershocks occur to the west of the source because the largest off-fault aftershock – the Mw6.5 Big Bear earthquake – is located in this area.

Most of aftershock productivity of the Northridge earthquake occur in the east part of the source, as shown in Figure 8b1–b4. The largest asperity formed in the productivity patterns does not vary much within 10 days. However, an apparent migration of all the triggered aftershocks is observed in Figure 9b1–b4, which starts in the northwest part in 1 day, and then migrates to the middle part in 100 days.

In Figure 8c1–c3, most of the productivity occurs in the middle part around 34.6°, and high productivity asperity extends to more than 20 km along the rupture. In the north end of the Hector Mine source (Figure 9c1 & c4), about 136 indirect aftershocks occur in 1–10 days. This asperity extends to a larger area in 10–100 days, indicating the later aftershocks in the northern part are more diffusive.

The Ridgecrest earthquake triggers most direct aftershocks among all the four mainshocks within 0.1 days, about 155 of all 287 aftershocks, mostly located in the north end of the source. This suggests that the detection rate of the Ridgecrest aftershock sequence is better than that of the other three mainshocks. High order aftershocks in this area exist in the whole period as Figure 9d1–d4 show. In Figure 8d1–d4, high
productivity values extend to the whole south part in 1 day, and then shrink into smaller areas at both ends of the rupture within 1–100 days.

It is worth pointing out that the aftershock productivity patterns reconstructed in 0.1 days may be incomplete due to the missing issue of early aftershocks. As a result, the spatial distribution and total numbers of aftershocks in Figures 8 and 9 only represent for the aftershocks recorded in the catalog.

4.4. Productivity on Fault Plane

To directly compare aftershock productivity with coseismic slip distributions, we interpolate aftershock productivity onto the mainshock fault planes. The results are shown in Figure 10 (also in Figures S1 and S2), and the slip distributions of four mainshocks are cited from Kaverina et al. (2002), Liu et al. (2019), and
Zeng and Anderson (2000). Note that due to the non-uniqueness of solutions to the inverse problem, there are usually several available slip distributions for the same mainshock. Generally, these solutions share same characteristics on a large scale but differ in detail. Here we focus on the large scale of slip distribution, and take three slip distributions as references for each mainshock.

In Figure 10, the patches with high productivity values form compensating patterns to the high slip areas on the whole scale. For the Landers earthquake, aftershock productivity is mainly distributed in the shallow part of the fault. The largest aftershock asperity (productivity value > 0.1) extends to more than 5 km in depth and 10 km along the fault trace, located between the two large slip asperities in Figure 10a.

High productivity values of the Northridge earthquake exist on the upper-right corner of the fault plane (Figure 10d), while large slip is found in the moderate-to deep-depth part of the fault, with the largest slip in the lower-left part. It is worth noting that the aftershock productivity distributes in a wider range as shown...
Figure 10. The coseismic slip and interpolated aftershock productivity density distributions of four mainshocks. (a) and (b), Landers; (c) and (d), Northridge; (e) and (f), Hector Mine; (g) and (h) Ridgecrest.
in Figure 4, the productivity pattern in Figure 10d only presents the top right part of the whole aftershock asperity.

From the results of the Hector Mine earthquake (Figure 10e and 10f) we can see that the two fault segments in the north produce most of direct aftershocks, and the largest productivity occurs at the place where the two segments are connected. Although the majority of slip is located on the western branch of the fault, the majority of aftershock productivity is located in the shallow part of the eastern branch, suggesting that the area near the eastern branch is subject to significant stress increases.

The slip and aftershock productivity patterns of the Ridgecrest earthquake in Figure 10g and 10h are somewhat compensating on the whole scale. The highest productivity value locates in the north end of the fault, where coseismic slip is relatively low. The productivity is mainly distributed between 5 km and 15 km, while the slip extends to as deep as 20 km in the middle part.

The productivity distributions of the mainshocks are interpolated on the fault planes of slip models obtained by other researchers. The productivity patterns show slight differences since the fault geometry and slip distributions of the same mainshock from different studies are different. However, the on-fault productivity patterns share the same feature: there are almost no significant aftershocks located in high slip areas (Figures S1 and S2).

To further quantify the productivity distribution with respect to coseismic slip, we plot the ROC curves in Figure 11. From Figure 11 we can see that nearly 60% of the aftershock productivity or more is located in the low slip areas (less than 0.3 times of the maximum slip) in all cases. For the Northridge earthquake, the productivity accumulates over 80% in areas with low slip in Zeng’s and Dreger’s models (Dreger, 1997; Zeng & Anderson, 2000). Similar high proportion of productivity in low slip areas also exists in Ross’s model (Figure 11d) for the Ridgecrest earthquake.

4.5. Background and Cluster Seismicity

The patterns of background seismicity rate at different depths obtained from the 3D-FS ETAS model are shown in Figure 12. The total number of background events from 1980 to 2019 in the region −119.5° to −115.5°W, 32.5° to 36.5°N is 2,127, making up of 23.6% of all events. The background seismicity shows a strong spatial heterogeneity in both the horizontal and depth components. In the Coso volcanic region near (−117.8°, 36.0°), high values mainly distribute in depth shallower than 10 km. While for the areas in southeast Ventura (−119.2°, 34.3°), around Salton Sea (−115.8°, 33.3°), and to the east of San Bernardino (−117.3°, 34.1°), high background seismicity rate extends to nearly 20 km in depth. The high background rate may correlate with local tomography: Regions of high background rate at deep layers correspond to high shear-wave velocity, and the Coso region where high background rate exists at shallow layers corresponds to low shear-wave velocity (Tape et al., 2009).

The numbers of cumulative triggered events versus time in the source regions of four mainshocks are presented in Figure 13. From the curves of the cumulative probabilities of triggered events (1 − φi, with φ defined in Equation 9), we can see that both the 3D-PS and 3D-FS ETAS models yield similar results for all the four mainshocks. This implies that the underestimated aftershock productivity of large earthquakes in the point source model is compensated by the overestimated high order aftershocks. Meanwhile, the numbers of aftershocks directly/indirectly triggered by the mainshocks are significantly enhanced in the 3D-FS ETAS model, which suggests that the 3D-FS ETAS model corrects the underestimated triggering ability of mainshocks in the 3D-PS ETAS model. As addressed in Zhuang et al. (2018), it is necessary to incorporate the rupture geometries of the mainshocks into the ETAS model to reasonably classify the family trees in an earthquake catalog.

In Figure 14, the cumulative ratios of aftershocks directly triggered by the mainshocks are presented. The ratio values are calculated by \[ \frac{\sum_{t_j<t}I(t_j)}{\sum_{t_j<t}I(t_j)} \] where \[ R(t_j) < t \] takes 1 if \( t_j < t \) holds true, and 0 otherwise. The ratios of aftershock productivity for the Landers, Northridge, Hector Mine, and Ridgecrest earthquakes are 44.5, 44.6, 44.6, and 36.5, respectively, in 50 days after the mainshocks. The low ratio in the rupture areas of Ridgecrest earthquake may be related to the high heat flow in this region (Blackwell et al., 2011), while the high ratios of the Landers and Hector Mine earthquakes are related to the low heat flow in their
source regions. The source region of the Northridge earthquake is characterized by relatively low heat flow, however, the ratio of direct aftershocks is low. We explain this discrepancy as the deep slip asperities of the mainshock (Figure 10c) and relatively small moment release ($\sim M_{w}6.7$), causing small changes in stress to the shallow part and thus leading to less direct aftershocks.

Our results are somewhat consistent with Zaliapin and Ben-Zion (2013), which claims that the proportion of direct aftershocks in the geothermal region is generally smaller than that in the non-geothermal region. Zaliapin and Ben-Zion (2013) categorize the seismicity in Southern California into burst-like and swarm-like clusters. The former having more direct aftershocks likely reflect brittle dominated failures in relatively cold regions, while the latter are more likely dominated by brittle-ductile failures in regions with high temperature or fluid content. They also point out that for aftershock sequences of M7+ mainshocks, both basic cluster types are supposed to exist simultaneously. Therefore, the low ratios of direct aftershocks of Northridge and Ridgecrest mainshocks suggest a higher proportion of swarm-like clusters in the aftershock
sequences. Additionally, the low ratio of direct aftershocks in the Ridgecrest rupture is consistent with the low $\alpha$ values obtained by Enescu et al. (2009) in the Coso region, who also attribute the low productivity efficiency of larger events to high heat flow in corresponding areas.

5. Simulation Test

To test whether the information of aftershock heterogeneity can improve the performance of aftershock forecasts, we carry out the simulation test for four mainshocks. The time start of the simulation is set to 0.1 days after the mainshock. For each time step 1 day, we simulate 10,000 times to reproduce the aftershocks within the target regions. The maximum and minimum magnitudes of aftershocks are 6.0 and 3.0, respectively. The $b$-values obtained from the four aftershock sequences are 0.9, 0.8, 0.9, and 0.9 for the Landers, Northridge, Hector Mine and Ridgecrest earthquakes, respectively. The corresponding $b$-value is used when sampling the magnitudes of simulated events of the mainshocks.
It is easy to carry out the simulation of the 3D-PS ETAS model by following the algorithm in Guo et al. (2018). However, it is unreasonable to use the patterns of aftershock productivity in Figures 3–6 as the proxies of mainshock sources in the simulation of the 3D-FS ETAS model, since these patterns are reconstructed from all the events till 2019. Therefore, we use the locations of early aftershocks to represent the centers of the patches on the finite source. In the first time step from 0.1 ∼ 1.1 days after the mainshock, the source is defined by aftershocks within 0.1 days, and the source in the second step is defined by aftershocks within 1.1 days, and so on. To restrict the locations of source patches in the rupture areas of mainshocks, we take the probabilities of events to be directly triggered by the mainshocks defined by Equation 10 as the weights of patches. Moreover, the weights are further revised by multiplying an Omori-Utsu type factor $(\Delta t + c)'$, where $\Delta t$ denotes the time difference between the aftershock and the mainshock, $c$ is a small value taken from Table 1, and $p'$ takes 0.0, 0.5, 1.0, 1.5. In this way, we would like to test whether the temporal migration of next-day aftershocks can be tracked and recovered by earlier aftershocks.

Figure 13. The cumulative probabilities of earthquakes that are triggered by any event, directly/indirectly triggered by the mainshock, for source region: (a) Landers, (b) Northridge, (c) Hector Mine, and (d) Ridgecrest. The meanings of the legends are: “point-*,” results from the 3D-PS ETAS model; “finite-*,” results from the 3D-FS ETAS model; “*-D”: directly triggered aftershocks of the mainshock; “*-A”: all triggered aftershocks of the mainshock. The lower panel of each plot shows the magnitude-time distribution of events.

Figure 14. The cumulative percentages of direct aftershocks in all triggered events of four mainshocks. The colored dots mark the occurrence times of aftershocks of corresponding mainshocks.
The simulation results from the 3D-PS and 3D-FS ETAS models are compared by estimating the information gain scores (e.g., Vere-Jones, 1998; Zhuang, 2011). The forecasting regions of four mainshocks are divided into 0.05° × 0.05° × 2.5 km cells. All cells are separated as event cells and non-event cells, with the former having at least one event and the latter having no events. The information gains in the two kinds of cells are separately calculated, that is to say, the better model should have overall positive information gains in the whole region. To make comparisons, we take the results from the 2D-PS ETAS model as reference, in which the focal depths are assumed to follow the uniform distribution. The information gains of the 3D ETAS models are shown in Figure 15. Overall speaking, the 3D versions perform significantly better than the 2D-PS version, and the 3D-FS version performs better than the 3D-PS version. For the 3D-FS ETAS model, the maximum values of total information gains for the Landers, Northridge, Hector Mine and Ridgecrest earthquakes are 300.4, 50.6, 215.8, and 200.9, when setting $p'$ values at 0.5, 1.5, 1.5, and 1.5, respectively. In contrast, the total information gains for the 3D-PS model in the four cases are 234.0, 28.2, 165.1, and 193.8, respectively.

The performances quite vary for the first day, especially for the Northridge earthquake, which shows negative or low positive information gains in the 3D versions (Figure 15b). Unlike the other three strike-slip mainshocks, the Northridge earthquake occurs on a buried thrust fault. As a result, its aftershocks distribute more diffusively in space. In Figures S4–S7, we plot the depth distributions of the simulated aftershocks in the first 3 days following the mainshocks. These plots suggest that the 3D versions can only reconstruct the depth distribution of aftershocks in the learning period, and may not perform well when the aftershocks migrate significantly in the forecasting period. In the case of Northridge earthquake, the aftershock distribution within 0.1–1.1 days cannot be well reconstructed by aftershocks within 0.1 days in the 3D versions because the diffusive distribution of aftershock locations is difficult to reproduce.
Among the 3D ETAS models, the FS version performs significantly better than the PS version, except for the case of the Ridgecrest earthquake. In the first day, the forecasted numbers of aftershocks from the 3D-FS version are about 2~3 times of those from the 3D-PS version (Figure S3). But for for the Ridgecrest earthquake, it is only about 1.3 times. The large aftershock number (287; Figure 9) within 0.1 days following the Ridgecrest mainshock enhances the forecasted numbers of the first-day aftershocks in the 3D-PS ETAS model.

For the second to the tenth days following the mainshock, the 3D-FS ETAS model always improves after-shock forecasts, with informance gain decreasing with time. This is because the direct triggering effect of the mainshock dominates the aftershock occurrences in this period and decays away after it.

The $p'$ also influences the forecasting performance. In most cases, a higher $p'$ yields higher performance gain. For the case of the Landers earthquake (Figure 15a), forecasts made by $p' = 1.5$ are just slightly worse than $p' = 0.5$.

In summary, the enhanced productivity of mainshocks in the 3D-FS ETAS model leads to larger numbers of simulated aftershocks within the first several days (Figure S3). The aftershock numbers decay rapidly and the forecasted aftershock numbers of the 3D ETAS models are almost the same after a few days of the mainshocks. Therefore, the information gains of the 3D-FS ETAS model with respect to the 3D-PS ETAS model is nearly gone to 0 after 2~4 days.

6. Remarks

In this study we fit the 3D-FS ETAS model to the Southern California catalog. This model incorporates the 3D fault geometries of large earthquakes when modeling the triggering relationships between mainshocks and aftershocks. Four mainshocks with magnitudes larger than M6.5 are taken as finite sources, which include three strike-slip events: The Landers, Hector Mine, Ridgecrest earthquakes, and one thrust event: the Northridge earthquake. We use early aftershocks as constraints for source boundaries of mainshocks, and introduce an iterative algorithm to estimate the spatial distribution of aftershock productivity within the mainshock source.

Compared to the result from the 3D-PS ETAS model, the 3D-FS ETAS model corrects underestimated aftershock productivity of large mainshocks by obtaining a higher $\alpha$-value. For an earthquake of magnitude 7.0, the expected number of direct aftershocks increases by more than 5 times in the finite source model. The reconstructed direct aftershocks occupy 21% to 41% of all the events within the source regions after the mainshocks.

Aftershock productivity within source regions shows a strong spatial heterogeneity with high values mainly distributing in depth layers shallower than 15 km. Aftershocks of the three strike-slip mainshocks mostly distribute along the rupture faults, particularly concentrating at the ends of mainshock ruptures. For the Northridge earthquake, many aftershocks composed of all faulting types occur in the shallow layers, of which the shallow thrust events reflect the folding formation of secondary faults in the hanging wall, the normal and strike-slip aftershocks release the extensional strains in the handing wall imposed by the mainshock (Hauksson et al., 1995). Note that the productivity results obtained from the FS ETAS models are subject to the missing aftershocks immediately following the mainshocks, which could provide more insights on the physical mechanisms of aftershock generations (Kato & Nakagawa, 2014; Peng & Zhao, 2009; Wu et al., 2017; Zhuang et al., 2017).

The hypothesis of larger aftershocks occurring in areas of relatively low productivity holds true for three mainshocks except for the Ridgecrest earthquake as shown by the ROC curves in Figure 7. We attribute the relatively dense distribution of large aftershocks in the north end of the mainshock source to the densely distributed sub-faults in this region, whose geometries are rather complex (Wang & Zhan, 2020). In the local scale, the M4.5+ aftershocks in the north end (Figure 6) still locate in the outer side of the productivity asperity.

Significant temporal migrations of aftershocks of the four mainshocks are observed in Figures 8 and 9, which is impossible to be solely explained by coseismic stress changes. In fact, the spatial temporal
migrations of aftershocks may be driven by more complex mechanisms, for example, the afterslip (Helmstetter & Shaw, 2009). A detailed migration pattern of aftershocks reconstructed from the finite source ETAS model would help us further understand the physical mechanisms of aftershock triggering.

The reconstructed patterns of aftershock productivity enable us to interpolate the productivity values onto the fault planes of four mainshocks, and make a direct comparison with coseismic slip distributions. On the whole scale, the slip and aftershock productivity form complementary patterns, which indicates that aftershocks play a role in the postseismic relaxation and respond to coseismic stress changes. Almost no aftershocks are observed inside the large slip areas, suggesting that these areas which are sufficiently ruptured during the mainshock have limited triggering potential. As the result, the subsequent aftershocks tend to nucleate near the periphery of large slip asperities. This result is consistent with previous studies about the aftershock distributions on/near the mainshock ruptures. We quantitatively verify this conclusion, with the uncertainty of slip distribution being taken into account.

High background seismicity rate extends to different depths at different places in Southern California, presenting a very heterogeneous distribution. A high resolution of background seismicity patterns in depth is beneficial to the seismic hazard assessment. Compared with the tomography data in Southern California, we find that the background rate at depth may correlate with shear-wave velocity in corresponding regions, with shallow background seismicity in low-velocity region and deep background seismicity in high-velocity region.

The proportions of direct aftershocks in the source regions of four mainshocks differ significantly. A relatively low proportion is found in the source region of Ridgecrest earthquake, which is near the Coso volcanic region. This finding is consistent with previous studies (Enescu et al., 2009; Zaliapin & Ben-Zion, 2013), which conclude that the aftershocks in the geothermal region are more likely belong to the swarm-like type, characterized by narrow angular distribution and larger proportions of high order aftershocks. Although the 3D-FS ETAS model significantly enhances the mainshock productivity, the indirect triggering still plays an important role in the whole aftershock sequence, as the productivity ratios decrease to below 50% in several days after the mainshocks (Figure 14).

In simulations, the 3D models fail to reproduce the first-day aftershocks of the Northridge earthquake. We argue that this is because the Northridge earthquake is a thrust event and thus has more diffusive aftershocks. The aftershock response kernels in the ETAS model decay rapidly according to the $p$ and $q$ values in the time and space PDF. Therefore, it is difficult for the ETAS model to reproduce the aftershock hypocenters by earlier aftershocks when the aftershock sequence is diffusively distributed in space. Moreover, the incorporation of the temporal factor $(\Delta t + c)^p$ in the 3D-FS model slightly improves the forecasting performance when the $p'$-value takes non-zero values. Note that a larger $p'$-value means the later aftershocks in the learning period weight more in producing aftershock locations. Hence, the latest aftershocks are important to capture the migration fronts of aftershocks.

In summary, the 3D-FS model enhances the triggering ability of large earthquakes, indicating a better capability of reproducing directly triggered aftershocks within the mainshock ruptures. Although it is still difficult to use productivity patterns in real-time aftershock forecasts, the FS ETAS model is capable of improving the forecasting performance by incorporating the constraints from early aftershocks and coseismic slip distributions.

Appendix A: Methodology in Model Estimation

A1. Estimation of $\mu(x, y, z)$ and $\tau(u, v, w)$

In the 3D ETAS model, the total seismicity rate $\hat{m}(x, y, z)$ is estimated through the kernel method,

$$
\hat{m}(x, y, z) = \frac{1}{T} \sum_{j=1}^{N} k_d (x - x_j, y - y_j) \left( \frac{z_j}{Z} \right) \left( \frac{1 - z_j}{Z} \right) ZB \left( \frac{z_j}{Z} + 1, d_z \left( \frac{1 - z_j}{Z} \right) + 1 \right).
$$

where \( d_z \) is the bandwidth of the Beta kernel for the depth component, and \( d_j \) is the bandwidth of the Gaussian kernel for the longitude and latitude components:

\[
k_{d_j}(x,y) = \frac{1}{2\pi d_j^2} \exp\left(-\frac{x^2 + y^2}{2d_j^2}\right).
\]

(A2)

Note that \( d_j \) takes different values for different events defined by the distance between event \( j \) and its \( n \)th nearest event with a minimum cut-off value (0.04° in this study), which should be comparable to the location errors of epicenters.

Combining Equations 9 and A1, we can estimate the background seismicity rate by

\[
\hat{\phi}(x,y,z) = \frac{1}{T} \sum_{j=1}^{N} \varphi \cdot k_{d_j}(x-x_j, y-y_j) \times \frac{\left(\frac{z_j}{Z} \left(1 - \frac{z_j}{Z}\right)\right)}{ZB \left(\frac{d_z}{Z} + 1, d_j \left(\frac{1-z_j}{Z}\right) + 1\right)}
\]

(A3)

To calculate \( \tau(u,v,w) \), we use the discrete form of Equation 7

\[
f_{sd}(x,y,\mathbf{z}; S, M) = \frac{q - 1}{\pi D^2} \sum_{i=1}^{n_i} \tau_i \left[ 1 + \frac{(x-u_i)^2 + (y-v_i)^2}{D^2} \right]^q h(z; w_i),
\]

(A4)

where \( \tau_i \) is the productivity of patch \( i \) on \( S_i \), with centroid located at \( (u_i, v_i, w_i) \), \( P_i = \sum_{i=1}^{n_i} \tau_i \) represents the total productivity of event \( i \), \( n_i \) is the number of patches, and \( i \) runs over all the patches on the finite source \( S_i \). Using Equation 11, it is easy to calculate \( \tau_i \) by the summation \( \sum_{j|\mathbf{r}_j} \rho_j \). However, according to our experience, such simple treatment may heavily weight those patches where aftershocks are highly clustered, and eventually lead to numerical instabilities in parameter estimations. To avoid this, we introduce a smoothness factor in the calculation of \( \tau_i \)

\[
\hat{\tau}_i = \sum_{k=1}^{n_i} \frac{z_d(u_k, v_k, w_k)}{Z_d(i)} = \sum_{k=1}^{n_i} \frac{z_d(u_k, v_k, w_k)}{Z_d(i)} \left[ \sum_{j=1}^{N} \rho_j \right].
\]

(A5)

where the smoothness kernel \( z_d \) takes the form

\[
\frac{\exp\left(-\frac{(u_k-u_j)^2 + (v_k-v_j)^2}{2d_1^2}\right)}{2\pi d_1^2} \left(\frac{w_j}{Z}\right)^d \left(\frac{1-w_j}{Z}\right)^d \frac{\left(1-\frac{w_j}{Z}\right)^d}{ZB \left(\frac{d_z}{Z} + 1, d_j \left(\frac{1-w_j}{Z}\right) + 1\right)},
\]

(A6)

where \( d_1 \) and \( d_2 \) are bandwidth values of the Gaussian kernel in the horizontal components and the Beta kernel in the depth component, respectively. Note that in Equation A5, \( k \) runs over all the patches on \( S_i \), and is normalized by

\[
Z_d(i) = \sum_{k=1}^{n_i} z_d(u_k, v_k, w_k) = \sum_{k=1}^{n_i} z_d(u_k, v_k, w_k) \left[ \sum_{j=1}^{N} \rho_j \right] = \sum_{k=1}^{n_i} z_d(u_k, v_k, w_k).
\]

A2. Maximum Likelihood Estimation

Given the background seismicity \( \mu(x,y,z) \) and the mainshock productivity \( \tau(u,v,w) \), the model parameters can be estimated using the maximum likelihood procedure. The log-likelihood function can be written as (Daley & Vere-Jones, 2003)
Guo et al. (2017) gives the iterative algorithm for simultaneously estimate model parameters, background seismicity $\mu(x, y)$, and mainshock productivity $n(u, v)$ in the 2D-3S ETAS model. In the 3D case, we use a similar algorithm to the model estimation, which is listed below:

1. Set variable Gaussian bandwidth $d_{xj}$ for event $j = 1, 2, \ldots, N$ and fixed Beta bandwidth $d_x$ for the depth component, calculate $\hat{m}(x, y, z)$ by Equation A1.

2. Set $l = 0$, $\mu^{(l)}(x, y, z) = \hat{m}(x, y, z)$, $\tau^{(l)}(u, v, w) = 1$, initial parameters $\theta^{(l)} = (A, \alpha, c, p, D, D', q, \gamma, \eta)$, and smoothness bandwidths $d_j$ and $d_x$ in Equation A6.

3. Renew $\tau^{(l)}(u, v, w)$ by Equation 11 and A5 several times to get stable $\tau^{(l)}(u, v, w)$. Fit $\tilde{\lambda}(x, y, z)$ through the Maximum Likelihood Estimation (MLE) procedure to update the model parameters $\hat{\theta}$.

4. Calculate $\varphi_j$ in Equation 9 and $\rho_{uj}$ in Equation 11 for $j = 1, 2, \ldots, N$, $i = 1, 2, \ldots, n_i$, and $l = 1, 2, \ldots, N_{\text{sub}}$, where $N_{\text{sub}}$ represents the total number of finite sources.

5. Recalculate $\mu(x, y, z)$ by Equation A3, record as $\hat{\mu}^{(l+1)}(x, y, z)$, and re-calculate $\tilde{\tau}(u, v, w)$ for $i = 1, 2, \ldots, N_{\text{sub}}$, record as $\hat{\tau}^{(l+1)}(u, v, w)$.

6. If $\max(\tau^{(l+1)}(x, y, z) - \hat{\tau}^{(l)}(x, y, z)) > \varepsilon$, where $\varepsilon$ is a positive number that small enough to ensure convergence, then set $l = l + 1$, $\theta^{(l+1)} = \hat{\theta}$, and go to step 3. Otherwise, output $\hat{\mu}^{(l+1)}(x, y, z)$, $\hat{\tau}^{(l+1)}(u, v, w)$, and stop.

### Data Availability Statement

The catalog data used in this study is provided by the Southern California Earthquake Data Center, which can be accessed by the following link: [https://service.scedc.caltech.edu/ftp/catalogs/](https://service.scedc.caltech.edu/ftp/catalogs/). The slip distributions of mainshocks are downloaded from [http://equake-re.info/srcmod/](http://equake-re.info/srcmod/).

### References

Barnhart, W. D., Hayes, G. P., & Gold, R. D. (2019). The July 2019 Ridgecrest, California, earthquake sequence: Kinematics of slip and stressing in cross-fault ruptures. *Geophysical Research Letters*, 46, 11859–11867. [https://doi.org/10.1029/2019GL084474](https://doi.org/10.1029/2019GL084474)

Blackwell, D., Richards, M., Frone, Z., Batir, J., Ruzo, A., Dingwall, R., & Williams, M. (2011). *Renew and Stop* runs over all the events in the catalog within the time window $[T_l, T_j]$, the study region $R$, and the depth range $[0, Z]$.

Freed, A. M. (2005). Earthquake triggering by static, dynamic, and postseismic stress transfer. *Annual Review of Earth and Planetary Sciences*, 33, 335–367. [https://doi.org/10.1146/annurev.earth.33.092203.122505](https://doi.org/10.1146/annurev.earth.33.092203.122505)

## Acknowledgments

This work is supported by National Key R&D Program of China (Grant No. 2020YFA0713401), and by National Natural Science Foundation of China (Grant No. 41725017). H. Zhang is partially supported by the Fundamental Research Funds for the Central Universities. J. Zhuang is partially supported by Grants-in-Aid no. 19H04073 for Scientific Research from the Japan Society for the Promotion of Science, and China Seismic Experimental Site (Grant No. 2019CSES0106). The authors would like to thank the Associate Editor and two anonymous reviewers for their constructive comments which have greatly improved the quality of this study.

---

\[ \log L(\theta) = \sum_j \log \lambda(t_j, x_j, y_j, z_j) - \int_0^T \int_R \int_{T_j} \lambda(t_j, x_j, y_j, z_j) dx dy dz dt, \]

(A8)

where $\theta = (A, \alpha, c, p, D, q, \gamma, \eta$) for the 3D-finite source ETAS model, and $j$ runs over all the events in the catalog within the time window $[T_l, T_j]$, the study region $R$, and the depth range $[0, Z]$.  

---

**Data Availability Statement**

The catalog data used in this study is provided by the Southern California Earthquake Data Center, which can be accessed by the following link: [https://service.scedc.caltech.edu/ftp/catalogs/](https://service.scedc.caltech.edu/ftp/catalogs/). The slip distributions of mainshocks are downloaded from [http://equake-re.info/srcmod/](http://equake-re.info/srcmod/).
