



HAL
open science

An overview of the extremal index

Nicholas Moloney, Davide Faranda, Yuzuru Sato

► **To cite this version:**

Nicholas Moloney, Davide Faranda, Yuzuru Sato. An overview of the extremal index. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 2019, 29 (2), pp.022101. 10.1063/1.5079656 . hal-02334271

HAL Id: hal-02334271

<https://hal.science/hal-02334271>

Submitted on 28 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

An overview of the extremal index

Nicholas R. Moloney,^{1, a)} Davide Faranda,^{2, b)} and Yuzuru Sato^{3, c)}

¹⁾*Department of Mathematics and Statistics, University of Reading, Reading RG6 6AX, UK^{d)}*

²⁾*Laboratoire de Sciences du Climat et de l'Environnement, UMR 8212 CEA-CNRS-UVSQ, IPSL, Université Paris-Saclay, 91191 Gif-sur-Yvette, France*

³⁾*RIES/Department of Mathematics, Hokkaido University, Kita 20 Nishi 10, Kita-ku, Sapporo 001-0020, Japan*

(Dated: 5 January 2019)

For a wide class of stationary time series, extreme value theory provides limiting distributions for rare events. The theory describes not only the size of extremes, but also how often they occur. In practice, it is often observed that extremes cluster in time. Such short-range clustering is also accommodated by extreme value theory via the so-called extremal index. This review provides an introduction to the extremal index by working through a number of its intuitive interpretations. Thus, depending on the context, the extremal index may represent (i) the loss of iid degrees of freedom, (ii) the multiplicity of a compound Poisson point process, (iii) the inverse mean duration of extreme clusters. More recently, the extremal index has also been used to quantify (iv) recurrences around unstable fixed points in dynamical systems.

Whether extreme events occur in isolation or in clusters is an important question for their prediction and mitigation. Extreme value theory can accommodate clustering via the extremal index which, heuristically, measures the size of the cluster. Mathematically, clustering is most conveniently understood within the framework of point processes. Without clustering, extreme events occur in the manner of a Poisson process. With clustering, extreme events are bunched together in a compound Poisson process. In order to develop an intuition for the extremal index, we survey some simple examples from stochastic processes, real-world time series and dynamical systems.

I. INTRODUCTION

Human societies are perpetually exposed to natural hazards, and an understanding of the statistics of extreme events is vital to predicting and mitigating their effects^{51,53}. Analogous to the central limit theorem for the sums of random variables, universal laws likewise exist for their maxima or minima, and extreme value theory now finds application in a broad range of areas including hydrology⁴¹, earth science⁴⁸, finance⁴⁶, meteorology and climate science^{1,13}. For recent reviews of modeling frameworks for extremes, see^{31,36}.

While extreme events are most often felt in terms of their impact, for the purposes of this review we will limit

ourselves to the technically more tractable definition of an extreme as something rare or large in magnitude (defined more precisely in Sec. II). We have in mind a stationary or easily detrendable process generating weakly correlated random variables, which are available in reasonably large numbers. In the 1920s, Fisher and Tippett³⁰ identified the possible limiting distributions for the rescaled maxima $M_n = \{X_1, \dots, X_n\}$ of a set of n independently and identically distributed (iid) random variables X_i . For parametric modeling, these generalised extreme value (GEV) cumulative distributions are most conveniently written in the form

$$G_{\xi;\mu,\sigma}(x) = \exp\left(-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right), \quad (1)$$

where μ and $\sigma > 0$ are location and scale parameters, respectively. The shape parameter ξ is determined by the tail behaviour of the parent distribution $F(x) = \mathbb{P}(X \leq x)$ ³⁸. These GEV distributions are closely related to generalised Pareto distributions (GPD), which describe the distribution of exceedances over high thresholds. Taken together, the GEV and GPD form the basis of extreme value theory. Details may be found in numerous textbooks^{8,16,19,20,42}. Importantly, extreme value theory covers non-iid series, provided they are stationary and weakly correlated. For example, for stationary Gaussian series, the Berman condition^{4,20} provides a particularly simple check for applicability in terms of the autocorrelation function. It states that if the autocorrelation r_h at time-lag h satisfies $r_h \log h \rightarrow 0$ as $h \rightarrow \infty$, then (i) the distribution of block maxima follows Eq. 1 with $\xi = 0$, (ii) extremes do not cluster.

Our focus here is on the clustering of extremes. It is often observed that the extremes of e.g. temperatures³, water levels²¹, wind speeds⁶ or financial time series⁴⁷ cluster in time. In other words, they do not occur randomly as would be expected on the basis of a Poisson process. Extreme value theory can accommodate such

^{a)}Electronic mail: n.moloney@lml.org.uk

^{b)}Electronic mail: davide.faranda@lsce.ipsl.fr

^{c)}Electronic mail: ysato@math.sci.hokudai.ac.jp

^{d)}Also at London Mathematical Laboratory, 8 Margravine Gardens, London, UK

clustering via the so-called extremal index, while leaving the shape of the GEV distribution in Eq. 1 unchanged (this is not to be confused with the ‘extreme value index’, which in some literature refers to what we here call the shape parameter). This index, denoted $\theta \in [0, 1]$, essentially quantifies the inverse of the mean cluster size, and thus has an appealing physical meaning. When modeling threshold-exceeding extremes with GPDs, the extremal index estimates the effective number of iid degrees of freedom, since all the exceedances in a cluster are highly correlated. Nevertheless, the clusters themselves do occur randomly in the manner of a Poisson process, and therefore the occurrence of exceedances overall can be described by a compound Poisson process. In this way, the extremal index fits naturally within the point process framework of extremes by quantifying the multiplicity of the underlying compound Poisson process. This perspective is particularly insightful when treating recurrences in chaotic dynamical systems: visits of a trajectory to rare sets on the attractor will cluster if they occur in the manner of a compound Poisson process. This is indeed the case if the rare set includes an unstable fixed point. The extremal index is thus a tool for probing local properties of the attractor. The aim of this review is to work through these various statistical and dynamical interpretations.

It is important to mention that there is no exclusive definition of clustering, and this review restricts its attention to clustering as measured by the extremal index. Other notions of clustering are also developed in the literature. For example, the index of dispersion (also known as the coefficient of variation or Fano factor) is a diagnostic used in point processes¹⁴. It measures the ratio of the variance of the number of points in an interval, to the mean of the number of points in the same interval. For a Poisson process, this ratio is one. If the ratio is greater than one, the point process is said to be “over-dispersed”. Visually, points appear to come in bursts. Clustering is also sometimes associated with the Hurst effect³², since this describes persistence in an underlying time series. The Hurst effect is typically exhibited by long-range autocorrelated time series. In such cases, authors report non-exponential waiting times between extremes for both synthetic and real-world time series^{7,18,50}. Meanwhile, fractional Poisson processes are also capable of producing non-exponential waiting times, again interpreted as a possible mechanism for clustering⁵.

The review is structured as follows: in Sec. II we define the regime in which extreme value theory operates; Secs. III, IV, V builds up an intuition for the extremal index in its traditional statistical setting. Sec. VI gives a brief overview of recent developments in dynamical systems. We conclude in Sec. VII, and describe one of the simplest estimators of the extremal index, together with the idea of declustering, in the Appendix.

II. POISSON APPROXIMATION TO EXTREMES

Apart from furnishing an estimate of the size of extreme events, the Poisson approximation makes clear how these events occur within the framework of a point process. The basic idea is that extreme events, by virtue of the rarity, occur in the manner of a Poisson process.

Consider a set of iid random variables $\{X_i\}_{i=1}^n$, and let $\bar{F}(x) = \mathbb{P}(X > x)$ be the probability that the random variable X exceeds a value x . The number of exceedances n_u over a threshold u from n trials with success probability $\bar{F}(u)$ is a random variable following a binomial distribution, $\text{Bin}(n, \bar{F}(u))$. Therefore, the expected number of exceedances is $n\bar{F}(u)$.

To make the passage to extremes, suppose now that the threshold u_n is no longer fixed but allowed to increase with n in order to probe the tail of the distribution \bar{F} . If u_n is chosen such that

$$\lim_{n \rightarrow \infty} n\bar{F}(u_n) = \tau, \quad (2)$$

for some constant τ , then it is well known that the binomial distribution converges to a Poisson distribution with intensity τ . Thus, in the extreme regime, the number of exceedances is a Poisson random variable. This in turn gives an approximation to the maximum $M_n = \max\{X_1, \dots, X_n\}$, since if even the maximum fails to exceed the threshold, $M_n \leq u_n$, then the number of exceedances $n_u = 0$. But since n_u is a Poisson random variable,

$$\mathbb{P}(M_n \leq u_n) = \mathbb{P}(n_u = 0) = e^{-\tau}. \quad (3)$$

This approximation can be further refined to recover the GEV distributions in Eq. (1) via a suitable choice of the intensity $\tau = \tau(u_n)$ ¹⁹. Note that the Poisson approximation applies not only to extremely high exceedances, but to any kind of rare event. The distinguishing feature is that recurrences to the rare set (i.e. recurrence of an extreme event) take place in the manner of a Poisson process.

To extend this approximation to stationary but non-iid sequences, a condition on the decay of correlations is required. In the literature this is usually denoted $D(u_n)$, and we refer the reader to textbooks for its technical definition⁴². In words, it is a mixing condition requiring that well-separated maxima in different blocks of the sequence are asymptotically independent. Provided this condition is satisfied, the same family of GEV distributions describe the distribution of maxima as in the iid case. As before, a sequence of thresholds u_n satisfying Eq. (2) picks out the rare extremal events. But in the presence of correlations, instead of Eq. (3), we now have

$$\mathbb{P}(M_n \leq u_n) = e^{-\theta\tau}, \quad (4)$$

for some $\theta \in [0, 1]$, which defines the so-called the extremal index. The aim of this review is to provide some intuition for this quantity in terms of clustering.

We conclude this section with some remarks. First, by comparison with Eq. (3), we see that for iid random variables $\theta = 1$. Second, if the details of the sequence X_n are known, it may be possible to check explicitly whether $\theta < 1$ via an anticlustering condition. In the literature, this is often denoted $D'(u_n)$ ⁴². Third, it is important to realize that clustering relies on sufficiently strong short-range correlations. Thus, clustering does not affect block maxima, which are assumed to be asymptotically independent for extreme value theory to work in the first place. It does, however, potentially affect a sequence of exceedances over a threshold: by virtue of short-range correlations, exceedances may cluster rather than appear as isolated events. From a modeling point of view, it is therefore important to check whether clustering is present in order to avoid biased fits to the generalised Pareto distribution.

III. LOSS OF DEGREES OF FREEDOM

In this section we consider the extremal index in terms of a loss of effective iid degrees of freedom. In the presence of correlations, extremes may cluster in time such that extremes inside the cluster cannot reasonably be treated as iid. Clustering therefore results in a loss of effective iid degrees of freedom, and the extremal index gives a measure of the fraction of extremes that are approximately iid.

As a concrete example, consider a series Y_i of iid random variables drawn from the distribution

$$F_Y(y) = \exp(-1/2y), \quad y \geq 0, \quad (5)$$

and construct from it the series X_i , according to

$$X_i = \max\{Y_i, Y_{i+1}\}. \quad (6)$$

Evidently, the series X_i is not iid but its marginal distribution can nevertheless be calculated:

$$F_X(x) = \mathbb{P}(X_i \leq x) = \mathbb{P}(Y_i \leq x, Y_{i+1} \leq x) \quad (7)$$

$$= \mathbb{P}(Y_i \leq x)\mathbb{P}(Y_{i+1} \leq x) \quad (8)$$

$$= \exp(-1/x) \quad (9)$$

For the sake of comparison, now introduce an iid series X_i^* with exactly the same marginal distribution as X_i , i.e. $F_{X^*}(x) = \exp(-1/x)$. Representative sequences of X_i and X_i^* are shown in Fig. 1. The dependent sequence X_i clearly shows clustering among large values, whereas the independent sequence X_i^* of course does not. We can make this observation explicit by calculating the extreme value distribution for each sequence. For the marginal distribution in this example, the typical size of maxima grows linearly with n . Indeed, by direct calculation, the properly normalized limiting distribution of $M_n^* = \max\{X_1^*, \dots, X_n^*\}$ is given by

$$\mathbb{P}(M_n^* \leq nx) = F_{X^*}^n(nx) \quad (10)$$

$$= \exp(-1/x), \quad (11)$$

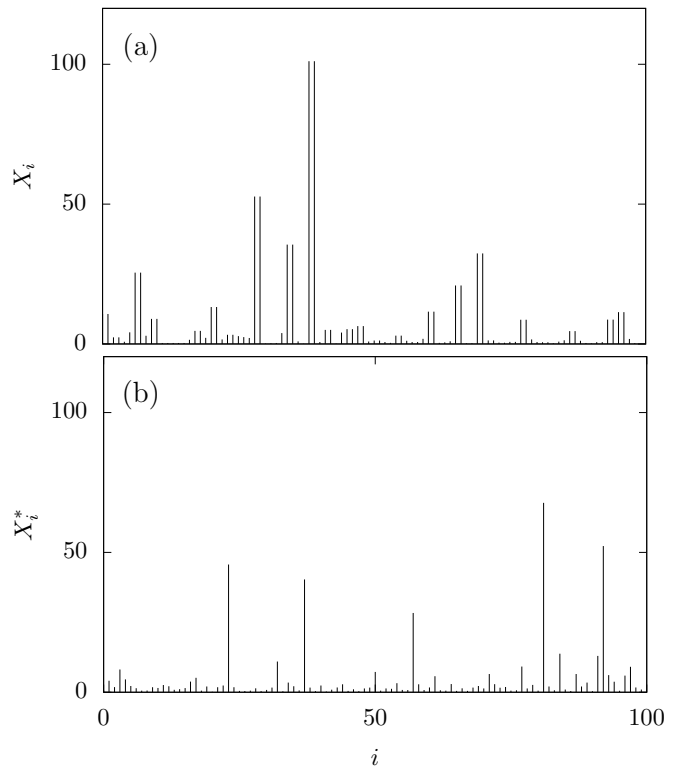


FIG. 1. (a) Time series of the clustered process $X_i = \max\{Y_i, Y_{i+1}\}$ (Eq. (6)) and (b) an iid process X_i^* having the same marginal distribution as X_i .

which is an example of a Fréchet distribution. A similar calculation for the dependent sequence X_i gives

$$\mathbb{P}(M_n \leq nx) = \mathbb{P}(X_1 \leq nx, \dots, X_n \leq nx) \quad (12)$$

$$= \mathbb{P}(Y_1 \leq nx, Y_2 \leq nx, \dots, Y_n \leq nx, Y_{n+1} \leq nx) \quad (13)$$

$$= F_Y^{n+1}(nx) \quad (14)$$

$$\rightarrow [\exp(-1/x)]^{1/2}, \quad \text{as } n \rightarrow \infty. \quad (15)$$

The distributions of maxima for the two series are related via

$$\mathbb{P}(M_n \leq nx) = [\mathbb{P}(M_n^* \leq nx)]^\theta \quad (16)$$

$$= \mathbb{P}(M_n^* \leq \theta nx), \quad (17)$$

with extremal index $\theta = 1/2$. Thus, we have the interpretation that the maxima of the dependent series are distributed in the same way as those of the independent series, but with a series length that has been reduced from n to θn . In this example, half the effective iid degrees of freedom have been lost. This is to be expected from the definition of the process X_i since values are correlated in pairs.

Turning now to a real-world example, let us consider a simplified extreme value analysis of the daily minimum temperatures from Wooster, Ohio — for a more careful treatment, see⁹. Fig. 2 shows the coldest 150 daily temperatures between 1983-1988. The temperature series

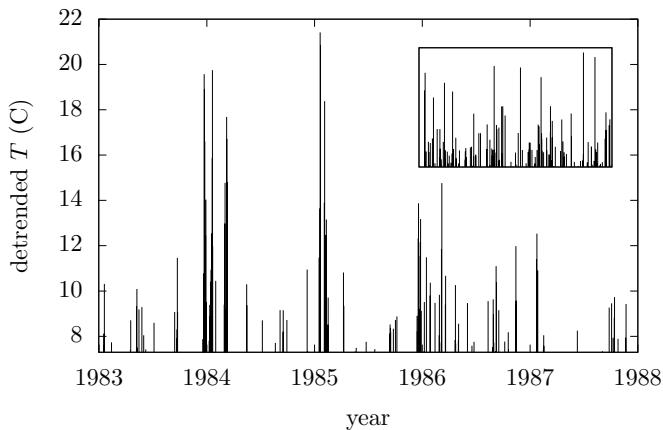


FIG. 2. Detrended coldest 150 days in Wooster, from 1983-1988. The inset shows a randomly shuffled version of the same data.

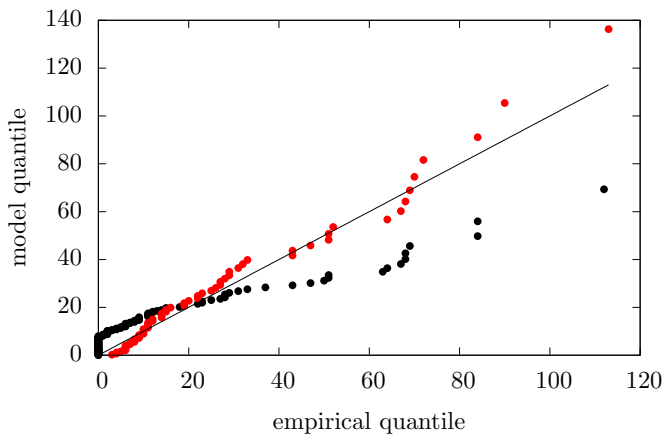


FIG. 3. Quantile plot for waiting times between exceedances, comparing empirical and exponential model quantiles for the original (black) and declustered (red) series.

has been detrended (by subtracting a seasonal sinusoidal component) and flipped in sign so that minima become maxima — a conventional choice. For comparison, the inset shows the randomly shuffled series of the same exceedances. Clustering is visually apparent in the original series. Indeed, we find $\theta = 0.54$ for the detrended series, and $\theta = 0.89$ for the shuffled series. This latter value is apparently less than one. We discuss this discrepancy as a finite-size effect in Sec. V.

Since unclustered exceedances arrive in the manner of a Poisson point process, they are separated by exponential waiting times. A diagnostic for comparing empirical data with a model distribution is to plot empirical versus model quantiles for a given probability in a quantile plot, see Fig. 3. Random variables drawn from an exponential distribution would scatter along the diagonal identity line. The original data (black points), however, deviate from the diagonal line for both small and large quantiles. To demonstrate that this is due to clustering, we pass the data through a declustering algorithm (described in

the Appendix), which identifies clusters of exceedances and replaces them by a single representative exceedance for each cluster. In this way, remaining exceedances may be treated as effectively iid. In Fig. 3, the declustered data (red points) lie much closer to the diagonal identity line. Declustering is often employed as a preprocessing step before fitting exceedances to e.g. GPDs.

IV. COMPOUND POINT PROCESS

We now consider our second interpretation of the extremal index in terms of compound point processes. A compound point process can be thought of as a point process in which points carry marks. For example, the mark might be the size of a claim made against an insurance company, or the size of a group of people arriving at a restaurant. In the present context, the mark corresponds to the weight of clusters in a mixture distribution, as we further develop here.

Adopting the language of statistical physics, it is instructive to examine extreme exceedances under the renormalization of an underlying point process^{11,12}. This consists of two steps: a thinning step that knocks out points that do not exceed the threshold, and a time rescaling step that restores the original intensity of the points. These two steps constitute a single renormalization operation. Applying this operation many times is akin to inspecting the process at long timescales. The question, first studied by Rényi⁴⁹, is whether the sequence of renormalized point process converges towards some fixed point. We answer this question in the following for iid random variables, before considering a non-random thinning procedure which gives rise to clustering.

Given that we are particularly interested in the manner of arrivals of exceedances, we will track the evolution of the probability density of waiting times between exceedances under renormalization. If we denote this by $f(t)$, then the waiting time density $f_p(t)$ between points that exceed the threshold with probability p is given by convolution. Specifically, we convolve over points separated by any number of intermediate waiting times, weighted appropriately. An example of a term that appears in this convolution is shown in Fig. 4, in which two exceedances are separated by a time which is the sum of three independent waiting times. Summing over all weighted possibilities, we have

$$f_p(t) = pf(t_1) + pqf(t_1+t_2) + pq^2f(t_1+t_2+t_3) + \dots, \quad (18)$$

where $q = 1-p$. Since the right-hand side is made up of a convolution of waiting times, it is convenient to introduce the Laplace transform

$$\hat{f}(s) = \int_0^\infty dt e^{-st} f(t), \quad (19)$$

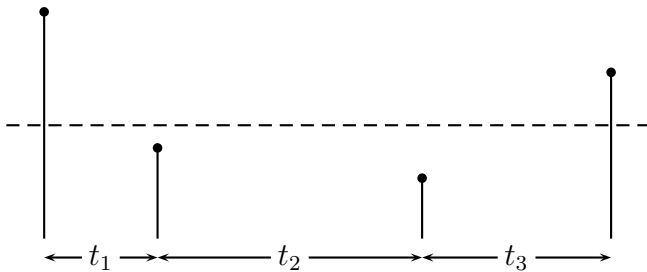


FIG. 4. Example term from the convolution sum in Eq. 18. The exceedances are separated by the sum of three independent waiting times, $t_1 + t_2 + t_3$. The weight of this contribution is $q^2 p$.

which, applied to Eq. (18), yields

$$\hat{f}_p(s) = p \left[\hat{f}(s) + q\hat{f}^2(s) + q^2\hat{f}^3(s) + \dots \right] \quad (20)$$

$$= \frac{p\hat{f}(s)}{1 - q\hat{f}(s)}. \quad (21)$$

As a result of thresholding, only a fraction p of the original points are retained. Consequently, the intensity of the process decreases. The second step of the renormalization procedure rescales time, $t \rightarrow t/p$, in order to restore the original intensity of the process. In Laplace space, this corresponds to the rescaling $s \rightarrow ps$. Thus, the combined steps of thresholding and rescaling time leads to

$$\hat{f}_p(s) = \frac{p\hat{f}(ps)}{1 - q\hat{f}(ps)}, \quad (22)$$

which defines a fixed point equation for the renormalized waiting time density. As can be checked by substitution, this functional equation is satisfied by

$$\hat{f}_p(s) = \frac{1}{1 + s/\lambda}, \quad (23)$$

where λ is a constant. It can be shown that this fixed point is attracting, such that upon repeated application of the renormalization procedure the waiting time density flows towards the form described in Eq. (23). Inverting Eq. (23), we arrive at

$$f_p(t) = \lambda e^{-\lambda t}, \quad (24)$$

which is the familiar exponential waiting time density for a Poisson point process of intensity λ . Although we assumed iid random variables at the outset, the above result is more general: short-range correlations that may be present in the original point process do not survive renormalization. A more mathematically rigorous treatment may be found in^{40,56}.

It would appear, then, that the Poisson point process is the attractor for renormalized high exceedances, and that there is no clustering among extremes. Indeed, this

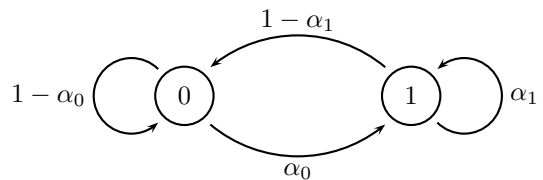


FIG. 5. Two states of a Markov chain that determine whether to delete (0) or retain (1) points. The transition probabilities α_0 and α_1 are tuned so that a fraction p of points are retained.

is true for a thinning procedure that knocks out points at random.

For a thinning procedure that produces clustering, we follow Isham³⁹ and introduce Markovian thinning. That is, we retain or delete points according to the state of a two-state Markov process. Let us label these states 1 (for retain) and 0 (for delete), and define transition probabilities α_1 for remaining in state 1, and $1 - \alpha_0$ for remaining in state 0 (see Fig. 5). We can tune these rates so as to retain a fraction p of points by requiring that

$$p = \alpha_1 p + \alpha_0 (1 - p), \quad (25)$$

which is essentially a stationarity condition imposed on the Markov process. Each time a point is visited, the internal state of the Markov process is updated according to the transition probabilities satisfying Eq. (25). With this thinning protocol set up, we can again examine the evolution of the renormalized waiting time density $f_p(t)$. The full details are given in³⁹, but here we just cite the main result that the fixed point equation now reads

$$\hat{f}_p(s) = \frac{\alpha_1 \hat{f}(ps) + (\alpha_0 - \alpha_1) \hat{f}^2(ps)}{1 - (1 - \alpha_0) \hat{f}(ps)}. \quad (26)$$

This more complicated functional equation requires an expansion in small p , which corresponds to retaining a vanishing number of points and is the appropriate limit for inspecting the renormalized point process. For $p \rightarrow 0$ we obtain

$$\lim_{p \rightarrow 0} \hat{f}_p(s) = \alpha_1 + (1 - \alpha_1) \frac{1}{1 + s/(\lambda(1 - \alpha_1))}, \quad (27)$$

which, after inverting, yields

$$f_p(t) = \alpha_1 \delta_{0,t} + (1 - \alpha_1) \left[\lambda(1 - \alpha_1) e^{-(\lambda(1 - \alpha_1)t)} \right]. \quad (28)$$

The second term on the right-hand side is the usual exponential waiting time for a Poisson point process. The first term represents a Dirac point mass at zero waiting time. This term corresponds to the clustering of points which, under renormalization, have become squashed together. Thus, the Markovian thinning procedure results in a mixture distribution for the renormalized waiting time density: clusters of (compounded) points are separated by exponential waiting times. The weight of these

clusters in the mixture distribution is given by α_1 . If we adopt the interpretation of Sec. III, then only a fraction $1 - \alpha_1$ of exceedances are effectively iid, and so heuristically we have that $\theta = 1 - \alpha_1$. It can be shown more generally that when exceedances cluster, a mixture distribution of the same form as Eq. (28) describes their intervals²⁹.

In summary, a compound Poisson process can be thought of as a generalization of a Poisson process. In particular, points are still separated by exponential waiting times. However, a random variable is now assigned to each point (independent of the waiting times) which represents the size of a cluster. Under renormalization, clusters occur instantaneously, giving rise to a Dirac point mass of zero waiting time (see Eq. (28)). In practice, clusters typically appear as a localised train of successive exceedances. Thus, the cluster size — previously assigned to a point — now represents the length of this train.

V. INVERSE MEAN CLUSTER SIZE

Our next interpretation of the extremal index relates clustering to the time spent above a threshold (also known as the sojourn time). In the case of the process described by Eq. (6), the fact that exceedances appear in clusters of size two is directly related to the halving of the effective iid degrees of freedom, $\theta = 1/2$. Indeed, the inverse of the extremal index θ^{-1} is approximately the mean cluster size or, equivalently, the mean time spent above threshold. Note that a cluster need not consist of strictly consecutive exceedances, but to simplify calculations we will assume this is the case in the following.

For iid random variables, it is possible by chance to have a consecutive sequence of threshold exceedances. But the higher the threshold, the more probable it is that exceedances occur as isolated events. A calculation of the mean sojourn time confirms this: an exceedance begins a cluster of size k if the subsequent $k - 1$ values exceed the threshold, and the k th value does not. This event occurs with probability $p^{k-1}q$. Summing over all possibilities, the average cluster size is

$$\frac{1}{\theta} \approx 1 \cdot q + 2 \cdot pq + 3 \cdot p^2q + \dots = \frac{1}{1-p}, \quad (29)$$

which tends to one (i.e. no clustering) in the limit of small p . For clustering to persist in the limit of high thresholds, exceedances must be strongly correlated. It can be shown, for example, that clustering does not occur in a Gaussian sequence with exponential or even power-law decaying autocorrelations¹⁹.

Nevertheless, in practice empirically estimated extremal indices may indicate clustering for some series even if it is known theoretically that $\theta = 1$. This is because arbitrarily high thresholds cannot be applied to series of finite length. It is instructive to work through

the details of this finite size effect for a Gaussian, exponentially correlated time series, as given by the AR(1) process

$$X_n = \phi X_{n-1} + \sqrt{1 - \phi^2} Z_n, \quad (30)$$

where $|\phi| < 1$ is the autoregressive parameter and Z_n is standard Gaussian white noise (zero mean, unit standard deviation). The prefactor $\sqrt{1 - \phi^2}$ in front of the noise term is purely for standardization and guarantees that the overall stationary marginal distribution of X_n is Gaussian with unit standard deviation.

For the purpose of calculating sojourn times, it is convenient to work with the transition kernel $W(x_{k+1}|x_k)$ which evolves the distribution of X_k at step k into the distribution X_{k+1} at step $k + 1$. For the AR(1) process driven by Gaussian noise, this is given by

$$W(x_{k+1}|x_k) = \frac{1}{\sqrt{2\pi(1 - \phi^2)}} \exp[-(x_{k+1} - \phi x_k)^2 / (2(1 - \phi^2))]. \quad (31)$$

Given the homogeneous Markovian property of the AR(1) process, a sequence of transitions can be chained together to evolve an AR(1) trajectory starting at x_0 according to

$$W(x_k|x_0) = \int_{\mathcal{C}} dx_1 \dots dx_{k-1} W(x_k|x_{k-1}) \dots W(x_1|x_0), \quad (32)$$

where the multidimensional integrals over intermediate points x_1, \dots, x_{k-1} are subject to path constraints \mathcal{C} , such as the process staying within some interval. If this interval is denoted by (a, b) , then the probability that the trajectory remains within the interval in the preceding n steps is

$$\mathbb{P}_k = \int_a^b dx_1 \dots dx_{k-1} W(x_k|x_{k-1}) \dots W(x_1|x_0). \quad (33)$$

Multidimensional integrals of this type are known as orthant probabilities, and a number of computational schemes have been developed for their numerical calculation^{2,15,17}. However, for present purposes it is sufficient to numerically compute Eq. (33) using *Mathematica*, since after a handful of steps k very little probability remains in the interval.

Large or small extremes may be considered by taking $b \rightarrow \infty$ or $a \rightarrow -\infty$, in which case the sojourns either exceed level a or fall below level b , respectively. Here, instead, we will consider small intervals (a, b) which distinguish extreme events by virtue of their rarity — such rare sets are of particular interest in dynamical systems, as discussed in Sec. VI. Thus, the sojourns we will consider are confined to a small interval (a, b) , in which the trajectory typically exits soon after entering, as illustrated in Fig. 6. \mathbb{P}_k decreases with every step since fewer trajectories remain within the interval. The change in probability leaving the interval thus gives the sojourn time f_k via

$$f_k = \mathbb{P}_{k-1} - \mathbb{P}_k, \quad (34)$$

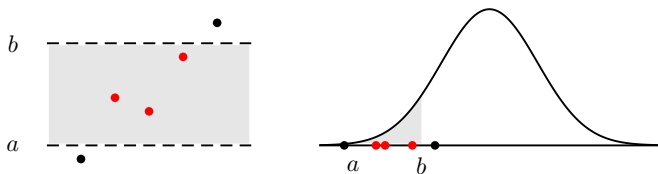


FIG. 6. (Left) Trajectory of an AR(1) process evolving in time, of which a portion sojourns inside the interval (a, b) , defining a cluster of size 3. (Right) The same trajectory, with the marginal density of the AR(1) process overlaid.

with initial condition $\mathbb{P}_0 = 1$.

Thus, by numerically computing the multidimensional integral in Eq. (33) and substituting into Eq. (34), we use the sojourn time probabilities f_k to approximate the extremal index in terms of the mean cluster size via

$$\frac{1}{\theta} \approx \sum_{k=1}^{\infty} k f_k. \quad (35)$$

Since the probability of long sojourns decays rapidly with k , mean sojourn times in small intervals are well estimated by retaining only the first handful of terms in the above sum.

In Figs. 7 and 8 we show how the theoretical no-clustering limit is approached when ϕ and q is changed. Note that the estimated extremal index depends on the location x_0 of the interval. This is because intervals further away from the centre of the distribution (where the density is highest) must widen in order to enclose the same integrated area $1 - q$. A wider interval in turn leads to a longer sojourn time, and hence more clustering.

The characteristic decay time of the exponential memory of the AR(1) process decreases with ϕ (vanishing in the white noise limit $\phi \rightarrow 0$). Thus, clustering too should likewise decrease. This is borne out in Fig. 7, which shows estimated extremal indices (points) approaching $\theta = 1$ with decreasing ϕ , together with theoretically predicted curves based on Eq. (35). Finally, increasing q (i.e. decreasing the size of sojourn intervals) also results in the same trend towards no clustering, as shown in Fig. 8. In such numerical experiments, time series with 10^7 points are easily obtainable. But for real-world time series very high quantiles are often unrealistic, and significant finite-size effects may be present. Curves such as those in Figs. 7 and 8 are therefore useful for discerning possible clustering in an underlying time series, beyond that given by finite-size artefacts²⁴.

It is also worth mentioning that, in practice, empirical values of the extremal index will depend on the temporal resolution of the time series. Implicitly, we have measured the size of clusters in units of the cadence of the (discrete) time series. If an observable is poorly resolved in time, then clusters may likewise be poorly resolved. Typically, this will tend to increase the empirical value of the extremal index.

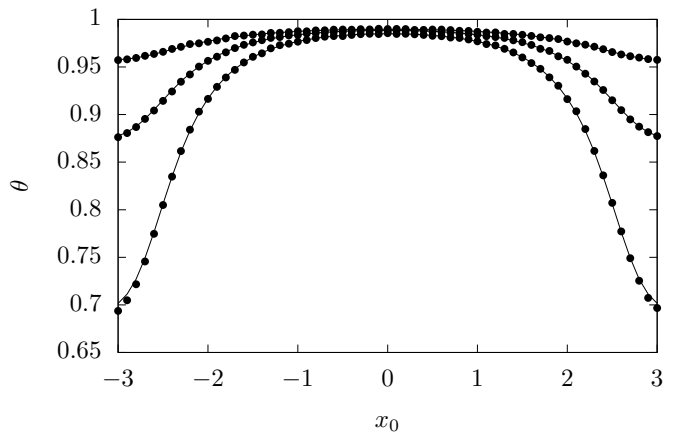


FIG. 7. Estimated extremal index (points), together with inverse mean cluster size based on Eq. (35) (solid lines), for an ensemble of 100 AR(1) time series of length 10^7 , with $\phi = 0.25, 0.5, 0.75$ (top-to-bottom) and $q = 0.99$.

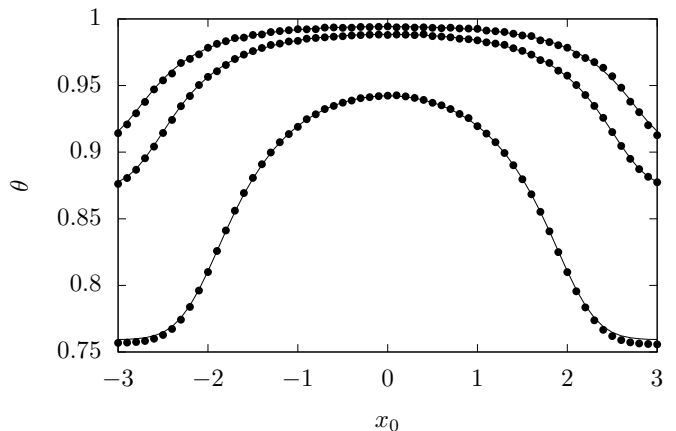


FIG. 8. Estimated extremal index (points), together with inverse mean cluster size based on Eq. (35) (solid lines), for an ensemble of 100 AR(1) time series of length 10^7 , with $\phi = 0.5$ and $q = 0.95, 0.99, 0.995$ (bottom-to-top).

VI. EXTREMAL INDEX IN DYNAMICAL SYSTEMS

A. Background

We have emphasized thus far that extreme value theory is underpinned by rare events within the framework of point processes. Put differently, a random variable visits a rare set in state space if its arrivals take place in the manner of a (possibly compound) Poisson process. Rarity, here, is expressed by the fact that the mean number of exceedances over a threshold u_n is fixed to some intensity τ , i.e. $n\bar{F}(u_n) \rightarrow \tau$.

These ideas fit naturally within dynamical systems satisfying Poincaré's recurrence theorem, since visits by orbits to rare sets typically take place after exponential waiting times with an intensity given by Kac's lemma³⁵. Strikingly, extreme value theory is able to provide lo-

cal dynamical information about these rare sets, such as their local dimension and ‘stickiness’. The local dimension, also known as the pointwise dimension⁵⁷, gives the dimension at a point on an attractor. It characterizes the unpredictability of the dynamics at that point based on the number of expanding directions. The local dimension is in fact estimated by the inverse of the scaling parameter σ of the GEV distribution⁴³.

Here, instead, we focus on ‘stickiness’, which is the inverse of the mean sojourn time near a point on an attractor. Stickiness can therefore be estimated by the extremal index θ . The main result is that generic points on the attractor do not admit clustering, so that recurrences to small balls around these points is Poisson. But at (periodic) unstable fixed points, recurrences cluster and are compound Poisson. We give a brief overview of this result in the following. For rigorous proofs, a recent book⁴³ provides a detailed survey.

Consider a deterministic dynamical system T which evolves an initial condition x_0 according to $x_{n+1} = T(x_n) = T^n(x_0)$ (our discussion can be generalized to flows). Assume that this map has a unique and compact attractor in phase space. Extreme value theory is applied to the hitting time statistics of an ϵ -ball $B_\epsilon(\zeta)$ centered on a point $x = \zeta$ on the attractor:

$$\mathbb{P}(T^n(x_0) \in B_\epsilon(\zeta)). \quad (36)$$

To make the connection to stationary stochastic series, we construct the random variables $X_n = g(T^n(x_0), \zeta)$, where the observable g takes its global maximum at the point ζ . The stochasticity in this series comes from the sampling of the initial condition x_0 from the invariant density of the map. A convenient choice for g is

$$X_n = -\log(\text{dist}(T^n(x_0), \zeta)), \quad (37)$$

where dist is any metric (e.g. Euclidean) between x_n and ζ . By construction, iterates $T^n(x_0)$ that land close to ζ will be associated with observables X_n with large positive values. Furthermore, it can be shown that taking the logarithm of the distance fixes the maxima of X_n to the Gumbel class¹⁰. This result hinges on the asymptotic independence of maxima. To this end, a mixing condition very similar to that used in stochastic processes can be developed for dynamical systems³³. Assuming this condition holds, an estimate for the maximum M_n takes the familiar form (see Sec. II)

$$\mathbb{P}(T^n(x_0) \in B_{\epsilon_n}(\zeta)) = \mathbb{P}(M_n \leq u_n; \zeta) = e^{-\theta\tau}. \quad (38)$$

In the dynamical systems context, ϵ_n plays the rôle of a threshold and is related to u_n via $\epsilon_n = e^{-u_n}$: as the length n orbit increases, the radius of the ϵ -balls shrink, thereby maintaining a constant intensity τ of recurrences. The sequence u_n will contain information about the location and scale parameters of the Gumbel distribution fitted to M_n ²³. In general, these will need to be estimated. It can be shown that the inverse scale parameter σ^{-1} of this fit is in fact an estimate of the local Hausdorff

dimension $d(\zeta)$ ⁴⁴. This quantity, averaged over all points ζ on the attractor, provides an alternative estimate of the attractor dimension²⁶.

Here, we focus our attention on the extremal index θ appearing in Eq. (38). Freitas-Freitas-Todd³⁴ have proved that, if ζ is an unstable periodic point with prime period k , then

$$\theta(\zeta) = 1 - |\text{Det} [DT^{-k}|_s(\zeta)]|, \quad (39)$$

where $DT^{-k}|_s$ is the derivative of the k th inverse of T restricted to the contracting subspaces. For all other points ζ on the attractor, $\theta(\zeta) = 1$. Put differently, if the $B_{\epsilon_n}(\zeta)$ is centered on an unstable periodic fixed point, recurrences to that ball arrive in the manner of a compound Poisson process. Otherwise, recurrences are Poisson.

For applications, Eq. (39) implies that ‘sticky’ regions of the dynamics can be identified by measuring the extremal index at various points on the attractor and checking for $\theta < 1$. Faranda et al.²⁶ have used this method to characterize atmospheric flows.

B. Examples

We now illustrate how Eq. (39) can be used to calculate the extremal index for some simple chaotic maps.

1. Uniformly chaotic maps

The Bernoulli map is given by

$$x_{n+1} = f(x_n) = ax_n \pmod{1}, \quad (40)$$

where $a > 1$, and $x \in [0, 1]$. For almost all initial conditions, the map generates uniformly chaotic dynamics. If we take, say, $a = 3$, then three unstable fixed points are located at $x = 0$, $x = 1/2$ and $x = 1$. By Eq. (39), the extremal index at these points is

$$\theta(0) = \theta(1/2) = \theta(1) = 1 - |Df^{-1}(0)| \quad (41)$$

$$= 1 - \frac{1}{3} = \frac{2}{3}. \quad (42)$$

For all aperiodic points, $\theta(x) = 1$, since clustering can only occur at unstable fixed points by the Freitas-Freitas-Todd theorem.

2. Intermittent maps

The Pomeau-Manneville map, a toy model for intermittency in turbulent fluids, is given by

$$x_{n+1} = f(x_n) = \begin{cases} x_n + 2^{\alpha-1}x_n^\alpha & x \in [0, \frac{1}{2}) \\ 2x_n - 1 & x \in [\frac{1}{2}, 1] \end{cases}, \quad (43)$$

where $\alpha > 1$, and $x \in [0, 1]$. For stationarity, we require that $\alpha < 2$. Two unstable fixed points are located at

$x = 0$ and $x = 1$. The vicinity of the neutrally stable fixed point $x = 0$ is associated with a sticky region of the ‘laminar’ phase. At the neutrally stable fixed point at $x = 0$, we have the extremal index

$$\theta(0) = 1 - |Df^{-1}(0)| = 1 - 1 = 0., \quad (44)$$

which implies a diverging mean sojourn time, consistent with the trapping of orbits for anomalous amounts of time. At the unstable fixed point at $x = 1$,

$$\theta(1) = 1 - |Df^{-1}(1)| = 1 - \frac{1}{2} = \frac{1}{2}. \quad (45)$$

For all aperiodic points, $\theta(x) = 1$.

3. Hénon map

The Hénon map is a generic two-dimensional invertible map, originally inspired by three-body Hamiltonian dynamics. It takes the form

$$(x_{n+1}, y_{n+1}) = f(x_n, y_n) = (1 - ax_n^2 + y_n, bx_n), \quad (46)$$

where a and b are parameters. The dissipation rate $b \in [0, 1]$ interpolates between one-dimensional logistic dynamics for $b = 0$, and Hamiltonian dynamics for $b = 1$. With standard parameter choices $a = 1.4$ and $b = 0.3$, the extremal index $\theta(x^*, y^*)$ at the unstable fixed point $(x^*, y^*) \approx (0.631854, 0.189406)$ is given by

$$\theta(x^*, y^*) = 1 - |Df^{-1}|_s(x^*, y^*)| = 1 - |\lambda_-| \approx 0.48, \quad (47)$$

where $\lambda_- \approx -0.52$ is the eigenvalue of the Jacobian matrix $Df^{-1}(x^*, y^*)$, whose absolute value is less than 1, corresponding to the contracting direction. For aperiodic points, $\theta(x, y) = 1$. Fig. 9 illustrates how a scan of the numerically estimated extremal index across the attractor picks out the unstable fixed point, based on the clustering of recurrences.

VII. CONCLUSION

The extremal index is one of a number of quantities that measure the clustering of extremes in stationary time series. If $\theta = 1$, extremes do not cluster and arrive in the manner of a Poisson process; if $\theta < 1$, extremes cluster and arrive in the manner of a compound Poisson process. In this review we have developed some intuition for this quantity in terms of the (i) loss of effective degrees of freedom, (ii) multiplicity of a compound Poisson point process, (iii) sojourn time inside a rare set, and the (iv) stickiness of unstable fixed points in dynamical systems.

The compound Poisson process that underpins the extremal index is asymptotic, in the sense of Eq. (2). In practice, it is not possible to apply arbitrarily high thresholds to series, and therefore one has to be alert

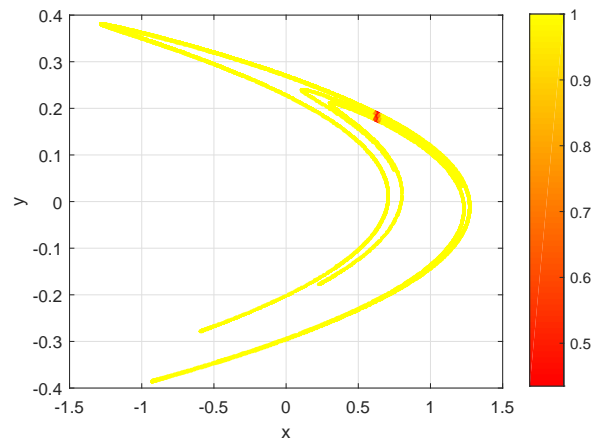


FIG. 9. The extremal index for different points on the Hénon attractor. For generic points, $\theta = 1$ (no clustering). But at the unstable fixed point $(0.631854, 0.189406)$, $\theta \approx 0.48$. These theoretical values are well recovered via numerical estimates, as indicated by the color scale.

to possible convergence issues, as explored in Sec. V for an AR(1) process. In future, it would be interesting to reconcile the implication of Berman’s condition with numerical studies on Gaussian long-range dependent time series¹⁸, where non-exponential waiting times between extremes are observed.

Although the extremal index is widely used as a statistical parameter in the modeling of extremes, we hope we have convinced the reader that it is an interesting quantity in its own right. For example, in the dynamical systems context, extreme value theory give access to local and global properties of the attractor. In this review, we have considered recurrences between trajectories and points in phase space⁴⁵. Instead, by considering recurrences between pairs of trajectories, it is possible to infer global dynamical quantities such as the correlation dimension and Lyapunov exponent²⁷. For systems with a space-time structure, the extremal index is related to synchronization of the dynamics²². All such dynamical quantities can be extracted by fits to universal extreme value distributions and are therefore readily available empirically²³. This versatility is reflected in a large number of applications to geophysical flows, where local dimensions and persistence are associated with typical atmospheric configurations^{25,26}, and extreme events³⁷ or dynamical shifts in turbulent boundary layers⁵⁵ can be detected.

ACKNOWLEDGMENTS

NRM would like thank Nicholas Watkins, Álvaro Corral and Víctor Navas-Portella for useful discussions.

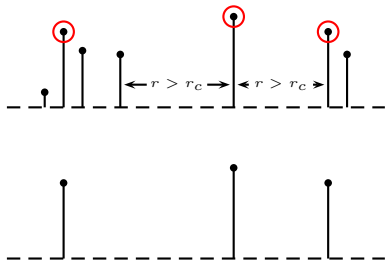


FIG. 10. Declustering of exceedances. (Top) Clusters are identified as distinct if they are separated by more than a run length r_c . (Bottom) Each cluster is reduced to a single exceedance.

Appendix A: Statistical estimation of the extremal index and declustering

Here we describe the ‘runs’ method⁵², one of the simplest ways of estimating the extremal index. It is computed by dividing the number of clusters, n_c , by the number of exceedances n_u :

$$\hat{\theta} = \frac{n_c}{n_u}. \quad (\text{A1})$$

If all clusters are composed solely of single exceedances, then this ratio is one, consistent with the absence of clustering. The runs method requires a definition of a cluster, which introduces a degree of arbitrariness. Pragmatically, a cluster is defined as terminated when the series has stayed below threshold for at least r_c steps. The extremal index can be recomputed for various choices of the run length r_c to check for robustness¹⁹.

Having defined a cluster, it is possible to decluster exceedances by replacing each cluster with one of its exceedances, see Fig. 10. In this way, the short-range correlated exceedances within a cluster are replaced by a single effective degree of freedom. It typically does not matter which exceedance is retained. For example, the largest or middle exceedance of a cluster is a reasonable choice.

This declustering procedure has been applied to exceedances in the Wooster temperature series in Sec. III. An indication of the procedure’s success is shown in Fig. 3, since the declustered exceedances are much closer to exponential model quantiles than the original exceedances.

There are other estimators of the extremal index which do not require the specification of a run length^{29,54}. In the experience of the authors, these tend to be more reliable for applications. For a review, see²⁸.

¹Barriopedro, D., Fischer, E. M., Luterbacher, J., Trigo, R. M., and García-Herrera, R. (2011). The hot summer of 2010: Redrawing the temperature record map of Europe. *Science*, 332(6026):220–224.

²Basak, G. K. and Ho, K. R. (2004). Level-crossing probabilities and first-passage times for linear processes. *Adv. Appl. Probab.*, 36:643–666.

- ³Beirlant, J., Goegebeur, Y., Segers, J., and Teugels (2004). *Statistics of extremes: theory and applications*. John Wiley & Sons Inc.
- ⁴Berman, S. M. (1964). Limit theorems for the maximum term in stationary sequences. *Ann. Math. Statist.*, 35:502–516.
- ⁵Blender, R., Raible, C. C., and Lunkeit, F. (2015). Non-exponential distributions for vorticity extremes explained by fractional Poisson processes. *Q. J. R. Meteorol. Soc.*, 141:249–257.
- ⁶Brabson, B. B. and Palutikof, J. P. (2000). Tests of the Generalized Pareto Distribution for Predicting Extreme Wind Speeds. *J. Appl. Meteorol.*, 39:1627–1640.
- ⁷Bunde, A., Eichner, J. F., Kantelhardt, J. W., and Havlin, S. (2005). Long-Term Memory: A Natural Mechanism for the Clustering of Extreme Events and Anomalous Residual Times in Climate Records. *Phys. Rev. Lett.*, 94:048701.
- ⁸Coles, S. (2001). *An introduction to statistical modeling of extreme values*. Springer-Verlag, London.
- ⁹Coles, S. G., Tawn, J. A., and Smith, R. L. (1994). A seasonal Markov model for extremely low temperatures. *Environmetrics*, 5:221–239.
- ¹⁰Collet, P. (2001). Statistics of closest return for some non-uniformly hyperbolic systems. *Ergod. Theory Dyn. Syst.*, 21(2):401–420.
- ¹¹Corral, A. (2009). Point-occurrence self-similarity in crackling-noise systems and in other complex systems. *J. Stat. Mech.*, page P01022.
- ¹²Corral, A. (2015). Scaling in the timing of extreme events. *Chaos Soliton. Fract.*, 74:99–112.
- ¹³Coumou, D. and Rahmstorf, S. (2012). A decade of weather extremes. *Nature Clim. Change*, 2(7):491–496.
- ¹⁴Cox, D. R. and Isham, V. (1980). *Point Processes*. Chapman and Hall, London and New York.
- ¹⁵Craig, P. (2008). A new reconstruction of multivariate normal orthant probabilities. *J. R. Statist. Soc. B.*, 70:227–243.
- ¹⁶de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory: An Introduction*. Springer, New York.
- ¹⁷Di Nardo, E. (2008). On the first passage time problem for linear processes. *Scientiae Math. Japonica*, 21:61–76.
- ¹⁸Eichner, J. F., Kantelhardt, J. W., Bunde, A., and Havlin, S. (2007). Statistics of return intervals in long-term correlated records. *Phys. Rev. E*, 75:011128.
- ¹⁹Embrechts, P., Klüppelberg, C., and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, Berlin Heidelberg.
- ²⁰Falk, M., Hüsler, J., and Reiss, R.-D. (2004). *Laws of Small Numbers: Extremes and Rare Events*. Birkhäuser Verlag, Basel, Boston, Berlin, 2nd edition.
- ²¹Fang, B., Guo, S., Wang, S., Liu, P., and Xiao, Y. (2007). Non-identical models for seasonal flood frequency analysis. *Hydrolog. Sci. J.*, 52:974–991.
- ²²Faranda, D., Ghoudi, H., Guiraud, P., and Vaienti, S. (2018). Extreme value theory for synchronization of coupled map lattices. *Nonlinearity*, 31(7):3326.
- ²³Faranda, D., Lucarini, V., Turchetti, G., and Vaienti, S. (2011). Numerical convergence of the block-maxima approach to the generalized extreme value distribution. *J. Stat. Phys.*, 145(5):1156–1180.
- ²⁴Faranda, D., Masato, G., Moloney, N., Sato, Y., Daviaud, F., Dubrulle, B., and Yiou, P. (2016). The switching between zonal and blocked mid-latitude atmospheric circulation: a dynamical system perspective. *Clim. Dyn.*, 47:1587.
- ²⁵Faranda, D., Messori, G., Alvarez-Castro, M. C., and Yiou, P. (2017a). Dynamical properties and extremes of northern hemisphere climate fields over the past 60 years. *Nonlinear Proc. Geophys.*, 24(4):713–725.
- ²⁶Faranda, D., Messori, G., and Yiou, P. (2017b). Dynamical proxies of North Atlantic predictability and extremes. *Sci. Rep.*, 7:41278.
- ²⁷Faranda, D. and Vaienti, S. (2018). Correlation dimension

- and phase space contraction via extreme value theory. *Chaos*, 28(4):041103.
- ²⁸Ferreira, M. (2018). Heuristic tools for the estimation of the extremal index: a comparison of methods. *RevStat*, 16:115–136.
- ²⁹Ferro, C. A. T. and Segers, J. (2003). Inference for clusters of extreme values. *J. R. Statist. Soc. B*, 65:545–556.
- ³⁰Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample. *Proc. Cambridge Philos. Soc.*, 24:180–190.
- ³¹Franzke, C. L. E. (2017). Extremes in dynamic-stochastic systems. *Chaos*, 27:012101.
- ³²Franzke, C. L. E., Osprey, S. M., Davini, P., and Watkins, N. W. (2015). A Dynamical Systems Explanation of the Hurst Effect and Atmospheric Low-Frequency Variability. *Sci. Rep.*, 5:9068.
- ³³Freitas, A. C. M., Freitas, J. M., and Todd, M. (2010). Hitting time statistics and extreme value theory. *Probab. Theory Relat. Fields*, 147(3):675–710.
- ³⁴Freitas, A. C. M., Freitas, J. M., and Todd, M. (2012). The extremal index, hitting time statistics and periodicity. *Adv. Math.*, 231(5):2626–2665.
- ³⁵Freitas, J. M. (2013). Extremal behaviour of chaotic dynamics. *Dynam. Syst.*, 28:302–332.
- ³⁶Ghil, M., Yiou, P., Hallegatte, S., Malamud, B. D., Naveau, P., Soloviev, A., Friederichs, P., Keilis-Borok, V., Kondrashov, D., Kossobokov, V., Mestre, O., Nicolis, C., Rust, H. W., Shebalin, P., Vrac, M., Witt, A., and Zaliapin, I. (2011). Extreme events: dynamics, statistics and prediction. *Nonlinear Proc. Geophys.*, 18(3):295–350.
- ³⁷Giamalaki, K., Beaulieu, C., Faranda, D., Henson, S., Josey, S., and Martin, A. (2018). Signatures of the 1976-77 regime shift in the North Pacific revealed by statistical analysis. *J. Geophys. Res.-Oceans*.
- ³⁸Gnedenko, B. V. (1943). Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. Math.*, 44:423–453.
- ³⁹Isham, V. (1980). Dependent thinning of point processes. *J. Appl. Prob.*, 17:987–995.
- ⁴⁰Kallenberg, O. (2002). *Foundations of Modern Probability*. Springer-Verlag, New York Berlin Heidelberg, 2nd edition.
- ⁴¹Katz, R. W., Parlange, M. B., and Naveau, P. (2002). Statistics of extremes in hydrology. *Adv. Water Resour.*, 25:1287–1304.
- ⁴²Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag, New York, Heidelberg, Berlin.
- ⁴³Lucarini, V., Faranda, D., de Freitas, J. M. M., Holland, M., Kuna, T., Nicol, M., Todd, M., Vaienti, S., et al. (2016). *Extremes and Recurrence in Dynamical Systems*. John Wiley & Sons.
- ⁴⁴Lucarini, V., Faranda, D., Turchetti, G., and Vaienti, S. (2012a). Extreme value theory for singular measures. *Chaos*, 22(2):023135.
- ⁴⁵Lucarini, V., Faranda, D., and Wouters, J. (2012b). Universal behaviour of extreme value statistics for selected observables of dynamical systems. *J. Stat. Phys.*, 147(1):63–73.
- ⁴⁶Malevergne, Y., Pisarenko, V., and Sornette, D. (2006). On the power of generalized extreme value (GEV) and generalized Pareto distribution (GPD) estimators for empirical distributions of stock returns. *Appl. Fin. Econ.*, 16(3):271–289.
- ⁴⁷McNeil, A. J., Frey, R., and Embrechts, P. (2015). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.
- ⁴⁸Pisarenko, V. and Sornette, D. (2003). Characterization of the frequency of extreme earthquake events by the generalized pareto distribution. *Pure Appl. Geophys.*, 160(12):2343–2364.
- ⁴⁹Rényi, A. (1956). A characterization of Poisson processes. *Magyar Tud. Akad. Mat. Kutató Int. Közl.*, 1:519–527.
- ⁵⁰Santhanam, M. S. and Kantz, H. (2008). Return interval distribution of extreme events and long-term memory. *Phys. Rev. E*, 78:051113.
- ⁵¹Smith, A. and Katz, R. (2013). US billion-dollar weather and climate disasters: data sources, trends, accuracy and biases. *Nat. Hazards*, 67:387–410.
- ⁵²Smith, R. L. and Weissman, I. (1994). Estimating the extremal index. *J. R. Statist. Soc. B*, 56:515–528.
- ⁵³Smolka, A. (2006). Natural disasters and the challenge of extreme events: risk management from an insurance perspective. *Philos. T. Roy. Soc. A*, 364(1845):2147–2165.
- ⁵⁴Süveges, M. (2007). Likelihood estimation of the extremal index. *Extremes*, 10:41–55.
- ⁵⁵Vercauteren, N., Boyko, V., Faranda, D., and Stiperski, I. (2018). Scale interactions and anisotropy in stable boundary layers. *arXiv preprint arXiv:1809.07031*.
- ⁵⁶Westcott, M. (1976). Simple proof of a result on thinned point processes. *Ann. Prob.*, 4:89–90.
- ⁵⁷Young, L.-S. (1982). Dimension, entropy and Lyapunov exponents. *Ergod. Theory Dyn. Syst.*, 2(1):109–124.