

Productivity within the epidemic-type seismicity model

G. Molchan,¹ E. Varini² and A. Peresan³

¹*Institute of Earthquake Prediction Theory and Mathematical Geophysics, Russian Academy of Science, 84/32 Profsoyuznaya st., 117997 Moscow, Russia*

²*Institute for Applied Mathematics and Information Technologies ‘Enrico Magenes’, National Research Council, via Corti 12, 20133 Milano, Italy*

³*National Institute of Oceanography and Applied Geophysics - OGS, Seismological Research Centre, via Treviso 55, 33100 Udine, Italy.*

E-mail: aperesan@inogs.it

Received 2022 June 24; in original form 2022 March 10

SUMMARY

The productivity of a magnitude m event can be characterized in term of triggered events of magnitude above $m - \Delta$: it is the number of direct ‘descendants’ ν_Δ and the number of all ‘descendants’ V_Δ . There is evidence in favour of the discrete exponential distribution for both ν_Δ and V_Δ with a dominant initial magnitude m (the case of aftershock cluster). We consider the general Epidemic Type Aftershock Sequence model adapted to any distribution of ν_Δ . Our first result shows that models with branching aftershock structure do not allow for the coincidence of distribution types of ν_Δ and V_Δ (say, the discrete exponential, as in the scientific literature). The second problem is related to the tail behaviour of the V_Δ distribution. We show the fundamental difference in tail behaviour of the V_Δ -distributions for general-type clusters and clusters with a dominant initial magnitude: the tail is heavy in the former case and light in the latter. The real data demonstrate the possibilities of this kind. This result provides theoretical and practical constraints for distributional analysis of V_Δ .

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

1 INTRODUCTION

Earthquakes clustering, aftershocks occurrence in particular, is a prominent feature in observed seismicity. Different models have been proposed to describe the features of triggered events, especially in the case of aftershocks clusters with an initial dominant magnitude earthquake; yet, several theoretical aspects still need to be explored. In this study, we investigate the distributions of the number of triggered events as described by a branching process, which might provide useful practical constraints for empirical analysis of earthquake data.

Let V_Δ denote the number of seismic events of magnitude $m \geq m_0 - \Delta$ caused by an earthquake of magnitude $m_0 \geq m_c + \Delta$, where m_c is the lower magnitude threshold and Δ a positive default value. V_Δ is further referred to as the *total Δ -productivity*.

Considering an earthquake cluster as the result of a branching process (Zaliapin *et al.* 2008), we can also deal with the directly triggered events of m_0 , that is the first generation events of an m_0 -event. The number ν_Δ of such events is in the same magnitude range $m \geq m_0 - \Delta$ and is later named *Δ -productivity* of m_0 .

The distributions of quantities V_Δ and ν_Δ have recently become the object of active analysis in the literature (Baranov & Shebalin 2019; Shebalin *et al.* 2020; Shebalin & Baranov 2021; Baranov *et al.* 2022). According to these publications, the ν_Δ -distribution $f(n) = P(\nu_\Delta = n)$ is *universal in shape*, namely, it is *geometric* (or *discrete exponential*):

$$f_G(n) = p^n (1 - p) \quad n = 0, 1, \dots, \quad (1)$$

where p does not depend on the magnitude of triggering event. This result is relevant in statistical seismology because the popular Epidemic Type Aftershock Sequence (ETAS) model (Ogata 1998) is based on the Poisson distribution

$$f_P(n) = e^{-\lambda} \lambda^n / n! \quad n = 0, 1, \dots \quad (2)$$

Similarly, the geometric distribution (eq. 1) was adopted for statistics V_Δ , provided that it refers to clusters with a dominant initial magnitude m_0 , that is, to aftershocks. For the first time this hypothesis appeared in the work by Soloviev & Solovieva (1962); it was based on relatively scarce data (1954–1961) for the Pacific Belt ($m_0 \approx 7$; $\Delta = 2$) and for the Kamchatka and Kuril Islands region ($m_0 \approx 6$; $\Delta = 2$). Shebalin *et al.* (2018) expanded the analysis and confirmed the hypothesis using 850 aftershock sequences with main shocks $m_0 \geq 6.5$, $\Delta = 2$ from the ANSS (1975–2018) catalogue. In the same time Kagan (2010), using aftershocks $m \geq 4.7$ from PDE (1977–2007) catalogue for $m_0 = 7.1-7.2$, found that the empirical distribution of V_Δ is bimodal with the dominant peak at $n = 0$.

Zaliapin & Ben-Zion (2016) provided a thorough analysis of earthquake clustering and showed its spatial dependence and connection with heat flow production. They showed that, on the global scale, the total Δ -productivity distribution (for clusters with $\max m \geq 6$ and $\Delta = 2$) looks like a power-law for locations with high heat flow ($H > 0.2 \text{ W m}^{-2}$), and as a log-normal distribution otherwise. These conclusions are qualitative because they are based on log–log plots. Nevertheless, the data for the regions of high heat flow differ significantly from the geometric hypothesis (eq. 1) for aftershocks. The same should be expected for any cluster with $\max m \leq 5$. In this case, according to Zaliapin & Ben-Zion (2016), the cluster size distribution in zones with both high and low heat flow is a power type distribution, $f(n) \propto n^{-\kappa-1}$, $\kappa \approx 2.5$.

These observations rise the following questions: (i) is it possible to consider the geometric distribution of ν_Δ and V_Δ (for aftershocks) as laws of seismicity and (ii) how to explain the observed difference in the tail behaviour of the V_Δ distribution for arbitrary clusters and aftershocks. A rigorous analysis of these issues is possible within the framework of a model adequately related to the concepts of ν_Δ and V_Δ . A model suitable for this purpose may be the general ETAS model, namely the ETAS model with an arbitrary (not only Poisson) distribution of direct aftershocks associated to any event (Saichev *et al.* 2005). The purpose of this note is twofold: first, to show the incompatibility of the hypotheses about the geometric distribution of ν_Δ and V_Δ (for aftershocks) within the general ETAS model; and second, to prove the fundamental difference between size distributions for general type clusters and aftershock clusters.

2 ETAS MODELS

Let us consider the seismic events $x = (t, g, m)$ occurred in the time–location–magnitude space $\mathcal{S} = (t \geq 0, g \in \mathbf{R}^2, m \geq 0)$. The lower magnitude threshold m_c is taken here as the zero reference point without loss of generality. According to the branching structure of the ETAS model, an earthquake cluster can be represented as follows (Saichev *et al.* 2005). The initial event $x_0 = (t_0, g_0, m_0)$ generates a random number $\nu(m_0)$ of events $\{x_i: i = 1, \dots, \nu(m_0)\}$, where $\nu(m_0)$ follows a special distribution F with an average $\lambda(m_0) = E\nu(m_0)$. The offspring events are distributed in \mathcal{S} as independent variables with the probability density

$$f(x | x_0) = f_1(m)f_2(t - t_0)f_3(g - g_0 | m_0) \quad (3)$$

Each new event independently generates its own family of events following the generation law described above, and so on. If the mean of the direct triggered events is $\Delta_1 = E\lambda(m) < 1$, then the mathematical expectation of the cluster size will be finite.

The independence of the magnitudes of direct ‘descendants’ between themselves and from the parent event x_0 is an important property of the model. This allows us to consider the distribution $f_1(m)$ as a frequency–magnitude law for the cluster as a whole. This also is important for understanding the choice of the distribution models for ν_Δ and V_Δ .

The components of the *regular ETAS model* are specified as follows:

$\nu(m)$ -distribution	- Poisson;
$\lambda(m) = \lambda e^{\alpha m}$, $m \geq 0$	- a copy of the Utsu law for aftershocks;
$f_1(m) = \beta e^{-\beta m}$, $m \geq 0$, $\beta > \alpha$	- the Gutenberg-Richter law,
$f_2(t) = (p - 1)(t/c + 1)^{-p}/c$	- the time scattering (a copy of the Omori law for aftershocks);
$f_3(g m) = (q - 1) [g/\psi_m ^2 + 1]^{-q} / (\pi \psi_m^2)$	- one of the laws of scattering in space;
$\psi_m = d e^{\gamma m/2}$	- the scale parameter.

Note: The exponents (α , β) in the laws of Utsu and Gutenberg–Richter are more convenient (Saichev *et al.* 2005) to use with the base ‘ e ’ instead of the usual 10.

A theoretical analysis of various aspects of the general ETAS model with the regular components $\lambda(m)$ and $f_1(m)$ is provided in the works of Saichev and co-authors (see Saichev & Sornette 2017, for further details). In particular, they proved the power law behaviour of the cluster size distribution $f_{cl}(n) \propto n^{-1-\kappa}$ with $\kappa \in [0.5, 1]$ (Saichev *et al.* 2005; Baró 2020). Note that, according to Zaliapin & Ben-Zion (2016), $\kappa \in [1, 2.5]$ for real data.

In the general case we need to specify the family of $\nu(m)$ distributions: $F = \{f(n|m)\}$. This choice will be reflected in the designation of the model as ETAS(F). In particular, we will set $F = P$ for Poisson, and $F = G$ for Geometric distribution. The component $\lambda(m) = E\nu(m)$ will be common to all models. The relation of $\lambda(m)$ with the parameter $p(m)$ of the geometric distribution is determined by the equality $\lambda = p/(1 - p)$, that is

$$p(m) = \lambda(m)/(1 + \lambda(m)). \quad (4)$$

Recall that the regular ETAS model can be defined in terms of the conditional intensity of events (Hawkes & Oakes 1974):

$$\lambda^*(x | \{x_i : t_i < t\}) = \mu(g)f_1(m) + \sum_i \lambda(m_i)f(x | x_i) \quad (5)$$

where $x = (t, g, m)$ and $\mu(g)$ is the rate of background seismicity. The general definition is broader because it allows for any $\nu(m)$ distribution.

To provide the G -distribution of Δ -productivity for any Δ , it is natural to consider the G -distribution of $\nu(m)$. It follows from the following seemingly obvious but useful statement.

Statement. Let the distribution type F of $\nu(m_0)$ be $F = G$ or P . Then $\nu_\Delta(m)$ -distribution will have the same type for $m_0 \geq \Delta$. In both cases the statistics $\nu_\Delta(m_0)$ and $\nu(m_0)$ differ only in average values:

$$\lambda_\Delta(m_0) = \lambda(m_0)\bar{F}_1(m_0 - \Delta)/\bar{F}_1(\Delta) \tag{6}$$

where $\bar{F}_1(m) = \int_m^\infty f_1(u)du$, $\bar{F}_1(0) = 1$. Let $a = f_1(0) > 0$ and $\lambda_\Delta(m_0)$ be independent of m_0 in the area $0 < \Delta \leq m_0 < M$. Then $\lambda(m) = ce^{am}$ and $\bar{F}_1(m) = e^{-am}$, $m \in (0, M)$.

The proof of the statement and all subsequent ones are contained in the Appendix.

Remark 1. The conditions for the independence of the $\nu_\Delta(m_0)$ -distribution from m_0 are very strict, and therefore, in contrast with Shebalin *et al.* (2020), this property should be unstable.

Remark 2. The Random Thinning (RT) property. Statement 1 allows the following generalization. Consider the $RT(\pi)$ operation of a random cluster thinning: $RT(\pi)$ saves each cluster event independently with probability π , and deletes it with probability $1 - \pi$; the transformed cluster size V is $RT(\pi)V$. Given the magnitude independence in the ETAS(F) model, the productivity ν_Δ has the same distribution as $RT(\pi)\nu$ provided that $\pi = P(m \geq m_0 - \Delta)$.

In the case $F = (P \text{ or } G)$, $E[RT(\pi)\nu] = \pi E\nu$ and the distribution type of ν and $RT(\pi)\nu$ is identical. We will call these two properties of the discrete distribution the RT property. We can formalize the RT property as follows.

Let V_λ and $V_\lambda(\pi) = RT(\pi)V_\lambda$ be the cluster sizes, where $\lambda = EV_\lambda$. If $Ez^{V_\lambda} = \varphi(z, \lambda)$, then by eq. (A1) from Appendix, $Ez^{V_\lambda(\pi)} = \varphi(1 - \pi + \pi z, \lambda)$, $EV_\lambda(\pi) = \lambda\pi$. The RT property of V_λ implies

$$\varphi(1 - \pi + \pi z, \lambda) = \varphi(z, \lambda\pi), \quad \lambda < \lambda_0 \leq \infty$$

Let for simplicity $\lambda_0 \geq 1$. Setting $\pi = \Lambda/n < 1$, $\lambda = n$ and $\psi^{(n)}(z) = \varphi(1 + z/n, n)$, we get:

$$\psi^{(n)}(\Lambda(z - 1)) = \varphi(z, \Lambda), \quad \Lambda \leq n < \lambda_0. \tag{7}$$

Due to the analyticity of $\varphi(z, \lambda)$ in the disk $|z| < 1$, $\psi^{(n)}(z)$ is analytical in $K_n = \{z: |1 + z/n| < 1\}$. Moreover, (7) implies that $\psi^{(n)}(z)$ and $\psi^{(n+1)}(z)$ are equal in $K_n \cap K_{n+1} = K_n$. Hence $\psi^{(n)}(z)$ is an analytical continuation of $\psi^{(1)}(z) = \varphi(1 + z, 1)$. The function $\psi(z) = \varphi(1 + z, 1)$ specifies the type of F in the ETAS(F) model for which Statement 1 remains true. For example, $\psi(z) = e^z$ and $\psi(z) = (1 - z)^{-1}$ correspond to the types of Poisson and Geometric distributions, respectively.

Examples of the RT property.

(i) *The Negative Binomial Distribution (NBD) for which $\varphi(z, \lambda) = (1 - \lambda(z - 1))^{-\tau}$, $\tau > 0$, holds the RT property.* In the case $\tau = 1$ it corresponds to the Geometric distribution.

(ii) *Combining N independent clusters with RT property also preserves this property.* Formally this follows from eq. (7) and the relation for the cluster characteristics $\varphi_i(z, \lambda_i)$:

$$\varphi(z, \lambda) = \prod_i \varphi_i(z, \lambda_i) = \prod_i \psi_i(p_i \lambda(z - 1)), \quad p_i \lambda = \lambda_i \quad \text{and} \quad \sum p_i = 1 \tag{8}$$

The NBD model was used by Kagan (2014) to describe the distribution of annual earthquake numbers for the CIT catalogue 1932–2001, $m \geq 5.0$. Below we will show that the RT property, which NBD possesses, is poorly compatible with the branching cluster structure. However, in the Kagan’s case, the statistics of events occurring worldwide in one year consists mainly of independent weakly grouped events. Such case is closer to the model (eq. 8) with components: $\varphi_i(z, \lambda_i) = 1 + \lambda_i(z - 1)$, $\lambda_i < 1$.

(iii) *A cluster selected randomly with probabilities $\{\rho_i > 0\}$ from N clusters with RT property retains RT property.* In this case

$$\varphi(z, \lambda) = \sum_i \varphi_i(z, \lambda_i)\rho_i = \sum_i \psi_i(p_i \lambda(z - 1))\rho_i, \quad \lambda p_i = \lambda_i \quad \text{and} \quad \sum_i \lambda_i \rho_i = \lambda. \tag{9}$$

However, if the cluster sizes distributions are of the same type, that is $\varphi_i(z, \lambda_i) = \psi(\lambda_i(z - 1))$, then the random cluster usually loses this type provided that $\lambda_i \neq \text{const}$.

For example, assuming a geometric distribution for both the size of a random cluster and the sizes of its components, (9) gives $(1 - w)^{-1} = \sum_i (1 - w p_i)^{-1} \rho_i$, $w = \lambda(z - 1)$, which is possible only when $p_i = 1$. This is also true in general case if $\varphi(z, \lambda_i)$ is N -smooth at $z = 1$. The considered case is typical for the observed clusters taken as one statistical population. Consequently, empirical distributions of ν_Δ or V_Δ obtained over large territories generally do not represent the true types of their distributions. This circumstance makes it difficult to prove the universality of the type of $\nu_\Delta(m_0)$ distribution.

Remark 3. ETAS* model. To get the geometric distribution of ν_Δ in framework of the ETAS model, Shebalin *et al.* (2020) used randomization of the parameter $\lambda(m_0)$. They used the fact that a Poisson distribution with an exponentially distributed random parameter is equivalent to a geometric distribution. The resulting model was named ETAS*. Its analogue is the ETAS(G) model in which the average productivity does not depend on the initial magnitude, namely $\lambda(m_0) = \lambda$. Because of (6), in this case $\lambda_\Delta(m_0)$ becomes dependent on m_0 for any $f_1(m) > 0$.

The analogue of the ETAS* model may be interesting due to its connection with the classical branching Galton–Watson model (Harris 1963). Both models are identical if we omit the space–time component in the model. This gives us an exact formula for the distribution of the

total productivity of any initial event (cluster size not including the initial event):

$$P(V = n) = \Gamma(1/2 + n) / \Gamma(2 + n) \times (4pq)^n (q/\sqrt{\pi}), \quad q = 1 - p, \tag{10}$$

where $\Gamma(x)$ is the Gamma function, and $p < 1/2$ is the parameter of the geometric distribution (Harris 1963). Formula (10) remains valid in the critical regime when $\lambda = p/(1 - p) = 1$ or $p = 1/2$. The distribution (eq. 10) has the following asymptotic behaviour:

$$P(V = n) \propto n^{-3/2} \exp(-C_1 n), \quad n \gg 1 \tag{11}$$

where $C_1 = -\ln(4pq) \geq 0$. Such asymptotic with the factor $n^{-3/2}$ is typical for the Galton–Watson models (Harris 1963) and, consequently, for any ETAS(F) model with $\lambda(m) = \text{const}$.

The generating function of the distribution (eq. 10) is

$$Ez^V = (1 + 2\Lambda - \sqrt{1 - 4(1 + \Lambda)\Lambda(z - 1)}) / (2\Lambda z), \tag{12}$$

where $EV = \Lambda = \lambda/(1 - \lambda) = p/(1 - 2p)$ (see eq. 4). It is obvious that eq. (12) cannot be represented as eq. (7). Therefore, the analogue of ETAS* model has the geometric distribution for the Δ -productivity ν_Δ , but does not have any RT property for the total Δ -productivity V_Δ .

3 THE TOTAL Δ -PRODUCTIVITY

We will consider the ETAS(F) model in a subcritical regime: $\Lambda_1 = E\lambda(m) < 1$. $Ev(m) = \lambda(m)$ is a non-decreasing, finite, positive function; the other components, namely $f_1(m)$, $f_2(t)$ and $f_3(g|m)$, are free from additional restrictions.

As above $m_0 > 0$ is the magnitude of an initial cluster event and $V_\Delta(m_0)$ is the total Δ -productivity of m_0 . Any initial event can be *fixed* or *random*, namely it has the $f_1(m)$ distribution. A further distinction concerns the *dominant* initial events, that is those initial events having maximum magnitude within their cluster. The case of a random m_0 is important because identifying the initial cluster event is an unstable practical procedure.

Statement 2. Consider the specified ETAS(F) model. Assume that all moments of F -distribution are finite and F has the RT property.

(i) Let $V_\Delta(m_0)$, $m_0 > \Delta > 0$, be the size of the cluster with any condition C : the initial magnitude m_0 is fixed or random. Then the number N_{V_Δ} of finite moments of $V_\Delta(m_0)$ is independent of C . Moreover,

$$N_{V_\Delta} = N_\lambda := \max \left\{ n : I_n = \int_0^\infty \lambda^n(m) f_1(m) dm < \infty \right\}.$$

(ii) Let F be Poisson or Geometric distribution and m_0 is fixed and dominant. Then $N_{V_\Delta} = \infty$ regardless of the N_λ -value. This is also true for the random dominant m_0 -event if additionally

$$\lambda(m) \int_0^\Delta f_1(m - u) du \leq \text{const} < \infty. \tag{13}$$

Remark 4. Statement 2 allows us to judge on the tail attenuation of the V_Δ distribution because, by the Chebyshev–Markov inequality, $P(\xi \geq u) \leq E\xi^n / u^n$ for any random variable $\xi \geq 0$ and $u > 0$. Distributions having all moments will be called *distributions with light tails*. Otherwise, we will talk about *heavy tails*.

Let’s consider clusters with fixed or random initial magnitude. The conditions of the Statement 2 are easily verified for the regular ETAS components: $\lambda(m) = \lambda e^{\alpha m}$, $\alpha > 0$ and $f_1(m) = \beta e^{-\beta m}$, $\beta > \alpha$. In this case we have $N_\lambda < \beta/\alpha$. This means that the total Δ -productivity distribution has only a heavy tail for small β/α . The data of Zaliapin & Ben-Zion (2016) demonstrate the possibilities of this kind. In the considered case Saichev *et al.* (2005) give more information on the tail behaviour of V_Δ distribution: if β/α is large, then the tail is subexponential and looks like eq. (11).

The case $\alpha = 0$ corresponds to the analogue of the ETAS* model; here $N_\lambda = \infty$ in accordance with eq. (10).

Given $\alpha \leq \beta$, condition (13) is satisfied. This guarantees the light tails in the distribution of the total Δ -productivity of main shocks. All the observations mentioned above support this fact.

Remark 5. The difference in tail behaviour of V_Δ statistics can be partially explained as follows. The total Δ -productivity of the main shock with a *fixed* magnitude m_0 is based on events in the finite range of magnitudes $(0, m_0)$. In this case, the frequency–magnitude law $f_1(m)$ is replaced by its truncated counterpart in the interval $(0, m_0)$, where all moments of $\lambda(m)$ are finite. Since $N_{V_\Delta} = N_\lambda$, the tail of the V_Δ distribution becomes typical for the case $N_\lambda = \infty$ that is quasi exponential. The similar effect for a random dominant magnitude is not so obvious and is a non-trivial analytical fact.

Statement 3. (1) Let m_0 be a dominant initial magnitude in the ETAS(G) cluster and let $f_1(m)$ be strictly positive. Then the statistics $V_\Delta(m_0)$ can have a geometric distribution for no more than one value $\Delta \in (0, m_0)$.

(2) Let ν and V be the sizes of clusters of direct descendants and all descendants, respectively, in the ETAS(F) cluster with the random initial event. Assume that both distributions of ν and V have the RT property, as well as $Ev^2 < \infty$ and $\int \lambda^2(m) f_1(m) dm < \infty$. Then the distribution types of ν and V cannot be the same.

Remark 6. According to Statement 3 we can say that, in the framework of generalized ETAS model, the geometric laws for ν_Δ and V_Δ , including the case of aftershocks, are incompatible. In the ETAS(G) model case with $\lambda(m) = \text{const.}$, the geometric law of ν_Δ completely excludes the RT property for V_Δ (see eq. 12). The reason for this lies in the branching structure of the cluster.

Numerical examples

To support the theoretical conclusions, we simulated a synthetic earthquake catalogue from the regular ETAS(P) model. To mimic the features of real seismicity, the set of ETAS parameters was mainly derived from the analysis of the earthquake catalogue of Northeastern Italy (Benali *et al.* 2020). The simulation region also corresponds to that area, namely: Lon 11.5–14.0 and Lat 45.5–47.0.

The model was based on the following parameters:

- (1) The uniform background seismicity with the rate $\mu = 0.04$ events (degree²·day)⁻¹, which roughly corresponds to the rate of background events identified in the earthquake catalogue of Northeastern Italy. The uniformity makes it easier the declustering process, as it allows to better identify the clustered component.
- (2) Magnitude range $m \geq 2$; the Gutenberg–Richter parameter $\beta = 2.07$, which corresponds to the b -value 0.9.
- (3) The Utsu Law parameters: $\lambda = 0.22$ (event/day), $\alpha = 1.54$. The estimated $\lambda = 0.67$ for Northeastern Italy was reduced to 0.22 to satisfy the stability conditions $\Lambda_1 = E\lambda(m) < 1$.
- (4) The Omori law parameters: $c = 0.015$ (day), $p = 1.037$. The estimated p parameter very close to 1 implies an extremely long tail of $f_2(t)$, so that offspring events may span several thousands years (Harte 2013); this practically hinders the analysis of cluster size for synthetic catalogues. However, we recall that, from the theoretical viewpoint, the temporal component of the ETAS model does not affect cluster size. Therefore, in order to prevent this inconvenience, instead of increasing the p parameter, we considered a truncated distribution $f_2(t)$ on the support $t < 0.7$ yr, which corresponds to the 95 per cent quantile of the cluster lifetime for Northeastern Italy.
- (5) The space scattering law parameters: $d = 0.0085$ (degree), $q = 2.25$, $\gamma = 0.62$.

The synthetic earthquake catalogue, consisting of 143 568 events, spans about 500 yr and includes information about the links between each event and its direct descendants. The synthetic time-series were processed by the nearest-neighbour method (Zaliapin & Ben-Zion 2013) to identify clusters of events. This method allows partitioning earthquakes into background and clustered components, based on nearest-neighbour distances between earthquakes in the space–time–magnitude domain (Baiesi & Paczuski 2004; Zaliapin *et al.* 2008). In this study, the parameters necessary for the computation of nearest-neighbour distances, namely the b -value and the fractal dimension of epicentres d_f , were set according to the values estimated from the analysis of Northeastern Italy catalogue (Benali *et al.* 2020, and references therein): b -value = 0.9 and $d_f = 1.1$. The threshold of the nearest-neighbour distance, which separates the clustered and background components, was automatically set ($\log_{10}\eta_0 = -4.2$) following a criterion based on a 1-D Gaussian mixture model with two modes, where the threshold is the maximum likelihood boundary between the two modes (Zaliapin & Ben-Zion 2013). The total Δ -productivity distribution was examined considering the clusters extracted from the synthetic earthquake catalogue.

Fig. 1 represents the total Δ -productivity distributions ($\Delta = 2$) (a) for arbitrary initial events and (b) for main shocks, as identified by nearest-neighbour method applied to the synthetic ETAS(P) seismicity. In the first case, the distribution V_Δ demonstrates a heavy tail (requiring the use of logarithmic values in the figure), which is also confirmed by the coefficient of variation $\sigma(V_\Delta)/EV_\Delta \approx 10.8$. On the contrary, the distribution V_Δ for main shocks (Fig. 1b) demonstrates the light tail behaviour; for comparison with the previous case $\sigma(V_\Delta)/EV_\Delta \approx 0.7$.

Fig. 2 shows similar results based on the clusters originally defined by the branching structure of the synthetic ETAS(P) catalogue. The coefficients of variation are also consistent with previous results: (a) $\sigma(V_\Delta)/EV_\Delta \approx 9.7$ for arbitrary initial events and (b) $\sigma(V_\Delta)/EV_\Delta \approx 0.8$ for main shocks.

A good agreement of the empirical data Figs 1 and 2 demonstrates the robustness of conclusions about the tail behaviour of V_Δ distribution.

4 CONCLUSION

The concept of the Δ -productivity ν_Δ is associated with a priori ideas about the structure of a cluster of seismic events in the form of a random branching tree. Such a representation is not the only possible one. Therefore, the Geometric distribution of ν_Δ can be considered as a useful alternative within the framework of the generalized ETAS(F) model. The other characteristic of the Δ -productivity, namely V_Δ for main shocks, is more objective and stable.

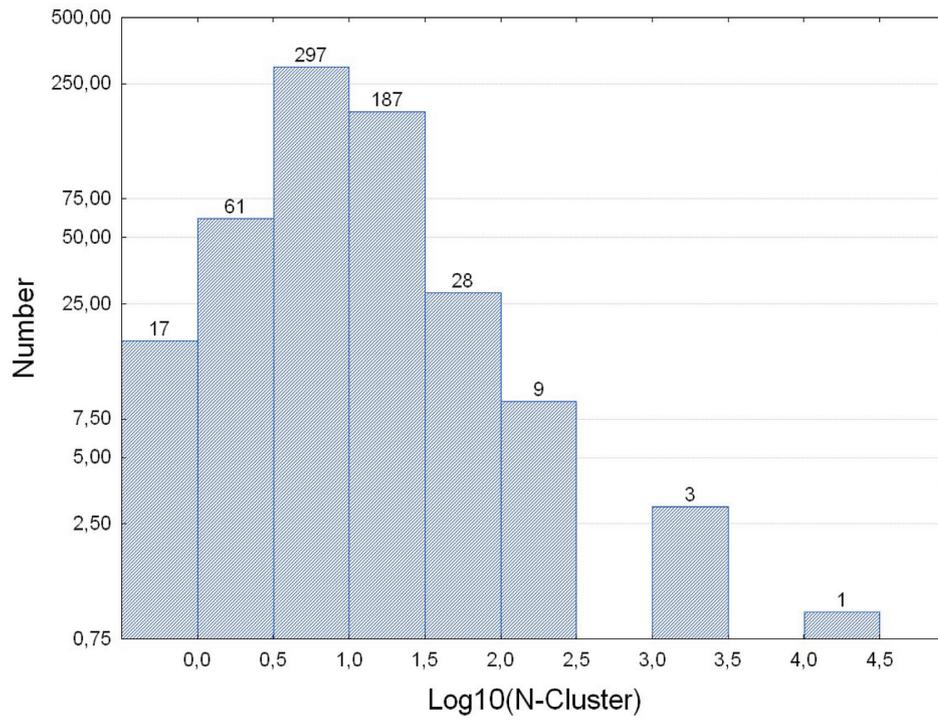
For large class of the ETAS(F) models, we answered the question when the total Δ -productivity distribution has light tails for main shocks and heavy ones for arbitrary events. The real data demonstrate the possibilities of this kind.

Note the difficulties of substantiating the mentioned geometric laws for ν_Δ and aftershock's V_Δ . They are associated both with the declustering of seismicity and with the effect of averaging the distributions of ν_Δ and V_Δ when grouping data.

The first difficulty can be judged by experiments of Zaliapin & Ben-Zion (2013) with the nearest-neighbour method. This method adapted to the ETAS structure has 40 per cent error of incorrectly determining the parent of an event. The stability of the conclusions in this case may depend on the strict similarity in the real hierarchical structure of the seismic cluster.

(a) The empirical data:

$$N = 612, EV_{\Delta} = 39.8, \sigma(V_{\Delta}) = 431.3, \max V_{\Delta} = 10079$$



(b) The empirical data:

$$N = 1074, EV_{\Delta} = 9.2, \sigma(V_{\Delta}) = 6.5, \max V_{\Delta} = 52$$

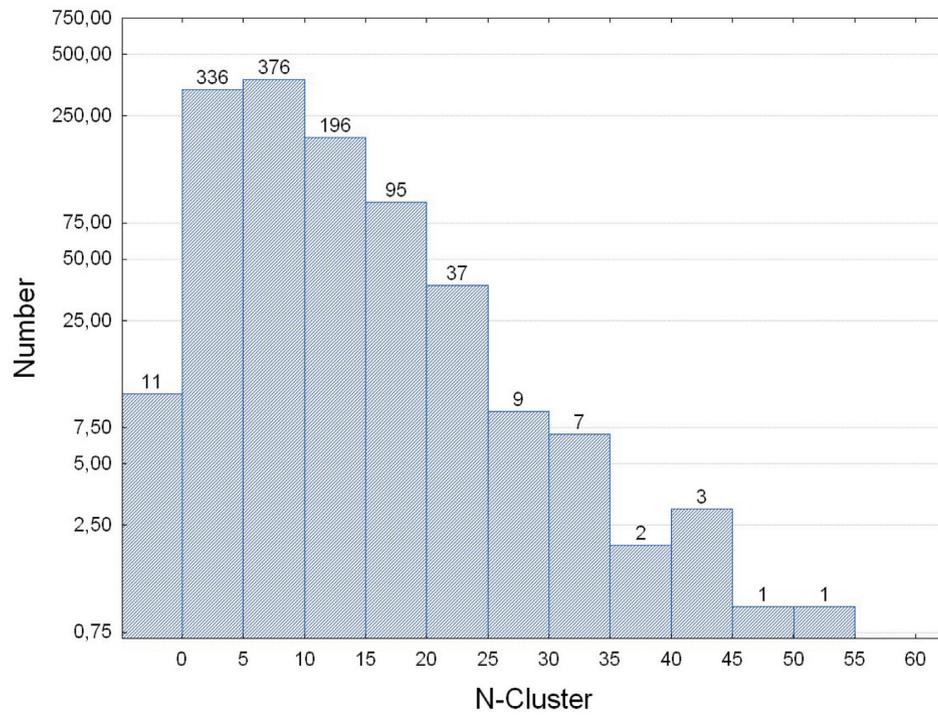
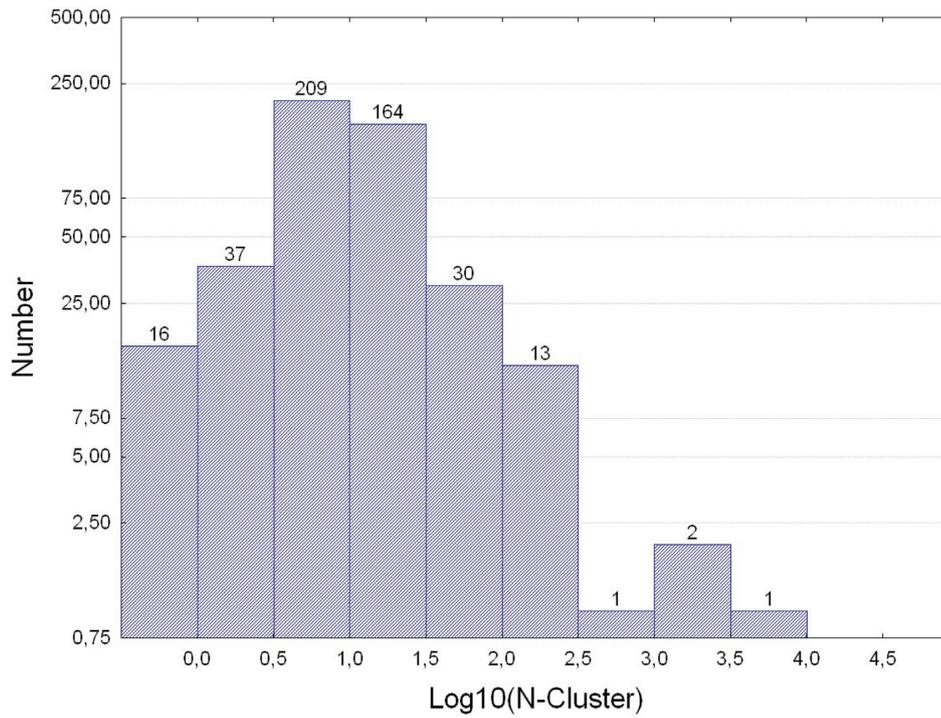


Figure 1. Clusters identified by nearest-neighbour method applied to the synthetic ETAS(P) seismicity: The V_{Δ} distributions, $\Delta = 2$, for (a) arbitrary initial events, $m_0 \geq 4$ and (b) main shocks, $m_0 \geq 4$.

(a) The empirical data:

$$N = 481, EV_{\Delta} = 46.5, \sigma(V_{\Delta}) = 452.3, \max V_{\Delta} = 9470$$



(b) The empirical data:

$$N = 1074, EV_{\Delta} = 9.9, \sigma(V_{\Delta}) = 7.9, \max V_{\Delta} = 56$$

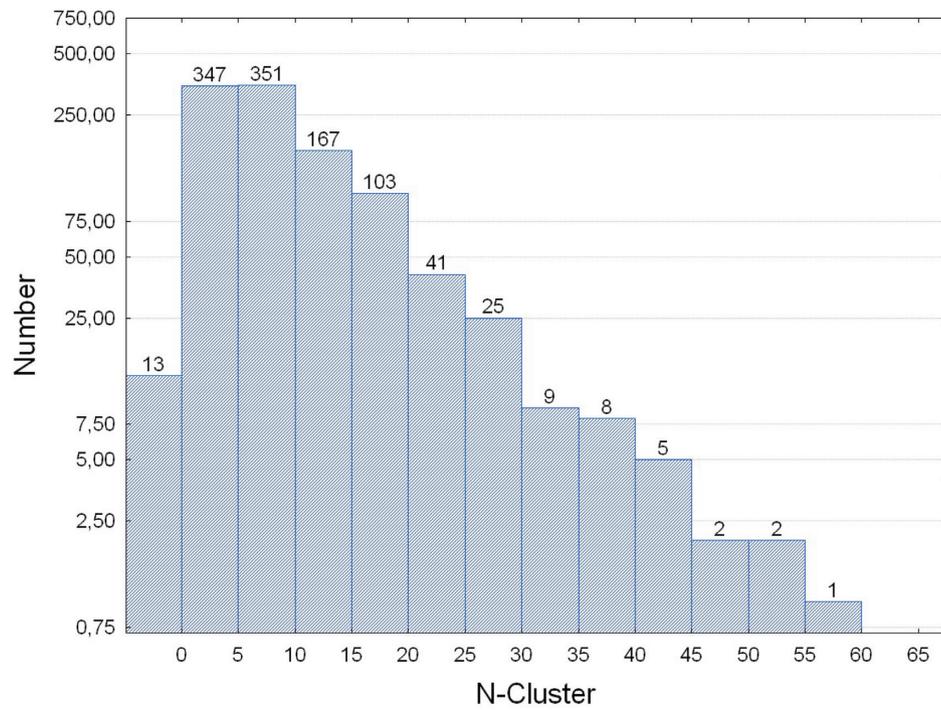


Figure 2. Clusters in the synthetic ETAS(P) seismicity: The V_{Δ} distributions, $\Delta = 2$, for (a) arbitrary initial events, $m_0 \geq 4$ and (b) main shocks, $m_0 \geq 4$.

The second difficulty is expressed in the fact that the mixture of geometric distributions retains the original property of monotonicity, but ceases to be geometric, that is loses the original G -type. This is unavoidable if the distribution parameter is dependent on location or the magnitude of the initial event (see eq. 6). Therefore, we can speak more confidently about the light tail of the V_Δ distribution for the main shocks.

Finally, the theoretical result for ETAS(F) model shows that the types of distributions of ν_Δ and V_Δ cannot be the same due to the branching structures of the clusters.

ACKNOWLEDGMENTS

We are grateful to Prof J. Zhuang and another anonymous reviewer for their careful reading of the manuscript and comments that allowed us to significantly revise and improve the text.

DATA AVAILABILITY

The data (synthetic earthquake catalogue) underlying this paper will be shared on reasonable request to the corresponding author.

REFERENCES

- Baiesi, M. & Paczuski, M., 2004. Scale-free networks of earthquakes and aftershocks, *Phys. Rev. E*, **69**, doi:10.1103/PhysRevE.69.066106.
- Baranov, S., Narteau, C. & Shebalin, P., 2022. Modeling and prediction of aftershock activity, *Surv. Geophys.*, doi:10.1007/s10712-022-09698-0.
- Baranov, S.V. & Shebalin, P.N., 2019. *Laws of the Post-Seismic Process and the Forecast of Strong Aftershocks*, pp. 218, RAS.
- Baró, J., 2020. Topological properties of epidemic aftershock processes, *J. geophys. Res.*, **125**(5), e2019JB018530, doi:10.1029/2019JB018530.
- Benali, A., Peresan, A., Varini, E. & Talbi, A., 2020. Modelling background seismicity components identified by nearest neighbour and stochastic declustering approaches: the case of Northeastern Italy, *Stoch. Environ. Res. Risk Assess.*, **34**, 775–791.
- Harris, T.E., 1963. *The Theory of Branching Processes*, Springer-Verlag.
- Harte, D.S., 2013. Bias in fitting the ETAS model: a case study based on New Zealand seismicity, *Geophys. J. Int.*, **192**(1), 390–412.
- Hawkes, A.G. & Oakes, D., 1974. A cluster process representation of a self-exciting process, *J. Appl. Probab.*, **11**(3), 493–503.
- Kagan, Y.Y., 2010. Statistical distributions of earthquake numbers: consequence of branching process, *Geophys. J. Int.*, **180**(3), 1313–1328.
- Kagan, Y.Y., 2014. *Earthquakes: Models, Statistics, Testable Forecasts*, pp. 283, AGU, Wiley.
- Ogata, Y., 1998. Space-time point-process models for earthquake occurrences, *Ann. Inst. Stat. Math.*, **50**, 379–402.
- Saichev, A. & Sornette, D., 2017. Statistics of seismic cluster durations, preprint, (arXiv:1708.06970).
- Saichev, A., Helmstetter, A. & Sornette, D., 2005. Power-law distributions of offspring and generation numbers in branching models of earthquake triggering, *Pure appl. Geophys.*, **162**, 1113–1134.
- Shebalin, P.N. & Baranov, S.V., 2021. Statistical laws of post-seismic activity, in *Statistical Methods and Modeling of Seismogenesis*, Chapter 3, pp. 63–103, eds Limmios, N., Papadimitriou, E. & Tsaklidis, G., John Wiley & Sons, Ltd..
- Shebalin, P.N., Baranov, S. & Dzeboev, B.A., 2018. The law of the repeatability of the number of aftershocks, *Doklady Akademii Nauk*, **481**, 963–966.
- Shebalin, P.N., Narteau, C. & Baranov, S.V., 2020. Earthquake productivity law, *Geophys. J. Int.*, **222**(2), 1264–1269.
- Soloviev, S.L. & Solovieva, O.N., 1962. Exponential distribution of the total number of subsequent earthquakes and decreasing with the depth of its average value, *Izv. AN USSR (geophys.)*, **12**, 1685–1694.
- Zaliapin, I. & Ben-Zion, Y., 2013. Earthquake clusters in southern California I: Identification and stability, *J. geophys. Res.*, **118**(6), 2847–2864.
- Zaliapin, I. & Ben-Zion, Y., 2016. A global classification and characterization of earthquake clusters, *Geophys. J. Int.*, **207**(1), 608–634.
- Zaliapin, I., Gabrielov, A., Keilis-Borok, V. & Wong, H., 2008. Clustering analysis of seismicity and aftershock identification, *Phys. Rev. Lett.*, **101**, doi:10.1103/PhysRevLett.101.018501.

APPENDIX

A1 Proof of Statement 1

Let $E(z^{\nu(m_0)} | m_0) = \varphi(z, \lambda(m_0))$ be the generating function of the $\nu(m_0)$ with the characteristic $\lambda(m_0) = E\nu(m_0)$. If $\{m_i: i = 1, \dots, \nu(m_0)\}$ are magnitudes of the direct descendants of the m_0 event, then the $\nu_\Delta(m_0)$ has the same distribution as $\sum_{i=1}^{\nu(m_0)} \varepsilon(m_i)$, where $\varepsilon(m) = \mathbf{1}_{m \geq m_0 - \Delta}$ is the logical 1-0 function. Since $\{m_i\}$ are independent,

$$\varphi_\Delta(z, \lambda(m_0)) := E(z^{\nu_\Delta(m_0)} | m_0) = \sum_{n \geq 0} (Ez^{\varepsilon(m)})^n P(\nu(m_0) = n) = \varphi(\pi_{m_0 - \Delta} + \bar{\pi}_{m_0 - \Delta} z, \lambda(m_0)) \quad , \quad (\text{A1})$$

where $\bar{\pi}_{m_0 - \Delta} = P(m \geq m_0 - \Delta | m_0 \geq \Delta)$ and $\pi_m = 1 - \bar{\pi}_m$.

Recall that $\varphi(z, \lambda(m)) = (1 - p(m))/(1 - p(m)z) = [1 - \lambda(m)(z - 1)]^{-1}$ for $F = G$ and $\varphi(z, \lambda(m)) = \exp(\lambda(m)(z - 1))$ for $F = P$. Substituting any of the functions $\varphi(z, \lambda(m))$ in the right-hand part of eq. (A1), we get the same function with the replacement of $\lambda(m_0)$ by

$$\lambda_\Delta(m_0) = \lambda(m_0) \bar{F}_1(m_0 - \Delta) / \bar{F}_1(\Delta) \quad . \quad (\text{A2})$$

Assume that the right-hand part of eq. (A2) is independent of m_0 in the interval $\Delta \leq m_0 < M$. Then $\lambda(m)$ and $F_1(m)$ are differentiable functions simultaneously. Since $\lambda_\Delta(m_0) = \text{const}$ for $m_0 \geq \Delta$, we have

$$0 = \left. \frac{d\lambda_\Delta(m_0)}{dm_0} \right|_{m_0 = \Delta} = [\dot{\lambda}(\Delta) - \lambda(\Delta) f_1(0)] / \bar{F}_1(\Delta) \quad .$$

Here we used condition $\bar{F}_1(0) = 1$. Solving the equation $\dot{\lambda}(\Delta) - \lambda(\Delta)f_1(0) = 0$ in the interval $0 < \Delta \leq m_0$, we get $\lambda(m) = \lambda e^{am}$, $m \in (0, M)$, where $a = f_1(0)$. Substitute in eq. (A2) $\lambda(m) = \lambda e^{am}$, $\lambda_\Delta(m_0) = C(\Delta)$ and $\Delta = 0$, to get $\bar{F}_1(m) = C(0)\lambda^{-1}e^{-am}$. Since $\bar{F}_1(0) = 1$, we have $\bar{F}_1(m) = e^{-am}$.

A2 Proof of the statement 2

A2.1 General remarks.

We will need the following notation: $N_M(m)$ is the total number of events of magnitude $\geq M$ triggered by an event of magnitude m . In particular, the total Δ -productivity $V_\Delta(m_0)$ of the initial event with magnitude m_0 is $N_{M(m_0)}(m_0)$, $M(m) = (m - \Delta)_+$, where $a_+ = \max(a, 0)$. The notation $E(*|m)$ denotes the conditional mean given m .

The spatio-temporal components of the ETAS model do not affect cluster size, since the law of generating new events depends only on the parent magnitude, and the problem of declustering does not arise in our theoretical analysis. Therefore, the spatio-temporal components in theoretical analysis can be any and the distribution of $N_M(m)$ is the same for any $x = (t, g, m)$.

By definition, if $\{m_i: i = 1, \dots, v(m_0)\}$ are offspring magnitudes of m_0 , then $\{N_M(m_i): i = 1, \dots, v(m_0)\}$ are independent for a given m_0 . Moreover, they are identically distributed together with $N_M(m_0)$ since the magnitudes $m_i, i \geq 0$, have a common distribution $f_1(m)$. Let us consider a new variable

$$\tilde{N}_M(m_0) = \sum_{i=1}^{v(m_0)} \{N_M(m_i) + \mathbf{1}_{m_i \geq M}\} \tag{A3}$$

where the symbol $\mathbf{1}_{a \geq b}$ denotes the logical function 1-0. Obviously, eq. (A3) is nothing more than counting events with magnitude $\geq M$ in the cluster through first-generation events and its triggered events of magnitude $\geq M$. The component $\mathbf{1}_{m_i \geq M}$ takes into account the i th event of the first generation if its value is greater than M . Therefore $N_M(m_0)$ and $\tilde{N}_M(m_0)$ are identically distributed, that is $N_M(m_0) \stackrel{law}{=} \tilde{N}_M(m_0)$.

The moments $\mathbf{E}V_\Delta^n$.

We will use the notation:

$$\Lambda_n = \int \lambda^n(m)f_1(m)dm, \quad \bar{\Lambda}_1 = 1 - \Lambda_1, \quad \lambda(m) = p(m)/(1 - p(m)) \text{ in the case } F = G,$$

$$b_M^{(n)}(m) = E(N_M^n(m) | m), \quad b_M^{(n)} = EN_M^n(m), \quad \pi_M = P(m \geq M).$$

Mean value: $\mathbf{E}V_\Delta$.

Averaging eq. (A3) given $m_0 = m$, we have

$$b_M^{(1)}(m) = \lambda(m)(b_M^{(1)} + \pi_M).$$

After averaging over m , we obtain the equation for $b_M^{(1)}$:

$$b_M^{(1)} = \Lambda_1(b_M^{(1)} + \pi_M).$$

For finite $b_M^{(1)}$, this equality can be true if $\Lambda_1 < 1$. Hence

$$b_M^{(1)} = \pi_M \cdot \Lambda_1 / \bar{\Lambda}_1, \quad b_M^{(1)}(m) = \lambda(m)\pi_M / \bar{\Lambda}_1. \tag{A4}$$

Since $V_\Delta = N_{M(m)}(m)$, $M(m) = (m - \Delta)_+$, we have

$$E(V_\Delta | m) = \lambda(m)\bar{F}_1(m - \Delta) / \bar{\Lambda}_1, \quad E(V_\Delta) \leq \Lambda_1 / \bar{\Lambda}_1, \\ E(V_\Delta | m \geq \Delta) = E(V_\Delta \mathbf{1}_{m \geq \Delta}) / \bar{F}_1(\Delta) \leq E(V_\Delta) / \bar{F}_1(\Delta) \leq (\Lambda_1 / \bar{\Lambda}_1) / \bar{F}_1(\Delta),$$

that is, V_Δ has finite mean value in conditional and unconditional situations if $\Lambda_1 < 1$.

The case $\mathbf{n} > 1$.

(a) Fixed/random initial \mathbf{m}_0

It is easy to see that for any $\lambda(m) \geq 0$ there exists $K \leq \infty$ such that $\Lambda_k = E\lambda^k(m)$ is finite for $k \leq K$ and is infinite for $k > K$. Let us show that there are finite constants B_k such that

$$b_M^{(k)} \leq B_k, \quad b_M^{(k)}(m) < \infty \quad \text{if and only if} \quad \Lambda_k < \infty, \quad \Lambda_1 < 1. \tag{A5}$$

We have shown (see above) that eq. (A5) is true for $k = 1$. We will prove eq. (A3) by the method of induction on k . Assume that all values in eq. (A5) are finite for $k \leq n - 1$. Putting $\xi_i = N_M(m_i) + \mathbf{1}_{m_i \geq M}$ in eq. (A3), one has

$$N_M^n(m) = \sum_{i_1 \neq i_2 \neq \dots \neq i_n}^{v(m)} \xi_{i_1} \dots \xi_{i_n} + R_M^{(n)} + \sum_{i=1}^{v(m)} \xi_i^n, \tag{A6}$$

where $R_M^{(n)}$ represents the secondary component of $N_M^n(m)$. The number of terms in the first sum is $\eta_n = v(m)(v(m) - 1) \dots (v(m) - n + 1)$. By the RT property of the F -distribution, $Ez^{v(m)} = \varphi(z, \lambda(m)) = \psi(\lambda(m)(z - 1))$, where $\varphi(z, \lambda(m))$ has all derivatives at $z = 1$ because all moments of $v(m)$ are finite. Hence

$$E[\eta_n | m] = \left(\frac{d}{dz} \right)^n \varphi(z, \lambda(m)) \Big|_{z=1} = a_n(F) \lambda^n(m), \tag{A7}$$

where $a_n(F) = \left(\frac{d}{dz} \right)^n \psi(z) \Big|_{z=0} < \infty$. Therefore

$$E \left(\sum_{i_1 \neq i_2 \neq \dots \neq i_n}^{v(m)} \xi_{i_1} \dots \xi_{i_n} \mid m \right) = a_n(F) \lambda^n(m) [b_M^{(1)} + \pi_M]^n.$$

Similarly

$$E \left(R_M^{(n)} \mid m \right) = \sum_{\Omega} c_n(\omega) a_k(F) \lambda^k(m) \tilde{b}_M^{(r_1)} \dots \tilde{b}_M^{(r_k)}, \tag{A8}$$

where

$$\tilde{b}_M^{(r)} = E \{ N_M(m_i) + \mathbf{1}_{m_i \geq M} \}^r,$$

and

$$c_n(\omega) = \frac{n!}{r_1! \dots r_k! k!}, \quad \Omega = \{ \omega = (k, r_1, \dots, r_k) : 1 < k < n, r_i \geq 1, r_1 + \dots + r_k = n \}.$$

Since for any random variable $\xi \geq 0$:

$$E \xi^n E \xi^m \leq E \xi^{n+m},$$

we get from eq. (A8):

$$E \left(R_M^{(n)} \mid m \right) \leq C_n (1 \vee \lambda(m))^{n-1} \max_{1 \leq k < n} \left(\tilde{b}_M^{(k)} \tilde{b}_M^{(n-k)} \right), \tag{A9}$$

where C_n is some finite constant. The expectation of eq. (A6), given m , is

$$b_M^{(n)}(m) = a_n(F) \lambda^n(m) \left(\tilde{b}_M^{(1)} \right)^n + E \left(R_M^{(n)} \mid m \right) + \lambda(m) \left[b_M^{(n)} + \delta_M^{(n)} \right], \tag{A10}$$

where $\delta_M^{(n)} = \tilde{b}_M^{(n)} - b_M^{(n)}$ is a secondary term of $\tilde{b}_M^{(n)}$. By eq. (A8),

$$\delta_M^{(n)} < E \left[(N_M(m) + 1)^n - N_M^{(n)}(m) \right] < n E \left[(N_M(m) + 1)^{n-1} \right] < n (2^{n-1} b_M^{(n-1)} + 1). \tag{A11}$$

Averaging eq. (A10) over m , we get

$$b_M^{(n)}(1 - \Lambda_1) = a_n(F) \Lambda_n \left(\pi_M / \bar{\Lambda}_1 \right)^n + E R_M^{(n)} + \Lambda_1 \delta_M^{(n)}. \tag{A12}$$

The relations (A11)–(A12) show that $b_M^{(n)}$ is finite if and only if $\Lambda_n < \infty$. Similarly, eqs (A10)–(A11) show that $b_M^{(n)}(m) < \infty$ if and only if $b_M^{(n)} < \infty$. Hence, at the n th step of induction, these three quantities are finite simultaneously, that is eq. (A5) holds. By setting $M = (m - \Delta)_+$, we get $E(V_\Delta^n | m) = b_{(m-\Delta)_+}^{(n)}(m) < \infty$ only when $\Lambda_n < \infty$. It remains to consider $E(V_\Delta^n)$. By eqs (A10) and (A12),

$$b_M^{(n)}(1 - \Lambda_1) = a_n(F) \Lambda_n \left(\pi_M / \bar{\Lambda}_1 \right)^n + E R_M^{(n)} + \Lambda_1 \delta_M^{(n)},$$

$$E(V_\Delta^n) = E b_{(m-\Delta)_+}^{(n)}(m) \geq E \lambda(m) b_{(m-\Delta)_+}^{(n)} \geq E \lambda(m) a_n(F) \Lambda_n \pi_{(m-\Delta)_+}^n / \bar{\Lambda}_1^{n+1}.$$

Therefore $E(V_\Delta^n) = \infty$ if Λ_n is infinite. Let $\Lambda_n < \infty$. Using (A10) again, we have

$$E(V_\Delta^n) \leq a_n(F) \Lambda_n (\bar{\Lambda}_1)^{-n} + E \left(R_{M(m)}^{(n)} \mid m \right) + E \lambda(m) \left[b_{M(m)}^{(n)} + \delta_{M(m)}^{(n)} \right].$$

Due to eq. (A9), we have $E \left(R_{M(m)}^{(n)} \mid m \right) \leq \tilde{C}_n \max(1, \Lambda_1, \dots, \Lambda_{n-1})$; by eq. (A9), $E \lambda(m) \delta_{M(m)}^{(n)} \leq D_n \Lambda_1$. Finally, $E \lambda(m) b_{M(m)}^{(n)} \leq B_n \Lambda_1$, because we have proved (A3) for index n . Hence, $E(V_\Delta^n) < \infty$ if $\Lambda_n < \infty$. The proof is complete.

(b) Fixed dominant \mathbf{m}_0

Let us consider the ETAS cluster with a dominant initial magnitude m_0 . Let $\hat{N}_{0M}(m)$ be the number of events of magnitude $>M$ in the cluster with the initial event $m \leq m_0$ in which all events have magnitude $< m_0$. In such cluster, direct descendants of any initial magnitude m are realized under the condition $\Omega(m) = \{m_i < m_0 : i = 1, \dots, \nu(m_0)\}$.

In term of the generating function $Ez^{\nu(m)} = \varphi(z, \lambda(m))$, the probability of $\Omega(m)$ is

$$P(\Omega(m)) = \varphi(F_1(m_0), \lambda(m)) = \psi(-\lambda(m)\bar{F}_1(m_0)). \tag{A13}$$

Therefore, the conditional probability density of the direct descendants is

$$P\{m_i = \mu_i : i = 1, \dots, n; \nu(m_0) = n \mid \Omega(m)\} = \hat{f}_{01}(\mu_1) \dots \hat{f}_{01}(\mu_n) P(\hat{\nu}(m) = n).$$

where

$$\hat{f}_{01}(m) = f_1(m)/F_1(m_0), \quad 0 \leq m \leq m_0, \quad F_1(m) = \int_0^m f_1(u)du \tag{A14}$$

$$P(\hat{\nu}(m) = n) = P(\nu(m) = n \mid \Omega(m)) = P(\nu(m) = n)F_1^n(m_0)/\psi(-\lambda(m)\bar{F}_1(m_0)) \tag{A15}$$

Substituting Geometric (eq. 1) and Poisson (eq. 2) distributions here, we get the distributions of the same type

$$P(\nu(m_0) = n \mid \Omega(m)) = \begin{cases} \hat{\lambda}_0^n(m)e^{-\hat{\lambda}_0(m)}/n! & [P\text{-model}] \\ \hat{p}_0^n(m)(1 - \hat{p}_0(m)) & [G\text{-model}] \end{cases}, \tag{A16}$$

where

$$\begin{aligned} \hat{p}_0(m) &= p(m)F_1(m_0) = \lambda(m)F_1(m_0)/(1 + \lambda(m)) & [F = G] \\ \hat{\lambda}_0(m) &= \lambda(m)F_1(m_0) \begin{cases} 1 & [F = P] \\ (1 + \lambda(m)\bar{F}_1(m_0))^{-1} & [F = G] \end{cases} \end{aligned} \tag{A17}$$

Therefore, given the dominance of the initial magnitude m_0 , we are again dealing with an ETAS model in which the magnitude m is transformed into \hat{m} with a distribution (eq. A14), and the productivity $\nu(m)$ into $\hat{\nu}(\hat{m})$ with the same distribution (Poisson/Geometric), but with new parameters $\hat{\lambda}_0(\hat{m})$ as in eq. (A17). Therefore, as above, we will have the following relations:

$$\hat{N}_{0M}(\hat{m}) \stackrel{law}{=} \sum_{i=1}^{\hat{\nu}(\hat{m})} (\hat{N}_{0M}(\hat{m}_i) + \mathbf{1}_{\hat{m}_i \geq M}), \tag{A18}$$

where $\{\hat{N}_{0M}(\hat{m}_i) : i = 1, \dots, \hat{\nu}(\hat{m})\}$ are independent and identical distributed given $\hat{\nu}(\hat{m})$ and $\{\hat{m}_i : i = 1, \dots, \hat{\nu}(\hat{m})\}$.

As a result

$$V_{\Delta}(m_0) = \hat{N}_{0(m_0-\Delta)_+}(m_0), \quad m_0 \geq \Delta. \tag{A19}$$

Since m_0 is a parameter, in the random case of initial event, m_0 has distribution $F_1(m)/\bar{F}_1(\Delta)$. As we can see, eq. (A18) is a special case of eq. (A3), in which the magnitude range is finite. By eq. (A17), $\hat{\lambda}_0(\hat{m}) \leq \lambda(m_0)$ and therefore $\hat{\lambda}_0(\hat{m})$ has all moments. Consequently, the total Δ -productivity will have all the moments at any fixed m_0 .

(c) Random dominant \mathbf{m}_0

We need additional notation:

$$\begin{aligned} \hat{b}_{0M}^{(n)}(\hat{m}) &= E \left[\hat{N}_{0M}^n(\hat{m}) \mid \hat{m} \right], \quad \hat{b}_{0M}^{(n)} = E \hat{N}_{0M}^n(\hat{m}), \\ \tilde{b}_{0M}^{(n)} &= E \left(\hat{N}_{0M}(\hat{m}) + \mathbf{1}_{\hat{m} \geq M} \right)^n, \end{aligned} \tag{A20}$$

$$\hat{\pi}_{0M} = P(\hat{m} \geq M) = [F_1(m_0) - F_1(M)]_+/F_1(m_0),$$

$$\hat{\Lambda}_{0n} = E \hat{\lambda}_0^n(\hat{m}).$$

Using eq. (A17) and non-decreasing function $\lambda(m)$, we have for the P-model

$$\hat{\Lambda}_{0n} = \int_0^{m_0} \lambda^n(x) F_1^{n-1}(m_0) f_1(x) dx \leq \hat{\lambda}_0^{n-1}(m_0) \Lambda_1. \tag{A21}$$

The same is true for the G-model. Really, by eq. (A17) we have the following representation:

$$\hat{\Lambda}_{0n} = \hat{\lambda}_0^{n-1}(m_0) \int_0^{m_0} [\psi^{n-1}(\lambda(x))/(1 + \lambda(m)\bar{F}(m_0))] \lambda(x) f_1(x) dx,$$

where

$$\psi(m) = (\lambda^{-1}(m_0) + \bar{F}_1(m_0))/(\lambda^{-1}(m) + \bar{F}_1(m_0)).$$

Since $\lambda(m)$ is a non-decreasing function, $\psi(m) \leq 1$ on the interval $(0, m_0)$. Therefore, the integrand in square brackets is less than 1, which proves eq. (A21). As above, the relation eq. (A18) gives,

$$\hat{b}_{0M}^{(1)} = \hat{\Lambda}_{01} \hat{\pi}_{0M} / (1 - \hat{\Lambda}_{01}) \leq \Lambda_1 \hat{\pi}_{0M} / (1 - \Lambda_1), \tag{A22}$$

$$EV_\Delta \leq \Lambda_1 / (1 - \Lambda_1).$$

Setting $\xi_i = \hat{N}_{0M}(\hat{m}_i) + \mathbf{1}_{\hat{m}_i \geq M}$ in eq. (A18) and using eq. (A6), we get an analogue of eq. (A10):

$$\hat{b}_{0M}^{(n)}(\hat{m}) = \sum_{k=1}^n \sum_{\mathbf{r}^{(k)}} c_k(r_1, \dots, r_k) \hat{\lambda}_0^k(\hat{m}) \tilde{b}_{0M}^{(r_1)} \dots \tilde{b}_{0M}^{(r_k)}, \tag{A23}$$

where the summation is over all integer vectors $\mathbf{r}^{(k)} = (r_1, \dots, r_k): r_1 + \dots + r_k = n, r_i \geq 1$. Averaging eq. (A23) over \hat{m} we will also have

$$\hat{b}_{0M}^{(n)} = \sum_{k=2}^n \sum_{\mathbf{r}^{(k)}} c_k(r_1, \dots, r_k) \hat{\Lambda}_{0k} \tilde{b}_{0M}^{(r_1)} \dots \tilde{b}_{0M}^{(r_k)} + \hat{\Lambda}_{01} \tilde{b}_{0M}^{(n)}. \tag{A24}$$

Let's extract the term $\hat{\Lambda}_{01} \tilde{b}_{0M}^{(n)}$ from $\hat{\Lambda}_{01} \tilde{b}_{0M}^{(n)}$ and move it to the left-hand part of eq. (A24); then apply eq. (A21) to $\hat{\Lambda}_{0k}$ and take into account the relation $\hat{\Lambda}_{01} / (1 - \hat{\Lambda}_{01}) \leq \Lambda_1 / (1 - \Lambda_1)$. Then multiplying eq. (A24) by $\hat{\lambda}_0(m_0)$, we get

$$\hat{\lambda}_0(m_0) \hat{b}_{0M}^{(n)} \leq \frac{\Lambda_1}{1 - \Lambda_1} \sum_{k=2}^n \sum_{\mathbf{r}^{(k)}} \left\{ c_k(r_1, \dots, r_k) (\hat{\lambda}_0(m_0) \tilde{b}_{0M}^{(r_1)}) \dots (\hat{\lambda}_0(m_0) \tilde{b}_{0M}^{(r_k)}) + \hat{\lambda}_0(m_0) (\tilde{b}_{0M}^{(n)} - \hat{b}_{0M}^{(n)}) \right\}. \tag{A25}$$

Since $\tilde{b}_{0M}^{(n)} = E \left(\hat{N}_{0M}(\hat{m}_i) + \mathbf{1}_{\hat{m}_i \geq M} \right)^n$, the elements of the right-hand part of eq. (A25) can be obtained by summing and multiplying elements of the form: $\hat{\lambda}_0(m_0) \hat{\pi}_{0M}$ and

$$D_{k,\varepsilon} = \hat{\lambda}_0(m_0) E \hat{N}_{0M}^k(\hat{m}) (\mathbf{1}_{\hat{m} \geq M})^\varepsilon, \quad k = 1, \dots, n - 1; \quad \varepsilon = 0, 1, \tag{A26}$$

where $D_{k,\varepsilon} \leq \hat{\lambda}_0(m_0) \hat{b}_{0M}^{(k)}$. We have assumed that $\lambda(m) \int_0^\Delta f_1(m-x) dx \leq C$. Hence

$$\hat{\lambda}_0(m_0) \hat{\pi}_0(m_0 - \Delta) = \lambda(m_0) F_1(m_0) [F_1(m_0) - F_1(m_0 - \Delta)] / F_1(m_0) \leq C$$

and by virtue of eq. (A22)

$$\hat{\lambda}_0(m_0) \hat{b}_{0M(m_0)}^{(1)} \leq C_1, \quad M(m_0) = (m_0 - \Delta)_+.$$

Using eq. (A25) and doing induction over n , we can conclude that $\hat{\lambda}_0(m_0) \hat{b}_{0M(m_0)}^{(n)} \leq C_n$ for any n . But then, due to eq. (A19), $EV_\Delta^n = E \hat{b}_{0M(m_0)}^{(n)}(m_0) \leq K_n < \infty$ for any n .

The proof is complete.

A3 Proof of the statement 3.1

Let ξ be a random variable with a geometric distribution. Then

$$E\xi(\xi - 1)/(E\xi)^2 = 2. \tag{A27}$$

Consider $\xi = V_\Delta(m_0)$, where m_0 is the initial dominant magnitude in the ETAS(G) cluster. We will show that the relation (A27) is possible for at most one parameter $\Delta \in (0, m_0)$. For simplicity of notation, we will temporarily proceed from eq. (A3). Recall that

$$\Lambda_n = \int \lambda^n(m) f_1(m) dm, \quad \pi_M = P(m \geq M),$$

$$b_M^{(n)}(m) = E(N_M^n(m) | m), \quad b_M^{(n)} = E N_M^n(m).$$

Above we show that

$$b_M^{(1)}(m) = \lambda(m) \pi_M / \bar{\Lambda}_1, \quad b_M^{(1)} = \pi_M \Lambda_1 / \bar{\Lambda}_1, \quad b_M^{(1)} + \pi_M = \pi_M / \bar{\Lambda}_1. \tag{A28}$$

According to eq. (A10) for the ETAS(G) model, we have

$$b_M^{(2)}(m) = 2(\lambda(m) \pi_M / \bar{\Lambda}_1)^2 + \lambda(m) (b_M^{(2)} + 2b_M^{(1)} \pi_M + \pi_M^2). \tag{A29}$$

By eq. (A28),

$$2b_M^{(1)} \pi_M + \pi_M^2 = (1 + \Lambda_1) \pi_M^2 / \bar{\Lambda}_1. \tag{A30}$$

Averaging eq. (A29) over m , we can find $b_M^{(2)}$; using eq. (A30), we will have

$$b_M^{(2)} = 2\Lambda_2\pi_M^2/\bar{\Lambda}_1^3 + \Lambda_1(1 + \Lambda_1)\pi_M^2/\bar{\Lambda}_1^2. \tag{A31}$$

If we substitute eq. (A31) into eq. (A29), we get

$$b_M^{(2)}(m) = 2(b_M^{(1)})^2 + \lambda(m)(\pi_M/\bar{\Lambda}_1)^2(2\Lambda_2 + 1 - \Lambda_1^2)/\bar{\Lambda}_1.$$

As a result,

$$(b_M^{(2)}(m) - b_M^{(1)}(m))/(b_M^{(1)}(m))^2 = 2 + \lambda^{-1}(m)[(2\Lambda_2 + 1 - \Lambda_1^2)/\bar{\Lambda}_1 - \bar{\Lambda}_1/\pi_M]. \tag{A32}$$

The condition (A27) is satisfied only if

$$\pi_M = (1 - \Lambda_1)^2/(2\Lambda_2 + 1 - \Lambda_1^2), \tag{A33}$$

where $\Lambda_2 - \Lambda_1^2 = \sigma^2(\lambda) > 0$. The eq. (A33) with respect to M has unique solution if π_M is strictly monotone. Now we can return to the ETAS(G) cluster with dominant initial magnitude m_0 . In this case the main equation is (A18). It is identical to eq. (A10), but requires the following replacement of the characteristics used (see A14 and A17):

$$f_1(m) \implies f_1(m)/F_1(m_0), \quad 0 \leq m \leq m_0,$$

$$\lambda(m) \implies \lambda(m)F_1(m_0)/(1 + \lambda(m)\bar{F}_1(m_0)),$$

$$\pi_M \implies [F_1(m_0) - F_1(M)]_+/F_1(m_0), \tag{A34}$$

$$\Lambda_n \implies \int_0^{m_0} \lambda^n(m)F_1^{n-1}(m_0)(1 + \lambda(m)\bar{F}_1(m_0))^{-n} f_1(m)dm. \tag{A35}$$

Since $V_\Delta(m_0) = N_{m_0-\Delta}(m_0)$, we have to consider eq. (A33) given the substitutions (A34) and (A35) and $M = m_0 - \Delta$. Given $f_1(m) > 0$, the function (A34) will be strictly monotonic with respect to Δ . Hence, the relation (A27) is possible for at most one value of $\Delta = \Delta(m_0)$.

The proof is complete.

A4 Proof of the statement 3.2

Let $v(m)$ and $V(m)$ be the sizes of clusters of direct descendants and all descendants, respectively, associated to an initial event with magnitude m . Let $\phi_V(z, \lambda(m)) = Ez^{V(m)}$ and $\phi_V(z) = \int_0^\infty Ez^{V(m)} f_1(m)dm$ be the conditional and unconditional generating functions of $V(m)$. Due to the RT property of $v(m)$ distribution, we have

$$Ez^{v(m)} = \psi(\lambda(m)(z - 1)), \quad \psi(0) = \psi'(0) = 1, \quad \psi''(0) < \infty.$$

Then by eq. (A3)

$$\phi_V(z, \lambda(m)) = \psi(\lambda(m)(z\phi_V(z) - 1)),$$

and after integration over m we get

$$\phi_V(z) = \int_0^\infty \psi(\lambda(m)(z\phi_V(z) - 1))f_1(m)dm. \tag{A36}$$

Since it is assumed that $\int \lambda^2(m)f_1(m)dm < \infty$ and $Ev^2 < \infty$, it follows $EV^2 < \infty$ from Statement 2. This means that $\phi_V''(1) < \infty$. Set $\mu = EV = \phi_V'(1)$, $\Lambda_k = \int_0^\infty \lambda^k(m)f_1(m)dm$. Differentiating eq. (A36) at point 1, we get

$$\mu = \Lambda_1(1 + \mu) \quad \text{or} \quad \mu = \Lambda_1/(1 - \Lambda_1). \tag{A37}$$

Differentiating eq. (A36) twice at point 1, we get

$$\phi_V''(1) = \psi''(0)\Lambda_2(1 + \mu)^2 + \Lambda_1(2\mu + \phi_V''(1)).$$

Given eq. (A37), we have

$$\phi_V''(1) = \psi''(0)\Lambda_2(1 - \Lambda_1)^{-3} + 2\mu^2. \tag{A38}$$

Assume that the distribution types of v and V are the same. Then $\phi_V(z) = \psi(\mu(z - 1))$ and $\phi_V''(1) = \psi''(0)\mu^2$. Combining this equality with eq. (A38) we have

$$\psi''(0) = \psi''(0)\Lambda_2\Lambda_1^{-2}(1 - \Lambda_1)^{-1} + 2. \tag{A39}$$

By definition, all terms in this equality are non-negative, $\Lambda_2\Lambda_1^{-2} \geq 1$, and $(1 - \Lambda_1)^{-1} > 1$. Therefore, eq. (A39) is contradictory.