Statistical performance of local attractor dimension estimators in non-Axiom A dynamical systems

Flavio Maria Emanuele Pons, Gabriele Messori, Davide Faranda

Statistical performance of local attractor dimension estimators in non-Axiom A dynamical systems

Flavio Pons,[1] Gabriele Messori,[2, a)] and Davide Faranda[1, b)]
[1)] *LSCE-IPSL, CEA Saclay l'Orme des Merisiers, CNRS UMR 8212 CEA-CNRS-UVSQ, Université Paris-Saclay, 91191 Gif-sur-Yvette, France*
[2)] *Department of Earth Sciences and Centre of Natural Hazards and Disaster Science (CNDS), Uppsala University, Uppsala, 752 36, Sweden*

(*Electronic mail: flavio.pons@lsce.ipsl.fr)

(Dated: 20 July 2023)

We investigate various estimators based on extreme value theory (EVT) for determining the local fractal dimension of chaotic dynamical systems. In the limit of an infinitely long timeseries of an ergodic system, the average of the local fractal dimension is the system's global attractor dimension. The latter is an important quantity that relates to the number of effective degrees of freedom of the underlying dynamical system, and its estimation has been a central topic in the dynamical systems literature since the 1980s. In this work, we propose a framework that combines phase space recurrence analysis with EVT to estimate the local fractal dimension around a particular state of interest. While the EVT framework allows for the analysis of high-dimensional complex systems, such as the Earth's climate, its effectiveness depends on robust statistical parameter estimation for the assumed extreme value distribution. In this study, we conduct a critical review of several EVT-based local fractal dimension estimators, analyzing and comparing their performance across a range of systems. Our results offer valuable insights for researchers employing the EVT-based estimates of the local fractal dimension, aiding in the selection of an appropriate estimator for their specific applications.

Dynamical systems are characterized by an attractor, i.e. a compact region embedded in the space of the physical variables, hosting all of the system's trajectories. The attractor is a fractal object, thus its dimension is not integer. In real-world systems, such as the Earth's climate, the local fractal dimension around a point corresponding to a state of interest provides insights on the predictability of said state. Recent applications of statistical extreme value theory have led to developing estimators of the local fractal dimension. These have been used in several studies, especially in geophysics, sometimes making strong simplifying assumptions that may not be met in the real world. In this article, we compare the performance of several of these estimators, providing some general guidelines for their use in applied studies.

I. INTRODUCTION

Dynamical systems theory is a powerful mathematical framework that provides information on the behavior of time-evolving systems, ranging from simple cases such as pendulum motion to extremely complex ones such as the Earth's climate system. The general idea is to characterize all possible trajectories of a system, including their density, recurrence properties and persistence. The attractor of the system is an object which hosts all the information regarding these trajectories. It resides within the system's phase space, i.e. the space spanned by all the variables of the system. For natural chaotic systems originating from forced-dissipative dynamics, the attractor is a compact object in phase space: the trajectories have finite energy and cannot escape from the preferential region defined by the attractor.

The geometrical properties of the attractor of a dynamical system provide a number of insights on the system's behaviour. Here, we specifically consider the local fractal dimension (hereinafter simply d or 'local dimension'), whose average in the limit of an infinitely long time series of an ergodic system, is the system's attractor dimension D. The attractor dimension is an important quantity that relates to the number of effective degrees of freedom of the underlying dynamical system, and its estimation has been a central topic in the dynamical systems literature since the 1980s. Recently, a technique combining the analysis of recurrences in phase space with extreme value theory (EVT), has allowed to estimate D without relying on a correlation sum approach of the type proposed by Grassberger and colleagues[1]. The key idea of this method is that the limiting distribution of suitably rescaled recurrences of the system around a state of interest ζ can be modelled by the generalized Pareto distribution, following the Pickands–Balkema–De Haan theorem[2,3]. The latter provides an asymptotic result for the tail distribution of a sequence of independent identically distributed (iid) random variables following an unknown probability law. The application of the block maxima and peak over threshold approaches to time-dependent stationary processes, such as time series of observables issued from dynamical systems, rely on the work of[4] and[5], respectively.

In other words, recurrences in phase space can be viewed as extreme events and follow an extreme value law. The scaling of the ball centered around ζ and containing the recurrences serves as an estimate of the *local* fractal dimension around the state ζ, $d(\zeta)$. When a sufficient number of states ζ is consid-

[a)] Also at Department of Meteorology and Bolin Centre for Climate Research, Stockholm University, 106 91, Stockholm, Sweden
[b)] Also at London Mathematical Laboratory, 14 Buckingham Street, London, WC2N 6DF, UK

ered, the balls densely cover the attractor and, by averaging d over several states ζ, one obtains an estimate for D. Therefore, the EVT approach offers a way to measure D, while providing valuable information on the dynamical system in a local (in phase space) sense. More specifically, the local dimension informs on the geometry of the system's trajectories in the neighborhood of the state of interest. In general terms, a larger local dimension is associated to a lack of predictability and a lower local dimension to higher predictability.

The EVT-based method has been applied to several high dimensional complex systems, and in particular to better understand the Earth system. These applications include: dynamical characterisations of global climate states[6–8], key mechanisms of climate change[9], evaluation of climate models[10,11], and investigations of atmospheric predictability[12–17], atmospheric dynamics[18–21], dynamics and attribution of climate extremes[22–25], palaeoclimates[26], slow earthquakes[27] and more. These studies chiefly leveraged the local information provided by d to discriminate between different states (or, more properly, special Poincaré sections) of the climate system. As a concrete example, low local dimensions of sea-level pressure maps in the North Atlantic region are associated with frequent extratropical storms. High dimensions, on the contrary, correspond to so-called blocked flows, atmospheric configurations in which the typical mid-latitude westerly flow is diverted by a persistent high pressure system[18].

Given the broad range of applications of EVT-based estimates of d, it is important to choose a robust statistical estimator. Indeed, a number of different estimators, all based on extreme value laws, can be used to obtain d, yet a systematic comparison is lacking in the literature. The goal of this paper is to present different EVT-based estimators for d, and provide a critical review of the drawbacks and advantages of each of them, by comparing their performances on a range of dynamical systems.

Our study is structured as follows: in Section II we introduce the concepts of local and global attractor dimensions and their estimation via the Generalized Pareto distribution, followed by Section III which provides an overview of different statistical estimators proposed in the literature. In Section IV we describe the synthetic and real-world data that we use for the statistical evaluation of all the estimators, whose results are described in Section V. Finally, Section VI contains our conclusions and recommendations on the estimation of the local dimension of dynamical systems.

II. THEORETICAL FRAMEWORK

A. Local and global attractor dimensions via EVT

The concept of "fractional" dimension is the extension of the classical concept of topological dimension to non-integer values, introduced in 1918 by Hausdorff[28] as a way to measure the "size" of fractal objects. For instance, the Hausdorff dimension of a straight line is 1, whereas for a volume it is 3. In the case of a fractal object, like the Koch curve, the Hausdorff dimension is greater than 1, but smaller than 2. While the Hausdorff dimension of an attractor is a well-defined mathematical concept, its numerical estimation poses a number challenges. Precise knowledge of the Hausdorff dimension requires infinitely long time series and computations, involving mathematical limits for infinitesimally small radii around the point of interest. This global view of a system's attractor can be contrasted with the local dimensions, which provide a means to assess the dimensionality of a fractal object at a specific location within the phase space.

A classical method to approximate global and local dimensions of a fractal object is box counting. This involves counting how many boxes of a given size are required to cover the fractal object. By repeating this process for different box sizes, one obtains a curve representing the relationship between the box size and the number of boxes needed to cover the object. The slope of this curve can be used to estimate the global dimension and local dimensions of the fractal object.

Despite its usefulness in estimating dimensions of low-dimensional systems, the box counting method is limited by the fact that it only provides information about the behavior of the fractal object at a finite range of scales. It is therefore heavily affected by the so-called curse of dimensionality. As a result, dimension estimates for complex systems, such as Earth's climate, are unrealistically small[29–31].

More recently, a methodology based on EVT has been introduced to estimate local and global dimensions. This approach leverages EVT to extrapolate the behavior of the fractal object to very large or very small scales. In other terms, the probability distribution of the distance between a reference point and nearby points on the attractor is modeled as a power law. The exponent of this power law is related to the local dimension at that specific point. By employing EVT, it becomes possible to obtain computationally fast estimates of the local and global attractor dimensions of the fractal object, even in high-dimensional systems. Nonetheless, it should be noted that EVT does not overcome the existence of a bias associated with the curse of dimensionality[32].

Consider a time series of length N, denoted by y_j for $j = 1,\ldots,N$, which corresponds to observed realizations of a dynamical system Y. We designate a specific realization, denoted as ζ, as the reference state. Following[33,34], we consider the distance observable defined as

$$g_j = -\ln(\mathrm{dist}(y_j,\zeta)), \tag{1}$$

where $\mathrm{dist}(y_j,\zeta)$ is the Euclidean distance between each state y_j and the reference state ζ. We define the *recurrences* of the reference state ζ as $\{y_j : \mathrm{dist}(y_j,\zeta) < r\}$, i.e. the ensemble of states of the system located in the neighbourhood of ζ defined by the ball $B(\zeta,r)$. Let us consider the sequence of N values of g, and let g_q be an extreme quantile of the probability density function of g, $f_G(g)$, corresponding to the probability q. We define the Peaks Over Threshold (POTs) as the ensemble of $n = [N(1-q)]$ values such that $g > g_q$, where $[\cdot]$ denotes the closest integer operator.

We stress here that there is no univocal definition of 'extreme quantile'. When applying EVT to POTs to character-

ize extreme events in a time series (e.g. to study heatwaves or extreme precipitation over a certain region) the value of what is considered 'extreme' may be linked to some objective quantity, such as the impacts of the extremes. In our case, the choice should be dictated by the capability of the resulting estimator to converge to the correct dimension. Previous analyses[32] have shown that the EVT estimator is always biased, and that the bias depends on the chosen extreme quantile, with results shown for $q \in [0.98, 0.999]$. However, for a chosen quantile, the relationship between the attractor dimensions of different systems, or local dimensions of different states of a system, is preserved. In light of this, the choice of the threshold quantile is in part arbitrary: a common sense indication would be to use the highest quantile that still guarantees a sufficient exceedance sample size n for accurate GPD parameter estimation. All results presented in the following sections of this paper, if not specified otherwise, are obtained with $q = 0.98$.

The observable g in Eq. 1 is defined in such a way that the recurrences, states characterized by minima of the distance function, correspond to POTs of the observable itself. Given an n-dimensional sequence of POTs, we define the *exceedances* X as the POTs shifted by subtracting the threshold g_q, so that $X = \{X_\ell\}_{\ell=1}^{n} = \{g - g_q : g \geq g_q\}$, where $\ell = 1, \ldots, n$ and $\min X = 0$.

The choice of the Euclidean distance to define g guarantees that the threshold g_q used to defined the exceedances is in direct relation to the scaling of the ball $B(\zeta, r)$, since $r = \exp\{-g_q\}$. In other words, the asymptotic frequency with which the system enters a ball of radius r centred on ζ can now be expressed as the probability of exceeding a threshold corresponding to a high quantile g_q of the distribution of g.

In the limit $N \to \infty$, the exceedances follow a two-parameter Generalized Pareto Distribution $GPD(\sigma, \xi)$[2,3]. This distribution constitutes a classical statistical model for time series POTs, making it naturally applicable in this setting, where recurrences of ζ correspond to extreme values of g. The local dimension around the state ζ is then given by[34]:

$$d(\zeta) = \frac{1}{\sigma}. \tag{2}$$

By averaging d over a sufficiently large ensemble of states ζ on the attractor, one then obtains an estimate of the global attractor dimension D[35].

It is important to acknowledge that the EVT approach is not the only recently developed method for estimating local dimensions using recurrences or related concepts. For instance,[36] proposed an estimator based on hitting times, that exhibits rapid convergence for low-dimensional dynamical systems with known global attractor dimension. However, we have chosen not to include this particular estimator in our study. Instead, our focus lies on assessing the effectiveness of different EVT estimation methods specifically based on generalized Pareto distribution (GPD) parameters. These estimators have been extensively employed in applied studies, often making the unrealistic assumption of an Axiom A system.

B. Axiom A systems and extreme value laws

The term "Axiom A" denotes a specific class of dynamical system behavior characterized by the presence of a uniform fractal set as an attractor. This uniform fractal set is a mathematical construct defined by self-similarity, where smaller portions of the set replicate the overall structure of the entire set. In Axiom A systems, the following characteristics hold true: i) the attractor represents a non-empty collection of points toward which the system tends to converge over time. ii) The attractor exhibits self-similarity, meaning that a closer examination of any small region within the attractor reveals a resemblance to the overall shape and structure of the entire attractor.

The fractal properties within Axiom A attractors are uniform, implying uniform predictability across the entire attractor. It is important to note that while Axiom A serves as a framework for understanding certain aspects of dynamical systems, natural systems such as physical, biological, or ecological systems do not strictly adhere to this paradigm. Natural systems often exhibit a broader range of behaviors encompassing both regular and irregular dynamics. They can involve chaos, intricate interactions, and non-linear relationships that may not conform to the strict requirements of Axiom A.

Nevertheless, certain aspects of Axiom A systems may be observed within natural systems. For example, specific patterns or structures within natural systems may exhibit self-similarity, fractal-like properties, or predictable behaviors across certain scales or under particular conditions.

For Axiom A systems in the asymptotic case $n \to \infty$, $\xi \to 0$, and the two-parameter $GPD(\sigma, \xi)$ converges to an exponential distribution, $Exp(\sigma)$[37]. In this case, the local dimension estimator is the inverse of the exceedances' sample mean (Eq. 18). However, as discussed above, real-world systems do not realistically satisfy the Axiom A assumption, and the exponential model for the exceedances distribution likely introduces some bias to the estimates.

C. Generalized Pareto Distribution

A random variable X follows a generalized Pareto distributon $GPD(\mu, \xi, \sigma)$ if its cumulative distribution function (CDF), giving the probability $P(X > x)$, is defined as

$$F(x|\xi, \mu, \sigma) = \begin{cases} 1 - \left(1 + \xi \frac{x-\mu}{\sigma}\right)^{-1/\xi}, & \xi \neq 0 \\ 1 - \exp\left(-\frac{x-\mu}{\sigma}\right), & \xi = 0 \end{cases} \tag{3}$$

and the corresponding probability density function (pdf) reads:

$$f(x|\xi, \mu, \sigma) = \begin{cases} \frac{1}{\sigma}\left(1 + \xi \frac{x-\mu}{\sigma}\right)^{-1/\xi - 1}, & \xi \neq 0 \\ \frac{1}{\sigma} \exp\left(-\frac{x-\mu}{\sigma}\right), & \xi = 0 \end{cases} \tag{4}$$

where $\xi \in \mathbb{R}$ is the shape parameter, $\mu \in \mathbb{R}$ is the location parameter and $\sigma > 0$ is the scale parameter. The support of X depends on the shape parameter: if $\xi \leq 0$, $\mu \leq x \leq \mu + \frac{\sigma}{\xi}$; if

$\xi > 0$, $\mu \leq x \leq \infty$. The moment generating function is:

$$M_X(t) = E_X[e^{tx}] = e^{t\mu} \sum_{j=0}^{\infty} \left[\frac{(t\sigma)^j}{\prod_{k=0}^{j}(1-k\xi)} \right], \quad \text{for } k\xi < 1$$

where $E_X[\cdot]$ denotes expectation computed with respect to the density of X. The r-th moment around zero can be obtained as:

$$E[X^r] = \left. \frac{d^r M_X(t)}{dt^r} \right|_{t=0} = r! \frac{\sigma^r}{(-\xi)^{r+1}} \frac{\Gamma\left(-\frac{1}{\xi} - r\right)}{\Gamma\left(1 - \frac{1}{\xi}\right)}, \quad \text{for } \xi < \frac{1}{r} \tag{5}$$

In the special case $\mu = 0$ the three-parameter $GPD(\mu, \xi, \sigma)$ reduces to a two-parameter $GPD(\sigma, \xi)$. In the following, we will consider this specific case, since the procedure described in Section III requires to shift the POTs so that their minimum is 0.

There are several methods to estimate the parameters of the GPD distribution, ranging from frequentist techniques such as Maximum likelihood (ML), the method of moments (MM), probability-weighted method of moments (PWM), linear combination of order statistics and others, to Bayesian methods and approaches based on entropy[38,39].

Here, we focus on the Axiom A degenerate case and on several frequentist methodologies, including ML, MM and PWM estimation, plus an alternative formulation of ML, which we will call pseudo-ML (PML). We choose these estimators based on their computational simplicity (MM and PWM), or because an implementation is readily available in an `R` package (ML, PML).

III. GPD PARAMETER ESTIMATION

In this section, we introduce five estimators of the parameter $\sigma = \frac{1}{d}$, namely the inverse local dimension. The first one (given in Eq. 18), is constructed assuming the degenerate case of POTs following a $GPD(\xi = 0, \sigma) \equiv Exp(\sigma)$. The others (given in Eq.s 19-22) are constructed under the assumptions of POTs following a $GPD(\sigma, \xi)$.

A. Estimation based on $Exp(\sigma)$

As already mentioned, for Axiom A systems in the asymptotic case $n \to \infty$, $\xi \to 0$, so that the distribution of the exceedances converges to an exponential distribution.

In such a case, a ML estimator for the local dimension is readily available in close form, given by $\hat{\sigma}_m = \bar{x}$. Using this estimator,[32] have shown that the estimated attractor dimension is biased, with a bias growing nonlinearly with the topological dimension of the system, as a particular case of the curse of dimensionality[40], due to the concentration of the Euclidean norm. This effect was analytically proven for very long time series of random vectors, and shown to equally affect real-world complex data, such as sea-level pressure fields or multivariate time series of financial indices.

However, for both real-world and simulated systems displaying chaotic behaviour, the asymptotic case may not be met as easily as for purely stochastic systems. Much longer sampling is required to observe the invariant distribution of chaotic systems compared to random vectors, and assuming exponential exceedances may constitute a serious model misspecification, which may introduce further bias.

B. Estimation based on $GPD(\sigma, \xi)$

Maximum Likelihood (ML)

Maximum Likelihood is one of the most popular methods to estimate the parameters of a statistical model, and it has been widely used since its first introduction[41]. It is beyond our scope to give an overview of ML, which can be found in many classic statistics textbooks (e.g.[42]). In brief, given a sample x consisting of n independent realizations of a random variable with probability density function $f(x|\theta)$, where θ denotes the vector of parameters, we call likelihood function the quantity:

$$L(\theta|x) = \prod_{i=1}^{n} f(x_i|\theta). \tag{6}$$

The ML estimator of the parameter vector is the point $\hat{\theta}$ of the parameter space at which $L(\theta|x)$ attains its maximum. Finding the ML parameter estimates can be a simple analytical task or a challenging numerical procedure, depending on the smoothness of the likelihood function. In the case of the GPD, it is not possible to obtain closed-form estimators analytically, so that alternative strategies must be found to maximize $L(\theta|x)$. To obtain ML estimates for the GPD, we use the `gpd.mle()` function from the `R` package `'mev'`[43], based on the profile likelihood maximization method of[44].

Pseudo Maximum Likelihood (PML)

An alternative way to obtain ML estimates of the GPD parameters has been introduced by Villaseñor-Alva and González-Estrada[45] and implemented in the `R` package `gPdtest`[46]. The estimator is constructed differently depending on the sign of ξ; given this composite strategy and the passage through intermediate steps not based on the likelihood function, we will refer to this method as "pseudo maximum likelihood". Here we only provide a sketch and refer the reader to the aforementioned article for full details.

For $\xi \geq 0$, `gPdtest` implements a maximization of the asymptotic likelihood, based on the logarithm of the k largest values in the ordered sample $x_{1:n}, x_{2:n}, \ldots, x_{n:n}$: $W_j = \log x_{j:n}$ with $j = n-k+1, n-k+2, \ldots, n$. For $n \to \infty$ the log-likelihood can be expressed as a function of W_j, and the two

ML estimators follow:

$$\hat{\xi}_k = -\left(W_{n-k+1} - \frac{1}{k}\sum_{j=1}^{k} W_{n-j+1}\right) \tag{7}$$

$$\hat{\sigma}_k = \hat{\xi}_k \exp\left\{W_{n-k+1} + \hat{\xi}_k \log\frac{k}{n}\right\}. \tag{8}$$

The case $\xi < 0$ requires a composite strategy, combining a moment equation step and a maximum likelihood estimator of the upper boundary of the GPD. Recalling that we consider the case $\mu = 0$, the variable $U = [1-F(X|\xi,\sigma)]^{-\xi} = 1 + \frac{\xi}{\sigma}X$ follows a Beta$(-\frac{1}{\xi}, 1)$ distribution. The first sample moment is $\overline{u} = \frac{1}{n}\sum_{i=1}^{n}\left(1+\frac{\xi}{\sigma}x_i\right) = 1 + \frac{\xi}{\sigma}\overline{x}$, to be equated to the expected value of the Beta distribution $E[U] = \frac{1}{1-\xi}$, resulting in $\xi = 1 - \frac{\sigma}{\overline{x}}$. Let us now recall that, for $\xi < 0$, $0 \le x \le \sigma/\xi$: then, the sample maximum $x_{n:n}$ is a ML estimator for the upper bound σ/ξ. Combining the two steps, one obtains:

$$\hat{\sigma}_{PML} = -\hat{\xi}x_{n:n} \tag{9}$$

$$\hat{\xi}_{PML} = \frac{\overline{x}}{\overline{x} - x_{n:n}} \tag{10}$$

Since the estimation strategy depends on the sign of the shape parameter, a priori knowledge on the value of ξ must be used. Results discussed in this paper from PML estimation are based on the sign of $\hat{\xi}_{PWM}$, introduced later. We have also tested the possibility to use $\hat{\xi}_{MM}$ (not shown in the article) obtaining equivalent results.

Method of Moments (MM)

The method of moments is another common parameter estimation method. It consists of expressing the theoretical moments as a function of the parameters and solving the resulting equations for the parameter estimates using the values of the sample moments.

Let X be a random variable $X \sim GPD(\sigma,\xi)$, $x = x_1, x_2, \ldots, x_n$ being a sample of n observations from X. The first two GPD moments, required to estimate the two parameters, can be obtained using Eq. 5:

$$\begin{aligned} E[X] &= \frac{\sigma}{1+\xi}, && \text{for } \xi < 1 \\ Var[X] &= E[X^2] - E[X]^2 = \frac{\sigma^2}{(1+\xi)^2(1+2\xi)}, && \text{for } \xi < \frac{1}{2} \end{aligned} \tag{11}$$

and the MM parameter estimators follow by equating sample and theoretical moments, and solving for σ, ξ:

$$\begin{aligned} \hat{\xi}_{MM} &= \frac{1}{2}\left(\frac{\overline{x}^2}{s^2} - 1\right) \\ \hat{\sigma}_{MM} &= \frac{\overline{x}}{2}\left(\frac{\overline{x}^2}{s^2} + 1\right) \end{aligned} \tag{12}$$

where $\overline{x} = n^{-1}\sum x_i$ is the sample mean and $s^2 = (n-1)^{-1}\sum(x_i - \overline{x})^2$ is the sample variance. Given the condition $\xi < \frac{1}{r}$ in Eq. 5, the estimators in Eq. 12 only hold for $\xi < 0.5$.

Notice that $\hat{\sigma} = \overline{x}(\hat{\xi}+1)$. If $\xi = 0$, the GPD reduces to a standard exponential distribution, $X \sim Exp(\sigma)$, and the MM estimator of σ is the sample mean. On the other hand, if it is erroneously assumed that $\xi = 0$, the estimate of σ is affected by a bias equal to $-\overline{x}\hat{\xi}$.

Probability Weighted Moments (PWM)

Finally, we consider a variation of the MM, based on probability weighted moments (PWM). This method is particularly suited for distributions that can be uniquely expressed in terms of their inverse CDF in a simple way, but it is also popular in hydrological modeling due to its computational simplicity. For a random variable X with CDF $F(X)$, the PWM are defined as

$$M_{p,r,s} = E[X^p (F(X))^r (1-F(X))^s], \qquad p,r,s \in \mathbb{R} \tag{13}$$

For the estimation of $GPD(\sigma,\xi)$, it is suggested[38] to use the moments defined by $p = 1$, $r = 0$:

$$\begin{aligned} \alpha_s &= M_{1,0,s} = E[X(1-F(X))^s] \\ &= \frac{\sigma}{(s+1)(s+1+\xi)}, \qquad \xi < 1, \quad s = 0,1,2,\ldots \end{aligned} \tag{14}$$

so that the PWM expression of the *GPD* parameters is:

$$\begin{aligned} \xi &= 2 - \frac{\alpha_0}{\alpha_0 - 2\alpha_1} \\ \sigma &= \frac{2\alpha_0\alpha_1}{\alpha_0 - 2\alpha_1}. \end{aligned} \tag{15}$$

Given a sample x of size n from $X \sim GPD(\sigma,\xi)$, the sample version of the PWM α_s is

$$a_s = \frac{1}{n}\sum_{i=1}^{n} x_{i:n}(1-p_{i:n})^s, \tag{16}$$

where $x_{i:n}$ is the i-th element of the ordered sample and $p_{i:n}$ are the corresponding plotting positions, such that $(1-p_{i:n})$ is a sample estimate of $(1-F(X))$. Notice that $a_0 = \overline{x}$. There are several possible ways to estimate the plotting positions. Here, we adopt the simple expression recommended in the literature[47,48]:

$$p_{i:n} = \frac{i - 0.35}{n}.$$

The two estimators can be obtained by plugging Eq. 16 in Eq.s 15. Once again, we recall that the upper limit of $GPD(\sigma,\xi)$ is given by $-\sigma/\xi$ and therefore estimated by $-\hat{\sigma}/\hat{\xi}$; however, if $-\hat{\sigma}/\hat{\xi} < x_{n:n}$, the sample maximum $x_{n:n}$ itself is a better estimator of the upper limit. The final expression of the PWM estimators follows:

$$\begin{aligned} \hat{\sigma}_{PWM} &= \frac{2a_0a_1}{a_0 - 2a_1}, \qquad \hat{\xi}' = 2 - \frac{a_0}{a_0 - 2a_1} \\ \hat{\xi}_{PWM} &= \begin{cases} \hat{\xi}' & -\hat{\sigma}/\hat{\xi}' \ge x_{n:n} \\ -\frac{\hat{\sigma}}{x_{n:n}} & -\hat{\sigma}/\hat{\xi}' < x_{n:n} \end{cases} \end{aligned} \tag{17}$$

The PWM estimators are unbiased, and for large samples are normally distributed with mean (σ, ξ) and known covariance matrix.

C. Summary of local dimension estimators

To summarize, we will consider the following five expressions of the inverse local dimension:

$$\hat{d}_m^{-1} = \hat{\sigma}_m = \bar{x} \tag{18}$$

$$\hat{d}_{ML}^{-1} = \hat{\sigma}_{ML} = \arg\max\{L(\sigma, \xi | x)\} \tag{19}$$

$$\hat{d}_{PML}^{-1} = \hat{\sigma}_{PML} = -\frac{\bar{x} x_{(n)}}{\bar{x} - x_{(n)}} \tag{20}$$

$$\hat{d}_{MM}^{-1} = \hat{\sigma}_{MM} = \frac{\bar{x}}{2}\left(\frac{\bar{x}^2}{s^2} + 1\right) \tag{21}$$

$$\hat{d}_{PWM}^{-1} = \hat{\sigma}_{PWM} = \frac{2\bar{x}a_1}{\bar{x} - 2a_1} \tag{22}$$

with $a_1 = \frac{1}{n}\sum_{i=1}^{n} x_{1:n}(\frac{i-0.35}{n})$. We do not provide an analytical expression for $\hat{\sigma}_{ML}$, as this is obtained by numerical optimization of the likelihood function.

IV. DATA

In order to evaluate and compare the statistical performance of the estimators, we use data from three sources. First, we perform a Monte Carlo study, so that we can compare estimates to known true values of σ, as a function of ξ and of the sample size. Then, we generate a very long trajectory from the well-known Lorenz63 attractor[49], whose global attractor dimension is known, and which is sufficiently low-dimensional that the curse of dimensionality is not relevant. Finally, we compare results obtained from the different estimators when applied to real-world climatological data, that are the main motivation for this study.

A. Monte Carlo experiment

As a first step, we conduct a Monte Carlo experiment to assess the performance of the estimators as a function of the true values of σ, ξ and of the number of exceedances n. In order to do so, for every combination (σ, ξ, n) we generate $M = 5000$ random samples drawn from a $GPD(\sigma, \xi)$ with sample size $n \in [10, 100]_{10} \cup \{150, 200, 500, 1000\}$, $\xi \in [-1, 1]_{0.2}$; $\sigma = 1/d$, with $d \in [1, 5]_{0.2} \cup [5, 100]_5$, where $[a, b]_\delta$ denotes a partition of the closed a, b interval with step δ. Notice that n is the number of available exceedances, so that the original sample size is implied, and in real cases it depends on the chosen POT threshold: for example, $n = 100$ and a threshold at the 99th percentile implies an initial time series of 10^4 observations.

B. Lorenz 1963

Moving further beyond Monte Carlo data, we consider a simple chaotic dynamical system: the well-known Lorenz63 model[49]. This is a three-variable system meant to describe Rayleigh-Bénard convection, and consisting of the following ordinary differential equations:

$$\begin{aligned} \frac{dX}{dt} &= \sigma(Y - X) \\ \frac{dY}{dt} &= X(\rho - Z) - Y \\ \frac{dZ}{dt} &= XY - \beta Z \end{aligned} \tag{23}$$

Here X represents convective overturning rate, Y horizontal temperature variations, and Z departure from a linear vertical temperature gradient. We use the standard parameter configuration to obtain chaotic behaviour (the well-known "butterfly"), namely $\sigma = 10$, $\rho = 28$ and $\beta = 8/3$. We use a timestep of 0.001 s and generate $4 \cdot 10^5$ iterations. We discard the first 2000 iterations and sub-sample the trajectory taking one every 4 iterations.

C. ERA5 climate reanalysis

As a final step, we consider climate reanalysis data. This is data issued from a full-complexity atmospheric global model, constrained by a very large amount of in-situ and remotely sensed data.

In particular, we choose the ERA5 reanalysis dataset[50], produced by the European Centre for Medium-Range Weather Forecasts (ECMWF). We use 500 hPa geopotential height data with a 0.25° horizontal resolution and daily temporal resolution. We consider data over the period 1979–2018 and a domain spanning 26° – 70° North and 22° West – 46° East. This provides a lat-lon grid of 273×177 points covering Europe, and parts of the Eastern Atlantic, Russia and North Africa.

This is a conceptually different case from the Lorenz63 model. In the latter, each phase space dimension corresponds to a physical variable. In the former, physical variables are measured on a regular spatial grid, and the phase space's dimension corresponds to the number of gridpoints in the data.

V. RESULTS

A. Monte Carlo experiment

As described in Section IV A, we specify a set of combinations (σ, ξ, n), for each combination we generate 5000 Monte Carlo samples, and for each sample we estimate the dimension with the five estimators described by Eq.s 18-22. All results discussed in this section are obtained as the average over the 5000 Monte Carlo samples for each (σ, ξ, n) combination,

and refer to the local dimension $d = 1/\sigma$. The dependence of the estimator performance on ξ is important to understand the actual applicability of each estimator to particular classes of systems.

In the following, we will consider two error measures to evaluate the performance of the different estimators: the root mean square error (RMSE) and the relative error. The RMSE is defined as the square root of the average squared deviation:

$$RMSE = \sqrt{E[(\hat{d}-d)^2]} = \sqrt{Bias(\hat{d})^2 + Var(\hat{d})} \qquad (24)$$

where the expectation is taken over the M Monte Carlo replicates, and it is thus a function of n, ξ and the true value d. Since the mean square error can be decomposed in terms of bias and variance of the estimator as shown in Eq. 24, the RMSE allows to find the most efficient estimator, in the sense of the best bias-variance trade-off. As an example, Figure 1 shows values of RMSE as a function of d and n, for $\xi = -0.5$. The complete set of results for all tested ξ is shown in supplementary figures S1-S21.

In general, larger sample sizes correspond to smaller RMSE, and larger RMSE are found for large values of the true dimension d. For the mean estimator, however, a strong positive effect of the sample size is observed only at $\xi \sim 0$, where the underlying assumption of an exponential distribution of the exceedances is approximately met. The MM estimator produces large RMSE for large n when $\xi \gtrsim 0.5$, which is expected as the second moment of the GPD only exists for $\xi < 0.5$. Notice that the PML shows a qualitatively similar dependence of RMSE on n, d, but the RMSE magnitude is much larger compared to the other estimators.

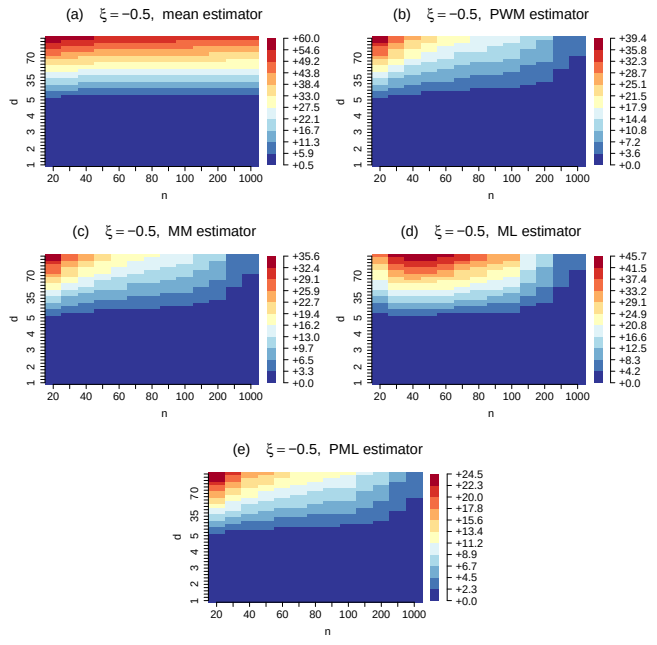

FIG. 1. Root mean square error of the mean estimator (a), PWM (b), MM (c), ML (d) and PML (e) on d as a function of d and n, for $\xi = -0.5$. Note that the colour ranges differ in each panel.

Figure 2 and Table I summarize the effect of the true value of ξ on RMSE for n = 20, 100, 1000, averaged over the true dimension. As expected, the mean estimator has the smallest RMSE when $\xi \simeq 0$, but it loses efficiency when this assumption is not met. The PML is the most efficient over a restricted range of negative values of ξ, however for $\xi \gtrsim -0.1$ it produces by far the largest RMSE values. The efficiency of the ML as a function of ξ changes visibly from small to large sample sizes, where it shows larger RMSE for $-0.5 < \xi < 0$. The MM estimator shows a good performance for $\xi \lesssim 0.5$, but it is affected by large RMSE values at larger values of ξ, which does not improve with sample size. Finally, the PWM is the most stable estimator: for small n it is not the most efficient, but its performance is largely independent of the true value of ξ. At larger sample sizes the RMSE is overall small, and again displays only small variations as a function of ξ – the most notable being a decreased efficiency for ξ close to 1.

Overall, it is difficult to point out the best estimator in terms of efficiency when the sample size is small. For larger samples, the PWM is the best choice; however, if it can be safely assumed that $\xi < 0$, MM and PWM are in close competition in terms of efficiency.

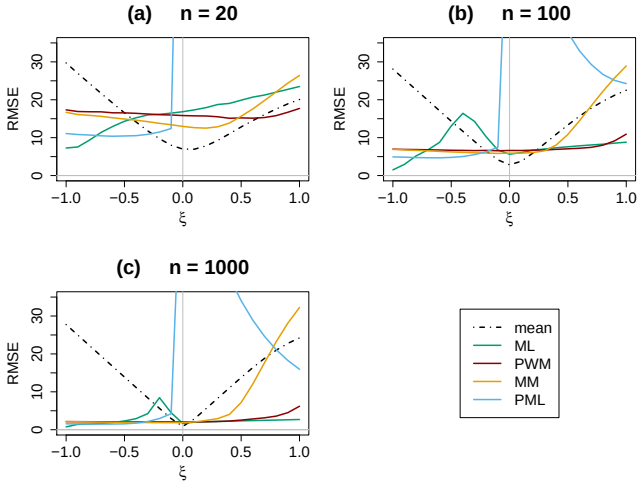

FIG. 2. Graphic visualization of the results shown in Table I. Root mean square error of the five considered estimators of the local dimension, averaged over d and shown as a function of ξ, for sample sizes n = 20 (a), 100 (b), 1000 (c).

As a further way to characterize the performance of the estimators, we consider the average relative error $E[(\hat{d}-d)/d]$, where the expectation is computed over the M Monte Carlo replicates as for the RMSE. This measure allows to evaluate the magnitude of the estimation error relative to the true value, while also assessing the sign of the bias.

As an example, Figure 3 shows the values of the relative error for the five considered estimators, as a function of the sample size n and of the true dimension d for $\xi = -0.5$. The complete set of results for all ξ is shown in supplementary figures S22-S42.

For all values of ξ and for all estimators, results improve with the sample size, with overall convergence at values

around $n \sim 200$. Results appear to be stable with respect to the true dimension d for all methods, except maximum likelihood. For $\xi > 0$, the error of the ML estimator seems to be independent of d, or shows a weak dependence for $0 < \xi \lessapprox 0.2$. However, for $\xi < 0$ the ML estimator shows a strong dependence on d, especially for sample sizes $50 \lessapprox n \lessapprox 200$: for $-1 \lessapprox \xi \lessapprox -0.5$ the relative error becomes more negative with increasing d, while for $-0.5 \lessapprox \xi \lessapprox 0$ it becomes more positive.

For the special case $\xi = 0$ (Fig. 3), when the GPD converges to an exponential distribution, the performance of the mean estimator shows a great improvement, with relative errors $\sim 0.5\%$ at $n = 100$ and $\sim 0.01\%$ for $n > 200$. This is expected, since in this case the sample mean is also the exact maximum likelihood estimator of σ.

The PML estimator has the overall worst performance, with relative errors comparable to other methods for $\xi < 0$ but jumping to errors of $100 - 200\%$ for $\xi \geq 0$. Even though the package authors recommend the use of `gPdtest` for all values of ξ, Eq.s 7-8 show that an unbiased estimation is not possible for $\xi = 0$, since if ξ was correctly estimated, it would necessarily result in $\hat{\sigma} = 0$, corresponding to a degenerate distribution with 0 variance, and to an infinite estimated dimension.

As done for RMSE, we next fix n and average over d to obtain a synthetic representation of the relationship between ξ and the relative estimation error affecting $\hat{d}$. As we pointed out earlier, the ML estimator error shows some dependence on d for $\xi \lessapprox 0.2$. However, the range of relative value errors is small, in particular compared to the mean and the PML estimators. We thus still include ML in this analysis for a global comparison. Results are summarized in Table II and shown in Fig. 4, where the relative error averaged over d for $n = 20, 100, 1000$ is plotted as a function of ξ.

The performance of PML shows large relative errors for $\xi > -0.2$, and a non-monotonic, discontinuous behaviour as a function of ξ. The ML estimator also shows a non-monotonic dependence of the relative error for negative values of ξ. This irregularity is likely due to the aforementioned dependence of the ML estimator on d, overlooked in this analysis. All the other estimators produce a relative error that is monotonically decreasing with the true value of ξ. The two moment-based estimators have a similar performance, with relative errors $\approx 0.1\%$ for $\xi \leq 0$. The performance of the MM estimator degrades for values larger than some value in the range $0 < \xi \lessapprox 0.5$ depending on the sample size, with large samples allowing for accurate estimations close to the limit of the domain of existence $\xi < 0.5$. Similarly, the PWM exist for $\xi < 1$ and the corresponding estimator shows a tendency to under-estimate σ when $\xi \to 1$, with smaller errors at larger sample size.

The performance of the mean estimator is particularly interesting. We recall that this is a well-specified unbiased estimator only under the assumption of an Axiom A system in the asymptotic case, condition that can be rarely satisfied in real-world data. On the other hand, it is very simple to compute, and it has been used in several studies, providing physically coherent results. Indeed, while the relative error of this estimator is large compared to the others, it has the peculiar property of being $(\hat{\sigma} - \sigma)/\sigma = -\xi$, except for $\xi \approx 1$, where it slightly deviates from this relationship.

Overall, for systems where it is expected that $\xi \leq 0$, the MM and PWM perform the best. ML performs slightly worse for $\xi \leq 0$, but it is the only estimator providing unbiased estimates for $\xi > 0$. All three estimators are quite stable with the sample size, providing fairly good estimates already for $n = 20$. The mean estimator is severely biased, but its error is monotonic and approximately linear with ξ, so that conclusions obtained from comparing local dimensions of different states or systems in relative terms still hold.

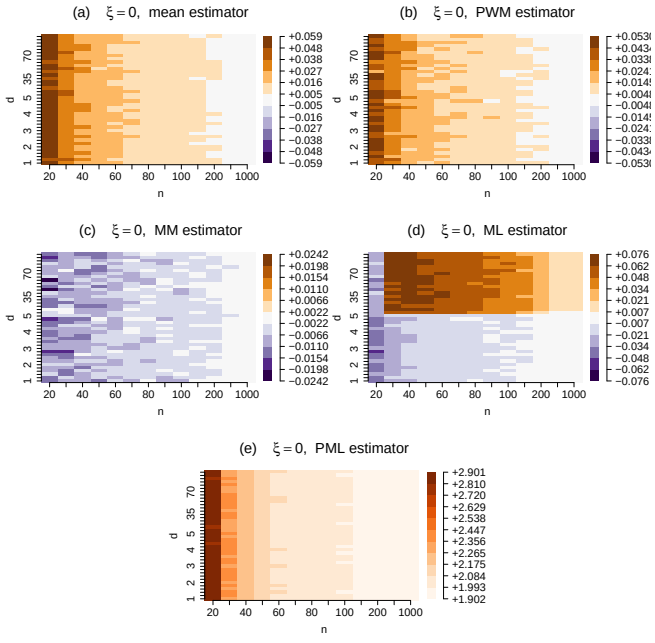

FIG. 3. Relative error of the mean estimator (a), PWM (b), MM (c), ML (d) and PML (e) on d as a function of d and n, for $\xi = 0$. Note that the colour ranges differ in each panel.

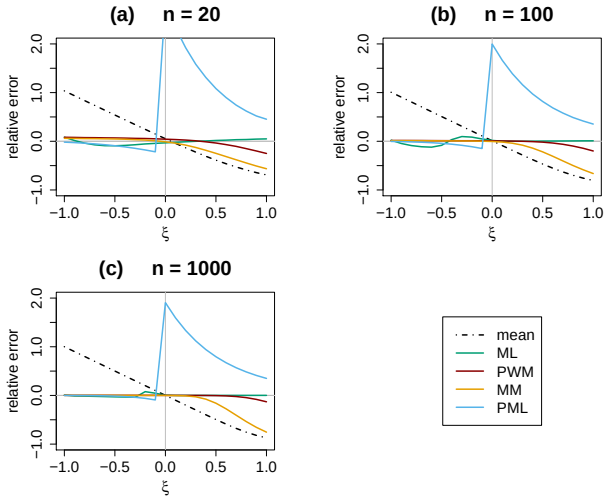

FIG. 4. Graphic visualization of the results shown in Table II. Relative error of the five considered estimators on the local dimension, averaged over d and shown as a function of ξ, for sample sizes n = 20 (a), 100 (b), 1000 (c).

B. Lorenz 1963 system

Next, we analyze the performance of the estimators on the local dimension of the L63 system. For this system, it is known that $D = 2.06 \pm 0.01$ since the early work of Grassberger and Procaccia [51], and more recently it has been estimated to be $D = 2.0627160$[52]. We choose the L63 system not only because its D is known, but also because in this case the estimation of the local dimension via Euclidean distance is not affected by the curse of dimensionality documented for higher-dimensional systems[32], since the phase space of L63 has dimension 3.

Before looking at the estimates of d and D, we consider the performance of the different estimators on the shape parameter ξ. For the MM, PWM and PML estimators, $\hat{\sigma}$ can be written as a function of $\hat{\xi}$, and analysing ξ may bring insights on the source of estimation errors on d. Figure 5 shows the probability density functions of $\hat{\xi}$ for all estimators, except the mean estimator which assumes $\xi = 0$. We can notice marked differences: PML results in a roughly symmetric distribution, centered around $\hat{\xi}_{PML} \sim -0.15$, while regular ML produces a much wider distribution with its peak just below zero. The MM and PWM estimators both produce a bi-modal distribution, with a peak very close to zero and another one between -0.1 and -0.15. To better characterize this result, we used the break-point search based on segmented regression implemented in the `R` package `segmented`[53] to separate the two modes of each distribution. The change points are located at -0.058 for the MM and at -0.049 for the PWM estimator. Fig. 6 shows the two probability distributions of $\hat{\xi}$ for the MM and PWM estimators (left panels), and the L63 attractor colored based on the local value of $\hat{\xi}$ obtained with the two estimators (right panels). More specifically, we separate the two modes of each distribution using the aforementioned break-point algorithm, and we color the histogram depending on the mode to which it is assigned. We then also color the attractor based on the mode to which each state is assigned. While the two modes of each distribution seem to separate similar regions of the attractor, the values are inverted, so that the regions producing the low mode of $\hat{\xi}_{MM}$ correspond to the high mode of $\hat{\xi}_{PWM}$, and vice versa. This is reflected in the negative correlation coefficient of -0.42 between $\hat{\xi}_{MM}$ and $\hat{\xi}_{PWM}$.

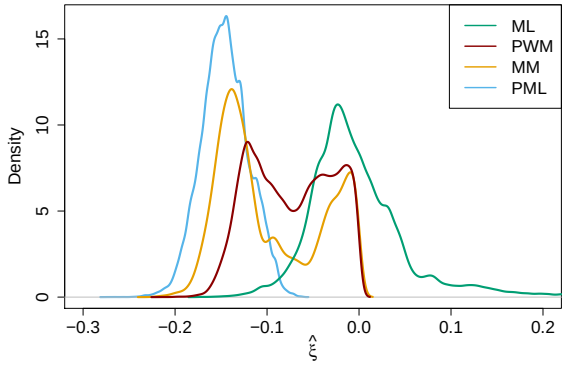

FIG. 5. Probability density functions of the shape parameter ξ for the L63 system estimated using the ML, PWM, MM and PML methods.

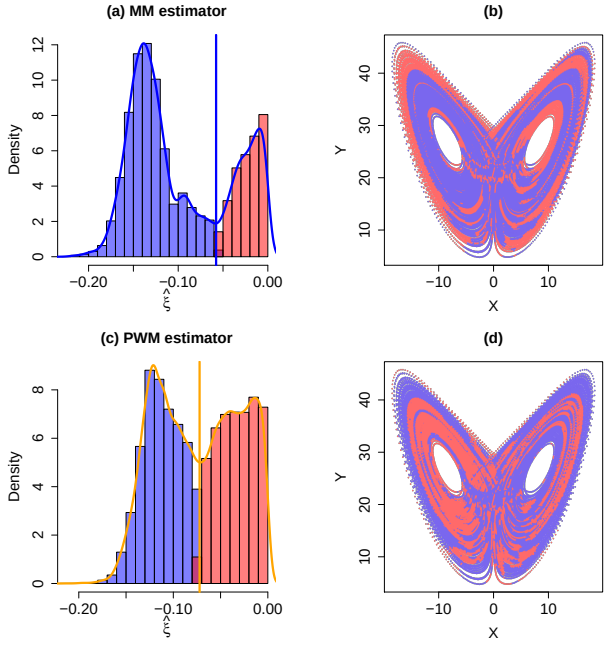

FIG. 6. Bimodal distributions of $\hat{\xi}$ estimated with the MM (a) and PWM (c) methods. The distributions have been partitioned with a breakpoint search algorithm. (b, d) Attractor of the L63 system colored according to the partitioning of the distributions shown in (a) and (c), respectively.

In Fig. 7 we show the probability density functions of the

local dimension d for all the estimators, in logarithmic scale to reduce the visual effect of the large skewness. As expected from the Monte Carlo results, the PML estimator has the worst performance. To obtain the shown results, we have used $\hat{\xi}_{PWM}$ as initial estimate. Using $\hat{\xi}_{MM}$ (not shown) provided almost identical results. The resulting fractal dimension estimate is $\hat{D}_{PML} = 1.801$, the only estimate showing a difference on the first digit.

The estimates of D obtained with the other estimators are overall much closer to the true value. We apply a Student's t test at the 5% level with Bonferroni correction to each mean, under the null hypothesis that $\hat{D} = 2.0627160$ against the bilateral alternative. All estimates are significantly different from the true value, except for $\hat{D}_{MM}$. The mean estimator has the smallest RMSE, thanks to its smaller variance compared to the competitors and to a statistically significant but small difference in estimated global attractor dimension. However, we note that the variability in the estimates of the local dimension in this experiment is partially due to the structure of the attractor, so that this metric is less important than in the case of the Monte Carlo experiment.

Figure 8 shows the $X-Y$ plot of the L63 attractor, colored according to the local dimension obtained with the five estimators. Values are saturated at 3, since larger dimensions are non-physical, but possible due to finite sampling effects. The estimators show comparable performances, with largely superimposed high- and low-dimensional values, although higher dimensional regions of the attractor are less evident using PML estimates. This is confirmed by the large positive correlations between time series of estimated local dimension, ranging from 0.89 for the correlation between $\hat{d}_{PML}$ and $\hat{d}_{PWM}$ to 0.99 for the correlation between $\hat{d}_{ML}$ and both $\hat{d}_{MM}$ and $\hat{d}_{PWM}$. Thus, the differences in the estimation of ξ, especially concerning the bimodality of the MM and PWM estimates discussed above, do not translate into equally visible differences in the estimation of d. These findings corroborate the result from the Monte Carlo experiment: for $\xi \leq 0$, the MM estimator provides the smallest bias and the best bias-variance trade-off.

Since the value of the global attractor dimension for the Lorenz attractor is known, and the low dimensionality of the system allows for an accurate EVT-based estimation as found in[32], the L63 system is a good candidate to check the stability of the estimators with respect to the sample size and the threshold quantile. We consider the same realization of the L63 system described in Section IV B, but sub-sampled with a time step equal to 0.01. Let N be the length of the original time series and L the chosen length for a sub-interval of the same time series: we generate M random indices m from a discrete uniform distribution in $[1, N-L]$, and for each one we consider the sub-sample of length L of the original time series given by $(X(m,m+L), Y(m,m+L), Z(m,m+L))$. For each of these time series we estimate the local and global attractor dimension using all the estimators. We choose the combinations $L = \{100, 500, 1000, 10000\}$ and $M = \{500, 200, 100, 20\}$. We repeat this procedure for three values of the threshold $q = \{0.95, 0.98, 0.99\}$. For $q = 0.99$ we do not consider sample size 100, which results in a single POT value. The distributions of estimated values of D as a function of L, q and the type of estimator are shown in Fig. 9. For the chosen thresholds, all estimators except PML converge to the true value. However, also ML has some convergence issues, with an extremely high variance of the estimates at low sample size, but also at relatively large sample sizes when $q = 0.99$.

Overall, the effect of q appears to be linked to the number of POTs, rather than to how extreme the quantile is. In other words, convergence is better guaranteed when the sample size is large: for example, convergence is reached more easily for $L = 1000$ with $q = 0.95$ than with $q = 0.99$. Following these results, we recommend to choose the highest value of q that guarantees a sufficient POTs sample size for consistent GPD parameter estimation, avoiding the PML estimator, which is affected by a negative bias that does not disappear even at large sample sizes.

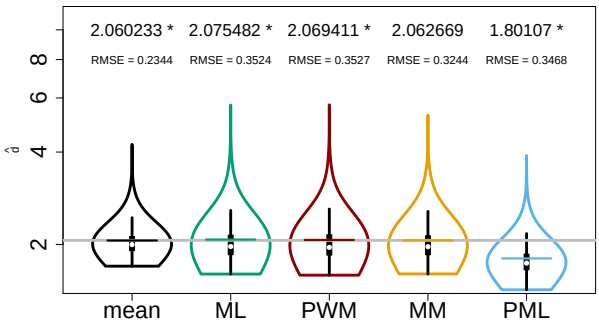

FIG. 7. Probability density functions of $\hat{d}$ for the L63 system in logarithmic scale obtained with different estimators. The horizontal grey line shows the value of the L63 global attractor dimension D_{L63}; horizontal coloured segments show estimates of $\hat{D}$ obtained as the average of all local dimensions. The values are shown above each plot; stars indicate a statistical difference with respect to the theoretical value of D_{L63}. Values of root mean square error are also shown.

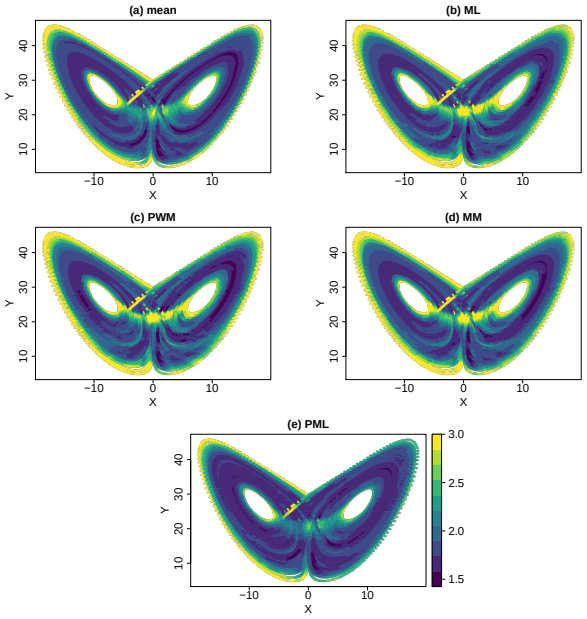

FIG. 8. Attractor of the L63 system colored according to the estimated local dimension obtained with the estimator based on the mean (a), ML (b), PWM (c), MM (d), and PML (e).

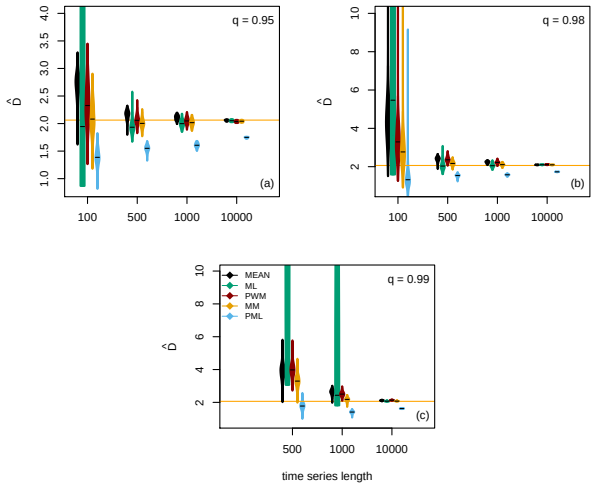

FIG. 9. Probability density functions of the estimated global attractor dimension $\hat{D}$ for the L63 system, for different time series lengths and values of the threshold quantile $q = 0.95$ (a), $q = 0.98$ (b), $q = 0.99$ (c)

C. Climate Data

Finally, we replicate the previous analysis for the ERA5 dataset. In this case, it is not possible to compare estimates to the true value, as in the Monte Carlo experiment, or to previous knowledge about the attractor dimension, as for the L63 system. The goal of considering real-world data is rather to check the stability of dimension estimates when using different estimators. This is particularly relevant because many existing climatological studies have been based on the mean estimator, while the climate system is expected to behave as a non-Axiom A system, characterized by $\xi < 0$. Moreover, time series issued from climate data are non-stationary due to internal variability and changes in external forcing, including the anthropogenic influence in the industrial period.

Figure 10 shows the probability density functions of estimated values of ξ with all methods, except the mean estimator which assumes $\xi = 0$. Qualitatively, the distributions of ξ_{MM} and ξ_{PML} present the highest resemblance, with the main difference due to a secondary peak of estimated values close to 0 for the MM estimator. All possible combinations of a two-sample Kolmogorov-Smirnov test reject the null hypothesis of identical distributions at the 5% significance level. The PWM and ML distributions are shifted towards larger values, with a large peak around 0 in the distribution of $\hat{\xi}_{ML}$.

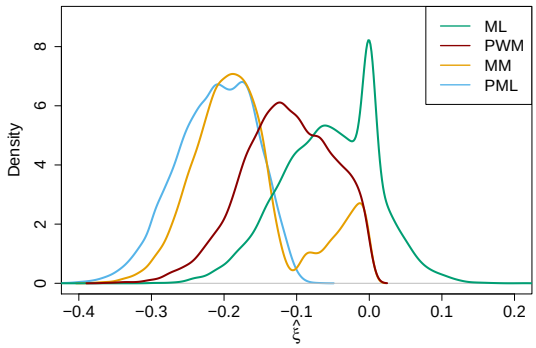

FIG. 10. Probability density functions of the shape parameter ξ for ERA5 500 hPa geopotential height data, estimated using the using the ML, PWM, MM and PML methods.

However, as in the case of L63, the differences in the estimation of ξ do not reflect on the dimension estimates. All the estimators produce values $8 \leq \hat{D} \leq 9$ (see Fig. 11), with the exception of PML which gives an estimate $\hat{D}_{PML} = 7.3$. Also in this case, all possible combinations of a Kolmogorov-Smirnov test reject the null hypothesis of identical distributions at the 5% significance level. We also test the differences among the means, i.e. the estimates of D, through pairwise Student's t tests with Bonferroni correction for multiple test comparisons, without assuming equal variances in the distributions of d. All comparisons reject the null hypothesis of equal means at the 5% level.

While these differences are statistically significant, from the point of view of many climatological applications, the estimates are equivalent. The one possible exception are the estimates produced by PML which are visibly smaller. As an example of this type of application,[54] use a machine learning algorithm to decompose the sea level pressure (slp) anomaly fields over the North Atlantic in a number of coherent structures. Each structure corresponds to a low or high pressure anomaly pattern, and can be seen as a degree of freedom of slp variability. To decide the optimal number of patterns, the

authors construct a few statistics that allow to draw scree plots, and choose 28. They note that this value is compatible with the range of local dimensions estimated for the same atmospheric fields using the mean estimator[18]. In the context of this analysis, where an integer approximation of $\hat{D}$ is used to gain heuristic understanding, the estimates depicted in Fig. 11 can be regarded as compatible, yielding consistent outcomes. The possible exception are the PML estimates, which have both a smaller mean and lower maximum local dimension values than the other estimators. Whether these differences may be overlooked or not would depend on the specific application being considered.

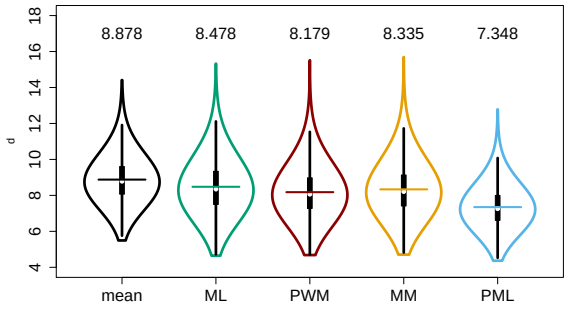

FIG. 11. Probability density functions of $\hat{d}$ obtained using the mean, ML, PWM, MM and PML methods. The horizontal color segments and the numbers above the violin plots show estimates of $\hat{D}$ obtained as the average of all local dimensions.

There are several other studies in the climate science literature that have used estimates of the local dimension. For example, investigating temporal properties of $\hat{d}_m$, such as its seasonality and relation to a dynamical measure of persistence[6,18], or long-term trends in reanalysis and climate projections[9]. These and other studies have typically analysed the local dimension in relative terms. A simple way to check whether biases in the estimates of d may have affected the above results it to verify the similarity between time series of the different estimators using Pearson's correlation. The latter is insensitive to mean shifts or scaling factors, but it attains small values if two time series are not synchronized. The upper triangle of the correlation plot in Fig. 12 shows values of Pearson's correlation between pairs of estimated local dimension time series. Corresponding scatterplots with regression lines are depicted in the lower triangle. The high correlation values and fairly linear scatterplots show that the differences between $\hat{d}_{PML}$ and the other estimators are mainly due to a scale factor.

The correlations among all estimators are positive and large, ranging between 0.88 and 0.97. This shows that, even under the erroneous assumption of $\xi = 0$, the mean estimator produces time series of $\hat{d}_m$ which are well-synchronized with the ones obtained by its competitors that assume a GPD as the exceedances distribution. To ensure that seasonality and modes of interannual variability are consistently reflected by all the estimators, we also inspect the autocorrelation functions (ACFs) and the spectral density functions (SDFs) of the corresponding time series of $\hat{d}$, both shown in Fig. 13. We show the ACF up to 730 daily lags, corresponding to 2 years, and the SDF at periods of up to 4 years. All estimators follow a similar behavior, with a clear yearly cycle. The SDF shows a dominant peak at 2 years, and two secondary peaks at 1 and 3 years. All estimators reflect these dominant frequencies, although the mean estimator shows lower peaks, especially at the yearly frequency. This suggest that the mean estimator may display a subdued seasonal cycle. However, the overall performance of the five methods is largely comparable.

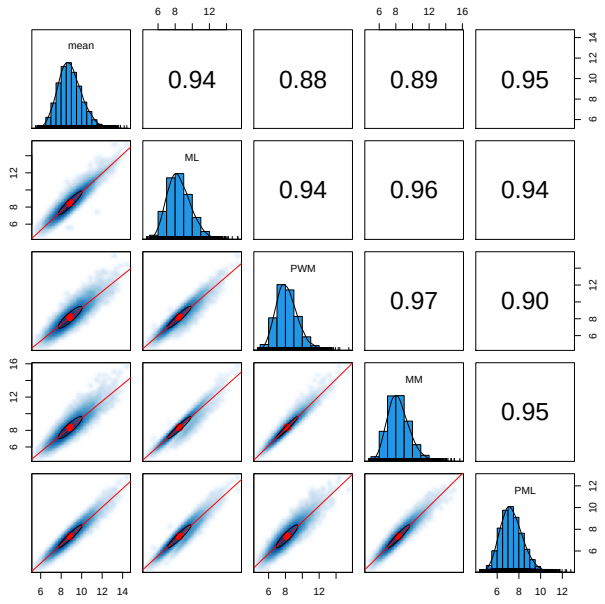

FIG. 12. Correlation plot among the time series of $\hat{d}$ obtained with the five considered estimators. Panels in the upper triangle show pairwise Pearson's correlation coefficients. Panels in the lower triangle show pairwise scatterplots, with superimposed correlation ellipse and linear regression line. The red dots denote the scatterplot centers of mass. Panels along the diagonal show the histograms of the marginal distributions of $\hat{d}$.

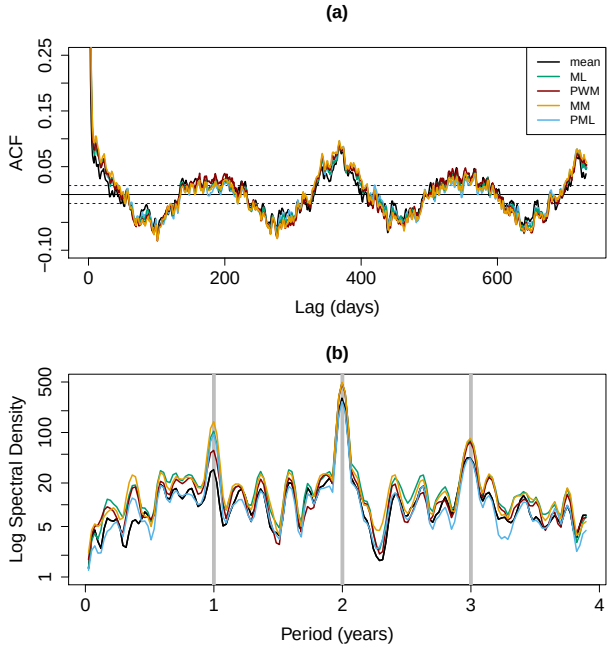

FIG. 13. (a) Global autocorrelation function (ACF) of $\hat{d}$ for all five considered estimators, at lags up to 730 days. Dashed blue lines show the 95% confidence interval constructed under the white noise null hypothesis. (b) Logarithm of the spectral density function (SDF) of $\hat{d}$ for all estimators, at periods up to 4 years. Vertical grey lines mark the main periodicities, showing annual and interannual variability.

VI. DISCUSSION

We have assessed different estimators of the local dimension of a dynamical system, by using extreme value theory for peaks over threshold (POTs). This method also provides a way to estimate the global attractor dimension, which can be obtained as the average over all local dimensions, provided that the observed trajectories are long enough. We have considered four methods assuming that POTs follow a Generalized Pareto Distribution (GPD) with parameters ξ, σ: the method of moments (MM), probability weighted method of moments (PWM), maximum likelihood (ML) and pseudo-ML (PML) estimators. The MM estimation is only defined for $\xi < 0.5$, while the PML estimation requires two different procedures depending on the sign of ξ. We also added a naive estimator, namely the inverse of the POT mean, based on the simplified assumption of exponentially distributed POTs. This is valid in the asymptotic case for Axiom A dynamical systems.

We have performed a Monte Carlo experiment by simulating the POTs directly from a GPD over a realistic range of true parameter values. In this way, we were able to measure the bias of each estimator as a function of the true parameter space. We assessed the performance in terms of the root mean square error and of the relative error on the local dimension, given by the inverse of the GPD scale parameter σ.

In terms of statistical efficiency as measured by the root mean square error (RMSE), the overall best estimates are provided by the PWM. However, for simulated POTs characterized by $\xi < 0$, the MM estimator always produces the best estimates and it is very simple to compute, without any tuning parameter such as the $c = 0.35$ coefficient in the definition of the plotting positions for the PWM. The mean estimator, ML and PML display large RMSEs over varying ranges of ξ, indicating a poorer performance in terms of statistical efficiency. The PML in particular displays very large RMSE values.

The Monte Carlo results also show that the relative error does not depend on the true value of d, and convergence towards a fixed value is generally reached for a POT sample size $\sim$200. The PWM estimator shows the overall best performance, and the MM estimator performs equally well in its domain of existence $\xi \lessapprox 0.5$, despite the theoretical superiority of ML estimators under well-specified statistical models. The PML performance degrades rapidly for growing $\xi \geq -0.2$. The average relative errors of the naive estimator display n almost perfectly linear relationship with the true dimension values as a function of the true value of ξ.

We further test the five estimators when applied to the Lorenz 1963 system – a 3-equation system in a 3-dimensional phase space – and some high-complexity climate data from the ERA5 reanalysis. For the Lorenz system, the global attractor dimension has been previously estimated, so we can use that as a reference value. The results are consistent with those for the Monte Carlo data, with the MM estimator showing the best performance for $\xi < 0$ and the PML estimator showing a negative bias over this range of ξ. For the climate data we have no prior estimate of the dimension, yet we know that the data is characterized by a negative estimated ξ. We thus consider the MM estimator as our baseline. We are particularly interested in testing the performance of the mean estimator, as it has been repeatedly used in the literature but there has been no systematic comparison of how it performs on complex systems relative to alternative estimators. The correlations between the mean and the MM estimator are very close to or even equal to 1, corroborating the idea that, even far from the asymptotic sampling of Axiom A systems, the simple mean estimator can be effectively used when interpreted in a relative sense. Correlations of MM with the other estimators are also very high. For L63, we have also tested the effect of the sample size and of the probability q used to define the threshold quantile. Overall, for *reasonable* choices of q, the effect of the threshold appears to be limited to its role in defining the sample size of the POTs given an initial time series length. The recommendation would be to choose the highest value of q that guarantees a sufficient POT sample size for consistent GPD parameter estimation.

Since for the MM, PWM and PML estimators $\hat{d}$ can be written as a function of $\hat{\xi}$, we have also inspected the performance of the different methods in estimating the shape parameter, to check whether it could be a source of estimation error on d. Results suggest the opposite: the important differences among the distribution of $\hat{\xi}$ do not seem to translate into differences in estimates of $\hat{d}$. At the same time, the two estimators with the most similar distribution of $\hat{\xi}$, i.e. MM and PML, respectively provide the best and worst estimate $\hat{d}$.

Overall, we find that if one is interested only in the esti-

mation of the local attractor dimension (or simply the σ parameter in a different setting), the performance of the chosen estimator on ξ does not seem to be crucial.

To conclude, our results can be used to draw simple, general guidelines for the choice of attractor dimension estimators. If it is known a priori that $\xi < 0$, the MM estimator is the best choice. However, if no prior assumption can be made on the true value of ξ, the mean estimator could be a suitable choice. In contrast, despite the popularity of maximum likelihood in statistical estimation and testing, it appears that the considered PML estimator only performs well in a narrow interval of values of ξ, which does not include ξ values usually observed in climate data. Moreover, the different estimators show a high correlation. When one is interested in relative values of the dimension, the choice of estimator thus does not appear crucial. This is often the case when studying the behaviour of a system through the attractor dimension, as the relationship between dimension values of different states is more important than the value itself. Indeed, dimensions are usually compared to one another to detect changes in system behaviour, rather than interpreted in terms of their absolute values.

SUPPLEMENTARY MATERIAL

The Supplementary Material associated with this article contains all the figures showing the statistical performance of the considered estimators in terms of root mean square error and relative error for the Monte Carlo experiment.

ACKNOWLEDGMENTS

GM acknowledges the support of the European Union's H2020 research and innovation programme under ERC grant no. 948309 (CENÆ). FP and DF received support from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 101003469 (XAIDA). DF and GM received further support from the European Union's Horizon 2020 Marie Sklodowska-Curie grant agreement No. 956396 (EDIPI) and the LEFE-MANU-INSU-CNRS grant "CROIRE".

The authors also thank the two anonymous reviewers for the constructive and insightful comments.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study, the simulation scripts and the data analysis scripts are openly available on Zenodo at https://doi.org/10.5281/zenodo.7944869.

Appendix A TABLES

[1] V. Lucarini, D. Faranda, A. de Freitas, J. de Freitas, M. Holland, T. Kuna, M. Nicol, M. Todd, and S. Vaienti, *Extremes and Recurrence in Dynamical Systems*, Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts (Wiley, 2016).

[2] A. A. Balkema and L. De Haan, "Residual life time at great age," The Annals of probability **2**, 792–804 (1974).

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

[4] M. Leadbetter and H. Rootzen, "Extremal theory for stochastic processes," The Annals of Probability , 431–478 (1988).

[5] R. L. Smith and I. Weissman, "Estimating the extremal index," Journal of the Royal Statistical Society: Series B (Methodological) **56**, 515–528 (1994).

[6] D. Faranda, G. Messori, M. C. Alvarez-Castro, and P. Yiou, "Dynamical properties and extremes of northern hemisphere climate fields over the past 60 years," Nonlinear Processes in Geophysics **24**, 713–725 (2017).

[7] S. Buschow and P. Friederichs, "Local dimension and recurrent circulation patterns in long-term climate simulations," Chaos: An Interdisciplinary Journal of Nonlinear Science **28**, 083124 (2018).

[8] M. Brunetti, J. Kasparian, and C. Vérard, "Co-existing climate attractors in a coupled aquaplanet," Climate Dynamics **53**, 6293–6308 (2019).

[9] D. Faranda, M. C. Alvarez-Castro, G. Messori, D. Rodrigues, and P. Yiou, "The hammam effect or how a warm ocean enhances large scale atmospheric predictability," Nature communications **10**, 1316 (2019).

[10] D. Rodrigues, M. C. Alvarez-Castro, G. Messori, P. Yiou, Y. Robin, and D. Faranda, "Dynamical properties of the north atlantic atmospheric circulation in the past 150 years in cmip5 models and the 20crv2c reanalysis," Journal of Climate **31**, 6097–6111 (2018).

[11] F. Falasca and A. Bracco, "Exploring the tropical pacific manifold in models and observations," Physical Review X **12**, 021054 (2022).

[12] G. Messori, R. Caballero, and D. Faranda, "A dynamical systems approach to studying midlatitude weather extremes," Geophysical Research Letters **44**, 3346–3354 (2017).

[13] S. Scher and G. Messori, "Predicting weather forecast uncertainty with machine learning," Quarterly Journal of the Royal Meteorological Society **144**, 2830–2841 (2018).

[14] A. Hochman, P. Alpert, T. Harpaz, H. Saaroni, and G. Messori, "A new dynamical systems perspective on atmospheric predictability: Eastern mediterranean weather regimes as a case study," Science advances **5**, eaau0936 (2019).

[15] A. Hochman, P. Alpert, P. Kunin, D. Rostkier-Edelstein, T. Harpaz, H. Saaroni, and G. Messori, "The dynamics of cyclones in the twentyfirst century: the eastern mediterranean as an example," Climate Dynamics , 1–14 (2019).

[16] A. Hochman, S. Scher, J. Quinting, J. G. Pinto, and G. Messori, "A new view of heat wave dynamics and predictability over the eastern mediterranean," Earth System Dynamics **12**, 133–149 (2021).

[17] A. Hochman, S. Scher, J. Quinting, J. G. Pinto, and G. Messori, "Dynamics and predictability of cold spells over the eastern mediterranean," Climate Dynamics **58**, 2047–2064 (2022).

[18] D. Faranda, G. Messori, and P. Yiou, "Dynamical proxies of north atlantic predictability and extremes," Scientific reports **7**, 41278 (2017).

[19] D. Faranda, Y. Sato, G. Messori, N. R. Moloney, and P. Yiou, "Minimal dynamical systems model of the northern hemisphere jet stream via embedding of climate data," Earth System Dynamics **10**, 555–567 (2019).

[20] G. Messori, N. Harnik, E. Madonna, O. Lachmy, and D. Faranda, "A dynamical systems characterization of atmospheric jet regimes," Earth System Dynamics **12**, 233–251 (2021).

[21] A. Hochman, G. Messori, J. F. Quinting, J. G. Pinto, and C. M. Grams, "Do atlantic-european weather regimes physically exist?" Geophysical Research Letters **48**, e2021GL095574 (2021).

[22] P. De Luca, G. Messori, F. Pons, and D. Faranda, "Dynamical systems theory sheds new light on compound climate extremes in europe and eastern north america," Quarterly Journal of the Royal Meteorological Society (2020).

[23] K. Giamalaki, C. Beaulieu, S. Henson, A. Martin, H. Kassem, and D. Faranda, "Future intensification of extreme aleutian low events and their climate impacts," Scientific reports **11**, 18395 (2021).

[24] D. Faranda, G. Messori, P. Yiou, S. Thao, F. Pons, and B. Dubrulle, "Dynamical footprints of hurricanes in the tropical dynamics," Chaos: An Interdisciplinary Journal of Nonlinear Science **33**, 013101 (2023).

[25] M. Ginesta, P. Yiou, G. Messori, and D. Faranda, "A methodology for attributing severe extratropical cyclones to climate change based on reanal-

ξ	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	1.0
						n = 20					
mean	29.72	24.31	19.10	14.13	9.71	6.98	7.85	11.03	14.50	17.55	20.06
MM	16.68	15.89	15.25	14.61	13.78	12.98	12.52	13.86	17.58	22.08	26.42
PWM	17.33	16.84	16.56	16.42	16.06	15.83	15.72	15.11	15.14	15.81	17.70
ML	7.27	9.35	12.97	15.25	16.16	16.85	17.93	19.04	20.70	22.01	23.50
PML	11.14	10.65	10.38	10.55	11.49	146.81	103.16	75.64	59.71	54.74	61.16
						n = 100					
mean	28.09	22.60	17.10	11.67	6.45	2.84	5.94	10.91	15.68	19.61	22.52
MM	6.91	6.66	6.38	6.12	5.95	5.91	6.08	8.07	14.49	22.23	28.92
PWM	6.96	6.86	6.75	6.65	6.64	6.62	6.70	6.91	7.21	8.02	10.92
ML	1.50	5.20	8.77	16.42	10.21	5.63	6.67	7.33	7.85	8.28	8.80
PML	4.92	4.76	4.69	5.04	6.29	85.97	60.54	43.70	32.95	26.70	24.06
						n = 1000					
mean	27.74	22.20	16.67	11.14	5.64	0.87	5.58	11.06	16.39	20.97	24.23
MM	2.13	2.06	1.98	1.91	1.87	1.88	2.44	4.05	12.06	23.20	32.24
PWM	2.14	2.11	2.08	2.06	2.06	2.04	2.10	2.29	2.84	3.58	6.20
ML	0.75	1.50	1.89	2.93	8.45	1.66	2.09	2.24	2.41	2.54	2.69
PML	1.55	1.50	1.49	1.74	2.95	80.94	56.75	40.24	28.89	21.12	15.95

TABLE I. Root mean square error of the five considered estimators of the local dimension, averaged over d and shown as a function of ξ, for sample sizes n = 20, 100, 1000.

ysis data: the case study of storm alex 2020," Climate Dynamics , 1–25 (2022).

[26] G. Messori and D. Faranda, "Characterising and comparing different palaeoclimates with dynamical systems theory," Climate of the Past **17**, 545–563 (2021).

[27] A. Gualandi, J.-P. Avouac, S. Michel, and D. Faranda, "The predictable chaos of slow earthquakes," Science advances **6**, eaaz5548 (2020).

[28] F. Hausdorff, "Dimension und äußeres maß," Mathematische Annalen **79**, 157–179 (1918).

[29] C. Nicolis and G. Nicolis, "Is there a climatic attractor?" Nature **311**, 529–532 (1984).

[30] P. Grassberger, "Do climatic attractors exist?" Nature **323**, 609 (1986).

[31] E. N. Lorenz, "Dimension of weather and climate attractors," Nature **353**, 241–244 (1991).

[32] F. M. E. Pons, G. Messori, M. C. Alvarez-Castro, and D. Faranda, "Sampling hyperspheres via extreme value theory: implications for measuring attractor dimensions," Journal of statistical physics **179**, 1698–1717 (2020).

[33] A. C. M. Freitas, J. M. Freitas, and M. Todd, "Hitting time statistics and extreme value theory," Probability Theory and Related Fields **147**, 675–710 (2010).

[34] D. Faranda, V. Lucarini, G. Turchetti, and S. Vaienti, "Numerical convergence of the block-maxima approach to the generalized extreme value distribution," J. Stat. Phys. **145**, 1156–1180 (2011).

[35] T. Caby, D. Faranda, G. Mantica, S. Vaienti, and P. Yiou, "Generalized dimensions, large deviations and the distribution of rare events," arXiv preprint arXiv:1812.00036 (2018).

[36] T. Carletti and S. Galatolo, "Numerical estimates of local dimension by waiting time and quantitative recurrence," Physica A: Statistical Mechanics and its Applications **364**, 120–128 (2006).

[37] R. Vitolo, M. P. Holland, and C. A. Ferro, "Robust extremes in chaotic deterministic systems," Chaos: An Interdisciplinary Journal of Nonlinear Science **19**, 043127 (2009).

[38] P. de Zea Bermudez and S. Kotz, "Parameter estimation of the generalized pareto distribution—part i," Journal of Statistical Planning and Inference **140**, 1353–1373 (2010).

[39] P. de Zea Bermudez and S. Kotz, "Parameter estimation of the generalized pareto distribution—part ii," Journal of Statistical Planning and Inference **140**, 1374–1388 (2010).

[40] R. Bellman, "Dynamic programming," Princeton university press, Princeton (1957).

[41] R. A. Fisher, "On an absolute criterion for fitting frequency curves," Messenger of Mathematics **41**, 155–156 (1912).

[42] G. Casella and R. L. Berger, *Statistical inference*, Vol. 2 (Duxbury Pacific Grove, CA, 2002).

[43] L. Belzile *et al.*, *mev: Modelling Extreme Values* (2022), r package version 1.14.

[44] S. D. Grimshaw, "Computing maximum likelihood estimates for the generalized pareto distribution," Technometrics **35**, 185–191 (1993).

[45] J. A. Villaseñor-Alva and E. González-Estrada, "A bootstrap goodness of fit test for the generalized pareto distribution," Computational statistics & data analysis **53**, 3835–3841 (2009).

[46] J. A. Villaseñor-Alva and E. González-Estrada, *gPdtest: Bootstrap goodness-of-fit test for the generalized Pareto distribution* (2012), r package version 0.4.

[47] J. M. Landwehr, N. Matalas, and J. Wallis, "Probability weighted moments compared with some traditional techniques in estimating gumbel parameters and quantiles," Water resources research **15**, 1055–1064 (1979).

[48] J. R. M. Hosking, J. R. Wallis, and E. F. Wood, "Estimation of the generalized extreme-value distribution by the method of probability-weighted moments," Technometrics **27**, 251–261 (1985).

[49] E. N. Lorenz, "Deterministic nonperiodic flow," J. Atmos. Sci. **20**, 130–141 (1963).

[50] H. Hersbach, B. Bell, P. Berrisford, S. Hirahara, A. Horányi, J. Muñoz-Sabater, J. Nicolas, C. Peubey, R. Radu, D. Schepers, *et al.*, "The era5 global reanalysis," Quarterly Journal of the Royal Meteorological Society **146**, 1999–2049 (2020).

[51] P. Grassberger and I. Procaccia, "Characterization of strange attractors," Phys. Rev. Lett. **50**, 346 (1983).

[52] D. Viswanath, "The fractal property of the lorenz attractor," Physica D: Nonlinear Phenomena **190**, 115–128 (2004).

[53] V. M. Muggeo, "Estimating regression models with unknown breakpoints." Statistics in Medicine **22**, 3055–3071 (2003).

[54] L. Fery, B. Dubrulle, B. Podvin, F. Pons, and D. Faranda, "Learning a weather dictionary of atmospheric patterns using latent dirichlet allocation,"

ξ	-1.0	-0.8	-0.6	-0.4	-0.2	0.0	0.2	0.4	0.6	0.8	1.0
	n = 20										
mean	1.035	0.836	0.639	0.441	0.246	0.052	-0.134	-0.308	-0.461	-0.589	-0.691
MM	0.061	0.055	0.051	0.042	0.026	-0.011	-0.078	-0.185	-0.314	-0.445	-0.564
PWM	0.082	0.075	0.070	0.065	0.057	0.043	0.025	-0.011	-0.068	-0.151	-0.250
ML	0.084	-0.038	-0.091	-0.085	-0.054	-0.032	-0.010	0.007	0.023	0.036	0.049
PML	-0.017	-0.040	-0.071	-0.116	-0.177	2.834	1.920	1.312	0.894	0.617	0.449
	n = 100										
mean	1.007	0.808	0.607	0.408	0.209	0.010	-0.187	-0.378	-0.552	-0.698	-0.807
MM	0.012	0.012	0.010	0.008	0.006	-0.004	-0.040	-0.140	-0.308	-0.498	-0.663
PWM	0.016	0.016	0.014	0.012	0.012	0.009	0.004	-0.007	-0.036	-0.098	-0.200
ML	0.023	-0.088	-0.121	0.034	0.085	0.017	0.000	0.001	0.005	0.008	0.010
PML	-0.003	-0.012	-0.029	-0.060	-0.112	1.997	1.388	0.972	0.686	0.489	0.353
	n = 1000										
mean	1.001	0.801	0.601	0.401	0.201	0.001	-0.199	-0.397	-0.589	-0.755	-0.873
MM	0.001	0.001	0.001	0.001	0.001	0.000	-0.008	-0.076	-0.275	-0.537	-0.754
PWM	0.002	0.001	0.002	0.001	0.001	0.001	0.000	-0.001	-0.010	-0.045	-0.131
ML	0.009	-0.015	-0.024	-0.033	0.078	0.005	0.000	0.000	0.001	0.001	0.001
PML	0.000	-0.002	-0.008	-0.023	-0.060	1.905	1.334	0.943	0.672	0.483	0.351

TABLE II. Relative error of the five considered estimators of the local dimension, averaged over d and shown as a function of ξ, for sample sizes n = 20, 100, 1000.