Hainzl, S., Christophersen, A. (2017): Testing alternative temporal aftershock decay functions in an ETAS framework. - *Geophysical Journal International*, *210*, 2, pp. 585—593.

DOI: http://doi.org/10.1093/gji/ggx184

Geophysical Journal International

Geophys. J. Int. (2017) **210,** 585–593
Advance Access publication 2017 May 4
GJI Seismology

doi: 10.1093/gji/ggx184

Testing alternative temporal aftershock decay functions in an ETAS framework

S. Hainzl[1] and A. Christophersen[2]

[1] *GFZ German Research Centre for Geosciences, Potsdam, Germany. E-mail: hainzl@gfz-potsdam.de*
[2] *GNS Science, Lower Hutt, New Zealand*

Accepted 2017 May 3. Received 2017 April 28; in original form 2017 January 3

SUMMARY

Earthquake clustering can be well described by the Epidemic Type Aftershock Sequence model (ETAS), where each earthquake potentially triggers its own aftershocks. The temporal decay of aftershocks is most commonly modelled with a power law, the so-called Omori–Utsu law. However, new results suggest that alternative decay functions may be more appropriate. One recent study found that a version of the ETAS model fitted the data better when the Omori–Utsu law was truncated in time. A finite triggering time is consistent with the rate-state model that expects an exponential roll-off after a finite time following the initial power law decay. Another recent study compared a power law, pure exponential and stretched exponential and found that the stretched exponential described the overall decay of aftershocks best. Our aim is to find the best temporal aftershock decay function within the ETAS model framework. We investigate six decay functions; three power laws and three exponential decays. The power laws are an unlimited Omori–Utsu law, a sharply truncated Omori–Utsu law, and an exponential roll-off consistent with the rate-state friction model. The exponential decay functions are the pure exponential, stretched exponential and a modified stretched exponential. We fit model parameters for each decay function to 326 individual earthquake sequences from four regional and one global earthquake catalogue. The three models that fit most of the sequences the best are the truncated Omori–Utsu law (32 per cent of sequences), the power law based on the rate-state friction model (26 per cent) and the unlimited Omori–Utsu law (23 per cent). When the parameters are not fitted individually but the median model parameters are used for each function, the modified stretched exponential function fits most (28 per cent) sequences the best, followed by the unlimited Omori–Utsu law (22 per cent) and the stretched exponential (18 per cent). However, the majority of sequences (53 per cent) is still best fit by a power law. Out of all the tested decay functions, the one based on the rate-state friction model is the only one that performs in a majority of cases better than the Omori–Utsu law for fixed parameters. This suggests that it could be a potential candidate to replace the unlimited Omori–Utsu law in ETAS-model-based earthquake forecasts.

Key words: Earthquake hazards; Earthquake interaction, forecasting, and prediction; Statistical seismology.

1 INTRODUCTION

Aftershocks occur after almost all large shallow earthquakes and can cause further damages as for example in the Canterbury sequence, New Zealand (Bannister & Gledhill 2012). Although aftershocks are most abundant shortly after the mainshock, they can continue to occur with decaying rate for months and years. Therefore, the specific functional form of the temporal decay is of crucial importance for time-varying seismic hazard assessments.

The decay of aftershocks is most commonly modelled by the Omori–Utsu law, a power law, which can present some challenges in practical implementation. The decay rate of aftershocks $R(t)$ at time t is given by

$$R(t) = K_0(c + t)^{-p}. \quad (1)$$

The decay exponent p typically takes values in the range 0.8–1.2; the time-offset parameter c avoids a singularity at time zero, and the productivity is described by K_0; see Utsu *et al.* (1995) for a

review. The proportionality factor K_0 increases exponentially with mainshock magnitude M according to $K_0 \sim 10^{\alpha(M-M_{\min})}$, where α is a constant around 1 and $M_{\min}$ is the lower magnitude cut-off of the earthquakes under consideration. The time offset parameter c is usually much smaller than one day, and is associated with reduced detection ability of the operating seismic network immediately after large events (Kagan 2004; Hainzl 2016a,b). The functional form of the Omori–Utsu law poses challenges: For $p \leq 1$ it does not integrate and has to be truncated at a finite triggering time T for the expected number of aftershocks to be finite. Even for $1 < p < 1.2$, the aftershock activity can continue for thousands if not millions of years according to the parameters fitted to the early part of aftershock decay (Harte 2013). However, we do not have the data available to test the applicability of the Omori–Utsu law for these time periods. Many aftershock sequences tend to last until the end the earthquake catalogues (Godano & Tramelli 2016). While the 1891 Great Nobi, Japan earthquake still had its aftershocks decaying according the Omori–Utsu law after 100 yr (Utsu *et al.* 1995; Hainzl & Christophersen 2016), most earthquake catalogues provide homogeneous data of sufficient quality for aftershock studies for not more than 30–50 yr, which complicates the analysis of the tail of the aftershock decay.

Recently two alternatives to the Omori–Utsu law have been suggested: The truncation of the Omori–Utsu law at a finite triggering time (Hainzl *et al.* 2016) and the stretched exponential (Mignan 2015). A finite triggering time is consistent with physics-based models (Hainzl *et al.* 2016). In particular, the rate-state friction model predicts that a population of faults will respond to a sudden stress increase by producing aftershocks with a rate equal to the Omori–Utsu decay with $p = 1$, but with an exponential roll-off at larger times (Dieterich 1994; Dieterich *et al.* 2000). In contrast, Mignan (2015) found that the stretched exponential function better describes the temporal decay of aftershocks than a power law. He thus inferred that aftershocks are caused by a simple relaxation process like many other processes observed in nature. Mignan (2015) fitted a power law, a pure exponential and a stretched exponential in their simplest form to the overall decay of individual aftershock sequences in three regional earthquake catalogues. Apart from a problem with the visual presentation of the data (Hainzl & Christophersen 2016; Mignan 2016), his statistical analysis showed that the overall aftershock decay for times $t > t_{\min}$ can be better modelled by the stretched exponential function of the form $t^{\beta-1}\exp(-\lambda t^{\beta})$ with the two parameters λ, β with $t_{\min}$ in the order of an hour. These results were confirmed for three methods of defining aftershocks sequences but the fitting did not account for background seismicity and secondary aftershock activity triggered by preceding aftershocks. Previous tests of the stretched exponential were less successful (Kisslinger 1993; Gross & Kisslinger 1994; Lolli & Gasperini 2006; Lolli *et al.* 2009), but improved when background seismicity was included (Gasperini & Lolli 2009). For the overall aftershock decay of the Nobi earthquake the Omori–Utsu law was found to be the best decay function (Hainzl & Christophersen 2016; Mignan 2016). Still, the findings from regional catalogues could have important consequences for understanding aftershock generation and modelling time-varying seismic hazard.

The functional form of the aftershock decay functions needs more thorough testing to include secondary triggering and potential background activity, which can affect the tail of the decay function when the aftershock rate is small. To account for independent events and secondary aftershocks in a consistent manner, we apply the Epidemic Type Aftershock Sequence (ETAS) model, which was developed by Ogata (1988) to model cascading earthquake occurrence. The total rate of earthquakes at time t is the sum of background seismicity and on-going aftershocks triggered by all past events and can be described by

$$R(\vec{x}, t) = \mu(\vec{x}) + \sum_{i:t_i<t} K 10^{\alpha(M_i - M_{\min})}(c + t - t_i)^{-p} g(\vec{x} - \vec{x}_i, M_i) \tag{2}$$

where $\mu(\vec{x})$ is the space-dependent background rate, K a proportionality factor, and g the spatial probability density function for triggered aftershocks. It is important to remember that for $p \leq 1$, the Omori–Utsu law would predict an infinite number of direct aftershocks and thus the total seismicity in the ETAS model would escalate with time (Harte 2013; Zhuang *et al.* 2013). Thus this model is not self-consistent for $p \leq 1$. Hainzl *et al.* (2016) introduced a truncation time T for the Omori–Utsu decay in the ETAS model, which leads to a finite aftershock number also for $p \leq 1$. By means of the analysis of synthetic simulations and empirical earthquake catalogues, they found that maximum likelihood fits often reveals T-values significantly shorter than the catalogue length and that falsely assuming $T = \infty$ can significantly bias the estimations of the other ETAS parameters. In general, the ETAS model can be rewritten in the case of self-consistent decay functions (that can be normalized) as

$$R(\vec{x}, t) = \mu(\vec{x}) + \sum_{i:t_i<t} N_0 10^{\alpha(M_i - M_{\min})} f(t - t_i) g(\vec{x} - \vec{x}_i, M_i) \tag{3}$$

with a temporal probability density function f. In this case, N_0 is the expected total number of $M \geq M_{\min}$ aftershocks triggered by an earthquake with magnitude $M_{\min}$.

The spatial aftershock triggering is usually anisotropic due to extended complex earthquake ruptures. Thus its parametrization might introduce some bias in the ETAS model estimation (Hainzl *et al.* 2008) as well as the unknown spatial variation of the background rate. Because we are mainly interested in the temporal decay, we follow the approach of Hainzl *et al.* (2016) and Zakharova *et al.* (2017) and only consider the temporal evolution of the integrated activity in a region A to reduce the effects of the unknown details of the spatial variation. The integration of eq. (3) yields

$$R_A(t) = \mu_A + \sum_{i:t_i<t} w_i N_0 10^{\alpha(M_i - M_{\min})} f(t - t_i) \tag{4}$$

with background rate $\mu_A = \int_A \mu(\vec{x})\,\mathrm{d}\vec{x}$ and the weighting factor $w_i = \int_A g(\vec{x} - \vec{x}_i, M_i)\,\mathrm{d}\vec{x}$ which represents the fraction of aftershocks of the ith event that is expected to occur in A. Note that the summation on the right side of the equation includes all earthquakes in the catalogue occurring before time t independent of whether the events occurred inside or outside of A.

In the following, we will use eq. (4) to systematically compare the performance of the infinite and truncated Omori–Utsu law, the stretched and simple exponential function, and the physics-based rate-state decay function for a large number of earthquake sequences recorded in global and regional catalogues. Our aim is to identify the decay function that best describes the temporal decay of aftershocks within the ETAS framework. The tested functions are summarized in Section 2. The data and method are introduced in Section 3. Finally, the results are presented and discussed in Sections 4–6, respectively.

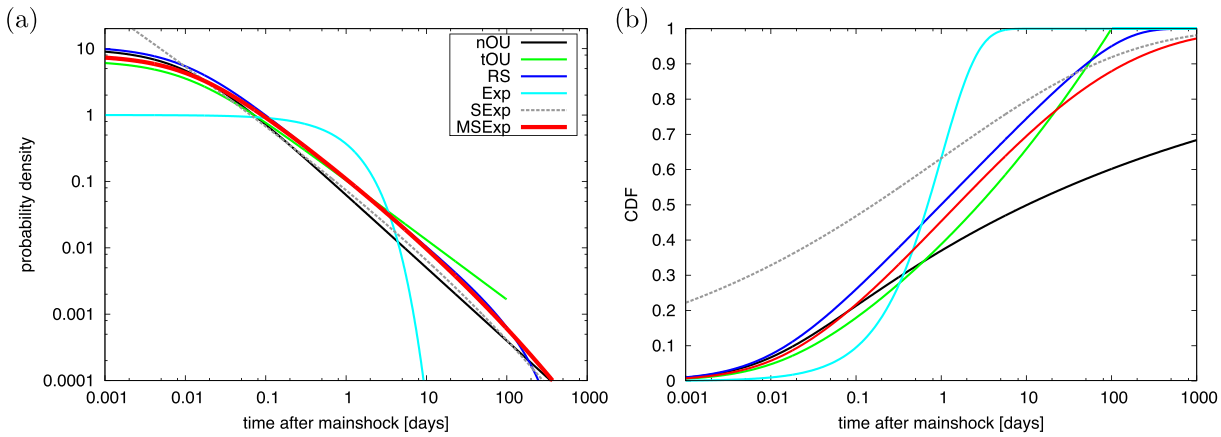

Figure 1. (a) Probability density functions of the different tested decay functions for exemplary parameter values. The corresponding cumulative distribution function are shown in plot (b). Parameters are $c_1 = c_2 = c_6 = 0.01$, $T = t_a = 100$, $a = 1$ (all in units of days), $p_1 = 1.1$, $p_2 = 0.9$, $\lambda_5 = \lambda_6 = 1.0$, and $\beta_5 = \beta_6 = 0.2$.

2 DECAY FUNCTIONS

We study six different decay functions:

Normalized Omori–Utsu law (nOU). For $p > 1$, the Omori–Utsu law can be normalized and its probability density function is given by

$$f_1(t) = (p_1 - 1)c_1^{p_1-1}(c_1 + t)^{-p_1} \tag{5}$$

with the two parameters c_1, p_1.

Truncated Omori–Utsu law (tOU). Assuming a truncation at time T, the Omori–Utsu law can be normalized for any value of p and its probability density function is given by

$$f_2(t) = C\,(c_2 + t)^{-p_2}\; H(T - t) \tag{6}$$

with H being the Heaviside-function ($H(s) = 1$ for $s \geq 0$ and 0 otherwise) and the normalization coefficient C being equal to

$$C = \begin{cases} (p_2 - 1)c_2^{p_2-1} / \left(1 - (1 + \frac{T}{c_2})^{1-p_2}\right) & \text{for } p_2 \neq 1, \\ [\ln(c_2 + T) - \ln(c_2)]^{-1} & \text{for } p_2 = 1. \end{cases}$$

The three parameters of this function are c_2, p_2, and T. Note that the non-normalized formulation of the Omori–Utsu decay in the standard ETAS model (eq. 2) can account for $p \leq 1$. However, in this case it is implicitly assumed that a truncation exists and the corresponding T-value exceeds the catalogue lengths.

Rate-state response (RS). The decay function of the RS model is identical to the Omori–Utsu law with $p = 1$ with an exponential roll-off at later times. The framework of rate-and-state friction (Dieterich 1994; Dieterich *et al.* 2000) takes into consideration the rate- and slip-dependence of frictional strength and time-dependent re-strengthening observed in laboratory experiments. Assuming a population of nucleation sites in a stationary regime with background rate μ, the rate of additionally triggered aftershocks in response to a single positive stress jump ΔCFS at time $t = 0$ evolves according to

$$R(t) = \frac{\mu}{1 - Be^{-\frac{t}{t_a}}} - \mu \tag{7}$$

where $B = 1 - \exp(-\Delta\text{CFS}/A\sigma)$ with $A\sigma$ being the frictional resistance and t_a the aftershock relaxation time which is inversely proportional to the tectonic stressing rate $\dot{\tau}$, i.e. $t_a = A\sigma/\dot{\tau}$. This decay function is identical to the Omori–Utsu law with $p = 1$ and $c = (1 - B)t_a/B$ at short times, but with an exponential roll-off at time t_a (Cocco *et al.* 2010).

Its normalized form is given by

$$f_3(t) = -\frac{B}{t_a \ln(1 - B)} \frac{1}{e^{\frac{t}{t_a}} - B} \tag{8}$$

with the two parameters t_a and B, where $0 < B < 1$.

Exponential function (Exp). The exponential function is the simplest decay function known to describe linear relaxation processes, and that is why we also test it. Its normalized form is given by

$$f_4(t) = ae^{-at} \tag{9}$$

with the parameter a.

Stretched exponential function (SExp). The normalized form of the stretched exponential form used by Mignan (2015) is given by

$$f_5(t) = \lambda_5\, \beta_5\, t^{\beta_5 - 1}\, e^{-\lambda_5 t^{\beta_5}} \tag{10}$$

with the two parameters λ_5 and β_5, where $0 < \beta_5 < 1$.

Modified stretched exponential function (MSExp). To allow for a time offset immediately following the mainshock and prior to the onset of the decay, Gross & Kisslinger (1994) introduced a parameter c_6-value similar to the c-value in the Omori–Utsu law. The normalized form of the modified stretched exponential form is given by

$$f_6(t) = \lambda_6\, \beta_6\, e^{\lambda_6 c_6^{\beta_6}}\, (c_6 + t)^{\beta_6 - 1}\, e^{-\lambda_6 (c_6 + t)^{\beta_6}} \tag{11}$$

with the three parameters c_6, λ_6, and β_6, where $0 < \beta_6 < 1$. The parameter c_6 can also account for the partial incompleteness of earthquake catalogues directly after mainshocks.

Fig. 1 shows a comparison of the different decay functions for exemplary parameter sets. Despite different parameters (e.g. $p_1 = 1.1$ in the case of nOU and $p_2 = 0.9$ in the case of tOU) and functional forms, the shapes of probability density functions look rather similar in the doubly-logarithmic presentation, except for the exponential function. However, differences become clear in the log-linear plot (Fig. 1b) of the cumulative distribution function which presents the fraction of aftershocks occurring on average before the given time. Thus the different decay functions are expected to lead to significantly different fits.

3 DATA AND METHOD

The empirical mainshock–aftershock sequences to be analysed are selected from the following five catalogues:

Southern California catalogue (SCA). We use the updated relocated Southern California catalogue containing earthquakes from 1981 to 2014 (Hauksson *et al.* 2012). We use a cut-off magnitude of $M_{\min} = 3.0$, which is well above the reported completeness of $M_c = 1.8$ for this time period (Hauksson *et al.* 2012). Except following a large earthquake, the selected sub-set is therefore complete and includes 12 947 earthquakes above this threshold.

Northern California catalogue (NCA). We use the updated relocated Northern California catalogue containing earthquakes from 1984 to 2011 (Waldhauser & Schaff 2008). We use a cut-off magnitude of $M_{\min} = 3.0$ leading to 9350 earthquakes above this threshold.

New Zealand catalogue (NZ). We use the GeoNet catalogue of New Zealand earthquakes which, until the end of 2011, was processed by the CalTech-USGS seismic processor (CUSP) system (Lee & Stewart 1989). We selected earthquakes inside the testing region of the Collaboratory for the Study of Earthquake Predictability (CSEP) which includes the main islands of New Zealand and extends about 50 km offshore with a depth cut-off of 40 km (Gerstenberger & Rhoades 2010). Considering changes in the magnitude of completeness over time, we use the cut-off magnitude of $M_{\min} = 3.5$ and select for our model fits only events that occurred after 1987. This data set includes 10,154 events. However, pre-1987 events are used as input for the ETAS rate estimate in the time period after 1987.

Taiwan catalogue (Taiwan). We use the relocated catalogue of Taiwan from 1991 to 2005 of Wu *et al.* (2008), with a cut-off magnitude of $M_{\min} = 3.0$ and a depth cut-off of 40 km, containing 24,631 events.

Global catalogue (global). We analyse the ISC-GEM Global Instrumental Earthquake Catalogue (www.isc.ac.uk/iscgem) (Storchak *et al.* 2013). We use the cut-off magnitude of $M_{\min} = 5.6$ and select only shallow events with a depth less than 80 km in the time period 1964–2012. This selection yields a catalogue of 12,780 events. However, pre-1964 events are used as input for the ETAS rate estimate in the time period after 1964.

3.1 Sequence selection

Instead of analysing the seismicity with universal parameters, we allow for potential variations of the ETAS parameters in space by analysing the different decay functions separately for different earthquake sequences. By using the ETAS rate integrated over a localized region A (eq. 4), we minimize the potential bias due to anisotropic spatial triggering and inhomogeneous background activity without ignoring the aftershocks triggered by events that occurred outside of A. The spatial areas A are circular areas centred around mainshocks defined by a simple window-based procedure introduced by Tahir *et al.* (2012). We use exactly the same procedure as in Hainzl *et al.* (2016), where it is described in detail. Here we only provide a brief summary.

An earthquake with magnitude M is defined as a mainshock if it is the largest earthquake within the time period of $\pm$ one year and distance range D, where the spatial window is set to be a multiple of the rupture length, that is, $D = \tilde{D}\, L(M)$. Here we use $L(M) = 10^{-2.44 + 0.59M}$ (km) according to the average empirical rupture length of an earthquake with magnitude M (Wells & Coppersmith 1994). The majority of aftershocks usually occur very close to the mainshock rupture, but remotely triggered aftershocks can also occur which will be missed for small values of $\tilde{D}$. As a compromise, we choose $\tilde{D} = 3$ but we show the robustness of our results also for different values (see the Supporting Information).

However, the choice of the parameter values for the mainshock selection is not crucial because we use the ETAS model, which does not require any pre-definitions of mainshocks, aftershocks and background events. Furthermore, we account for aftershocks triggered by earthquakes outside the circular spatial region. For this purpose, we calculate for all earthquakes in the catalogue, the fraction w of aftershocks which is expected to occur in the study region by means of the empirical probability density distribution recently derived for California seismicity (Moradpour *et al.* 2014), which has been also found to be in agreement with static stress triggering (Hainzl *et al.* 2014). Although the spatial distribution function has so far been fitted only to California data, Hainzl *et al.* (2014) showed in their electronic material that the distribution is not strongly dependent on the focal mechanism. Thus we use it for all empirical data.

3.2 Parameter estimation

After the identification of a mainshock with magnitude M, we fit the ETAS model to earthquakes occurring in the circular area A with radius D around the mainshock epicentre in the time interval from 1 year before the mainshock until the end of the catalogue. Within this period we exclude time intervals of incompleteness in the catalogue known to exist after mainshocks (Kagan 2004). For that we use the estimated incompleteness function for California, $M_c(M, \Delta t) = M - 4.5 - 0.75 \log_{10}(\Delta t)$, where Δt is the time (in days) after an earthquake with magnitude M (Helmstetter *et al.* 2006). Earthquakes in time periods with $M_c > M_{\min}$ are not considered as target events, but still contribute to the predicted ETAS rate in later time periods. To ensure some statistical significance, we analyse only sequences with mainshock magnitudes $M \geq M_{\min} + 1.5$ and $N \geq 50$ events occurring in the fitting period.

For the N observed earthquakes occurring within the area A in one of the N_k subperiods with complete recordings (see above), we estimate ETAS parameters by maximizing the Log-Likelihood function $\mathcal{LL}$

$$\mathcal{LL} = \sum_{j=1}^{N} \ln(R_A(t_j)) - \sum_{k=1}^{N_k} \int_{t_s(k)}^{t_e(k)} R_A(t)\ \mathrm{d}t \tag{12}$$

where $t_s(k)$ and $t_e(k)$ refer to the start and end times of the kth complete subinterval. We consider the incomplete periods after all $M \geq M_{\min} + 2$ events. The parameters are optimized by the Davidon–Fletcher–Powell optimization algorithm (Press *et al.* 1992). An exception is the tOU function, where a grid-search is performed for the truncation time T, because the sharp truncation precludes the calculation of the derivative. For a given T the other parameters are estimated by the Davidon–Fletcher–Powell optimization algorithm. In the latter case, we allow as T-values only mid-points between the sorted interevent-times of the fitted events to any preceding earthquake in the sequence. Selecting mid-points avoids a systematic overestimation of the $\mathcal{LL}$-value, which has its maximum at one of these interevent-times, without paying the price of predicting a greater total number of earthquakes during the subsequent quiet time interval.

3.3 Model comparison

For model comparison, the number of free parameters has to be taken into account. We use the corrected Akaike information criterion (cAIC), which indicates the best model by the minimum

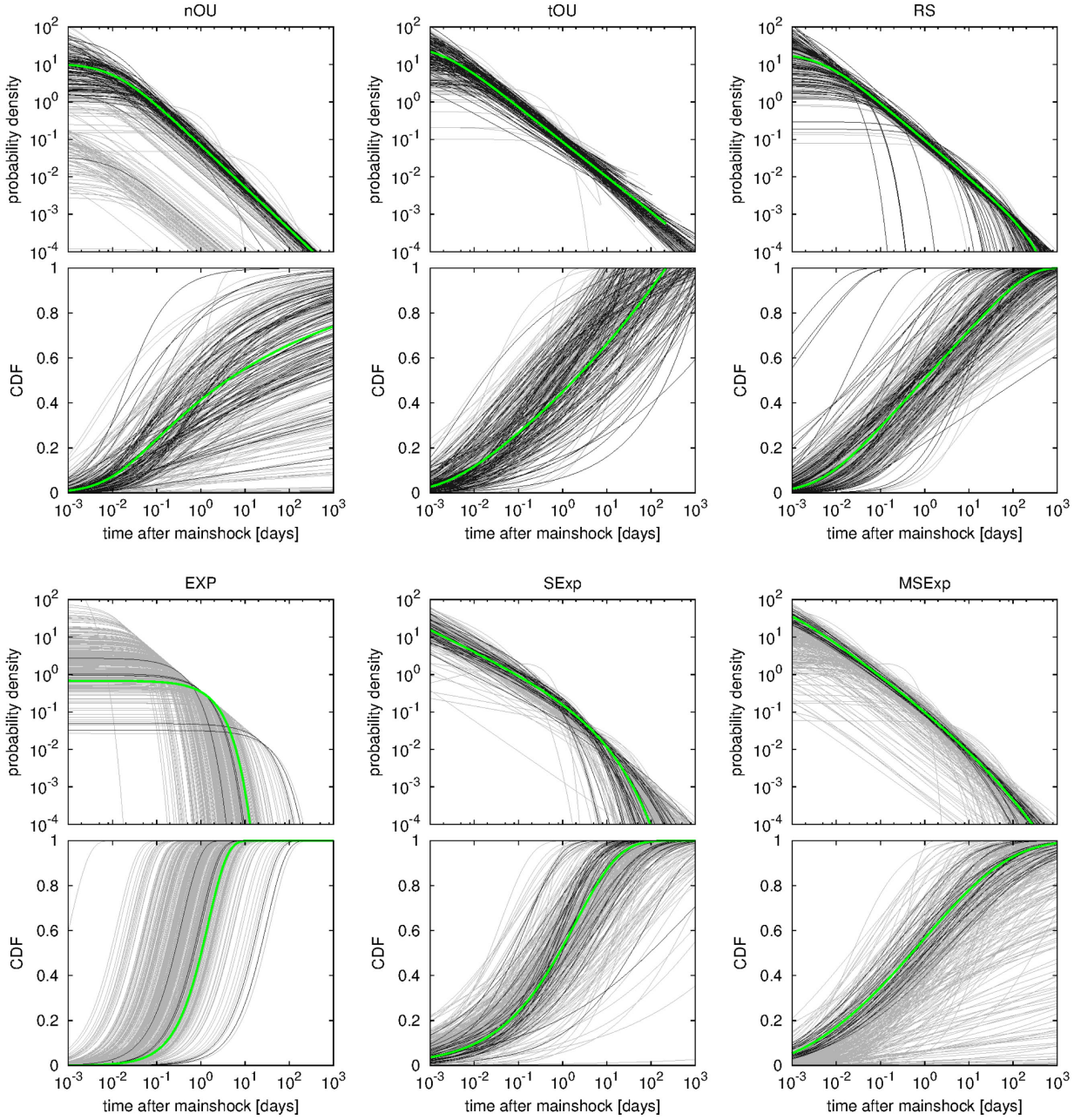

Figure 2. The maximum likelihood estimates of the decay functions for the different analysed sequences. In the case of each decay function, the upper plot refers to the probability density function and the bottom plot shows the corresponding cumulative distribution function. Black lines indicate the result for sequences where the corresponding function provides the best fit based on the cAIC-value, while the grey lines represent the results for the other cases. The green lines illustrate the decay functions with median values provided in Table 1.

cAIC-value (see the Supporting Information for the application of an alternative information criterion). The corrected Akaike information criterion accounts for the sample size and is defined as $\mathrm{cAIC} = 2\,(n + n(n+1)/(N-n-1) - \mathcal{LL})$ with n being the number of free model parameters and $\mathcal{LL}$ being the log-likelihood value defined in eq. (12) (Hurvich & Tsai 1989). The number of free parameters varies for the different models according to $n_{\mathrm{nOU}} = 2$, $n_{\mathrm{tOU}} = 3$, $n_{\mathrm{RS}} = 2$, $n_{\mathrm{Exp}} = 1$, $n_{\mathrm{SExp}} = 2$, and $n_{\mathrm{MSExp}} = 3$. For example, if the true decay follows an unlimited $p > 1$ decay, the nOU function as well as the tOU function with T larger than the catalogue length will lead to identical $\mathcal{LL}$-values, but the nOU function is selected because it has only two instead of three free parameters.

4 RESULTS

The total number of sequences selected according to the defined selection criteria is 326; 146 in the global catalogue, 38 in the SCA-catalogue, 39 in the NCA-catalogue, 78 in Taiwan and 25 in the NZ-catalogue. For each sequence, we calculate the maximum likelihood value and the corresponding model parameters for each decay function. Furthermore, the best model is determined for each sequence by means of the cAIC-values. Fig. 2 shows the maximum likelihood estimates of the decay functions and the cumulative distribution functions for all sequences. For each decay function, the black lines indicate the result for sequences where the corresponding function provides the best fit, while the grey lines represent the

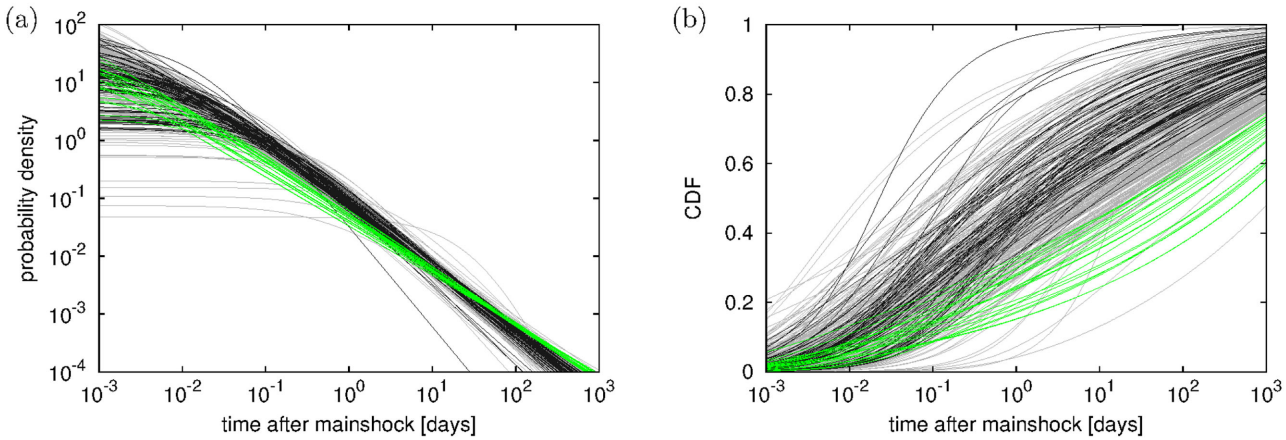

Figure 3. The resulting decay curves for the nOU function shown in Fig. 2, but now renormalized for a finite aftershock duration of T: (a) probability density function and (b) cumulative distribution function. Black lines indicate the result for sequences where the corresponding function provides the best fit. Grey lines represent the results for the other cases. In addition, the green lines provide the result of the cases where the non-normalized OU function with $p \leq 1$ fits best. Note that in this case it is implicitly assumed that a truncation time T exists and exceeds the catalogue lengths. T is set to 100 yr for all curves.

Table 1. Parameters for sequences where the corresponding decay function fits best according to the corrected AIC-criterion. In brackets, the corresponding results for all sequences are provided. All values are in time units of days.

No.	Function		Functional form	Parameter	Median	Mean	Standard deviation
1	nOU	eq. (5)	$\sim(c_1+t)^{-p_1}$	c_1	0.011 (0.008)	0.029 (0.55)	0.04 (7.1)
				p_1	1.12 (1.08)	1.15 (1.46)	0.11 (6.4)
2	tOU	eq. (6)	$\sim(c_2+t)^{-p_2}\ H(T-t)$	c_2	0.002 (0.003)	0.006 (0.25)	0.01 (3.5)
				p_2	0.94 (1.01)	0.95 (1.56)	0.13 (9.3)
				T	218 (226)	991 (864)	2559 (1999)
3	RS	eq. (8)	$\sim \frac{1}{e^{\frac{t}{t_a}}-B}$	B	0.99998 (0.99998)	0.99917 (0.99950)	0.0029 (0.0036)
				t_a	188 (271)	10497 (11977)	89705 (73824)
4	Exp	eq. (9)	$\sim e^{-at}$	a	0.7 (1.9)	0.9 (7.8)	0.90 (51.0)
5	SExp	eq. (10)	$\sim t^{\beta_5-1}e^{-\lambda_5 t^{\beta_5}}$	λ_5	0.75 (0.76)	0.90 (0.89)	0.58 (0.68)
				β_5	0.44 (0.36)	0.45 (0.39)	0.10 (0.12)
6	MSExp	eq. (11)	$\sim(c_6+t)^{\beta_6-1}e^{-\lambda_6(c_6+t)^{\beta_6}}$	c_6	0.0004 (0.002)	0.0005 (0.04)	0.0004 (0.48)
				λ_6	1.01 (1.12)	1.03 (21.4)	0.30 (76.1)
				β_6	0.22 (0.17)	0.22 (0.19)	0.05 (0.16)

results for the other cases. The results are found to scatter quite significantly, however most solutions lie close to each other. Note that in the case of nOU, the separated bottom cluster of curves represent the solutions with p-value close to one, in which case only a small fraction of events are expected to occur in the first 10 d. There is a cross-over of the curves from the bottom cluster with the higher cluster, which is outside the range shown in these graphs. Note that the nOU function for $p > 1$ has identical data fits as the tOU function with any triggering time T longer than the catalogue length $T_{\rm cat}$. However, it has one less parameter. We also tested the non-normalized OU decay function used in eq. (2). For $p > 1$, the function can be normalized and has identical results as the nOU function. The non-normalized function can also be applied for $p \leq 1$ when implicitly assuming that the aftershock duration is finite but longer than the catalogue length, $T > T_{\rm cat}$. This is equivalent to the tOU function with $T > T_{\rm cat}$ but preferred by the cAIC-value for the same $\mathcal{LL}$-values due to the reduced number of parameters. The non-normalized OU decay with $p \leq 1$ is the best of all tested decay functions for 22 out of the 326 sequences (6.7 per cent) according to the cAIC-value. For an arbitrary truncation time of 100 yr (approximately the length of our longest tested catalogue), these 22 curves are shown by green lines in Fig. 3 together with the nOU results re-normalized to the same time period. The two populations of the nOU results found in Fig. 2 (top left) converge in this case. Due to the arbitrariness of setting T, we exclude the non-normalized OU function from the following analysis.

While the application of the nOU function leads to the expectation that for many sequences a significant fraction of direct aftershocks occurs after 1000 d, all other models forecast only a minor number of direct aftershocks after this time. A large percentage of late aftershocks can significantly alter seismic hazard estimations. Table 1 provides the median, mean and standard deviation of the estimated parameters for the case that the corresponding function fits best. The mean values are partly dominated by some outliers, as indicated by a significantly higher mean than median value in the case of c, T, a and t_a. The decay functions based on the median values of its individual parameters are shown as green lines in Fig. 2. Since the median value is chosen separately for each parameter, the green lines do not necessarily agree with any fitted solutions. Nevertheless, the median parameters lead to decay curves centred in the cloud of solutions.

The resulting percentage of sequences in which each function is best fitting is given in Table 2 and shown in Fig. 4(a). The results for all analysed sequences show that the tOU function performs best overall with a success rate of 32.2 per cent, followed by the RS function with 26.1 per cent and the nOU with 22.7 per cent.

Table 2. Percentage of sequences best fitted by the different decay functions, where model parameters are estimated for each individual sequence. The combined result for all sequences as well as the results for each individual catalogue are provided. In brackets, the corresponding results for the fits with median parameter values are given.

Function	all ($N = 326$)	NCA ($N = 39$)	SCA ($N = 38$)	Taiwan ($N = 78$)	NZ ($N = 25$)	ISC-GEM ($N = 146$)
nOU	22.7 (21.8)	7.7 (15.4)	10.5 (18.4)	11.5 (10.3)	24.0 (36.0)	35.6 (28.1)
tOU	32.2 (15.6)	38.4 (23.1)	31.6 (5.3)	35.9 (28.2)	64.0 (8.0)	23.3 (11.0)
RS	26.1 (16.0)	30.8 (25.6)	31.6 (15.8)	29.5 (3.8)	4.0 (16.0)	25.3 (19.8)
EXP	1.8 (0.6)	2.6 (0.0)	0.0 (0.0)	1.3 (0.0)	0.0 (0.0)	0.0 (1.4)
SExp	9.5 (18.1)	7.7 (5.1)	0.0 (0.0)	9.0 (6.4)	8.0 (16.0)	13.0 (32.9)
MSExp	7.7 (27.9)	12.8 (30.8)	26.3 (60.5)	12.8 (51.3)	0.0 (24.0)	2.8 (6.8)

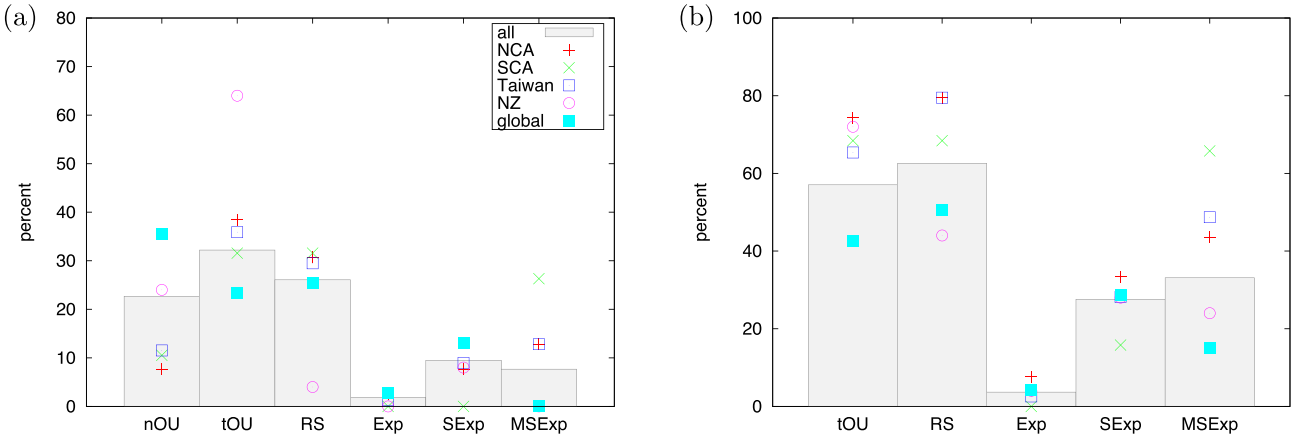

Figure 4. Percentages of sequences which are best fitted by the decay functions noted by their abbreviation on the abscissa: (a) best fit of all tested functions and (b) better fit than the normalized Omori–Utsu (nOU) function.

All exponential functions perform significantly worse. The SExp and MSExp function are best in only 9.5 per cent and 7.7 per cent, respectively, and the Exp function only in 1.8 per cent of the cases. The performance is similar for all catalogues. Altogether, the best function belongs in 81 per cent of the cases to the power law family (nOU, tOU, RS) and only in 19 per cent to the exponential model family (EXP, SExp, MSExp). For each sequence we compare the performance of each decay function with that of the Omori–Utsu decay function (Fig. 4b) to evaluate whether there is evidence for changing the decay function in the ETAS model. We find that the tOU and RS function both outperform the nOU function in the majority of cases. In this comparison the RS function does best with a success rate of 63 per cent of all analysed sequences, while the tOU function has a success rate of 57 per cent. All exponential functions are fitting worse than the nOU function.

The above analysis shows that the Omori–Utsu decay function, with or without truncation, generally fits individual sequences better than the stretched exponential function. However, the fitted parameters vary significantly from sequence to sequence, as discussed above. Therefore, the predictability for future sequences might be lower for the Omori-type functions than for the stretched exponential function. To test this using our existing data sets, we repeat the data fits with all parameters fixed to their median values provided in Table 1. For each sequence, we compare the log-likelihood values and select the best function. The result is provided in Table 2 in brackets and shown in Fig. 5. The MSExp function is better than all other decay functions with a success rate of 28 per cent followed by the nOU function with 22 per cent, and the SExp function with 18 per cent. Here the RS and tOU functions perform best only for approximately 16 per cent of the sequences. Again, the pure exponential fits the worst and only describes one sequence best. In summary, the best function belongs to the power law family (nOU, tOU, RS) for the majority of sequences (53.4 per cent) and to the exponential family (EXP, SExp, MSExp) for 46.6 per cent of the sequences. Comparing the fit quality directly with the nOU function, only the RS function leads to an improved fit in the majority (69 per cent) of the sequences. All other decay functions are found to be better than nOU in less than half of the analysed sequences. The discrepancy between the observed high score for the direct comparison to the nOU function and the only moderate performance of the RS function in the comparison to all decay functions (Fig. 5a) indicates that the exponential truncation of the Omori–Utsu law with fixed $p = 1$, seems to be often better than an unlimited Omori–Utsu decay with higher p-value (nOU), but that other functions are frequently even better for explaining individual sequences. While the nOU function significantly outperforms the two functions SExp and Exp for all catalogues, the MSExp function is found to be better than the nOU function in the majority of the sequences in California and Taiwan. The same holds for the tOU function with the exception of southern California.

5 DISCUSSION

From a physical point of view, the underlying p-value of an infinite power law aftershock decay has to be greater than one in order to lead to a finite energy release. However, ETAS model fits are known to be subject to many problems. This can lead to some biased parameter estimations, e.g. due to anisotropic spatial clustering, data incompleteness, and finite size effects (Hainzl *et al.* 2008; Seif *et al.* 2017). In particular, missing small earthquakes below the completeness magnitude can lead to an underestimation of the p-value (Harte 2015). Thus an estimated value of $p \leq 1$ for the infinite OU decay function does not necessarily reject an infinite decay, but instead be related to a biased p-value estimation while

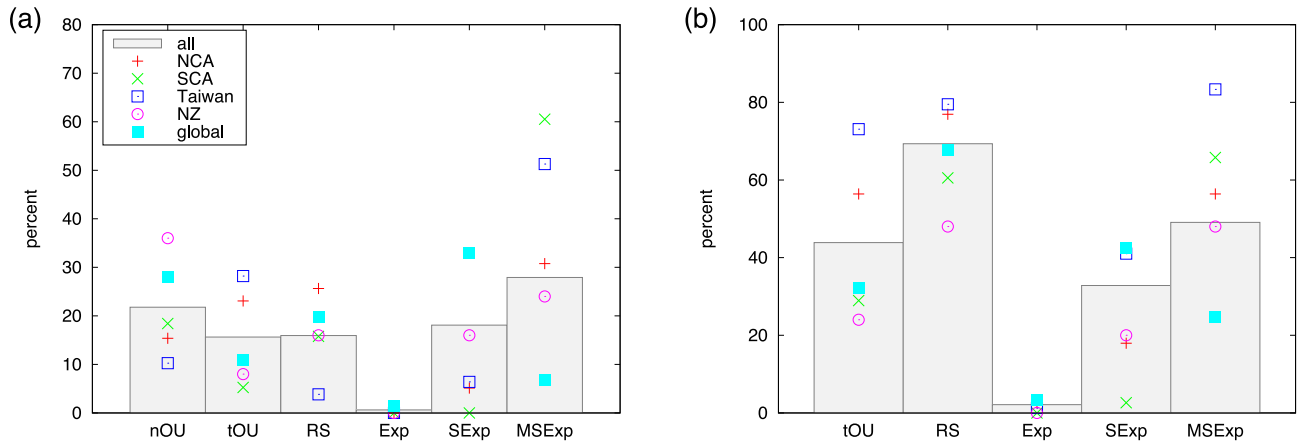

Figure 5. Percentage of sequences, which are best fitted by the decay functions with its median values: (a) best fit of all tested decay functions and (b) better fit than the normalized Omori–Utsu (nOU) function.

the true underlying value is $p > 1$. Our application of the non-normalized OU function shows that approximately 7 per cent of all analysed sequences can be best fitted according to the cAIC-value by an OU decay with $p \leq 1$ and no truncation (i.e. T larger than the catalogue length $T_{\rm cat}$). The consideration of these cases only slightly affects the estimated median parameters, $c = 0.010$ d and $p = 1.09$, compared to $c = 0.011$ d, $p = 1.12$ for the nOU function alone (see Table 1). However, the non-normalized OU function with $p \leq 1$ formally belongs to the tOU class of functions with $T > T_{\rm cat}$. For the purpose of seismicity forecasts based on the ETAS model, a rather arbitrary setting of a T-value with $T > T_{\rm cat}$ is necessary in these cases to avoid inconsistent simulations in the long run.

Our results favour a power law decay for aftershock rate. The total percentage of sequences fitted best by one of the functions OU, nOU, tOU, or RS is 81 per cent whether the non-normalized OU decay is considered as a separate function or not. This result is consistent with earlier works comparing the fits of power law and exponential decay functions for the overall decay of the aftershock activity ignoring ETAS-type secondary aftershock triggering (Kisslinger 1993; Gasperini & Lolli 2009). The observation that the percentage of sequences that are best described with a power law decay drops from more than 80 per cent to around 53 per cent with fixed parameters is an indication that the model parameters are more consistent for the exponential decays. In particular, tOU, which is the best performing decay function in all but one catalogue with fitted parameters, scores much less with fixed parameters. This can be explained by the large variation of the triggering time that ranges from less than 10 d to around 30 yr (Hainzl *et al.* 2016), i.e. it is close to the duration of some of the earthquake catalogues, and thus practically the same as being infinite. When the actual triggering time for a sequence is shorter than the fixed median value ($T = 218$ d), tOU forecasts too many late aftershocks relative to nOU because the fixed median p-value ($p_2 = 0.94$) is smaller than for the nOU function ($p_1 = 1.12$). On the contrary, when the actual triggering time for a sequence is larger than the fixed value, nOU with an infinite triggering time also tends to perform better than tOU.

Out of all tested functions, the RS function is the only function that systematically performs better than the nOU function for fixed parameters. This suggests that it could be a potential candidate function to replace the unlimited OU decay in ETAS-based forecasts. In the rate-and-state model the duration of the triggering is related to the stressing rate (Dieterich 1994). Aftershock duration has been found to correlate inversely with fault loading rate (Stein & Liu 2009). Therefore, including strain-rate measurements as a proxy for fault loading may account for the large variability in triggering time and might even improve the fits.

6 CONCLUSIONS

The stretched exponential function was recently found to fit the overall aftershock decay better than a power law (Mignan 2015). In comparison to a power law, the stretched exponential implies a faster decay of aftershock rates at longer times. Confirming this faster decay in an ETAS framework, which can be used to model seismicity for earthquake forecasting, would have important consequences for time-varying seismic hazard assessments. We have implemented the stretched exponential function as well as five other decay functions in the framework of the ETAS model, which includes background and secondary triggering and is known to describe earthquake clustering well. Analysing several hundreds of mainshock–aftershock sequences recorded in global and regional catalogues, we find that a power law decay fits a larger percentage of sequences better than the stretched exponential function. In particular, the truncated Omori–Utsu law which has been studied by Hainzl *et al.* (2016) fits most sequences the best, followed by the physics-based rate-state response function and the unlimited Omori–Utsu law. The stretched exponential function fits only a small fraction of the analysed sequences best. However, its parameters are found to be more stable. This leads to a comparable performance as the power law decay in fits with universal parameters. Nevertheless, the difference is only minor and does not justify the replacement of the Omori–Utsu decay function by the stretched exponential function in ETAS-type seismicity models. In contrast, the rate-state function is found to be superior to the unlimited Omori–Utsu decay function in the majority of cases. It represents a Omori–Utsu decay with $p = 1$ and an exponential roll-off at a time which should be theoretically inversely related to the stressing rate. Including other observables in aftershock modelling such as strain-rate as proxy for fault loading rate may in future improve aftershock modelling and lead to better time-varying hazard estimates.

ACKNOWLEDGEMENTS

We are grateful to two anonymous reviewers for their very helpful recommendations. We thank our colleague David Rhoades for critically reading the final version of the manuscript. The New Zealand

Earthquake Commission Time Varying Hazard Programme provided financial support through projects 14/TV682 and 14/TV695.

REFERENCES

Bannister, S. & Gledhill, K., 2012. Evolution of the 2010–2012 Canterbury earthquake sequence, *N.Z.J. Geol. Geophys.,* **55,** 295–304.

Cocco, M., Hainzl, S., Catalli, F., Enescu, B., Lombardi, A.M. & Woessner, J., 2010. Sensitivity study of forecasted aftershock seismicity based on Coulomb stress calculation and rate- and state-dependent frictional response, *J. geophys. Res.,* **115,** B05307, doi:10.1029/2009JB006838.

Dieterich, J.H., 1994. A constitutive law for rate of earthquake production and its application to earthquake clustering, *J. geophys. Res.,* **99,** 2601–2618.

Dieterich, J.H., Cayol, V. & Okubo, P., 2000. The use of earthquake rate changes as a stress meter at Kilauea volcano, *Nature,* **408,** 457–460.

Gasperini, P. & Lolli, B., 2009. An empirical comparison among aftershock decay models, *Phys. Earth planet. Inter.,* **175,** 183–193.

Gerstenberger, M.C. & Rhoades, D.A., 2010. New Zealand earthquake forecast testing centre, *Pure appl. Geophys.,* **167**(8/9), 877–892.

Godano, C. & Tramelli, A., 2016. How Long is an Aftershock Sequence?, *Pure appl. Geophys.,* **173,** 2295–2304.

Gross, S.J. & Kisslinger, C., 1994. Test of models of aftershock rate decay, *Bull. seism. Soc. Am.,* **84,** 1571–1579.

Hainzl, S., 2016a. Rate-dependent incompleteness of earthquake catalogs, *Seismol. Res. Lett.,* **87,** 337–344.

Hainzl, S., 2016b. Apparent triggering function of aftershocks resulting from rate-dependent incompleteness of earthquake catalogs, *J. geophys. Res.,* **121,** 6499–6509.

Hainzl, S. & Christophersen, A., 2016. Comment on 'Revisiting the 1894 Omori Aftershock Dataset with the Stretched Exponential Function' by A. Mignan, *Seismol. Res. Lett.,* **87,** 1130–1133.

Hainzl, S., Christophersen, A. & Enescu, B., 2008. Impact of earthquake rupture extensions on parameter estimations of point-process models, *Bull. seism. Soc. Am.,* **98,** 2066–2072.

Hainzl, S., Moradpour, J. & Davidsen, J., 2014. Static stress triggering explains the empirical aftershock distance decay, *Geophys. Res. Lett.,* **41,** doi:10.1002/2014GL061975.

Hainzl, S., Christophersen, A., Rhoades, D. & Harte, D., 2016. Statistical estimation of the duration of aftershock sequences, *Geophys. J. Int.,* **205**(2), 1180–1189.

Harte, D.S., 2013. Bias in fitting the ETAS model: a case study based on New Zealand seismicity, *Geophys. J. Int.,* **192,** 390–412.

Harte, D.S., 2015. Model parameter estimation bias induced by earthquake magnitude cut-off, *Geophys. J. Int.,* **204,** 1266–1287.

Hauksson, E., Yang, W. & Shearer, P., 2012. Waveform relocated earthquake catalog for southern California (1981 to 2011), *Bull. seism. Soc. Am.,* **102,** 2239–2244.

Helmstetter, A., Kagan, Y.Y. & Jackson, D.D., 2006. Comparison of short-term and time-independent earthquake forecast models for southern California, *Bull. seism. Soc. Am.,* **96**(1), 90–106.

Hurvich, C.M. & Tsai, C.-L., 1989. Regression and time series model selection in small samples, *Biometrika,* **76,** 297–307.

Kagan, Y.Y., 2004. Short-term properties of earthquake catalogs and models of earthquake source, *Bull. seism. Soc. Am.,* **94**(4), 1207–1228.

Kisslinger, C., 1993. The stretched exponential function as an alternative model for aftershock decay rate, *J. geophys. Res.,* **98,** 1913–1922.

Lee, W.H.K. & Stewart, S.W., 1989. Large-scale processing and analysis of digital waveform data from the USGS Central California microearthquake network, in *Observatory Seismology: An Anniversary Symposium on the Occasion of the Centennial of the University of California at Berkeley Seismographic Stations,* p. 86, ed. Litehiser, J.J., University of California Press.

Lolli, B. & Gasperini, P., 2006. Comparing different models of aftershock rate decay: the role of catalog incompleteness in the first times after main shock, *Tectonophysics,* **423,** 43–59.

Lolli, B., Boschi, E. & Gasperini., P., 2009. A comparative analysis of different models of aftershock rate decay by maximum likelihood estimation of simulated sequences, *J. geophys. Res.,* **114,** B01305, doi:10.1029/2008JB005614.

Mignan, A., 2015. Modeling aftershocks as a stretched exponential relaxation, *Geophys. Res. Lett.,* **42,** 9726–9732.

Mignan, A., 2016. Reply to "Comment on 'Revisiting the 1894 Omori Aftershock Dataset with the Stretched Exponential Function' by A. Mignan" by S. Hainzl and A. Christophersen, *Seismol. Res. Lett.,* **87,** doi:10.1785/0220160110.

Moradpour, J., Hainzl, S. & Davidsen, J., 2014. Non-trivial decay of aftershock density with distance in southern California, *J. geophys. Res.,* doi:10.1002/2014JB010940.

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

Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P., 1992. *Numerical Recipes in C: The Art of Scientific Computing,* 2nd edn, Cambridge Univ. Press.

Seif, S., Mignan, A., Zechar, J.D., Werner, M.J. & Wiemer, S., 2017. Estimating ETAS: the effects of truncation, missing data, and model assumptions, *J. geophys. Res.,* **122,** 449–469.

Stein, S. & Liu, M., 2009. Long aftershock sequences within continents and implications for earthquake hazard assessment, *Nature,* **462,** 87–89.

Storchak, D.A., Di Giacomo, D., Bondar, I., Engdahl, E. R., Harris, J., Lee, W. H. K., Villasenor, A. & Bormann, P., 2013. Public release of the ISC-GEM global instrumental earthquake catalogue (1900–2009), *Seismol. Res. Lett.,* **84**(5), 810–815.

Tahir, M., Grasso, J.-R. & Amorèse, D., 2012. The largest aftershock: how strong, how far away, how delayed?, *Geophys. Res. Lett.,* **39,** L04301, doi:10.1029/2011GL050604.

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

Waldhauser, F. & Schaff, D.P., 2008. Large-scale relocation of two decades of Northern California seismicity using cross-correlation and double-difference methods, *J. geophys. Res.,* **113,** B08311, doi:10.1029/2007JB005479.

Wells, D.L. & Coppersmith, K.J., 1994. New empirical relationships among magnitude, rupture length, rupture width, rupture area, and surface displacement, *Bull. seism. Soc. Am.,* **84,** 974–1002.

Wu, Y.-M., Chang, C.-H., Zhao, L., Teng, T.-L. & Nakamura, M., 2008. A comprehensive relocation of earthquakes in Taiwan from 1991 to 2005, *Bull. seism. Soc. Am.,* **98,** 1471–1481.

Zakharova, O., Hainzl, S., Lange, D. & 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 appl. Geophys.,* **174,** 77–102.

Zhuang, J., Werner, M.J. & Harte, D.S., 2013. Stability of earthquake clustering models: criticality and branching ratios, *Phys. Rev. E,* **88,** 062109, doi:10.1103/PhysRevE.88.062109.

SUPPORTING INFORMATION

Supplementary data are available at *GJI* online.

supplement.pdf

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.