MVA 2023 FINAL REPORT
EXTREME VALUE THEORY : WHEN REGULAR VARIATION AND PRINCIPAL COMPONENT ANALYSIS MEET AT INFINITY

GABRIEL SINGER
MATHIAS VIGOUROUX

Summary

We studied the article [4]. We illustrate our understanding of the theorems quoted with examples and counter-examples. We've also tried to do our best to make relevant remarks and to make links between the course and the notions encountered in [4]. Using the examples in the article 4, we construct processes with regular variation in $L^2[0,1]$ but not in $L^1[0,1]$ or $C[0,1]$. From a numerical point of view, we put into practice the characterization of regular variation of Theorem 1 by following the method proposed in [4], for synthetic data and meteorological data measured in Spain [1]. We have also plotted the eigenvectors associated with the extreme covariance operator for weather data 3.

1. Introduction

We live in a world where collecting data has become very simple. Let's take the example of temperatures in a city: with the number of measurements being taken, the representation of temperatures as a function of time looks more like a curve than a set of points. More generally, the nature of measurements is changing, with more and more measurements taking the form of random curves rather than random vectors. The data produced are of infinite dimensions and live in spaces of infinite dimensions.

The complexity of algorithms rises with increasing data, underscoring the fundamental role of data reduction. This concern is explored across various mathematical domains like algebraic geometry, analysis, and statistics. In scenarios like anomaly detection or network monitoring without known structures, dimensionality reduction emerges as a crucial initial step.

This is why functional approaches in extreme value theory are an active field. In [4] author considers Hilbert spaces. In extreme value theory, they have been much less studied than the continuous or semi-continuous spaces [3]. However, from a modeling point of view, L^p spaces for $p = 1, 2$ make sense in physics. For example, if we have two mass densities f, g and want to estimate the difference, we look at $||f - g||_{L^1}$ or if we consider some volume Ω and mass density $\rho : \Omega \mapsto \mathbb{R}_+$ the massic volume is $||\rho||_{L^1(\Omega)}$and for the norm L^2 we often find it associated with the energy of a physical quantity such as cinetic energy. Knowing that a physical measurement is very often the convolution of two quantities f and g if we had more time, we would have liked to study the impact of convolution on regular variation.

We place ourselves in the peak over threshold (POT) framework, i.e. we are interested in the distribution of a normalized observation knowing that its norm is greater than an arbitrarily large threshold. The standard hypothesis is the regular variation of the random element. This notion has been studied for function and for measures in a finite-dimensional context [2]. It has been extended to measures on complete, separable metric spaces, [3], and in what follows, the definition of a regular random variable comes from this work.

The contributions of the studied article [4] are multiple. Among other things, they demonstrate a characterization of regular variation in a Hilbert space and study the link between regular variation in $L^2[0,1]$ and in $C[0,1]$ with their usual norms.

They also extend Principal Component Analysis (PCA) for extremes to infinite dimension, which can be seen as an extension of [6]. The main result gives theoretical guarantees concerning the empirical subspaces associated with the covariance operator of an extreme.

We have chosen to illustrate the regular variation part with examples and counter-examples. As for the PCA part, we focused on a numerical illustration.

2. Regular Variation In infinite-dimensional case

Let $(E, ||\cdot||)$ be a separable Banach space and note $E_0 = E \setminus \{0\}$. Actually, in the studied paper [4], the authors consider a complete separable metric space (E, d), which requires the definition of continuous scalar multiplication and the existence of 0. Since in this report, we consider $E = C[0,1]$ or $E = L^2[0,1]$, which are Banach spaces, these two points are automatically verified.

A measure ν is regularly varying i.f.f. there exists a nonzero measure μ and a regularly varying function b such that $b(n)\nu(n\cdot) \to \mu(\cdot) \quad \text{as} \quad n \to \infty$. In finite-dimensional case, a convenient way to show that a measure μ converges to l is to show that $\{\mu_n\}_n$ is relatively compact and that the limits of two subsequences must coincide. In our course, this is done in Proposition 2.5.2 to prove the characterization of weak convergence in terms of Laplace transform and also in Proposition 3.1.7. By the theorem 4.2 of [3] we see that the regular variation is easier to prove in the finite-dimensional case than in function spaces. A characterization of regular variation in infinite dimensions is therefore welcome. This is done in the paper we are studying.

Given $(\Omega, A, \mathbb{P})$ a probability space, a random variable X with value in E is a Borel measurable application of $(\Omega, A, \mathbb{P})$ to E. X will be regularly varying in E if the measure $\mu(\cdot) = \mathbb{P}(X \in \cdot)$ is regularly varying in M_0.

Example: Let $(E, d) = (\mathbb{R}, |\cdot|)$, consider $X \sim \text{Pareto}(\alpha)$ for $\alpha > 0$. If $x > 0$, define $A_x :=]x, \infty[$, and $b : t \mapsto t^\alpha \in RV_\alpha$ it follows that $\forall n > 0 \quad b(n)P(X \in nA_x) = b(n)P(X \geq nx) = x^{-\alpha} \to x^{-\alpha} := \eta(A_x)$.

Thus, by theorem 4.2 of [3], $b(n)\mathbb{P}(X \in n\cdot) \to \eta(\cdot)$ and X is regular varying. This is exactly what is difficult to check in an infinite-dimensional case. The authors of [4] solved this problem by demonstrating the characterization of regular variation in $L^2[0,1]$ and $C[0,1]$.

In what follows, E may also designate the Hilbert space $L^2[0,1]$, the set of square functions integrable on $[0,1]$, with its canonical scalar product or $C[0,1]$ be the vector space of real-valued continuous functions on $[0,1]$.

The authors introduce the polar decomposition of X, i.e. $T : E_0 \to \mathbb{R}_+^\star \times \mathbb{S} \mapsto (r(x), \theta(x))$, where $\mathbb{S} = \{x \in E \quad r(x) = 1\}$, $r(x) = ||x||$, $\theta(x) = \frac{x}{||x||}$, and $\Theta(X) = \theta(X)$ and $R(X) = r(X)$.

Theorem 1 ([4]). *Let X be a random element in $\mathbb{H}$, and let θ_t be a random element in $\mathbb{H}$ distributed on the sphere $\mathbb{S}$ according to the conditional angular distribution $P_{\theta,t}$. Let $P_{\theta,\infty}$ denote a probability measure on $(\mathbb{S}, B(\mathbb{S}))$, and let Θ_∞ be a random element distributed according to $P_{\Theta,\infty}$. Where for any $t > 0$ $\mathbb{P}_{t,\Theta}(.) = \mathbb{P}(\Theta \in . \parallel R > t)$. The following statements are equivalent.*

(1) X is regularly varying with index α with limit angular measure $P_{\theta,\infty}$, so that $P_{\Theta,t} \xrightarrow{w} P_{\Theta,\infty}$.
(2) $\|X\|$ is regularly varying in $\mathbb{R}$ with index α, and $\forall h \in H, \langle \Theta_t, h\rangle \xrightarrow{w} \langle \Theta_\infty, h\rangle$ as $t \to \infty$.
(3) $\|X\|$ is regularly varying in $\mathbb{R}$ with index α, and $\forall N \in \mathbb{N}, \pi_N(\Theta_t) \xrightarrow{w} \pi_N(\Theta_\infty)$ as $t \to \infty$.

Where for any $N \in \mathbb{N}^\star$ $\pi_N : x \mapsto \mathbb{R}^N$ is defined by $\forall x \in L^2[0,1] \quad \pi_N(x) = (< x, e_i >)_{1\leq i\leq N}$, for $(e_i)_i$ an Hilbertian base of $\mathbb{H}$.

Theorem 1 gives, by the points (2) or (3) a more tractable tool to prove or disprove and to check numerically the regular variation of a random variable. From a theoretical point of view, we do not have to work with the borelian of $L^2[0,1]$, which is in our opinion not easy to manipulate. Plus, point (2) requires "only" checking the regular variation of $||X||_{\mathbb{H}}$ for which, we have plenty of strategies and we have to check the weak convergence of the random real-valued variable $< \Theta_t, h >$ for any $h \in \mathbb{H}$. This can be done numerically for a finite number of functions h. Notice that numerically, the basis $(e_i)_i$ can be chosen appropriately. As it is done in [4] when we model periodic measure it makes sense to choose periodic functions $\left(\sin 2k\pi \cdot|_{[0,1]}\right)_{k=1,..,N}$ for N equal to 2 or 3.

Another important consequence of this characterization is that, since the index of regular variation is the same for X as for $||X||_{\mathbb{H}}$, we can determine the index by returning to the one-dimensional case (for which we may use Hill-estimator) and, by the same token, we can determine the threshold.

Another very useful consequence of this theorem is that, in our opinion, is to shows that a random $\mathbb{H}$ valued variable is not regularly varying. Let $X \in \mathbb{R}^+$ by Proposition 1.3.2 of [5] we have that if $\forall \delta > 0 \quad \mathbb{E}X^\delta = \infty$, then $X \notin RV$. In fact,

$$\underbrace{\exists \alpha > 0, X \in RV_{-\alpha}}_{P0} \quad \underbrace{\Longrightarrow}_{\text{Proposition 1.3.2 of [5]}} \quad \underbrace{\forall u < \alpha \quad \mathbb{E}X^\alpha < \infty \quad \text{and} \quad \forall u > \alpha \quad \mathbb{E}X^\alpha = \infty}_{P_1}$$

$$\Longrightarrow \underbrace{\exists \delta (:= \frac{\alpha}{2}) > 0 \quad EX^\delta < \infty}_{P2}.$$

So we find the claim by taking the converse[1] of $P_0 \implies P_2$, combining it with the statement (3) of 1, it suffices to show that

$$\forall \delta > 0 \quad \mathbb{E}\,||X||_E^\delta = \infty. \tag{1}$$

Which is in practice an effective characterization of not being of regular variation. The next proposition gives us a way to construct random processes that are in $RV(L^2[0,1])$.

Proposition 1 ([4]). *[Construction of a regular varying variable in L^2] Let $Z = (Z_1, ..., Z_d) \in \mathbb{R}^d$ be regularly varying with index $\alpha > 0$ and limit measure μ, and let*

[1]("contraposée" in french)

$A = (A_1, ..., A_d)$ be a random vector of $\mathbb{L}^2[0,1]$-valued variables A_i, independent of Z, such that

$$\mathbb{E}\left(\sum_{1\leq i\leq d} ||A_i||^2_{L^2[0,1]}\right)^{\frac{\gamma}{2}} < \infty \tag{2}$$

for some $\gamma > \alpha$. Then, $X = \sum_{1\leq j\leq d} Z_j A_j$ is regularly varying in $\mathbb{L}^2[0,1]$ with limit measure $\tilde{\mu}(.) = \mathbb{E}\left(\mu(x \in \mathbb{R}^d \quad \sum_{1\leq j\leq d} A_j x \in (.)\,\{\}\right)$

The proposition below enables us to generate processes with regular variation in $L^2[0,1]$ in a very simple way since in practice the condition (2) is verified for many processes. For instance, it is of the type $(t,\omega) \mapsto tX(\omega)$ with X admitting a moment of order $\alpha + \varepsilon > 0$. It's a challenge to have enough correct data to train machine-learning models, especially GAN-type models (see [7]) and in the context of regular variation time series in $L^2[0,1]$ this proposition provides an answer.

They then show proposition 3.4, that if X is a continuous process on $[0,1]$ and with regular variation in $C[0,1]$ with convergence in the law of $\Theta(X)|\,||X||_\infty > t$ when $t \to \infty$ then X has regular variation in $L^2[0,1]$. They also show that the reciprocal is false. The counter-example used is defined for all $t \in [0,1]$ by

$$X(t) = \rho\left(1 - \frac{t}{3Z^2\exp(-2Z)}\right)\exp(Z)\mathbb{I}_{0,3Z^2\exp(-2Z)} \quad Z \sim Par(\alpha_Z) \quad \rho \sim Par(\alpha) \quad 0 < \alpha < \alpha_Z. \tag{3}$$

To show that this process is not regularly varying in $C[0,1]$ they show that X verifies (1) while to show that X is regularly varying in L^2 they show that X verifies Proposition 1.

2.1. Example and counter-example to the regular variation.

- Can we do without the condition (2)? The answer is no.

  Let $Z \sim Pareto(1) \implies Z \in RV_{-1}$ and A_1, defined by

  $$\forall t \in [0,1], \quad A_1(t) = 2tY \quad Y \sim \text{Log-Cauchy}(1). \tag{4}$$

  Since $||A_1||_{L^2[0,1]} = Y$, notice that, for any $\gamma > 0$, by Bertrand criterion:

  $$\mathbb{E}\,||A_1||^\gamma_{L^2[0,1]} = \int_{\mathbb{R}} \frac{t^{\gamma-1}}{\pi(1+\ln(t)^2)}\mathbb{I}_{t>0}dt < \infty \iff \gamma = 0.$$

  So [2] (1) is verified and by independence of A and Z, it implies that (1) is also verified for $||ZA_1||_{L^2[0,1]}$. Hence $||ZA_1||_{L^2[0,1]}$ is not regular varying and thus (2) is not satisfied. Therefore $ZA \notin RV(L^2[0,1])$

  Furthermore, to show that ZA_1 is not regularly varying in $L^2[0,1]$, it suffices to check: (1) with $E = L^2[0,1]$ which is clearly the case.

  The article [4] also addresses the question of regular variation in $(C([0,1]), ||\cdot||_\infty)$ and shows that if X is regularly varying in $C[0,1]$, then it is also in $L^2[0,1]$. However, the converse is not true. To demonstrate it, the authors provide a counterexample, which has been numerically illustrated in 4. An attempt to

[2] If the index of variation is greater than one, since for any $\delta \geq 1$, $t \in \mathbb{R}_+ \mapsto t^\delta$ is convex, by Jensen inequality we have, for $X \in L^2[0,1]$ that

$$\mathbb{E}\,||X||^\delta_{L^2[0,1]} \geq \left(\mathbb{E}\,||X||_{L^2[0,1]}\right)^\delta. \tag{5}$$

It follows that if $||X||_{L^2[0,1]}$ has no moment of order 1 it doesn't verify (2) (for $d = 1$).

construct our counterexample is also made:

- $RV(L^2[0,1]) \implies RV(C[0,1])$? The answer is no although $C[0,1]) \subset L^2[0,1]$.
  Let $Z \sim Pareto(4)$ and $Y \sim$ Log-Cauchy(1), independent of Z. Define A_2 as follows:

$$\forall t \in [0,1] \quad A_2(t) = tY \exp(Y)\mathbb{I}_{[0;\exp(-2^{-1}Y)[}(t). \tag{6}$$

  Since, $||A_2||_{L^2[0,1]} = 3^{-\frac{1}{2}} Y \exp(-\frac{Y}{2})$. For any $\gamma > 4$, $\mathbb{E}\,||A_2||^{\gamma}_{L^2[0,1]} < \infty$, we choose $\gamma = 5 > 4$ to verify the hypothesis (2) of Proposition 1.
  Thus, according to 1 with $d = 1$, we have: $ZA_2 \in RV(L^2[0,1])$.
  Additionally, $||ZA_2||_\infty = ZY$, so by the same argument using Bertrand integral we find that (1) is verified with $E = C[0,1]$ so $ZA_2 \notin RV(C[0,1])$.

- $RV(L^2[0,1]) \implies RV(L^1[0,1])$? Consider $A := A_2$ defined in (6) and $Z \sim$ Pareto(4). It follows that $A \in L^1[0,1]$ and $||A||_{L^1[0,1]} = \frac{Y}{2}$. Again, (1) is verified with $E = L^1[0,1]$ so $ZA \notin RV(L^1[0,1])$. Thus, the answer is also no even though (by Hölder and the fact that $[0,1]$ is bounded), $L^2[0,1] \subset L^1[0,1]$.

By combining the two previous examples we have constructed a process $U \in L^1[0,1]$ which is regularly varying in $(L^2[0,1], ||\cdot||_{L^2[0,1]})$ but not in $(C[0,1], ||\cdot||_\infty)$ nor in $(L^1[0,1], ||\cdot||_{L^1[0,1]})$ when we have $C[0,1] \subset L^2[0,1] \subset L^1[0,1]$.

3. Principal component analysis of extreme functions

3.1. **Theoritical part.** The regular variation hypothesis is often used to study extremes. In class, we saw, among other things, that in the case where $X \in \mathbb{R}$, F has regular variation of index $\alpha > 0$ if and only if F is in the domain of attraction of a Weibull. We've also seen how regular variation provides a link between machine learning and extreme value theory.

Let $\mathbb{H} = L^2[0,1]$, and X a random element variable $\mathbb{H}$-valued and regular varying. We also assume that $\mathbb{E}\,||X||^2_{\mathbb{H}} < \infty$[3] We assume that the context of theorem 1 is satisfied.

Let $\Theta_\infty \sim P_{\Theta,\infty}$ and $C = \mathbb{E}[\Theta_\infty \otimes \Theta_\infty] = \sum_{j\in\mathbb{N}} \lambda^j_\infty \phi^\infty_j \otimes \phi^\infty_j$ the covariance operator associated, where the integral over $\mathbb{H}$ are understood as in Böchner sense, $(\phi^\infty_j)_j$ and $(\lambda^j_\infty)_j$ denotes respectively eigenfunctions and eingenvalues associated to Θ_∞. Furthermore, for two given subspace V, W of $\mathbb{H}$ we measure their distance by $\rho(V,W) := ||\Pi_V - \Pi_W||_{HS}$ where Π_V is the orthogonal projection operator on V.

Let $X_1, \ldots, X_n$ an independent sample of X. Then fixe a number of excesses k above a random radial threshold chosen as the empirical $1-\frac{k}{n}$ quantile of the norm, with $k \ll n$. Denote $X_{(1)}, \ldots, X_{(n)}$ the permutation of the sample such that $\|X_{(1)}\| \geq \|X_{(2)}\| \geq \ldots \geq \|X_{(n)}\|$, and accordingly, let $\Theta_{(i)}, R_{(i)}$ denote the angular and radial components of $X_{(i)}$. Then $\|X_{(k)}\| = R_{(k)}$ is an empirical version of the $(1-\frac{k}{n})$ quantile of the norm R. The choice of k is a difficult subject as we have experienced in finite-dimensional cases in this course. We choose k by visual techniques. They define the empirical covariance of extreme angles

$$\hat{C}_k := \frac{1}{k}\sum_{i=1}^{k} \Theta_{(i)} \otimes \Theta_{(i)} = \frac{1}{k}\sum_{i}^{k} \hat{\lambda}_i \hat{\phi}_i \otimes \hat{\phi}_i.$$

[3]This assumption is necessary to ensure the definiteness of the covariance operator. See [4] Page 9 for more details.

Then define V_∞^p the space generated by the p first eigenvectors of C_∞ and $\hat{V}_k^p$ the space generated by the p first eigenfunction of $\hat{C}_k$. The authors show that $\hat{C}_k$ is consitent and that $(\hat{V}_k^p)_k$ converge to V_∞^p as $k \to \infty$. More precisely :

Proposition 2 ([4]). *The empirical covariance of extreme angles $\hat{C}_k$ is consistent, i.e. as $n, k \to \infty$ with $\frac{k}{n} \to 0$,*

$$\left|\left|\hat{C}_k - C_\infty\right|\right|_{HS} \to 0 \quad in \quad probability,$$

and

$$\lim_{k\to\infty} \rho(\hat{V}_k^p, V_\infty^p) = 0 \quad in \quad probability.$$

In the article [4], it is not stated like that, in particular, they have an explicit bound on the deviation of the eigenvectors spaces. The proofs are essentially based on Bernstein's type inequality and involve an intermediate pseudo-empirical covariance operator see Proposition 4.1 and 4.2 of [4] for more details. From an application point of view, this result has drawbacks and assets. One limitation is that it assumes that $(X_j)_{1\le j\le n}$ are iid which is in practice not the case at all and also the PCA lacks explicability.

However, the result is very interesting, as it shows by projecting from an infinite dimension to a finite dimension and by tending the number of samples toward infinity, little information is lost. In the article, the error regarding the empirical estimation and eigenspaces is in $O\left(\frac{1}{\sqrt{k}}\right)$, (see Theorem 4.2, [4]). In practice, we choose $p = 2$ or $p = 3$ to represent the variables.

4. Numerical illustration

We used meteorological data [1] and we generated some thanks to the proposition 2. The dataset we took is composed of time series. Each time series represents the temperature measured by a sensor during one year. There are 73 observations and each observation is in $\mathbb{R}^{365}$.

Consider $Z \sim$ Pareto(2) and $h_k(t) = \sin 2k\pi t_{k=0,1}$ and $A(t) = t$ and $X = ZA \in L^2[0,1]$. We want to show the convergence of $< \Theta_t, h_k >$ as $t \to \infty$ we illustrate it (as it is done in [4]) by plotting for a fixed k, $\frac{1}{n}\sum_{i=1}^n \left|< \Theta_{(i)}, h_k >\right|$.

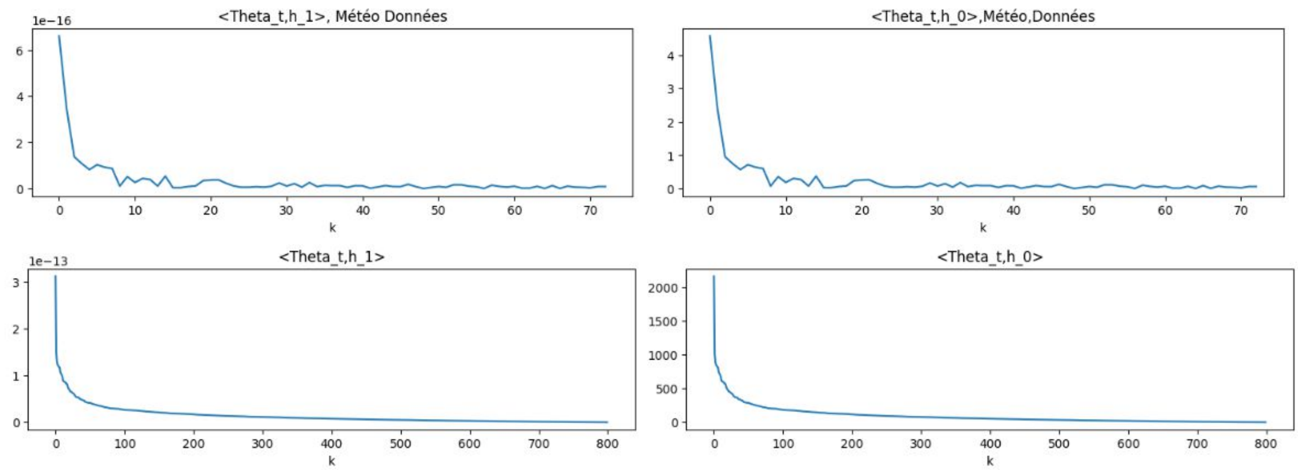

Figure 1. Numerical Illustration of the regular variation in $L^2[0,1]$ according to Theorem 1 for meteorological data and synthetic data.

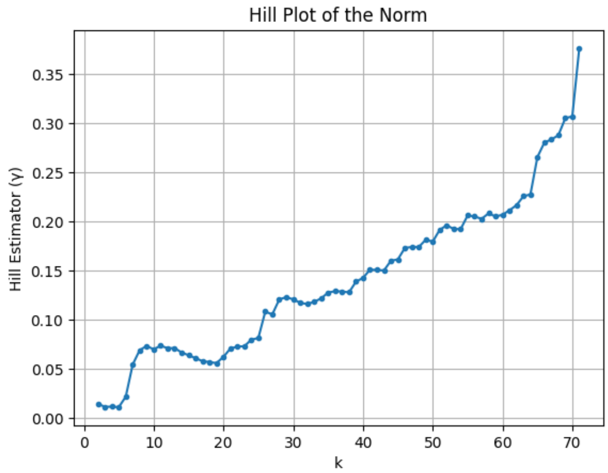

FIGURE 2. Hill plot for $||X_{\mathbf{meteo}}||_{L^2[0,1]}$

We estimate visually $\hat{\alpha} = \frac{1}{\hat{\gamma}} \approx \frac{1}{0{,}2} = 5$.

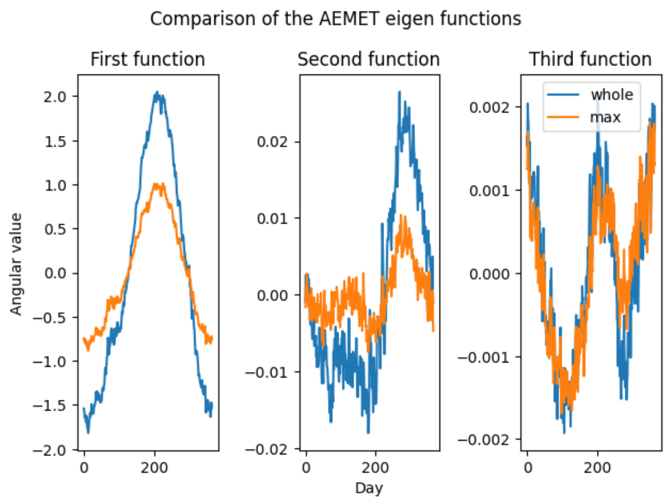

FIGURE 3. Eigenfunctions for the AEME set

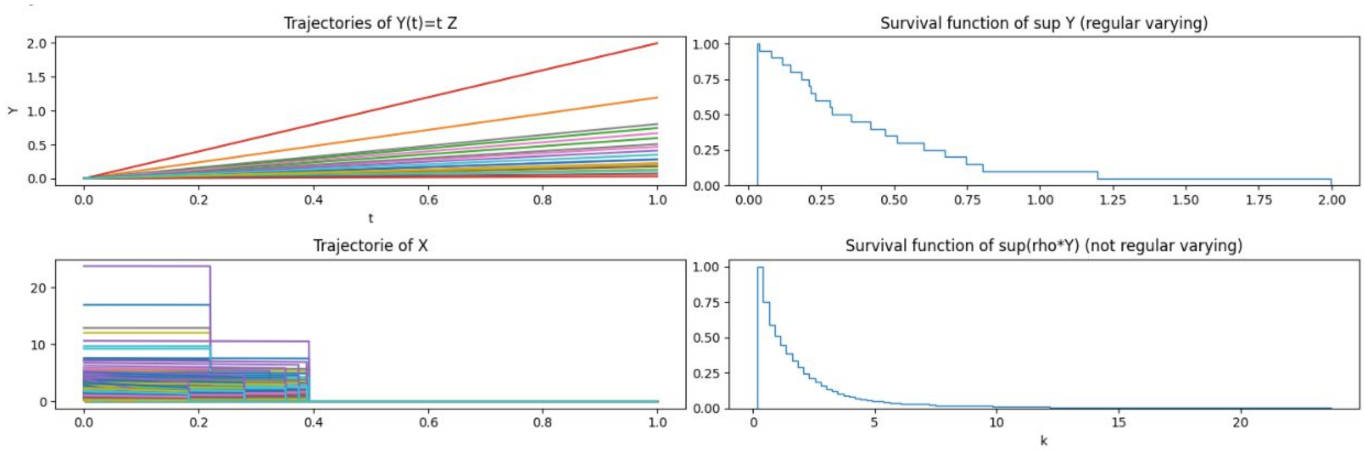

FIGURE 4. Numerical illustration of 3 and 4

References

[1] Agence d'État de la Météorologie de l'Espagne (AEMET), *Meteorological data of Spanish weather stations between years 1980 and 2009*, 2009. http://www.aemet.es/

[2] Resnick, S. I. (1987). *Extreme Values, Regular Variation and Point Processes.* Springer-Verlag New York.

[3] Hult, H. and Lindskog, F. (2006). *Regular variation for measures on metric spaces.* Publications de l'Institut Mathematique, 80(94), 121–140.

[4] Clémençon, S., Huet, N., Sabourin, A. (2023). *Regular Variation in Hilbert Spaces and Principal Component Analysis for Functional Extremes. arXiv:2308.01023* [math.ST].

[5] Mikosch, T. (1999). *Regular variation, subexponentiality and their applications in probability theory.*

[6] Drees, H., Sabourin, A. (2021). *Principal component analysis for multivariate extremes. Electronic Journal of Statistics, 15*, 908–943.

[7] Jinsung Yoon, Daniel Jarrett, Mihaela van der Schaar, *Time-series Generative Adversarial Networks*, in *Advances in Neural Information Processing Systems*, eds. H. Wallach et al., vol. 32, Curran Associates, Inc., 2019. Available at: https://proceedings.neurips.cc/paper_files/paper/2019/file/c9efe5f26cd17ba6216bbe2a7d26d490-Paper.pdf.