J. Appl. Probab. **60,** 367–381 (2023)
*doi:*10.1017/jpr.2022.52

ON EXTREMES OF RANDOM CLUSTERS AND MARKED RENEWAL CLUSTER PROCESSES

BOJAN BASRAK,* **
NIKOLINA MILINČEVIĆ,* *** AND
PETRA ŽUGEC,**** *University of Zagreb*

Abstract

This article describes the limiting distribution of the extremes of observations that arrive in clusters. We start by studying the tail behaviour of an individual cluster, and then we apply the developed theory to determine the limiting distribution of $\max\{X_j : j = 0, \ldots, K(t)\}$, where $K(t)$ is the number of independent and identically distributed observations (X_j) arriving up to the time t according to a general marked renewal cluster process. The results are illustrated in the context of some commonly used Poisson cluster models such as the marked Hawkes process.

Keywords: Renewal cluster processes; Poisson cluster processes; Hawkes process; maximal claim size; extreme value distributions; random maxima; limit theorems

2020 Mathematics Subject Classification: Primary 60G70
Secondary 60G55; 91G30

1. Introduction

In many real-life situations one encounters observations which tend to cluster when collected over time. This behaviour is commonly seen in various applied fields, including, for instance, non-life insurance, climatology, and hydrology (see e.g. [24], [30], [29]). This article aims to describe the limiting distribution for the extremes of such observations over increasing time intervals.

In Section 2 we study a simpler question concerning the tail behaviour of the maximum in one random cluster of observations. More precisely, consider

$$H = \bigvee_{j=1}^{K} X_j,$$

where we assume that the sequence (X_j) of independent and identically distributed (i.i.d.) random variables belongs to the maximum domain of attraction of some extreme value distribution

Received 22 September 2021; revision received 8 April 2022.
* Postal address: Department of Mathematics, University of Zagreb, Bijenička 30, Zagreb, Croatia.
** Email address: bbasrak@math.hr
*** Email address: nikolina.milincevic@math.hr
**** Postal address: Faculty of Organization and Informatics, University of Zagreb, Pavlinska 2, Varaždin, Croatia. Email address: petra.zugec@foi.hr

G, or MDA(G) for short, and K is a positive random integer, possibly dependent on the observations themselves. For an introduction to MDAs and extreme value theory in general we refer to [26], [14], or [13]. In the case of non-random K, H belongs to the same MDA as X_1 by standard extreme value theory. The case of K independent of the sequence (X_j) has been subject of several studies, including [20] and [28]; see also [15], where the tail behaviour of the randomly indexed sums is studied in a similar setting. The same problem in a multidimensional setting has recently been considered in [18]. In the sequel, we allow, for instance, for K to be a stopping time with respect to the sequence (X_j), and we show that H remains in the same MDA as the observations as long as K has a finite mean. This is the content of our main theorem in Section 2. For this result we provide an original and relatively simple proof based on [5].

In Section 3 we consider observations (X_j) which are i.i.d. but arrive in possibly overlapping groups at times $\tau_1, \tau_2, \ldots$. We show how one can determine the asymptotic distribution of $M(t) = \sup\{X_k : \tau_k \le t\}$ under certain mild conditions on the clustering among the observations. Thanks to the results in Section 2, it turns out that the effects of clustering often remain relatively small in the limit; cf. Corollary 3.1. Processes of the form $M(t) = \bigvee_{j=0}^{K(t)} X_j$, where $K(t)$ is a stochastic process possibly dependent on the observations X_j, have received considerable attention over the years. For some of the earliest contributions see [8] and [2]. More recently, [23] and [25] studied the convergence of the process $(M(t))$ towards an appropriate extremal process. For the study of all upper order statistics up to time $K(t)$, see [5], and for the more general weak convergence of extremal processes with a random sample size, see [27].

Section 4 is dedicated to the application of our main results to some frequently used stochastic models of clustering. In particular, we study variants of Neyman–Scott, Bartlett–Lewis, and randomly marked Hawkes processes. For each of the three clustering mechanisms we find sufficient conditions which imply that $M(t)$ properly centred and normalized, roughly speaking, stays in MDA(G).

Throughout, let $\mathbb{S}$ denote a general Polish space and $\mathcal{B}(\mathbb{S})$ a Borel σ-algebra on $\mathbb{S}$. The space of boundedly finite point measures on $\mathbb{S}$ is denoted by $M_p(\mathbb{S})$. For this purpose $\mathbb{S}$ is endowed with a family of so-called *bounded* sets; see [3]. We use the standard vague topology on the space $M_p(\mathbb{S})$ (see [26] or [21]). Recall that $m_n \xrightarrow{v} m$ in $M_p(\mathbb{S})$ simply means that $\int f dm_n \longrightarrow \int f dm$ for any bounded continuous function $f : \mathbb{S} \to \mathbb{R}$ whose support is *bounded* in the space $\mathbb{S}$.

The Lebesgue measure on $[0, \infty)$ will be denoted by Leb, whereas the Poisson random measure with mean measure η will be denoted by PRM(η). To simplify the notation, for a generic member of an identically distributed sequence or an array, say (X_j), $(A_{i,j})$, throughout we write X, A, etc. The set of natural numbers will be denoted by $\mathbb{N} = \{1, 2, \ldots\}$. The set of non-negative integers we denote by $\mathbb{Z}_+$.

2. Random maxima

Let $(X_j)_{j\in\mathbb{N}}$ be an i.i.d. sequence with distribution belonging to MDA(G) where G is one of the three extreme value distributions, and let K denote a random non-negative integer. We are interested in the tail behaviour of

$$H = \bigvee_{j=1}^{K} X_j.$$

In the sequel we allow for K to depend on the values of the sequence $(X_j)_{j\in\mathbb{N}}$ together with some additional sources of randomness. Assume that $((W_j, X_j))_{j\in\mathbb{N}}$ is a sequence of i.i.d. random elements in $\mathbb{S} \times \mathbb{R}$. For the filtration $(\mathcal{F}_n)_{n\in\mathbb{N}} = (\sigma\{(W_j, X_j) : j \le n\})_{n\in\mathbb{N}}$ we assume that K is a stopping time with respect to $(\mathcal{F}_n)_{n\in\mathbb{N}}$. Already in this case, H can be a rather complicated distribution, as one can see from the following.

Example 2.1.

(a) Assume $(W_j)_{j\in\mathbb{N}}$ is independent of $(X_j)_{j\in\mathbb{N}}$ and integer-valued. When $K = W_1$, H has been studied already in the references mentioned in the introduction.

(b) Assume $((W_j, X_j))_{j\in\mathbb{N}}$ is i.i.d. as before (note that some mutual dependence between W_j and X_j is allowed) and $\mathbb{P}(X > W) > 0$. Let $K = \inf\{k \in \mathbb{N} : X_k > W_k\}$. Clearly K has geometric distribution, and we will show that this implies that H is in the same MDA as X.

(c) Assume $(W_j)_{j\in\mathbb{N}}$ and $(X_j)_{j\in\mathbb{N}}$ are two independent i.i.d. sequences. Let $K = \inf\{k \in \mathbb{N} : X_k > W_1\}$. Clearly $H = X_K > W_1$. Therefore, H has a tail at least as heavy as W.

Recall (see Chapter 1 in [26] by Resnick) that the assumption that X belongs to MDA(G) is equivalent to the existence of a sequence of positive real numbers $(a_n)_{n\in\mathbb{N}}$ and a sequence of real numbers $(b_n)_{n\in\mathbb{N}}$ such that for every $x \in \mathbb{E} = \{y \in \mathbb{R} : G(y) > 0\}$

$$n \cdot \mathbb{P}(X > a_n \cdot x + b_n) \to -\log G(x) \quad \text{as } n \to \infty, \tag{2.1}$$

and it is further equivalent to

$$\mathbb{P}\left(\frac{\bigvee_{i=1}^n X_i - b_n}{a_n} \le x\right) \to G(x) \quad \text{as } n \to \infty.$$

We denote by μ_G the measure $\mu_G(x, \infty) = -\log G(x)$, $x \in \mathbb{E}$. Consider point processes

$$N_n = \sum_{i\in\mathbb{N}} \delta_{\left(\frac{i}{n}, \frac{X_i - b_n}{a_n}\right)}, \quad n \in \mathbb{N}.$$

It is well known (again again [26]) that $X \in \text{MDA}(G)$ is both necessary and sufficient for weak convergence of N_n towards a limiting point process, N say, which is a PRM(Leb $\times\, \mu_G$) in $M_p([0, \infty) \times \mathbb{E})$, where both $\mathbb{E}$ and the concept of *boundedness* depend on G. For instance, in the Gumbel MDA, $\mathbb{E} = (-\infty, \infty)$, and sets are considered *bounded* in $[0, \infty) \times \mathbb{E}$ if contained in some set of the type $[0, T] \times (a, \infty)$, $a \in \mathbb{R}$, $T > 0$; cf. [4].

Denote by $m|_A$ the restriction of a point measure m to a set A, i.e. $m|_A(B) = m(A \cap B)$. Denote by $\mathbb{E}'$ an arbitrary measurable subset of $\mathbb{R}^d$. The following simple lemma (see Lemma 1 in [5]) plays an important role in a couple of our proofs.

Lemma 2.1. *Assume that N, $(N_t)_{t\ge 0}$ are point processes with values in $M_p([0, \infty) \times \mathbb{E}')$. Assume further that Z, $(Z_t)_{t\ge 0}$ are $\mathbb{R}_+$-valued random variables. If $P(N(\{Z\} \times \mathbb{E}') > 0) = 0$ and $(N_t, Z_t) \xrightarrow{d} (N, Z)$, in the product topology as $t \to \infty$, then*

$$N_t|_{[0,Z_t]\times\mathbb{E}'} \xrightarrow{d} N|_{[0,Z]\times\mathbb{E}'} \quad \text{as } t \to \infty.$$

Suppose that the stopping time K is almost surely finite. Our analysis of H depends on the following simple observation: since $((W_j, X_j))_{j\in\mathbb{N}}$ is an i.i.d. sequence, by the strong Markov property, after the stopping time $K_1 = K$, the sequence $((W_{K_1+j}, X_{K_1+j}))_{j\in\mathbb{N}}$ has the same distribution as the original sequence. Therefore it has its own stopping time K_2, distributed as K_1,

https://doi.org/10.1017/jpr.2022.52 Published online by Cambridge University Press

such that $((W_{K_1+K_2+j}, X_{K_1+K_2+j}))_{j\in\mathbb{N}}$ again has the same distribution. Using the shift operator ϑ, one can also write $K_2 = K \circ \vartheta^{K_1}(((W_j, X_j))_j)$. Applying this argument iteratively, we can break the original sequence into i.i.d. blocks

$$((W_{T(l-1)+1}, X_{T(l-1)+1}),\ (W_{T(l-1)+2}, X_{T(l-1)+2}),\ \ldots,\ (W_{T(l)}, X_{T(l)}))_{l\in\mathbb{N}},$$
$$\text{where} \quad T(0)=0, \quad T(n) = K_1 + K_2 + \cdots + K_n.$$

Clearly,

$$H_l = \bigvee_{j=T(l-1)+1}^{T(l)} X_j, \qquad l \in \mathbb{N},$$

are i.i.d. with the same distribution as the original compound maximum H. Assume that $((W_{i,j}, X_{i,j}))_{i,j\in\mathbb{N}}$ is an i.i.d. array of elements as above, and let $(K_i')_{i\in\mathbb{N}}$ be an i.i.d. sequence of stopping times such that for each $l \in \mathbb{N}$, $(K_l', (W_{l,j}, X_{l,j})_{j\in\mathbb{N}}) \stackrel{d}{=} (K, (W_j, X_j)_{j\in\mathbb{N}})$. Then

$$H_l' = \bigvee_{j=1}^{K_l'} X_{l,j}$$

are also i.i.d. with the same distribution as H. Before stating the main theorem, we prove a simple lemma.

Lemma 2.2. *Assume that $\xi = \mathbb{E}[K] < \infty$. Then*

$$\sum_{i=1}^{n} \sum_{j=1}^{K_i'} \delta_{\frac{X_{i,j} - b_{\lfloor n\xi \rfloor}}{a_{\lfloor n\xi \rfloor}}} \xrightarrow{d} \mathrm{PRM}(\mu_G) \quad \textit{as } n \to \infty.$$

Proof. First note that

$$\sum_{i=1}^{n} \sum_{j=1}^{K_i'} \delta_{\frac{X_{i,j} - b_{\lfloor n\xi \rfloor}}{a_{\lfloor n\xi \rfloor}}} \stackrel{d}{=} \sum_{i=1}^{T(n)} \delta_{\frac{X_i - b_{\lfloor n\xi \rfloor}}{a_{\lfloor n\xi \rfloor}}}.$$

To use Lemma 2.1, let $Z = 1$, $(Z_n)_{n\in\mathbb{N}} = (T(n)/(n\xi))_{n\in\mathbb{N}}$ be $\mathbb{R}_+$-valued random variables, $N = \mathrm{PRM}(\mathrm{Leb} \times \mu_G)$ as before, and define point processes $(N_n')_{n\in\mathbb{N}}$, where

$$N_n' = \sum_{i\in\mathbb{N}} \delta_{\left(\frac{i}{n\xi}, \frac{X_i - b_{\lfloor n\xi \rfloor}}{a_{\lfloor n\xi \rfloor}}\right)},$$

with values in the space $[0, \infty) \times \mathbb{E}$, where $\mathbb{E}$ depends on G as before. By the weak law of large numbers and by Proposition 3.21 from [26], since $X_1 \in \mathrm{MDA}(G)$, we have

$$Z_n \xrightarrow{P} Z = 1 \quad \text{and} \quad N_n' \xrightarrow{d} N \quad \text{as } n \to \infty.$$

Hence, by the standard Slutsky argument (Theorem 3.9 in [9]),

$$(N_n',\ Z_n) \xrightarrow{d} (N,\ Z) \quad \text{as } n \to \infty.$$

Note that $\mathbb{P}(N(\{Z\} \times \mathbb{E}) > 0) = 0$, so by Lemma 2.1,

$$N_n'\Big|_{[0,Z_n]\times\mathbb{E}} \xrightarrow{d} N\Big|_{[0,Z]\times\mathbb{E}}.$$

https://doi.org/10.1017/jpr.2022.52 Published online by Cambridge University Press

We conclude that

$$N'_n\Big|_{\left[0,\frac{T(n)}{n\xi}\right]\times\mathbb{E}}([0,\infty)\times\cdot)=\sum_{i=1}^{T(n)}\delta_{\frac{X_i-b_{\lfloor n\xi\rfloor}}{a_{\lfloor n\xi\rfloor}}}(\cdot)\xrightarrow{d}N\Big|_{[0,1]\times\mathbb{E}}([0,\infty)\times\cdot)\quad\text{as } n\to\infty,$$

where the point process on the right is a PRM(μ_G); see Theorem 2 in [5] for details. □

Theorem 2.1. *Assume that K is a stopping time with respect to the filtration $(\mathcal{F}_j)_{j\in\mathbb{N}}$ with a finite mean. If X belongs to* MDA(G)*, then the same holds for $H=\bigvee_{j=1}^{K}X_j$.*

Proof. For (H_i) i.i.d. copies of H, using Lemma 2.2 and the notation therein,

$$\mathbb{P}\left(\frac{\bigvee_{i=1}^{n}H_i-b_{\lfloor n\xi\rfloor}}{a_{\lfloor n\xi\rfloor}}\le x\right)=\mathbb{P}\left(\sum_{i=1}^{n}\sum_{j=1}^{K'_i}\delta_{\frac{X_{i,j}-b_{\lfloor n\xi\rfloor}}{a_{\lfloor n\xi\rfloor}}}(x,\infty)=0\right)$$
$$\to\mathbb{P}(\text{PRM}(\mu_G)(x,\infty)=0)=G(x).$$ □

Example 2.2. *(Example 2.1 continued.)* Provided $\mathbb{E}[W]<\infty$, we recover known results for Example 2.1(a). Since $\mathbb{E}[K]<\infty$, in the case (b) H belongs to the same MDA as X. As we have seen, the case (c) is more involved, but the theorem implies that if W_1 has a heavier tail index than X, then $\mathbb{E}[K]=\infty$ and $H\notin$ MDA(G). On the other hand, for bounded or lighter-tailed W, we can still have $H\in$ MDA(G).

3. Limiting behaviour of the maximal claim size in the marked renewal cluster model

To describe the marked renewal cluster model, consider first an independently marked renewal process N^0. Let $(Y_k)_{k\in\mathbb{N}}$ be a sequence of i.i.d. non-negative inter-arrival times in N^0, and let $(A_k)_{k\in\mathbb{N}}$ be i.i.d. marks independent of $(Y_k)_{k\in\mathbb{N}}$ with distribution Q on $(\mathbb{S},\mathcal{B}(\mathbb{S}))$. Throughout we assume that

$$0<\mathbb{E}[Y]=\frac{1}{\nu}<\infty.$$

If we denote by $(\Gamma_i)_{i\in\mathbb{N}}$ the sequence of partial sums of $(Y_k)_{k\in\mathbb{N}}$, the process N^0 on the space $[0,\infty)\times\mathbb{S}$ has the representation

$$N^0=\sum_{i\in\mathbb{N}}\delta_{\Gamma_i,A_i}.$$

Processes of this type appear in non-life insurance mathematics, where marks are often referred to as claims. They can represent the size of the claim, type of the claim, severity of the accident, etc.

Assume that at each time Γ_i with mark A_i another point process in $M_p([0,\infty)\times\mathbb{S})$, denoted by G_i, is generated. All G_i are mutually independent and intuitively represent clusters of points superimposed on N^0 after time Γ_i. Formally, there exists a probability kernel K from $\mathbb{S}$ to $M_p([0,\infty)\times\mathbb{S})$ such that, conditionally on N^0, the point processes G_i are independent, almost surely finite, and with distribution equal to $K(A_i,\cdot)$. Note that this permits dependence between G_i and A_i.

In this setting, the process N^0 is usually called the parent process, while the G_i are called the descendant processes. We can write

$$G_i=\sum_{j=1}^{K_i}\delta_{T_{i,j},A_{i,j}},$$

https://doi.org/10.1017/jpr.2022.52 Published online by Cambridge University Press

where $(T_{i,j})_{j\in\mathbb{N}}$ is a sequence of non-negative random variables and K_i is a $\mathbb{Z}_+$-valued random variable. If we count the original point arriving at time Γ_i, the actual cluster size is $K_i + 1$.

Throughout, we also assume that the cluster processes G_i are independently marked with the same mark distribution Q independent of A_i, so that all the marks $A_{i,j}$ are i.i.d. Note that K_i may possibly depend on A_i. We assume throughout that

$$\mathbb{E}[K_i] < \infty.$$

Finally, to describe the size and other characteristics of all the observations (claims) together with their arrival times, we use a marked point process N as a random element in $M_p([0, \infty) \times \mathbb{S})$ of the form

$$N = \sum_{i=1}^{\infty} \sum_{j=0}^{K_i} \delta_{\Gamma_i + T_{i,j}, A_{i,j}}, \tag{3.1}$$

where we set $T_{i,0} = 0$ and $A_{i,0} = A_i$. In this representation, the claims arriving at time Γ_i and corresponding to the index $j = 0$ are called ancestral or immigrant claims, while the claims arriving at times $\Gamma_i + T_{i,j}$, $j \in \mathbb{N}$, are referred to as progeny or offspring. Note that N is almost surely boundedly finite, because $\Gamma_i \to \infty$ as $i \to \infty$, and K_i is almost surely finite for every i, so one could also write

$$N = \sum_{k=1}^{\infty} \delta_{\tau_k, A^k}, \tag{3.2}$$

with $\tau_k \le \tau_{k+1}$ for all $k \in \mathbb{N}$ and A^k being i.i.d. marks which are in general not independent of the arrival times (τ_k). Observe that this representation ignores the information regarding the clusters of the point process. Note also that eventual ties turn out to be irrelevant asymptotically.

In the special case, when the inter-arrival times are exponential with parameter ν, the renewal counting process which generates the arrival times in the parent process is a homogeneous Poisson process. The associated marked renewal cluster model is then called a marked Poisson cluster process (see [12]; cf. [6]).

Remark 3.1. In all our considerations we take into account the original immigrant claims arriving at times Γ_i as well. One could of course ignore these claims and treat Γ_i as times of incidents that trigger, with a possible delay, a cluster of subsequent payments, as in the model of the so-called incurred but not reported (IBNR) claims; cf. [24].

The numerical observations, i.e. the sizes of the claims, are produced by the application of a measurable function on the marks, say $f : \mathbb{S} \to \mathbb{R}_+$. The maximum of all claims due to the arrival of an immigrant claim at time Γ_i equals

$$H_i = \bigvee_{j=0}^{K_i} X_{i,j}, \tag{3.3}$$

where $X_{i,j} = f(A_{i,j})$ are i.i.d. random variables for all i and j. The random variable H_i has an interpretation as the maximal claim size coming from the ith immigrant and its progeny. If we denote $f(A^k)$ by X^k, the maximal claim size in the period $[0, t]$ can be represented as

$$M(t) = \sup \left\{ X^k : \tau_k \le t \right\}.$$

https://doi.org/10.1017/jpr.2022.52 Published online by Cambridge University Press

In order to bring the model into the context of Theorem 2.1, observe that one can let $W_k = A^k$, for $k \in \mathbb{N}$. Introduce the first-passage-time process $(\tau(t))_{t\geq 0}$ defined by

$$\tau(t) = \inf\{n : \Gamma_n > t\}\,, \qquad t \geq 0.$$

This means that $\tau(t)$ is the renewal counting process generated by the sequence $(Y_n)_{n\in\mathbb{N}}$. According to the strong law for counting processes (Theorem 5.1 in [16, Chapter 2]), for every $c \geq 0$,

$$\frac{\tau(tc)}{\nu t} \xrightarrow{as} c \quad \text{as } t \to \infty.$$

Denote by

$$M^\tau(t) = \bigvee_{i=1}^{\tau(t)} H_i$$

the maximal claim size coming from the maximal claim sizes in the first $\tau(t)$ clusters. Now we can write

$$M^\tau(t) = M(t) \bigvee H_{\tau(t)} \bigvee \varepsilon_t, \quad t \geq 0, \tag{3.4}$$

where the last error term represents the leftover effect at time t, i.e. the maximum of all claims arriving after t which correspond to the progeny of immigrants arriving before time t; more precisely,

$$\varepsilon_t = \max\{X_{i,j} : 0 \leq \Gamma_i \leq t,\ t < \Gamma_i + T_{i,j}\}, \quad t \geq 0.$$

Denote the number of members in the set above by

$$J_t = \#\{(i,j) : 0 \leq \Gamma_i \leq t,\ t < \Gamma_i + T_{i,j}\}. \tag{3.5}$$

We study the limiting behaviour of the maximal claim size $M(t)$ up to time t and aim to find sufficient conditions under which $M(t)$ converges in distribution to a non-trivial limit after appropriate centring and normalization.

Recall that H belongs to MDA(G) if there exist constants $c_n > 0$, $d_n \in \mathbb{R}$ such that for each $x \in \mathbb{E} = \{y \in \mathbb{R} : G(y) > 0\}$,

$$n \cdot \mathbb{P}(H > c_n x + d_n) \to -\log G(x) \quad \text{as } n \to \infty. \tag{3.6}$$

An application of Lemma 2.1 yields the following result.

Proposition 3.1. *Assume that H belongs to* MDA(G), *so that* (3.6) *holds, and that the error term in* (3.5) *satisfies*

$$J_t = o_P(t).$$

Then

$$\frac{M(t) - d_{\lfloor \nu t\rfloor}}{c_{\lfloor \nu t\rfloor}} \xrightarrow{d} G \quad \textit{as } t \to \infty. \tag{3.7}$$

Proof. Using the equation (3.4),

$$\frac{M^\tau(t) - d_{\lfloor \nu t\rfloor}}{c_{\lfloor \nu t\rfloor}} = \frac{M(t) - d_{\lfloor \nu t\rfloor}}{c_{\lfloor \nu t\rfloor}} \bigvee \frac{H_{\tau(t)} - d_{\lfloor \nu t\rfloor}}{c_{\lfloor \nu t\rfloor}} \bigvee \frac{\varepsilon_t - d_{\lfloor \nu t\rfloor}}{c_{\lfloor \nu t\rfloor}}.$$

https://doi.org/10.1017/jpr.2022.52 Published online by Cambridge University Press

Since for $x \in \mathbb{E}$

$$0 \le \mathbb{P}\left(\frac{M^\tau(t) - d_{\lfloor \nu t \rfloor}}{c_{\lfloor \nu t \rfloor}} > x\right) - \mathbb{P}\left(\frac{M(t) - d_{\lfloor \nu t \rfloor}}{c_{\lfloor \nu t \rfloor}} > x\right)$$
$$\le \mathbb{P}\left(\frac{H_{\tau(t)} - d_{\lfloor \nu t \rfloor}}{c_{\lfloor \nu t \rfloor}} > x\right) + \mathbb{P}\left(\frac{\varepsilon_t - d_{\lfloor \nu t \rfloor}}{c_{\lfloor \nu t \rfloor}} > x\right),$$

it suffices to show that

$$\frac{M^\tau(t) - d_{\lfloor \nu t \rfloor}}{c_{\lfloor \nu t \rfloor}} \xrightarrow{d} G \quad \text{as } t \to \infty, \tag{3.8}$$

$$\lim_{t\to\infty} \mathbb{P}\left(\frac{H_{\tau(t)} - d_{\lfloor \nu t \rfloor}}{c_{\lfloor \nu t \rfloor}} > x\right) = 0, \quad \text{and} \quad \lim_{t\to\infty} \mathbb{P}\left(\frac{\varepsilon_t - d_{\lfloor \nu t \rfloor}}{c_{\lfloor \nu t \rfloor}} > x\right) = 0. \tag{3.9}$$

Recall that H_i represents the maximum of all claims due to the arrival of an immigrant claim at time Γ_i, and by (3.3) it equals

$$H_i = \bigvee_{j=0}^{K_i} X_{i,j}.$$

Note that (H_i) is an i.i.d. sequence, because the ancestral mark in every cluster comes from an independently marked renewal point process. As in the proofs of Lemma 2.2 and Theorem 2.1,

$$\mathbb{P}\left(\frac{M^\tau(t) - d_{\lfloor \nu t \rfloor}}{c_{\lfloor \nu t \rfloor}} \le x\right) = \mathbb{P}\left(\sum_{i=1}^{\tau(t)} \delta_{\frac{H_i - d_{\lfloor \nu t \rfloor}}{c_{\lfloor \nu t \rfloor}}}(x, \infty) = 0\right)$$
$$\to \mathbb{P}(\mathrm{PRM}(\mu_G)(x, \infty) = 0) = G(x),$$

as $t \to \infty$, which shows (3.8). To show (3.9), note that $\{\tau(t) = k\} \in \sigma(Y_1, \ldots Y_k)$ and by assumption $\{H_k \in A\}$ is independent of $\sigma(Y_1, \ldots Y_k)$ for every k. Therefore, $H_{\tau(t)} \stackrel{d}{=} H_1 \in \mathrm{MDA}(G)$, so the first part of (3.9) easily follows from (3.6). For the second part of (3.9), observe that the leftover effect ε_t admits the representation

$$\varepsilon_t \stackrel{d}{=} \bigvee_{i=1}^{J_t} X_i,$$

for $(X_i)_{i\in\mathbb{N}}$ i.i.d. copies of $X = f(A)$. Hence,

$$\frac{\varepsilon_t - d_{\lfloor \nu t \rfloor}}{c_{\lfloor \nu t \rfloor}} \stackrel{d}{=} \frac{\bigvee_{i=1}^{J_t} X_i - d_{\lfloor \nu t \rfloor}}{c_{\lfloor \nu t \rfloor}}.$$

Since $J_t = o_P(t)$, for every fixed $\delta > 0$ and t large enough, $\mathbb{P}(J_t > \delta t) < \delta$. For measurable $A = \{J_t > \delta t\}$ we have

$$\mathbb{P}\left(\frac{\bigvee_{i=1}^{J_t} X_i - d_{\lfloor \nu t \rfloor}}{c_{\lfloor \nu t \rfloor}} > x\right) \le \mathbb{P}(A) + \mathbb{P}\left(\left\{\frac{\bigvee_{i=1}^{J_t} X_i - d_{\lfloor \nu t \rfloor}}{c_{\lfloor \nu t \rfloor}} > x\right\} \cap A^C\right)$$
$$< \delta + \mathbb{P}\left(\frac{\bigvee_{i=1}^{\lfloor \delta t \rfloor} X_i - d_{\lfloor \nu t \rfloor}}{c_{\lfloor \nu t \rfloor}} > x\right),$$

which converges to 0 as $\delta \to 0$. □

https://doi.org/10.1017/jpr.2022.52 Published online by Cambridge University Press

As we have seen above, it is relatively easy to determine the asymptotic behaviour of the maximal claim size $M(t)$ as long as one can determine the tail properties of the random variables H_i and the number of points in the leftover effect at time t, J_t in (3.5). An application of Theorem 2.1 immediately yields the following corollary.

Corollary 3.1. *Let $J_t = o_P(t)$, and let $(X_{i,j})$ satisfy (2.1) and the assumptions from the proof of Theorem 2.1. Then (3.7) holds with (c_n) and (d_n) defined by*

$$(c_n) = (a_{\lfloor (\mathbb{E}[K]+1)\cdot n \rfloor}), \quad (d_n) = (b_{\lfloor (\mathbb{E}[K]+1)\cdot n \rfloor}). \tag{3.10}$$

As we shall see in the following section, showing that $J_t = o_P(t)$ holds remains a rather technical task. However, this can be done for several frequently used cluster models.

4. Maximal claim size for three special models

In this section we present three special models belonging to the general marked renewal cluster model introduced in Section 3. We try to find sufficient conditions for these models in order to apply Proposition 3.1.

Remark 4.1. In any of the three examples below, the point process N can be made stationary if we start the construction in (3.1) on the state space $\mathbb{R} \times \mathbb{S}$ with a renewal process $\sum_i \delta_{\Gamma_i}$ on the whole real line. For the resulting stationary cluster process we use the notation N^*. Still, from the applied perspective, it seems more interesting to study the nonstationary version, where both the parent process N^0 and the cluster process itself have arrivals only from some point onwards, e.g. in the interval $[0, \infty)$.

4.1. Mixed binomial cluster model

Assume that the renewal counting process which generates the arrival times in the parent process $(\Gamma_i)_{i\in\mathbb{N}}$ is a homogeneous Poisson process with mean measure (νLeb) on the state space $[0, \infty)$ for $\nu > 0$, and that the individual clusters have the form

$$G_i = \sum_{j=1}^{K_i} \delta_{V_{i,j}, A_{i,j}}.$$

Assume that $(K_i, (V_{i,j})_{j\in\mathbb{N}}, (A_{i,j})_{j\in\mathbb{Z}_+})_{i\in\mathbb{N}}$ constitutes an i.i.d. sequence with the following properties for fixed $i \in \mathbb{N}$:

- $(A_{i,j})_{j\in\mathbb{Z}_+}$ are i.i.d.;
- $(V_{i,j})_{j\in\mathbb{N}}$ are conditionally i.i.d. given $A_{i,0}$;
- $(A_{i,j})_{j\in\mathbb{N}}$ are independent of $(V_{i,j})_{j\in\mathbb{N}}$;
- K_i is a stopping time with respect to the filtration generated by the $(A_{i,j})_{j\in\mathbb{Z}_+}$, i.e. for every $k \in \mathbb{Z}_+$, $\{K_i = k\} \in \sigma(A_{i,0}, \ldots A_{i,k})$.

Notice that we do allow possible dependence between K_i and $(A_{i,j})_{j\in\mathbb{Z}_+}$. Also, we do not exclude the possibility of dependence between $(V_{i,j})_{j\in\mathbb{N}}$ and the ancestral mark $A_{i,0}$ (and consequently K_i). Recall that K is an integer-valued random variable representing the size of a cluster, such that $\mathbb{E}[K] < \infty$. Observe that we use the notation $V_{i,j}$ instead of $T_{i,j}$ to emphasize the relatively simple structure of clusters in this model, in contrast with the other two models

https://doi.org/10.1017/jpr.2022.52 Published online by Cambridge University Press

in this section. Such a process N is a marked version of the so-called Neyman–Scott process; e.g. see [12, Example 6.3(a)].

Corollary 4.1. *Assume that* $f(A) = X$ *belongs to* MDA(G), *so that* (2.1) *holds. Then* (3.7) *holds for* (c_n) *and* (d_n) *defined in* (3.10).

Proof. Using Theorem 2.1 we conclude that the maximum H of all claims in a cluster belongs to the MDA of the same distribution as X. Apply Proposition 3.1 after observing that $J_t = o_P(t)$. Using Markov's inequality, it is enough to check that $\mathbb{E}[J_t] = o(t)$,

$$\mathbb{E}[J_t] = \mathbb{E}[\#\{(i, j) : 0 \le \Gamma_i \le t,\ t < \Gamma_i + V_{i,j}\}]$$
$$= \mathbb{E}\left[\sum_{0\le\Gamma_i\le t}\sum_{j=1}^{K_i} \mathbb{I}_{t\le\Gamma_i+V_{i,j}}\right].$$

Using Lemma 7.2.12 in [24] and calculations similar to those in the proofs of Corollaries 5.1 and 5.3 in [6], we have

$$\mathbb{E}\left[\sum_{0\le\Gamma_i\le t}\sum_{j=1}^{K_i} \mathbb{I}_{t<\Gamma_i+V_{i,j}}\right] = \int_0^t \mathbb{E}\left[\sum_{j=1}^{K_i} \mathbb{I}_{V_{i,j}>t-s}\right] \nu ds = \int_0^t \mathbb{E}\left[\sum_{j=1}^{K_i} \mathbb{I}_{V_{i,j}>x}\right] \nu dx.$$

Now note that as $x \to \infty$, by the dominated convergence theorem,

$$\mathbb{E}\left[\sum_{j=1}^{K_i} \mathbb{I}_{V_{i,j}>x}\right] \to 0.$$

An application of a Cesàro argument now yields that $\mathbb{E}[J_t]/t \to 0$. □

4.2. Renewal cluster model

Assume next that the clusters G_i have the following distribution:

$$G_i = \sum_{j=1}^{K_i} \delta_{T_{i,j},A_{i,j}},$$

where $(T_{i,j})$ represents the sequence such that

$$T_{i,j} = V_{i,1} + \cdots + V_{i,j}, \quad 1 \le j \le K_i.$$

We keep all the other assumptions from the model in the previous subsection.

A general unmarked model of a similar type, called the Bartlett–Lewis model, is analysed in [12]; see Example 6.3(b). See also [15] for an application of a similar point process to modelling of teletraffic data. By adapting the arguments from Corollary 4.1 we can easily obtain the next corollary.

Corollary 4.2. *Assume that* $f(A) = X$ *belongs to* MDA(G), *so that* (2.1) *holds. Then* (3.7) *holds for* (c_n) *and* (d_n) *defined in* (3.10).

https://doi.org/10.1017/jpr.2022.52 Published online by Cambridge University Press

4.3. Marked Hawkes processes

Another example in our analysis is the so-called (linear) marked Hawkes process. These processes are typically introduced through their stochastic intensity (see, for example, [22] or [12]). More precisely, a point process $N = \sum_k \delta_{\tau_k, A^k}$ represents a Hawkes process of this type if the random marks (A^k) are i.i.d. with distribution Q on the space $\mathbb{S}$, while the arrivals (τ_k) have stochastic intensity of the form

$$\lambda(t) = \nu + \sum_{\tau_i < t} h(t - \tau_i, A^i),$$

where $\nu > 0$ is a constant and $h : [0, \infty) \times \mathbb{S} \to \mathbb{R}_+$ is assumed to be integrable in the sense that $\int_0^\infty \mathbb{E}[h(s, A)]ds < \infty$. On the other hand, Hawkes processes of this type have a neat Poisson cluster representation due to [19]. For this model, the clusters G_i are recursive aggregations of Cox processes, i.e. Poisson processes with random mean measure $\tilde{\mu}_{A_i} \times Q$ where $\tilde{\mu}_{A_i}$ has the form

$$\tilde{\mu}_{A_i}(B) = \int_B h(s, A_i)ds,$$

for some fertility (or self-exciting) function h; cf. Example 6.4(c) of [12]. It is useful to introduce a time shift operator θ_t, by defining

$$\theta_t m = \sum_j \delta_{t_j + t, a_j},$$

for an arbitrary point measure $m = \sum_j \delta_{t_j, a_j} \in M_p([0, \infty) \times \mathbb{S})$ and $t \geq 0$. Now, for the parent process $N^0 = \sum_{i \in \mathbb{N}} \delta_{\Gamma_i, A_i}$, which is a Poisson point process with mean measure $\nu \times Q$ on the space $[0, \infty) \times \mathbb{S}$, the cluster process corresponding to a point (Γ_i, A_i) satisfies the following recursive relation:

$$G_i = \sum_{l=1}^{L_{A_i}} \left(\delta_{\tau_l^1, A_l^1} + \theta_{\tau_l^1} G_l^1 \right), \tag{4.1}$$

where, given A_i,

$$\tilde{N}_i = \sum_{l=1}^{L_{A_i}} \delta_{\tau_l^1, A_l^1}$$

is a Poisson process with mean measure $\tilde{\mu}_{A_i} \times Q$, and the sequence $(G_l^1)_l$ is i.i.d., distributed as G_i and independent of $\tilde{N}_i$.

Thus, at any ancestral point (Γ_i, A_i), a cluster of points appears as a whole cascade of points to the right in time generated recursively according to (4.1). Note that L_{A_i} has Poisson distribution conditionally on A_i, with mean $\kappa_{A_i} = \int_0^\infty h(s, A_i)ds$. It corresponds to the number of first-generation progeny (A_l^1) in the cascade. Note also that the point processes forming the second generation are again Poisson conditionally on the corresponding first-generation mark A_l^1. The cascade G_i corresponds to the process formed by the successive generations, drawn recursively as Poisson processes given the former generation. The marked Hawkes process is obtained by attaching to the ancestors (Γ_i, A_i) of the marked Poisson process $N^0 = \sum_{i \in \mathbb{N}} \delta_{\Gamma_i, A_i}$ a cluster of points, denoted by C_i, which contains the point $(0, A_i)$ and a whole cascade G_i

of points to the right in time generated recursively according to (4.1) given A_i. Under the assumption

$$\kappa = \mathbb{E}\left[\int h(s, A) ds\right] < 1, \tag{4.2}$$

the total number of points in a cluster is generated by a subcritical branching process. Therefore, the clusters are finite almost surely. Denote their size by K_i+1. It is known (see Example 6.3(c) in [12]) that under (4.2) the clusters always satisfy

$$\mathbb{E}[K_i]+1 = \frac{1}{1-\kappa}. \tag{4.3}$$

Note that the clusters C_i, i.e. point processes which represent a cluster together with the mark A_i, are independent by construction. They can be represented as

$$C_i = \sum_{j=0}^{K_i} \delta_{\Gamma_i + T_{i,j}, A_{i,j}},$$

with $A_{i,j}$ being i.i.d., $A_{i,0} = A_i$, $T_{i,0} = 0$, and $T_{i,j}$, $j \in \mathbb{N}$, representing arrival times of progeny claims in the cluster C_i. Observe that in the case when marks do not influence conditional density, i.e. when $h(s, a) = h(s)$, the random variable K_i+1 has a so-called Borel distribution with parameter κ; see [17]. Notice also that in general, marks and arrival times of the final Hawkes process N are not independent of each other; rather, in the terminology of [12], the marks in the process N are only unpredictable.

As before, the maximal claim size in one cluster is of the form

$$H \stackrel{d}{=} \bigvee_{j=0}^{K} X_j.$$

Note that K and (X_j) are not independent. In this case, thanks to the representation of Hawkes processes as the recursive aggregation of Cox processes (4.1), the maximal claim size can also be written as

$$H \stackrel{d}{=} X \vee \bigvee_{j=1}^{L_A} H_j.$$

Recall from (4.2) that $\kappa = \mathbb{E}[\kappa_A] < 1$. The H_j on the right-hand side are independent of κ_A and i.i.d. with the same distribution as H. Conditionally on A, the waiting times are i.i.d. with common density

$$\frac{h(t, A)}{\kappa_A}, \quad t \geq 0; \tag{4.4}$$

see [22] or [6]. In order to apply Proposition 3.1, first we show that H is in MDA(G), using the well-known connection between branching processes and random walks; see for instance [1], [7], or the quite recent [11]. This is the subject of the next lemma.

Lemma 4.1. *Let X belong to* MDA(G) *in the marked Hawkes model. Then H also belongs to the same* MDA(G).

Proof. By the recursive relation (4.1), each cluster can be associated with a subcritical branching process (Bienaymé–Galton–Watson tree) where the total number of points in a

https://doi.org/10.1017/jpr.2022.52 Published online by Cambridge University Press

cascade (cluster) corresponds to the total number of vertices in such a tree. It has the same distribution as the first hitting time of level 0,

$$\zeta = \inf\{k : S_k = 0\} ,$$

by a random walk (S_n) defined as

$$S_0 = 1, \quad S_n = S_{n-1} + L_n - 1,$$

with i.i.d. $L_n \stackrel{d}{=} L$. Notice that (S_n) has negative drift, which leads to the conclusion that ζ is a proper random variable. Moreover, since $\mathbb{E}[L] < 1$, an application of Theorem 3 from [16] gives $\mathbb{E}[\zeta] < \infty$ and implies that we can use (4.3) since $\zeta = K + 1$.

If we write, for arbitrary $k \in \mathbb{N}$,

$$\begin{aligned}\{\zeta = k\} &= \{S_0 > 0, S_1 > 0, \dots, S_{k-1} > 0, S_k = 0\} \\ &= \left\{1 > 0, L_1 > 0, \dots, \sum_{i=1}^{k-1} L_i - (k-2) > 0, \sum_{i=1}^{k} L_i - (k-1) = 0\right\} \\ &\in \sigma(L, A_0, A_1, \dots, A_k) ,\end{aligned}$$

we see that ζ is a stopping time with respect to $(\mathcal{F}'_j)_{j\in\mathbb{Z}_+}$, where $\mathcal{F}'_j = \sigma(L, A_0, A_1, \dots, A_j)$, and where L has conditionally Poisson distribution with random parameter κ_A and is independent of the sequence $(A_j)_{j\in\mathbb{Z}_+}$. By Theorem 2.1 we conclude that H is also in MDA(G). □

Remark 4.2. The equation (4.3) implies that the sequences (c_n) and (d_n) in the following corollary have the representations

$$(c_n) = \left(a_{\lfloor \frac{1}{1-\kappa} n \rfloor}\right), \qquad (d_n) = \left(b_{\lfloor \frac{1}{1-\kappa} n \rfloor}\right).$$

Corollary 4.3. *Assume that X belongs to* MDA(G)*, so that* (2.1) *holds, and*

$$\mathbb{E}\big[\tilde{\mu}_A(t, \infty)\big] \to 0 \quad \text{as } t \to \infty.$$

Then (3.7) *holds for* (c_n) *and* (d_n) *defined in* (3.10).

Proof. Recall from (3.2) that one can write

$$N = \sum_{i=1}^{\infty} \sum_{j=0}^{K_i} \delta_{\Gamma_i + T_{i,j}, A_{i,j}} = \sum_{k=1}^{\infty} \delta_{\tau_k, A^k},$$

without loss of generality assuming that $0 \le \tau_1 \le \tau_2 \le \dots$. At each time τ_j, a claim arrives generated by one of the previous claims, or an entirely new (immigrant) claim appears. In the former case, if τ_j is the direct offspring of a claim at time τ_i, we will write $\tau_i \to \tau_j$. The progeny τ_j then potentially creates further claims. Notice that $\tau_i \to \tau_j$ is equivalent to $\tau_j = \tau_i + V_{i,k}$, $k \le L^i = L_{A^i}$, where $V_{i,k}$ are waiting times which, according to the discussion above (4.4), are i.i.d. with common density $h(t, A^i)/\kappa_{A^i}$, $t \ge 0$, and independent of L^i conditionally on the mark A^i of the claim at τ_i. Moreover, conditionally on A^i, the number of direct progeny of the claim at τ_i, denoted by L^i, has Poisson distribution with parameter $\tilde{\mu}_{A^i}$. We denote by K_{τ_j} the total number of points generated by the arrival at τ_j. Clearly, the K_{τ_j} are identically distributed as K and even mutually independent if we consider only points which are not offspring of one another.

It is enough to check $\mathbb{E}[J_t]/t = o(1)$ and see that

$$\mathbb{E}[J_t] = \mathbb{E}\Big[\sum_{\Gamma_i \le t}\sum_j \mathbb{I}_{\Gamma_i + T_{i,j} > t}\Big] = \mathbb{E}\Big[\sum_{\tau_i \le t}\sum_{\tau_j > t} (K_{\tau_j} + 1)\, \mathbb{I}_{\tau_i \to \tau_j}\Big]$$
$$= \mathbb{E}\Big[\sum_{\tau_i \le t} \mathbb{E}\Big[\sum_{k=1}^{L^i} (K_{\tau_i + V_{i,k}} + 1)\mathbb{I}_{\tau_i + V_{i,k} > t} \,|\, (\tau_i, A^i)_{i \ge 0}; \tau_i \le t\Big]\Big]$$
$$= \frac{1}{1-\kappa}\mathbb{E}\left[\int_0^t \int_{\mathbb{S}} \tilde{\mu}_a((t-s, \infty))N(ds, da)\right],$$

where $\tilde{\mu}_a((u, \infty)) = \int_u^\infty h(s, a)ds$. Observe that from the projection theorem (see Theorem 3 in [10, Chapter 8]), the last expression equals

$$\frac{1}{1-\kappa}\mathbb{E}\left[\int_0^t \int_{\mathbb{S}} \tilde{\mu}_a((t-s, \infty))Q(da)\lambda(s)ds\right].$$

Recall from Remark 4.1 that N has a stationary version, N^*, such that the expression $\mathbb{E}\big[\lambda^*(s)\big]$ is a constant equal to $\nu/(1-\kappa)$. Using Fubini's theorem, one can further bound the last expectation from above by

$$\mathbb{E}\left[\int_0^t \int_{\mathbb{S}} \tilde{\mu}_a((t-s, \infty))Q(da)\lambda^*(s)ds\right] = \int_0^t \int_{\mathbb{S}} \tilde{\mu}_a((t-s, \infty))Q(da)\mathbb{E}[\lambda^*(s)]ds$$
$$= \frac{\nu}{1-\kappa}\int_0^t \int_{\mathbb{S}} \tilde{\mu}_a((t-s, \infty))Q(da)ds.$$

Now we have

$$\mathbb{E}J_t \le \frac{\nu}{(1-\kappa)^2}\int_0^t \int_{\mathbb{S}} \tilde{\mu}_a((t-s, \infty))Q(da)ds = \frac{\nu}{(1-\kappa)^2}\int_0^t \int_s^\infty \mathbb{E}[h(u, A)]duds.$$

Dividing the last expression by t and applying L'Hôpital's rule proves the theorem for the nonstationary or pure Hawkes process. □

Acknowledgements

We sincerely thank the anonymous reviewers for suggestions which led to simplified proofs and improved layout of the article.

Funding information

The work of Bojan Basrak and Nikolina Milinčević was supported by the Swiss Enlargement Contribution in the framework of the Croatian-Swiss Research Programme (project number IZHRZ0_180549).

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

[1] ASMUSSEN, S. AND FOSS, S. (2018). Regular variation in a fixed-point problem for single- and multi-class branching processes and queues. *Adv. Appl. Prob.* **50**, 47–61.

[2] BARNDORFF-NIELSEN, O. (1964). On the limit distribution of the maximum of a random number of independent random variables. *Acta Math. Acad. Sci. Hungar.* **15**, 399–403.

[3] BASRAK, B. AND PLANINIĆ, H. (2019). A note on vague convergence of measures. *Statist. Prob. Lett.* **153**, 180–186.

[4] BASRAK, B. AND PLANINIĆ, H. (2021). Compound Poisson approximation for regularly varying fields with application to sequence alignment. *Bernoulli* **27**, 1371–1408.

[5] BASRAK, B. AND ŠPOLJARIĆ, D. (2015). Extremes of random variables observed in renewal times. *Statist. Prob. Lett.* **97**, 216–221.

[6] BASRAK, B., WINTENBERGER, O. AND ŽUGEC, P. (2019). On the total claim amount for marked Poisson cluster models. *Adv. Appl. Prob.* **51**, 541–569.

[7] BENNIES, J. AND KERSTING, G. (2000). A random walk approach to Galton–Watson trees. *J. Theoret. Prob.* **13**, 777–803.

[8] BERMAN, S. M. (1962). Limiting distribution of the maximum term in sequences of dependent random variables. *Ann. Math. Statist.* **33**, 894–908.

[9] BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd edn. John Wiley, New York.

[10] BREMAUD, P. (1981). *Point Processes and Queues*. Springer, New York.

[11] COSTA, M., GRAHAM, C., MARSALLE, L. AND TRAN, V. (2020). Renewal in Hawkes processes with self-excitation and inhibition. *Adv. Appl. Prob.* **52**, 879–915.

[12] DALEY, D. J. AND VERE-JONES, D. (2003). *An Introduction to the Theory of Point Processes*, Vol. I–II, 2nd edn. Springer, New York.

[13] DE HAAN, L. AND FERREIRA, A. (2006). *Extreme Value Theory: An Introduction*. Springer, New York.

[14] EMBRECHTS, P., KLÜPPELBERG, C. AND MIKOSCH, T. (1997). *Modelling Extremal Events*. Springer, Berlin.

[15] FAŸ, G., GONZÁLEZ-ARÉVALO, B., MIKOSCH, T. AND SAMORODNITSKY, G. (2006). Modeling teletraffic arrivals by a Poisson cluster process. *Queueing Systems* **54**, 121–140.

[16] GUT, A. (2009). *Stopped Random Walks*, 2nd edn. Springer, New York.

[17] HAIGHT, F. A. AND BREUER, H. A. (1960). The Borel–Tanner distribution. *Biometrika* **47**, 143–150.

[18] HASHORVA, E., PADOAN, S. A. AND RIZZELLI, S. (2021). Multivariate extremes over a random number of observations. *Scand. J. Statist.* **48**, 845–880.

[19] HAWKES, A. G. AND OAKES, D. (1974). A cluster process representation of a self-exciting process. *J. Appl. Prob.* **11**, 493–503.

[20] JESSEN, A. H. AND MIKOSCH, T. (2006). Regularly varying functions. *Publ. Inst. Math.* **80**, 171–192.

[21] KALLENBERG, O. (2017). *Random Measures, Theory and Applications*. Springer, Cham.

[22] KARABASH, D. AND ZHU, L. (2015). Limit theorems for marked Hawkes processes with application to a risk model. *Stoch. Models* **31**, 433–451.

[23] MEERSCHAERT, M. M. AND STOEV, S. A. (2009). Extremal limit theorems for observations separated by random power law waiting times. *J. Statist. Planning Infer.* **139**, 2175–2188.

[24] MIKOSCH, T. (2009). *Non-life Insurance Mathematics*, 2nd edn. Springer, New York.

[25] PANCHEVA, E., MITOV, I. K. AND MITOV, K. V. (2009). Limit theorems for extremal processes generated by a point process with correlated time and space components. *Statist. Prob. Lett.* **79**, 390–395.

[26] RESNICK, S. I. (1987). *Extreme Values, Regular Variation, and Point Processes*. Springer, New York.

[27] SILVESTROV, D. S. AND TEUGELS, J. L. (1998). Limit theorems for extremes with random sample size. *Adv. Appl. Prob.* **30**, 777–806.

[28] TILLIER, C. AND WINTENBERGER, O. (2018). Regular variation of a random length sequence of random variables and application to risk assessment. *Extremes* **21**, 27–56.

[29] TOWE, R., TAWN, J., EASTOE, E. AND LAMB, R. (2020). Modelling the clustering of extreme events for short-term risk assessment. *J. Agric. Biol. Environm. Statist.* **25**, 32–53.

[30] VOGEL, M. M., HAUSER, M. AND SENEVIRATNE, S. I. (2020). Projected changes in hot, dry and wet extreme events' clusters in CMIP6 multi-model ensemble. *Environm. Res. Lett.* **15**, article no. 094021.