

Peaks over Manifold (POM): A Novel Technique to Analyze Extreme Events over Surfaces

Gonzalo Perera, Angel M. Segura 

CURE, Rocha, Uruguay
Email: gperera@cure.edu.uy

How to cite this paper: Perera, G. and Segura, A.M. (2022) Peaks over Manifold (POM): A Novel Technique to Analyze Extreme Events over Surfaces. *Advances in Pure Mathematics*, 12, 48-62.
<https://doi.org/10.4236/apm.2022.121004>

Received: December 14, 2021

Accepted: January 24, 2022

Published: January 27, 2022

Copyright © 2022 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

We present a novel method to analyze extreme events of flows over manifolds called Peaks Over Manifold (POM). Here we show that under general and realistic hypotheses, the distribution of affectation measures converges to a Generalized Pareto Distribution (GPD). The method is applicable to floods, ice cover extent, extreme rainfall or marine heatwaves. We present an application to a synthetic data set on tide height and to real ice cover data in Antarctica.

Keywords

Peaks Over Thresholds (POT), Generalized Pareto Distribution, Extreme Events

1. Introduction

Extreme events such as Tsunamis or floods are crucial for the impact they have in societies and ecosystems human developments as financial markets, industrial methods of production, telecommunication systems, and, at last but not least, on natural phenomena and its relationship with antropogenic factors [1]-[9]. The appropriate characterization of the impact of these events requires novel theoretical developments. Techniques such as the peak over threshold (POT) for univariate time series, which describe the distribution of the excess of a time series with respect to a given (*large*) threshold are well known from decades ago [10] [11] [12]. The change from univariate to n-dimensional data implies substantial additional complexity and in particular an unsolved problem is to characterize the distribution of excesses of the progressive development of mass, physical, biological or chemical properties on affected areas (tides, floods, ice cover, etc.) over surfaces or manifolds. We will call *flow* such kind of phenomena. Here, we

derive a theoretical model on the distribution of extremes of flows over a surface or manifold. We also test the proposed model using synthetic data on tide heights and a real data set containing time series of ice cover over Antarctica.

The paper is organized as follows. First we will expose the framework providing a precise model for flow data. We shall always consider the case of flows over surfaces, but with no particular effort, everything can be extended to Riemannian manifolds. Then, we will introduce two preliminary indexes of impact of a set of data flow, in Theorems 1 and 2, helping to arrive to the main result of this paper, Theorem 3. In our main result, we consider an impact measure, whose density may be chosen according to subjective valuation of risks. We then compute this impact of the flow expansion over the region with an (intrinsic) distance greater than u to the base line of the flow. In terms of tides, we measure affectation of areas at a distance greater than u meters from the tides base line. We then show that, for u tending to infinity. The distribution of this impact conditioned to the fact that the impact is positive, converges to a Generalized Pareto Distribution. We finally show the application of this result to simulated data on tide height and beach cover and to a real data set of ice cover in Antarctica.

2. POM Derivation

Let S be a smooth (*differentiable*) surface with intrinsic distance d and intrinsic area measure σ . Let us assume that S is simply connected and that C is a finite-length differentiable curve. We will say that C *splits* S in the following sense:

1) C defines two disjoint sets, namely C_{int} and C_{ext} .

2) C_{ext} is defined at each point x of C by $n(x)$, the unitary vector on the tangent plane of S at x that is perpendicular to the tangent vector $v(x)$ of the curve C at x , $n(x)$ taking the anti-clockwise direction (**Figure 1**). Notice that $S = C_{int} \cup C \cup C_{ext}$ and recall that C is not included neither in C_{int} nor C_{ext} .

Given two differentiable curves \tilde{C} and C^* that splits S , we will say that they define a *coherent splitting* if any of the following conditions is matched (**Figure 2**):

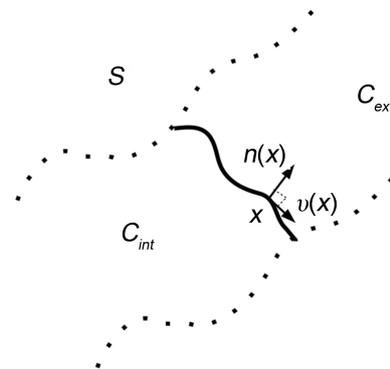


Figure 1. Schematic representation of a surface S and a curve C that splits the space into two regions. The external region (C_{ext}) and an internal region C_{int} . At each point (x) of curve C the tangent vector ($v(x)$) and a perpendicular vector ($n(x)$) defines the regions.

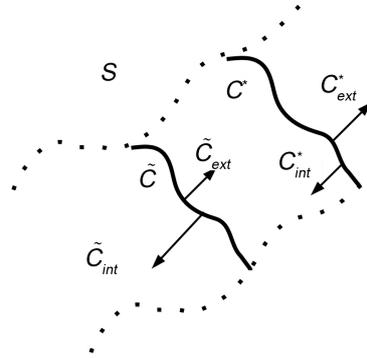


Figure 2. Schematic example of a coherent splitting of the space.

$$1) \tilde{C}_{ext} \subset C^*_{ext}, C^*_{int} \subset \tilde{C}_{int}$$

$$2) \tilde{C}_{int} \subset C^*_{int}, C^*_{ext} \subset \tilde{C}_{ext}$$

Given any $r > 0$, we define

$$C(r) = \{x \in C_{ext} : d(x, C) \leq r\}$$

We then have that

$$\partial C(r) = C \cup C_r$$

where $C_r = \{x \in C_{ext} : d(x, C) = r\}$ that we assume to be a differentiable curve.

We will assume that C, C_r define a coherent splitting (**Figure 3**).

We will now consider a function a that quantifies, in terms of density, the impact of the flow at each point. More precisely, we will assume:

$$i) a(x) \geq 0 \quad \forall x \in S$$

$$ii) a(x) > 0 \quad \forall x \in C_{ext}$$

$$iii) a(x) = 0 \quad \forall x \in C_{int}$$

$$iv) a \text{ is continuous in } C_{ext} \cup C$$

$$v) \int_{C_{ext}} a(x) d\sigma(x) = +\infty$$

$$vi) \int_{C(r)} a(x) d\sigma(x) \text{ is finite } \forall r > 0$$

Let us explain this in intuitive terms. C is the baseline for the flow. C_{int} represents the “normal status”, where there is no danger or impact. C_{ext} is the region exposed to risk by the flow. Given any measurable set A on S :

$$\mu(A) = \int_A a(x) d\sigma(x)$$

quantifies the impact produced by the flow over the region A .

Thus, condition (v) above means that if the flow covers C_{ext} there is a total disaster where condition (vi) implies that if the flow covers $C(r), r > 0$ there is a serious event but not a total disaster.

Finally, it is straightforward to check that if

$$I(r) = \mu(C_r) = \int_{C(r)} a(x) d\sigma(x)$$

then, $I(r)$ is strictly increasing with r and that

$$\lim_{r \rightarrow \infty} I(r) = \int_{C_{ext}} a(x) d\sigma(x) = +\infty$$

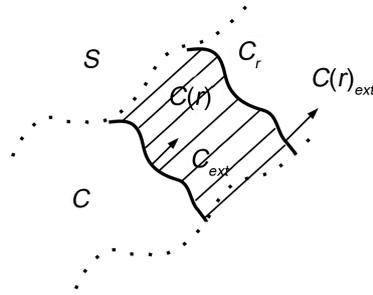


Figure 3. Coherent splitting of $C, C(r)$.

Remark 1: The density is subjective as it should be, since different criteria may be applied to different contexts or by different agents. For instance, if a is constant and equals 1, the measure of affectation is nothing but the affected area. But if we are considering tides on a coast with human inhabitants, and $d(x)$ denotes the population density at point x , if $a = d$, the affectation measure is no longer homogeneous on the coast and it will give greater values as more human beings are affected by tides. Similarly, if we have a measure of the relevance for ecosystem conservation at each point x , $c(x)$, due to relevant natural areas that may be seriously affected by marine water, taking a linear combination of d and c we will try to protect human populations and natural areas as well, with relative weights depending on the coefficients of the linear combination.

Consider now C_1, \dots, C_n *iid* differentiable curves on S , such that:

- i) $C_i \in C_{ext} \quad \forall i$
- ii) C, C_i define a coherent splitting $\forall i$

We will think C_1, \dots, C_n as n empirical realizations of the border of the flow.

For any $u > 0$, let us define $O_i(u) = C_{i_{int}} \cap C(u)_{ext}$ (**Figure 4**).

Based on the previous concepts and as a first auxiliary tool, we are now able to define a first impact index X_i as:

$$X_i = \mu((C_i)_{int}) = \int_{(C_i)_{int}} a(x) d\sigma(x) \quad (1)$$

Let us define

$$D_i = \max(d(x, C) : x \in C_i) \quad (2)$$

which is assumed to be a finite random variable.

In this first and auxiliary step, we will be interested in the asymptotic behaviour of the tail probability:

$$P(X_i > t + I(u) / D_i > u) \quad \forall t \geq 0 \quad (3)$$

The question we are interested here is: If the flow exceeds C_u , what is the probability that the impact exceed by t the impact of total covering of $C(u)$?

Observe that:

$$P(X_i > t + I(u) / D_i > u) = P(X_i > t + I(u), D_i > u) / P(D_i > u) \quad (4)$$

But if $D_i \leq u$ it is easy to check that $\sigma(O_i(u)) = 0$, therefore $X_i \leq I(u)$

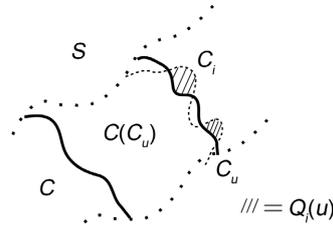


Figure 4. Affected region exceeding the curve C_u defined by C_i .

and $X_i > t + I(u)$ does not hold. Thus we conclude that:

$$\{X_i > t + I(u)\} \subset \{D_i > u\} \tag{5}$$

and that

$$P(X_i > t + I(u), D_i > u) = P(X_i > t + I(u)) \tag{6}$$

taking into account that

$$P(X_i > t + I(u)) = P(X_i > t + I(u) / X_i > I(u)) \times P(X_i > I(u)) \quad \forall t \geq 0 \tag{7}$$

and since as u goes to infinity, $I(u)$ increases to infinity (recall that $I(u)$ is deterministic), we conclude that if we assume:

(H₁): X_i belongs to the domain of attraction of an extremal distribution, then the Pickands-Balkema-De Haan Theorem [10] [11] applies to X_i and

$$\lim_{u \rightarrow \infty} P(X_i > t + I(u) / X_i > I(u)) = Q(t) \tag{8}$$

with Q the tail of a **Generalized Pareto Distribution (GPD)**.

Coming back to (4), (5), (6), (7) and (8) we finally find out that if we assume:

(H₂): There exists a positive constant M such that:

$$\lim_{u \rightarrow \infty} P(X_i > I(u)) / P(D_i > u) = M \tag{9}$$

then,

$$\lim_{u \rightarrow \infty} P(X_i > t + I(u) / D_i > u) = MQ(t) \tag{10}$$

where Q is the tail of a **GPD**. We then have the following first result:

Theorem 1:

Under H_1 and H_2 , as u goes to infinity, $P(X_i \leq t + I(u) / D_i > u)$ approaches to $1 - MQ(t)$ where Q is the tail of a **(GPD)**. \diamond

3. Example 1

Let $S = \{(x, y) \in \mathbb{R}^2; 0 \leq y \leq 1\}$ and let C be a curve of length L splitting S . Take σ the Lebesgue measure in the plane and

$$a(x) = \begin{cases} 1 & \forall x \in C_{ext} \cup C \\ 0 & \forall x \in C_{int} \end{cases}$$

Then $I(u) = \sigma(C(u)) = Lu$. If $\alpha_i(t), t \in [0, 1]$ is a parametrization of C_i ,

$$X_i = \int_{(C_i)_{int}} 1 dx dy = \int_0^1 \left(\int_0^{\alpha_i(t)} dy \right) dt = \int_0^1 \alpha_i(t) dt \tag{11}$$

and

$$D_i = \max_{t \in [0,1]} \alpha_i(t)$$

then H_1 reads: $X_i = \int_0^1 \alpha_i(t) dt$ belongs to an extremal domain of attraction. On

the other hand, H_2 becomes: $\lim_{u \rightarrow \infty} \frac{P(X_i > Lu)}{P(D_i > u)} = M$ where M is a positive constant.

Since $X_i \leq D_i$, then $P(X_i > Lu) \leq P(D_i \geq Lu)$. If we assume that α_i has derivative bounded by K , $|\alpha_i(t) - D_i| \leq Kt$ and $X_i \geq D_i - K \int_0^1 t dt = D_i - \frac{K}{2}$ and hence

$$P(X_i > Lu) \geq P\left(D_i > Lu + \frac{K}{2}\right) \quad (12)$$

Therefore, H_2 is satisfied if for any positive constants $\mathbb{C}_1, \mathbb{C}_2$, there exists a positive constant M such that the following holds true:

$$\lim_{u \rightarrow \infty} \frac{P(D_i > \mathbb{C}_1 u)}{P(D_i > u)} = \frac{P(D_i > \mathbb{C}_1 u + \mathbb{C}_2)}{P(D_i > u)} = M > 0 \quad (13)$$

For instance, if for t tending to infinity $P(D_i > t) \approx \frac{1}{t^\beta}, \beta > 1$.

$$\lim_{u \rightarrow \infty} \frac{P(D_i > \mathbb{C}_1 u)}{P(D_i > u)} = \lim_{u \rightarrow \infty} \frac{u^\beta}{\mathbb{C}_1^\beta u^\beta} = \frac{1}{\mathbb{C}_1^\beta} \quad (14)$$

and

$$\lim_{u \rightarrow \infty} \frac{P(D_i > \mathbb{C}_1 u + \mathbb{C}_2)}{P(D_i > u)} = \lim_{u \rightarrow \infty} \frac{1}{\left(\frac{\mathbb{C}_1 u + \mathbb{C}_2}{u}\right)^\beta} = \frac{1}{\mathbb{C}_1^\beta} \quad (15)$$

and H_2 is fulfilled. \diamond

We consider now our main index of impact $\mu(O_i(u))$. But we will yet provide an intermediate result concerning the asymptotic behaviour of

$$P(\mu(O_i(u)) > t / D_i > u) \quad \forall t > 0 \quad (16)$$

for u tending to infinity.

Observe that $\{D_i > u\} = \{\sigma(O_i(u)) > 0\} = \{\mu(O_i(u)) > 0\}$, and thus, (16) may be rewritten as

$$P(\mu(O_i(u)) > t / \mu(O_i(u)) > 0) \quad \forall t > 0 \quad (17)$$

Before going to (16), we will discuss the asymptotic behaviour, for u tending to infinity, of:

$$P(\mu(O_i(u)) > t / d_i > u) \quad \forall t > 0 \quad (18)$$

where

$$d_i = \min \{d(x, C) : x \in C_i\} \leq D_i \quad (19)$$

We will assume:

H₃: $d_i > 0$ a.s.

If $d_i > u$ it is then clear that

$$C(u) \subset C(d_i) \subset (C_i)_{int} \cap C(u)_{ext} \subset C(D_i) \cap C(u)_{ext} \tag{20}$$

Taking into account (19) and (20) and the basic properties of measures we get,

$$\begin{aligned} & \mu(C(d_i)) - \mu(C(u)) \\ &= \mu(C(d_i) \cap C(u)_{ext}) \leq \mu(O_i(u)) \\ &\leq \mu(C(D_i) \cap C(u)^c) = \mu(C(D_i)) - \mu(C(u)) \end{aligned} \tag{21}$$

Therefore,

$$\begin{aligned} & P(\mu(C(d_i)) - \mu(C(u)) > t/d_i > u) \\ &\leq P(\mu(O_i(u)) > t/d_i > u) \leq P(\mu(C(D_i)) - \mu(C(u)) > t/d_i > u) \end{aligned} \tag{22}$$

Hence

$$\begin{aligned} & P(\mu(C(d_i)) > t + \mu(C(u))/d_i > u) \\ &\leq P(\mu(O_i(u)) > t/d_i > u) \leq P(\mu(C(D_i)) > t + \mu(C(u))/d_i > u) \end{aligned} \tag{23}$$

Lets us focus first on the left hand term of (23). Taking into account that:

$$\{d_i > u\} = \{C(u) \subset C(d_i)\} = \{\mu(C(u)) < \mu(C(d_i))\}, \tag{24}$$

that term may be written as

$$\begin{aligned} & P(\mu(C(d_i)) > t + \mu(C(u))/d_i > u) \\ &= P(\mu(C(d_i)) > t + \mu(C(u)) / \mu(C(d_i)) > \mu(C(u))) \end{aligned} \tag{25}$$

Therefore, if we assume:

(H4) $\mu(C(d_i))$ belongs to an extremal domain of attraction, then by the Theorem of Picklands-Balkema-DeHaan, since $\lim_{u \rightarrow +\infty} \mu(C(u)) = +\infty$, we have that

$$\lim_{u \rightarrow \infty} P(\mu(C(d_i)) > t + \mu(C(u)) / \mu(C(d_i)) > \mu(C(u))) = Q(t), \tag{26}$$

where $Q(t)$ is the tail of a **GPD**. Thus the left-hand term of (23) satisfies that:

$$\lim_{u \rightarrow \infty} P(\mu(C(d_i)) > t + \mu(C(u))/d_i > u) = Q(t) \tag{27}$$

with $Q(t)$ the tail of a **GPD**.

Let us now consider the right-hand term of (23):

$$\begin{aligned} & P(\mu(C(D_i)) > t + C(u)/d_i > u) \\ &= \frac{P(\mu(C(D_i)) > t + C(u), d_i > u)}{P(d_i > u)} \\ &= \frac{P(\mu(C(D_i)) > t + C(u), D_i > u) P(D_i > u)}{P(D_i > u) P(d_i > u)} \\ &= \frac{P(\mu(C(D_i)) > t + C(u), D_i > u \geq d_i)}{P(d_i > u)} \\ &= (I) - (II) \end{aligned} \tag{28}$$

Assume that:

$$(H_5) \lim_{u \rightarrow +\infty} P(D_i > u) / P(d_i > u) = 1$$

This implies that:

$$\mu(C(D_i)) \text{ and } \mu(C(d_i)) \quad (29)$$

belong to the same extremal domain of attraction.

Therefore, by (28), H_5 , (29) and Picklands-Balkeman-De Haan Theorem we get:

$$\begin{aligned} \lim(I) &= \lim_{u \rightarrow \infty} P(\mu(C(D_i)) > t + \mu(C(u)) / \mu(C(D_i)) > \mu(C(u))) \\ &\times \lim_{u \rightarrow \infty} \frac{P(D_i > u)}{P(d_i > u)} = Q(t) \times 1 = Q(t) \end{aligned} \quad (30)$$

where Q is the same GPD as in (27).

On the other hand,

$$\begin{aligned} 0 \leq (II) &\leq \frac{P(D_i > u > d_i)}{P(d_i > u)} = \frac{P(D_i > u)}{P(d_i > u)} - \frac{P(D_i > u, d_i > u)}{P(d_i > u)} \\ &= \frac{P(D_i > u)}{P(d_i > u)} - \frac{P(d_i > u)}{P(d_i > u)} = \frac{P(D_i > u)}{P(d_i > u)} - 1 \end{aligned} \quad (31)$$

and it turns out from H_5 that

$$\lim_{u \rightarrow \infty} (II) = 0 \quad (32)$$

therefore, by (28), (29) and (32), we have

$$\lim_{u \rightarrow \infty} P(\mu(C(D_i)) > t + \mu(C(u)) / d_i > u) = Q(t) \quad (33)$$

Then, adding (33) to (23) and result (27), we obtain the following result.

Theorem 2

Under H_3 , H_4 and H_5 ,

$$\lim_{u \rightarrow \infty} P(\mu(O_i(u)) > t / d_i > u) = Q(t) \quad (34)$$

with Q the tail of a GPD. \diamond

Now let us come back to (16). We have that:

$$\begin{aligned} &P(\mu(O_i(u)) > t / \mu(O_i(u)) > 0) \\ &= P(\mu(O_i(u)) > t / D_i > u) = \frac{P(\mu(O_i(u)) > t, D_i > u)}{P(D_i > u)} \\ &= \frac{P(\mu(O_i(u)) > t, D_i > u, d_i > u)}{P(d_i > u)} \frac{P(d_i > u)}{P(D_i > u)} \\ &\quad + \frac{P(\mu(O_i(u)) > t, D_i > u \geq d_i)}{P(d_i > u)} \frac{P(d_i > u)}{P(D_i > u)} \\ &= (III) + (IV) \end{aligned} \quad (35)$$

But,

$$\begin{aligned}
 (III) &= \frac{P(\mu(O_i(u)) > t, D_i > u, d_i > u)}{P(d_i > u)} \frac{P(d_i > u)}{P(D_i > u)} \\
 &= P(\mu(O_i(u)) > t / d_i > u) \frac{P(d_i > u)}{P(D_i > u)}
 \end{aligned}
 \tag{36}$$

Then using **Theorem 2**, and H_5 we get

$$\lim_{u \rightarrow \infty} (III) = Q(t) \times 1 = Q(t)
 \tag{37}$$

On the other hand,

$$0 \leq (IV) \leq \frac{P(D_i > u \geq d_i)}{P(d_i > u)} \frac{P(d_i > u)}{P(D_i > u)}
 \tag{38}$$

By (30) and (H_5), the right-hand term on (38), as u tends to infinity, tends to $0 \times 1 = 0$ and therefore:

$$\lim_{u \rightarrow \infty} (IV) = 0
 \tag{39}$$

and we then have the main result of this paper:

Theorem 3

Under H_3, H_4 and H_5

$$\lim_{u \rightarrow \infty} P(\mu(O_i(u)) > t / \mu(O_i(u)) > 0) = Q(t) \quad \forall t > 0
 \tag{40}$$

with $Q(t)$ the tail of a GPD. \diamond

Remark 2: This result is a clear extension of the POT method, since it gives the conditional distribution of the measure of affectation given that the area $O_i(u)$ is reached by the flow. We will derive a statistical application of this Theorem that we defined as Peak Over Manifold (**POM**).

3.1. Example 1, Synthetic Data

In the first example, we simulate a data set on tide height along 100 km of beach monitored for three months under different tidal conditions. Then we evaluate the probability of the sea exceeding a fixed threshold distance (u) over the sand. The threshold was set as the height exceeding 20% the maximum average tide ($u = 3.5$ m). We estimated the tidal height using the function *waves* (<https://beckmw.wordpress.com/2017/04/12/predicting-tides-in-r/>) that models a sinusoidal function with a defined period (T in hours) and amplitude (A in meters). We chose to combine functions with three different tidal configurations, $T = [24, 48, 12]$ and $A = [1, 0.5, 2]$ respectively. This raises a cyclic pattern (**Figure 5**) of average tide (tide). For each time, we simulated distance from shoreline over the beach for each of the 100 kilometer along the beach ($s = 1, \dots, 100$ in km). The distance from the shore (d in m) was simulated by combining a sinusoidal function with variable spatial amplitude (A_s in meters) and fixed spatial period ($T_s = 20$ in km) with iid with the noise ($e \approx \text{Normal}(0, 0.5)$) as follows:

$$d(s) = \overline{\text{tide}} + A_s \sin\left(\frac{2\pi}{T_s} s\right) + e$$

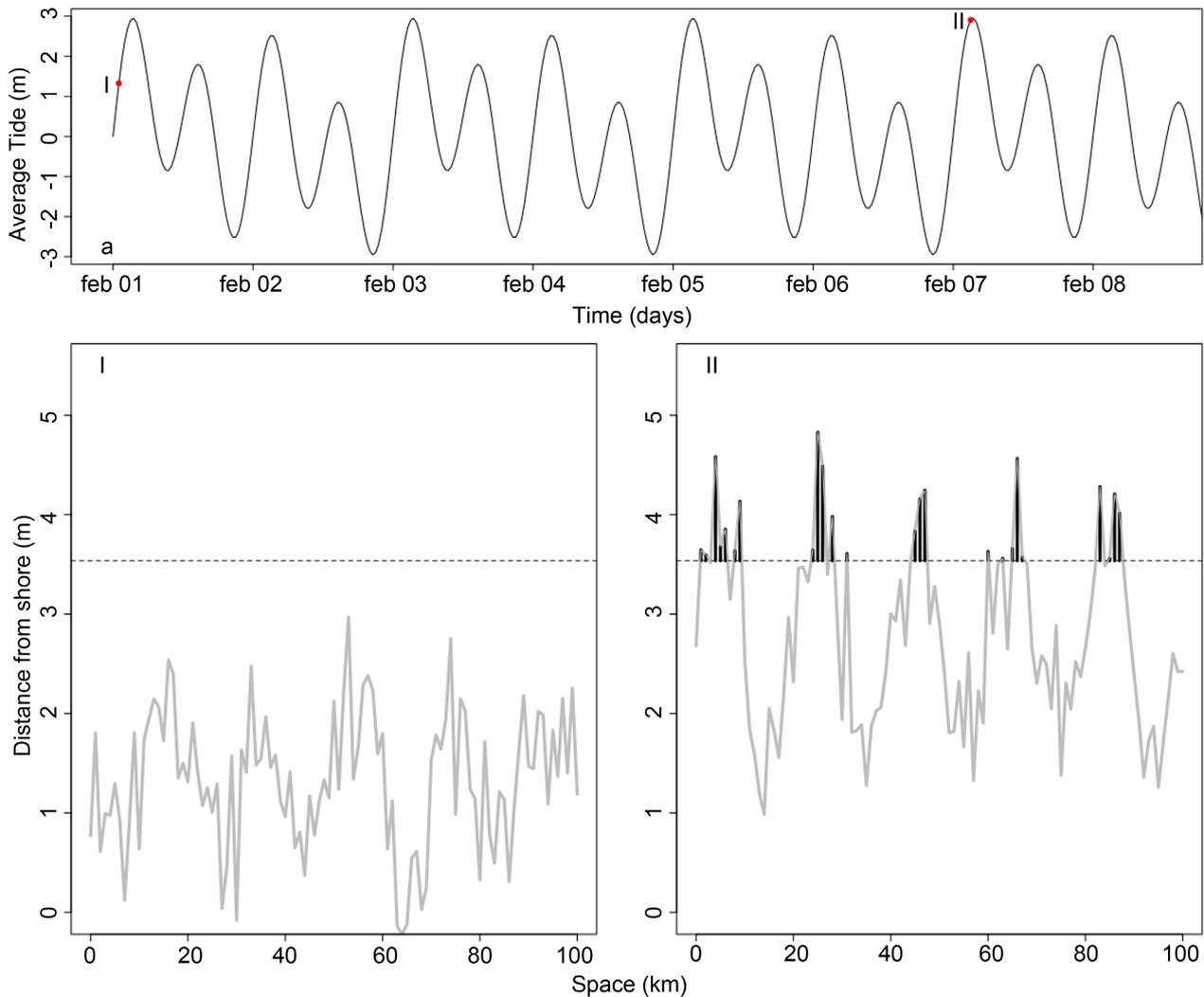


Figure 5. Temporal trend in average tidal height in a week. Red dots indicate examples of points where detailed spatial patterns in tidal height is shown. I) An example where tide did not exceed the threshold (u) and II) an example where tide exceeds the predefined threshold. The dashed area is the observed excess ($O_i(u)$).

We registered dates when, at any point, the distance from the shore exceeded threshold ($d > u$). For those dates, the area exceeding the threshold ($O_i(u)$) was calculated (dashed vertical lines in **Figure 5**, II). Excess ($d > u$) were registered in 16% ($N = 685$) of the sampled events. We assumed affectation was independent of the point of the beach reached by high tide, and thus the affected area was used as indicator of effect. We fitted by maximum likelihood estimation a GPD using the *fitGPD* function from the *{POT}* package in **R** to affected areas larger than 1000 m^2 . The distribution of areas exceeding the threshold ($O_i(u)$) showed a GPD as expected on theoretical grounds (**Figure 6**).

Scale parameter (average [95% CI]; $\sigma = 0.007[0.006, 0.009]$) and a shape parameter (average [95% CI]; $\xi = 1 \times 10^{-13}[-0.14, 0.11]$) suggested an Exponential distribution (**Figure 6**). Notice the confidence interval of the shape parameter overlapping zero is useful to detect exponential distribution.

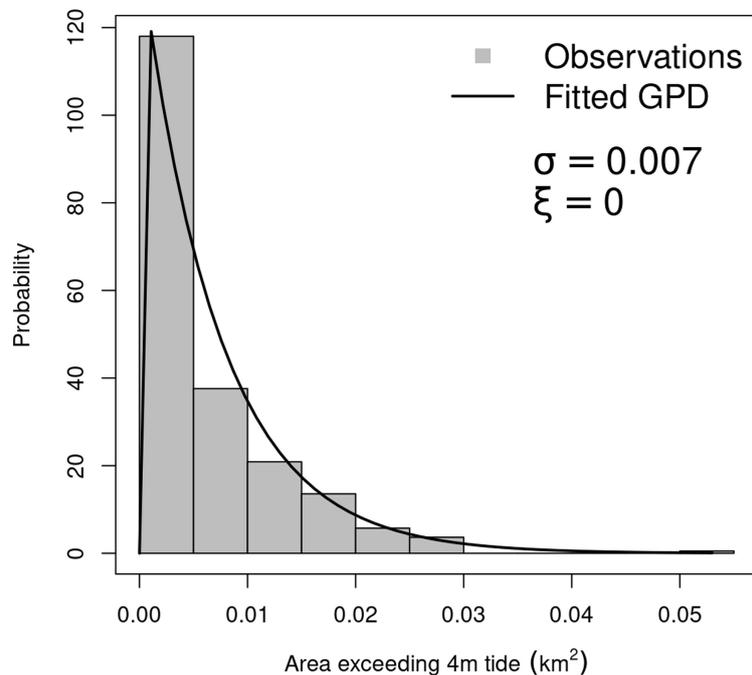


Figure 6. Simulated area exceeding a threshold ($O_i(u)$) and fitted Generalized Pareto Distribution (GPD) fitted. The values of the fitted parameters are also shown.

3.2. Example 2: Ice Cover in Antarctica

We explored ice cover data in Antarctica using reanalysis data from the Environmental Modeling Center of the National Centers for Environmental Prediction (NCEP) from the National Oceanic and Atmospheric Administration (NOAA) using the second version [13]. Data on monthly sea ice concentration (Figure 7) was downloaded from

http://iridl.ldeo.columbia.edu/SOURCES/NOAA/NCEP/EMC/CMB/GLOBAL/Reyn_SmithOIv2/monthly/T/722.5/VALUE/X/150.0/360.0/RANGEEDGES/Y/-90/-60/RANGEEDGES/sea_ice/index.html.

Sea ice concentration data were obtained from -89.5 to -60.5 degrees South of latitude and -149.5 to 0.5 degrees West in Longitude for cell of 1×1 degree. Monthly data consisted in 461 months starting in October 1981 to February 2020. Sea ice concentration was expressed as percentage (0 - 100%) of cell area.

We calculated the average monthly ice cover for the whole period (*i.e.* climatology, Figure 8) and defined a threshold ($u < 10\%$). The cells with less than 10% of ice cover were selected to evaluate the dynamic of ice cover in the extreme regions. Then, for each month we explored whether the fraction of ice cover exceeded the climatological threshold (u) previously estimated. For the months in which the percentage of ice cover exceeded 10%, we recorded the area of ice cover exceeding threshold. The area of 1×1 change with latitude. Thus we used the function *area* from {raster} package to convert from percentage of ice cover to area of the ice cover (in km^2) for each cell. Despite this function is not

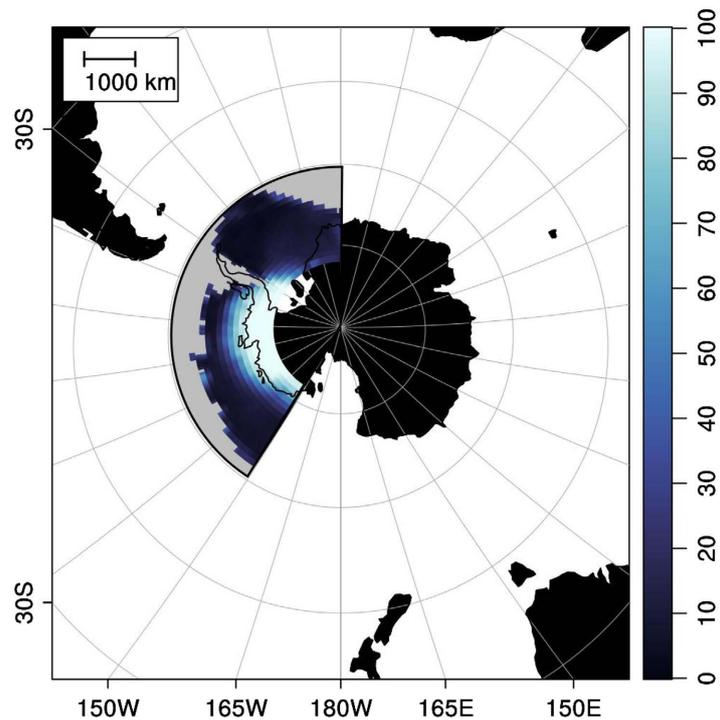


Figure 7. Ice cover (in %) in the studied region of the Antarctica.

precise for high latitudes, it represent a first order approximation. Moreover, most of the points were located mostly in the lower latitudinal regions and thus this does not introduce a relevant bias.

For each month were some ice cover exceeds climatological ice cover, we estimated the area of the ice cover exceeding the threshold and stored in a separate vector. We fitted by maximum likelihood estimation a GPD to the area exceeding threshold using the *fitgpd* function from the $\{POT\}$ package in **R** to areas larger than 1000 km^2 . We found a good fit to a GPD (**Figure 9**) with a positive shape parameter (average [95% CI] $\xi = 0.158[0.07, 0.26]$).

4. Discussion & Conclusions

This work introduces a novel technique to evaluate extreme processes over a surface or manifold. The reasoning expands previous univariate approaches to the topic (POT) by demonstrating under realistic hypothesis that the resultant distribution of the area exceeded by a given phenomena follows a Generalized Pareto Distribution (GPD). The applicability of this novel approach is demonstrated under simulated data on tidal height and ice cover in a relevant region in the Antarctica. Results confirm the application of a GPD to explore problems of extreme over surfaces. This opens a new avenue in the analisis of extreme flows which can be applied in ice cover in the poles [14], to areas covered by flood, the incidence of heatwaves among other examples. Under resonable conditions, the distribution converges to a GDP, but exploring if other distribution can emerge is interesting and also to compare the method using

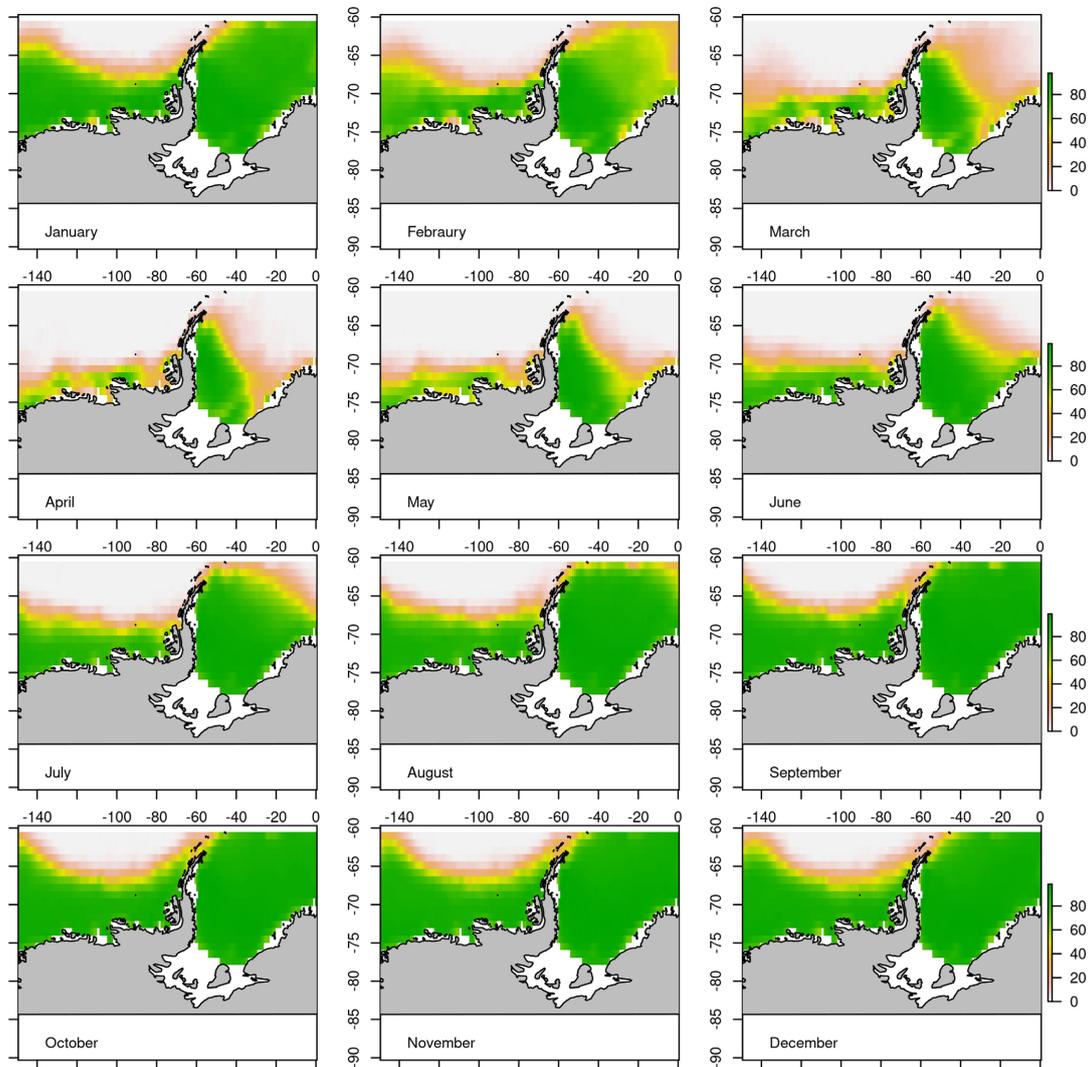


Figure 8. Monthly climatological values of ice cover extent in Antarctica.

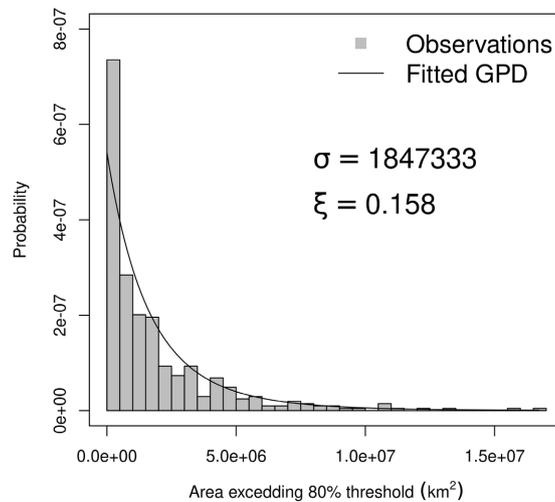


Figure 9. Area covered by ice in the extreme range of ice cover. Superimposed is the Generalized Pareto Distribution fitted to data. The value of the shape (σ) and scale parameters (ξ) are shown.

more data-sets, as has been done in recent articles [15]. In real applications, the evaluation over surfaces (2D) will be common, but the theorem allows to expand the results to a more general N-dimensional manifold. This avenue has to be further explored.

Acknowledgements

Sincere thanks to the participants of the “Jornadas de Estadística Aplicada 2019” in La Paloma-Uruguay, for suggestions on a previous version of this results. This is a contribution of the project entitled “Mecanismos determinantes de la estructura comunitaria y la variabilidad espacio-temporal en comunidades acuáticas” founded by ANII-Uruguay (FCE_3_2020_1_162710).

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Batt, T.D., Carpenter, S.R. and Ives, A.R. (2017) Extreme Events in Lake Ecosystem Time Series. *Limnology and Oceanography Letters*, **2**, 63-69. <https://doi.org/10.1002/lol2.10037>
- [2] Bellanger, L. and Perera, G. (2003) Compound Poisson Limit Theorems for High-Level Exceedances of Some Non-Stationary Processes. *Bernoulli*, **9**, 497-515. <https://doi.org/10.3150/bj/1065444815>
- [3] Bousquet, N. and Bernardara, P. (2021) Extreme Value Theory with Applications to Natural Hazards: From Statistical Theory to Industrial Practice. Springer, Berlin.
- [4] Hannesdóttir, Á., Kelly, M.C. and Dimitrov, N.K. (2019) Extreme Wind Fluctuations: Joint Statistics, Extreme Turbulence, and Impact on Wind Turbine Loads. *Wind Energy Science*, **4**, 325-342. <https://doi.org/10.5194/wes-4-325-2019>
- [5] Jiménez, E., Cabañas, B. and Lefebvre, G. (2015) Environment, Energy and Climate Change I: Environmental Chemistry of Pollutants and Wastes. Springer, Berlin. <https://doi.org/10.1007/978-3-319-12907-5>
- [6] Pechiar, J., Perera, G. and Simon, M. (2002) Effective Bandwidth Estimation and Testing for Markov Sources. *Wind Energy Science*, **48**, 157-175. [https://doi.org/10.1016/S0166-5316\(02\)00035-4](https://doi.org/10.1016/S0166-5316(02)00035-4)
- [7] Katz, R.W., Brush, G.S. and Parlange, G.S. (2005) Statistics of Extremes: Modeling Ecological Disturbances. *Ecology*, **86**, 1124-1134. <https://doi.org/10.1890/04-0606>
- [8] Reiss, R.-D. and Thomas, M. (2007) Statistical Analysis of Extreme Values: With Applications to Insurance, Finance, Hydrology and Other Fields. Birkhäuser, Basel.
- [9] Embrechts, P., Klüppelberg, C. and Mikosch, T. (2014) Modelling Extremal Events: For Insurance and Finance. Springer, Berlin.
- [10] Balkema, A.A. and de Haan, L. (1974) Residual Life Time at Great Age. *The Annals of Probability*, **5**, 13-15. <https://doi.org/10.1214/aop/1176996548>
- [11] Pickands, J. (1975) Statistical Inference Using Extreme Order Statistics. *The Annals of Statistics*, **3**, 119-131. <https://doi.org/10.1214/aos/1176343003>
- [12] Far, S.S. and Wahab, A.K.A. (2016) Evaluation of Peaks-over-Threshold Method.

Ocean Science Discussions, **47**, 1-25.

- [13] Reynolds, R.W., Rayner, N.A., Smith, T.M., Stokes, D.C. and Wang, W. (2002) An Improved *in Situ* and Satellite SST Analysis for Climate. *Journal of Climate*, **15**, 1609-1625. [https://doi.org/10.1175/1520-0442\(2002\)015<1609:AIISAS>2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015<1609:AIISAS>2.0.CO;2)
- [14] Segura, A., Crisci, A. and Perera, G. (in prep.) A Statistical Approach to Quantify the Impact of Melting Areas in Polar Regions under Different Greenhouse Emission Scenarios.
- [15] Arshad, M., Iqbal, M. and Ahmad, M. (2018) Transmuted Exponentiated Moment Pareto Distribution. *Open Journal of Statistics*, **8**, 939-961. <https://doi.org/10.4236/ojs.2018.86063>