

Invariant Galton-Watson Branching Process for Earthquake Occurrence

Yevgeniy Kovchegov¹, Ilya Zaliapin², and Yehuda Ben-Zion³

¹Department of Mathematics, Oregon State University, Corvallis, OR 97331.

E-mail: kovchegy@math.oregonstate.edu

²Department of Mathematics and Statistics, University of Nevada Reno, Reno, NV 89557.

E-mail: zal@unr.edu

³Department of Earth Sciences and Southern California Earthquake Center, University of Southern California, Los Angeles, 90089.

E-mail: benzion@usc.edu

26 May 2022

SUMMARY

We propose a theoretical modeling framework for earthquake occurrence and clustering based on a family of invariant Galton-Watson (IGW) stochastic branching processes. The IGW process is a rigorously defined approximation to imprecisely observed or incorrectly estimated earthquake clusters modeled by Galton-Watson branching processes, including the Epidemic Type Aftershock Sequence (ETAS) model. The theory of IGW processes yields explicit distributions for multiple cluster attributes, including magnitude-dependent and magnitude-independent offspring number, cluster size, and cluster combinatorial depth. Analysis of the observed seismicity in southern California demonstrates that the IGW model provides a close fit to the observed earthquake clusters. The estimated IGW parameters and derived statistics are robust with respect to the catalog lower cutoff magnitude. The proposed model facilitates analyses of multiple quantities of seismicity based on self-similar tree attributes, and may be used to assess the proximity of seismicity to criticality.

Key words: Statistical seismology; Earthquake dynamics; Persistence, memory, correlations, clustering; Earthquake interaction, forecasting, and prediction

1 INTRODUCTION AND MOTIVATION

Stochastic branching processes have been widely used for modeling earthquake occurrence since the 1970s, starting with the pioneering works of Kagan (1973), Kagan & Knopoff (1976), and Vere-Jones (1976). The most widespread and mathematically developed approach builds on the *self-exciting Hawkes point process* (Hawkes 1971; Adamopoulos 1976; Daley & Vere-Jones 2003), which is equivalent to a *branching process with immigration* (Hawkes & Oakes 1974; Saichev et al. 2005; Baró 2020). Informally, a stochastic branching process describes a population (here – a sequence of earthquakes) where each member generates offspring according to a predefined set of rules. Accordingly, the earthquake population is decomposed into a collection of clusters, each of which starts with an immigrant and includes all its offspring, offspring of offspring, etc. Every such cluster can be represented by a tree graph (tree). The root of the tree corresponds to the first cluster event (immigrant), the other vertices to triggered earthquakes, and edges to triggering relations imposed by the model (which may or may not correspond to actual physical triggering if the model is fitted to data). An inflow of immigrants (back-

ground events) ensures that the population does not disappear with time. During 1970–1990 powerful probabilistic tools have been developed for working with space-time-magnitude generalizations of branching processes, and a tradition of their seismological applications has been established in a series of studies by Vere-Jones, Ogata, and coauthors (e.g., Vere-Jones 1970, 1978; Vere-Jones & Ozaki 1982; Ogata et al. 1982; Ogata 1983; Ogata & Vere-Jones 1984; Ogata & Katsura 1986, 1988; Musmeci & Vere-Jones 1992).

The Epidemic Type Aftershock Sequence (ETAS) model introduced by Ogata (1988) builds on these developments by synthesizing the key empirical laws of statistical seismology with rigorous stochastic modeling and estimation tools; see also Ogata (1998, 1999). The model describes a regional flow of earthquakes with magnitudes above M_0 . Background events are modeled by a Poisson process with intensity $\mu(t)$. Each earthquake generates offspring according to a modified Omori law (Omori 1894; Utsu 1970; Utsu et al. 1995). Specifically, an earthquake with magnitude M_i that occurred at time t_i produces offspring, also called the first-generation after-

shocks, according to a Poisson process with intensity

$$\nu(t|t_i, M_i) = \frac{K_0 10^{\alpha(M_i - M_0)}}{(t - t_i + c)^p}, \quad t > t_i \quad (1)$$

parameterized by positive constants K_0 , α , c and $p > 1$. The offspring intensity ν combines an exponential productivity law (numerator) and a power-law temporal decay (denominator). Every newly generated event, background or offspring, is assigned a magnitude M_i independently of other events (including its parent) according to the Gutenberg-Richter law (Gutenberg & Richter 1944):

$$P(M_i > M) = 10^{-b(M - M_0)}, \quad M > M_0. \quad (2)$$

The combined earthquake flow includes the background events, their first-generation aftershocks, offspring of these aftershocks (called second generation aftershocks), offspring of the second-generation aftershocks (third generation aftershocks), and so on. The combined flow is a point process specified by its conditional intensity

$$\lambda(t|\mathcal{H}_t) = \mu(t) + \sum_{i:t_i < t} \nu(t|t_i, M_i), \quad (3)$$

with a process history $\mathcal{H}_t = \{(t_i, M_i) : t_i < t\}$. This modeling framework may include a space component, by considering a point field with the background intensity $\mu(t, \mathbf{x})$ and a conditional space-time distribution of offspring given by a density $\nu(t, \mathbf{x}|t_i, \mathbf{x}_i, M_i)$; see details in Ogata (1998, 1999).

A Galton-Watson (GW) stochastic branching process describes a population that begins with a single progenitor at step $s = 0$ and develops in discrete steps. At every time step each existing member gives birth to $k = 0, 1, \dots$ offspring, independently of other members, according to a distribution $\{p_k\}$, and terminates. If we only focus on the parent-offspring earthquake relations in the ETAS model, leaving aside for now the time and space attributes, then a single cluster that begins with an earthquake of a random magnitude and includes its aftershocks of all generations is described by a GW process. If the average offspring number is unity ($\sum k p_k = 1$) the process is called critical, and if it is less than unity ($\sum k p_k < 1$) it is called sub-critical. The average progeny (population size) of a critical process is unity at each step, and it can be shown that the average progeny of a subcritical process vanishes exponentially. Critical and subcritical GW processes produce finite populations with probability 1. At the same time, the average size of a critical population is infinite. Within the ETAS framework, the offspring distribution p_k is obtained by taking the conditional Poisson distribution of offspring numbers for a parent of a given magnitude and integrating it with respect to the magnitude distribution (2).

The ETAS model and its multiple ramifications demonstrated an unprecedented success in approximating and forecasting observed earthquake rates in various seismically active regions. It provides a foundational tool for the Uniform California Earthquake Rupture Forecasts (e.g., Field et al. 2017) and facilitates multiple forecast models tested within the Collaboratory for the Study of Earthquake Predictability (e.g., Zechar et al. 2010).

Despite the overall effectiveness of the ETAS model, it has several well recognized features that complicate analyses of observed seismicity. First, it is well known that “[p]arameters of the modified Omori formula and the ETAS model may [...] vary spatially and in some cases temporally”

(Utsu et al. 1995, p. 26). For example, Page et al. (2016) estimated regional aftershock parameters in the Omori law using global regionalization of García et al. (2012). They found that while the mean aftershock productivity varies from region to region by a factor of 10, the reported ratios of the aftershock productivities for distinct sequences within the same region may be as high as 1000 (Page et al. 2016, Fig. 8). Using notations of the present work, this refers to the sequence-to-sequence ratio of parameter K_0 from (1). One may wonder if an alternative parameterization may help decreasing this large variability of parameters. An independent evidence of an insufficient model flexibility against data comes from analysis of offspring (first generation aftershock) numbers. The total number of offspring generated by a single earthquake according to (1) is a Poisson random variable with mean

$$N^{\text{off}}(M) = \int_0^\infty \nu(t|0, M) dt = R_0 10^{\alpha(M - M_0)}, \quad (4)$$

where $R_0 = K_0 c^{1-p} (p - 1)^{-1}$. However, the offspring numbers estimated in observations may deviate from the Poisson distribution. The data commonly exhibit over-dispersion, with the negative binomial distribution providing a much closer fit (e.g., Zaliapin & Ben-Zion 2013a). Kagan (2010) performed an alternative (non-ETAS) analysis suggesting importance of the over-dispersed negative binomial distribution for space-time earthquake counts. Second, the model fitting involves estimation of multiple (typically, six to eight) parameters, which is known to be unstable with respect to the lower magnitude cutoff M_0 , and other catalog specifications and uncertainties (Veen & Schoenberg 2008; Wang et al. 2010). Finally, most of the results related to model forecast and estimation are obtained numerically, which is challenging in short time intervals right after large earthquakes, and more generally in a near-critical process that often produces very large clusters.

We propose a new theoretical framework for modeling earthquake clustering that overcomes the above discrepancies while preserving the main postulates of stochastic branching theory. The framework is based on a family of invariant Galton-Watson (IGW) branching processes (Neveu 1986; Duquesne & Winkel 2007, 2019; Kovchegov & Zaliapin 2021). The trajectories of IGW processes induce a one-parameter family of critical Galton-Watson tree distributions that is invariant under the operations of tree erasure, thinning, hereditary reduction, Horton pruning, and generalized dynamical pruning. These operations mimic tree deformation caused by imprecise observations or estimations. Moreover, the IGW trees are the only attractors of other critical Galton-Watson trees under the tree-deforming operations (Kovchegov & Zaliapin 2021). This means that if any sufficiently large critical tree is consecutively deformed, the resulting tree necessarily approaches an IGW tree; see more in Sect. 2.3. Accordingly, the class of IGW processes is a convenient approximation to general imprecisely observed Galton-Watson branching processes, including the ETAS model.

Seismicity in the crust is not expected to operate at all places and times precisely at criticality (e.g., Ben-Zion 2008, Sects. 4 and 5), and can exhibit deviations from criticality related to values of physical parameters governing seismicity and the time within large earthquake cycle. Examples of key controlling parameters, which likely vary among faults and in time, include the difference between dynamic and static frictions on a fault (Fisher et al. 1997), conservation of elastic

stress transfer (Dahmen et al. 1998) and evolving geometrical and stress heterogeneities (Ben-Zion et al. 2003). The closeness of fitness of our model results to data may reflect the distance from criticality of the analyzed data in specific space-time domain in relation to these and other parameters.

The theory of IGW processes yields explicit distributions for various tree (i.e., cluster) statistics, including magnitude-dependent and magnitude-independent offspring number, tree size, and tree combinatorial depth. Moreover, it suggests new observed statistics based on the Horton-Strahler analysis of earthquake clusters, such as distribution of Horton-Strahler tree orders, number of branches and side-branches of different orders, and Tokunaga coefficients; see additional details in Sect. 6 and Kovchegov & Zaliapin (2020).

We show that the IGW process closely approximates a theoretical ETAS model, using a single parameter $q = \alpha/b$. The respective theoretical distributions of IGW cluster attributes provide a close fit to their empirical counterparts estimated for the crustal seismicity in southern California. The analysis suggests that the cluster statistics of the observed seismicity are well approximated by those of a critical branching process, implying that the observed seismicity operates on average close to criticality. The near criticality further reduces the model parameterization by imposing a constraint $R_0 = 1 - q$. Notably, the estimated process parameter q , and other examined statistics, are stable with respect to the lower magnitude threshold M_0 .

The rest of the paper is organized as follows. Section 2 defines the IGW process, derives the distributions of its key attributes, and discusses its invariance and attraction properties. The earthquake branching process – the main focus of this study – is introduced in Sect. 3. The practical problems of assessing process criticality in finite data, and some ways of overcoming these problems, are discussed in Sect. 4. Section 5 illustrates a close fit provided by the IGW process for selected cluster attributes estimated using the observed catalog of seismicity in southern California (Hauksson et al. 2012, extended). The results are discussed and some future research directions are outlined in Sect. 6.

2 INVARIANT GALTON-WATSON BRANCHING PROCESS

2.1 Definition

An invariant Galton-Watson (IGW) process with parameters $q \in (1/2, 1)$ and $r \in [0, 1)$ is defined as a critical Galton-Watson process with offspring distribution given by $q_1 = r$ and

$$q_k = (1 - r) \frac{(1 - q)\Gamma(k - 1/q)}{q\Gamma(2 - 1/q)k!}, \quad k = 0, 2, 3, \dots \quad (5)$$

This is a Zipf-type distribution with

$$q_k \sim Ck^{-(1+q)/q}, \quad C = (1 - r) \frac{1 - q}{q\Gamma(2 - 1/q)}. \quad (6)$$

Here the power index $(1 + q)/q$ decreases from 3 to 2 as the model parameter q increases from $q = 1/2$ to $q = 1$ (not including the boundary points). The IGW family also includes the critical binary Galton-Watson process with offspring distribution

$$\{q_0 = q_2 = \frac{1 - r}{2}, \quad q_1 = r\}, \quad (7)$$

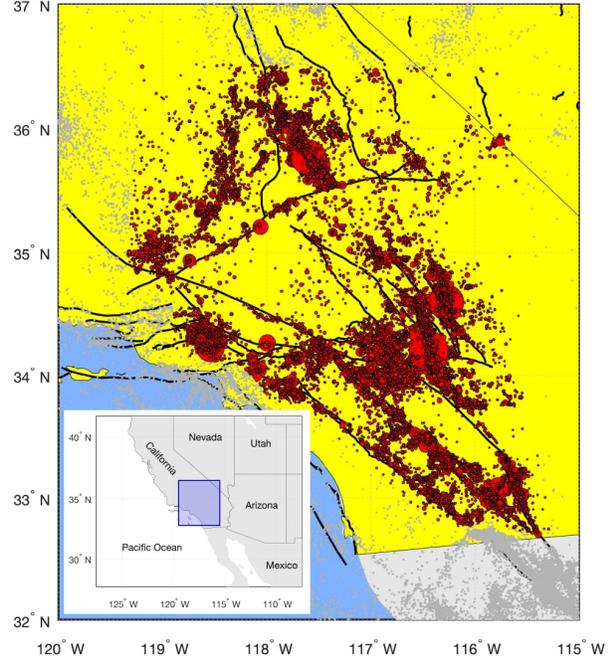


Figure 1. Seismicity of Southern California examined in this work. Earthquakes with magnitude $M \geq 2$ in Hauksson et al. (2012) catalog extended to 1981-2019 are shown by gray dots, the examined earthquakes in the central part of the catalog are shown by red circles whose size is proportional to magnitude. Black lines show the major faults.

which is the only member of the family with a finite branching.

Most concisely, one can define the IGW process as a critical Galton-Watson process whose offspring distribution $\{q_k\}$ has the generating function

$$Q(z) = \sum_{k=0}^{\infty} q_k z^k = z + (1 - r)q(1 - z)^{1/q}, \quad (8)$$

with some $q \in [1/2, 1)$ and $r \in [0, 1)$. Recall that the offspring probabilities are related to $Q(z)$ as

$$q_k = \frac{1}{k!} \left. \frac{d^k Q(z)}{dz^k} \right|_{z=0}.$$

Accordingly, the case $q = 1/2$ corresponds to the critical binary Galton-Watson distribution (7).

Sometimes it is more convenient to use the following equivalent expressions for the IGW offspring probabilities, which can be easily calculated for large k :

$$q_0 = (1 - r)q, \quad (9)$$

$$q_1 = r, \quad (10)$$

$$q_2 = (1 - r) \frac{1 - q}{2q}, \quad (11)$$

$$q_k = (1 - r) \frac{1 - q}{kq} \prod_{i=2}^{k-1} \left(1 - \frac{1}{iq}\right), \quad k \geq 3. \quad (12)$$

We observe that the parameter r only affects the distribution of linear chains – sequences of degree-2 vertices – within an IGW tree. The tree combinatorial structure – the distribution of IGW trees after removing the chains of degree-2 vertices – is completely determined by a single parameter q and remains the same for all $r \in [0, 1)$.

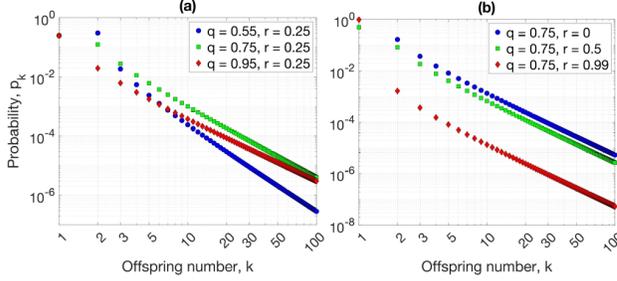


Figure 2. Examples of the IGW offspring distribution (5) with different parameters q and r . (a) Varying q for a fixed $r = 0.25$. (b) Varying r for a fixed $q = 0.75$.

Figure 2 shows several examples of the IGW offspring distribution. Each illustrated distribution has a clear power-law tail with index $(1 + q)/q$, expressed in the double-logarithmic plot as a line with slope $-(1 + q)/q$. Panel (a) illustrates that at small offspring numbers a family with a lower q (and hence a faster decaying power-law tail) has higher probabilities q_k . Panel (b) illustrates that higher values of r lead to smaller offspring probabilities q_k , $k > 1$, for the same q .

2.2 The depth and size distribution of IGW trees

Consider the generating function (8) of an IGW tree with offspring distribution $\{q_k\}$. Let $Q_0 \equiv 0$ and, for any $k \geq 1$, define the composite function

$$Q_k(z) = \underbrace{Q \circ \dots \circ Q}_{k \text{ times}}(z). \quad (13)$$

For $k = 0, 1, \dots$, let d_k denote the probability that a random IGW tree has combinatorial depth k , that is the earthquake cluster described by this tree has k generations of aftershocks. The depth distribution d_k is given by (Appendix A)

$$d_k = (1 - r)q(1 - Q_k(0))^{1/q} \quad \text{for } k = 0, 1, \dots \quad (14)$$

Let v_n , $n = 1, 2, \dots$, denote the probability that a random IGW tree has n vertices (i.e., the earthquake cluster represented by this tree is comprised of n events). Then (Appendix B)

$$v_n = \sum_{k=1}^n (-1)^{k-1} \binom{n-1}{k-1} \frac{\Gamma(k/q + 1)}{k! \Gamma(\frac{1-q}{q}k + 2)} (1-r)^k q^k. \quad (15)$$

The expression (15) involves summation of terms with very large absolute values and interchanging signs, which complicates its numerical evaluation. For example, for $r = 0$, $q = 0.75$ and $n = 30$, the maximal term in the summation is 5.5×10^9 , while the resulting value $v_n = 7.5 \times 10^{-4}$, with the ratio between these two values of the order of 10^{12} . The ratio between the largest summation terms and the final results increases with n . Say, for $n = 50$ it increases to 10^{20} . As a result, the equation (15) cannot be used (without a special precision treatment) for large n .

Appendix C shows that the tail of the cluster size distribution is approximated by

$$u_n = \sum_{k=n+1}^{\infty} v_k \sim \frac{n^{-q}}{(1-r)^q q^q \Gamma(1-q)} \quad \text{as } n \rightarrow \infty. \quad (16)$$

2.3 Invariance and attraction property

Here we informally discuss the characteristic property of the IGW processes – invariance and attraction with respect to multiple transformations of their trajectories.

Consider a Galton-Watson process with offspring distribution $\{p_k\}$ and the respective generating function $g(z) = \sum p_k z^k$. A process trajectory is a tree T , which is finite as soon as the process is critical or subcritical. Consider now a transformation of the tree T that eliminates some of its subtrees. It is often the case that the transformed tree can be represented as a trajectory of another Galton-Watson process with the new offspring distribution \tilde{p}_k and generating function $\tilde{g}(z)$ such that (Duquesne & Winkel 2019)

$$\tilde{g}(z) = z + (1 - \rho) \frac{g(z + (1 - z)a) - a - (1 - a)z}{(1 - a)(1 - g'(a))} \quad (17)$$

for some $a \in [0, 1]$ that often represents a fraction of eliminated edges, and some $\rho \in [0, 1]$. Such transformation include a continuous *erasure* of Neveu (1986) that eliminates a tree from leaves down as a constant rate; a *thinning* of Duquesne & Winkel (2007) that transforms a tree to its minimal subtree that contains the root and a set of randomly and independently selected leaves; a *Horton pruning* as in Burd et al. (2000) and Kovchegov & Zaliapin (2020, 2021) that removes all *leaf chains*, each consisting of a leaf together with the maximal possible adjacency chain of degree-2 vertices directly connected to the leaf; a *generalized dynamical pruning* of Kovchegov & Zaliapin (2020) and Kovchegov, Xu, & Zaliapin (2021) that selects an increasing function on subtrees and removes all subtrees whose value is below a threshold; and a *hereditary reduction* introduced by Duquesne & Winkel (2019).

Observe that for IGW process with generating function $Q(z)$ defined in (8) equation (17) yields

$$\begin{aligned} \tilde{Q}(z) &= z + (1 - \rho) \frac{Q(z + (1 - z)a) - a - (1 - a)z}{(1 - a)(1 - Q'(a))} \\ &= z + (1 - \rho) \frac{(1 - r)q(1 - a)^{1/q}}{(1 - a)(1 - Q'(a))} (1 - z)^{1/q} \\ &= z + (1 - \rho)q(1 - z)^{1/q}. \end{aligned} \quad (18)$$

Hence, IGW trees are invariant under a variety of thinnings, prunings, and hereditary reductions. Moreover, the parameter q stays the same under these transformations.

Furthermore, Kovchegov & Zaliapin (2021) and Kovchegov, Xu, & Zaliapin (2021) have shown that the IGW trees are the only possible attractors of critical Galton-Watson trees under the generalized dynamical pruning, subject to mild technical conditions. This means that if the transformation is applied to a (random) critical Galton-Watson tree consecutively, then the transformed tree distribution approaches that of the IGW process. An analogous result is expected to hold for other types of tree transformations.

3 EARTHQUAKE BRANCHING PROCESS

Here we describe a temporal version of the IGW model for earthquakes: A marked point process that represents earthquake occurrence times t_i and magnitudes M_i . Informally, the process consists of background events (immigrants), each

of which starts a cluster whose combinatorial part is approximated by an IGW process.

The background events are modeled by a Poisson process with intensity $\mu(t)$. Every event is assigned a magnitude $M_i \geq M_0$ independently of other events and in accordance with the Gutenberg-Richter distribution (2). Next we describe the combinatorial and temporal structure of a cluster – collection of events triggered by a single background earthquake. One can also readily introduce a spatial component in a way that is used in the existing spatial ETAS models.

3.1 Combinatorial structure of a cluster

Every event i in the process triggers N_i offspring according to a conditional offspring distribution

$$p_k(M) = \mathbb{P}(N_i = k | M_i = M),$$

which is interpreted as the probability that event with magnitude M triggers $k \geq 0$ offspring. The triggering event is called the parent with respect to its offspring. The offspring are also known as aftershocks of the first generation. Each offspring is assigned a magnitude according to (2) independently of other events (including its parent).

As in the ETAS model, the conditional average offspring number follows the Utsu scaling (Utsu et al. 1995):

$$y := \mathbb{E}[N_i | M] = R_0 10^{\alpha(M-M_0)}, \alpha > 0. \quad (19)$$

Combining this with the Gutenberg-Richter magnitude distribution (2) we obtain

$$\begin{aligned} \mathbb{E}[N_i] &= \mathbb{E}[\mathbb{E}[N_i | M_i]] \\ &= R_0 \mathbb{E} \left[10^{\alpha(M_i - M_0)} \right] = R_0 \frac{b}{b - \alpha}. \end{aligned} \quad (20)$$

The constraint $b > \alpha$ is required for a finite average progeny; the process is critical if $R_0 = (b - \alpha)/b$ and subcritical if $R_0 < (b - \alpha)/b$.

The unconditional branching probabilities p_k for a randomly selected earthquake i to have k offspring are obtained by integrating with respect to the event magnitude M_i :

$$p_k = \int_{M_0}^{\infty} p_k(M) d\mathbb{P}(M_i \leq M). \quad (21)$$

Using the magnitude distribution (2) and changing the integration variable to y of (19) we obtain

$$p_k = \frac{b}{\alpha} R_0^{\frac{b}{\alpha}} \int_{R_0}^{\infty} p_k(y) y^{-\frac{b+\alpha}{\alpha}} dy. \quad (22)$$

Hence, for each fixed conditional distribution $p_k(M)$, a cluster that starts with a random background event is a Galton-Watson process with offspring probabilities p_k parameterized by the triplet (b, α, R_0) .

Appendix D shows that commonly used conditional critical offspring distributions $p_k(M)$ – Poisson, Geometric, and Negative Binomial – correspond to an unconditional distribution p_k that can be represented as

$$p_k = q_k + \mathcal{E}_k,$$

where q_k are the IGW offspring probabilities given by (5) with $q = \alpha/b$, and \mathcal{E}_k are distribution-specific error terms such that $\mathcal{E}_k = o(q_k)$ as $k \rightarrow \infty$ and $\mathcal{E}_k \rightarrow 0$ as $q \rightarrow 1$. In other words,

the IGW process provides a close approximation to the commonly used earthquake branching model for a range of conditional offspring probabilities, and the approximation quality improves for α/b close to unity and for large offspring numbers.

Furthermore, Appendix E shows that if $p_k(M)$ is either Poisson, Geometric, or Negative Binomial critical conditional offspring distribution, then there exist $\tilde{p}_k(M)$ such that

$$\tilde{p}_k(M) \rightarrow p_k(M) \quad \text{as } q \rightarrow 1,$$

and the unconditional distribution of (22) that corresponds to $\tilde{p}_k(M)$ is the IGW distribution q_k of (5) with $q = \alpha/b$. This provides further support to an informal statement that the IGW process is a slight variation of a branching earthquake process with Poisson, Geometric, or Negative Binomial conditional offspring distribution.

3.2 Temporal structure of a cluster

Consider an earthquake i with occurrence time t_i and magnitude M_i . We assume that the offspring of event i occur according to a Cox point process, that is a Poisson process with a stochastic intensity

$$V(t | t_i, M_i) = R_0 10^{\alpha(M_i - M_0)} \ell(t - t_i) = y_i \ell(t - t_i).$$

Given a particular realization of the intensity process, the random number N_i of offspring of event i is a Poisson random variable with intensity $y_i \Lambda$, where

$$\Lambda = \int_0^{\infty} \ell(t) dt.$$

Selecting a process $\ell(s)$ such that $\mathbb{E}[\Lambda] = 1$ we ensure that

$$\mathbb{E}[N_i] = \mathbb{E}[\mathbb{E}[N_i | \Lambda]] = \mathbb{E}[y_i \Lambda] = y_i = R_0 10^{\alpha(M_i - M_0)}.$$

In particular, if Λ is a Gamma random variable with variance $1/\eta$, that is $\Lambda \sim \text{Gamma}(\eta, \eta)$, then $p_k(M)$ is a Negative Binomial distribution given by (Appendix F)

$$p_k(M_i) = \binom{k + \eta - 1}{k} \frac{(y_i/\eta)^k}{(1 + y_i/\eta)^{k+\eta}}.$$

A simple way to construct a process ℓ that satisfies the above requirements is to select some $p > 1$, $c > 0$, and $\eta > 0$ and put

$$\ell(t) = \ell_0 \frac{(p-1)c^{p-1}}{(t+c)^p}, \quad \ell_0 \sim \text{Gamma}(\eta, \eta), \quad (23)$$

where the only stochastic element is the random variable ℓ_0 used to scale the familiar Omori-Utsu offspring decay function.

4 DETECTING CRITICALITY

Branching criticality (unit average progeny) is an important property of a stochastic branching process that affects its applied statistical analysis. The IGW processes are critical, and their attraction property discussed above only applies to other critical processes. In particular, one expects that an IGW approximation to the earthquake branching process would be useful only if the latter exhibits critical-like behavior. Subcritical processes behave rather differently; for instance, similar transformations of subcritical trees converge to a trivial empty

tree. Accordingly, it is important to test criticality in observations. Such testing is affected by what we call the *curse of criticality*, which may bias assessment of criticality by using the empirical offspring numbers. Here we discuss this effect and suggest alternative ways of detecting criticality in observations.

4.1 Curse of criticality

Consider a tree with N vertices (i.e. an earthquake cluster with N events). Since each vertex, except the root, represents a triggered event and has a single parent, the total number of offspring in the tree is $N - 1$. Accordingly, the empirical average offspring number is

$$\bar{N}^{\text{off}} = \frac{\# \text{ offspring}}{\# \text{ parents}} = \frac{N - 1}{N} = 1 - \frac{1}{N}.$$

Observe that this expression is completely determined by the tree size and is independent of the actual offspring distribution. It is a standard practice to select large aftershock sequences for statistical analysis, assuming that they are most informative of the earthquake cluster properties. The above result indicates that every large aftershock sequence suggests a nearly critical branching, which can mislead the interpretation of results.

More generally, in a finite forest that consists of M trees with total of N vertices, we have

$$\bar{N}^{\text{off}} = \frac{\# \text{ offspring}}{\# \text{ parents}} = \frac{N - M}{N} = 1 - \frac{M}{N}.$$

Hence, even when working with multiple clusters, the empirical average offspring number may not be informative of the actual offspring average. This is the case when the cluster selection is subject to size constraining. For example, if one decides to only examine clusters with over 50 events, then the empirical average is bounded within $0.98 < \bar{N}^{\text{off}} < 1$.

Appendix G shows that the offspring statistics $\{k_1, \dots, k_N\}$ collected in a finite forest with M trees and N vertices have the same conditional distribution for a range of Galton-Watson processes with the actual average progeny anywhere within $(0, 1)$. Accordingly, these statistics alone may be not informative about the actual offspring probabilities.

In this work, we examine the criticality condition of earthquake trees by several alternative techniques that avoid the above curse of criticality. One of them is related to the cluster size distribution, which we discuss below.

4.2 Scaling of cluster size: A cure to the curse of criticality

The proposed branching process predicts three types of cluster size scaling that we summarize and compare in this section. Notably, these scalings are different in critical and sub-critical cases and are not affected by the curse of criticality discussed in Sect. 4.1.

First is the scaling (19) of the average number of offspring of earthquake with parent magnitude M_i . This is an exponential scaling of the form $10^{\alpha M}$ with index α . This scaling is independent of criticality or subcriticality of the process.

Second is the scaling of the average size of (number of events within) the aftershock sequence generated by a mainshock with magnitude M . Recall that mainshock is defined as

the earliest cluster event with the largest magnitude (Zaliapin & Ben-Zion 2013a). Accordingly, the aftershock sequence of a mainshock can be modeled as the earthquake branching process conditioned on having magnitudes not exceeding that of its first event. Appendix H shows that the average number of events in such a process, given that it begins with an earthquake of magnitude M , scales exponentially as 10^{bM} in a critical case and $10^{\alpha M}$ in a subcritical case, where b is the b -value of the Gutenberg-Richter law (2) and α is the exponent of the offspring scaling (19).

Third, one can consider the average size of a cluster that begins with an event of magnitude M and may include events of larger magnitudes. The first event of such a cluster produces a random number of offspring according to $p_k(M)$. Next, each of these offspring produce their own sequences that include aftershocks of all generations. Each such sequence is an independent realization of the earthquake branching process. In a subcritical case, the average size of each sequence is constant, so the final scaling is the same as that for the number of first generation aftershocks, $10^{\alpha M}$. In a critical case, the average size of each sequence is infinite, so the average cluster size is also infinite. This suggests that in a critical case the empirical cluster sizes conditioned on the magnitude of the first event should exhibit an erratic behavior.

4.3 Critical process that deviates from Utsu scaling

We have shown in (20) that the criticality of an earthquake branching process implies $R_0 = (b - \alpha)/b$. It is important to emphasize that this regularity does not affect any of our qualitative conclusions regarding the behavior of a critical process. This equality is a consequence of the explicit form of the Utsu scaling (19), and may not hold in a general critical process. Indeed, the Utsu scaling only applies to sufficiently large parent magnitudes and may not hold for $M \approx M_0$. Such deviations, however, do not affect more fundamental cluster properties, including the scaling of the size of an aftershock sequence. For example, Appendix H shows that if one considers the offspring scaling of the form

$$E[N_i | M_i = M] = R_0 10^{\alpha(M - M_0)} + o\left(10^{\alpha M}\right),$$

then the scaling of the average size of an aftershock sequence is still given by 10^{bM} . In summary – some of the process quantitative properties, for example the equality $R_0 = (b - \alpha)/\alpha$, should not be used to test for criticality in data. The focus should be on more robust characteristics, for example scaling of the cluster size.

5 ANALYSIS OF SOUTHERN CALIFORNIA SEISMICITY

We examine observed seismicity in southern California during 1981-2019 using an extended version of the catalog produced by Hauksson et al. (2012). The completeness magnitude in this catalog is known to be between $M_c = 2$ and $M_c = 3$ and may change in time and space (Hauksson et al. 2012). We examine the central region of the observational domain with higher density of stations, which does not include the offshore seismicity and earthquakes south of the Mexican border (Fig. 1). The catalog for this part has better location and magnitude quality. The lower magnitude for analysis is $M_0 = 2$. The examined

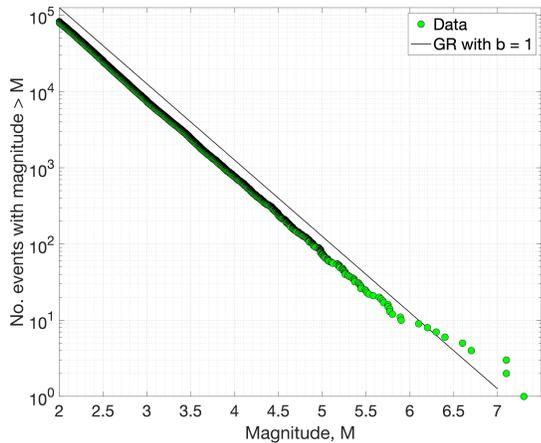


Figure 3. Magnitude distribution of the examined earthquakes (green circles) is well fit by the Gutenberg-Richter law with $b = 1$ (black line).

seismicity includes 82 716 earthquakes. The average horizontal location error for the analyzed events is 237m, the average vertical error is 928m.

The magnitude-frequency distribution of the examined seismicity (Fig. 3) is closely approximated by the Gutenberg-Richter law (2) with $b \approx 1$. Specifically, the Aki-Utsu maximum likelihood estimation is $\hat{b} = 1.003$ with a 99% confidence interval of $[0.99, 1.01]$ (Shi & Bolt 1982).

We identify (estimate) offspring and earthquake clusters using the nearest-neighbor approach of Zaliapin & Ben-Zion (2013a) that is summarized in Appendix I. The analysis is done for hypocenters with $w = 1$, $d_f = 2.6$, and $\eta_0 = 10^{-4}$. This results in 21 385 clusters with sizes varying between 1 and 14 390 earthquakes. The offspring numbers vary between 1 and 2 268.

Figures 4a,b illustrate how the empirical offspring numbers of earthquakes with magnitude $M \geq 2$ are approximated by the IGW offspring distribution (5). The fit minimizes the total variation distance between the theoretical (q_k) and empirical (\hat{q}_k) distributions with respect to the parameters (r, q):

$$\delta(q_k, \hat{q}_k) = \frac{1}{2} \sum_{k=0}^{K_0} |q_k - \hat{q}_k| \rightarrow \min. \quad (24)$$

Here K_0 denotes the maximal observed offspring number; both distributions are conditioned on this maximum number. We notice that the main contribution to the total variation distance is made by the largest probabilities, q_0 and q_1 , which, according to (9) and (10), can be used to uniquely solve for (r, q).

Overall, the IGW framework provides a reasonably close fit to the data, although there are some notable deviations that are emphasized by the c.d.f. representation of Fig. 4b. The empirical offspring distribution seems to have a slightly higher power law index at the small-to-intermediate offspring numbers $k < 100$, and a heavier tail at the largest numbers $k > 100$ than that predicted by the best fit IGW distribution with parameters $r = 0.24$ and $q = 0.86$. These deviations might be caused by a combination of the following factors. The observed proportion of earthquakes with $M \geq 6$ in the examined catalog is slightly higher than that predicted by the

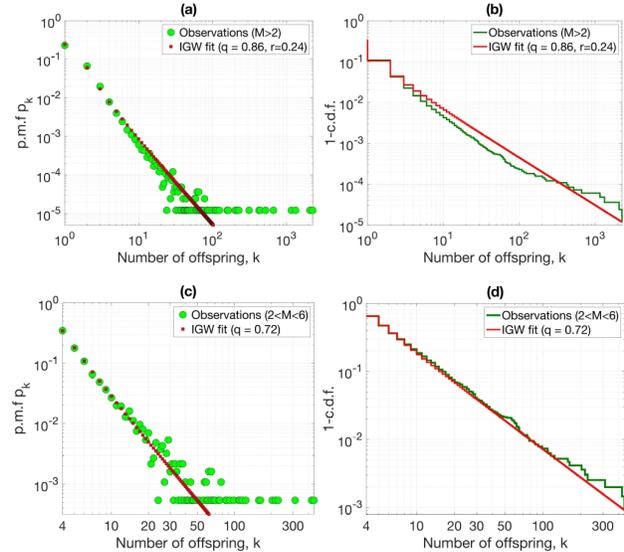


Figure 4. IGW fit to the empirical offspring numbers. (a) p.m.f. for events with $M \geq 2$; the horizontal pattern of green circles in the bottom right corner corresponds to large offspring numbers that have been only observed once in the examined catalog; (b) survival function (1-c.d.f.) for $M \geq 2$, (c) p.m.f. conditioned on $k \geq 4$ for $2 \leq M \leq 6$, (d) survival function (1-c.d.f.) conditioned on $k \geq 4$ for $2 \leq M \leq 6$. Catalog of (Hauksson et al. 2012, extended) in southern California.

Gutenberg-Richter law (Fig. 3). This is often the case, and it may be related to possible relevance of the characteristic earthquake distribution on large individual faults (Ben-Zion 2008). Hence, it is expected that the proportion of very large offspring numbers, which are associated with such events, is also higher than that in a model that assumes an exact Gutenberg-Richter law. It is also known that the clusters associated with the largest earthquakes ($M \geq 6$) include a mixture of more elementary clusters, and behave differently from the majority of clusters associated with small-to-medium magnitude events (Zaliapin & Ben-Zion 2013a). Next, the IGW approximation may not be expected to apply to very small offspring numbers. Indeed, if the observed clusters were a realization of a Galton-Watson process (not necessarily critical, or IGW), then the empirical proportions of single-event clusters (among all clusters) and events with no offspring (among all individual events) should be close, since their theoretical counterparts match exactly. This, however, is not the case in the examined data set, where the empirical proportion of single event clusters is 0.77 and the empirical proportion of earthquakes with no offspring is 0.66. The observed difference is highly significant for the large number of the examined clusters and events. Accordingly, one may assume that the observed process is a mixture of a population of very small clusters that deviate from the IGW predictions and a population that follows the IGW branching dynamics. If this is the case, the IGW approximation should apply to the empirical offspring distribution conditioned on sufficiently large offspring numbers. Figure 4c,d shows the IGW fit to the offspring numbers of earthquakes with $2 \leq M \leq 6$ conditioned on $k \geq 4$ (i.e., we eliminated events with 0, 1, 2 or 3 offspring from analysis). Such conditioning results in a very close fit.

Figure 5 shows the IGW fit of (15,16) to the empirical

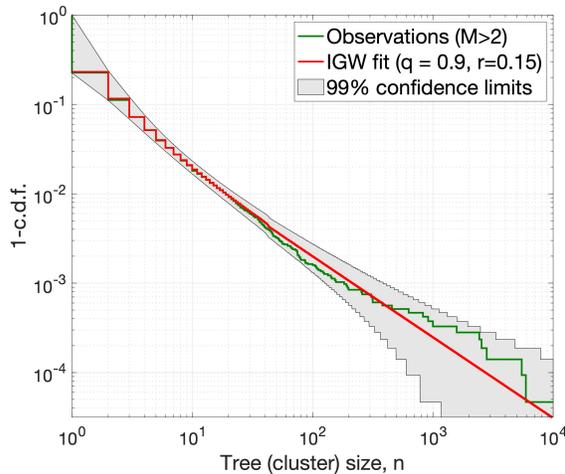


Figure 5. IGW fit to the empirical cluster sizes (number of tree vertices). The empirical survival function of cluster size (green) and its IGW fit (red). Catalog of (Hauksson et al. 2012, extended) in southern California.

cluster size distribution. We notice a very close approximation for cluster sizes $n \leq 20$ and a proper power law decay of the distribution tail. There exist slight fluctuations in the empirical tail, reminiscent of those observed in the offspring distribution of Fig. 4b; and they can be interpreted in a similar fashion. Notably, these fluctuations do not exceed the expected statistical fluctuations of a sample distribution. This suggests that the IGW process provides a useful projection of the expected cluster sizes.

The scaling of the three cluster attributes discussed in Sect. 4.2 is examined in Fig. 6. Figure 6a compares the scaling of the number of offspring of an event of magnitude M (green circles) with that of the total number of aftershocks triggered by the cluster mainshock (red squares). Both attributes are examined in catalogs with different M_0 and are shown as a function of $M - M_0$. The results exhibit a well-defined scaling for both attributes, expressed in a linear form of the plots. Notably, the scaling exponents are significantly different for the two examined attributes. The number of aftershocks (red squares) scales as $10^{b(M-M_0)}$ with $b \approx 1$, which matches the regional b -value of the Gutenberg-Richter law; see Fig. 3. The offspring numbers scale as $10^{\alpha(M-M_0)}$ with $\alpha \approx 0.83$, which is close to the IGW parameter q estimated in empirical offspring numbers; see Fig. 4a,b. The relation $q = \alpha/b$ (Sect. 3.1) indicates that the scaling of Fig. 6a is consistent with that predicted by a *critical* branching process. Furthermore, Fig. 6b illustrates the same scaling analysis for the cluster sizes averaged for a given magnitude of the first event in the cluster. The results fluctuate as expected for a near critical branching process (Sect. 4.2).

Finally, we illustrate stability of the IGW estimation. Figure 7 shows the values of the four key IGW parameters estimated with different cutoff magnitude M_0 that varies between $M_0 = 2$ and $M_0 = 4$. This 2-unit change in the cutoff magnitude corresponds to a 100-fold decrease of the total number of examined events (from 82 716 to 773) and the total number of detected clusters (from 21 385 to 236). The offspring and cluster identification was performed independently for each examined cutoff value using the nearest-neighbor technique of Zaliapin & Ben-Zion (2013a). Despite a dramatic change in the

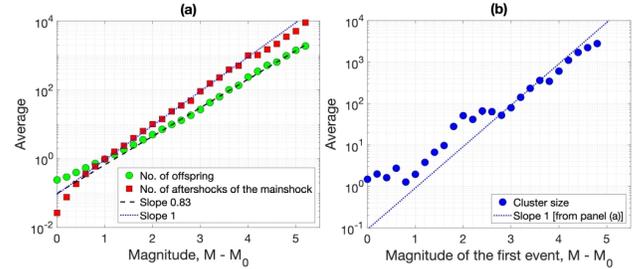


Figure 6. Cluster size scaling. (a) Average number of offspring of an earthquake of magnitude M (green) and average number of aftershocks of all generation of a mainshock of magnitude M (red) as a function of $M - M_0$. Dashed lines show the least square fits done for $2 \leq M - M_0 \leq 4$. (b) Average size of a cluster with the first event of magnitude M as a function of $M - M_0$. Dashed line is taken from panel (a) for visual comparison. Averaging is done in a sliding magnitude window of width 0.3, with a step of 0.2. The analysis uses catalogs with cutoff magnitudes $M_0 = 2.0, 2.1, \dots, 2.9, 3.0$. Catalog of (Hauksson et al. 2012, extended) in southern California.

examined catalogs, the estimated IGW parameters remain approximately the same for each cutoff magnitude, and preserve the same relative order. This indicates a structural stability of the earthquake clustering, as analyzed using the proposed IGW process.

6 DISCUSSION

We propose the IGW branching process (Sect. 2) as a model for earthquake occurrence (Sect. 3), and illustrate performance of the model with several types of analysis of seismicity from southern California (Sect. 5). The IGW process models the combinatorial part of earthquake clusters (parent-offspring relations between earthquakes), and the magnitude, time, and space earthquake attributes are added as marks of the process.

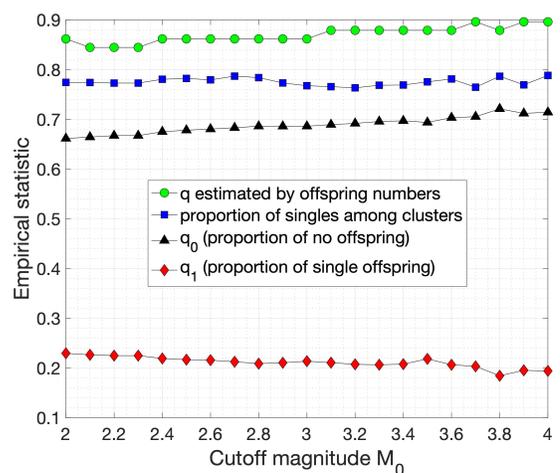


Figure 7. Stability of empirical IGW statistics with respect to the cutoff magnitude M_0 . The IGW parameter q estimated for empirical offspring numbers (green circles), proportion of singles among clusters (blue squares), proportion q_0 of events with no offspring, and proportion $q_1 = r$ of events with a single offspring as a function of the cutoff magnitude $M_0 = 2.0, 2.1, \dots, 3.9, 4.0$. Catalog of (Hauksson et al. 2012, extended) in southern California.

The key justification for the proposed framework comes from the following two properties of the IGW process. First, the IGW is theoretically shown (Appendix D) to approximate a number of branching processes with magnitude-dependent offspring generation mechanism, including the Poisson offspring distribution that is used in the ETAS model (Ogata 1988) and the negative binomial distribution that is proposed in several observational studies and in this work. Second, the IGW is a unique sub-family of Galton-Watson branching processes that is invariant with respect to numerous prunings and other operations that mimic imprecise observations and errors of cluster reconstruction in observed data (Sect. 2.3). This is also the only sub-family that can be an attractor of the tree-deforming operations applied to *any* critical Galton-Watson tree distribution (Kovchegov & Zaliapin 2021).

The main practical advantage of the proposed approach is the existence of an analytical framework that yields theoretical distributions for multiple cluster statistics. We present explicit expressions for the *offspring number* distribution q_k (5), *tree (cluster) size* distribution v_n (15), and *combinatorial tree depth* d_k (14). The theoretical IGW distributions provide a close fit to the observations in Southern California (Figs. 4, 5). The combinatorial depth distribution suggests a theoretical foundation for examining the *average leaf depth* (Zaliapin & Ben-Zion 2013b) that has been shown useful in identifying burst-like and swarm-like clusters in natural and laboratory seismicity and discriminating between natural and induced earthquakes (Reverso et al. 2015; Zaliapin & Ben-Zion 2016a,b; Davidsen et al. 2017; Bayliss et al. 2019; Martínez-Garzón et al. 2019; Peresan & Gentili 2018, 2020; Baró 2020; Kothari et al. 2020; Verdecchia et al. 2021). We also emphasize the concise specification of the IGW process that only uses two parameters r and q of (5) and stability of parameter estimation with respect to the catalog cutoff magnitude M_0 that has been illustrated in Sect. 5, Fig. 7.

The branching process framework allows one to evaluate the hypothesis of branching *criticality* without being affected by the curse of criticality (Sect. 4.1). This can be done by examining the magnitude dependence of different forms of clusters (Sect. 4.2). The results of Fig. 6 suggests that the earthquake clusters in southern California correspond to a realization of a near-critical process. We emphasize that exact criticality is not expected to exist in all space-time domains, and is also unrealizable in a finite observed system that cannot produce clusters of arbitrarily large size. Various studies suggested that large earthquakes are associated with evolving conditions that approach criticality (e.g., Main 1995; Sammis & Smith 1999; Rundle et al. 2000; Zöller & Hainzl 2002; Ben-Zion et al. 2003; Girard et al. 2010; Renard et al. 2018). Assessing deviations from criticality in different space-time domains of observed seismicity (using the discussed and additional tools) may be a useful diagnostic tool for seismicity analysis.

The IGW analyses in this study are based on a nearest-neighbor identification of parent-offspring relations between earthquakes, which is subject to errors of different types. Zaliapin & Ben-Zion (2013a) showed (their Suppl. Inf., Sect. D) that the proportion of misidentified event types (foreshock, mainshock, aftershock) in a synthetic ETAS catalog with parameters fit for southern California is about 10%, and it decreases with the cutoff magnitude M_0 . At the same time, the proportion of misidentified parent-offspring relations is about

40%, independently of the cutoff magnitude. These identification errors have the largest effect on the offspring numbers of individual events. In particular, within clusters (which comprise about 75% of the examined seismicity, with the rest 25% being single events) it is very likely to overestimate the proportion of events with no offspring (because such events have a large chance of being assigned some offspring of other events) and underestimate the proportion of events with a large number of offspring (because many of these offspring are incorrectly assigned to other events). Overall, this should inflate the relative proportion of the small offspring numbers in the offspring distribution q_k , and hence deflate the estimated parameter q . The cluster size v_k seems to be a better data source for estimating q . In agreement with this discussion, our analysis of offspring numbers (Fig. 4) suggests $\hat{q} \approx 0.72$, while analysis of cluster sizes (Fig. 5) suggests a higher estimation of $\hat{q} \approx 0.9$. A direct statistical estimation of the IGW process, e.g. via the likelihood approach, may help avoiding discrepancies caused by an independent estimation of parent-offspring relations. Such an estimation would also open a way for a direct comparison of the data fit between the IGW and other models, including ETAS, using point process residuals (Clements et al. 2011; Gordon et al. 2015) or other well-developed tests (Zechar et al. 2013). Such estimation and comparison will be a subject of a follow-up study.

Section 2.3 discusses invariance of the IGW process with respect to multiple prunings and thinnings. The Horton pruning (cutting the leaf chains) is particularly notable. The Horton prune-invariance makes the IGW a *self-similar* process (Kovchegov & Zaliapin 2020, 2021), which has several practically relevant implications. Informally, statistical properties of a self-similar process do not change after its trajectories are zoomed in or out. Such a scale change may correspond, for example, to changing the catalog cut-off magnitude M_0 . Accordingly, self-similarity ensures that the process parameters are robust with respect to M_0 , which we have illustrated in Fig. 7. Similarly, self-similarity ensures stability with respect to various data imprecisions as discussed in Sect. 2.3. Furthermore, self-similarity allows adopting multiple related tools for analysis of earthquake clusters (e.g., Holliday et al. 2008; Yoder et al. 2013). A self-similar analysis of directed tree graphs is based on Horton-Strahler orders defined as the number of Horton prunings needed to completely eliminate a tree or subtree (Horton 1945; Strahler 1957; Peckham 1995; Burd et al. 2000; Kovchegov & Zaliapin 2020; Kovchegov et al. 2021). Accordingly, a tree is decomposed into a set of linear segments (branches) of a given order, from $K = 1$ at the leaves to the largest order $K \geq 1$ at the root. Each branch represents a linear chain of earthquakes where each event triggers exactly one later event. The Horton-Strahler order K provides a logarithmic measure of a tree size N . More precisely, Horton's law of branch numbers states that the numbers N_k of order- k branches are related as $N_k/N_{k+1} \approx R_B$ and the tree size N scales with tree order K as $N \sim 10^{R_B K}$ for some *Horton exponent* $R_B \geq 2$. Scalings of the same form (called Horton's laws), with different Horton exponents, have been established for multiple other tree attributes (Rodríguez-Iturbe & Rinaldo 2001; Kovchegov et al. 2021). A rigorous treatment of Horton's laws in self-similar trees, which yields explicit values of Horton exponents, can be found in a recent survey (Kovchegov et al. 2021). The Hortonian analyses of scalings in a tree structure, and in the processes that operate on the tree, are among

the key tools in hydrogeomorphological modeling (Peckham 1995; Burd et al. 2000; Peckham & Gupta 1999; Dodds & Rothman 2000; Rodriguez-Iturbe & Rinaldo 2001; Kovchegov et al. 2021), and we expect they can be informative in analysis of seismicity.

Applying the self-similar analyses to the trees that describe earthquake clusters may advance quantification of regional earthquake clustering style (e.g., finding additional discriminants between induced and natural earthquakes or between environments with different temperature-fluid conditions). An additional important application is analysis of potential premonitory cluster attributes in zones showing localization of seismicity (Ben-Zion & Zaliapin 2020), as part of an effort to track preparation processes leading to large earthquakes. Specific quantities to explore include time- and space-dependent branch numbers N_K (the number of branches of order K in a tree), branch magnitudes M_K (the average number of leaves in a tree of order K), branch lengths (combinatorial or in years), side-branch numbers $N_{i,j}$ (the number of branches of order i that merge with branches of order j , $j > i$), and Tokunaga coefficients $T_{i,j}$ (the average number of branches of order i that merge with a random branch of order j , $j > i$). The theory of self-similar trees predicts a number of power-law relations between distinct tree attributes, with specific values of power exponents (Kovchegov et al. 2021). The famous Hack's law in hydrology (Hack 1957; Rodriguez-Iturbe & Rinaldo 2001) that relates the length L of the longest river stream to the basin area A via $L \sim A^h$ with $h \approx 0.6$ is an example of such a power law. It would be interesting to explore and interpret similar relations in earthquake data, examining cluster duration, cluster area, cluster size, and total cluster seismic moment.

Kovchegov & Zaliapin (2021) established that in an IGW process with parameter q , we have $T_{i,i+k} = T_k = ac^{k-1}$ for any $i \geq 1$ and $k = 2, 3, \dots$ with $c = 1/(1-q)$ and $a = (c-1)(c^{1/(c-1)} - 1)$. Moreover, the Horton exponent is given by $R_B = (1-q)^{-1/q}$. These results provide an alternative, and probably more robust, way of estimating the parameter q from observations, by inverting R_B and $T_{i,j}$. Recall that the parameter q combines the two key seismicity exponents, $q = \alpha/b$, so its robust estimation has immediate practical implications for analysis of seismicity. The above expressions also may help evaluating deviations of the observed process from a pure IGW.

The proposed IGW modeling framework can readily include a spatial component, by assigning offspring locations according to a space kernel centered at the parent (e.g., Ogata 1988, 1998, 1999). A fuller space-time model will be developed in a follow-up study.

ACKNOWLEDGMENTS

We thank two anonymous referees for constructive comments that helped to improve the manuscript. The study was supported by the National Science Foundation (Grants DMS-1412557, EAR-1723033, EAR-2122191, EAR-1722561, EAR-2122168) and the Southern California Earthquake Center (based on NSF Cooperative Agreement EAR-1600087 and USGS Cooperative Agreement G17 AC00047).

DATA AVAILABILITY

The Hauksson et al. (2012) catalog of southern California earthquakes, extended to 2019, is available via the Southern California Earthquake Data Center: <https://scedc.caltech.edu/index.html>

REFERENCES

- Abramowitz, M. & Stegun, I. A., 1964. *Handbook of mathematical functions with formulas, graphs, and mathematical tables*, vol. 55, US Government printing office.
- Adamopoulos, L., 1976. Cluster models for earthquakes: Regional comparisons, *Journal of the International Association for Mathematical Geology*, **8**(4), 463–475.
- Baiesi, M. & Paczuski, M., 2004. Scale-free networks of earthquakes and aftershocks, *Physical review E*, **69**(6), 066106.
- Baró, J., 2020. Topological properties of epidemic aftershock processes, *Journal of Geophysical Research: Solid Earth*, **125**(5), e2019JB018530.
- Bayliss, K., Naylor, M., & Main, I. G., 2019. Probabilistic identification of earthquake clusters using rescaled nearest neighbour distance networks, *Geophysical Journal International*, **217**(1), 487–503.
- Ben-Zion, Y., 2008. Collective behavior of earthquakes and faults: Continuum-discrete transitions, progressive evolutionary changes, and different dynamic regimes, *Reviews of Geophysics*, **46**(4).
- Ben-Zion, Y. & Zaliapin, I., 2020. Localization and coalescence of seismicity before large earthquakes, *Geophysical Journal International*, **223**(1), 561–583.
- Ben-Zion, Y., Eneva, M., & Liu, Y., 2003. Large earthquake cycles and intermittent criticality on heterogeneous faults due to evolving stress and seismicity, *Journal of Geophysical Research: Solid Earth*, **108**(B6).
- Burd, G. A., Waymire, E. C., & Winn, R. D., 2000. A self-similar invariance of critical binary Galton-Watson trees, *Bernoulli*, pp. 1–21.
- Clements, R. A., Schoenberg, F. P., & Schorlemmer, D., 2011. Residual analysis methods for space-time point processes with applications to earthquake forecast models in California, *The Annals of Applied Statistics*, **5**(4), 2549–2571.
- Dahmen, K., Ertas, D., & Ben-Zion, Y., 1998. Gutenberg-richter and characteristic earthquake behavior in simple mean-field models of heterogeneous faults, *Physical Review E*, **58**(2), 1494.
- Daley, D. J. & Vere-Jones, D., 2003. *An introduction to the theory of point processes. Volume I: Elementary theory and methods*, Springer.
- Davidson, J., Kwiatek, G., Charalampidou, E.-M., Goebel, T., Stanchits, S., Rück, M., & Dresen, G., 2017. Triggering processes in rock fracture, *Physical review letters*, **119**(6), 068501.
- Dodds, P. & Rothman, D., 2000. Scaling, universality, and geomorphology, *Ann Rev. Earth and Planet Sci.*, **28**, 571–610.
- Duquesne, T. & Winkel, M., 2007. Growth of Lévy trees, *Probability Theory and Related Fields*, **139**(3-4), 313–371.
- Duquesne, T. & Winkel, M., 2019. Hereditary tree growth and Lévy forests, *Stochastic Processes and their Applications*, **129**(10), 3690–3747.
- Field, E. H., Jordan, T. H., Page, M. T., Milner, K. R., Shaw, B. E., Dawson, T. E., Biasi, G. P., Parsons, T., Hardebeck, J. L., Michael, A. J., et al., 2017. A synoptic view of the third uniform California earthquake rupture forecast (UCERF3), *Seismological Research Letters*, **88**(5), 1259–1267.
- Fisher, D. S., Dahmen, K., Ramanathan, S., & Ben-Zion, Y., 1997. Statistics of earthquakes in simple models of heterogeneous faults, *Physical review letters*, **78**(25), 4885.
- García, D., Wald, D. J., & Hearne, M., 2012. A global earthquake discrimination scheme to optimize ground-motion prediction equa-

- tion selection, *Bulletin of the Seismological Society of America*, **102**(1), 185–203.
- Girard, L., Amitrano, D., & Weiss, J., 2010. Failure as a critical phenomenon in a progressive damage model, *Journal of Statistical Mechanics: Theory and Experiment*, **2010**(01), P01013.
- Gordon, J. S., Clements, R. A., Schoenberg, F. P., & Schorlemmer, D., 2015. Voronoi residuals and other residual analyses applied to csep earthquake forecasts, *Spatial Statistics*, **14**, 133–150.
- Greenwood, M. & Yule, G. U., 1920. An inquiry into the nature of frequency distributions representative of multiple happenings with particular reference to the occurrence of multiple attacks of disease or of repeated accidents, *Journal of the Royal statistical society*, **83**(2), 255–279.
- Gutenberg, B. & Richter, C. F., 1944. Frequency of earthquakes in California, *Bulletin of the Seismological Society of America*, **34**(4), 185–188.
- Hack, J. T., 1957. *Studies of longitudinal stream profiles in Virginia and Maryland*, vol. 294, US Government Printing Office.
- Hauksson, E., Yang, W., & Shearer, P. M., 2012. Waveform relocated earthquake catalog for southern California (1981 to June 2011), *Bulletin of the Seismological Society of America*, **102**(5), 2239–2244.
- Hawkes, A. G., 1971. Spectra of some self-exciting and mutually exciting point processes, *Biometrika*, **58**(1), 83–90.
- Hawkes, A. G. & Oakes, D., 1974. A cluster process representation of a self-exciting process, *Journal of Applied Probability*, **11**(3), 493–503.
- Holliday, J. R., Turcotte, D. L., & Rundle, J. B., 2008. Self-similar branching of aftershock sequences, *Physica A: Statistical Mechanics and its Applications*, **387**(4), 933–943.
- Horton, R. E., 1945. Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology, *Bulletin of Geophysical Society of America*, **56**, 275–370.
- Kagan, Y. & Knopoff, L., 1976. Statistical search for non-random features of the seismicity of strong earthquakes, *Physics of the earth and planetary interiors*, **12**(4), 291–318.
- Kagan, Y. Y., 1973. A probabilistic description of the seismic regime, *Izv. Acad. Sci. USSR, Phys. Solid Earth*, **213**, 219.
- Kagan, Y. Y., 2010. Statistical distributions of earthquake numbers: Consequence of branching process, *Geophysical Journal International*, **180**(3), 1313–1328.
- Kothari, S., Shcherbakov, R., & Atkinson, G., 2020. Statistical modeling and characterization of induced seismicity within the Western Canada Sedimentary Basin, *Journal of Geophysical Research: Solid Earth*, **125**(12), e2020JB020606.
- Kovchegov, Y. & Zaliapin, I., 2020. Random self-similar trees: A mathematical theory of Horton laws, *Probability Surveys*, **17**, 1–213.
- Kovchegov, Y. & Zaliapin, I., 2021. Invariance and attraction properties of Galton-Watson trees, *Bernoulli*, **27**(3), 1789–1823.
- Kovchegov, Y., Zaliapin, I., & Fofoula-Georgiou, E., 2021. Random self-similar trees: Emergence of scaling laws, *Surveys in Geophysics* (accepted).
- Main, I. G., 1995. Earthquakes as critical phenomena: implications for probabilistic seismic hazard analysis, *Bulletin of the Seismological Society of America*, **85**(5), 1299–1308.
- Martínez-Garzón, P., Ben-Zion, Y., Zaliapin, I., & Bohnhoff, M., 2019. Seismic clustering in the Sea of Marmara: Implications for monitoring earthquake processes, *Tectonophysics*, **768**, 228176.
- Musmeci, F. & Vere-Jones, D., 1992. A space-time clustering model for historical earthquakes, *Annals of the Institute of Statistical Mathematics*, **44**(1), 1–11.
- Neveu, J., 1986. Erasing a branching tree, *Advances in applied probability*, pp. 101–108.
- Ogata, Y., 1983. Estimation of the parameters in the modified omori formula for aftershock frequencies by the maximum likelihood procedure, *Journal of Physics of the Earth*, **31**(2), 115–124.
- Ogata, Y., 1988. Statistical models for earthquake occurrences and residual analysis for point processes, *Journal of the American Statistical Association*, **83**(401), 9–27.
- Ogata, Y., 1998. Space-time point-process models for earthquake occurrences, *Annals of the Institute of Statistical Mathematics*, **50**(2), 379–402.
- Ogata, Y., 1999. Seismicity analysis through point-process modeling: A review, *Seismicity patterns, their statistical significance and physical meaning*, pp. 471–507.
- Ogata, Y. & Katsura, K., 1986. Point-process models with linearly parametrized intensity for application to earthquake data, *Journal of applied probability*, **23**(A), 291–310.
- Ogata, Y. & Katsura, K., 1988. Likelihood analysis of spatial inhomogeneity for marked point patterns, *Annals of the Institute of Statistical Mathematics*, **40**(1), 29–39.
- Ogata, Y. & Vere-Jones, D., 1984. Inference for earthquake models: a self-correcting model, *Stochastic processes and their applications*, **17**(2), 337–347.
- Ogata, Y., Akaike, H., & Katsura, K., 1982. The application of linear intensity models to the investigation of causal relations between a point process and another stochastic process, *Annals of the Institute of Statistical Mathematics*, **34**(2), 373–387.
- Omori, F., 1894. On the after-shocks of earthquakes, *J. Coll. Sci., Imp. Univ., Japan*, **7**, 111–200.
- Page, M. T., Van Der Elst, N., Hardebeck, J., Felzer, K., & Michael, A. J., 2016. Three ingredients for improved global aftershock forecasts: Tectonic region, time-dependent catalog incompleteness, and intersequence variability, *Bulletin of the Seismological Society of America*, **106**(5), 2290–2301.
- Peckham, S. D., 1995. New results for self-similar trees with applications to river networks, *Water Resources Research*, **31**(4), 1023–1029.
- Peckham, S. D. & Gupta, V. K., 1999. A reformulation of Horton's laws for large river networks in terms of statistical self-similarity, *Water Resources Research*, **35**(9), 2763–2777.
- Peresan, A. & Gentili, S., 2018. Seismic clusters analysis in Northeastern Italy by the nearest-neighbor approach, *Physics of the Earth and Planetary Interiors*, **274**, 87–104.
- Peresan, A. & Gentili, S., 2020. Identification and characterisation of earthquake clusters: a comparative analysis for selected sequences in Italy and adjacent regions., *Bollettino di Geofisica Teorica ed Applicata*, **61**(1).
- Renard, F., Weiss, J., Mathiesen, J., Ben-Zion, Y., Kandula, N., & Cordonnier, B., 2018. Critical evolution of damage toward system-size failure in crystalline rock, *Journal of Geophysical Research: Solid Earth*, **123**(2), 1969–1986.
- Reverso, T., Marsan, D., & Helmstetter, A., 2015. Detection and characterization of transient forcing episodes affecting earthquake activity in the Aleutian Arc system, *Earth and Planetary Science Letters*, **412**, 25–34.
- Rodriguez-Iturbe, I. & Rinaldo, A., 2001. *Fractal river basins: chance and self-organization*, Cambridge University Press.
- Rundle, J. B., Klein, W., Turcotte, D. L., & Malamud, B. D., 2000. Precursory seismic activation and critical-point phenomena, in *Microscopic and Macroscopic Simulation: Towards Predictive Modelling of the Earthquake Process*, pp. 2165–2182, Springer.
- Saichev, A., Helmstetter, A., & Sornette, D., 2005. Power-law distributions of offspring and generation numbers in branching models of earthquake triggering, *Pure and Applied Geophysics*, **162**(6), 1113–1134.
- Sammis, C. G. & Smith, S. W., 1999. Seismic cycles and the evolution of stress correlation in cellular automaton models of finite fault networks, in *Seismicity Patterns, their Statistical Significance and Physical Meaning*, pp. 307–334, Springer.
- Shi, Y. & Bolt, B. A., 1982. The standard error of the magnitude-frequency b value, *Bulletin of the Seismological Society of America*, **72**(5), 1677–1687.
- Strahler, A. N., 1957. Quantitative analysis of watershed geomorphology, *Trans. Am Geophys. Un.*, **38**, 913–920.

- Utsu, T., 1970. Aftershocks and earthquake statistics (1): Some parameters which characterize an aftershock sequence and their interrelations, *Journal of the Faculty of Science, Hokkaido University. Series 7, Geophysics*, **3**(3), 129–195.
- Utsu, T., Ogata, Y., & Matsuura, R., 1995. The centenary of the omori formula for a decay law of aftershock activity, *Journal of Physics of the Earth*, **43**(1), 1–33.
- Veen, A. & Schoenberg, F. P., 2008. Estimation of space–time branching process models in seismology using an EM–type algorithm, *Journal of the American Statistical Association*, **103**(482), 614–624.
- Verdecchia, A., Cochran, E. S., & Harrington, R. M., 2021. Fluid-earthquake and earthquake-earthquake interactions in southern Kansas, USA, *Journal of Geophysical Research: Solid Earth*, **126**(3), e2020JB020384.
- Vere-Jones, D., 1970. Stochastic models for earthquake occurrence, *Journal of the Royal Statistical Society: Series B (Methodological)*, **32**(1), 1–45.
- Vere-Jones, D., 1976. A branching model for crack propagation, *pure and applied geophysics*, **114**(4), 711–725.
- Vere-Jones, D., 1978. Earthquake prediction—a statistician’s view, *Journal of Physics of the Earth*, **26**(2), 129–146.
- Vere-Jones, D. & Ozaki, T., 1982. Some examples of statistical estimation applied to earthquake data, *Annals of the Institute of Statistical Mathematics*, **34**(1), 189–207.
- Wang, Q., Schoenberg, F. P., & Jackson, D. D., 2010. Standard errors of parameter estimates in the etas model, *Bulletin of the Seismological Society of America*, **100**(5A), 1989–2001.
- Yoder, M. R., Aalsburg, J. V., Turcotte, D. L., Abaimov, S. G., & Rundle, J. B., 2013. Statistical variability and Tokunaga branching of aftershock sequences utilizing BASS model simulations, *Pure and Applied Geophysics*, **170**(1-2), 155–171.
- Zaliapin, I. & Ben-Zion, Y., 2013a. Earthquake clusters in southern California i: Identification and stability, *Journal of Geophysical Research: Solid Earth*, **118**(6), 2847–2864.
- Zaliapin, I. & Ben-Zion, Y., 2013b. Earthquake clusters in southern California ii: Classification and relation to physical properties of the crust, *Journal of Geophysical Research: Solid Earth*, **118**(6), 2865–2877.
- Zaliapin, I. & Ben-Zion, Y., 2016a. Discriminating characteristics of tectonic and human-induced seismicity, *Bulletin of the Seismological Society of America*, **106**(3), 846–859.
- Zaliapin, I. & Ben-Zion, Y., 2016b. A global classification and characterization of earthquake clusters, *Geophysical Journal International*, **207**(1), 608–634.
- Zechar, J. D., Schorlemmer, D., Liukis, M., Yu, J., Euchner, F., Maechling, P. J., & Jordan, T. H., 2010. The laboratory for the study of earthquake predictability perspective on computational earthquake science, *Concurrency and Computation: Practice and Experience*, **22**(12), 1836–1847.
- Zechar, J. D., Schorlemmer, D., Werner, M. J., Gerstenberger, M. C., Rhoades, D. A., & Jordan, T. H., 2013. Regional earthquake likelihood models i: First-order results, *Bulletin of the Seismological Society of America*, **103**(2A), 787–798.
- Zöller, G. & Hainzl, S., 2002. A systematic spatiotemporal test of the critical point hypothesis for large earthquakes, *Geophysical research letters*, **29**(11), 53–1.

APPENDIX A: DEPTH DISTRIBUTION OF IGW TREES

Consider the generating function (8) of an IGW tree with offspring distribution $\{q_k\}$ and recall the definition of a compos-

ite generating function Q_k of (13). Let $D_{-1} = 0$ and

$$D_k = \sum_{j=0}^k d_j \text{ for } k = 0, 1, \dots$$

Then $D_k = Q(D_{k-1})$ and therefore, $D_k = Q_{k+1}(0)$. Hence,

$$\begin{aligned} d_k &= D_k - D_{k-1} = Q_{k+1}(0) - Q_k(0) \\ &= Q(Q_k(0)) - Q_k(0) = (1-r)q(1-Q_k(0))^{1/q}. \end{aligned}$$

APPENDIX B: SIZE DISTRIBUTION OF IGW TREES

Recall that v_n , $n = 1, 2, \dots$, denotes the probability that a random IGW tree has n vertices. Let V_k be the distribution of the total size of k independent trees, that is the sum of k independent identically distributed random variables with common distribution $\{v_n\}$:

$$V_k = \underbrace{v * \dots * v}_{k \text{ times}}$$

Observe that since the number of vertices equals 1 plus the number of vertices in all subtrees branching from the root, we have

$$v_{n+1} = \sum_{k=n}^{\infty} q_k V_k(n).$$

Therefore, the generating function $v(z) = \sum_{n=1}^{\infty} z^n v_n$ satisfies

$$\begin{aligned} v(z) &= zQ(v(z)) \\ &= z \left(v(z) + (1-r)q(1-v(z))^{1/q} \right). \end{aligned} \quad (\text{B.1})$$

Letting $v = v(z)$ and $w = (1-r)qz(1-z)^{-1}$, we find $w = v(1-v)^{-1/q}$. Next, we find $v(z)$ by inverting the function $w(v)$. The Lagrange Inversion Theorem (Abramowitz & Stegun 1964) states that

$$v(z) = \sum_{k=1}^{\infty} \frac{d^{k-1}}{dv^{k-1}} \left(\frac{v}{w(v)} \right)^k \frac{w^k}{k!}.$$

Since

$$\begin{aligned} \frac{d^{k-1}}{dv^{k-1}} \left(\frac{v}{w(v)} \right)^k &= \frac{d^{k-1}}{dv^{k-1}} (1-v)^{k/q} \\ &= (-1)^{k-1} (k/q)(k/q-1) \dots (k/q-k+2) \\ &= (-1)^{k-1} \frac{\Gamma(k/q+1)}{\Gamma(k/q-k+2)}, \end{aligned}$$

the Lagrange Inversion Theorem implies

$$v(z) = \sum_{k=1}^{\infty} (-1)^{k-1} \frac{\Gamma(k/q+1)}{\Gamma(\frac{1-q}{q}k+2)} \frac{w^k}{k!}, \quad (\text{B.2})$$

where

$$\begin{aligned} w^k &= (1-r)^k q^k \frac{z^k}{(1-z)^k} \\ &= (1-r)^k q^k \sum_{n=k}^{\infty} \binom{n-1}{k-1} z^n. \end{aligned} \quad (\text{B.3})$$

Substituting (B.3) into (B.2) yields the following expression for $v(z)$:

$$\begin{aligned} & \sum_{k=1}^{\infty} \sum_{n=k}^{\infty} (-1)^{k-1} \binom{n-1}{k-1} \frac{\Gamma(k/q+1)}{k! \Gamma(\frac{1-q}{q}k+2)} (1-r)^k q^k z^n \\ &= \sum_{n=1}^{\infty} z^n \sum_{k=1}^n (-1)^{k-1} \binom{n-1}{k-1} \frac{\Gamma(k/q+1)}{k! \Gamma(\frac{1-q}{q}k+2)} (1-r)^k q^k. \end{aligned}$$

This implies the identity (15), as $v(z) = \sum_{n=1}^{\infty} z^n v_n$.

APPENDIX C: TAIL OF THE IGW TREE SIZE DISTRIBUTION

Here we find the asymptotic expression for the tail of the IGW tree size distribution

$$u_n = \sum_{k=n+1}^{\infty} v_k, \quad n = 0, 1, \dots$$

Notice that the generating functions $v(z)$ and $u(z) = \sum_{n=0}^{\infty} u_n z^n$ satisfy the following relation

$$\begin{aligned} 1 - v(z) &= \sum_{k=1}^{\infty} (1 - z^k) v_k = (1 - z) \sum_{k=1}^{\infty} \sum_{n=0}^{k-1} z^n v_k \\ &= (1 - z) \sum_{n=0}^{\infty} \sum_{k=n+1}^{\infty} v_k z^n \\ &= (1 - z) \sum_{n=0}^{\infty} u_n z^n = (1 - z)u(z). \end{aligned}$$

Therefore, by (B.1), we have

$$\begin{aligned} (1 - z)v(z) &= (1 - r)q(1 - v(z))^{1/q} \\ &= (1 - r)q(1 - z)^{1/q}u(z)^{1/q}. \end{aligned}$$

Thus,

$$u(z) = \frac{1}{(1 - z)^{1-q}} \frac{(v(z))^q}{(1 - r)^q q^q},$$

where

$$\lim_{z \rightarrow 1^-} \frac{(v(z))^q}{(1 - r)^q q^q} = \frac{1}{(1 - r)^q q^q}.$$

Hence, by the Tauberian theorem for power series,

$$u_n \sim \frac{1}{(1 - r)^q q^q \Gamma(1 - q)} n^{-q}$$

as the sequence $u_n \geq 0$ is monotone non-increasing.

APPENDIX D: IGW APPROXIMATION TO EARTHQUAKE BRANCHING PROCESS

This section shows that the combinatorial part of a critical earthquake branching process with Negative Binomial, Geometric (a special case of Negative Binomial), or Poisson conditional offspring distribution $p_k(M)$ is closely approximated by the IGW. Throughout this section, we suppose

$$\frac{1}{2} < q = \frac{\alpha}{b} < 1$$

and assume criticality, i.e. $R_0 = 1 - q$.

D1 Negative Binomial distribution

For a given $\eta > 0$, consider

$$\begin{aligned} p_k(M) &= \mathbf{P}(N_i = k | M_i = M) \\ &= \binom{k + \eta - 1}{k} \frac{(y/\eta)^k}{(1 + y/\eta)^{k+\eta}} \quad k = 0, 1, 2, \dots \end{aligned}$$

where $y = R_0 10^{\alpha(M - M_0)} = (1 - q) 10^{\alpha(M - M_0)}$.

For $k = 2, 3, \dots$ we have

$$\begin{aligned} p_k &= \frac{b}{\alpha} R_0^{b/\alpha} \binom{k + \eta - 1}{k} \int_{R_0}^{\infty} \frac{\eta^\eta y^k}{(\eta + y)^{k+\eta}} y^{-\frac{\alpha+b}{\alpha}} dy \\ &= \frac{1}{q} (1 - q)^{1/q} \binom{k + \eta - 1}{k} \int_{1-q}^{\infty} \frac{\eta^\eta y^{k-1/q-1}}{(\eta + y)^{k+\eta}} dy \\ &= \eta^{-1/q} \frac{1}{q} (1 - q)^{1/q} \binom{k + \eta - 1}{k} \int_{(1-q)/\eta}^{\infty} \frac{y^{k-1/q-1}}{(1 + y)^{k+\eta}} dy \\ &= \eta^{-1/q} \frac{1}{q} (1 - q)^{1/q} \binom{k + \eta - 1}{k} \int_0^{\infty} \frac{y^{k-1/q-1}}{(1 + y)^{k+\eta}} dy - \mathcal{E}_k, \end{aligned}$$

with the error term

$$\begin{aligned} \mathcal{E}_k &= \eta^{-1/q} \frac{1}{q} (1 - q)^{1/q} \binom{k + \eta - 1}{k} \int_0^{(1-q)/\eta} \frac{y^{k-1/q-1}}{(1 + y)^{k+\eta}} dy \\ &< \eta^{-1/q} \frac{1}{q} (1 - q)^{1/q} \binom{k + \eta - 1}{k} \int_0^{(1-q)/\eta} y^{k-1/q-1} dy \\ &= \frac{\eta^{-k}}{kq - 1} \binom{k + \eta - 1}{k} (1 - q)^k. \end{aligned} \quad (\text{D.1})$$

Recall that

$$B(x, y) = \int_0^{\infty} \frac{t^{x-1}}{(1 + t)^{x+y}} dt$$

and hence

$$\begin{aligned} \int_0^{\infty} \frac{y^{k-1/q-1}}{(1 + y)^{k+\eta}} dy &= B\left(k - \frac{1}{q}, \eta + \frac{1}{q}\right) \\ &= \frac{\Gamma(k - 1/q) \Gamma(\eta + 1/q)}{\Gamma(k + \eta)}. \end{aligned}$$

Hence, for $k = 2, 3, \dots$ we have

$$p_k = (1 - r) \frac{(1 - q) \Gamma(k - 1/q)}{q \Gamma(2 - 1/q) k!} - \mathcal{E}_k, \quad (\text{D.2})$$

where the first term on the right hand side is the IGW offspring probability q_k of (5) and

$$r = 1 - (1 - q)^{1/q-1} \frac{\Gamma(2 - 1/q) \Gamma(\eta + 1/q)}{\eta^{1/q} \Gamma(\eta)}. \quad (\text{D.3})$$

It also can be shown that

$$p_1 = r + \mathcal{E}_1, \quad (\text{D.4})$$

with

$$\mathcal{E}_1 = (\eta + 1)(1 - q) \int_{\eta}^{\infty} \frac{w^{1/q-1} - 1}{(1 - q + w)^3} w dw$$

where $w = (1 - q)/y$. In particular,

$$\begin{aligned} \mathcal{E}_1 &\leq \left(1 + q \frac{\eta^{1/q-1} - 1}{1 - q}\right) \frac{\eta + 1}{\eta(2q - 1)} (1 - q)^2 \\ &\approx (1 + \ln \eta) \frac{\eta + 1}{\eta(2q - 1)} (1 - q)^2, \quad \text{when } q \approx 1, \end{aligned}$$

Since both branching distributions, p_k and q_k , add up to one, we have

$$p_0 = (1 - r)q - \mathcal{E}_0,$$

where

$$\mathcal{E}_0 = \mathcal{E}_1 - \sum_{k=2}^{\infty} \mathcal{E}_k.$$

The total variation distance between distributions $\{p_k\}$ and $\{q_k\}$ with $q = \frac{\alpha}{b}$ and r in (D.3) equals $\delta(p_k, q_k) = \mathcal{E}_1$. For example, when $\eta = 1$ (Geometric offspring distribution) and $q = 0.92$, we have $\mathcal{E}_1 = 0.0144223 \dots$

Next, we establish a similar approximation result assuming that parameter η depends on the productivity y . Suppose that $\eta = \eta(y)$ is a decreasing function converging rapidly to $\eta_0 > 1$, so that η is indistinguishable from η_0 for $y \geq y_0$, where y_0 is sufficiently small. Then, for $k = 2, 3, \dots$, we have

$$\begin{aligned} p_k &= \frac{b}{\alpha} R_0^{b/\alpha} \int_{R_0}^{\infty} \binom{k + \eta - 1}{k} \frac{\eta^\eta y^k}{(\eta + y)^{k+\eta}} y^{-\frac{\alpha+b}{\alpha}} dy \\ &\approx \frac{1}{q} R_0^{1/q} \binom{k + \eta_0 - 1}{k} \int_{y_0}^{\infty} \frac{\eta_0^{\eta_0} y^{k-1/q-1}}{(\eta_0 + y)^{k+\eta_0}} dy \\ &= \frac{1}{q} R_0^{1/q} \binom{k + \eta_0 - 1}{k} \int_{y_0/\eta_0}^{\infty} \frac{y^{k-1/q-1}}{(1 + y)^{k+\eta_0}} dy \\ &\approx \frac{1}{q} R_0^{1/q} \binom{k + \eta_0 - 1}{k} \int_0^{\infty} \frac{y^{k-1/q-1}}{(1 + y)^{k+\eta_0}} dy \\ &= (1 - r) \frac{(1 - q)\Gamma(k - 1/q)}{q\Gamma(2 - 1/q)k!} = q_k \end{aligned}$$

with

$$r = 1 - R_0^{1/q} \frac{\Gamma(2 - 1/q)\Gamma(\eta + 1/q)}{(1 - q)\eta_0^{1/q}\Gamma(\eta)}.$$

D2 Poisson distribution

Consider

$$p_k(M) = \frac{1}{k!} y^k e^{-y} \quad k = 0, 1, 2, \dots$$

where $y = R_0 10^{\alpha(M - M_0)} = (1 - q) 10^{\alpha(M - M_0)}$.

If $q = \frac{\alpha}{b}$ is sufficiently close to 1, or $k = 2, 3, \dots$ we

have

$$\begin{aligned} p_k &= \frac{b}{\alpha} R_0^{b/\alpha} \int_{R_0}^{\infty} \frac{1}{k!} y^k e^{-y} y^{-\frac{\alpha+b}{\alpha}} dy \\ &= \frac{1}{q} (1 - q)^{1/q} \frac{1}{k!} \int_{1-q}^{\infty} e^{-y} y^{k-1/q-1} dy \\ &= \frac{1}{q} (1 - q)^{1/q} \frac{1}{k!} \int_0^{\infty} e^{-y} y^{k-1/q-1} dy - \mathcal{E}_k, \end{aligned}$$

where the error term

$$\begin{aligned} \mathcal{E}_k &= \frac{1}{q} (1 - q)^{1/q} \frac{1}{k!} \int_0^{1-q} e^{-y} y^{k-1/q-1} dy \\ &< \frac{1}{q} (1 - q)^{1/q} \frac{1}{k!} \int_0^{1-q} y^{k-1/q-1} dy \\ &= \frac{1}{k!q} (1 - q)^k. \end{aligned} \tag{D.5}$$

Therefore, from

$$\Gamma(x) = \int_0^{\infty} e^{-t} t^{x-1} dt,$$

we obtain

$$\begin{aligned} p_k &= \frac{1}{q} (1 - q)^{1/q} \frac{1}{k!} \int_0^{\infty} e^{-y} y^{k-1/q-1} dy - \mathcal{E}_k \\ &= \frac{1}{q} (1 - q)^{1/q} \frac{\Gamma(k - 1/q)}{k!} - \mathcal{E}_k \\ &= (1 - q)^{1/q-1} \Gamma(2 - 1/q) \frac{(1 - q)\Gamma(k - 1/q)}{q\Gamma(2 - 1/q)k!} - \mathcal{E}_k. \end{aligned}$$

Hence, for $k = 2, 3, \dots$,

$$p_k = (1 - r) \frac{(1 - q)\Gamma(k - 1/q)}{q\Gamma(2 - 1/q)k!} - \mathcal{E}_k, \tag{D.6}$$

where the first term in the right hand side equals to the IGW offspring probability q_k and

$$r = 1 - (1 - q)^{1/q-1} \Gamma(2 - 1/q). \tag{D.7}$$

It also can be shown that

$$p_1 = 1 - (1 - q)^{1/q-1} \Gamma(2 - 1/q) + \mathcal{E}_1,$$

where the error term

$$\begin{aligned} \mathcal{E}_1 &= \int_0^{1-q} e^{-y} [(1 - q)^{1/q-1} y^{1-1/q} - 1] dy \\ &< \int_0^{1-q} [(1 - q)^{1/q-1} y^{1-1/q} - 1] dy \\ &= \frac{1}{2q - 1} (1 - q)^2. \end{aligned} \tag{D.8}$$

Since both branching distributions, p_k and q_k , add up to one, we have

$$p_0 = (1 - r)q - \mathcal{E}_0,$$

where

$$\begin{aligned} \mathcal{E}_0 &= \mathcal{E}_1 - \sum_{k=2}^{\infty} \mathcal{E}_k \\ &< \frac{1}{2q-1} (1-q)^2 - \frac{1}{q} (1 - (2-q)e^{q-1}). \end{aligned} \quad (\text{D.9})$$

The total variation distance between distributions $\{p_k\}$ and $\{q_k\}$ with $q = \frac{\alpha}{b}$ and r in (D.7) equals

$$\begin{aligned} \delta(p_k, q_k) &= \mathcal{E}_1 \\ &= e^{q-1} - 1 + (1-q)^{1/q-1} \int_0^{1-q} e^{-y} y^{1-1/q} dy. \end{aligned}$$

For example, when $q = 0.92$, the total variation distance equals $\mathcal{E}_1 = 0.0074761 \dots$

APPENDIX E: EXACT IGW REPRESENTATION OF EARTHQUAKE BRANCHING PROCESS

This section introduces two conditional offspring distributions $p_k(M)$ that correspond to an exact IGW process. This means that the unconditional offspring distribution p_k of (21) that corresponds to $p_k(M)$ coincides with the IGW offspring distribution q_k of (5). The two distributions considered here are close to the Negative Binomial and Poisson.

Throughout this section we assume $1/2 < q = \alpha/b < 1$ and criticality, i.e. $R_0 = 1 - q$. We also consider productivity

$$y = r_0 10^{\alpha m} > r_0 10^{\alpha m_0} = R_0 = 1 - q,$$

and a constant r such that $0 < r < 1$.

E1 Quasi-negative binomial offspring distribution

Consider

$$\begin{aligned} p_k(M) &= \frac{(1-r)\eta^{\eta+1/q} \Gamma(\eta)}{(1-q)^{1/q-1} \Gamma(2-1/q) \Gamma(\eta+1/q)} \times \\ &\quad \binom{k+\eta-1}{k} \frac{(q-1+y)^{k-1/q-1}}{(\eta+q-1+y)^{k+\eta}} y^{1+1/q} \end{aligned} \quad (\text{E.1})$$

for $k = 2, 3, \dots$. For all $k \geq 2$, the unconditional offspring probabilities p_k are given by

$$\begin{aligned} p_k &= \frac{b}{\alpha} R_0^{b/\alpha} \frac{(1-r)\eta^{\eta+1/q} \Gamma(\eta)}{(1-q)^{1/q-1} \Gamma(2-1/q) \Gamma(\eta+1/q)} \times \\ &\quad \binom{k+\eta-1}{k} \int_{R_0}^{\infty} \frac{(q-1+y)^{k-1/q-1}}{(\eta+q-1+y)^{k+\eta}} y^{1+1/q} y^{-\frac{\alpha+b}{\alpha}} dy \\ &= \frac{(1-r)(1-q)\eta^{\eta+1/q} \Gamma(\eta)}{q \Gamma(2-1/q) \Gamma(\eta+1/q)} \times \\ &\quad \binom{k+\eta-1}{k} \int_{R_0}^{\infty} \frac{(q-1+y)^{k-1/q-1}}{(\eta+q-1+y)^{k+\eta}} dy \\ &= \frac{(1-r)(1-q) \Gamma(\eta)}{q \Gamma(2-1/q) \Gamma(\eta+1/q)} \times \\ &\quad \binom{k+\eta-1}{k} \int_0^{\infty} \frac{w^{k-1/q-1}}{(1+w)^{k+\eta}} dw \\ &= (1-r) \frac{(1-q) \Gamma(k-1/q)}{q \Gamma(2-1/q) k!} = q_k, \end{aligned}$$

where we let $w = (q-1+y)/\eta$.

We observe that the distribution (E.1) approaches negative binomial as $q \rightarrow 1$.

E2 Quasi-Poisson offspring distribution

Consider

$$\begin{aligned} p_k(M) &= (1-r) \frac{(1-q)^{1-1/q}}{\Gamma(2-1/q)} e^{-y+1-q} \times \\ &\quad \frac{(q-1+y)^{k-1/q-1}}{k!} y^{1+1/q} \quad k = 2, 3, \dots \end{aligned} \quad (\text{E.2})$$

For all $k \geq 2$ we have

$$\begin{aligned} p_k &= (1-r) \frac{(1-q)^{1-1/q}}{\Gamma(2-1/q)} \frac{b}{\alpha} R_0^{b/\alpha} \times \\ &\quad \int_{R_0}^{\infty} e^{-y+1-q} \frac{(q-1+y)^{k-1/q-1}}{k!} y^{1+1/q} y^{-\frac{\alpha+b}{\alpha}} dy \\ &= (1-r) \frac{(1-q)}{q \Gamma(2-1/q)} \int_0^{\infty} e^{-w} w^{k-1/q-1} dw \\ &= (1-r) \frac{(1-q) \Gamma(k-1/q)}{q \Gamma(2-1/q) k!}, \end{aligned}$$

where we let $w = q-1+y$.

Observe that we need r to be large enough so that

$$\begin{aligned} \sum_{k=2}^{\infty} p_k(M) &= (1-r) \frac{(1-q)^{1-1/q}}{\Gamma(2-1/q)} \times \\ &\quad (1 - e^{-y+1-q} (q+y)) (q-1+y)^{-1-1/q} y^{1+1/q} < 1 \end{aligned}$$

for all $y > 1 - q$. In this case, there are multiple choices for $p_0(M)$ and $p_1(M)$ such that $p_0 = (1-r)q$, $p_1 = r$, and

$$p_0(M) + p_1(M) = 1 - \sum_{k=2}^{\infty} p_k(M).$$

APPENDIX F: NEGATIVE BINOMIAL DISTRIBUTION AS A POISSON WITH GAMMA INTENSITY

The following result goes back to Greenwood & Yule (1920). Let the number N_i of offspring of earthquake i be a Poisson random variable with intensity $y\Lambda$, where $y = R_0 10^{\alpha(M_i - M_0)}$. Assume that $\Lambda \sim \text{Gamma}(\eta, \eta)$ with some $\eta > 0$. Then, for each integer $k \geq 0$,

$$\begin{aligned} p_k(M) &= P(N_i = k \mid M_i = M) \\ &= \int_0^{\infty} \frac{y^k x^k}{k!} e^{-yx} \frac{1}{\Gamma(\eta)} \eta^\eta x^{\eta-1} e^{-\eta x} dx \\ &= \frac{(y/\eta)^k}{(1+y/\eta)^{k+\eta}} \frac{\int_0^{\infty} (\eta+y)^{k+\eta} x^{k+\eta-1} e^{-(\eta+y)x} dx}{k! \Gamma(\eta)} \\ &= \frac{(y/\eta)^k}{(1+y/\eta)^{k+\eta}} \frac{\Gamma(k+\eta)}{k! \Gamma(\eta)} \\ &= \binom{k+\eta-1}{k} \frac{(y/\eta)^k}{(1+y/\eta)^{k+\eta}}, \end{aligned}$$

which is a Negative Binomial distribution with parameters η and $(1 + y/\eta)^{-1}$.

APPENDIX G: CONDITIONAL EQUIVALENCE

Consider a subcritical Galton-Watson process with offspring distribution $\{p_k\}$ and generating function $g(z) = \sum p_k z^k$. Assume that a finite forest generated by this process is observed and we record the number M of trees in the forest and the total number N of vertices. Observe that the maximal branching number in the observed forest cannot exceed $N - M$. The process that generated this forest cannot be distinguished from a subcritical Galton-Watson process with a finite offspring distribution

$$\tilde{p}_k = \frac{p_k}{\sum_{k=1}^{N-M} p_k}, \quad k = 0, \dots, N - M.$$

Formally, conditioned on the observed statistics (M, N) , the distribution of the forest is the same for the process with offspring distributions $\{p_k\}$ and $\{\tilde{p}_k\}$.

Next, we construct a critical Galton-Watson process with the same conditional distribution. For that we fix $w > 1$ and define offspring probabilities

$$q_k = \frac{\tilde{p}_k w^k}{g(w)}, \quad k = 0, \dots, N - M, \quad (\text{G.1})$$

where $g(z) = \sum z^k \tilde{p}_k$ is the generating function of $\{\tilde{p}_k\}$; it exists on $z \in \mathbb{R}$ because of the finiteness of the distribution support. The offspring probabilities q_k are well defined since

$$\sum_{k=0}^{N-M} q_k = \frac{1}{g(w)} \sum_{k=0}^{N-M} \tilde{p}_k w^k = \frac{g(w)}{g(w)} = 1.$$

Let X be a random variable with probability mass function \tilde{p}_k . Then, $E[X] < 1$, and the Fortuin-Kasteleyn-Ginibre (FKG) inequality yields

$$E[X w^X] - E[w^X]E[X] = \text{Cov}(X, w^X) > 0$$

as w^x is a strictly increasing function. Hence,

$$\sum_{k=0}^{N-M} k q_k = \frac{w g'(w)}{g(w)} = \frac{E[X w^X]}{E[w^X]} > E[X].$$

Observe that the transformation (G.1) is a multiplicative semi-group. Specifically, if we denote by $G(w) : \mathcal{P} \rightarrow \mathcal{P}$ the respective transformation on a space \mathcal{P} of finite offspring distributions

$$\mathcal{P} = \{p_k, k = 1, \dots, K : \sum_{k=1}^K p_k = 1\},$$

then

$$G(w_1 \cdot w_2) = G(w_1) \circ G(w_2).$$

Thus, the expected progeny in the branching process with branching probabilities q_k

$$\sum_{k=0}^{N-M} k q_k = \frac{\sum_{k=0}^{N-M} k w^k \tilde{p}_k}{\sum_{k=0}^{N-M} w^k \tilde{p}_k}$$

is a strictly increasing continuous function of w , which converges to $N - M$ as $w \rightarrow \infty$. Hence, as soon as $N - M > 1$,

that is the forest includes at least one tree with more than 2 vertices, there exists $w > 1$ such that the process with offspring probabilities q_k is critical, i.e., $\sum_{k=0}^K k q_k = 1$.

It is left to establish the conditional equivalence. Fix a pair (M, N) such that $N > M$ and consider the probability of observing a forest with M trees, N vertices, and offspring numbers $\{k_1, \dots, k_N\}$ such that $k_1 + \dots + k_N = N - M$ in a Galton-Watson process with offspring distribution $\{q_k\}$ of (G.1):

$$P(\{k_1, \dots, k_N\}) = \prod_{k=1}^N q_{k_i} = \frac{w^{N-M}}{g(w)^N} \prod_{k=1}^N \tilde{p}_{k_i}. \quad (\text{G.2})$$

This probability differs from its counterpart in the process with offspring distribution $\{\tilde{p}_k\}$ by a constant term that depends on w, M and N . Accordingly, the probabilities conditioned on (M, N) coincide.

APPENDIX H: AVERAGE SIZE OF AN AFTERSHOCK SEQUENCE

Consider an earthquake branching process that begins with an earthquake of magnitude M and condition this process on all aftershocks having magnitude smaller than M . There is an average of $R_0 10^{\alpha(M-M_0)}$ offspring of the first event, each generating a subtree of aftershocks. These subtrees are independent and identically distributed. In every subtree, the magnitudes are exponential random variables, conditioned on being less than M . Thus, the average progeny of an event in a subtree equals

$$\begin{aligned} & \frac{b \ln 10}{1 - 10^{-b\Delta}} \int_{M_0}^M R_0 10^{\alpha(x-M_0)} 10^{-b(x-M_0)} dx \\ &= R_0 \frac{b}{b-\alpha} \frac{1 - 10^{-(b-\alpha)\Delta}}{1 - 10^{-b\Delta}}, \end{aligned}$$

where $\Delta = M - M_0$. Therefore, the mean size of each subtree T_i is equal to

$$\begin{aligned} & \sum_{k=0}^{\infty} \left(R_0 \frac{b}{b-\alpha} \frac{1 - 10^{-(b-\alpha)\Delta}}{1 - 10^{-b\Delta}} \right)^k \\ &= \frac{1 - 10^{-b\Delta}}{1 - R_0 \frac{b}{b-\alpha} + R_0 \frac{b}{b-\alpha} 10^{-(b-\alpha)\Delta} - 10^{-b\Delta}} \end{aligned}$$

Hence, the average size of a tree generated by earthquake of magnitude M equals

$$1 + R_0 10^{\alpha\Delta} \frac{1 - 10^{-b\Delta}}{1 - R_0 \frac{b}{b-\alpha} + R_0 \frac{b}{b-\alpha} 10^{-(b-\alpha)\Delta} - 10^{-b\Delta}}.$$

In a subcritical case, the fraction in the above expression converges to $(1 - R_0 b/(b-a))^{-1}$ as M increases, and hence the average cluster size scales as $10^{\alpha M}$. In a critical case we have $R_0 = \frac{b-\alpha}{b}$, and the average size of a tree generated by an earthquake of magnitude M event is

$$\begin{aligned} & 1 + R_0 10^{\alpha\Delta} \frac{1 - 10^{-b\Delta}}{10^{\alpha\Delta} - 1} 10^{b\Delta} \\ &= 1 + R_0 \frac{1 - 10^{-b\Delta}}{1 - 10^{-\alpha\Delta}} 10^{b\Delta}, \end{aligned}$$

which increases as 10^{bM} .

Assume now that the Utsu scaling (19) only holds for the sufficiently large magnitudes. For example, assume that

$$E[N_i | M_i = M] = R_0 10^{\alpha(M-M_0)} + C_0 10^{\kappa(M-M_0)},$$

with $\kappa < \alpha$. Assume criticality:

$$R_0 \frac{b}{b-\alpha} + C_0 \frac{b}{b-\kappa} = 1.$$

Then, the average progeny for an event in a subtree equals

$$\begin{aligned} & \frac{b \ln 10}{1 - 10^{-b\Delta}} \int_{m_0}^m (R_0 10^{\alpha\Delta_x} + C_0 10^{\kappa\Delta_x}) 10^{-b\Delta_x} dx \\ &= R_0 \frac{b}{b-\alpha} \frac{1 - 10^{-(b-\alpha)\Delta}}{1 - 10^{-b\Delta}} + C_0 \frac{b}{b-\kappa} \frac{1 - 10^{-(b-\kappa)\Delta}}{1 - 10^{-b\Delta}}, \end{aligned}$$

where $\Delta_x = x - M_0$. Thus, the mean size of each subtree T_i equals

$$\begin{aligned} & \sum_{k=0}^{\infty} \left(R_0 \frac{b}{b-\alpha} \frac{1 - 10^{-(b-\alpha)\Delta}}{1 - 10^{-b\Delta}} + C_0 \frac{b}{b-\kappa} \frac{1 - 10^{-(b-\kappa)\Delta}}{1 - 10^{-b\Delta}} \right)^k \\ &= \frac{1 - 10^{-b\Delta}}{R_0 \frac{b}{b-\alpha} 10^{-(b-\alpha)\Delta} + C_0 \frac{b}{b-\kappa} 10^{-(b-\kappa)\Delta} - 10^{-b\Delta}} \end{aligned}$$

Hence, the average size of the tree generated by an earthquake of magnitude M equals

$$1 + R_0 \frac{1 - 10^{-b\Delta}}{R_0 \frac{b}{b-\alpha} + C_0 \frac{b}{b-\kappa} 10^{-(\alpha-\kappa)\Delta} - 10^{-\alpha\Delta}} 10^{b\Delta}.$$

This shows that low-magnitude deviations from the Utsu scaling (19) do not affect the general scaling of the average size of an aftershock sequence, which is still given by 10^{bM} .

APPENDIX I: NEAREST-NEIGHBOR ANALYSIS OF EARTHQUAKES

Consider a catalog of earthquakes where each event i is characterized by its occurrence time t_i , hypocenter \mathbf{x}_i , and magnitude M_i . The proximity of earthquake j to previous earthquake i is asymmetric in time and is formally defined as (Baiesi & Paczuski 2004; Zaliapin & Ben-Zion 2013a)

$$\eta_{i,j} = \begin{cases} t_{i,j} r_{i,j}^{d_f} 10^{-wM_i}, & t_{i,j} > 0, \\ \infty, & t_{i,j} \leq 0, \end{cases} \quad (\text{I.1})$$

where $t_{i,j} = t_j - t_i$ is the event interoccurrence time, which is positive if event j occurred after event i ; $r_{i,j} \geq 0$ is the spatial distance between the earthquake hypocenters; d_f is the fractal dimension of the hypocenters; and w is the parameter that introduces exponential weight of the earlier event i by its magnitude M_i .

The nearest neighbor (parent) for event j is the event i that minimizes the proximity $\eta_{i,j}$. The events to which event i is the parent are called *offspring* of i . The nearest-neighbor proximity (proximity to the parent) of the event j is denoted by η_j . Connecting each event to its parent creates a spanning tree over the examined earthquakes. Removing the weak parent links, defined by the condition $\eta_j \geq \eta_0$ for a given $\eta_0 > 0$, partitions the spanning tree into individual *earthquake clusters*. Clusters that consist of a single earthquake are called *singles*. We refer to Zaliapin & Ben-Zion (2013a,b) for a detailed description of the nearest-neighbor methodology.