Grimm, C., Käser, M., Hainzl, S., Pagani, M., Küchenhoff, H. (2022): Improving Earthquake Doublet Frequency Predictions by Modified Spatial Trigger Kernels in the Epidemic-Type Aftershock Sequence (ETAS) Model. - Bulletin of the Seismological Society of America, 112, 1, 474-493.

https://doi.org/10.1785/0120210097

Improving earthquake doublet frequency predictions by modified spatial trigger kernels in the Epidemic Type Aftershock Sequence (ETAS) model

Christian Grimm[a], **Martin Käser**[b,e], **Sebastian Hainzl**[c], **Marco Pagani**[d], **Helmut Küchenhoff**[a]

[a]Ludwig-Maximilians-University Munich, Department of Statistics, Ludwigstraße 33, 80539 Munich, Germany

[b]Ludwig-Maximilians-University Munich, Department of Earth and Environmental Sciences, Geophysics, Theresienstraße 41, 80333 Munich, Germany

[c]GFZ German Research Centre for Geoscience, Physics of Earthquakes and Volcanoes, Helmholtzstraße 6/7, 14467 Potsdam, Germany

[d]Global Earthquake Model Foundation, Via Ferrata 1, 27100 Pavia, Italy

[e]also at Munich Re, Section GeoRisks, Königinstr. 107, 80802 Munich, Germany

Declaration of Competing Interests

The authors acknowledge there are no conflicts of interest recorded.

Abstract. Earthquake sequences add a substantial hazard beyond the solely declustered perspective of common probabilistic seismic hazard analysis (PSHA). A particularly strong driver for both social and economic losses are so-called earthquake doublets (more generally multiplets), i.e. sequences of two (or more) comparatively large events in spatial and temporal proximity. Without differentiating between foreshocks and aftershocks, we hypothesize three main influencing factors of doublet occurrence: (1) the number of direct and secondary aftershocks triggered by an earthquake; (2) the occurrence of independent clusters and seismic background events in the same time-space window; and (3) the magnitude size distribution of triggered events (in contrast to independent events). We tested synthetic catalogs simulated by a standard epidemic type aftershock sequence (ETAS) model for both Japan and Southern California. Our findings show that the common ETAS approach significantly underestimates doublet frequencies compared to observations in historical catalogs. In combination with that, the simulated catalogs show a smoother spatiotemporal clustering compared to the observed counterparts. Focusing on the impact on direct aftershock productivity and total cluster sizes, we propose two modifications of the ETAS spatial kernel in order to improve doublet rate predictions: (a) a restriction of the spatial function to a maximum distance of 2.5 estimated rupture lengths; and (b) an anisotropic function with contour lines constructed by a box with two semicircular ends around the estimated rupture segment. These modifications shift the triggering potential from weaker to stronger events and consequently improve doublet rate predictions for larger events, despite still underestimating historic doublet occurrence rates. Besides, the results for the restricted spatial functions fulfill better the empirical Bath's law for the maximum aftershock magnitude. The tested clustering properties of strong events are not sufficiently incorporated in typically used global catalog scale measures, such as log-likelihood values, which would favor the conventional, unrestricted models.

Keywords: earthquake doublets, ETAS, productivity, anisotropy.

Main author contact information: Christian.Grimm@stat.uni-muenchen.de

Introduction

Sequences of strong earthquakes within a relatively narrow time-space window can cause dramatic social and economic damage to our society. The financial losses produced by such multiplets are of particular interest to the risk assessment of governments and in the insurance industry. Recent examples of short-term clusters containing several strong, damaging earthquakes are the Kumamoto (Japan, 2016) sequence with a magnitude $M_{JMA} = 7.3$ mainshock preceded by $M_{JMA} = 6.4$ and $M_{JMA} = 6.5$ foreshocks within 28 hours (Zhuang et al., 2017), and the Ridgecrest (California, 2019) sequence with a mainshock $M_w = 7.1$ preceded by a $M_w = 6.4$ event about 34 hours earlier (Hauksson et al., 2020).

Most typically, sequences of strong and destructive foreshocks, mainshocks, and aftershocks occur within several hours or few days and can therefore be assumed to be controlled by a physical triggering mechanism. However, it is well-known that aftershock sequences can increase seismicity locally for years or even decades. In case that two strong events occur in spatial proximity but months apart, the second event may be an offspring of the on-going sequence of the first, or may have happened coincidentally due to independent background seismicity or as a part of an unrelated sequence.

However, from a risk management perspective, the question of physical causality and the particular interevent time seems rather irrelevant. In both cases, the repeated destruction may affect the same governmental budgets and (re-)insurance contracts within a relatively short time and thus presents a comparably severe risk. Reliable predictions of the likelihood of any strong event cluster, both triggered and coincidental, are therefore an important task for risk managers in governments and the insurance industry.

A suitable term for strong event clusters is given by so-called earthquake *doublets*, sometimes more generally referred to as *multiplets*. While exact specifications are highly inconsistent in the literature, they are generally defined as pairs (doublets) or sets (multiplets) of similarly strong earthquakes in spatiotemporal proximity (Felzer et al., 2004; Gibowicz and Lasocki, 2005; Kagan and Jackson, 1999; Lay and Kanamori, 1980).

Kagan and Jackson (1999) defined doublets as pairs of earthquakes with magnitude $M_w \geq 7.5$, that are no more than one rupture size apart and whose interevent time is less than their recurrence time derived from plate motion. They found that approximately 22% of worldwide events with $M_w \geq 7.5$ occur in doublets, with a maximum interevent time of doublet pairs of almost 17 years.

In contrast, Felzer et al. (2004) specified multiplets as a potential mainshock together with all aftershocks within 0.4 magnitude units, occurring during the following two days and within a spatial box centered in the mainshock's epicenter. The distance of the mainshock's epicenter to the sides of the box is set to 2.5 times the estimated fault length, which is justified by the hypothesis that aftershocks are generally expected to occur within two fault lengths, with an extra half a length accounting for location uncertainty. They demonstrated statistical evidence that foreshocks, aftershocks, and multiplets occur due to the same physical triggering mechanism and that the number of times that multiplets occur increases linearly with the number of aftershocks observed. Felzer et al. (2004) infer that certain regions in the world, such as Solomon Islands, show an increased multiplet rate due to higher aftershock rates and earthquake density, rather than unique seismic fault structures that support the occurrence of multiplets.

Gibowicz and Lasocki (2005) defined doublets as a pair of trigger-related earthquakes with no more than 0.25 magnitude units difference, applying magnitude-dependent stepwise spatial and temporal constraints of 40-90 kilometers and 200-450 days.

Although the concept of earthquake triggering is well known and the potential of additional damage due to on-going seismic sequences has been shown in recent studies (Abdelnaby, 2012; Kagermanov and Gee, 2019; Papadopoulos et al., 2020), seismic hazard is typically computed considering only independent (i.e. mainshock) earthquakes, e.g. in probabilistic seismic hazard analysis (PSHA) approaches (Cornell, 1968; McGuire, 2008). PSHA traditionally not only neglects contributions to hazard from supposedly triggered sequences and therefore underestimates chances of doublet and multiplet occurrences, but it is also based on the highly subjective and influential selection of a declustering method (Marzocchi et al., 2014; van Stiphout et al., 2011; Zhang et al., 2018).

A prominent and extensively studied method to analyse earthquake sequences is the epidemic type aftershock sequence (ETAS) model (Ogata, 1988, 1998). ETAS accounts for earthquake clustering in terms of a branching process and models the number of aftershocks as well as their spatial and temporal distribution depending on the magnitude of the trigger. The spatiotemporal event rate is formed by the sum of a triggered rate and a time-independent seismic background rate contribution (Chu et al., 2011; Jalilian, 2019; Kagan et al., 2010; Zhuang et al., 2002). ETAS model estimations can be used for both short-term aftershock forecasts and the simulation of long-term synthetic catalogs.

The goodness of ETAS model fits is typically assessed by the log-likelihood function, Akaike's information criterion (AIC), or the degree of spatial clustering, expressed by Ripley's K-function

(Chu et al., 2011; Veen, 2006). Besides, visual tools such as spatial plots of the estimated conditional intensity and a comparison of the ETAS triggering function with observed aftershock rates in the historic catalog can be used (Chu et al., 2011). All of the above have in common that they assess the model fit on a global catalog scale, i.e., they test whether the synthetic catalogs sufficiently well represent the observed spatiotemporal clustering behavior in the full magnitude range.

The log-likelihood and AIC measures are related to the joint probability of all earthquakes and thus mainly determined by the fit to the more numerous smaller magnitude events. This might be problematic concerning earthquake risk, which is mainly related to large events. For example, Hainzl et al. (2008, 2013) showed that the common ETAS assumption of isotropic aftershock triggering leads to a biased magnitude-scaling of the aftershock productivity where the trigger potential of small magnitudes is overestimated to better adapt to realistic anisotropic aftershock distributions. Therefore, it is also desirable to assess synthetic ETAS catalogs on their capability to predict realistic occurrence rates of large magnitude doublets and multiplets.

In this paper, we present a new concept of assessing the quality of synthetic catalogs generated by ETAS with respect to doublet and multiplet rates. We introduce three novel and more realistic designs of the ETAS spatial kernel that improve predictions of the respective rates: (1) an anisotropic spatial distribution, (2) an isotropic but finite spatial distribution, and (3) a finite anisotropic spatial distribution. We then test our new model approaches for 24 and 39 years lasting earthquake catalogs recorded in Japan and Southern California, respectively.

In the following section, we derive a doublet and multiplet definition that is used in this paper and comprehensively discuss the main influencing factors for doublet and multiplet occurrences. Next, we briefly describe the utilized earthquake catalogs. We then describe the common ETAS model, define the tested variants of the spatial kernel and introduce the quality measures applied in our analysis of model fits and simulation results. Finally, we present and discuss the results of all four studied ETAS model versions, and we interpret the findings related to the initial motivation in the Conclusion section.

Earthquake doublets

Definition

For the sake of simplicity, in this work, we waive the term multiplet and define an earthquake doublet more generally as a pair or set of events with a magnitude difference of less than 0.4, occurring within one year (starting from the occurrence time of the earlier event) and within a circular radius of 2.5 times the estimated rupture length of the earlier event.

The temporal constraint of one year is derived from the typical length of a (risk) budget period or reinsurance contract. We limit our investigation to strong events with magnitude $M_w \geq 5.9$, therefore allowing for doublet and multiplet partner events down to $M_w \geq 5.5$.

Doublets may either occur within a supposed triggered sequence (mainshock and aftershock) or among independent clusters. In order to avoid doublets built by two aftershocks, being both related to a stronger mainshock prior to them, we only count doublets where the earlier event is not contained in the time-space domain of a previous, stronger event. This is consistent with our motivation drawn from a risk management perspective, since the damage caused by an aftershock-aftershock doublet is likely to be overshadowed by the mainshock.

Main influencing factors

Assuming equal physical triggering mechanisms of foreshocks and aftershocks (Felzer et al., 2004), we propose the following three main factors for doublet occurrences:

(1) the *aftershock productivity*, i.e. the number of direct and secondary offsprings triggered by an earthquake,

(2) the *number of independent events* in the same time-space window, i.e., the occurrence of clustered and background events that are unrelated to the triggering of the event under consideration, and

(3) the *magnitude size distribution* of triggered and independent background events.

It is evident that a higher amount of earthquakes within the time-space window of an investigated event increases the probability of a doublet occurrence. Therefore, an increased aftershock productivity and background activity (the first two factors above), increase the likelihood that a doublet partner is found. Clearly, the aftershock productivity has a much stronger effect than the time-homogeneous seismic background rate since it directly increases the local and short-term cluster size. It is important to mention that triggered and background seismicity are interacting like competing contributors to event rates in ETAS, so an increase of aftershock productivity is generally going along with a decrease of background seismicity and vice versa. The overlapping of the considered event sequence by a second, unrelated cluster evolving in the same time-space window increases the doublet probability substantially if the second cluster is approximately equal or larger.

Regarding the magnitude size distribution (the third factor above), it is still controversially debated in the literature whether the magnitude size distribution of a triggered event depends on the magnitude of its trigger. While Felzer et al. (2004) assume that the magnitude size distribution of a triggered event follows a constant Gutenberg-Richter relationship and therefore is independent of the trigger's magnitude, Nandan et al. (2019) find triggered magnitudes clustering around the triggering magnitude by a kinked magnitude size distribution, which would mean that the triggering event tends to reproduce similar magnitudes with increased probability. In contrast to that, based on a stacking-approach analysis of Gutenberg-Richter a and b value time series, Gulia et al. (2018) argue that the b-value on average shows a temporal 20-30% increase compared to the pre-mainshock time, with more significant increases for stronger events nearby the mainshock epicenter location. We point out that the kinked magnitude size distribution by Nandan et al. (2019) would increase chances of doublet and multiplet occurrence, whereas the temporal b-value increase suggested by Gulia et al. (2018) significantly lowers their likelihood.

In this paper, however, we assume a unique magnitude size distribution for all events according to the Gutenberg-Richter relationship (Gutenberg and Richter, 1944) as done so in the vast majority of ETAS studies. Instead, we are focusing our study on the impact of the aftershock productivity in ETAS on doublet and multiplet occurrence rates.

Selection of earthquake catalogs

We perform our study in two regions with distinct tectonic environments and faulting types, Japan and Southern California. The seismicity in Japan is complex, hosting reverse faulting subduction zone events (particularly along the coast) with relatively flat dips and broader, more isotropically shaped spatial distributions of aftershock epicenters, as well as in-slab normal faulting earthquakes and crustal events with varying depths and mechanisms. Southern California has mostly steep faults with strike-slip rupturing mechanisms in a continental tectonic regime, promoting narrower, elongate distributions of epicenters.

Regional catalogs

In the following, we describe the regional earthquake catalogs used for the estimations of the ETAS model. For each data set, we define a time-space *target* window, which is constructed by a time span and a geographical polygon. This window comprises the so-called *target events* that are used to fit the model. The additional *complementary* window is built by the preceding six months and a one-degree bounding box around the polygon in the geographic coordinate system. The so-called *complementary events* are not fitted by the ETAS model estimations but may contribute to the estimated trigger rate of events in the target domain.

We downloaded the Japan earthquake catalog from the *National Research Institute for Earth Science and Disaster Resilience (NIED)* (see Data and Resources; Kubo et al. (2002)). The catalog provides both moment tensor magnitudes and Japanese Meteorological Agency (JMA) scale magnitudes. For our study, we chose the moment magnitude data, which is complete from $M_c = 4.0$ according to the fit of the model of Ogata and Katsura (1993). We define the time-space target window from July 1, 1997, until October 31, 2020, and for a longitude-latitude range from 129°E to 144°E and from 28°N to 44°N, respectively. Figure 1(a) shows the selected event locations with the corresponding boundaries of the spatial polygon.

The focal mechanism catalog for Southern California was obtained from the *Southern California Earthquake Data Center (SCEDC)* (see Data and Resources; Hauksson et al. (2012); Yang et al. (2012)). Magnitudes are provided in moment magnitude scale. The completeness magnitude is estimated to be $M_c = 2.8$ using the Ogata and Katsura (1993) model. We defined the target window from July 1, 1981, until December 31, 2019, and by a hexagonal polygon (Hutton et al., 2010) which is depicted in figure 1(b) together with all event locations.

Both catalogs provide nodal plane solutions for each event. Since the accuracy of focal mechanisms cannot be guaranteed, especially for smaller magnitude events, we used the given sets only as additional candidates in our algorithm to determine the strike angle needed for the anisotropic ETAS model version (see section "ETAS model").

Short-term incompleteness

Short-term incompleteness in earthquake catalogs can be defined as the deficiency of events above the general completeness level M_c for a limited time after a relatively large event. The phenomenon appears to mainly result from the overlap of seismic records that are dominated by the coda waves of the preceding strong event and therefore let subsequent, weaker events remain undetected (de Arcangelis et al., 2018).

Short-term incompleteness in the underlying earthquake catalogs has been identified as a major source of bias in the ETAS estimation process (Hainzl, 2016a,b; Kagan, 2004; Page et al., 2016; Seif et al., 2017). For $m \geq 6$ earthquakes in Southern California, Helmstetter et al. (2006) estimated the duration of temporary catalog incompleteness (in days) above a given magnitude threshold M_c as

$$t = 10^{(m-4.5-M_c)/0.75}. \tag{1}$$

For instance, that means that a catalog with cut-off magnitude M_c is incomplete for about one day after an event with magnitude $m = M_c + 4.5$. The duration of incompleteness exceeds one minute for magnitudes $m \geq M_c + 2.2$.

In our study, we assume that relation (1) is approximately valid for the region of Japan as well. Events occurred during periods of temporary incompleteness are not used for the maximum likelihood ETAS fit, but still contribute to the ETAS event rates of future target events, which means, technically speaking, that they are downgraded from target to complementary events. To avoid excessive fragmentation of the target time window, we applied short-term incompleteness only to events with magnitudes $m \geq 6.2$ for Japan and $m \geq 5.0$ for Southern California, which is 2.2 magnitude units above the respective catalog thresholds.

Global ISC-GEM catalog

For the comparison with more long-term regional and global doublet occurrence rates, we utilize the *International Seismological Centre - Global Earthquake Model (ISC-GEM) Global Instrumental Earthquake Catalogue* with events from January 1, 1904 (see Data and Resources; Di Giacomo et al. (2018); Storchak et al. (2015)). Magnitudes are provided in moment magnitude scale. According to the catalog description and Di Giacomo et al. (2018), the ISC-GEM catalog is step-wise complete from $M_c = 7.5$ (before 1918), $M_c = 6.25$ (from 1918 to 1959) and $M_c = 5.5$ (since 1960). Significant continental earthquakes with magnitude 6.5 or larger are included before 1918.

ETAS model

The initial ETAS model implemented in this study is based on the *R* package *ETAS* as presented by Jalilian (2019) (see Data and Resources Section). It estimates the model parameters using a maximum likelihood approach and the stochastic declustering method introduced by Zhuang et al. (2002).

In ETAS, the occurrence rate of an earthquake at a given time t and location (x, y) corresponds to the sum of two overlaying components: (a) the coincidental, time-independent background seismicity rate and (b) the sum of dynamic trigger rate contributions from all events occurred before time t (i.e. the event history H_t). The combined occurrence rate is therefore modeled by a non-homogeneous Poisson process with intensity function

$$\lambda(t, x, y|H_t) = \mu\, h(x, y) + \sum_{i:t_i<t} \kappa_{A,\alpha}(m_i)\; g_{c,p}(t - t_i)\; f_{D,\gamma,q}(x, y, i) \tag{2}$$

where μ is the total rate of $m \geq M_c$ background events in the whole region and $h(x, y)$ denotes the spatial probability density function of the background seismicity.

The term within the sum describes the trigger rate contribution of an event i, occurred at time $t_i < t$ and location (x_i, y_i) with magnitude m_i, to the rate of $m \geq M_c$ events at time t and location (x, y).

The *aftershock productivity* function

$$\kappa_{A,\alpha}(m_i) = A\; \exp(\alpha(m_i - M_c)) \qquad (m_i \geq M_c;\; A, \alpha > 0) \tag{3}$$

describes the average number of *direct* aftershocks (offsprings) triggered by an event i with magnitude m_i. Such an exponential growth of the productivity is in good agreement with observations (see e.g. the summary provided by Hainzl and Marsan (2008)).

The *temporal trigger function*

$$g_{c,p}(t - t_i) = (t - t_i + c)^{-p} \qquad (t \geq t_i;\; c, p > 0) \tag{4}$$

is the well-known empirical Omori-Utsu law for the decay of aftershock rates with increasing time t after the occurrence time t_i of the triggering event i (Utsu et al., 1995). The c-value defines the delay of the onset of the power-law decay and is typically much less than 1 day. It is likely related to short-time incompleteness of earthquake catalogs after mainshocks (Hainzl, 2016a). The p value is in the range 0.8– 1.2 in most cases (Utsu et al., 1995).

Finally, the spatial trigger function $f_{D,\gamma,q}(x, y, i)$ is conventionally designed as an isotropic probability density function (pdf) and models the decay of aftershock rates depending on the distance of (x, y) to the epicenter of the triggering event, (x_i, y_i). The ETAS model with an isotropic spatial kernel is the hereinafter called isotropic reference model M_0.

However, the assumption of an isotropic distribution is considered to be a weak point in many publications throughout the literature (Bach and Hainzl, 2012; Hainzl et al., 2008, 2013; Ogata, 1998, 2011; Ogata and Zhuang, 2006; Seif et al., 2017; Zakharova et al., 2017; Zhang et al., 2018, 2020). To name a few, Zhang et al. (2018) emphasize that isotropy may be specifically unsuitable for subduction zone events above a magnitude of approximately $M_w = 7.5$ since estimated rupture lengths and widths are diverging increasingly. They suggest a uniform spatial density in the rupture area with power-law decay outside. Moreover, since ETAS usually neglects the depth dimension, increasing dip angles can already lead to a clearly elongate and thus anisotropic projection shapes of the rupture plane for even smaller events. Another prominent design is the elliptic Gaussian distribution introduced by Ogata (1998) and further studied by Ogata and Zhuang (2006) and Ogata (2011).

In the above references, anisotropic models are generally found to lead to more accurate ETAS model estimates. In particular, Hainzl et al. (2008, 2013) emphasize that the assumption of isotropy can lead to an underestimation of the aftershock productivity parameter α, resulting in underpredicted cluster sizes of stronger events. Given our particular interest in strong events, this gives the motivation to apply an anisotropic alternative in this study.

Besides, in preliminary analyses of a standard ETAS model, we observed that small events are typically assigned a much wider reach of spatial triggering relative to their estimated rupture size than large events. We hypothesize that this might similarly promote disproportionate triggering of smaller events, since it might be easier for the ETAS algorithm to model unique spatial cluster patterns by the overlapping spatial kernels of a large number of smaller events than by the rather inflexible spatial shapes of fewer, but stronger events.

Therefore, in this paper, we propose two modifications of the conventional, isotropic reference model M_0: Firstly, we apply an anisotropic spatial kernel constructed around the surface-projection of the estimated rupture segment, which is assumed to be parallel to the strike and passing through the epicenter. Secondly, we introduce a magnitude-dependent spatial restriction threshold to the spatial kernel that prevents events from triggering outside of the specified surrounding area.

In the following, we introduce the finite and infinite isotropic and anisotropic kernels. Next, we present the algorithm to estimate the rupture length as well as the strike angle and epicenter position along the rupture line in the anisotropic model case. Then we define the set of four models that were tested in this study. Ultimately, we account for a re-scaling of the aftershock productivity.

Isotropic versus anisotropic spatial kernel

Consider a triggering event i with magnitude m_i and epicenter location (x_i, y_i). Furthermore, in the *isotropic case*, let $r_i(x, y)$ be the *point-to-point* distance of a point (x, y) to the epicenter location of event i. We define the standard isotropic spatial kernel following Jalilian (2019) by

$$f_{D,\gamma,q}(x, y, i) := \frac{q-1}{D\exp(\gamma(m_i - M_c))} \left(1 + \frac{\pi r_i(x, y)^2}{D\exp(\gamma(m_i - M_c))}\right)^{-q} \tag{5}$$

with spatial parameters $q > 1$ and $D, \gamma > 0$. Note that the characteristic length of the power-law decay, $\sqrt{D\exp(\gamma(m_i - M_c))/\pi}$, scales with the trigger magnitude which accounts for the observed exponential increase of the rupture dimensions with earthquake magnitude (Wells and Coppersmith, 1994).

For the *anisotropic case*, let l_i be the estimated rupture length of event i and $r_i(x, y)$ denote the nearest *point-to-segment* distance of a point (x, y) to the estimated rupture segment of event i. Then we construct the anisotropic spatial kernel by

$$f_{D,\gamma,q}(x, y, i) := \frac{q-1}{D\exp(\gamma(m_i - M_c))} \left(1 + \frac{2\,l(m_i)\,r_i(x, y) + \pi r_i(x, y)^2}{D\exp(\gamma(m_i - M_c))}\right)^{-q}. \tag{6}$$

with the same parameter constraints $q > 1$ and $D, \gamma > 0$.

The anisotropic kernel (6) is a generalisation of the isotropic kernel (5) for rupture lengths $l(m_i) > 0$. In contrast to the isotropic function, the contour lines of the anisotropic kernel are not centered around the epicenter but constructed as a box with two semicircular ends around the estimated rupture line of the triggering event. Both kernels are probability density functions (pdf) over infinite space.

Spatial restriction

We can restrict the spatial extent of both the isotropic and anisotropic spatial kernel by setting $f_{D,\gamma,q}$ equal to 0 if the respective distance term exceeds a certain magnitude-dependent threshold $\tilde{r}(m_i)$, i.e.

$$\tilde{f}_{D,\gamma,q}(x, y, i) = \begin{cases} \frac{f_{D,\gamma,q}(x,y,i)}{F_{D,\gamma,q}(m_i)} & \text{if} \quad r_i(x, y) \le \tilde{r}(m_i) \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

where $\tilde{f}_{D,\gamma,q}(x,y,i)$ is normalized by the integral of $f_{D,\gamma,q}(x,y,i)$ over the area up to the cutoff distance $\tilde{r}(m_i)$ in order to retain a pdf,

$$F_{D,\gamma,q}(m_i) = \begin{cases} 1 - \left(1 + \frac{\pi \tilde{r}(m_i)^2}{D \exp(\gamma(m_i - M_c))}\right)^{1-q} & \text{(isotropic model)} \\ 1 - \left(1 + \frac{2\, l(m_i)\ \tilde{r}(m_i) + \pi \tilde{r}(m_i)^2}{D \exp(\gamma(m_i - M_c))}\right)^{1-q} & \text{(anisotropic model).} \end{cases}$$

In this study, we use a threshold which is proportional to the magnitude-dependent rupture length $l(m_i)$ of event i, i.e. $\tilde{r}(m_i) = k \cdot l(m_i)$, in order to correlate the spatial trigger extent to the estimated rupture dimension.

Figure 2 visualizes the shapes of isotropic and anisotropic spatial kernels, restricted to a distance of $\tilde{r}(m_i) = 2.5 \cdot l(m_i)$, for the exemplary magnitudes $m = 5.0$ and $m = 7.5$, using initial spatial parameter guesses $D = 2.0$, $\gamma = 2.1$ and $q = 1.5$.

Estimation of rupture length, strike, and position of rupture line

The anisotropic spatial kernel defined in (6) requires an estimation of the ruptured segment, in particular its central location, the length, and the strike angle in order to locate the rupture line segment of an earthquake.

In order to obtain magnitude-dependent estimates of the subsurface rupture lengths l of all events, we use the scaling relations

$$\log_{10}(l(m)) = \begin{cases} -2.37 + 0.57m & \text{reverse faulting} \\ -2.57 + 0.62m & \text{strike-slip faulting} \end{cases} \tag{8}$$

where, for the sake of simplicity, we selected the reverse faulting scaling relations for subduction environments, provided by Blaser et al. (2010), for all events in the Japan catalog and the strike-slip faulting equations for continental regimes, given by Wells and Coppersmith (1994), for all Southern Californian events.

The strike angles are selected such that the corresponding rupture line fits well to the cloud of potential aftershocks. Therefore, we test the given focal mechanism data in the earthquake catalogs and compute the summed trigger rates for the subsequent events like in equation (2), assuming initial parameter guesses:

$$(A_0, \alpha_0, c_0, p_0, D_0, \gamma_0, q_0) = (0.02, 1.6, 0.02, 1.0, 2.0, 2.1, 1.5)\,.$$

Additionally, we go through all strikes from 0° to 175° in 5° steps and compute the initial trigger rates accordingly. From all candidates, we choose the one that leads to the maximum sum of occurrence rate contributions to future events and therefore is in best agreement with presumed offsprings. Note that, since we do not consider the rupture's dip, strikes above 180° coincide with the tested set of angles. Once we have optimized the strikes via the above approach, we additionally test five different positions of the rupture line relative to the corresponding epicenter location of the trigger event. Thus, we allow the epicenter to lie either right at the start, center, or end of the rupture line; or a quarter or three quarters along the rupture line.

The above-mentioned selection algorithm clearly represents a manipulation of the initial model conditions. In fact, the so-selected strike angles show only moderate agreement with the originally provided strikes. Compared to using only nodal plane solutions given in the catalogs, we observed negligible effects for Japan and moderately increasing estimates of the aftershock productivity in Southern California, where potential aftershocks were more likely to scatter along a clearly identifiable line. In any case, the impact of optimized strike selection was much smaller than the effect of the introduced spatial restrictions or the anisotropic shape of the spatial kernel itself.

Mismodeling of the spatial aftershock distribution leads to biased model estimates (Hainzl et al., 2008). To minimize this problem, we refrained from directly using the strike values provided in the catalogs due to the large uncertainties in the source inversions. Instead, the optimized selection of strike angles assures that the event's rupture line passes through the cloud of its potential aftershocks, which we visually confirmed for individual sequences.

Choice of four model designs

In this paper, we analyze four different variants of the ETAS model regarding their ability to predict realistic doublet and multiplet rates. Table 1 lists the model design specifications made for each approach.

The reference model M_0 represents the standard isotropic design in equation (5) with event-specific spatial restriction

$$\tilde{r}_0(m_i) = 100 \cdot l(m_i).$$

The restriction $\tilde{r}_0(m_i)$ is only of technical nature and has negligible impact on results while considerably improving code performance by avoiding the computation of extremely distant inter-event triggering relations over the entire catalog size. Hereinafter, we will therefore refer to models with spatial extent $\tilde{r}_0(m_i)$ as *unrestricted*.

Using the same isotropic kernel (5), in model M_1 we test the spatial restriction

$$\tilde{r}_1(m_i) = min\{2.5 \cdot l(m_i), 1\}$$

where the lower limit of one kilometer guarantees a minimum spatial extent to the smallest events in the catalog. The aim of the restricted extent of the spatial kernel is to avoid wrong associations of distant events as aftershocks. It gives more triggering power to the stronger events (that may trigger in a larger area) and takes away triggering potential from the weaker events. The metric of 2.5 rupture lengths goes back to the assumption in Felzer et al. (2004) that aftershocks are expected to mainly occur within this distance, including a buffer of half a rupture length for location uncertainties.

Model M_2 builds upon the anisotropic spatial kernel (6) with optimized strikes and relative rupture locations and is unrestricted ($\tilde{r}_0(m_i)$).

Finally, model M_3 tests the anisotropic spatial kernel with restriction $\tilde{r}_1(m_i)$.

Note that, for the sake of consistency, we applied the anisotropic spatial kernels to all events disregarding their magnitude in models M_1 and M_3. For small rupture lengths, however, the shape is similar to an isotropic kernel.

Subsequent re-scaling of ETAS functions

Note that the temporal trigger function (4) is not a pdf since its integral over infinite time typically amounts to a number larger than 1 (for $p > 1$) or infinity (for $p < 1$). Therefore, the excessive density in (4) down-scales the estimates of parameter A in the productivity function (3).

In favor of better interpretability of the model results, it is useful to cut off the temporal trigger function (4) at the length of the entire catalog T (in days) and normalize it by the integral over the time range from 0 to T, i.e.

$$G_{c,p}(T) = \frac{1}{1-p}\left((T+c)^{1-p} - c^{1-p}\right).$$

Accordingly, we re-scale the absolute aftershock productivity parameter A by

$$\tilde{A} = A\, G_{c,p}(T). \tag{9}$$

Quality measures

In this section, we introduce the quality measures used to assess and compare the goodness of the selected models. We start with a short description of the log-likelihood value and branching ratio, designed to assess the goodness of fit and the detected trigger portion on a global catalog scale. These properties are widely used in ETAS analysis but have the disadvantage that they do not provide any detailed information on how well the model represents the critical triggering behavior of particularly strong earthquakes, which is of interest in this study.

Therefore, we add tools to more specifically evaluate the models' capability of representing strong event clusters. First, the expected, magnitude-dependent cluster size is derived. Next, we outline the ETAS forward simulation procedure for both single sequences and synthetic catalogs based on the model estimates. Then, we suggest visual and semi-quantitative measures (e.g. Bath's law, degree of temporal and spatial clustering) that help understand clustering properties in the simulated catalogs. Finally, we describe the evaluation of doublet probabilities from simulated catalogs and sequences.

Log-likelihood function and integrated event rate

The set of ETAS parameters, $\theta = (\mu, A, \alpha, c, p, D, \gamma, q)$, is optimized by maximizing the log-likelihood function (LLF)

$$l(\theta|H_T) = \sum_{j=1}^{N} ln(\lambda_\theta(t_j, x_j, y_j|H_{t_j})) - \Lambda_\theta(\mathbb{T}, \mathbb{S}|H_\mathbb{T}) \tag{10}$$

where the first term sums up the logarithmic event rates (2) at the exact times t_j and locations (x_j, y_j) of the N target events that occurred in the time-space window specified for each catalog. The second term

$$\Lambda_\theta(\mathbb{T}, \mathbb{S}|H_\mathbb{T}) = \int_\mathbb{T} \iint_\mathbb{S} \lambda_\theta(t, x, y|H_t) dxdydt \tag{11}$$

represents the total event rate integrated over the (step-wise) target time window $\mathbb{T}$ and the target space window $\mathbb{S}$ based on the estimated background seismicity rate and the triggering-induced rate resulting from contributions of both target and complementary events in the original catalog. In other words, $\Lambda_\theta(\mathbb{T}, \mathbb{S}|H_\mathbb{T})$ represents the expected total number of events to occur within the modeled target time-space domain (Jalilian, 2019; Ogata, 1988, 1998).

In general, a larger LLF value $l(\theta|H_T)$ implies a better fit to the event occurrence in the original catalog. Note that the LLF value is comparable only for identical data inputs, i.e. model runs for Japan and Southern California cannot be cross-compared. Since all model approaches are based on the same number of free parameters, the information from the AIC criterion is redundant and therefore not shown.

Branching ratio

We modeled the magnitude size distribution by the pdf derived from the Gutenberg-Richter relationship, i.e.

$$\rho(m) = \begin{cases} \beta exp(-\beta(m - M_c)) & \text{if } m \geq M_c \\ 0 & \text{otherwise,} \end{cases} \quad (12)$$

thus assuming $M_{max} = \infty$ as the maximum magnitude for each region. The maximum-likelihood estimator for parameter β is

$$\hat{\beta} = \frac{N}{\sum_{i=1}^{N}(m_i - M_c)}$$

where N is the number of fitted events and m_i denotes the respective event magnitudes (Jalilian, 2019). Applied to the magnitudes of all target events in our regional catalogs, we obtained $\hat{\beta}_{JPN} = 2.36$ for Japan and $\hat{\beta}_{CAL} = 2.73$ for Southern California.

The *branching ratio* measures the mean direct aftershock productivity of an arbitrary event, averaged over the entire magnitude range. It is computed by the integral of the estimated aftershock productivity with parameters $alpha$ and re-scaled $\tilde{A}$ weighted by the pdf of the magnitude size distribution $\rho(m)$, i.e. (Jalilian, 2019; Seif et al., 2017)

$$\nu_{branch} = \int_{m_c}^{\infty} \tilde{A}\, e^{\alpha(m-M_c)} \rho(m) dm = \frac{\tilde{A}\,\beta}{\beta - \alpha} \quad (13)$$

for $\alpha < \beta$.

Cluster size

Based on the estimates of the (direct) aftershock productivity function (3) and the branching ratio (13), we obtain the expected cluster size

$$\widehat{N_c}(m) = \frac{\tilde{A}\; e^{\alpha(m-M_c)}}{1-\nu_{branch}} \tag{14}$$

including secondary triggering by use of the geometric series (Helmstetter and Sornette, 2003).

ETAS forward simulation process

For every model and region, we used the fitted ETAS parameters to forward-simulate both single synthetic sequences and entire catalogs. We generated single trigger sequences to study the results without the impact of background seismicity and independent clusters. Each of these sequences is initiated by a mainshock of varying magnitude, starting from $M_w = 5.5$ and incrementally increasing in tenths of a magnitude unit. For each region and model, a set of 5,000 sequences was simulated for each mainshock magnitude.

Additionally, we simulated 10,000 realizations of an entire synthetic catalog, including background seismicity and simultaneously evolving trigger sequences. As a time-space window for the simulations, we chose the identical constraints for which the ETAS models were fitted (see section *Selection of earthquake catalogs*), including the semi-year complementary time window as an initialization period of pre-existing seismicity. The background seismicity rate is distributed over the spatial window by a superposition of bivariate, isotropic Gaussian kernels, centered in the original event occurrences (Jalilian, 2019).

In both types of simulations, the number of offsprings is drawn from a Poisson distribution with an expected value equal to the magnitude-dependent aftershock productivity estimate. We used the inversion method to sample the spatial and temporal distance of an offspring to the trigger and then sampled uniformly from the respective contour line of the spatial distribution. The magnitudes of both triggered and independent events were sampled from the Gutenberg-Richter distribution (eq. 12) with β as estimated for the respective region.

Since the original Japan catalog contains the extreme Tohoku earthquake (March 11, 2011; according to the catalog $M_w = 8.7$) that is very unlikely to be sampled from the Gutenberg-Richter distribution, we manually added the Tohoku event to all synthetic catalogs for Japan.

Bath's law

An important property of an earthquake cluster is the magnitude difference between the mainshock and the strongest aftershock, as it can serve as an indicator of how much hazard is added by the on-going triggering of a sequence.

Historical observations show that this magnitude difference is, on average, approximately 1.2 magnitude units independently of the absolute magnitude of the trigger event, which is referred to as *Bath's Law* (Helmstetter and Sornette, 2003; Shearer, 2012; Vere-Jones, 1969).

For observed and synthetic catalogs, we approximate the magnitude difference by applying the time-space constraints of our doublet definition to any event under consideration and computing the magnitude difference between the considered event and the strongest of all events that occurred in the specified time-space domain. Clearly, this selection can include independent background events or events occurred in unrelated clusters. In order to constrain the Bath law statistics to mainshocks, we skip earthquakes that are supposed aftershocks (i.e. that are contained in the time-space range of a previous, stronger event) or foreshocks (i.e. that contain a stronger event in their own time-space domain).

For synthetic sequences, we apply the same filtering algorithm to each simulated sequence with its known initiating magnitude.

Coefficient of variation

We measure the degree of temporal clustering of event occurrences by decomposing the time domain into a monthly grid and computing the variation of the numbers of events falling into the time intervals. In order to account for varying overall catalog sizes, we use the coefficient of variation (CV), which is a measure of the relative dispersion of a random distribution sample X standardized by its mean. It is computed as $CV = \frac{\sqrt{Var(X)}}{Mean(X)}$, where $Var(X)$ denotes the variance of the sample X.

Ripley's K

The degree of spatial clustering of the event locations can be expressed by Ripley's K function (Ripley, 1976; Veen, 2006). The K-function computes the average number of additional event locations within a distance h of any given event, normalized by the overall number of events per space unit $\frac{N}{A}$, i.e.

$$K(h) = \frac{A}{N^2} \sum_i \sum_{j \neq i} \mathbb{1}\left(r(i,j) \leq h\right) \tag{15}$$

where $\mathbb{1}$ is the indicator function.

If the investigated catalog was produced by a homogeneous Poisson process with no spatial clustering inherent, $K(h)$ would be asymptotically normal with $K(h) \sim N\left(\pi h^2, \frac{2\pi h^2 A}{N^2}\right)$ (Chu et al., 2011). The more $K(h)$ exceeds πh^2, the more clustered the event locations are. Values of $K(h) < \pi h^2$ signify inhibition.

Doublets probability

The most important measure for our study's purpose is the probability that an event is part of an earthquake doublet according to our definition. Similarly to the Bath law evaluation, we searched all events within the specified time-space window spanned by the earthquake under consideration in the synthetic catalogs. We counted the earthquake as a doublet event if any of the potential partners fulfill the magnitude criterion.

Similarly, for synthetic sequences, we applied the above algorithm to the known sequence initiating events.

Results and Discussion

In the following, we discuss the results obtained from the four tested models. We start by comparing the ETAS estimation results on a global catalog and model scale by looking at the log-likelihood values, the branching ratios, the general shapes of the fitted spatial kernels, and the average cluster sizes depending on the trigger magnitude. Then we move on to the analysis of the synthetic results from simulated sequences and catalogs. Herein, we first analyze the consistency of simulation results with Bath's law and observed magnitude differences in the original catalogs, respectively. We continue with an analysis of the degree of temporal and spatial clustering in simulated catalogs compared to the original event sets. Finally, we evaluate doublet frequencies in simulated catalogs and compare them to historical observations.

Model fit

Table 2 lists the results from models M_0, M_1, M_2 and M_3 for both regions, Japan and Southern California, including the log-likelihood function values and the branching ratios.

Regarding the log-likelihood values $l(\theta|H_T)$, in both regions, we observe the order $M_1 > M_3 > M_0 > M_2$. We can conclude that, according to the log-likelihood measure, the anisotropic shape of the spatial kernel leads to an improved performance, while the spatial restriction detracts the quality of the model fits.

One reason for the better performance of the anisotropic models can be found in the optimization process used to define the strike. In fact, the advantage of the anisotropic over the isotropic models was moderately reduced when we ran the models with the originally provided strike angles rather than the optimized ones. However, also in the case of original strikes, the anistropic models were superior with regard to the log-likelihood value.

On the other hand, more generally, the anisotropic shape of the spatial kernel leads to an improved adaptation to the aftershock clouds for most events. For two exemplary magnitudes $m = 5.0$ and $m = 7.5$, figure 3(a) and (b) depicts the cumulative distribution functions of the spatial kernels against the normalized distance to the event location (for isotropic models) or rupture segment (for anisotropic models). We can see that in both regions, the anisotropic models show a significantly narrower shape, which suggests that the estimated rupture segments fit the potential aftershock clouds better than the isotropic point sources and therefore tend to bring possible offsprings closer to the trigger source. While in the Japan models, this narrowing effect is characterized by the dramatic decrease of parameter D, in the Southern California results it is modeled by the increase of parameter q.

Note that the characteristic length of the power-law decay, $\sqrt{Dexp(\gamma(m_i - M_c))/\pi}$, has unit km^2, so it really has the dimension of an area. Therefore, its exponential increase does not fully compensate for the faster exponential growth of the one-dimensional rupture length estimates in equation (8). Consequently, especially the anisotropic spatial distributions tend to get narrower relative to the rupture length with increasing trigger magnitude, as observable in figure part (b). We conclude from this that the anisotropic kernel gains relevance in the upper magnitude ranges. Moreover, we observe that Southern Californian models generally fit narrower shapes than Japan's models. This agrees with the predominant faulting style. In California, strike-slip events on approximately vertical faults dominate, while shallow-dipping mechanisms are common in Japan, widening the epicentral aftershock distributions.

The generally inferior log-likelihood values of the restricted models can be explained by the additional constraint imposed to the model by the limitation of the extent of the spatial kernels. Any decline of flexibility inevitably leads to a lesser (or equal) overall model performance.

In this context, we observe that the parameter estimates of $\tilde{A}$, which represent the average number of aftershocks triggered by an event with threshold magnitude $m = M_c$, are substantially lower, in Japan even more than halved, when comparing a restricted model to the according unrestricted model. On the other hand, the restricted models lead to highly increased estimates for parameter α signifying an acceleration of the exponential increase of aftershock productivity with growing trigger magnitudes. Figure 4 displays the exponential relation of the expected cluster sizes according to equation (14), i.e. including direct and secondary aftershocks, to the initiating mainshock magnitude on a logarithmic scale. While the restricted models start at a lower base, they cross the lines of unrestricted models at about magnitude $M_{cross} = 5.6$ for Japan models and $M_{cross} = 4.4$ for Southern California models. In other words, events with $m \geq M_{cross}$ are expected to trigger more aftershocks, on average, in restricted models than in unrestricted models.

Since the vast majority of events in the original catalogs have magnitudes $m < M_{cross}$, we may expect that the observed shift of aftershock productivity from small to large events leads to, in total, fewer identified trigger relations in restricted models. Indeed, the restricted models reveal smaller branching ratios and, contrarily, larger background rates. Therefore, we may conclude that the spatial restriction eliminates some trigger relations between more distant events with relatively small magnitudes, that are consequently either associated to the background seismicity or to another, stronger trigger event with larger spatial extent. In particular, the latter case provides an explanation for the greater estimates of parameter α. Furthermore, under the realistic assumption that there were more trigger relations in reality than identified in the models, the absolute loss of identified trigger relations to background seismicity would explain the inferior log-likelihood values.

We further notice in figure 4 and table 2 that, in Southern California models, the anisotropy of the spatial kernels has far more impact on expected cluster sizes than in Japan models. Note that cross-comparisons of cluster sizes between the two regions are only valid if the cluster sizes of Southern California are down-scaled by $exp(-1.2\alpha)$ accounting for the difference of 1.2 of the magnitude thresholds. The clustering is on a generally comparable level, despite we note a more gradual growth due to smaller α estimates for restricted models in Southern California.

Bath law

Figure 5 depicts the mean magnitude differences between an earthquake and the strongest event following in the specified time-space domain in simulated catalogs in comparison to those in the respective original catalog. The corresponding algorithm is outlined in the section *Quality measures*. The top-left panel (a) presents the results for the unrestricted models M_0 and M_1 in Japan. Both models appear to estimate almost identical magnitude differences, with a significant slope for increasing reference magnitudes of the triggering event. On average, the simulated catalogs seem to continuously overestimate the magnitude difference for magnitudes $m > 6.8$ compared to the original data set, with some data points located even outside of the $10\% - 90\%-$ confidence interval for model M_1. The divergence between the results of the simulated catalogs and sequences can be explained by the impact of independent events, that are not contained in the pure sequences. The effect intensifies with increasing magnitudes due to the exponential growth of the spatial window size.

According to the top-right panel (b), Japan's restricted models show a better agreement with the original catalog. There are no data points for magnitudes $m \geq 6.8$ outside of the $10\% - 90\%-$ confidence interval for model M_3. The slope of the curves is smaller, which suggests better accordance with Bath's law hypothesis that the magnitude difference is independent of the trigger magnitude. The smaller divergence between catalog and simulation results emphasizes that the improvement is caused by the increase of the average cluster sizes for the investigated magnitude ranges, as shown in figure 4. This increases the chance of strong aftershocks, and at the same time, it reduces the relative impact of independent events in the considered time-space domain.

The results for Southern California, depicted in panels (c) and (d), show similar trends. Approximately half of the historic events have magnitude differences outside of the $10\% - 90\%-$ confidence interval in both models. In general, Southern California models estimate considerably larger and faster-growing magnitude differences than Japan models, reaching up to 2 magnitude units for the maximum magnitude $M_w = 7.5$. This observation can be explained by the more moderate increase of cluster sizes due to smaller estimates of α. Comparing the two regions, we conclude that the restricted models work better and lead to more pronounced improvements in Japan than in California.

Spatial and temporal clustering

Figure 6 analyses the degree of temporal and spatial clustering in synthetic catalogs compared to the respective original catalog. For these plots only, we generated the synthetic catalogs with magnitudes sampled from the empirical magnitude distribution observed in the respective original catalogs, instead of using the Gutenberg-Richter distribution (12) with estimated parameter β.

The reason is that, by using magnitudes sampled from (12), we observed a deficiency of extremely strong events in the synthetic catalogs compared to the original catalogs, which suggests that (12) tends to underestimate the tail of the empirical magnitude size distribution in the observational data. Consequently, the synthetic catalogs would lack some influential trigger events, that would otherwise cause sporadic peaks in the spatio-temporal distribution of event occurrences.

The top-left panel (a) depicts boxplots of the CV of event occurrence numbers in monthly time intervals of synthetic catalogs for Japan. We observe that, on average, the variance of monthly event occurrences in simulations is by factors smaller than in the original catalog, displayed by the horizontal black line. However, the restricted models tend to produce considerably more temporal variation, with some pronounced outliers, than the unrestricted models. The same observation is made for Southern California in the top-right panel (b). Furthermore, the CVs seem to correlate with the expected cluster sizes of strong events, shown in figure 4. For instance, the anisotropy of the spatial kernel leads to a stronger increase of both productivity parameter α and the temporal clustering in Southern California than in Japan.

Panel (c) demonstrates that the observed smoothing of temporal event occurrences is not a pure side-effect of catalog simulations. Exemplary for model M_3 in Japan, we plotted the curve of monthly event occurrences in the original catalog against the expected number of event occurrences predicted by the ETAS event rate. More precisely, the latter is computed as the total ETAS event rate (see $\Lambda_\theta(\mathbb{T}, \mathbb{S}|H_\mathbb{T})$ in equation (11)), step-wise integrated over the monthly intervals instead of the entire target time window $\mathbb{T}$, which provides us an estimate of the expected number of events occurring in the considered month. This monthly forecast is thus purely based on the fit of the model parameters and the original, non-simulated history of events.

On the one side, the integrated rate clearly underpredicts event occurrences in peak months; on the other hand, it overrates the seismicity in relatively calm months. This contrast is an immediate consequence of the log-likelihood-based model estimation algorithm, which requires that for the optimal set of parameters, the ETAS rate, integrated over the entire time-space target window, equalizes the exact number of target events. Thus, once underestimating the pronounced peaks in the most active months, the rate needs to compensate for this inaccuracy by predicting larger occurrence rates in rather inactive months, which ultimately leads to a clear smoothing of the temporal occurrence curve. We hypothesize that this compensation is caused by an overprediction of both the background seismicity and the triggering potential of small events and an underprediction of the triggering power of strong events.

Finally, the bottom-right panel (d) sheds light on the degree of spatial clustering, measured by Ripley's K function (15). For the sake of clearer visualization, we only present the spatially most strongly clustered models M_3 in both regions. Nevertheless, we observe that, in both regions, spatial clustering is underestimated compared with the respective original catalogs. Generally, event occurrences in Southern Californian seem less intensely clustered in space than in Japan. The kink in the curves, which in the case of Southern California even suggests inhibition, is a boundary effect due to the limited polygon areas.

Earthquake doublets

Finally, figure 7 analyzes the occurrence rates of doublets in the simulated catalogs and sequences. The top-left panel (a) compares the percentage that an event finds a doublet partner depending on its magnitude for the four models and both simulated sequences and catalogs in Japan. Note that, for the sake of clarity, the data is smoothed by aggregating magnitude intervals.

We observe that the restricted models show substantially larger doublet chances than the unrestricted models, which is consistent with our previous findings regarding the larger cluster sizes, the larger degree of temporal and spatial clustering, and the lower average magnitude differences to the strongest event in the time-space domain spanned by an event. Also, doublet percentages decrease with growing magnitudes, which accompanies the earlier observation of increasing Bath law magnitude differences.

It is also worth mentioning that the proportion of events that find a doublet partner is considerably larger within a simulated catalog than in a synthetic sequence. This implies that independent seismic background events or unrelated clusters generate a non negligible fraction of doublets.

The top-right panel (b) shows this aspect in more detail for models M_0 and M_3 in Japan. Conditional on realized doublet pairs, it shows the inverse proportions of doublets consisting of two events from the same cluster and doublets composed by two independent events. The corresponding triggering relationship is known in simulations. For small triggering magnitudes, in-cluster doublets make up a much larger proportion. The share declines with increasing trigger magnitude, however much stronger for model M_0 than for model M_3. In the case of model M_0, independent doublets get even more likely than in-cluster doublets for triggering magnitudes larger than $M_w = 7.6$.

These observations can be explained by the more rapid exponential growth of the area of the spatial window than the aftershock productivity and expected cluster sizes. According to the scaling relations (8), the area of the spatial window covering the surrounding of two and a half rupture lengths is $\pi(2.510^{-2.37+0.57m})^2$. Consequently, the area grows by factor $(10^{0.57})^2 = 10^{1.14} \approx 13.8$ which is faster than the magnitude-dependent growth of aftershock productivity and expected cluster sizes, $exp(\alpha)$, for all $\alpha < 2.62$. Following this line of argument, we can explain the growing impact of independent and unrelated events, with increasing trigger magnitudes. The curves for model M_3 are more robust, since the larger aftershock productivity and expected cluster sizes, resulting from greater estimates of parameter α, better balance out the growth of the spatial window, compared to M_0.

The bottom panels (c) and (d) of figure 7 compare the doublet rate predictions for the Japan models M_0 and M_3 to analogously measured doublet percentages in historic catalogs. As benchmarks, we use the original NIED Japan catalog used for the ETAS model estimation, as well as a regional and a global extract from the ISC-GEM catalog. Respecting the step-wise completeness levels in the ISC-GEM catalog, we counted doublets for events with magnitudes $M_w \geq 5.9$ from the year 1960 and for events with magnitudes $M_w \geq 6.7$ starting in 1918. In particular, this allows for a reliable search of doublet partners with a maximum magnitude difference of $0.4M_w$ units. Due to the relatively small sample sizes in historical data, we grouped the events in the four magnitude intervals $[5.9, 6.0]$, $[6.1, 6.2]$, $[6.3, 6.6]$ and $[6.7, \infty)$. Due to its limited time (24 years), the regional NIED catalog provides only between 18 and 37 events in the respective magnitude intervals and therefore has limited statistical significance, especially in the higher magnitude ranges. Furthermore, we obtained 70 to 105 events in the regional extract for Japan of the ISC-GEM catalog from 1918, and 1362 to 2219 events in the entire ISC-GEM data set. In the simulated catalogs, we isolated all events with magnitude $M_w = 8.7$ from the last interval, since they would dominate the statistic because of the manual sampling of the Tohoku event.

Panel (c) demonstrates that model M_0 tends to underestimate the doublet occurrence probabilities observed in the three benchmark catalogs. The simulations accurately fit two out of four data points of the original NIED catalog, which however is an uncertain statistic due to its small sample size. The more stable curves of the long-term Japan and global benchmark catalogs are mostly located outside of the 10-90% confidence interval.

Model M_3, shown in panel (d), moves considerably closer to the long-term benchmark catalog curves and appears to provide a rather adequate prediction of doublet probabilities compared to the original NIED event set. The 10-90% confidence interval covers all data points of the global catalog, and two out of four samples from the long-term Japan data set.

Both models show a downtrend of doublet rates with increasing magnitudes, which reveals itself particularly in the small fraction of doublets initiated by the sampling of the Tohoku event. In contrast, the probability of doublet occurrences seems magnitude-independent, at least in the lower three magnitude ranges, for the ISC-GEM catalog extracts, which reminds us of the self-similarity of earthquake clustering observed according to Bath's law. The comparison, however, is unavoidably biased due to the subjective specification of our time-space domain in the doublet definition and because of the fact that we do not prohibit doublets produced by independent events not belonging to the same triggered sequence.

The historical observations for Southern California do not provide a sufficient database for benchmarking. We only observed seven events in the overall magnitude range from $M_w \geq 5.9$ in the original catalog, with two of them being a doublet (both in magnitude range $[6.1, 6.2]$). In the regional extract of the ISC-GEM catalog since 1918, we found an overall number of 15 events, with one of them being doublets (in the third magnitude range). In the latter, this would signify a chance of 6.7% that an event finds a doublet partner, which is less than half of the global percentages shown in the bottom-line panels of figure 7. However, the Californian models predict a chance of only 3% (model M_3) or even 1.6% (model M_0) for doublet occurrences.

Sensitivity of results

The results described above, especially the estimated doublet probabilities, are clearly dependent on the rather subjective definition of the temporal and spatial constraints of 365 days and 2.5 rupture lengths as well as the magnitude window of 0.4 M_w units. In accordance with intuition, sensitivity tests have shown that a decrease of one of the three criteria led to lower doublet probabilities in both the simulated and historical data, and vice versa. However, the relative behavior of the four models under consideration, among each other and in comparison with historical catalogs, and therefore the central conclusions, remain the same.

Summary and Conclusion

We compared seismicity generated with four variants of the ETAS model to earthquake catalogs for Japan and Southern California. More precisely, we tested isotropic and anisotropic as well as unrestricted and restricted spatial kernels. The central objective of this study was to find out which of the four models best describes the clustering of particularly strong events and leads to the most realistic predictions of the occurrence probabilities of earthquake doublets. Rather subjectively, we defined a doublet as a pair of an earthquake with any other event occurring during the next 365 days and within a distance of 2.5 rupture lengths to the considered event, with a magnitude difference of no more than 0.4 units. By assuming an identical magnitude size distribution for triggered and independent events, we analyze the impact of aftershock productivity and cluster sizes on cluster properties and doublet occurrences.

The results indicate that the conventional, unrestricted isotropic model poorly represents clusters triggered by particularly large magnitude earthquakes. We found that this model estimates too large magnitude differences between a strong earthquake and the largest event in the specified time-space window, that it tends to highly underestimate the degree of temporal and spatial clustering by smoothing out the occurrence times and locations, and that it tends to underestimate the chances of doublet occurrence. This stands in contrast to global catalog scale measures such as the log-likelihood value, which do not incorporate these weaknesses, and that would attest to the conventional model a comparatively high quality.

The anisotropic spatial kernel improves the overall fit of the model but cannot noticeably alleviate the weaknesses of the unrestricted model variants. Perhaps, it shows its strengths primarily in combination with UCERF3-ETAS type models where crustal fault structures, subduction zones and multi-segment ruptures are incorporated on a detailed level (Field et al., 2017).

By shifting triggering potential from smaller to larger events and therefore increasing cluster sizes of strong trigger events, the restriction of the spatial kernel to 2.5 rupture lengths promotes more realistic estimations of the magnitude difference to the strongest following event and of the doublet probability, compared to historical observations. The temporal and spatial variability of event occurrences rises, additionally indicating more pronounced clustering. However, the improvements in the representation of strong earthquake clusters are at the expense of a decline of the log-likelihood value since trigger relations in the smallest magnitude ranges get lost.

Again, the anisotropic model variant improves the overall fit of the model but has negligible impact on the temporal and spatial clustering and the doublet's occurrence.

We conclude that global catalog scale measures such as the log-likelihood value or the AIC criterion are not an adequate tool for evaluating ETAS model fits if the representation of strong event clusters is of particular interest. It is in the nature of these measures, that they show better performance when more trigger relations are detected. Consequently, a model that is given more freedom, such as the unrestricted variants, will always outperform the more conditioned variants, such as the restricted variants in our study. However, this may lead to trigger relations between events that are, from a standpoint of reason, improbable. In other words, the conventional model does a good job in identifying triggered events, but it does a relatively poor job in assigning the aftershocks to their most realistic triggers, which goes to the benefit of the smaller events.

Certainly, this deficiency can be partly explained by the well-known and extensively studied biases in the use of the ETAS model, such as earthquake location uncertainty, the catalog cut-off magnitude, and short-term incompleteness. In our study, we have accounted for the latter by applying blind periods after strong events according to Helmstetter et al. (2006).

The spatial restriction tested in our models, however, demonstrates that we can improve aftershock to trigger assignments and therefore strengthen the aftershock productivity of strong events by giving the ETAS model more guidance in terms of conditions. Given the assumption of an identical magnitude size distribution for triggered and independent events, aftershock productivity becomes the dominant driver for cluster properties. The larger the size of a cluster, the smaller the magnitude difference to the strongest following event and the larger the chance of a doublet to occur. At the same time, a larger cluster size decreases the relative relevance of independent seismicity in the considered time-space window around an earthquake.

Even the restricted models reveal a persistent underestimation of the cluster properties of large earthquakes. We hypothesize that, in reality, the exponential growth of the aftershock productivity with increasing trigger magnitudes should be even larger. This would also increase the underrepresented clustering of events both in time and space.

Future work should emphasize the importance of a correct representation of strong event clusters by the ETAS model. Using only goodness of fit measures operating on a global catalog scale provides an inherent risk that a poor representation of extreme clusters remains undetected.

This work has analyzed the impact of aftershock productivity and cluster sizes on the occurrence of earthquake doublets. It has, however, neglected the influence of potentially varying magnitude size distributions, that may lead to a correlation of triggering and triggered magnitudes (Gulia et al., 2018; Nandan et al., 2019) and may therefore result in modified doublet occurrence probabilities. Positively correlated magnitudes could therefore contribute to closing the gap between simulated and observed doublet frequencies. Another, however more profound, research topic is the further evaluation of the impact of faulting types, event characteristics (e.g. dip, rake, and depth, etc.), and local geophysical parameters (e.g. strain rates, heat flow, tectonic plate velocities etc.) on the aftershock productivity and ultimately strong event clustering. This could also close the current gap in most seismic hazard models and lead to a better risk assessment by considering modeled damage based on more realistic, synthetic catalogs, including increased earthquake clustering and doublet occurrences.

Data and Resources

The National Research Institute for Earth Science and Disaster Resilience (NIED) earthquake mechanism catalog for Japan (Kubo et al., 2002) was downloaded from www.fnet.bosai.go.jp/event/search.php?LANG=en (last accessed on January 3, 2021). The Southern California Earthquake Data Center (SCEDC) focal mechanism catalog (Hauksson et al., 2012) was searched using scedc.caltech.edu/data/alt-2011-yang-hauksson-shearer.html (last accessed on January 3, 2021). Global earthquake data were obtained from the International Seismological Centre - Global Earthquake Model (ISC-GEM) Global Instrumental Earthquake Catalogue (Di Giacomo et al., 2018) at www.isc.ac.uk/iscgem/download.php (last accessed on January 3, 2021).

The ETAS model code used for this research was initially based on the CRAN R package repository *ETAS* (Jalilian, 2019) available at https://CRAN.R-project.org/package=ETAS (last accessed on January 3, 2021). The package is based on the original Fortran implementation *etas8p*, available at http://bemlar.ism.ac.jp/zhuang/software.html (last accessed on January 3, 2021).

Acknowledgments

We thank the anonymous reviewers for their helpful and constructive feedback that has significantly improved the work. Financial support for this work was provided by Munich Re through a scholarship granted to the first author, and by the Department of Statistics at Ludwig-Maximilians-University Munich.

References

Abdelnaby, A. E. (2012). Multiple earthquake effects on degrading reinforced concrete structures. PhD thesis, University of Illinois at Urbana-Champ.

Bach, C. and Hainzl, S. (2012). Improving empirical aftershock modeling based on additional source information. *Journal of Geophysical Research: Solid Earth*, **117**, B04312. doi: 10.1029/2011JB008901.

Blaser, L., Krüger, F., Ohrnberger, M., and Scherbaum, F. (2010). Scaling relations of earthquake source parameter estimates with special focus on subduction environment. *Bulletin of the Seismological Society of America*, **100**(6), 2914–2926. doi: 10.1785/0120100111.

Chu, A., Schoenberg, F. P., Bird, P., Jackson, D. D., and Kagan, Y. Y. (2011). Comparison of ETAS parameter estimates across different global tectonic zones. *Bulletin of the Seismological Society of America*, **101**(5), 2323–2339. doi: 10.1785/0120100115.

Cornell, C. A. (1968). Engineering seismic risk analysis. *Bulletin of the Seismological Society of America*, **58**(5), 1583–1606.

de Arcangelis, L., Godano, C., and Lippiello, E. (2018). The overlap of aftershock coda waves and short-term postseismic forecasting. *Journal of Geophysical Research: Solid Earth*, **123**, 5661–5674. doi: 10.1029/2018JB015518.

Di Giacomo, D., Robert Engdahl, E., and Storchak, D. A. (2018). The ISC-GEM earthquake catalogue (1904-2014): Status after the extension project. *Earth System Science Data*, **10**, 1877–1899. doi: 10.5194/essd-10-1877-2018.

Felzer, K. R., Abercrombie, R. E., and Ekström, G. (2004). A common origin for aftershocks, foreshocks, and multiplets. *Bulletin of the Seismological Society of America*, **94**(1), 88–98. doi: 10.1785/0120030069.

Field, E. H., Milner, K. R., Hardebeck, J. L., Page, M. T., van der Elst, N., Jordan, T. H., Michael, A. J., Shaw, B. E., and Werner, M. J. (2017). A spatiotemporal clustering model for the third uniform California earthquake rupture forecast (UCERF3-ETAS): Toward an operational earthquake forecast. *Bulletin of the Seismological Society of America*, **107**(3), 1049–1081. doi: 10.1785/0120160173.

Gibowicz, S. J. and Lasocki, S. (2005). Earthquake Doublets and Multiplets in the Fiji-Tonga-Kermadec Region. *Acta Geophysica Polonica*, **53**(3), 239–274.

Gulia, L., Rinaldi, A. P., Tormann, T., Vannucci, G., Enescu, B., and Wiemer, S. (2018). The effect of a mainshock on the size distribution of the aftershocks. *Geophysical Research Letters*, **45** (24), 13,277–13,287. doi: 10.1029/2018GL080619.

Gutenberg, B. and Richter, C. F. (1944). Frequency of earthquakes in California. *Bulletin of the Seismological Society of America*, **34**, 185–188. doi: 10.1038/156371a0.

Hainzl, S. and Marsan, D. (2008). Dependence of the Omori-Utsu law parameters on main shock magnitude: Observations and modeling. *Journal of Geophysical Research: Solid Earth*, **113**, B10309. doi: 10.1029/2007JB005492.

Hainzl, S., Christophersen, A., and Enescu, B. (2008). Impact of earthquake rupture extensions on parameter estimations of point-process models. *Bulletin of the Seismological Society of America*, **98**(4), 2066–2072. doi: 10.1785/0120070256.

Hainzl, S. (2016a). Rate-dependent incompleteness of earthquake catalogs. *Seismological Research Letters*, **87**(2A), 337–344. doi: 10.1785/0220150211.

Hainzl, S. (2016b). Apparent triggering function of aftershocks resulting from rate-dependent incompleteness of earthquake catalogs. *Journal of Geophysical Research: Solid Earth*, **121**(9), 6499–6509. doi: 10.1002/2016JB013319.

Hainzl, S., Zakharova, O., and Marsan, D. (2013). Impact of aseismic transients on the estimation of aftershock productivity parameters. *Bulletin of the Seismological Society of America*, **103**(3), 1723–1732. doi: 10.1785/0120120247.

Hauksson, E., Yang, W., and Shearer, P. M. (2012). Waveform relocated earthquake catalog for Southern California (1981 to June 2011). *Bulletin of the Seismological Society of America*, **102** (5), 2239–2244. doi: 10.1785/0120120010.

Hauksson, E., Yoon, C., Yu, E., Andrews, J. R., Alvarez, M., Bhadha, R., and Thomas, V. (2020). Caltech/USGS Southern California seismic network (SCSN) and Southern California earthquake data center (SCEDC): Data availability for the 2019 Ridgecrest sequence. *Seismological Research Letters*, **91**(4), 1961–1970. doi: 10.1785/0220190290.

Helmstetter, A. and Sornette, D. (2003). Båth's law derived from the Gutenberg-Richter law and from aftershock properties. *Geophysical Research Letters*, **30**(20), 2069. doi: 10.1029/2003GL018186.

Helmstetter, A., Kagan, Y. Y., and Jackson, D. D. (2006). Comparison of short-term and time-dependent earthquake forecast models for Southern California. *Bulletin of the Seismological Society of America*, **96**(1), 90–106. doi: 10.1785/0120050067.

Hutton, K., Woessner, J., and Hauksson, E. (2010). Earthquake monitoring in Southern California for seventy-seven years (1932-2008). *Bulletin of the Seismological Society of America*, **100**(2), 423–446. doi: 10.1785/0120090130.

Jalilian, A. (2019). ETAS: An R package for fitting the space-time ETAS model to earthquake data. *Journal of Statistical Software*, **88**(1). doi: 10.18637/jss.v088.c01.

Kagan, Y. Y. and Jackson, D. D. (1999). Worldwide doublets of large shallow earthquakes. *Bulletin of the Seismological Society of America*, **89**(5), 1147–1155.

Kagan, Y. Y., Bird, P., and Jackson, D. D. (2010). Earthquake patterns in diverse tectonic zones of the globe. *Pure and Applied Geophysics*, **167**(6), 721–741. doi: 10.1007/s00024-010-0075-3.

Kagan, Y. Y. (2004). Short-term properties of earthquake catalogs and models of earthquake source. *Bulletin of the Seismological Society of America*, **94**(4), 1207–1228. doi: 10.1785/012003098.

Kagermanov, A. and Gee, R. (2019). Cyclic pushover method for seismic assessment under multiple earthquakes. *Earthquake Spectra*, **35**(4), 1541–1558. doi: 10.1193/010518EQS001M.

Kubo, A., Fukuyama, E., Kawai, H., and Nonomura, K. (2002). NIED seismic moment tensor catalogue for regional earthquakes around Japan: Quality test and application. *Tectonophysics*, **356**, 23–48. doi: 10.1016/S0040-1951(02)00375-X.

Lay, T. and Kanamori, H. (1980). Earthquake doublets in the Solomon Islands. *Physics ofthe Earth and Planetary Interiors*, **21**, 283–304.

Marzocchi, W., Lombardi, A. M., and Casarotti, E. (2014). The establishment of an operational earthquake forecasting system in Italy. *Seismological Research Letters*, **85**(5), 961–969. doi: 10.1785/0220130219.

McGuire, R. K. (2008). Probabilistic seismic hazard analysis: Early history. *Earthquake Engineering and Structural Dynamics*, **37**, 329–338. doi: 10.1002/eqe.765.

Nandan, S., Ouillon, G., and Sornette, D. (2019). Magnitude of earthquakes controls the size distribution of their triggered events. *Journal of Geophysical Research: Solid Earth*, **124**(3), 2762–2780. doi: 10.1029/2018JB017118.

Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, **83**(401), 9–27.

Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Annals of the Institute of Statistical Mathematics*, **50**(2), 379–402.

Ogata, Y. (2011). Significant improvements of the space-time ETAS model for forecasting of accurate baseline seismicity. *Earth, Planets and Space*, **63**(3), 217–229. doi: 10.5047/eps.2010.09.001.

Ogata, Y. and Katsura, K. (1993). Analysis of temporal and spatial heterogeneity of magnitude frequency distribution inferred from earthquake catalogues. *Geophysical Journal International*, **113**(3), 727–738.

Ogata, Y. and Zhuang, J. (2006). Space-time ETAS models and an improved extension. *Tectonophysics*, **413**(1-2), 13–23. doi: 10.1016/j.tecto.2005.10.016.

Page, M. T., van Der Elst, N., Hardebeck, J., Felzer, K., and Michael, A. J. (2016). Three ingredients for improved global aftershock forecasts: Tectonic region, time-dependent catalog incompleteness, and intersequence variability. *Bulletin of the Seismological Society of America*, **106**(5), 2290–2301. doi: 10.1785/0120160073.

Papadopoulos, A. N., Bazzurro, P., and Marzocchi, W. (2020). Exploring probabilistic seismic risk assessment accounting for seismicity clustering and damage accumulation: Part I. Hazard analysis. *Earthquake Spectra*. doi: 10.1177/8755293020957338.

Ripley, B. D. (1976). The second-order analysis of stationary point processes. *Journal of Applied Probability*, **13**(2), 255–266.

Seif, S., Mignan, A., Zechar, J. D., Werner, M. J., and Wiemer, S. (2017). Estimating ETAS: The effects of truncation, missing data, and model assumptions. *Journal of Geophysical Research: Solid Earth*, **122**(1), 449–469. doi: 10.1002/2016JB012809.

Shearer, P. M. (2012). Self-similar earthquake triggering, Bath's law, and foreshock/aftershock magnitudes: Simulations, theory, and results for southern California. *Journal of Geophysical Research: Solid Earth*, **117**, B06310. doi: 10.1029/2011JB008957.

Storchak, D. A., Di Giacomo, D., Engdahl, E. R., Harris, J., Bondár, I., Lee, W. H., Bormann, P., and Villaseñor, A. (2015). The ISC-GEM global instrumental earthquake catalogue (1900-2009): Introduction. *Physics of the Earth and Planetary Interiors*, **239**, 48–63. doi: 10.1016/j.pepi.2014.06.009. URL http://dx.doi.org/10.1016/j.pepi.2014.06.009.

Utsu, T., Ogata, Y., and Matsu'ura, R. S. (1995). The centenary of the Omori formula for a decay law of aftershock activity. *J. Phys. Earth*, **43**, 1–33.

van Stiphout, T., Schorlemmer, D., and Wiemer, S. (2011). The effect of uncertainties on estimates of background seismicity rate. *Bulletin of the Seismological Society of America*, **101**(2), 482–494. doi: 10.1785/0120090143.

Veen, A. (2006). Some Methods of Assessing and Estimating Point Processes Models for Earthquake Occurrences. PhD thesis, University of California, Los Angeles.

Vere-Jones, D. (1969). A note on the statistical interpretation of Bath's law. *Bulletin of the Seismological Society of America*, **59**(4), 1535–1541.

Wells, D. L. and Coppersmith, K. J. (1994). New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacements. *Bulletin of the Seismological Society of America*, **84**(4), 974–1002.

Yang, W., Hauksson, E., and Shearer, P. M. (2012). Computing a large refined catalog of focal mechanisms for Southern California (1981-2010): Temporal stability of the style of faulting. *Bulletin of the Seismological Society of America*, **102**(3), 1179–1194. doi: 10.1785/0120110311.

Zakharova, O., Hainzl, S., Lange, D., and Enescu, B. (2017). Spatial variations of aftershock parameters and their relation to geodetic slip models for the 2010 Mw8.8 Maule and the 2011 Mw9.0 Tohoku-oki earthquakes. *Pure and Applied Geophysics*, **174**, 77–102. doi: 10.1007/s00024-016-1408-7.

Zhang, L., Werner, M. J., and Goda, K. (2018). Spatiotemporal seismic hazard and risk assessment of aftershocks of M 9 megathrust earthquakes. *Bulletin of the Seismological Society of America*, **108**(6), 3313–3335. doi: 10.1785/0120180126.

Zhang, L., Werner, M. J., and Goda, K. (2020). Variability of ETAS parameters in global subduction zones and applications to mainshock–aftershock hazard assessment. *Bulletin of the Seismological Society of America*, **110**, 191–212. doi: 10.1785/0120190121.

Zhuang, J., Ogata, Y., and Vere-Jones, D. (2002). Stochastic declustering of space-time earthquake occurrences. *Journal of the American Statistical Association*, **97**(458), 369–380. doi: 10.1198/016214502760046925.

Zhuang, J., Ogata, Y., and Wang, T. (2017). Data completeness of the Kumamoto earthquake sequence in the JMA catalog and its influence on the estimation of the ETAS parameters. *Earth, Planets and Space*, **69**, 36. doi: 10.1186/s40623-017-0614-6.

Contact to authors

Christian Grimm (main author), *Ludwig-Maximilians-University Munich, Department of Statistics, Ludwigstraße 33, 80539 Munich, Germany; Christian.Grimm@stat.uni-muenchen.de*,

Martin Käser, *Ludwig-Maximilians-University Munich, Department of Earth and Environmental Sciences, Geophysics, Theresienstraße 41, 80333 Munich, Germany*, also at *Munich Re, Section GeoRisks, Königinstr. 107, 80802 Munich, Germany*,

Sebastian Hainzl, *GFZ German Research Centre for Geoscience, Physics of Earthquakes and Volcanoes, Helmholtzstraße 6/7, 14467 Potsdam, Germany*,

Marco Pagani, *Global Earthquake Model Foundation, Via Ferrata 1, 27100 Pavia, Italy*,

Helmut Küchenhoff, *Ludwig-Maximilians-University Munich, Department of Statistics, Ludwigstraße 33, 80539 Munich, Germany*

List of Tables

Table 1: Overview of the specifications of the four ETAS model variants tested in this paper.

Model	spatial design	restriction factor	strike estimation	epicenter location
M_0	isotropic	100	-	-
M_1	anisotropic	100	optimized	optimized
M_2	isotropic	2.5	-	-
M_3	anisotropic	2.5	optimized	optimized

Table 2: Overview of model fit results for Japan and Southern California. The parameter D has been scaled to $D_{M=4.0} = D\ exp(\gamma(4.0 - M_c))$ in order to make results cross-comparable between regions (for Japan $D_{M=4.0} = D$ since $M_c = 4.0$).

Outcomes	Japan				Southern California			
	M_0	M_1	M_2	M_3	M_0	M_1	M_2	M_3
$l(\theta\|H_T)$	-21063	-18626	-22684	-19814	28444	30266	27144	30003
ν_{branch}	0.52	0.52	0.45	0.45	0.60	0.57	0.54	0.53
$\mu\ (day^{-1})$	0.51	0.54	0.61	0.64	0.18	0.19	0.21	0.21
$\tilde{A}$	0.26	0.24	0.12	0.11	0.34	0.27	0.25	0.22
$\alpha\ (mag^{-1})$	1.21	1.28	1.78	1.84	1.18	1.41	1.48	1.59
$c\ (days)$	0.015	0.017	0.013	0.014	0.011	0.012	0.013	0.012
p	1.02	1.05	1.00	1.03	1.07	1.08	1.08	1.09
$D_{M=4.0}\ (km^2)$	2.274	0.194	2.466	0.117	0.441	0.584	0.403	0.849
$\gamma\ (mag^{-1})$	1.72	1.73	2.05	2.48	1.37	1.78	1.86	1.95
q	1.43	1.20	1.60	1.21	1.48	1.71	1.19	1.95

List of Figures

6 Boxplot representation of the coefficients of variation (CV) of monthly numbers of event occurrences in the simulated catalogs, based on the four estimated models, for (a) Japan and (b) Southern California. The black horizontal line represents the CV of the respective original earthquake catalog. The red '+'-symbols represent outliers. (c) Comparison of monthly event occurrences between the original Japan catalog (black line) and the ETAS rate for Japan's model M_3, integrated piece-wise for the monthly integrals, based on trigger contributions of the original history of events. (red line). (d) Analysis of the degree of spatial clustering by Riley's K function. Solid lines represent results for synthetic catalogs, generated by model M_3 for Japan (JPN) and Southern California (CAL). Dashed lines show results for the respective original earthquake catalogs. The dotted black line represents Riley's K function values for a homogeneous Poisson process. Values above indicate clustering, values below signify inhibition.

7 (a) Percentages of doublet occurrences, depending on the considered event magnitude, for the four model variants in Japan. Solid lines represent simulated catalogs. Dashed lines show simulated sequences. Magnitudes are aggregated in 0.2-magnitude unit steps from $M_w = 5.9$ to $M_w = 7.1$, then in 0.3-unit steps up to $M_w = 8.0$, followed by one interval for all magnitudes above. (b) Proportions of doublet pairs generated by (i) independent seismic background events or unrelated clusters (dash-dotted lines) or (ii) events of the same cluster (solid lines). Results are presented for models M_0 and M_3 in Japan. (c,d) Comparison of the doublet occurrence frequencies in synthetic catalogs (blue lines) to historic catalogs (black lines), for (c) model M_0 and (d) model M_3, both Japan. Shaded ranges represent 10/90% confidence interval (CI) of the synthetic catalogs. Events are aggregated in the magnitude intervals labeled on the x-axis. Tohoku events are extracted in simulated catalogs.

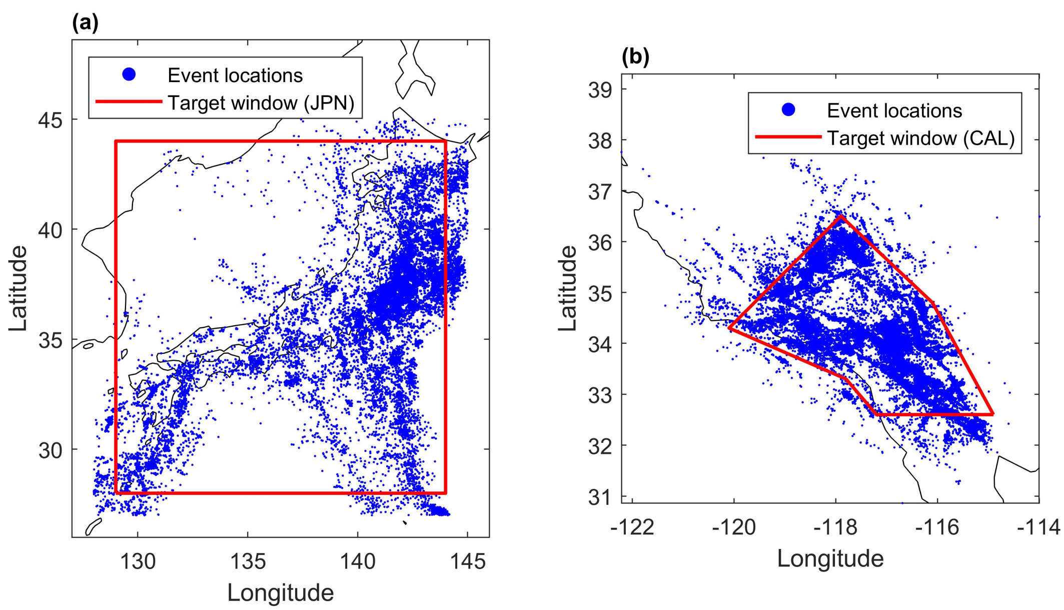

Fig 1: Event locations in the two utilized earthquake catalogs, including both target and complementary events. Red polygons represent the respective spatial target window. (a) Events in NIED catalog for Japan, $M_w \geq 4.0$, target period from 07/01/1997 until 10/31/2020, complementary period from 01/01/1997 until 06/30/1997; (b) Events in SCEDC catalog for Southern California, $M_w \geq 2.8$, target period from 07/01/1981 until 12/31/2019, complementary period from 01/01/1981 until 06/30/1981

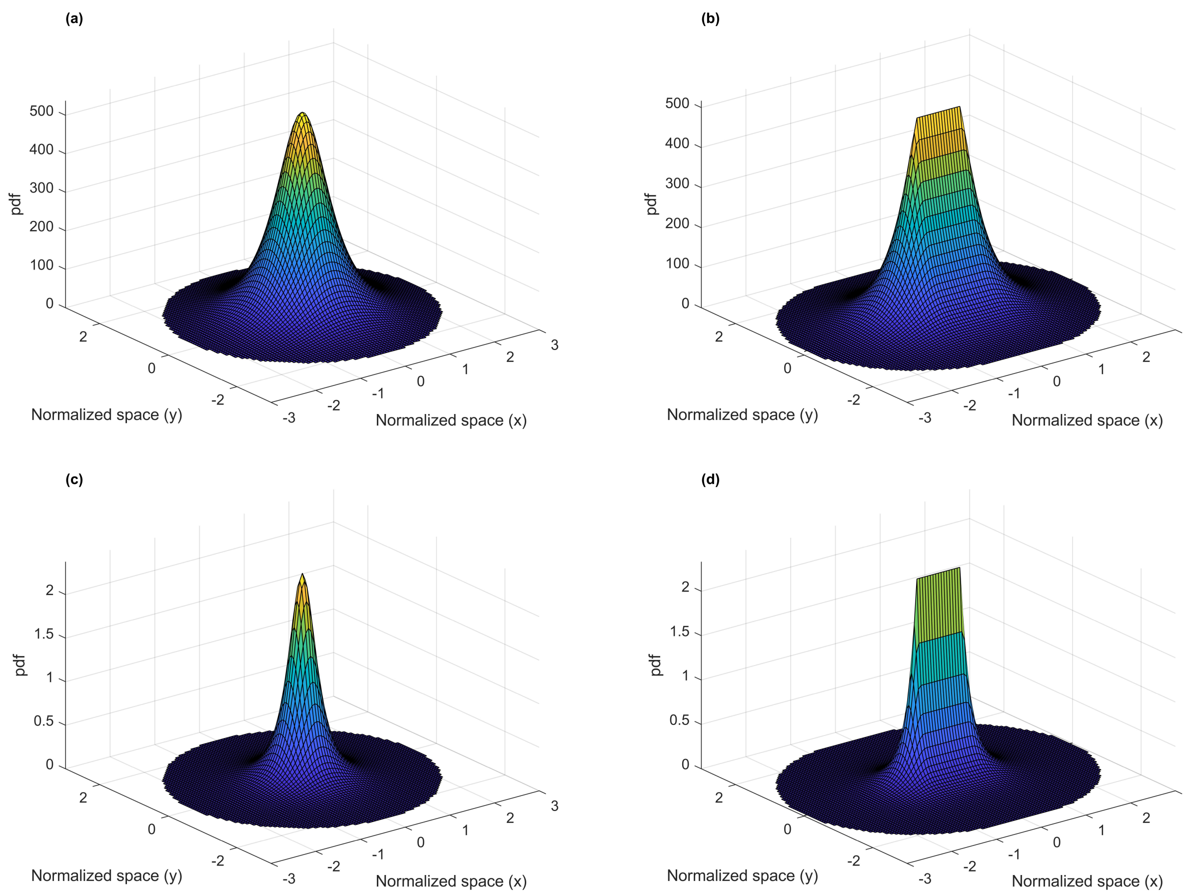

Fig 2: Visualization of the spatial kernels restricted to a distance of $\tilde{r}(m_i) = 2.5 \cdot l(m_i)$: (a) isotropic kernel for magnitude $m = 5.0$, (b) anisotropic kernel for magnitude $m = 5.0$, (c) isotropic kernel for magnitude $m = 7.5$ and (d) anisotropic kernel for magnitude $m = 7.5$. The 3D pdfs result from equation 7, using the initial spatial parameter guesses $D = 2.0$, $\gamma = 2.1$ and $q = 1.5$.

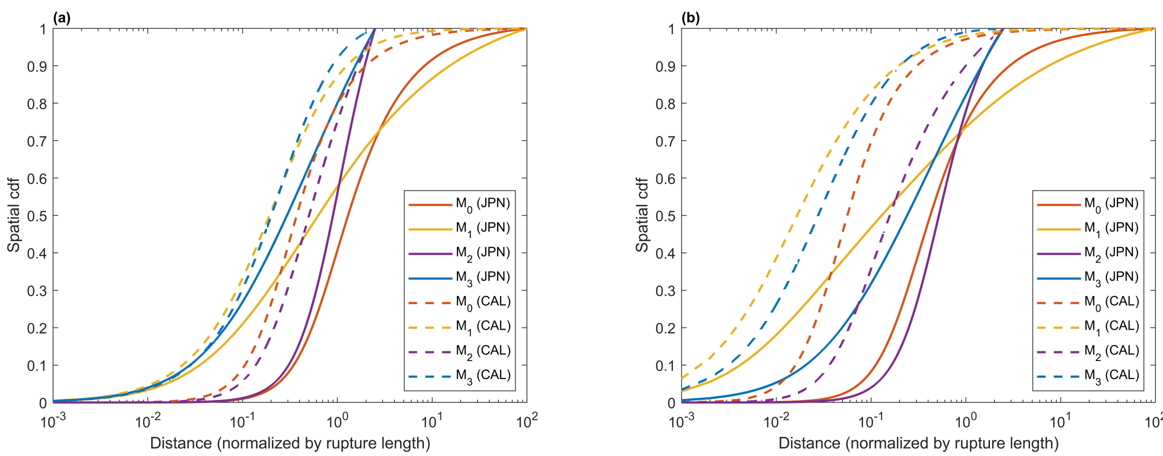

Fig 3: Cumulative distribution functions of spatial kernels for trigger magnitudes (a) $m = 5.0$ and (b) $m = 7.5$. Solid lines show Japan (JPN) models. Dashed lines represent Southern California (CAL) models. The x-axis is defined as the distance to the point source location (for isotropic models M_0 and M_2) or rupture line (for anisotropic models M_1 and M_3), normalized by the rupture length estimate for the respective region.

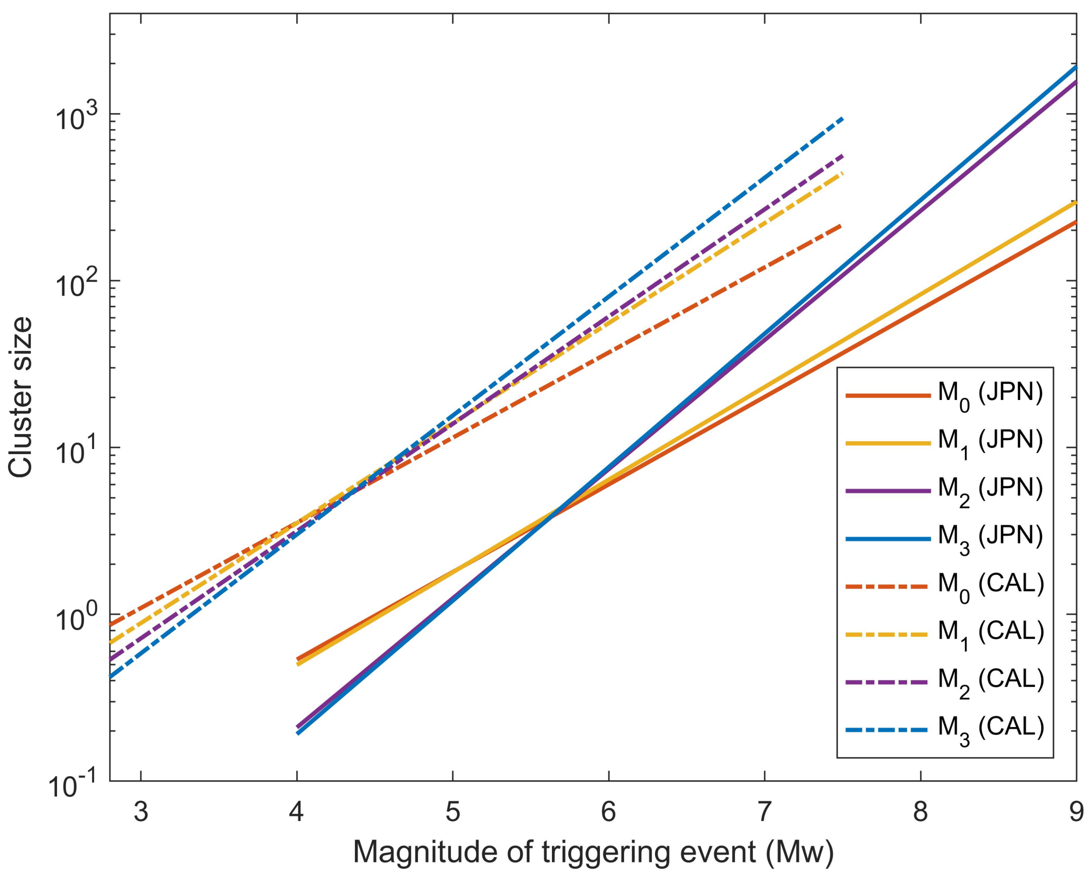

Fig 4: Expected cluster sizes according to equation (14). The x-axis states the magnitude of the sequence-initiating mainshock event. The y-axis is on logarithmic scale and denotes the average number of cluster members. Solid lines show Japan (JPN) models, starting from catalog cut-off magnitude $M_c = 4.0$. Dashed lines represent Southern California (CAL) models, starting from catalog cut-off magnitude $M_c = 2.8$ and ending at the assumed maximum magnitude $m = 7.5$.

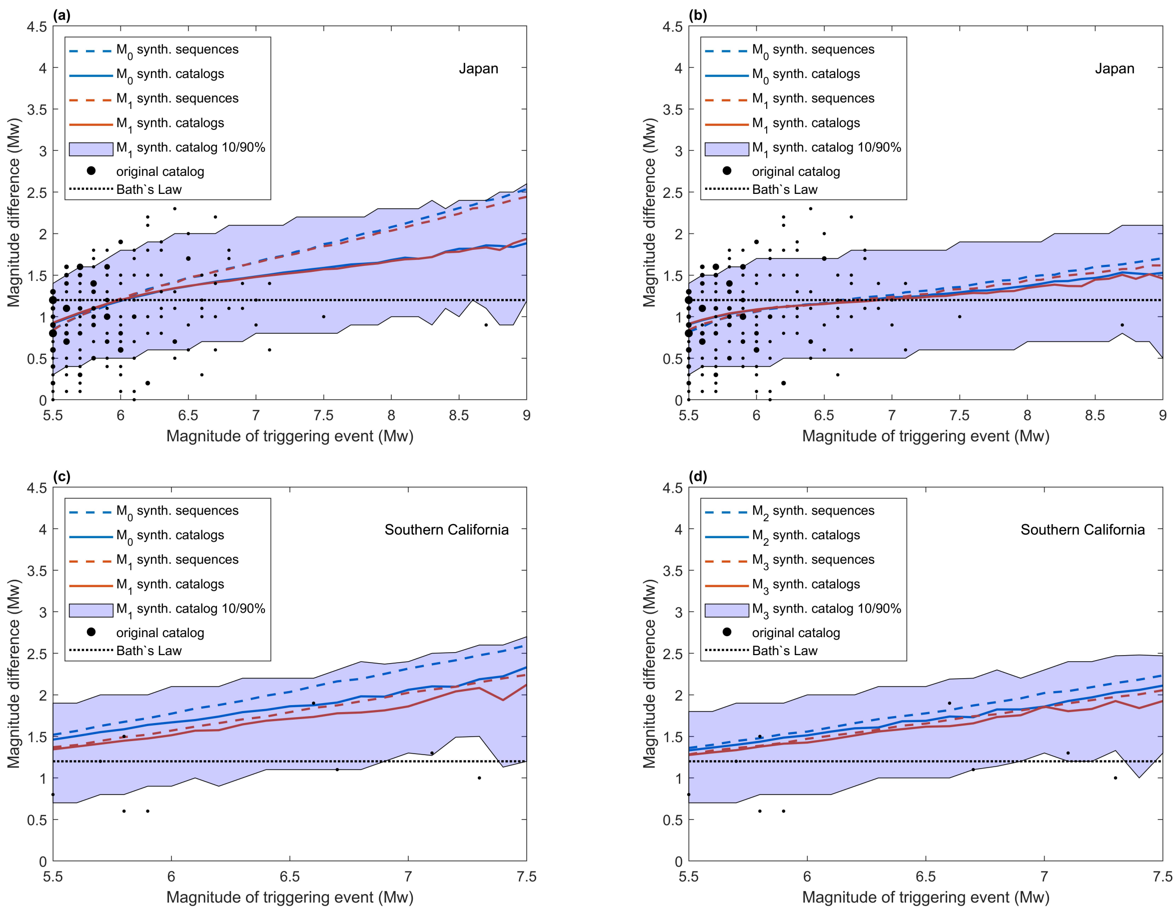

Fig 5: Approximations of the average magnitude difference between a considered mainshock event and the strongest event following in the specified time-space window, for (a) unrestricted models M_0 and M_1 in Japan (JPN), (b) restricted models M_2 and M_3 in JPN, (c) unrestricted models M_0 and M_1 in Southern California (CAL) and (d) restricted models M_2 and M_3 in CAL. Solid lines show catalog simulations, dashed lines represent sequence simulations. The shaded range visualizes the $10\% - 90\%$ confidence interval of the respective catalog simulation. Black dots represent observations in the underlying original catalogs, and are sized according to the number of points stacked. The horizontal dotted line is consistent with the Bath's law prediction of a magnitude difference of 1.2 units independent of the absolute size of the trigger magnitude.

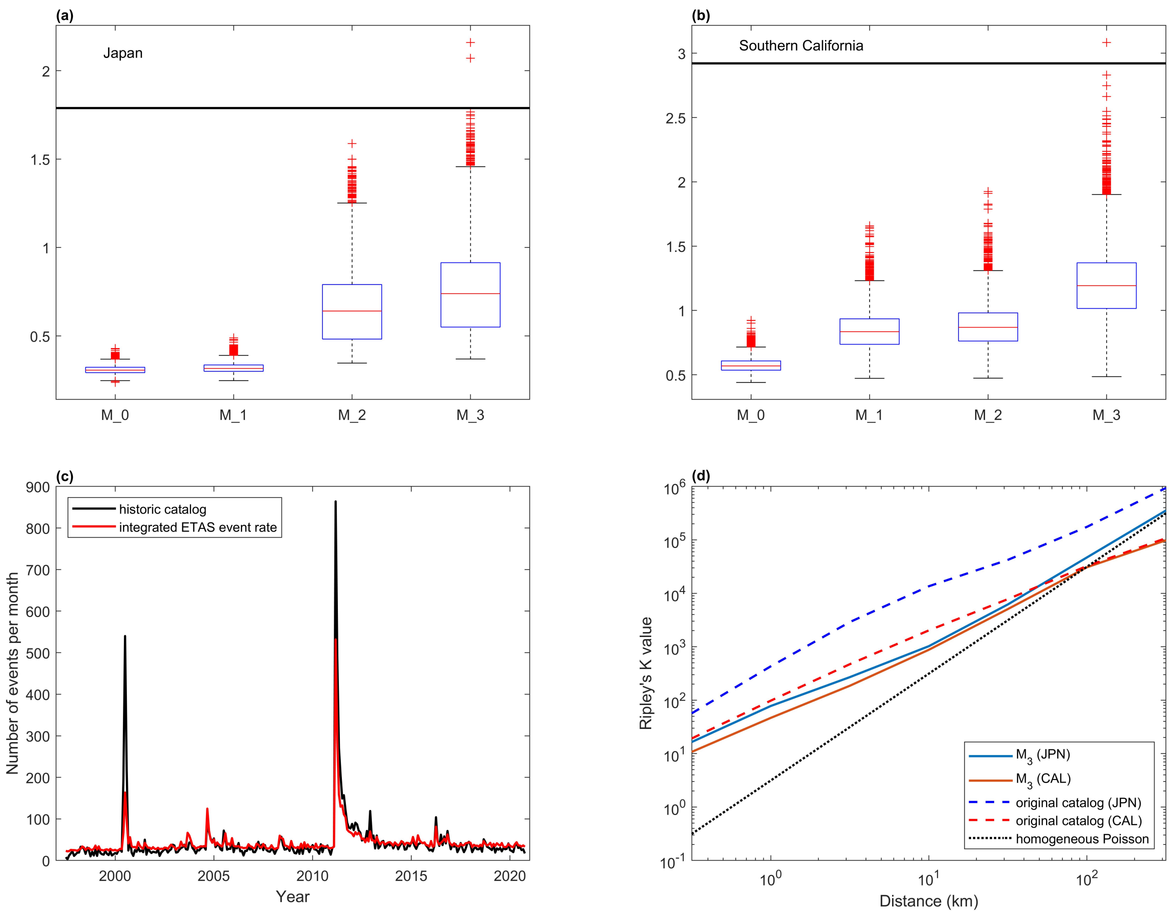

Fig 6: Boxplot representation of the coefficients of variation (CV) of monthly numbers of event occurrences in the simulated catalogs, based on the four estimated models, for (a) Japan and (b) Southern California. The black horizontal line represents the CV of the respective original earthquake catalog. The red '+'-symbols represent outliers. (c) Comparison of monthly event occurrences between the original Japan catalog (black line) and the ETAS rate for Japan's model M_3, integrated piece-wise for the monthly integrals, based on trigger contributions of the original history of events. (red line). (d) Analysis of the degree of spatial clustering by Riley's K function. Solid lines represent results for synthetic catalogs, generated by model M_3 for Japan (JPN) and Southern California (CAL). Dashed lines show results for the respective original earthquake catalogs. The dotted black line represents Riley's K function values for a homogeneous Poisson process. Values above indicate clustering, values below signify inhibition.

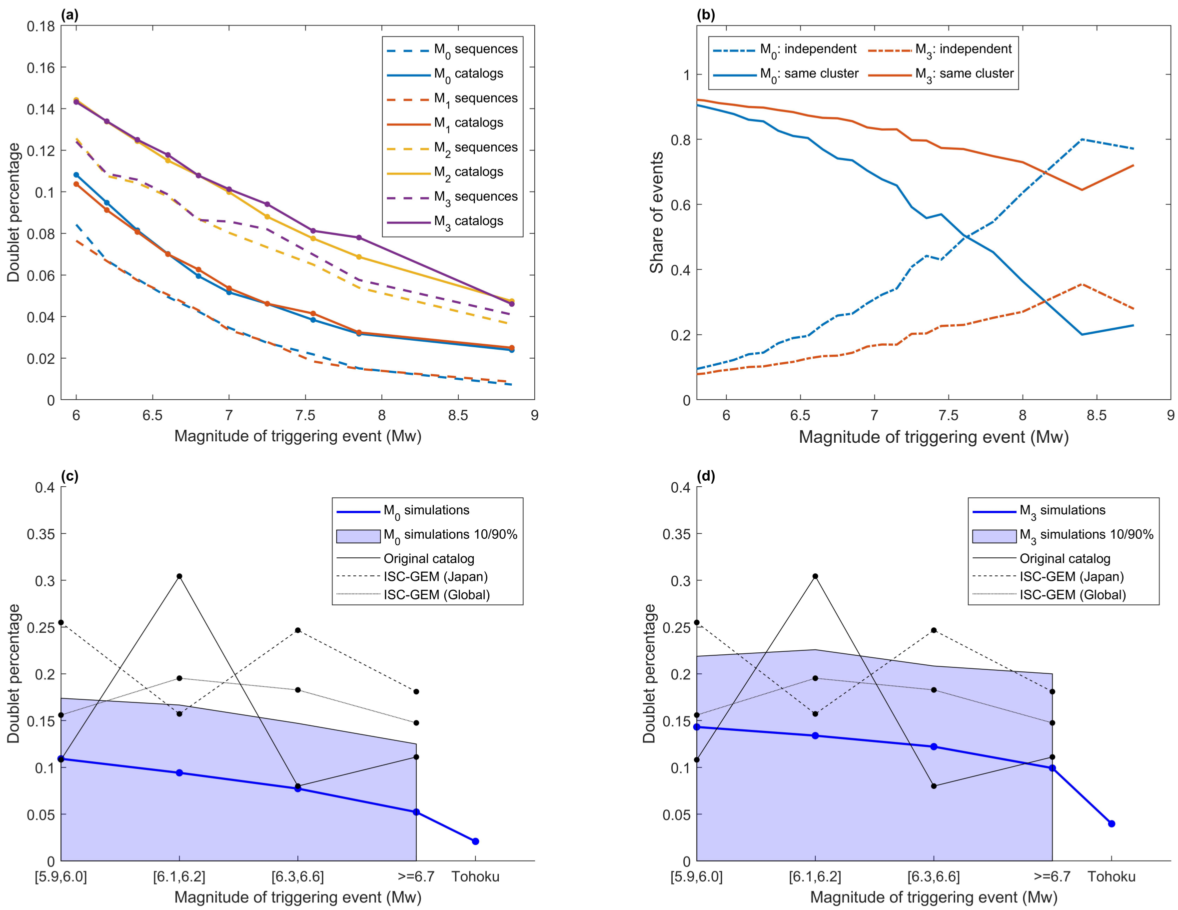

Fig 7: (a) Percentages of doublet occurrences, depending on the considered event magnitude, for the four model variants in Japan. Solid lines represent simulated catalogs. Dashed lines show simulated sequences. Magnitudes are aggregated in 0.2-magnitude unit steps from $M_w = 5.9$ to $M_w = 7.1$, then in 0.3-unit steps up to $M_w = 8.0$, followed by one interval for all magnitudes above. (b) Proportions of doublet pairs generated by (i) independent seismic background events or unrelated clusters (dash-dotted lines) or (ii) events of the same cluster (solid lines). Results are presented for models M_0 and M_3 in Japan. (c,d) Comparison of the doublet occurrence frequencies in synthetic catalogs (blue lines) to historic catalogs (black lines), for (c) model M_0 and (d) model M_3, both Japan. Shaded ranges represent 10/90% confidence interval (CI) of the synthetic catalogs. Events are aggregated in the magnitude intervals labeled on the x-axis. Tohoku events are extracted in simulated catalogs.