**Noname manuscript No.**
(will be inserted by the editor)

# Peaks Over Threshold for Bursty Time Series

**Katharina Hees · Smarak Nayak · Peter Straka* ·**

Received: date / Accepted: date

**Abstract** In many complex systems studied in statistical physics, inter-arrival times between events such as solar flares, trades and neuron voltages follow a heavy-tailed distribution. The set of event times is fractal-like, being dense in some time windows and empty in others, a phenomenon which has been dubbed "bursty".

This article generalizes the Peaks Over Threshold (POT) model to the setting where inter-event times are heavy-tailed. For high thresholds and infinite-mean waiting times, we show that the times between threshold crossings are Mittag-Leffler distributed, and thus form a "fractional Poisson Process" which generalizes the standard Poisson Process. We provide graphical means of estimating model parameters and assessing model fit. Along the way, we apply our inference method to a real-world bursty time series, and show how the memory of the Mittag-Leffler distribution affects the predictive distribution for the time until the next extreme event.

Peter Straka was supported by the Discovery Early Career Research Award DE160101147 on the Project "Predicting Extremes when Events Occur in Bursts" by the Australian Research Council. Katharina Hees was supported by the DAAD co-financed by the German Federal Ministry of Education and Research (BMBF).

Katharina Hees
University of Dortmund
E-mail: `hees@statistik.uni-dortmund.de`

Smarak Nayak
National Australia Bank
E-mail: `smarak.nayak@nab.com.au`

Peter Straka*
UNSW Sydney
E-mail: `p.straka@unsw.edu.au`

**Keywords** heavy tails · renewal process · extreme value theory · peaks over threshold ·

## 1 Introduction

Time series displaying temporally inhomogeneous behaviour have received strong interest in the recent statistical physics literature (Barabsi 2005; J. Oliveira and Barabsi 2005; Vasquez et al. 2006; Vazquez et al. 2007; Omi and Shinomoto 2011; Min, Goh, and Vazquez 2011; Karsai et al. 2011; Bagrow and Brockmann 2013). They have been observed in the context of earthquakes, sunspots, neuronal activity and human communication (see Karsai et al. 2012; Vajna, Tth, and Kertsz 2013; Mark M Meerschaert and Stoev 2008 for a list of references). Such time series exhibit high activity in some 'bursty' intervals, which alternate with other, quiet intervals. Although several mechanisms are plausible explanations for bursty behaviour (most prominently self-exciting point process by Hawkes (1971)), there seems to be one salient feature which very typically indicates the departure from temporal homogeneity: a heavy-tailed distribution of waiting times (Vasquez et al. 2006; Karsai et al. 2012; Vajna, Tth, and Kertsz 2013). As we show below in simulations, a simple renewal process with heavy-tailed waiting times can capture this type of dynamics. For many systems, the renewal property is appropriate; a simple test of the absence of correlations in a succession of waiting times can be undertaken by randomly reshuffling the waiting times (Karsai et al. 2012).

Often a magnitude, or mark can be assigned to each event in the renewal process, such as for earthquakes, solar flares or neuron voltages. The Peaks-Over-Threshold model (POT, see e.g. Coles 2001) applies a threshold to the magnitudes, and fits a Generalized Pareto distribution to the threshold exceedances. A commonly made assumption in POT models is that times between events are either fixed or light-tailed, and this entails that the threshold crossing times form a Poisson process (Hsing, Hsler, and Leadbetter 1988). Then as one increases the threshold and thus decreases the threshold crossing probability $p$, the Poisson process is rarefied, i.e. its intensity decreases *linearly* with $p$ (see e.g. Beirlant et al. 2006).

As will be shown below, in the heavy-tailed waiting time scenario threshold crossing times form a *fractional Poisson process* (Laskin 2003; Mark M Meerschaert, Nane, and Vellaisamy 2011), which is a renewal process with Mittag-Leffler distributed waiting times. The family of Mittag-Leffler distributions nests the exponential distribution (Haubold, Mathai, and Saxena 2011), and hence the fractional Poisson process generalizes the standard Poisson process. Again as the threshold size increases and the threshold crossing probability $p$ decreases, the fractional Poisson process is rarefied: The scale parameter of the Mittag-Leffler inter-arrival times of threshold crossing times increases, but *superlinearly*; see the Theorem below.

Maxima of events which occur according to a renewal process with heavy-tailed waiting times have been studied under the names "Continuous Time

Random Maxima process" (CTRM) (Benson, Schumer, and Meerschaert 2007; Mark M Meerschaert and Stoev 2008; Hees and Scheffler 2016; Hees and Scheffler 2017), "Max-Renewal process" (Silvestrov 2002; Silvestrov and Teugels 2004; Basrak and poljari 2015), and "Shock process" (Esary and Marshall 1973; Shanthikumar and Sumita 1983; Shanthikumar and Sumita 1984; Shanthikumar and Sumita 1985; Anderson 1987; Gut and Hsler 1999). The existing literature focuses on probabilistic results surrounding these models. In this work, however, we introduce a method of inference for this type of model, which is seemingly not available in the literature.

We review the marked renewal process in Section 2, and derive a scaling limit theorem for inter-exceedance times in Section 3. We give a statistical procedure to estimate model parameters via stability plots in Section 5, but to set the stage we first discuss inference for the Mittag-Leffler distribution in Section 4. (A simulation study of the effectiveness of our statistical procedure is given in the appendix.) Diagnostic plots for model criticism are discussed in Section 6. In Section 7, we discuss the memory property of the Mittag-Leffler distribution, and how it affects the predictive distribution for the time until the next threshold crossing event. Section 8 concludes. For all statistical computations we have used R (R Core Team 2018). All code and data used for the analysis in this article has been organized into an R package `CTRE` (`https://github.com/UNSW-MATH/CTRE`). The source code for the figures generated in this manuscript is available online at `https://github.com/UNSW-MATH/bursty-POT`.

## 2 Continuous Time Random Exceedances (CTRE)

As a model for extreme observations, we use a Marked Renewal Process (MRP):

**Definition (MRP):** Let $(W, J), (W_1, J_1), (W_2, J_2), \dots$ be i.i.d. pairs of random variables, where the $W_k > 0$ are interpreted as the *waiting times* and $J_k \in [x_L, x_R]$ as the *event magnitudes* ($x_L \in [-\infty, +\infty), x_R \in (-\infty, +\infty]$). If $W$ and $J$ are independent, the Marked Renewal Process is said to be *uncoupled*. □

Note that the $k$-th magnitude $J_k$ occurs at time $T_k = W_1 + \ldots + W_k$. Based on an MRP, we define the Continuous Time Random Exceedance model (CTRE) as follows:

**Definition (CTRE):** Given a threshold $\ell \in (x_L, x_R)$, consider the stopping time

$$\tau(\ell) := \min\{k : J_k > \ell\}, \quad \ell \in (x_L, x_R).$$

Define the pair of random variables $(X(\ell), T(\ell))$ via

$$X(\ell) = J_{\tau(\ell)} - \ell, \quad T(\ell) = \sum_{k=1}^{\tau(\ell)} W_k.$$

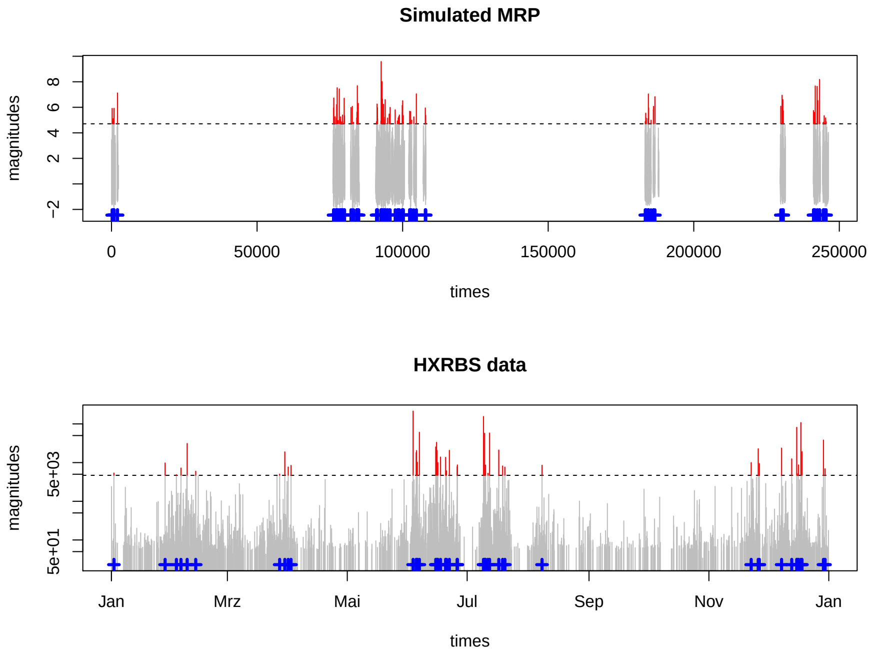


**Fig. 1** Exceedances (red) and Times until Exceedance (durations between blue crosses) for a given threshold $\ell$ (dashed line).

By restarting the MRP at $\tau(\ell)$, inductively define the two i.i.d. sequences $T(\ell, n)$ and $X(\ell, n)$, $n \in \mathbb{N}$, called the "interarrival times" and the "exceedances", respectively. The pair sequence $(T(\ell, n), W(\ell, n))_{n\in\mathbb{N}}$ is called a Continuous Time Random Exceedance model (CTRE). If the underlying MRP is uncoupled, then the CTRE is also called uncoupled. $\square$

In this article, we restrict ourselves to the uncoupled case, where the two sequences $X(\ell, n)_{n\in\mathbb{N}}$ and $T(\ell, n)_{n\in\mathbb{N}}$ are independent.[1] Figure 1 shows a simulated dataset in the top panel, where $W$ has a stable distribution with tail parameter $\beta = 0.8$ (and skewness 1 and location 0), and where $J$ is from a standard Gumbel distribution. In the bottom panel, we plot a time series of solar flare intensities derived from a NASA dataset (Dennis et al. 1991)[2]. Clearly, the simulated data exhibit long intervals *without any* events, whereas in the real-world dataset events appear continuously. The threshold exceedances, however,

[1] To see why, note that $X(\ell)$ is, in distribution, simply equal to $J|J > \ell$, independent of any waiting time $W_k$.

[2] The "complete Hard X Ray Burst Spectrometer event list" is a comprehensive reference for all measurements of the Hard X Ray Burst Spectrometer on NASA's Solar Maximum Mission from the time of launch on Feb 14, 1980 to the end of the mission in Dec 1989. 12,776 events were detected, with the "vast majority being solar flares". The list includes the start time, peak time, duration, and peak rate of each event. We have used "start time" as the variable for event times, and "peak rate" as the variable for event magnitudes.

appear to have similar statistical behaviour in both models. Observations below a threshold are commonly discarded in Extreme Value Theory; likewise, the CTRE model interprets these observations as noise and filters them out.

## 3 Scaling limit of Exceedance Times

In this section we state and prove the key theorem, see below. For an accessible introduction to regular variation and stable limit theorems, we recommend the book by Mark M Meerschaert and Sikorskii (2012).

**Theorem:** Let the waiting times $J_k$ be in the domain of attraction of a positively skewed sum-stable law with stability parameter $0 < \beta < 1$; more precisely,

$$\frac{W_1 + \ldots + W_n}{b(n)} \xrightarrow{d} D, \quad n \to \infty \tag{1}$$

for a function $b(n)$ which is regularly varying at $\infty$ with parameter $1/\beta$, and where $\mathbf{E}[\exp(-sD)] = \exp(-s^\beta)$. Write $p := \mathbf{P}(J > \ell)$. Then the weak convergence

$$\frac{T(\ell)}{b(1/p)} \to W_\beta \quad \text{as} \quad \ell \uparrow x_R$$

holds, where the Mittag-Leffler random variable $W_\beta$ is defined on the positive real numbers via

$$\mathbf{E}[\exp(-sW_\beta)] = \frac{1}{1+s^\beta}.$$

□

For a scale parameter $\sigma > 0$, we write $\mathrm{ML}(\beta, \sigma)$ for the distribution of $\sigma W_\beta$. The Mittag-Leffler distribution with parameter $\beta \in (0, 1]$ is a heavy-tailed positive distribution for $\beta < 1$, with infinite mean. However, as $\beta \uparrow 1$, $\mathrm{ML}(\beta, \sigma)$ converges weakly to the exponential distribution $\mathrm{Exp}(\sigma)$. This means that although its moments are all infinite, the Mittag-Leffler distribution may (if $\beta$ is close to 1) be indistinguishable from the exponential distribution, for the purposes of applied statistics. For a detailed reference on the Mittag-Leffler distribution, see e.g. Haubold, Mathai, and Saxena (2011), and for algorithms, see e.g. the R package `MittagLeffleR` (Gill and Straka 2017).

Proof of Theorem: We interpret the threshold crossing time $T(\ell)$ as the hitting time of the underlying CTRM (Continuous Time Random Maxima) or “max-renewal process”, and then utilize a result by Mark M Meerschaert and Stoev (2008). The running maximum process is defined as

$$M(c) := J_1 \vee \ldots \vee J_{\lfloor c \rfloor},$$

and since we assume that the $J_k$ have a continuous distribution, there exist norming functions $a(c)$ and $d(c)$ such that

$$\mathbf{P}\left[\frac{M(c)-d(c)}{a(c)} \le \ell^*\right] \longrightarrow F(\ell^*), \quad t \to \infty$$

where $F$ is a generalized extreme value distribution, and $\ell^*$ is any value from the support of $F$. The CTRM process is then defined via

$$V(t) = M(N(t)), \quad t \ge 0$$

where $N(t)$ is the renewal process associated with the waiting times $W_k$:

$$N(t) = \max\{n : W_1 + \ldots + W_n \le t\}.$$

Now a key observation is that

$$T(\ell) = \inf\{t : V(t) > \ell\},$$

and that

$$T(\ell) > t \quad \text{if and only if} \quad V(t) \le \ell.$$

By (Theorem 3.1, Mark M Meerschaert and Stoev 2008), we have the stochastic process convergence

$$\frac{V(ct) - d(\tilde{b}(c))}{a(\tilde{b}(c))} \xrightarrow{d} Y(t), \quad t > 0.$$

where $Y(t)$ is a time-changed ("subordinated") extremal process, and where $\tilde{b}(c)$ is a regularly varying norming function which is *inverse* to $b(c)$, in the sense that $b(\tilde{b}(c)) \sim c \sim \tilde{b}(b(c))$.

Without loss of generality, we choose $\ell^*$ such that $F(\ell^*) = 1/e$, and let $\ell = a(\tilde{b}(c))\ell^* + d(\tilde{b}(c))$. We may then calculate

$$\mathbf{P}\left[\frac{T(\ell)}{b(1/p)} > t\right] = \mathbf{P}[T(\ell) > b(1/p)t] = \mathbf{P}[V(ct) \le \ell]$$

where we have substituted $c = b(1/p)$. Moreover

$$\mathbf{P}[V(ct) \le \ell] = \mathbf{P}\left[\frac{V(ct) - d(\tilde{b}(c))}{a(\tilde{b}(c))} \le \frac{\ell - d(\tilde{b}(c))}{a(\tilde{b}(c))}\right] \longrightarrow \mathbf{P}\left[Y(t) \le \ell^*\right]$$

Defining the hitting time of level $\ell^*$ by $Y(t)$ as $\xi_{\ell^*} := \inf\{t : Y(t) > \ell^*\}$, we then have

$$P\left[Y(t) \le \ell^*\right] = \mathbf{P}[\xi_{\ell^*} > t] = \mathbf{P}[(-\log F(\ell^*))^{-1/\beta} X^{1/\beta} D > t]$$

by (Proposition 4.2, Mark M Meerschaert and Stoev 2008), where $X$ is an exponential random variable with mean 1. Using (Theorem 19.1, Haubold, Mathai, and Saxena 2011), we see that $X^{1/\beta}D \sim \mathrm{ML}(\beta, 1)$, concluding the proof. □

**Remark:** If $\beta = 1$, the result of the Theorem above is standard, see e.g. Equation (2.2) in Gut and Hsler (1999).

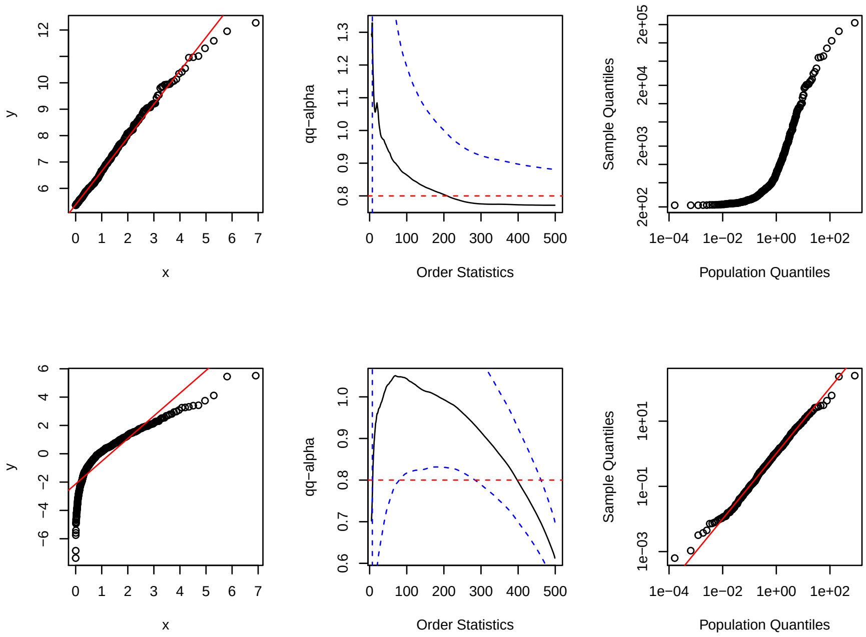


**Fig. 2** Pareto vs. Mittag-Leffler distribution, both with same tail exponent 0.8. Top row: solar flare magnitudes data, fitting nicely with a Pareto distribution. Bottom row: simulated Mittag-Leffler data. Left column: static QQ-Estimator plots, assessing a fit against the Pareto distribution. Centre column: dynamic QQ-Estimator plots. Right column: a usual QQ-Plot based on Mittag-Leffler quantiles.

## 4 Inference for the Mittag-Leffler distribution

Since the Mittag-Leffler distribution is heavy-tailed, many researchers would intuitively give the highest importance to the tail behaviour of the distribution, and estimate the exponent of the tail function with established methods such as the Hill estimator. The QQ-estimator (Kratz and Resnick 1996) is closely related to the Hill estimator and fits a least squares line through the logarithms of the ordered statistics (y-axis) and the corresponding quantiles of the exponential distribution (x-axis). The reciprocal slope is returned as the estimate of the tail exponent. For instance, the top-left panel in Figure 2 shows the (static) QQ-estimator for the magnitudes of the solar flare data, returning a good fit to a Pareto distribution with tail parameter 0.79. The dynamic QQ-estimator plot (top-center panel) plots the tail exponent estimate for the largest values at cutoffs varying up to the 5th order statistic, and the region of stability at 0.8 indicates a recommended estimate.

However, as described nicely by Resnick (1997), the less similar a heavy-tailed distribution is to a Pareto distribution, the less useful a Hill or QQ-Plot estimator becomes. The bottom-left panel of Figure 2, for instance, shows the static QQ-estimator plot for 2000 draws from the Mittag-Leffler distribution with tail parameter 0.8. The stretched exponential shape means that the Mittag-Leffler distribution has more probability near 0 than the Pareto distribution, severely biasing the estimator downwards (0.62). Even by looking at different cutoffs via the dynamic QQ-estimator plot (bottom-center panel) one hardly identifies 0.8 as a clear candidate for a tail parameter estimate.

QQ-estimators, and the closely related Hill estimator, are hence not suitable to detect a heavy-tailed Mittag-Leffler distribution. Moreover, QQ-estimator plots of exponentially distributed data are virtually indistinguishable from the bottom-left panel. This shows that *Mittag-Leffler distributed data may look like exponentially distributed data if examined via a QQ-estimator.*

Hence if there is some prior expectation that the data are drawn from the Mittag-Leffler distribution (as is the the case for threshold exceedance times), then we recommend avoiding QQ-estimators and Hill plots altogether, and instead examining QQ-plots directly, on a logarithmic scale (see Figure 2, right column). The scale parameter $\sigma$ is irrelevant for a QQ-Plot. The tail parameter $\beta$ can be estimated quickly via the log-moment method by Cahoy (2013), or via maximum likelihood. Both estimators are implemented in `MittagLeffleR` (Gill and Straka 2017). Since the exponential distribution is nested in the Mittag-Leffler family of distributions, a standard likelihood ratio test can be performed, with the exponential distribution as a null model against a Mittag-Leffler distribution as the alternative. As an example, the threshold crossing times for the solar flare dataset (Figure 5, right panel) yield a difference in deviance of $\approx 324$, which evaluates to a strongly significant $\chi_1^2$ test statistic (p-value $= 10^{-72}$) for the null hypothesis $\beta = 1$.

## 5 Inference on Exceedance times

The Theorem in Section 3 implies that for a high threshold $\ell$ we may approximate the distribution of $T(\ell)$ with an $\mathrm{ML}(\beta, b(1/p))$ distribution, where $b(\,)$ varies regularly at $\infty$ with parameter $1/\beta$. Building on the POT (Peaks Over Threshold) method, we propose the following estimation procedure for the distribution of $T(\ell)$:

1. For a range of thresholds $\ell$ near the largest order statistics, extract datasets of exceedance times $\{T(\ell, i)\}_i$.
2. For each choice of threshold $\ell$, fit a Mittag-Leffler distribution to the resulting dataset $\{T(\ell, i)\}_i$. This results in the estimates $\{\hat\beta(\ell)\}_\ell$ and $\{\hat\sigma(\ell)\}_\ell$.
3. Plot $\ell$ vs. $\hat\beta(\ell)$. As $\ell$ increases towards $x_R$, $\hat\beta(\ell)$ *stabilizes* around a constant $\hat\beta$. Use $\hat\beta$ as an estimate for the tail parameter $\beta$ of the Mittag-Leffler distribution of exceedance times.
4. Approximate $p \approx |\{k : J_k > \ell\}|/n$. Recall that $b(c)$ is regularly varying with parameter $1/\beta$, and hence has the representation $b(c) = L(c)c^{1/\beta}$ for some slowly varying function $L(c)$. Assuming that the variation of $L(c)$ is

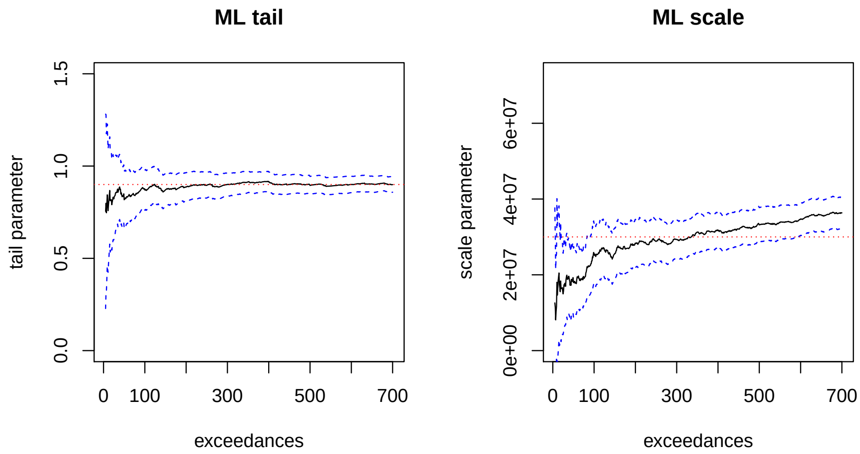


**Fig. 3** Stability plots for the tail and scale parameter of the Mittag-Leffler distribution of the Solar Flare dataset. Dotted horizontal lines are at $\beta = 0.85$ and $\sigma_0 = 3 \times 10^7$ seconds $\approx 0.95$ years.

negligible, we hence plot $\ell$ vs. $p^{1/\hat{\beta}}\hat{\sigma}(\ell)$. Again as $\ell$ increases towards $x_R$, $p^{1/\hat{\beta}}\hat{\sigma}(\ell)$ is expected to stabilize around a constant $\hat{\sigma}_0$. We then use $p^{-1/\hat{\beta}}\hat{\sigma}_0$ as an estimate of the scale parameter of the Mittag-Leffler distribution of exceedance times for the level $\ell$.

The above approach, though theoretically sound, benefits from the following practical adjustments (compare with Figure 3):

- We choose $\ell$ from the order statistics, i.e. $\ell$ is the $k$-th largest of the observations $X_j$, where $k$ runs from $k_{\min}, k_{\min}+1, \ldots, k_{\max}$. The datasets are then of length $k-1$.
- We use $k$ rather than $\ell$ for the horizontal axis of our plots.
- In Step 4, rather than plotting $p^{1/\hat{\beta}}\hat{\sigma}(\ell)$ we plot $k^{1/\hat{\beta}}\hat{\sigma}(\ell)$. This changes $\hat{\sigma}_0$ by the multiplicative constant $n^{1/\hat{\beta}}$, but has the advantage that $\hat{\sigma}_0$ does not change if one pre-processes the data by removing all observations below a certain threshold.

The estimates $\hat{\beta}$ and $\hat{\sigma}_0$ give an estimate of the distribution of exceedance times, dependent on the threshold $\ell$:

$$T(\ell) \sim \mathrm{ML}(\hat{\beta}, k^{-1/\hat{\beta}}\hat{\sigma}_0).$$

For quick estimates of the Mittag-Leffler parameters we have used the method of log-transformed moments by Cahoy (2013). We have verified the validity of our estimation algorithm via simulations, see the appendix.

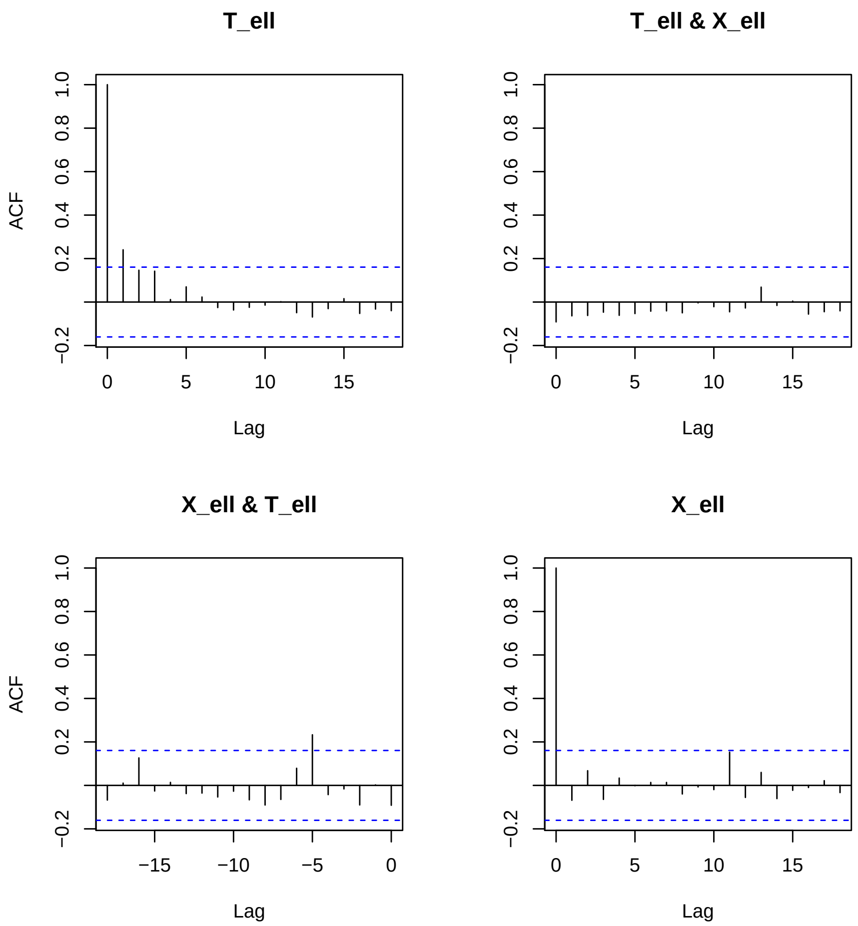


**Fig. 4** Diagnostic plots for the solar flare data: auto-correlation function.

## 6 Checking Model Assumptions

The CTRE model is based on three main assumptions, which are repeated below. For each assumption, we suggest one means of checking if it holds:

i.i.d.: After removing the "noise observations" below the smallest threshold $\ell_0$, the pair sequence $(T(\ell_0, i), X(\ell_0, i))$ is i.i.d. An indication if this is true is given by an auto-correlation plot for the logarithms (to ensure finite moments) of the two time series.

Uncoupled: Each $T(\ell, i)$ is independent of each $X(\ell, i)$. We propose an empirical copula plot to check for any dependence.

ML$(\beta, \sigma)$ distribution of $T(\ell, i)$: Apply a cutoff at the lowest threshold $\ell_0$, extract the threshold crossing times, and create a QQ Plot for the Mittag-

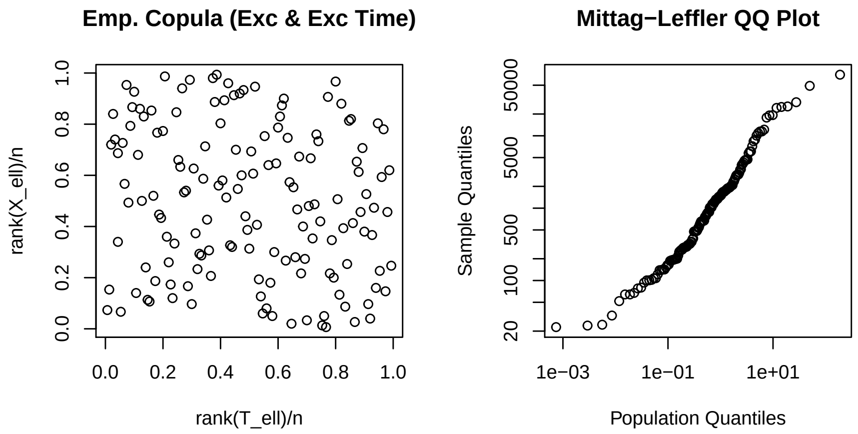


**Fig. 5** Diagnostic plots for the solar flare data: empirical copula and QQ Plot.

Leffler distribution. Use a log-Moment estimate of the tail parameter for the theoretical / population quantiles of the plot.

Figures 4 and 5 show the diagnostic plots for a minimum threshold chosen at the 200th order statistic. There is some residual autocorrelation for the sequence of threshold exceedance times that is not accounted for by the CTRE model. The fit with a Mittag-Leffler distribution ($\beta = 0.8$) is good, though there are signs that the power-law tail tapers off for very large inter-threshold crossing times. There is no apparent dependence between threshold exceedance times and event magnitudes seen in the copula plot.

## 7 Predicting the time of the next threshold crossing

According to Figure 3, for a threshold $\ell$ at the $k$-th order statistic, the fitted threshold exceedance time distribution is

$$T_\ell \sim \mathrm{ML}(\beta, k^{1/\beta}\sigma_0),$$

where $\beta = 0.85$ and $\sigma_0 = 3.0 \times 10^7$sec. Unlike the exponential distribution, the Mittag-Leffler distribution is not memoryless, and the probability density of the time $t$ until the next threshold crossing will depend on the time $t_0$ elapsed since the last threshold crossing. This density equals

$$p(t|\beta, \sigma_0, \ell, t_0) = \frac{f(t + t_0|\beta, k^{1/\beta}\sigma_0)}{\mathbf{P}[T_\ell > t_0]}$$

where $f(\,\cdot\,|\beta, k^{1/\beta}\sigma_0)$ is the probability density of $\mathrm{ML}(\beta, k^{1/\beta}\sigma_0)$. The more time has passed without a threshold crossing, the more the probability distribution shifts towards larger values for the next crossing (see Figure 6, left panel). The hazard rate

$$h(t) = \frac{f(t|\beta, k^{1/\beta}\sigma_0))}{\int_t^\infty f(\tau|\beta, k^{1/\beta}\sigma_0))\,d\tau}$$

represents the risk of a threshold crossing per unit time, and is a decreasing function for the Mittag-Leffler distribution. The closer $\beta$ is to 1, the more the hazard rate mimics that of an exponential distribution (a constant function, see Figure 6, right panel).

It is beyond the scope of the current paper to incorporate parameter uncertainty into our predictive distribution for the next threshold crossing; however, methods as described by Scarrott and Macdonald (2012) and Lee, Fan, and Sisson (2015) are likely to extend to our setting.

## 8 Discussion & Conclusion

We have extended the POT (Peaks over Threshold) model, a mainstay of extreme value theory, to "bursty" time series, which have been studied intensively in statistical physics. Burstiness is characterized by power-law waiting times between events, and we have shown that the Mittag-Leffler distribution arises naturally as a scaling limit for the inter-exceedance times of high thresholds. Moreover, we have derived the following non-linear scaling behaviour: $\sigma \sim p^{-1/\beta}$, where $\sigma$ is the scale parameter of the distribution of threshold exceedance times, $p$ is the fraction of magnitudes above the threshold, and $\beta$ the exponent of the power law.

The "anomalous" scaling behaviour in the bursty setting entails two phenomena: i) a heavy tail of the interarrival time distribution of threshold crossings (long rests), and ii) a high propensity for more threshold crossing events immediately after each threshold crossing event (bursts). The Mittag-Leffler

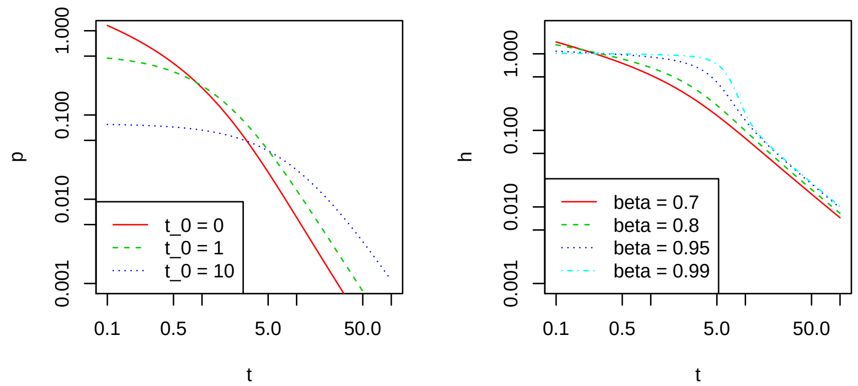


**Fig. 6** Left: Conditional distribution of time until next threshold crossing, depending on elapsed time $t_0$ since last crossing ($\beta = 0.8$, $\sigma_0 = 1$). Right: Hazard rate depending on tail parameter $\beta$.

distribution captures both phenomena, due to its heavy tail as well as its stretched exponential (peaked) asymptotics for small times. It generalizes the exponential distribution, and in the solar flare data example, this generalization is warranted, because the likelihood-ratio test is strongly significant.

When we introduced the CTRE model, we assumed that all events are i.i.d. This assumption is likely sufficient but not necessary for our limit theorem to hold. Moreover, any data below a (minimum) threshold $\ell_0$ is discarded in our inference procedure, and hence need not satisfy the i.i.d. assumption. For the purposes of statistical inference, we merely require that the inter-threshold-crossing times are i.i.d. However, it might be interesting to generalize our theory to the case of a stationary sequence of magnitudes. In case of fixed times between observations, the distribution of the sequence of (correctly normalized) inter-exceedance times converges under certain (mixing) conditions to a mixture of an exponential distribution and a point mass in zero (Ferro and Segers 2003). Hence, this also leads to a model for extreme events that occur in cluster. Allowing for both, dependence between the magnitudes and heavy tailed waiting times, could maybe further improve modelling of extreme events like earthquakes, that occur in clusters.

For some bursty time series, the current CTRE model may be too rigid, because often the heavy-tailed character of the inter-arrival time distribution does not hold at all time scales; rather, it often applies at short and intermediate time scales, and is truncated or tempered (reverting to an exponential distribution) at very long time scales, see e.g. Mark M. Meerschaert, Roy, and Shao (2012) and Aban, Meerschaert, and Panorska (2006). In such situations, a "tempered" Mittag-Leffler distribution may provide a more realistic fit, which we aim to introduce in follow-up work.

Acknowledgements

The authors would like to thank Prof. Peter Scheffler for insights on stochastic process limits for CTRMs, and Gurtek Gill who helped create the MittagLeffleR R-package.

## A Validating our inference method on simulated data

To test our inference method via stability plots, we have simulated 10000 independent waiting time and magnitude pairs $(W_k, J_k)$ (see upper panel in Figure 1). In order to have exact analytical values available for $\beta$ and $\sigma_0$, a distribution for $W_k$ needs to be chosen for which $b(n)$ from (1) is known. If we choose $W_k \stackrel{d}{=} D$, where $D$ is as in (1), then due to the stability property we have the *equality* of distribution $W_1 + \ldots + W_n \stackrel{d}{=} b(n)D$, for $b(n) = n^{1/\beta}$. Using the parametrisation of Samorodnitsky and Taqqu (1994), a few lines of calculation (see e.g. the vignette on parametrisation in Gill and Straka (2017)) show that $D$ must have the stable distribution $S_\beta(\cos(\pi\beta/2)^{1/\beta}, +1, 0)$, which is implemented in the R package `stabledist` by Wuertz, Maechler, and members. (2016).

By the Theorem, the distribution of $T(\ell)$ is approximately

$$\mathrm{ML}(\beta, p^{-1/\beta}) = \mathrm{ML}(\beta, k^{-1/\beta} n^{1/\beta}),$$

which means $\sigma_0 = n^{1/\beta}$. The distribution of $J_k$ is irrelevant for the inference on $\beta$ and $\sigma_0$ (we have chosen unit exponential random variables). Figure 7 displays plots of $\hat\beta(\ell)$ and $\hat\sigma(\ell)$ vs. $k$; recall that $k$ is the index of the order statistics of $J_k$ at which the threshold $\ell$ is placed. Dotted lines show 95% confidence intervals, which are derived from the asymptotic normality of the log-moments estimators (Cahoy 2013) and the $\delta$-method (Gill and Straka 2017). The dashed lines show the actual values of $\beta$ resp. $\sigma_0$, showing that our inference method identifies the parameters correctly.

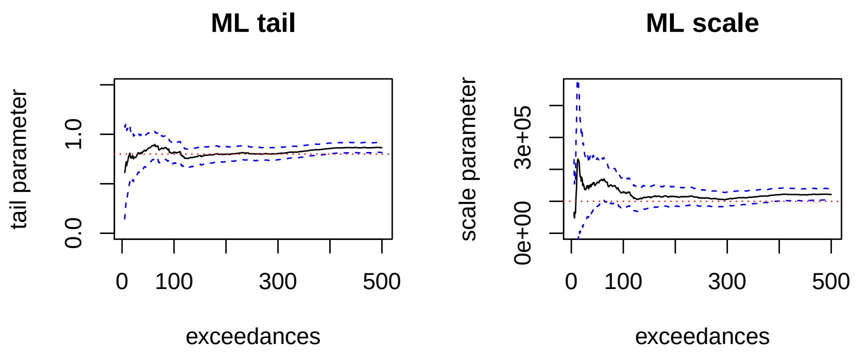


**Fig. 7** Tail and scale estimates for simulated data, with waiting times drawn from the stable distribution $S_\beta(\cos(\pi\beta/2)^{1/\beta}, +1, 0)$ with $\beta = 0.8$. Dashed lines are 95% confidence intervals, dotted lines are the known theoretical values (0.8 and $10000^{1/0.8}$).

## References


Aban, Inmaculada B, Mark M Meerschaert, and Anna K Panorska. 2006. "Parameter Estimation for the Truncated Pareto Distribution." *J. Am. Stat. Assoc.* 101 (473): 270–77. doi:10.1198/016214505000000411.

Anderson, Kevin K. 1987. "Limit Theorems for General Shock Models with Infinite Mean Intershock Times." *J. Appl. Probab.* 24 (2): 449–56. `http://www.jstor.org/stable/3214268`.

Bagrow, James P., and Dirk Brockmann. 2013. "Natural emergence of clusters and bursts in network evolution." *Phys. Rev. X* 3 (2): 1–6. doi:10.1103/PhysRevX.3.021016.

Barabsi, Albert Lszl. 2005. "The origin of bursts and heavy tails in human dynamics." *Nature* 435 (May): 207–11. doi:10.1038/nature03459.

Basrak, Bojan, and Drago poljari. 2015. "Extremes of random variables observed in renewal times." *Stat. Probab. Lett.* 97. Elsevier B.V.: 216–21. doi:10.1016/j.spl.2014.11.025.

Beirlant, Jan, Yuri Goegebeur, Johan Segers, and Jozef Teugels. 2006. *Statistics of extremes: theory and applications*. John Wiley & Sons.

Benson, David A, Rina Schumer, and Mark M Meerschaert. 2007. "Recurrence of extreme events with power-law interarrival times." *Geophys. Res. Lett.* 34 (l16404): DOI:10.1029/2007GL030767. doi:10.1029/2007GL030767.

Cahoy, Dexter O. 2013. "Estimation of Mittag-Leffler Parameters." *Commun. Stat. - Simul. Comput.* 42 (2): 303–15. doi:10.1080/03610918.2011.640094.

Coles, S. 2001. *An Introduction to Statistical Modelling of Extreme Values*. London: Springer-Verlag.

Dennis, Brian R, L E Orwig, G S Kennard, G J Labow, R A Schwartz, A R Shaver, and A K Tolbert. 1991. "The complete Hard X Ray Burst Spectrometer event list, 1980-1989." `https://umbra.nascom.nasa.gov/smm/hxrbs.html`.

Esary, J. D., and A. W. Marshall. 1973. "Shock Models and Wear Processes." doi:10.1214/aop/1176996891.

Ferro, Christopher AT, and Johan Segers. 2003. "Inference for Clusters of Extreme Values." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 65 (2). Wiley Online Library: 545–56.

Gill, Gurtek, and Peter Straka. 2017. *MittagLeffleR: Using the Mittag-Leffler Distributions in R*. `https://strakaps.github.io/MittagLeffleR/`.

Gut, Allan, and Jrg Hsler. 1999. "Extreme Shock Models." *Extremes*, no. 1983: 295–307. `http://link.springer.com/article/10.1023/A:1009959004020`.

Haubold, H.J., A. M. Mathai, and R. K. Saxena. 2011. "Mittag-Leffler Functions and Their Applications." *J. Appl. Math.* 2011: 1–51. doi:10.1155/2011/298628.

Hawkes, Alan G. 1971. "Point spectra of some mutually exciting point processes." *J. R. Stat. Soc. Ser. B Stat. Methodol.* JSTOR, 438–43.

Hees, Katharina, and Hans-Peter Scheffler. 2016. "On joint sum/max stability and sum/max domains of attraction," June, 1–31. `http://arxiv.org/abs/1606.03109`.

———. 2017. "Coupled Continuous Time Random Maxima." *Extremes*, no. June (September). Extremes: 1–24. doi:10.1007/s10687-017-0304-6.

Hsing, T, J Hsler, and M R Leadbetter. 1988. "On the Exceedance Point Process for a Stationary Sequence" 78: 97–112.

Karsai, Mrton, Kimmo Kaski, Albert Lszl Barabsi, and Jnos Kertsz. 2012. "Universal features of correlated bursty behaviour." *Sci. Rep.* 2. doi:10.1038/srep00397.

Karsai, Mrton, M. Kivel, R. K. Pan, K. Kaski, J. Kertsz, Albert Lszl Barabsi, and J. Saramki. 2011. "Small but slow world: How network topology and burstiness slow down spreading." *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.* 83: 1–4. doi:10.1103/PhysRevE.83.025102.

Kratz, Marie, and Sidney I. Resnick. 1996. "The qq-estimator and heavy tails." *Commun. Stat. Part C Stoch. Model.* 12 (4): 699–724. doi:10.1080/15326349608807407.

Laskin, Nick. 2003. "Fractional Poisson process." *Commun. Nonlinear Sci. Numer. Simul.* 8 (3-4): 201–13. doi:10.1016/S1007-5704(03)00037-6.

Lee, J., Y. Fan, and S. A. Sisson. 2015. "Bayesian threshold selection for extremal models using measures of surprise." *Comput. Stat. Data Anal.* 85. Elsevier B.V.: 84–99. doi:10.1016/j.csda.2014.12.004.

Meerschaert, Mark M, and Alla Sikorskii. 2012. *Stochastic Models for Fractional Cal-*