Received: 20 May 2020 | Accepted: 30 June 2021

DOI: 10.1002/env.2697

RESEARCH ARTICLE

WILEY

A self-exciting marked point process model for drought analysis

Xiaoting Li[1] | **Christian Genest**[1] | **Jonathan Jalbert**[2]

[1]Department of Mathematics and Statistics, McGill University, Montréal, Québec, Canada

[2]Département de mathématiques et de génie industriel, Polytechnique Montréal, Montréal, Québec, Canada

Correspondence
Christian Genest, Department of Mathematics and Statistics, McGill University, 805, rue Sherbrooke ouest, Montréal, Québec, Canada H3A 0B9.
Email: christian.genest@mcgill.ca

Funding information
Canada Research Chairs; Natural Sciences and Engineering Research Council of Canada; Trottier Institute for Science and Public Policy

Abstract

A self-exciting marked point process approach is proposed to model clustered low-flow events. It combines a self-exciting ground process designed to capture the temporal clustering behavior of extreme values and an extended Generalized Pareto mark distribution for the exceedances over a subasymptotic threshold. The model takes into account the dependence between the magnitude and occurrence time of exceedances and allows for closed-form inference on tail probabilities and large quantiles. It is used to analyze daily water levels from the Rivière des Mille Îles (Québec, Canada) and to characterize drought patterns in the Montréal area. The model is useful to generate short-term probability forecasts and to estimate the return period of major droughts. This information on the drought events is critical to water resource professionals in planning, designing, building, and managing more efficient water resource systems to hedge against the water shortage in case of extreme droughts.

KEYWORDS

drought, extended generalized Pareto, extreme-value inference, point process

1 | INTRODUCTION

The Rivière des Mille Îles is a channel of the Ottawa River in Southwestern Québec. It rises at the narrowing of the Lac des Deux Montagnes and runs into the Rivière des Prairies, which in turn flows into the mighty St. Lawrence (or Magtogoek) River. Although it is only 42 km long, the Rivière des Mille Îles is of crucial importance in providing drinking water for nearly half a million residents in the greater Montréal area. Low-flow conditions, especially in prolonged dry spells, imperil safe drinking-water supplies and increase the risk of waterborne infectious disease due to short-term peak concentrations of dangerous contaminants such as bacteria, viruses, and protozoans.

Hydrologically based design flow statistics are often used to set water quality standards. In Québec, the vulnerability to microorganisms is typically assessed using the 7Q2 standard, which refers to the lowest mean discharge for seven consecutive days at a 2-year recurrence interval. In the USA, about half the states rely on the 7Q10. Motivated by a need to make better-informed decisions regarding water resource management in the Montréal area, this article describes a statistical model that successfully captures drought attributes and occurrence patterns on the Rivière des Mille Îles. However, the methodology applies more generally and could prove helpful in other contexts.

The left panel of Figure 1 shows the daily water level (in meters) of the Rivière des Mille Îles from April 10 to August 10, 1988. A dashed line highlights the critical level $u = 13.66$ m selected by City of Montréal water engineers based on the

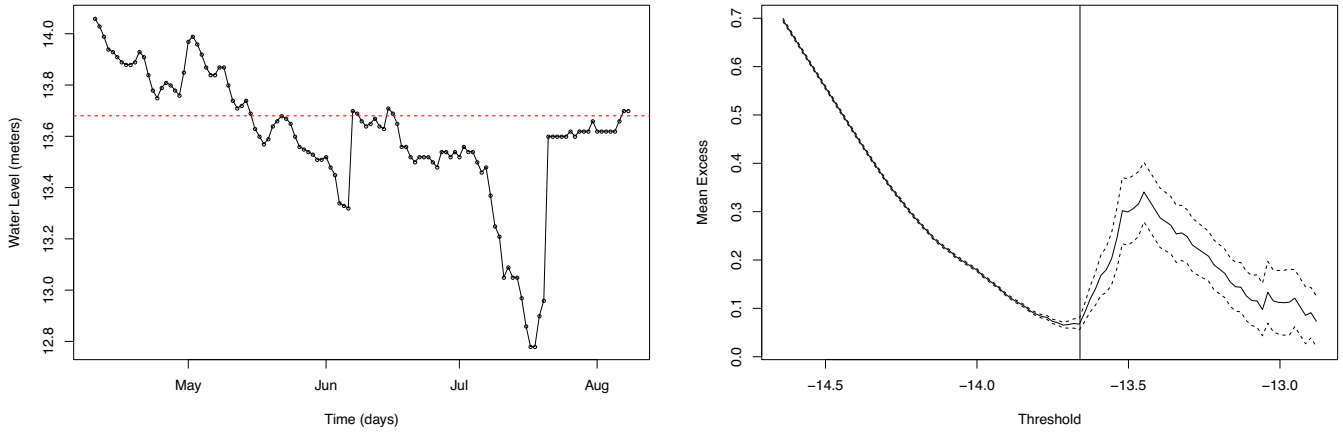

FIGURE 1 Water level (left, in meters) of the Rivière des Mille Îles (Québec, Canada) from April 10 to August 10, 1988 and threshold $u = 13.66$ m chosen by City of Montréal water engineers based on the plot (right) of estimates of the mean excess of minima as a function of the threshold (with reversed sign)

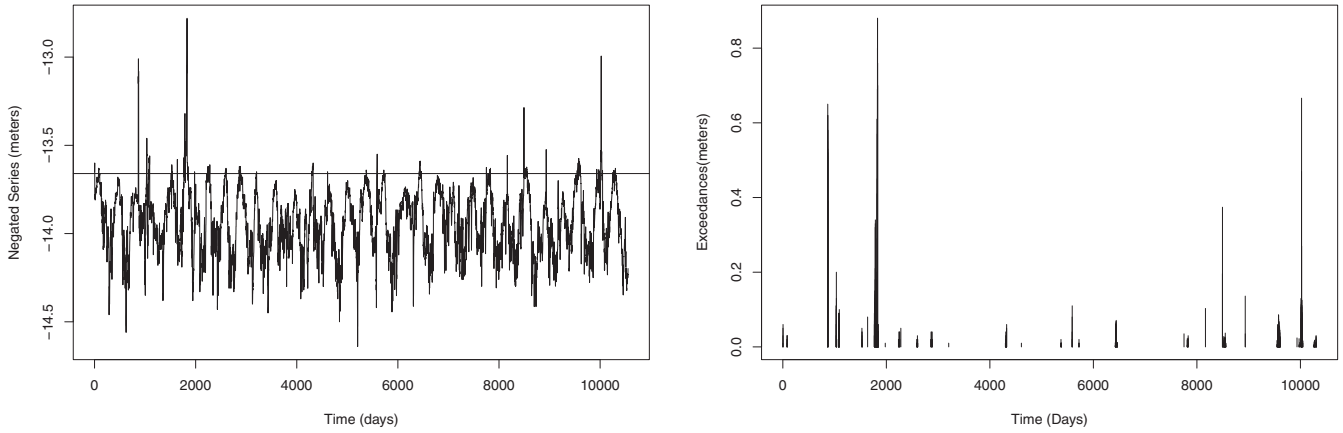

FIGURE 2 Plot (left) of the sign reversed daily water levels at the Rivière des Mille Îles (1983–2013), along with time and size (right) of the observations in excess of the 96% quantile

mean excess plot for minima (treated as maxima of data with reversed sign) shown in the right panel.[1] As can be seen, four dry spells occurred that summer when the water level fell below the set threshold for extended periods of time. This is typical of extended droughts (here the 1988–1989 North American Drought) where dependent dry spells are separated by above-threshold episodes that are too short to alleviate the effect of water shortage.

The tendency of extreme low-flows to occur in clusters is further documented in Figure 2, which exhibits the daily water levels (multiplied by -1) recorded at the Rivière des Mille Îles between 1983 and 2013. The left panel shows the entire series and the right panel shows the time and size of the extreme low-flow observations, defined as the exceedances over the threshold $u = -13.66$ m corresponding to the 96% quantile of this series.

To ensure reliable predictions, low-flow clustering should be taken into account in water level exceedance modeling. To illustrate this point, a classical extreme-value analysis was performed on the Rivière des Mille Îles data by fitting a generalized Pareto distribution (GPD) to the values beyond the 96% quantile. This model, which involves a scale parameter $\beta \in (0, \infty)$ and a shape parameter $\xi \in \mathbb{R}$, is based on the assumption that the exceedances are mutually independent and identically distributed (iid). Its cumulative distribution is given, for all $x \in (0, \infty)$, by

$$\mathrm{GPD}_{\beta,\xi}(x) = \begin{cases} 1 - (1 + \xi x/\beta)^{-1/\xi} & \text{if } \xi \neq 0, \\ 1 - \exp(-x/\beta) & \text{if } \xi = 0. \end{cases} \tag{1}$$

[1]For the construction of mean excess plots and an introduction to classical extreme-value theory, see, for example, the book by Coles (2001).

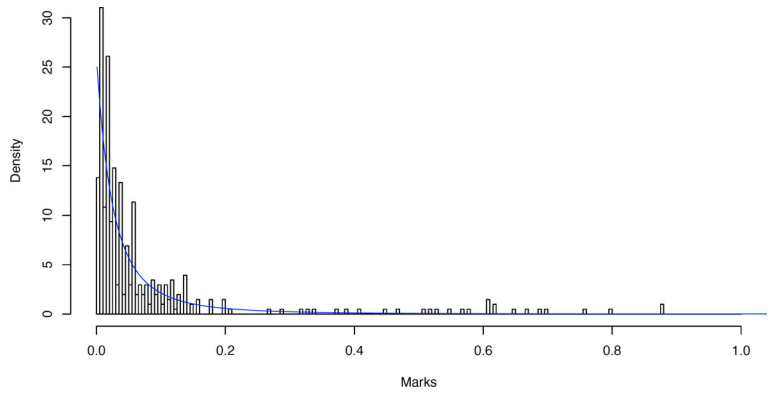

FIGURE 3 Fitted generalized Pareto distribution density overlaying a histogram of the (negated) exceedances beyond the 96th quantile of water levels observed at the Rivière des Mille Îles (1983–2013)

For the data at hand, the GPD parameter estimates (and their standard error [SE]) are $\hat{\beta} = 0.0384$ (0.003) and $\hat{\xi} = 0.537$ (0.070). The model's inadequacy transpires from the probability and quantile plot (not shown) and in Figure 3, where the estimated GPD density clearly fails to capture the heavy tail of the overlaid excess-amount histogram.

The standard way of dealing with clustered extremes in a dependent stationary sequence X_n is the peak over threshold (POT) approach, in which the original series is declustered and the GPD model is fitted to the resulting set of approximately independent peak excesses. This approach is justified by the fact that the conditional distribution of the maximum M_{C_n} of a cluster C_n in excess of an appropriate threshold u_n is asymptotically the same as that of any individual observation X_i (Davison & Smith, 1990), that is, for all $x \in \mathbb{R}$ and a suitable sequence $s_n \in (0, \infty)$,

$$\Pr\{(M_{C_n} - u_n)/s_n > x \,|\, M_{C_n} > u_n\} = \Pr\{(X_i - u_n)/s_n > x \,|\, X_i > u_n\} + o(1). \tag{2}$$

Therefore, the marginal excess distribution can be estimated by fitting a GPD to the set of cluster maxima. However, the declustering procedure on which this approach relies may induce estimation bias. In the runs declustering method (O'Brien, 1987), for example, two exceedances are deemed to belong to the same cluster if they are separated by less than r nonexceedances. However, the choice of the run-length parameter r is delicate. If r is too small, a major drought can be mistakenly divided into several dependent small droughts. In contrast if r is too large, two independent droughts could be mistakenly pooled together, leading to loss of information on extremes. Moreover, different choices of r may lead not only to very different model estimates for the cluster statistics but also for the exceedances.

In the present case, there are 74 clusters when $r = 1$ and the mean cluster size is then 5.48. When $r = 14$, however, there are only 32 clusters and their average size is 17.12. Similarly, the GPD parameter estimates (and their SE) are $\hat{\beta} = 0.032$ (0.006) and $\hat{\xi} = 0.544$ (0.151) when $r = 1$ and $\hat{\beta} = 0.063$ (0.018) and $\hat{\xi} = 0.54$ (0.24) when $r = 14$.

Given the importance of low-flow clustering in estimating the risk of drought and microorganism contamination, it is proposed here to model these exceedances with a self-exciting marked point process (SEMPP), which captures the clustering behavior of threshold exceedances through a self-exciting component embedded in its intensity function. The proposal is inspired by the work of Chavez-Demoulin et al. (2005) and Chavez-Demoulin and McGill (2012), who used this approach to capture volatility clustering in extreme log-return modeling. SEMPPs have since been adapted to other financial settings; see Stindl and Chen (2019) and references therein. As will be seen, this approach can be transposed rather straightforwardly to hydrological contexts, particularly when the data are gathered on a fine time grid.

There are, however, at least two major ways in which hydrological applications differ from financial contexts. First, observations are typically made on a coarse grid, which may be monthly or daily, the latter being the case for the Rivière des Mille Îles data. In such circumstances, a discrete version of the SEMPP may be preferable. This variant will also be presented and fitted to the data, and contrasted with the continuous-time version. Second, hydrological time series are often relatively short, and practical considerations often dictate a threshold which is too low to ensure the validity of relation (2). This is apparent in Figure 3, where significant lack of fit is actually observed in both tails of the distribution.

As mentioned by Fawcett and Walshaw (2007) and Eastoe and Tawn (2012), the GPD model with cluster maxima tends to overestimate the marginal excess distribution at subasymptotic threshold levels, thereby affecting return level estimation. In the POT model for the Rivière des Mille Îles data with $r = 14$, for example, the estimate of the average

excess amount (0.14) is 75% larger than for the marginal GPD (0.08). A much smaller discrepancy would be expected if approximation (2) held at the selected threshold level.

To address the latter issue, we incorporate into discrete and continuous versions of the SEMPP a time-varying mark distribution from the class of extended generalized Pareto distributions (EGPD) due to Papastathopoulos and Tawn (2013). This wide class of distributions has already been used in environmental applications, see Naveau et al. (2016), but seemingly not in conjunction with the SEMPP model. Its use makes it possible to circumvent the bias-variance dilemma often met in choosing a threshold in practice, and hence improves drought risk estimation and event forecasting.

The way in which the extremes of an iid sequence can be represented through a marked Poisson point process (MPPP) with independent GPD marks is sketched in Section 2. In Section 3, SEMPP are seen to be a natural extension of MPPPs to dependent sequences with clusters. Section 4 makes a short incursion into the distribution of run statistics for the discrete-time SEMPP model. Inference tools for the two models are then described in Section 5 and possible uses of the models are discussed in Section 6. While SEMPPs have found many applications in finance since the work of Chavez-Demoulin et al. (2005), their value for hydrological modeling is explored here for the first time and illustrated through an analysis of the Rivière des Mille Îles data presented in Section 7. The article concludes with a short discussion in Section 8.

2 | POINT PROCESS FRAMEWORK

The extreme values of a stationary time series are traditionally characterized either through block maxima or threshold exceedances. The fact that generalized extreme-value (GEV) distributions are the only possible limits of properly normalized sequences of maxima (Fisher & Tippett, 1928) lies at the root of the well-known block maxima approach to extreme-value analysis. The alternative POT approach advocated by Davison and Smith (1990) is based on a characterization of the asymptotic tail distribution of random variables due to Pickands III (1975).

As initiated by Pickands himself and further discussed by Leadbetter et al. (1983) and subsequent authors, point processes provide a convenient way of characterizing extreme values that unifies the block maxima and POT approaches. This is accomplished by taking into account both the size of the exceedances and the time at which they occur. While this approach is probabilistically more complex, it allows for the more advanced type of modeling required in the presence of clusters of observations. For the reader's convenience, this section provides a short introduction to this approach, based on standard references such as the books by Embrechts et al. (1997), Coles (2001), and Resnick (2007).

2.1 | Point processes

A point process is a random counting measure which captures the pattern of a sequence $Z_1, Z_2, \ldots$ of (almost surely) distinct observations in a given state space $\mathcal{E}$. To fix ideas, think of the random times, measured on a continuous scale, when exceedances occur in the interval $\mathcal{E} = [0, T]$. The corresponding point process N is defined by associating to any (measurable) set $A \subseteq \mathcal{E}$ the number $N(A)$ of observations within A. It is required that $N(A) < \infty$ whenever A is compact, in which case $N(A)$ is an integer-valued random variable; see, for example, chapter 3 of Resnick (2007). The mean measure of the process N, given by $\Lambda(A) = \mathrm{E}\{N(A)\}$ for any compact set $A \subseteq \mathcal{E}$, is then well defined but not necessarily finite.

A prime example is the Poisson point process (PPP), which is such that (i) for any set $A \subseteq \mathcal{E}$ and integer n, $\Pr\{N(A) = n\} = e^{-\Lambda(A)}\{\Lambda(A)\}^n/n!$ if $\Lambda(A) < \infty$ and zero otherwise; (ii) for any integer $k \geq 2$, the random variables $N(A_1), \ldots, N(A_k)$ are mutually independent whenever $A_1, \ldots, A_k \subseteq \mathcal{E}$ are mutually disjoint. The intensity function of this process is the (Radón–Nikodym) derivative λ of Λ, which satisfies, for any (measurable) set $A \subseteq \mathcal{E}$, $\Lambda(A) = \int_A \lambda(z)dz$.

When time is measured on a discrete scale, say $\mathcal{E} = \mathbb{N} = \{1, 2, \ldots\}$, the state space is countable and a set $A \subseteq \mathcal{E}$ is compact if and only if it is finite. The variables $Z_1, Z_2, \ldots$ are then the (almost surely) distinct random times in $\mathcal{E}$ at which the event of interest, say an exceedance, occurs. Such a discrete-time point process can also be represented as a sequence $Y_1, Y_2, \ldots$ of random variables with $Y_n = 1$ if an event occurs at time $n \in \mathcal{E}$ and $Y_n = 0$ otherwise (Lewis, 1970).

2.2 | Marked point processes

The above construction remains valid for a sequence $Z_1, Z_2, \ldots$ of (almost surely) distinct observations taking values in more general state spaces. In particular, suppose that for every $n \in \mathbb{N}$, $Z_n = (T_n, M_n)$, where $T_n \in [0, T]$ denotes the time

at which an event occurs, say again an exceedance, and $M_n \in \mathcal{M}$, say $\mathcal{M} = (0, \infty)$, is the magnitude of this exceedance. The resulting process N is then called a marked point process (MPP) in continuous time, and the value of M_n is called the mark associated with time T_n. This process is well-defined provided that the associated marginal process defined by $N_g(A) = N(A \times \mathcal{M})$ for any (measurable) $A \subseteq [0, T]$, known as the ground process, satisfies the finiteness condition for compact sets. Taking $T_n \in \mathbb{N}$ instead leads to a discrete-time MPP; see, for example, Sigman and Whitt (2019).

MPP form a broad class of models induced by a ground process, a mark structure, and the relation between them. The behavior of an MPP on $\mathcal{E} = [0, T] \times \mathcal{M}$ is characterized by the joint intensity function given, for all $(t, m) \in \mathcal{E}$, by $\lambda(t, m) = \lambda_g(t) f(m \mid t)$, where λ_g is the intensity function of the ground process and f is the conditional distribution of the mark on $\mathcal{M}$. In this continuous-time model, λ is the instantaneous risk of observing an event at time t with associated mark in the neighborhood of m conditionally on the past. The discrete analog has a similar interpretation.

An MPP is said to have predictable marks if the distribution of the mark at time T_n depends on the path of the process N up to that point. By contrast, the marks are said to be unpredictable if the distribution of the mark at time T_n is independent of history, that is, of all pairs (T_k, M_k) with $T_k < T_n$. Nevertheless, unpredictable marks can still affect the future evolution of the ground process as the ground intensity function of N_g could depend on past marks. The assumption that the mark process is completely independent of the ground process is a restrictive special case.

2.3 | MPP characterizations of extreme values

In continuous time, the MPP methodology encompasses both the block maxima and threshold exceedance approaches to the analysis of extreme values. To see why, suppose that observations $X_1, X_2, \ldots$ form an iid sequence whose distribution F belongs to the maximum domain of attraction of a GEV distribution $H_{\mu,\sigma,\xi}$ with location parameter $\mu \in \mathbb{R}$, scale $\sigma > 0$, and shape $\xi \in \mathbb{R}$. Recall that for all $x \in \mathbb{R}$, $H_{\mu,\sigma,\xi}(x) = H_\xi\{(x - \mu)/\sigma\}$, where

$$H_\xi(x) = \begin{cases} \exp\left\{-(1 + \xi x)^{-1/\xi}\right\} & \text{if } \xi \neq 0, \\ \exp(-e^{-x}) & \text{if } \xi = 0, \end{cases}$$

whenever $1 + \xi x > 0$. Setting $M_n = \max(X_1, \ldots, X_n)$ for each $n \in \mathbb{N}$, one can find sequences of constants $a_n \in \mathbb{R}$ and $b_n \in (0, \infty)$ such that $\Pr\{(M_n - a_n)/b_n \leq x\} \to H_{\mu,\sigma,\xi}(x)$ as $n \to \infty$ for all $x \in \mathbb{R}$. Given μ, σ and ξ, let $\mathcal{M}_{\mu,\sigma,\xi} = (m_1, m_2]$, where m_1 equals $\mu - \sigma/\xi$ if $\xi > 0$ and $-\infty$ otherwise, while m_2 equals $\mu - \sigma/\xi$ if $\xi < 0$ and ∞ otherwise. Consider the time-rescaled MPP defined on the space $\mathcal{E} = [0, 1] \times \mathcal{M}_{\mu,\sigma,\xi}$, for each $n \in \mathbb{N}$, by

$$N_n = \sum_{i=1}^{n} \epsilon_{(i/n,\ (X_i - a_n)/b_n)}, \tag{3}$$

where ϵ_a denotes a point mass at $a \in \mathcal{E}$.[2] The result below, adapted from theorem 5.7.2 of Leadbetter et al. (1983) or corollary 4.19 of Resnick (2007), specifies the weak limit of the sequence $N_1, N_2, \ldots$

Proposition 1. *Consider an iid sequence $X_1, X_2, \ldots$ with maximum attractor $H_{\mu,\sigma,\xi}$. Then as $n \to \infty$, the sequence N_n defined in (3) converges weakly to a PPP N with intensity measure Λ given by $\Lambda(A) = -(t_2 - t_1)\ln\{H_{\mu,\sigma,\xi}(x)\}$ for any $A = [t_1, t_2] \times (x, m_2]$ with $0 \leq t_1 < t_2 \leq 1$ and $x \in (m_1, m_2)$.*

Given that $(M_n - a_n)/b_n \leq x \Leftrightarrow N_n\{[0, 1] \times (x, m_2]\} = 0$, it follows from Proposition 1 that, for all $x \in (m_1, m_2)$,

$$\lim_{n \to \infty} \Pr\{(M_n - a_n)/b_n \leq x\} = e^{-\Lambda\{[0,1] \times (x, m_2]\}} = H_{\mu,\sigma,\xi}(x),$$

which is the classical result upon which rests the block maxima approach. To see how the threshold exceedance model also derives from Proposition 1, note that for any $x \in (m_1, m_2)$, $N_n\{[0, 1] \times (x, m_2]\}$ is a sequence of binomial random variables. Hence when $n \to \infty$, $\mathrm{E}[N_n\{[0, 1] \times (x, m_2]\}] \to \mathrm{E}[N\{[0, 1] \times (x, m_2]\}] = \Lambda\{[0, 1] \times (x, m_2]\} = -\ln\{H_{\mu,\sigma,\xi}(x)\}$. In particular if $\xi > 0$, $\mu = 0$ and $\sigma = 1$, then for sufficiently large threshold $u \in (0, \infty)$ and large n, one has

$$\Pr(X_i > u + x \mid X_i > u) \approx \frac{-\ln[H_{\mu,\sigma,\xi}\{(u + x - a_n)/b_n\}]}{-\ln[H_{\mu,\sigma,\xi}\{(u - a_n)/b_n\}]} = (1 + \xi x/\beta_n)^{-1/\xi}$$

[2]The points such that $(X_i - a_n)/b_n \notin \mathcal{M}_{\mu,\sigma,\xi}$ are ignored for simplicity; the probability of such occurrences tends to zero as $n \to \infty$.

with $\beta_n = b_n[\sigma + \xi\{(u - a_n)/b_n - \mu\}]$, which establishes the GPD as an approximate distribution for threshold exceedances.

2.4 | Implication for extreme-value modeling

Smith (1989) pioneered the use of Proposition 1 for extreme-value modeling. For large n, the MPP

$$N_n = \epsilon_{(1/n,X_1)} + \ \ldots \ + \epsilon_{(n/n,X_n)}$$

is well approximated by a Poisson process N on regions of the form $[0,1] \times (u, m_2]$ whenever the threshold u (whose choice is constrained by the sample size) is sufficiently close to m_2. The intensity function of N is then given, for all $t \in [0,1]$ and $x \in (u, m_2)$, by

$$\lambda(t,x) = \lambda(x) = \begin{cases} \sigma^{-1}\{1 + \xi(x-\mu)/\sigma\}^{-1-1/\xi} & \text{if } \xi \neq 0, \\ \sigma^{-1}\exp\{-(x-\mu)/\sigma\} & \text{if } \xi = 0. \end{cases} \tag{4}$$

Loosely speaking, $\lambda(t,x)$ is the probability of an exceedance in a small interval around $t \in (0,1)$ with mark in the neighborhood of x. As the function λ in Equation (4) does not depend on t, the intensity of the time process is constant. That is, the ground process N_g is a homogeneous PPP, and $\tau = \mathrm{E}\{N_g([0,1])\} = -\ln\{H_{\mu,\sigma,\xi}(u)\}$ is the expected number of exceedances above u. Moreover, the conditional mark distribution $f(x \mid t) = f(x)$ is given, for all $x \in (u, m_2)$, by $f(x) = \beta^{-1}\{1 + \xi(x-u)/\beta\}^{-1-1/\xi}$, where $\beta = \sigma + \xi(u - \mu)$ with $\xi \neq 0$, as will be assumed in the data application.

3 | SELF-EXCITING MARKED POINT PROCESSES

While exceedances in an iid sequence can be modeled with an MPPP with GPD marks that are independent of Poisson occurrence times, water levels and other hydrological phenomena often involve dependent sequences with clustered extremes. Happily, an SEMPP makes it possible to model dependent sequences of exceedances and hence to account simultaneously for clustering patterns in the time dimension and dependence between marks and time.

If exceedance times can be measured on a scale which is such that the probability of observing more than one event within a measurement interval is negligible, the point process of exceedances of a dependent sequence can be captured by a continuous-time SEMPP. In many hydrological applications, however, data are measured at a fixed frequency (e.g., daily or hourly) and thus the exceedances only occur at integer multiples of the sampling intervals. In such cases, it may be more appropriate to model the exceedance process with a discrete-time SEMPP.

In the continuous case, the prototype of an SEMPP is the process introduced by Hawkes (1971) as a special class of a continuous temporal point process. It is characterized by an intensity function of the form $\lambda(t) = \zeta + \nu(t)$ at any time $t \in [0, T]$, where ζ is the background rate of the process which mirrors the average arrival rate of clusters, and

$$\nu(t) = \sum_{i:\, t_i < t} \phi(t - t_i) = \int_0^t \phi(t-s)dN(s), \tag{5}$$

is the self-exciting component, which involves a function ϕ governing the clustering density of N. This self-exciting point process (SEPP) has been widely used in seismology and earthquake modeling, notably by Ogata (1988), see also Ogata (1998), and in many other fields where clustered events are common. Financial applications are particularly abundant, as documented by Embrechts et al. (2011) or Chavez-Demoulin and McGill (2012) and references therein. SEMPPs expand on the original concept of SEPP and as for regular MPP, they are specified by a ground process and a mark structure.

3.1 | Ground process

The ground process of an SEMPP is a temporal point process intended to model the occurrence times of exceedances. Like an SEPP, an SEMPP seeks to capture clustering effects through a self-driven mechanism but it expands on the concept

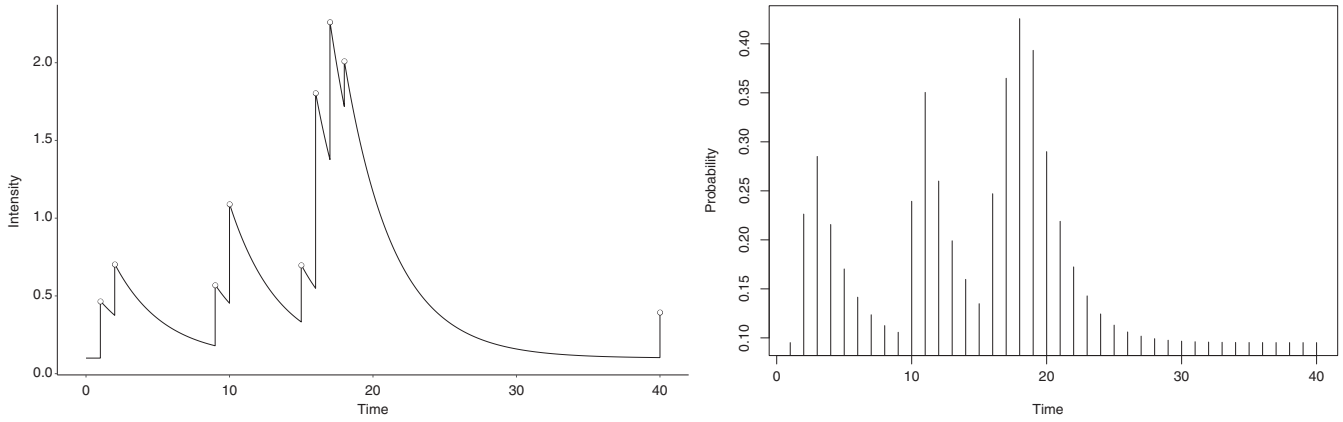

FIGURE 4 Typical path of a generic ground intensity function (left panel) and probability sequence (right panel)

of SEPP by taking into account the interplay between the ground process and the mark structure. This is appealing for hydrological applications because in many natural phenomena, past events tend to influence the system for some time as a function of their magnitude. The ground intensity function of a continuous-time SEMPP is again expressed at any $t \in [0, T]$ in the form $\lambda_g(t) = \zeta + \nu(t)$ with ζ denoting once more the background intensity, but where the self-exciting component is an extension of Equation (5) given by

$$\nu(t) = \sum_{i:t_i<t} \phi(t - t_i)g(m_i) = \iint_{(0,t)\times(0,\infty)} \phi(t - s)g(m)dN(s, m), \tag{6}$$

with the function g quantifying the impact of event i with mark m_i occurring at time t_i, and ϕ determining for how long this effect is felt as a function of the time $t - t_i$ elapsed since that event took place.

Formally, $\lambda_g(t) = \lim_{\Delta t\to 0} \Pr[N_g\{[t, t + \Delta t)\} > 0 \mid \mathcal{H}_t]/\Delta t$ is conditional on the history $\mathcal{H}_t$ of the process up to time t, where $\{\mathcal{H}_t : t \in [0, T]\}$ is a filtration on the same probability space on which the point process is defined. The list-history H_t is a set $\{(t_1, m_1), \ldots, (t_n, m_n)\}$ that records the actual times and marks of events observed up to time t.

For a discrete-time SEMPP, the event times $t_1, t_2, \ldots$ only take integer values, so the ground process becomes a discrete-time point process. As mentioned in Section 2.1, it can be represented as a finite binary sequence $Y_1, \ldots, Y_T$. The discrete analog of the ground intensity function is a sequence $p_1, p_2, \ldots$ of conditional probabilities, where for each integer $t \in \mathbb{N}$, $p_t = \mathrm{E}(Y_t \mid \mathcal{H}_{t-1}) = \Pr(Y_t = 1 \mid \mathcal{H}_{t-1})$. Daley and Vere-Jones (2003) considered a discrete Hawkes process with autoregressive probabilities $p_t = \mathrm{logit}^{-1}(a_0 + a_1 Y_{t-1} + \ldots + a_K Y_{t-K})$. More generally, one can set

$$p_t = \varphi\{\lambda_g(t)\} = \varphi\{\zeta + \nu(t)\}, \tag{7}$$

where the ground intensity function is mapped to the probability range through a nonlinear transfer function $\varphi : \mathbb{R} \to [0, 1]$. In this formulation, the autoregressive order K is modeled stochastically by the memory kernel ϕ and the set of autoregression coefficients $a_1, a_2, \ldots$ is accounted for by the impact function g.

Whether for the discrete- or continuous-time SEMPP, the past events have an excitation effect on future events by increasing their probability of occurrence. This effect is additive over the past events and decays with time. The memory kernel ϕ and the impact function g are introduced to modulate the magnitude and the decaying rate of the excitation effect. Figure 4 shows a realization of the ground intensity function (left) and the probability sequence (right) for a generic list history H_t for an exponential decay kernel ϕ and the commonly used linear impact function g. These are, respectively, defined, for all $(t, m) \in [0, T] \times (0, \infty)$, by $\phi(t) = e^{-\gamma t}$ and $g(m) = 1 + \delta m$ with parameters $\gamma, \delta \in (0, \infty)$.

Looking at Figure 4, one can see that each event induces a jump whose size $\phi(0)g(m)$ depends on the mark size m. Each jump is followed by an exponential decay but different events have different decay rates influenced by past history; this occurs even though γ is constant because of the superposition of different exponential decays. Another popular choice of kernel is the power law decay kernel used in epidemic-type after-shock models for earthquakes; this kernel is defined, for all $t \in [0, T]$, by $\phi(t) = 1/(t + \gamma)^{\rho+1}$ for some parameters $\rho, \gamma \in (0, \infty)$. Both of the above decay functions have explicit distributions, which is convenient for computing the induced likelihood.

Generally speaking, a jump in intensity increases the probability of arrivals of secondary events triggered by the first. If no event occurs over a sufficiently long period of time, the intensity gradually reduces to the background intensity ζ, which signals that the triggering effect of the previous cluster has subsided. The memory kernel ϕ determines how fast this triggering effect decays over time. If ϕ decays very quickly, the process only depends on the recent events but if ϕ decays slowly, the process may exhibit long-range dependence.

As remarked in proposition 6.4 VII of Daley and Vere-Jones (2003), to guarantee the existence of a stationary version of the marked Hawkes process, one should have $\rho = \int g(m)\phi(t)dtF(dm) < 1$.[3] The impact function should thus be chosen to ensure that its expectation with respect to the mark distribution is finite, that is, $\mathrm{E}\{g(m)\} = \int g(m)F(dm) < \infty$. Linear impact functions defined earlier are a common choice; another option is to take, for all $m \in (0, \infty)$, $g(m) = e^{\delta m}$ for some parameter $\delta \in (0, \infty)$; see, for example, Embrechts et al. (2011). In the data application, however, the marks are extreme exceedances that are heavy-tailed, so a bounded impact function must be selected to meet the stationary condition. For this reason, we will take $g(m) = 1 + \tanh(\delta m)$ for all $m \in (0, \infty)$.

3.2 | Extended GPD for marks

In specifying the mark structure of an SEMPP, we aim to construct a model that can capture the excess amount at different event times. The classical GPD is typically the first choice to model threshold exceedances, as justified by the Pickands–Balkema–de Haan theorem. This result states that if X_n is an iid sequence with common marginal F and there exists a scaling function h for which the scaled excess distribution converges to a nondegenerate distribution as the sequence u_n of thresholds converges to the upper end point x_F of F, then the limit must be of the form (1). The scaling function h depends on F but because the latter is unknown in practice, h is commonly absorbed into the estimation of the scale parameter σ_u, which depends on the threshold u.

While the exceedances of the Rivière des Mille Îles data displayed in Figure 3 are broadly consonant with the use of GPD marks, it was already mentioned in the Introduction that two characteristics of the phenomenon fail to be captured by a GPD model fitted through maximum likelihood. Specifically, (A) the accumulated mass observed in the neighborhood of the origin; (B) the heaviness of the right tail of the distribution.

The lack of fit of the GPD model arises here in part because the 96% quantile threshold selected by the City of Montréal water engineers based on the right panel of Figure 1 is too low for the asymptotic approximation to be reasonable. This phenomenon is frequent in hydrology, where the paucity of data forces practitioners to opt for subasymptotic thresholds for which the extreme-value limit is a poor approximation for the conditional excess distribution. Another reason might be the seasonality present in the data, which was not modeled as such.

To overcome issue (A), we opt for an SEMPP based on the EGPD class due to Papastathopoulos and Tawn (2013). Each EGPD is constructed through a probability integral transform $X = \sigma H_\xi^{-1}(V)$, where H_ξ^{-1} is the quantile function of a standard GPD and V is a random variable on $(0, 1)$ with distribution G and density g. One then has $F_\eta(x) = G\{H_\xi(x/\sigma)\}$ with $\eta = (\xi, \sigma)$ and the corresponding density function is given, for all $x \in (0, \infty)$, by $f_\eta(x) = \sigma^{-1}h_\xi(x/\sigma)$ $g\{H_\xi(x/\sigma)\}$.

The key in constructing a successful EGPD is to find a suitable parametric class G_κ of transition functions such that F preserves the GPD upper tail while characterizing the lower-tail behavior. To satisfy the upper tail equivalency, one must have $\{1 - F_\eta(x; \kappa)\}/\{1 - H_\xi(x/\sigma)\} \to c \in (0, \infty)$ as $x \to \infty$. Upon setting $v = H_\xi(x/\sigma)$, one can verify that this is equivalent to the requirement that $\{1 - G_\kappa(v)\}/(1 - v) \to c$ as $v \to 1$, that is, $1 - G_\kappa$ decays to zero at a linear rate.

As the support of F is bounded below, the lower tail can either be regularly varying and belong to the Weibull domain, or it is rapidly varying and belongs to the Gumbel domain. Naveau et al. (2016) gave conditions on the transition function G which lead to the former scenario. This brought them to recommend a Beta transition function defined, for all $v \in (0, 1)$, by $G_\kappa(v) = v^\kappa$. This choice preserves the GPD upper tail while providing flexibility in, and ensuring a smooth transition to, the lower tail. The latter is GPD when $\kappa = 1$ while other values of κ make it possible to model exceedances over a subasymptotic threshold without affecting the upper tail. For lower tails in the Gumbel domain, one could use an exponential transition function of the form $G_\kappa(v) = e^{\kappa - \kappa/v}$ with $v \in (0, 1)$. The EGPD is then such that, as $x \to 0$, $F_{\eta,\kappa}(tx)/F_{\eta,\kappa}(x) \to \infty$ if $t > 1$ and goes to zero when $t \in (0, 1)$.

Given that the EGPD and the GPD share the same upper tail behavior, issue (B) above remains unresolved. Accordingly, an SEMPP with iid marks would be insufficient to explain the heavy tail in the marks, even if their distribution were

[3]No such restriction is needed in the discrete-time case because the number of events is always finite.

EGPD. To address Issue (B), therefore, we will rely on a model where the mark density can be affected by the evolution of the ground process through the linking self-exciting function, so that clustering information in the time dimension can be embodied in the tail inference on the mark space. This can be achieved by incorporating a time-varying scaling parameter σ_t as a monotonic increasing function of the self-exciting function. Instead of being deterministic, the self-exciting function is then a Markov process that depends on the realization path of previous events. In what follows, we will set $\sigma_t = \beta_0 + \beta_1 \nu(t)$ and $\nu(t)$ be defined as in Equation (6), so that the mark structure of the SEMPP model is specified by the density such that, for all $m \in (0, \infty)$,

$$f_M(m \mid t, \mathcal{H}_t) = \sigma_t^{-1} \mathrm{g}_\kappa\{H_\xi(m/\sigma_t)\} h_\xi(m/\sigma_t). \tag{8}$$

3.3 | Cluster process representation

One of the advantages of SEPP is that they have an equivalent cluster process representation that turns out be very useful in interpreting, fitting, and simulating the process. Hawkes and Oakes (1974) showed that any stationary SEPP with finite intensity can be represented as a Poisson cluster process or generalized branching PPP. The equivalence stems from the one-to-one relationship uncovered by Daley and Vere-Jones (2007) between Cox shot-noise processes and Neyman–Scott cluster processes.

In the cluster representation, each event can be classified as either an immigrant or an offspring. The immigrants arrive according to a homogeneous PPP with intensity parameter ζ, forming the background immigrant process N_c. Each immigrant in N_c generates offspring events according to a hidden branching process structure: given an immigrant event (t_j, m_j), the first generation arrives according to a nonhomogeneous PPP with intensity function $\phi(t - t_j)\mathrm{g}(m_j)$. The offspring of subsequent generations are produced recursively and an immigrant's offspring form a cluster.

The cluster representation permits easy simulation of the SEMPP model and also leads to useful interpretations of the parameters. A key parameter under this branching structure is the branching ratio ρ, which is defined in terms of the mean measure of the Poisson offspring process, namely

$$\rho = \iint g(m)\phi(t)dtF(dm).$$

The total number of events in a cluster N is equivalent to the total size of a branching process in which the number of offspring of a single event follows a Poisson distribution with mean ρ. When $\rho < 1$, the expected total cluster size is the limiting sum of a geometric series, namely, $\mathrm{E}(N) = 1/(1 - \rho)$. The background or immigration intensity ζ measures the expected number of arrivals of immigrant events per unit of time. Accordingly, ζT, where T is the length of the observed time, is the expected total number of clusters. The branching ratio ρ gives the proportion of events that are triggered by previous events, so it can be alternatively used as an indicator of the clustering effect among events. For processes in subcritical regime, the branching ratio is between 0 and 1, approaching the latter as the clustering effect strengthens. Therefore, for an SEPP, crucial information on the cluster statistics can be deduced from the cluster process representation without manually declustering the original series.

The cluster process interpretation remains justified for the discrete-time SEMPP. The background immigrant process is then a homogeneous Bernoulli process with constant occurrence probability $\varphi(\zeta)$ and as a result, the number of clusters is binomial with mean $\varphi(\zeta)T$. Each immigrant gives rise to a sequence of offspring events and the offspring process is a dependent Bernoulli process where the occurrence probability is modulated by the self-exciting component. A cluster is deemed to terminate as the probability of occurrence returns to the background level. As the probability structure of the discrete-time point process is simpler, the exact distribution of some cluster statistics can be derived in the form of run statistics, as detailed in the following section.

4 | DISTRIBUTION OF THE RUN STATISTICS

In the discrete-time SEMPP model, a run is a sequence of consecutive exceedances preceded and succeeded by at least one nonexceedance. The statistics of interest include the run length, the length of the longest run, and the waiting time to a run of a given length. Each of these statistics is considered below in turn.

4.1 | Distribution of the run length

Let X_t be a stationary sequence. Given a threshold u, its exceedance process can be represented as a binary sequence Y_t. Let Y_1 denote the first exceedance of an isolated cluster (i.e., all the observations prior to it are nonexceedances), and N be the total number of consecutive exceedances following it (including Y_1). Then for all $k \in \{1, 2, \dots\}$,

$$\Pr(N = k) = \Pr(Y_2 = 1, \dots, Y_k = 1, Y_{k+1} = 0 \mid Y_1 = 1) = (1 - p_k)\, p_1 \dots p_{k-1}, \tag{9}$$

where p_i denotes the success probability of exceedance given there are i consecutive exceedances before it. Therefore, N follows a geometric-like distribution with nonidentical probabilities. For the constant impact model, $g(m) = \psi$ and p_i is a deterministic quantity, namely, $p_i = \varphi[\zeta + \psi\{\phi(1) + \dots + \phi(i)\}]$. Otherwise, one must take into account the extra randomness introduced by the marks by integrating over the mark distribution, that is,

$$p_i = \int \varphi \left\{ \zeta + \sum_{j=1}^{i} g(m_j)\phi(i + 1 - j) \right\} dF(m_1, \dots, m_i),$$

where F is the joint distribution of the marks. Hence if $\lim_{n\to\infty} p_n < 1$, the expected run length exists and is given by

$$\mathrm{E}(N) = \sum_{k=1}^{\infty} k(1 - p_k) \prod_{i=1}^{k-1} p_i = 1 + \sum_{n=1}^{\infty} \prod_{i=1}^{n} p_i. \tag{10}$$

4.2 | Waiting time distribution and return period

In deriving the distribution of the other two run statistics, we adopt the Markov chain approach introduced by Fu and Koutras (1994). The main idea is to embed the run statistics into a Markov chain with finite state space, so that the distribution of the run statistics is transformed to the probability distribution of the Markov chain.

Let T_{k*} be the waiting time to the first consecutive exceedances of at least length k. To find the distribution of T_{k*}, we consider the Markov chain W_t, where for each integer $t \geq 0$, W_t is the number of consecutive exceedances counting backward from t. The associated state space is $\Omega = \{\omega_0, \dots, \omega_k, \omega_{k*}\}$, with ω_i indicating that the chain has just reached i consecutive exceedances, ω_k indicating that the chain has seen k or more consecutive exceedances, and ω_{k*} indicating that the chain has seen k or more consecutive exceedances followed by a nonexceedance.

Given that W_t is in state ω_i, there is a probability p_i that the next observation Y_{t+1} is an exceedance. If an exceedance occurs, W_{t+1} moves to state ω_{i+1}; if a nonexceedances occurs, the chain jumps back to state ω_0. If the chain has reached state ω_k, it will move to ω_{k*} only if the run ends with a nonexceedance. Finally, once the chain enters state ω_{k*}, it will never leave, that is, ω_{k*} is an absorbing state. Let P be the $(k + 2) \times (k + 2)$ one-step transition matrix associated with this Markov chain on Ω. The probability of having seen at least k consecutive exceedances up to time n is equivalent to the probability that the chain has reached state ω_{k*} at time $n + 1$. Therefore, for every integer $n \geq 1$, one has

$$\Pr(T_{k*} \leq n) = \Pr(W_{n+1} = \omega_*) = \pi_0 \left(\prod_{t=1}^{n+1} P_t \right) U^\top = \pi_0 P^{n+1} U^\top, \tag{11}$$

where $\pi_0 = (1, 0, \dots, 0)$ denotes the initial distribution and $U = (0, 0, \dots, 1)$. In formula (11), the probability transition matrix is homogeneous in time. For the SEMPP model with seasonal effect, however, the transition matrix P_t can depend on the seasonal variations. The expectation of T_{k*} can be computed from Equation (11), namely,

$$\mathrm{E}(T_{k*}) = \sum_{n=0}^{\infty} \Pr(T_{k*} > n) = \sum_{n=0}^{\infty} \{1 - \Pr(T_{k*} \leq n)\}. \tag{12}$$

Alternatively, $\mathrm{E}(T_{k*})$ can be viewed as the mean hitting time for the state ω_{k*}, which yields

$$\mathrm{E}(T_{k*}) = \left\{ 1 + \sum_{n=0}^{k-2} \prod_{i=0}^{n} p_i + \prod_{i=0}^{k-1} p_i/(1 - p_k) \right\} \Big/ \left\{ 1 - \sum_{n=0}^{k-1} (1 - p_n) \prod_{j=1}^{n-1} p_j \right\}.$$

The Markov chain method can also be used to derive the distribution of the longest run L_n of exceedances observed over n trials. It can be shown that L_n can be embedded into the same Markov chain defined above. The event $\{L_n < k\}$ implies that the chain has not seen any consecutive exceedances of length greater than or equal to k until time n, which is equivalent to saying that W_n is not in state ω_k nor ω_{k*}. Thus, setting $U = (1, \dots, 1, 0, 0)$, one finds

$$\Pr(L_n < k) = \Pr(W_n \in \Omega \setminus \{\omega_k, \omega_{k*}\}) = \pi_0 P^n U^\top. \tag{13}$$

It should be noted that in the SEMPP model, W_t is not exactly a Markov chain, because the probability of exceedance p_t depends on the whole past history $\mathcal{H}_{t-1}$. But because the dependence is decaying to zero as specified by the decaying kernel ϕ, it can be well approximated by a Kth order autoregressive model for some large number K. As K increases, the probability estimate given by the Markov chain approach converges to the true quantity under the fitted SEMPP model. The value of K can be chosen as the smallest number achieving a certain convergence tolerance. Given the chosen value of K, one can enrich the state space Ω to take into account at least the past K observations; the dimension of the state space then increases to $k - K + 2^K + 1$. For example, ω_2 is now expanded to 2^{K-3} states because if state ω_2 is reached, the three observations prior to it are $(0, 1, 1)$, so there are 2^{K-3} scenarios in ω_2 and each of them has a different probability of exceedance depending on the permutations of the $K - 3$ digits to the left.

5 | INFERENCE FOR THE PROPOSED MODEL

This section describes how to make inference for the continuous- and discrete-time SEMPP models based on a realization $(t_1, m_1), \dots, (t_n, m_n)$ on $[0, T] \times [0, \infty)$ of the marked exceedance process of a dependent time series X_t above a large threshold u, where t_i records the time of the ith exceedance and m_i the corresponding amount.

The continuous-time SEMPP model is completely specified by the product of the ground intensity function (6) and the conditional mark density function (8). For all $t \in [0, T]$ and $m \in (0, \infty)$, one has

$$\lambda(t, m_t) = \{\zeta + \nu(t)\} f(m_t \mid t, \mathcal{H}_t), \tag{14}$$

where $\sigma_t = \beta_0 + \beta_1 \nu(t)$ and $\nu(t)$ is as defined in Equation (6).[4] The discrete-time analogue is given, for $t \in \{1, \dots, T\}$, by

$$f_{Y_t, M_t \mid \mathcal{H}_{t-1}}(t, m) = \Pr(Y_t = y_t \mid \mathcal{H}_{t-1}) f_{M_t}(m_t \mid y_t, \mathcal{H}_{t-1}) = \left\{p_t f_{M_t}(m_t \mid Y_t = 1, \mathcal{H}_{t-1})\right\}^{y_t} (1 - p_t)^{1 - y_t}.$$

Given a threshold u and n_u observations above it, the parameters of the model can be estimated by maximizing the log-likelihood function $\ln(L_1) + \ln(L_2)$, where

$$\ln(L_1) = \sum_{i=1}^{n_u} \ln(1/\sigma_{t_i}) + \sum_{i=1}^{n_u} \ln\left[g_\kappa\left\{1 - (1 + \xi\, m_i/\sigma_{t_i})^{-1/\xi}\right\}\right] - (1 + 1/\xi) \sum_{i=1}^{n_u} \ln(1 + m_i/\sigma_{t_i})$$

and $\ln(L_2)$ differs depending whether the SEMPP is in continuous- or discrete-time. In the former case,

$$\ln(L_2) = \sum_{i=1}^{n_u} \ln\{\lambda_g(t_i)\} - \int_0^T \lambda_g(u) du = -\zeta T - \sum_{i=1}^{n_u} g(m_i; \delta) B(T - t_i; \gamma) + \sum_{i=1}^{n_u} \ln\{\zeta + \nu(t_i)\}$$

with $B(t) = \int_0^t \phi(u) du$ while in the latter case,

$$\ln(L_2) = \sum_{t=1}^{T} \{y_t \ln(p_t) + (1 - y_t) \ln(1 - p_t)\}$$

with $p_t = \varphi\{\zeta + \nu(t)\}$. Note that the terms $\ln(L_1)$ and $\ln(L_2)$ are dependent and must be maximized jointly. Indeed, the ground process and the mark process are linked with a common self-exciting component $\nu(t)$ and the parameters relating

[4] To avoid negative values of σ_t, one could impose a constraint or work with $\ln(\sigma_t)$ instead; in the present application, this was unnecessary.

to it appear in both L_1 and L_2. This is in contrast to the classical case discussed in Section 2, where the marks are independent of the occurrence times, which leads to a likelihood factorizing into two independent parts. For results on the consistency and asymptotic normality of maximum likelihood estimators in the present context, see Ogata (1978).

The fit of a point process model can be assessed using residual analysis and model diagnostics for the marks. These two techniques are briefly described below.

5.1 | Point process residual analysis

Point process residual analysis is traditionally based on the random time transformation, which turns a point process with continuous conditional intensity function into a unit rate PPP. The following result, which supports this procedure, is excerpted from Daley and Vere-Jones (2003); see proposition 7.4.IV on p. 261.

Proposition 2. *Let N be a point process with continuous intensity function λ and compensator function $\Lambda(t) = \int_0^t \lambda(s)ds$. If the original sequence t_i is a realization from the point process N, then, with probability 1, the transformed sequence $\tau_i = \Lambda(t_i)$ is a realization of a unit rate PPP.*

For the discrete-time point process, the integral is simply replaced with a finite sum, that is, $\Lambda(t_i) = p_1 + \ldots + p_{t_i}$, which is the expected number of events observed on the time interval $[0, t_i]$. Therefore, when the intensity function of a point process is fixed, goodness-of-fit testing amounts to testing for departure from the corresponding unit rate PPP. A well-known property of the latter process is that conditionally on the number of arrivals, the arrival times are mutually independently and uniformly distributed. Thus, the transformed time sequence $\tau_1 = \Lambda(t_1), \ldots, \tau_n = \Lambda(t_n)$ are arrival times from a unit rate Poisson process. Equivalently, the rescaled ordered variables $w_1 = \tau_{(1)}/\Lambda(T) < \cdots < w_n = \tau_{(n)}/\Lambda(T)$ are order statistics from a uniform distribution on the interval $[0, 1]$.

When Λ is replaced by a consistent estimate $\hat{\Lambda}$, the statistics $\hat{w}_1 = \hat{\tau}_{(1)}/\hat{\Lambda}(T) < \cdots < \hat{w}_n = \hat{\tau}_{(n)}/\hat{\Lambda}(T)$ should behave approximately like order statistics from a standard uniform distribution. This can be checked visually with a probability plot based on the pairs $(1/n, \hat{w}_1), \ldots, (n/n, \hat{w}_n)$. A Kolmogorov–Smirnov test can also be used to assess departure from uniformity, although the p-value must be interpreted with caution given that $\hat{w}_1, \ldots, \hat{w}_n$ are dependent through $\hat{\Lambda}$.

The random time transformation can be extended to the case of a marked temporal point process, based on proposition 7.4.VI on p. 266 of Daley and Vere-Jones (2003). Indeed if N is an MPP with conditional intensity function $\lambda(t, m) = \lambda_g(t)f(m \mid t)$, then under the random time transformation with respect to the ground process, namely, $(t, m) \mapsto (\Lambda_g(t), m)$ with $\Lambda_g(t) = \int_0^t \lambda_g(s)ds$, the MPP N is transformed into an MPPP with unit ground rate. Goodness-of-fit can then be assessed by checking the extent to which the transformed time sequence $\hat{\tau}_1 = \hat{\Lambda}_g(t_1), \ldots, \hat{\tau}_n = \hat{\Lambda}_g(t_n)$ deviates from a unit rate PPP on $\left[0, \hat{\Lambda}_g(T)\right]$.

5.2 | Model diagnostics for the marks

The model for the mark process is based on an EGPD whose shape parameter is constant in time but whose scale parameter is a time-varying function depending on past self-excitement. This induces heterogeneity in the distribution of marks which must be factored out before one can check whether the model fits well with the observed marks. A natural way to proceed is to standardize the marks through the probability integral transform, namely,

$$U_t = G_\kappa\{H_\xi(X_t/\sigma_t)\} = G_\kappa\left\{1 - (1 + \xi X_t/\sigma_t)^{-1/\xi}\right\}.$$

Once the parameters are estimated, the standardized marks are approximately independent and uniform on the interval $[0, 1]$. To construct diagnostic plots, one can then choose any reference distribution F_Y. The transformation $\tilde{X}_t = F_Y^{-1}(U_t)$ leads to marks $\tilde{X}_t$ that are (approximately) distributed as F_Y. Thus, if $\tilde{x}_{(1)} < \cdots < \tilde{x}_{(n)}$ denote the ordered values of the transformed marks, a probability plot can be based on the n pairs defined, for each $i \in \{1, \ldots, n\}$, by

$$\left(\frac{i}{n+1}, F_Y(\tilde{x}_i)\right) = \left(\frac{i}{n+1}, F_Y\{F_Y^{-1}(u_{(i)})\}\right) = \left(\frac{i}{n+1}, u_{(i)}\right).$$

Clearly, this plot is invariant to the choice of reference measure F_Y, which is an advantage. By contrast, the quantile plot, which consists of the pairs $(F_Y^{-1}\{i/(n+1)\}, F_Y^{-1}(u_{(i)}))$ with $i \in \{1, \ldots, n\}$, does depend on the choice of F_Y^{-1}.

The unit exponential distribution is arguably a natural option for F_Y given that it is a special case of the EGPD as $\xi \to 0$ and $\kappa \to 1$. With this choice of reference distribution, the transformed variables $\tilde{X}_t = -\ln(1 - U_t)$ have a unit exponential distribution so that, for all $x \in (0, \infty)$, $\Pr(\tilde{X}_t \leq x) = 1 - e^{-x}$. The probability plot, which consists of the pairs $(i/(n+1), 1 - e^{-\tilde{x}_{(i)}})$ with $i \in \{1, \ldots, n\}$, is invariant to this choice but not the quantile plot, which is based on the pairs

$$\left(-\ln\left(1 - \frac{i}{n+1}\right), -\ln(1 - u_{(i)})\right) = \left(-\ln\left(1 - \frac{i}{n+1}\right), \tilde{x}_{(i)}\right)$$

with $i \in \{1, \ldots, n\}$. Both of these plots can be used for model validation. They basically convey the same information but on different scales. The quantile plot can be especially useful when interest focuses on goodness-of-fit in the upper tail.

6 | MODEL USE

The SEMPP model can be viewed as a unifying MPP for extreme-value analysis given the familiar models which it includes as special cases. Setting $\kappa = 1$ yields the self-exciting model with GPD marks, also called the self-exciting POT model by McNeil et al. (2015) in section 16.2 of their book. If $\kappa = 1$ and $\beta_1 = 0$, one finds the self-exciting model with independent GPD marks. If in addition $\nu(t) = 0$, there is then no self-exciting component and the ground process is reduced to a PPP with constant intensity, reducing the model to the standard MPPP with independent GPD marks.

In this section, various uses of the proposed SEMPP model are discussed. More specifically, a closed-form expression for the tail quantile is presented in Section 6.1 as an instantaneous risk measure of water level. Section 6.2 describes the modified thinning algorithm to simulate the fitted point process. It is also shown how one can use the simulated records to generate short-term forecasts and estimate the return period of major droughts.

6.1 | Tail estimation of the mark distribution

If the ground intensity indicates that there is a high probability of observing an event at time t, it is of great importance to make inference on its associated mark distribution to assess how extreme the mark can be. The estimation of large quantiles of the mark distribution, which measures the instantaneous tail risk conditionally on the past history, can be particularly useful in terms of hedging against extreme events. It gives information on the extreme water levels that are expected to be exceeded with small probability under current conditions but whose occurrence may have substantial influence on the planning and management of water supply.

Let X_{t+1} be the water level observed at time $t+1$, given a high threshold $u \in (0, \infty)$ and the information $\mathcal{H}_t$ observed up to time t. One then has, for all $x \in (u, \infty)$,

$$\overline{F}_{X_{t+1} \mid \mathcal{H}_t}(x) = \Pr(X_{t+1} > x \mid \mathcal{H}_t) = \Pr(X_{t+1} - u > x - u \mid X_{t+1} > u, \mathcal{H}_t) \Pr(X_{t+1} > u \mid \mathcal{H}_t).$$

In the discrete-time SEMPP model, the exceedance probability $\Pr(X_{t+1} > u \mid \mathcal{H}_t)$ is $p_{t+1} = \Pr(Y_{t+1} = 1 \mid \mathcal{H}_t) = \varphi\{\lambda_g(t+1)\}$ according to Equation (7) while the conditional excess distribution of the mark needed to compute $\Pr(X_{t+1} - u > x - u \mid X_{t+1} > u, \mathcal{H}_t)$ is an EGPD distribution with scaling parameter $\sigma_{t+1} = \beta_0 + \beta_1 \nu(t+1)$, as per Equation (8). Combining these two components, one finds that the predictive survival distribution is given, for all $x \in (0, \infty)$, by

$$1 - F_{X_{t+1} \mid \mathcal{H}_t}(x) = p_{t+1}\left[1 - G_\kappa\left[1 - \{1 + \xi(x-u)/\sigma_{t+1}\}^{-1/\xi}\right]\right].$$

So the qth quantile can be obtained by solving the equation $\Pr(X_{t+1} > x \mid \mathcal{H}_t) = 1 - F_{X_{t+1} \mid \mathcal{H}_t}(x) = 1 - q$. This yields

$$z_{t+1}^q = u + \sigma_{t+1}\xi^{-1}\left[\left[1 - \{1 - (1-q)/p_{t+1}\}^{1/\kappa}\right]^{-\xi} - 1\right], \tag{15}$$

if the Beta transition function is used, that is, $G_\kappa(v) = v^\kappa$ for all $v \in (0,1)$ and some $k \in (0, \infty)$. These formulas are valid under the condition that $p_{t+1} > 1 - q$. This can be understood in the sense that if p_{t+1} is negligible, that is, the probability of observing a mark exceeding the threshold u is very small, then it will not be meaningful to estimate the tail risk of the mark because with high probability it will be bounded above by the threshold.

In the continuous-time SEMPP model, given the information $\mathcal{H}_t$ observed up to time t, one can make inference on the extreme quantiles at the time point just after t, denoted as t^+. Formula (15) becomes

$$z^q_{t+} = u + \sigma_{t+}\xi^{-1}\left[\left[1 - \{1 - (1-q)/p_{t^+}\}^{1/\kappa}\right]^{-\xi} - 1\right],$$

where the exceedance probability at time t^+ can be approximated as

$$p_{t^+} = \Pr\{N_g(t, t+dt) > 0 \mid \mathcal{H}_t\} = 1 - \exp\left\{-\int_t^{t+dt} \lambda_g(s)ds\right\},$$

for some sufficiently small time increment dt, with the probability of more than one exceedances being negligible.

Equation (15) leads to a point estimate for the upper tail quantile of the negated series, which makes it possible to infer the extreme water levels for the next time period given the historical information. Furthermore, one can construct a confidence interval for z^q_{t+1}. Indeed, conditionally on $\mathcal{H}_t$, the self-exciting function at time $t+1$ is deterministic, so the point estimator $\hat{z}^q_{t+1}$ is a given function of the maximum likelihood estimators. One can then use the Delta method to obtain the SEs. By the Delta method, the variance of $\hat{z}^q_{t+1}$ can be estimated as $\text{var}(\hat{z}^q_{t+1}) \approx \nabla(\hat{z}^q_{t+1})^\top V \nabla(\hat{z}^q_{t+1})$, where the gradient vector is evaluated at the maximum likelihood estimates.

One could also base a Wald confidence interval for this quantile on an asymptotic normal approximation. However, Coles (2001) mentions that the likelihood surface of the extreme quantile can be highly asymmetric. Therefore, a symmetric Wald confidence interval from the Delta method could be misleading. A more reliable confidence interval can be derived from the profile likelihood function. Rearranging Equation (15), one can write

$$\beta_0 = \frac{\xi(z^{t+1}_q - u)}{\left[\left[1 - \{1 - (1-q)/p_{t+1}\}^{1/\kappa}\right]^{-\xi} - 1\right]} - \beta_1 \nu(t+1).$$

Moreover, the log-likelihood function with can be reparametrized with the quantile z^q_{t+1} becoming one of the parameters. Fixing $z^q_{t+1} = z$ for a range of values, the profile likelihood curve $\ell_p(z)$ is obtained by maximizing the likelihood with respect to the remaining parameters. The deviance statistic $D_p(z) = 2\{\ell_s - \ell_p(z)\}$, where ℓ_s refers to the log-likelihood value of the saturated model, has an asymptotic chi-square distribution with 1 degree of freedom. The $1-\alpha$ profile likelihood confidence interval is the set of values $\{z : D_p(z) \le c_\alpha\}$.

6.2 | Simulation-based inference

In general, simulation can be used as a tool for model assessment. By comparing the observed series with the simulated series, one can check whether the proposed model captures the key features of the data. In addition, the simulation-based method can be used to estimate the statistical quantities if the closed-form expressions are not available. It can be particularly useful for the continuous-time SEMPP as it has a more complex stochastic structure.

Given a fitted model for the drought event process in the form of a stationary MPP with explicitly specified conditional intensity and mark density function, a method of simulating event records from the model and using them for statistical inference is described in Section 6.2.1. It is also shown in Section 6.2.2 how simulations can be used to estimate the return period of major droughts. A method of generating probability forecasts of drought events over a prediction interval is then detailed in Section 6.2.3.

6.2.1 | Simulation algorithm

An SEMPP can be simulated by either the modified thinning algorithm or the cluster generation algorithm. The modified thinning algorithm of Ogata (1981) is based on the conditional intensity function. The general idea is to simulate candidate points from a homogeneous PPP with some upper-bound intensity and then accept the points based on a specific stochastic rule to effectively thin out the points. The cluster generation method relies on the cluster process representation of the SEMPP described in Section 3.3, which applies to both continuous- and discrete-time SEMPPs.

In continuous time, the immigrant process is generated from a homogeneous PPP with intensity ζ. Given an immigrant event (t_j, m_j), the times of its first-generation offspring are simulated by the standard thinning algorithm (Lewis & Shedler, 1979) and the associated marks are generated by the inverse transform sampling method. Offspring for subsequent generations are simulated recursively until the branching process becomes extinct, which occurs almost surely as the branching ratio is strictly smaller than 1. All the immigrant's descendants form a single cluster.

The approach is even more straightforward in discrete time. The immigrant process is generated from a homogeneous Bernoulli process with $\varphi(\zeta)$ as the probability of success. The offspring process is simulated as a dependent Bernoulli process where the probability of success is calculated based on Equation (7). To avoid the edge effect, one can assume that the intensity is zero before time zero in all simulations.

6.2.2 | Return period estimation

In the analysis of extremes, it is common to quantify the frequency of extreme events through a return period. Given a level r, the corresponding return period is defined as the expected number of observations until the level r is exceeded for the first time. In the analysis of extreme droughts, the return period to a single exceedance is not very informative as a severe drought is typically characterized with a prolonged period below a critical threshold, rather than a single extreme value. So a more reasonable indicator is the return period for a sequence of consecutive exceedances. As the process does not have long-range dependence, the distribution of the waiting time can approximate the distribution of the return period. It was seen in Section 4 that for a discrete-time SEMPP, the waiting time can be embedded into a finite state space Markov chain, which makes it possible to derive the exact distribution of the return period under this model.

For the continuous-time SEMPP, the sampling distribution can be obtained based on simulation results of the fitted SEMPP model. By simulating events on a sufficiently long time horizon, one can simulate the waiting time to the next targeted event. The return period of the targeted event can thus be estimated with an appropriate location measure of the sampling distribution. In most cases, the location measure is the sample mean but in the simulation of rare events, the simulated times are right-censored by the time horizons, so the sample median may be preferred. Moreover, the simulation-based technique can be applied to estimate other numerical quantities that can be attained from the list-history, for example, the expected number of events of a given magnitude over a fixed time interval.

6.2.3 | Drought event forecasting

Observations simulated from an SEMPP conditionally on the list of recently observed events can be used to deduce short-term forecasts. The approach is inspired by the work of Vere-Jones (1998) but cast in the context of drought analysis and extended through dynamic incorporation of information and prediction. In this setting, the aim of the forecasting exercise is to provide estimates for the probability that a drought of a given magnitude will occur within a predetermined prediction interval and, more importantly, to revise the probability estimates in response to new information as the process evolves in time. For drought analysis, magnitude can be a broader notion that refers to any key measure that can be computed based on simulated data, for example, the duration or minimal flow of a drought.

Suppose that information on past events is summarized in the list-history $\{(t_1, m_1), \ldots, (t_n, m_n)\}$. The simulation proceeds in a similar way as described in Section 6.2.1, but it begins with the simulation of the $(n+1)$st event with the time variable initialized at $t = t_n$. The initial intensity parameter needs to be computed conditionally on the observed history but the updating recurrence relation remains the same. Each simulation yields a possible realization path for the prediction interval. By repeating the process, one can estimate the proportion of paths containing an event within a given magnitude range. As new information about the events emerges, the simulations are generated, with revised prediction interval, conditionally on the augmented list history, leading to updated probability estimates. Given a sequence $(M_0, M_1), \ldots, (M_{K-1}, M_K)$ of magnitude ranges, one can eventually produce probability forecasts $p_{0,i}, \ldots, p_{K-1,i}$ summing up to 1 at time point t_i. Given the information up to that time, the model predicts that an event with magnitude between M_k and M_{k+1} occurs with probability $p_{k,i}$.

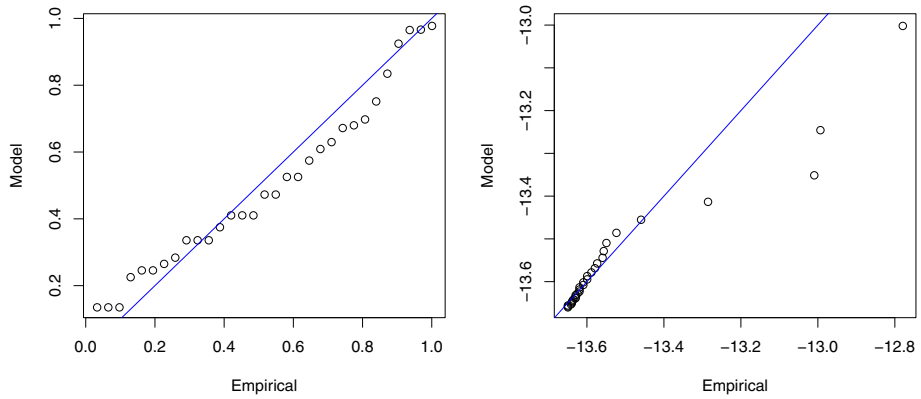

FIGURE 5 Probability plot (left panel) and quantile plot (right panel) for the peak over threshold model fitted to the declustered series of the exceedances above the 96% quantile of the negated daily water levels at the Rivière des Mille Îles between 1983 and 2013

7 | APPLICATION TO THE RIVIÈRE DES MILLE ÎLES DATA

To illustrate the use of the SEMPP model to the analysis of drought events, it was applied to the water level series of the Rivière des Mille Îles mentioned in the Introduction. The analysis is summarized below, along with the main findings.

7.1 | Data preprocessing and clustering of exceedances

To study dry spells, the right tail of the negated series is considered, which is equivalent to the left tail of the original series. Statistically speaking, a drought then corresponds to a cluster of exceedances over a high threshold, u. Once the latter is selected, the time series is viewed as a realization $(t_1, m_1), \ldots, (t_n, m_n)$ of a MPP on $[0, T] \times [0, \infty)$, where time t_i records the time where the ith exceedance occurs and mark m_i is the excess amount over the threshold. Figure 2 shows the original times series (left) and the transformed data (right) when the 96th centile $u = -13.66$ m of the negated series is chosen as the threshold, as dictated by the City of Montréal water engineers. The 406 exceedances are represented in the MPP plot, which provides a handy summary of the information available: t records the time at which droughts occur, and m measures their intensity. The information regarding clusters of events is also preserved.

It was already mentioned in the Introduction that a classical univariate GPD does not fit this series. As explained in Section 2, this model corresponds to an MPPP with homogeneous ground PPP for time and independent GPD marks. Under this scenario, the occurrence times of the exceedances should be randomly scattered on the scaled time axis and the interexceedance times should be approximately iid exponential random variables. However, this hypothesis is rejected on the basis of Cramér–von Mises test statistic, whose value is 87.028, leading to a tiny p-value.

There is also evidence that the exceedances are not randomly distributed over time. Indeed, the interexceedance times are spread out over a very wide range from 1 to 1285 days and more than 90% of them are less than 7 days long. Interexceedance times can be categorized as either intercluster or intracluster times. The large probability mass at very short interexceedance times implies that a large proportion of them are intracluster times, suggesting the presence of clusters of extremes. The extent of clustering can be measured with the extremal index, which is estimated at 0.08 by the interval method (Ferro & Segers, 2003). As the clustering degree increases, the extremal index approaches 0, so a value of 0.08 suggests a very strong clustering tendency of the exceedances.

The interval method gives rise to a natural way of declustering the original series, given that the 31 largest interexceedance times divide the exceedances into 32 clusters of different sizes. As the MPP plot and exponential quantile plot of the declustered series (not shown) do not provide any compelling evidence of residual clustering, the declustered series clearly allows for the application of the standard POT approach. However, the drawback of the declustering procedure is made clear with the diagnostic plots provided in Figure 5. While the dependence between exceedances is removed, all other large values within the same cluster are ignored and much information about the extreme values of the underlying series is wasted. The model is fitted with only 32 of the 406 exceedances, resulting in a large SE and an approximate 95% confidence interval for ξ ranging from 0.065 to 1.

TABLE 1 Maximum likelihood estimates of the parameters in the SEMPP model with time-varying impact function

Parameters	ζ	ψ	γ	δ	κ	ξ	β_0	β_1
Estimates	0.0065	2.2823	0.9948	11.8563	2.5855	0.5598	0.0036	0.0187
Standard error	0.0011	0.4305	0.1059	2.5275	0.3744	0.0635	0.0008	0.0047
Student *t* statistic	5.7988	5.3016	9.3942	4.6909	6.9046	8.8121	4.5962	3.9581

Abbreviation: SEMPP, self-exciting marked point process.

TABLE 2 Maximum likelihood estimates of the parameters in the SEMPP model with constant impact function

Parameters	ζ	ψ	γ	κ	ξ	β_0	β_1
Estimates	0.0064	2.8236	0.9461	2.3237	0.6771	0.0044	0.0216
Standard error	0.0011	0.5331	0.1043	0.3548	0.0722	0.0010	0.0061
Student *t* statistic	5.7503	5.2965	9.0664	6.5477	9.3773	4.1163	3.5828

Abbreviation: SEMPP, self-exciting marked point process.

7.2 | Model estimation

To specify the ground process, one must select an impact function g, a memory kernel ϕ in Equation (6), and a transfer function φ in Equation (7). Following the discussion in Section 3.1, an exponential decaying function is chosen, as well as a bounded impact function. The ground intensity function is then formulated as

$$\lambda_g(t) = \zeta + \psi v(t) = \zeta + \psi \sum_{i:t_i<t} \{1 + \tanh(\delta m_i)\} \exp\{-\gamma(t - t_i)\}, \tag{16}$$

where $\psi \in [0, \infty)$, $\gamma \in (0, \infty)$, and $\delta \in (0, \infty)$. The parameter ψ is introduced here so that it reduces to the constant impact model if $\delta = 0$. Given that the function λ_g is nonnegative, the transfer function φ is taken to be the standard exponential rather than the logistic distribution because it has a nonnegative support. Equation (7) then becomes $p_t = 1 - e^{-\lambda_g(t)}$.

Moreover, as evidenced by Figure 2 (left panel), the raw series exhibits some seasonality, which leads to extreme values (right panel) occurring mainly in the summer months, that is, from May to November. Among the 406 exceedances of the negated daily water level series, only 10 occurred in the winter months (from December to April), and most of them are isolated rather than clustered. To account for this seasonal variation, the ground intensity for the summer months was taken to be of the form given in Equation (16) but assumed constant in the winter months by constraining the self-exciting component to be zero for that period. The winter exceedances are then taken to be iid and to occur as per a Poisson process. Thus, unless otherwise stated, the rest of this section concerns the summer months.

To specify the mark structure, a transition function G must be specified which should be determined based on the left-tail behavior of the data. It can be checked visually that when m is close to zero, $\ln\{\hat{F}(m)\}$ is nearly linear in $\ln(m)$. This is an indication of a regularly varying left tail, which justifies a Beta transition function. Taking $g_\kappa(v) = \kappa v^{\kappa-1}$ for all $v \in (0, 1)$ and some choice of parameter $\kappa \in (0, \infty)$ to be determined, one finds that the conditional distribution of the marks specified in Equation (8) is given, for all $m \in (0, \infty)$, by

$$f_M(m \mid t, \mathcal{H}_t) = \kappa \sigma_t^{-1} \left\{1 - (1 + \xi m/\sigma_t)^{-1/\xi}\right\}^{\kappa-1} (1 + \xi m/\sigma_t)^{-1/\xi-1}.$$

Both the discrete- and continuous-time SEMPP model were fitted. In the continuous-time model, the estimated probability of observing more than one exceedance in a single day in a large cluster of exceedances is at least 20%, as reported in the Appendix. Therefore, the discrete-time model was deemed more suitable.

Two discrete-time SEMPP models were fitted, one with a time-varying impact function (no constraint on δ in Equation 16) and the other with a constant impact function ($\delta = 0$). Parameter estimates for these two SEMPP models and their associated SEs are reported in Tables 1 and 2, respectively.

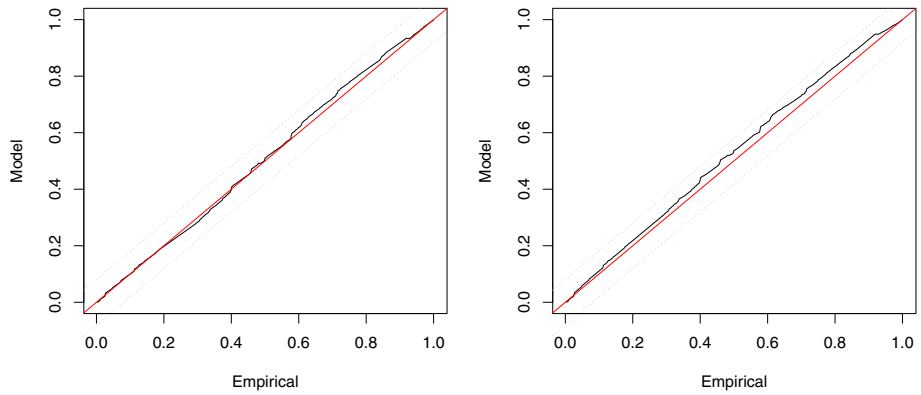

FIGURE 6 Probability plots of the transformed temporal residuals for the self-exciting marked point process model of the exceedances above the 96% quantile of the negated daily water levels at the Rivière des Mille Îles between 1983 and 2013: Constant impact function (left panel); time-varying impact function (right panel)

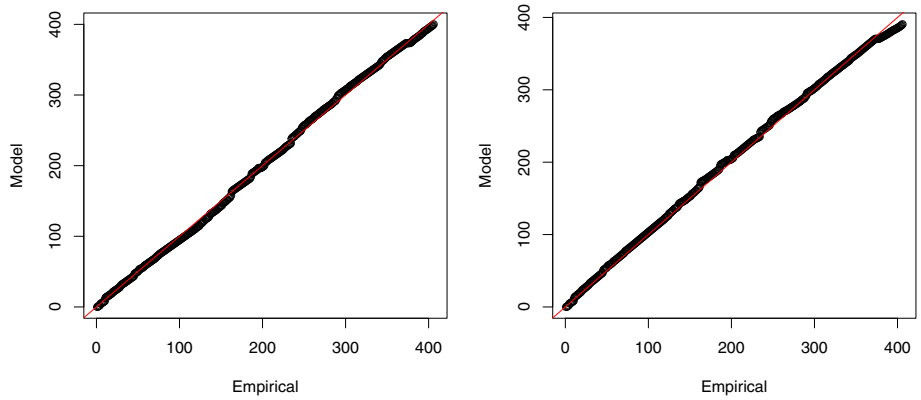

FIGURE 7 Observed number of exceedances versus the expected number of exceedances under the self-exciting marked point process model fitted to the exceedances above the 96% quantile of the negated daily water levels at the Rivière des Mille Îles between 1983 and 2013: Constant impact function (left panel); time-varying impact function (right panel)

Figure 6 shows the diagnostic plots of the rescaled residuals for the two models, together with 95% pointwise confidence intervals computed based on the Kolmogorov–Smirnov test of uniformity. Figure 7 compares the expected number of exceedances under the two models with the observed cumulative number of exceedances at each time point. With these diagnostic plots, one can assess visually whether the estimated ground process fits well. For both models, the empirical curve is within the 95% pointwise Kolmogorov–Smirnov bounds. For the model with $\delta = 0$ (left panel of Figure 6), the residuals lie very close to the diagonal reference line, suggesting a good fit in the time dimension. Slightly larger deviations are observed for the model with unconstrained δ (right panel of Figure 6).

Figure 8 presents diagnostic plots in the mark space for the two models, namely, the probability plots for the standardized EGP marks. It is noteworthy that the estimated κ in Tables 1 and 2 are significantly different from 1, which confirms the inadequacy of the GP distribution as a model for the exceedances over the chosen threshold. For better comparison, the self-exciting POT model, which is a special case of model (14), was fitted. The corresponding diagnostic plots appear in Figure 9. The probability plot of the self-exciting POT model exhibits a striking departure from the reference line, especially in the bottom left corner. The use of the EGPD family in place of the GPD clearly improves the fit, notably by injecting skewness in the left tail to adapt to the subasymptotic threshold.

As evidenced by the model diagnostics, the SEMPP model with time-varying impact function provides little improvement. Therefore, it seems preferable to assume a constant impact in specifying the SEMPP, which implies that, with respect to this particular event process, the triggering effect is dominated by the temporal proximity (distance) rather than by the mark size. Ultimately, it transpires that a discrete-time SEMPP with exponential decay kernel and constant impact function is the most appropriate choice for the data at hand. The subsequent analysis is based on it.

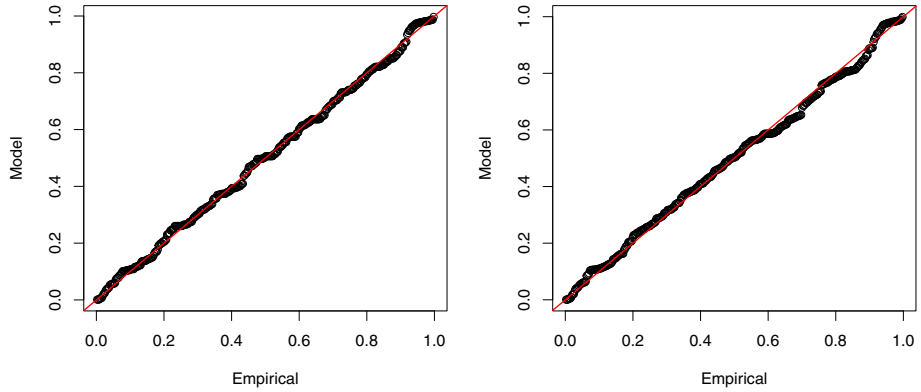

FIGURE 8 Probability plots of the transformed marks for the self-exciting marked point process model of the exceedances above the 96% quantile of the negated daily water levels at the Rivière des Mille Îles between 1983 and 2013: Constant impact function (left panel); time-varying impact function (right panel)

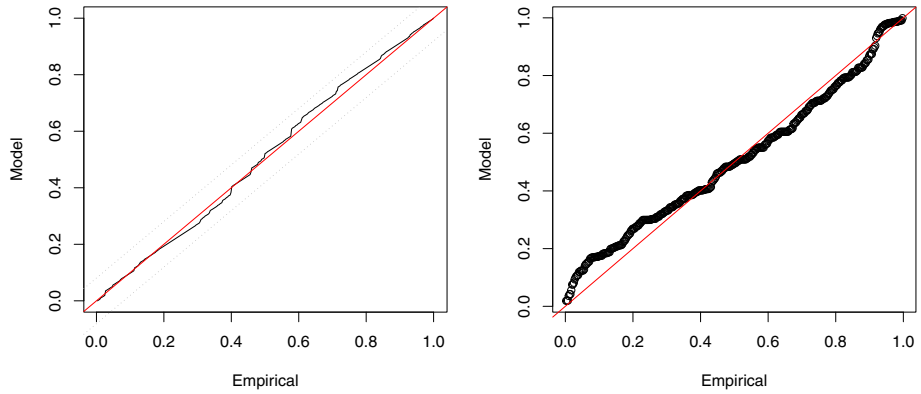

FIGURE 9 Probability plot of the transformed temporal residuals (left panel), probability plot of the transformed marks (right panel) for the self-exciting peak over threshold model with constant impact function fitted to the exceedances above the 96% quantile of the negated daily water levels at the Rivière des Mille Îles between 1983 and 2013

7.3 | Predictions regarding future dry spells

In this section, the SEMPP with exponential decay kernel and constant impact function is used to gain insight into the occurrence of dry spells in the Montréal area. The techniques described in Sections 4 and 6 are drawn upon to answer question such as how long droughts are expected to persist, how severe they can be, and how often they will recur.

7.3.1 | Drought severity

The severity of a drought can be measured with the number of consecutive days that the water level is below a critical threshold. Mathematically, it is a run statistic of the exceedance process of the negated water level series. Based on the formulas in Section 4.1, it is possible to compute the distribution of the run length, which is the number of consecutive exceedances not separated by any nonexceedance. The left panel of Figure 10 is the probability mass function of the run length, based on Equation (9). The expected run length under the fitted model is estimated to be 4.74 through Equation (10). As computed in Section 1, the mean run length observed in the data is 5.48, which is consistent with the estimate. Intuitively, it means that, according to the model, given that the water level is below the critical threshold on a given day, it is expected to be followed immediately by 3.74 below-threshold days.

As shown in Section 1, a major drought typically consists of several below-threshold dry spells separated by short above-threshold episodes. Thus, a useful measure of the drought persistence is the length of a cluster that can be separated

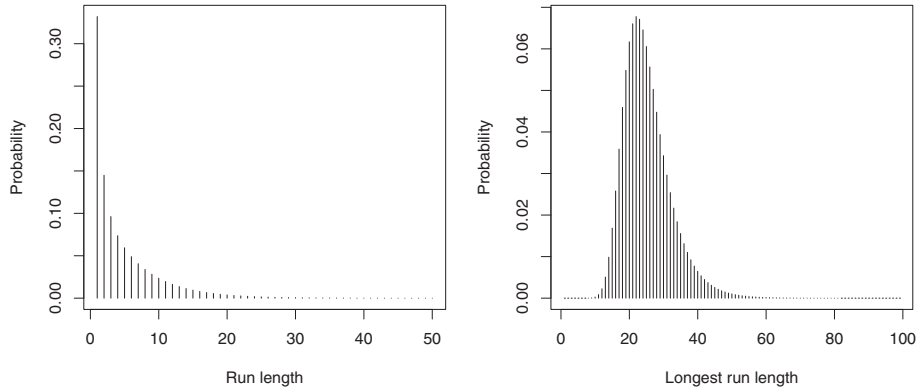

FIGURE 10 Probability mass function of the run length (left panel) and the longest run length (right panel)

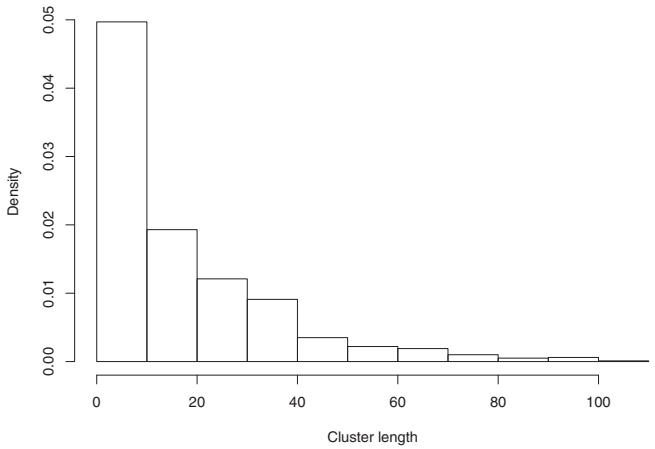

FIGURE 11 Histogram of the simulated length of a cluster of consecutive exceedances separated by no more than 14 consecutive nonexceedance days under the discrete-time self-exciting marked point process model with constant impact

by no more than r nonexceedances. The run length above corresponds to the case $r = 1$. As r increases, the computation of the distribution of the cluster length becomes combinatorially more complex. Alternatively, one can use simulations to assess the sampling properties of the cluster length. The cluster-based simulation method described in Section 6.2.1 is used to generate the clusters and compute the cluster length. Figure 11 shows the histogram of the simulated cluster length of sample size 1000. It is noteworthy that the point estimate given by the model, which is the sample mean of the simulated statistics, is 17.05. It is very close to the empirical estimate of 17.12 computed in Section 1, which suggests that the fitted model performs well in capturing the clustering pattern of the exceedances.

Using the finite Markov-chain approach described in Section 4.2, one can also estimate the distribution of the longest run of exceedances for, say, a 30-year time span. The right panel of Figure 10 displays an estimate of this distribution, based on Equation (13); it is unimodal and right-skewed. The expected longest run length is estimated to be 25.21, which means that within 30 years, the longest consecutive days of exceedances that can be observed is on average 25.21 days. But as the distribution is right-skewed, upper quantiles would be more informative. Under the fitted model, the 99% and 99.9% quantiles of the longest run length distribution were estimated to be 48 and 60 days, respectively.

In addition to the run statistics, the low quantiles of the water level distribution are a natural indicator of the risk of drought. These are equivalent to upper quantiles of the negated series for which the SEMPP model was developed. More specifically, if w_t^{1-q} denotes the quantile of order $1 - q$ of the water level distribution at time t, then $z_t^q = -w_t^{1-q}$ is the quantile of order q of the distribution of X_{t+1} conditionally on past history $\mathcal{H}_t$ at time t.

Consider the last day of the observation window where an exceedance occurs, which is the end of the last cluster displayed in Figure 2. Suppose that it is desired to exploit the recent past information to evaluate the likelihood of observing very extreme observations for the next day. A good indicator of the instantaneous tail risk would be the value of

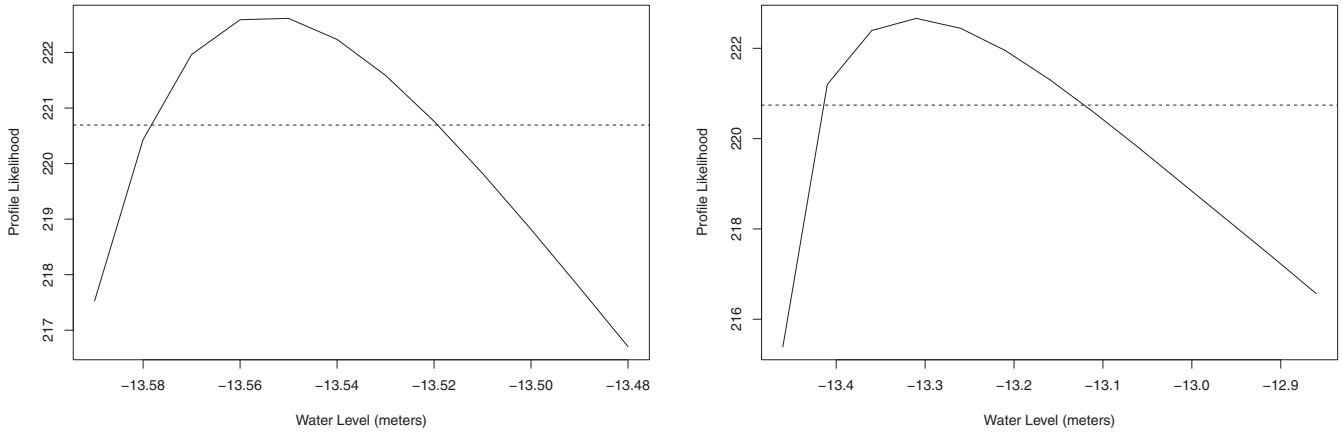

FIGURE 12 Profile likelihood curve of the 95% upper tail quantile $z_{t+1}^{0.95}$ (left panel) and the 99% upper tail quantile $z_{t+1}^{0.99}$ (right panel)

z_{t+1}^{q} for some large q, which can be estimated using Equation (15). Point estimates for the 95% and 99% upper quantiles are $\hat{z}_{t+1}^{0.95} = -13.55$ and $\hat{z}_{t+1}^{0.99} = -13.31$ m, respectively; the corresponding 95% profile likelihood confidence intervals are $[-13.57, -13.52]$ and $[-13.41, -13.16]$ m. In both cases, the confidence interval were constructed using the profile likelihood method described in Section 6.1 and the profile likelihood curves are presented in Figure 12.

The merit of the extreme quantile estimator stemming from the SEMPP model is to absorb the real-time information from p_{t+1} and σ_{t+1}. The quantile estimates obtained from the univariate GPD model and the POT model with declustering are constant in time, as the conventional models assume uniformity in the time dimension and iid marks. By contrast, the SEMPP model takes into account their interaction and can predict the marks with the cluster information.

7.3.2 | Drought frequency

As described in Section 3.3, the SEMPP model has an equivalent cluster process representation. The behavior of the immigrant process can shed light on the cluster arrival rate and hence the drought frequency. The estimated background intensity in the summer is 0.0064. Thus, the cluster arrivals follow a Bernoulli process with occurrence probability $\varphi(\hat{\zeta}) = 1 - e^{-0.0064} \approx 0.0064$. For a 30-year time span, the water level process is expected to have $\varphi(\hat{\zeta})T = 40.19$ clusters of extreme exceedances in the summer, where $T = 210 \times 30$ days. In the winter, the estimated background intensity is 0.001 with no clustering effect, so only 4.49 isolated exceedances are expected.

Another useful measure for drought risk management is the return period of a severe drought. The 1988–1989 North American Drought is one of the worst droughts on record. It was triggered by heat waves in most parts of North America, as can be seen by the particularly large cluster of peak exceedances around $n = 2000$ in Figure 2. The drought was characterized by extreme lack of precipitation for a prolonged time, resulting in losses to agriculture, forestry, natural ecosystems, and water supplies reaching 60 billion dollars in the USA and 1.8 billion in Canada.

As shown in Figure 1, the 1988 drought featured four dependent dry spells totaling 82 days; the longest one lasted over 1 month. To assess the return period of a drought of this magnitude, the distribution of the return period for a run T_{30^*} of at least 30 consecutive exceedances was examined. An estimate of the distribution of T_{30^*} can be derived from Equation (11). As the SEMPP model allows for seasonal variation, however, the transition matrix P_t is not homogeneous in time but alternating between the summer and winter. The probability that T_{30^*} is less than 30 years was found to be 0.28. The return period, as the expectation of the waiting time, was estimated to be 96.64 years, based on Equation (12).

7.3.3 | Drought event forecasting

Continuing with the analysis of the 1988–1989 North American Drought, imagine that at the initial stage of the drought, say in May 1988, one might be interested in predicting how long this drought might persist. Duration is measured by cluster length, where a cluster is defined as a succession of threshold exceedances, none of which is separated by more

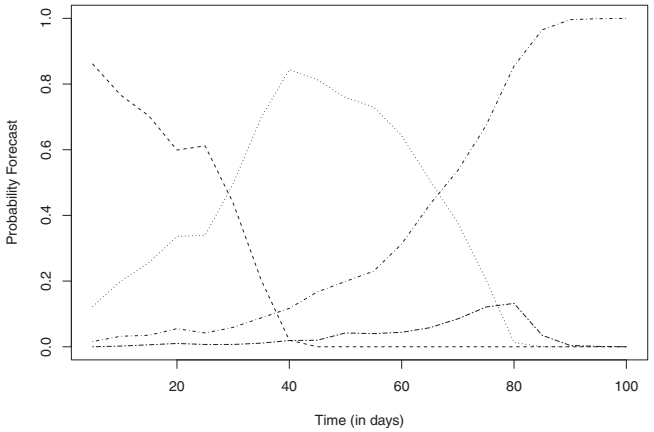

FIGURE 13 Evolution over time of the probability forecast for the duration (in days) of the drought from May 16 to August 25, 1988; dashed line: (0, 40]; dotted line: (40, 80]; dot-dashed line: (80, 120]; two-dash line: (120, ∞)

TABLE 3 Probability forecast

Duration (in Days)	(0, 40]	(40, 80]	(80, 120]	(120, ∞)
Day 5 (May 20, 1988)	0.861	0.123	0.016	0.00
Day 25 (June 9, 1988)	0.612	0.339	0.042	0.007
Day 50 (July 4, 1988)	0.000	0.759	0.199	0.042
Day 75 (July 29, 1988)	0.000	0.205	0.673	0.121
Day 100 (August 25, 1988)	0.000	0.000	1.000	0.000

than 14 nonexceedances. Starting with the first threshold exceedance, probability forecasts for the drought were generated from the SEMPP for the following duration spans: (0, 40], (40, 80], (80, 120], (120, ∞). The forecasts were updated every 5 days by incorporating new events into the history.

Table 3 gives probability estimates at different phases in the development of the drought and Figure 13 shows how these estimates evolve in time. Initially, there are 86.1% chances that the drought will last no more than 40 days. This probability decreases steadily at first but increases again on Day 25 after three consecutive days of nonexceedance; the model interprets this as a sign that the cluster may terminate soon. After Day 30, the probability drops quickly to zero. Similarly, the probability that the drought lasts between 40 and 80 days starts at 12.3%, gradually climbs and peaks in the middle phase of the drought, and ultimately vanishes by day 81. The other two curves can be interpreted in a similar way.

8 | DISCUSSION

This article shows how SEMPPs can be used for the analysis of clustered extremes in hydrology. A major benefit of this approach over the POT method is that the temporal dependence of extremes is modeled explicitly, thereby making the inference on cluster statistics possible without artificially declustering the data. A key component of the model is its reliance on the EGPD family, which allows for flexible joint modeling of extreme and nonextreme data. Overall, the proposed modeling strategy led to an effective modeling of the Rivière des Mille Îles data which motivated this study, leading to useful inference on the duration, return period, and extreme flows of dry spells in the Montréal area.

This work could be extended in various ways. Four options are briefly mentioned here. First, data-based strategies could be designed to help in the selection and validation of a suitable impact function in the self-exciting component of the model. Second, greater flexibility in the EGPD could be obtained by resorting to a nonparametric transition function, as recently explored by Tencaliec et al. (2020). Third, the model could be expanded to account for spatial interaction between data series at various sites as described by Reinhart (2018). Finally, it would be valuable to develop a Bayesian framework for the analysis of clustered hydrological extremes based on SEMPP methodology. The work of Rasmussen (2013), which

considers Bayesian inference for Hawkes processes, provides a starting point. By sampling from the posterior distribution of both the parameters and clustering structure, it would be possible to account for model uncertainty in making predictions and assessing uncertainty via the posterior predictive distribution.

ORCID

Christian Genest https://orcid.org/0000-0002-1764-0202

REFERENCES

Chavez-Demoulin, V., Davison, A. C., & McNeil, A. J. (2005). Estimating value-at-risk: A point process approach. *Quantitative Finance*, *5*, 227–234.

Chavez-Demoulin, V., & McGill, J. (2012). High-frequency financial data modeling using Hawkes processes. *Journal of Banking and Finance*, *36*, 3415–3426.

Coles, S. G. (2001). *An introduction to statistical modeling of extreme values*. Springer.

Daley, D. J., & Vere-Jones, D. (2003). *An introduction to the theory of point processes: Volume I: Elementary theory and methods*. Springer-Verlag.

Daley, D. J., & Vere-Jones, D. (2007). *An introduction to the theory of point processes: Volume. II: General theory and structure*. Springer Science & Business Media.

Davison, A. C., & Smith, R. L. (1990). Models for exceedances over high thresholds. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, *52*, 393–425.

Eastoe, E. F., & Tawn, J. A. (2012). Modelling the distribution of the cluster maxima of exceedances of subasymptotic thresholds. *Biometrika*, *99*, 43–55.

Embrechts, P., Klüppelberg, C., & Mikosch, T. (1997). *Modelling extremal events for insurance and finance*. Springer.

Embrechts, P., Liniger, T., & Lin, L. (2011). Multivariate Hawkes processes: An application to financial data. *Journal of Applied Probability*, *48*, 367–378.

Fawcett, L., & Walshaw, D. (2007). Improved estimation for temporally clustered extremes. *Environmetrics*, *18*, 173–188.

Ferro, C. A., & Segers, J. (2003). Inference for clusters of extreme values. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*, *65*, 545–556.

Fisher, R. A., & Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Mathematical Proceedings of the Cambridge Philosophical Society*, *24*, 180–190.

Fu, J. C., & Koutras, M. V. (1994). Distribution theory of runs: A Markov chain approach. *Journal of the American Statistical Association*, *89*, 1050–1058.

Hawkes, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika*, *58*, 83–90.

Hawkes, A. G., & Oakes, D. (1974). A cluster process representation of a self-exciting process. *Journal of Applied Probability*, *11*, 493–503.

Leadbetter, M. R., Lindgren, G., & Rootzén, H. (1983). *Extremes and related properties of random sequences and processes*. Springer Science & Business Media.

Lewis, P. (1970). Remarks on the theory, computation and application of the spectral analysis of series of events. *Journal of Sound and Vibration*, *12*, 353–375.

Lewis, P. W., & Shedler, G. S. (1979). Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistics*, *26*, 403–413.

McNeil, A. J., Frey, R., & Embrechts, P. (2015). *Quantitative risk management: Concepts, techniques and tools* (revised ed.). Princeton University Press.

Naveau, P., Huser, R., Ribereau, P., & Hannart, A. (2016). Modeling jointly low, moderate, and heavy rainfall intensities without a threshold selection. *Water Resources Research*, *52*, 2753–2769.

O'Brien, G. L. (1987). Extreme values for stationary and Markov sequences. *Annals of Probability*, *15*, 281–291.

Ogata, Y. (1978). The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics*, *30*, 243–261.

Ogata, Y. (1981). On Lewis' simulation method for point processes. *IEEE Trans Information Theory*, *27*, 23–31.

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

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

Papastathopoulos, I., & Tawn, J. A. (2013). Extended generalised Pareto models for tail estimation. *Journal of Statistical Planning and Inference*, *143*, 131–143.

Pickands, J., III. (1975). Statistical inference using extreme order statistics. *Annals of Statistics*, *3*, 119–131.

Rasmussen, J. G. (2013). Bayesian inference for Hawkes processes. *Methodology and Computing in Applied Probability*, *15*, 623–642.

Reinhart, A. (2018). A review of self-exciting spatio-temporal point processes and their applications. *Statistical Science*, *33*, 299–318.

Resnick, S. I. (2007). *Extreme values, regular variation and point processes*. Springer.

Sigman, K., & Whitt, W. (2019). Marked point processes in discrete time. *Queueing Systems*, *92*, 47–81.

Smith, R. L. (1989). Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone. *Statistical Science*, *4*, 367–377.

Stindl, T., & Chen, F. (2019). Modeling extreme negative returns using marked renewal Hawkes processes. *Extremes*, *22*, 705–728.

Tencaliec, P., Favre, A.-C., Naveau, P., Prieur, C., & Nicolet, G. (2020). Flexible semiparametric generalized Pareto modeling of the entire range of rainfall amount. *Environmetrics*, *31*(2), e2582.

Vere-Jones, D. (1998). Probabilities and information gain for earthquake forecasting. *Computational Seismology*, *30*, 248–263. https://doi.org/10.1029/CS005p0104

How to cite this article: Li, X., Genest, C., & Jalbert, J. (2021). A self-exciting marked point process model for drought analysis. *Environmetrics*, e2697. https://doi.org/10.1002/env.2697

APPENDIX A

Let t_n be the time of the last observed exceedance, and T_{n+1} be the random time mark of the next exceedance in the continuous-time SEMPP. The probability of observing more than one exceedance in a given day is the same as the probability that the interexceedance time is less than 1 day. That is,

$$\Pr\left(T_{n+1} \leq t_n + 1 \mid \mathcal{H}_{t_n}\right) = 1 - \exp\left\{-\int_{t_n}^{t_n+1} \lambda_g(t)dt\right\},$$

where $\lambda_g(t) = \zeta + \psi\nu(t) = \zeta + \psi \sum_{i:t_i<t} e^{-\gamma(t-t_i)}$. This probability can be computed easily, given that the integral has an explicit primitive but it depends on the past history $\mathcal{H}_{t_n}$. For example, say there are no exceedances prior to t_n. Given that the estimated parameters of the ground process of the continuous SEMPP model fitted to the summer exceedances are $(\hat{\zeta}, \hat{\psi}, \hat{\gamma}) = (0.0051, 0.2660, 0.2862)$, a plug-in estimate of the probability of observing more than one exceedance in a given day is computed as 0.2106. So based on the fitted continuous SEMPP model, the probability of observing more than one exceedances in a given day is at least 0.2106, which increases with the number of exceedances prior to t_n.