HAL
open science

Sampling hyperspheres via extreme value theory: implications for measuring attractor dimensions

Flavio Maria Emanuele Pons, Gabriele Messori, M Carmen Alvarez-Castro, Davide Faranda

To cite this version:

Flavio Maria Emanuele Pons, Gabriele Messori, M Carmen Alvarez-Castro, Davide Faranda. Sampling hyperspheres via extreme value theory: implications for measuring attractor dimensions. Journal of Statistical Physics, 2020, 10.1007/s10955-020-02573-5 . hal-01650250v3

HAL Id: hal-01650250

https://hal.science/hal-01650250v3

Submitted on 6 Jun 2019

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

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

Sampling hyperspheres via extreme value theory: implications for measuring attractor dimensions

Flavio Maria Emanuele Pons *[1], Gabriele Messori[1,2,3], M. Carmen Alvarez-Castro[1,4], and Davide Faranda[1,5]

[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 Meteorology and Bolin Centre for Climate Research, Stockholm University, 106 91, Stockholm, Sweden*

[3] *Department of Earth Sciences, Uppsala University, Uppsala, Sweden*

[4] *Climate Simulation and Prediction Division, Centro Euro-Mediterraneo sui Cambiamenti Climatici, Bologna, 40127, Italy*

[5] *London Mathematical Laboratory, 14 Buckingham Street, London, WC2N 6DF, UK*

Abstract

The attractor dimension is an important quantity in information theory, as it is related to the number of effective degrees of freedom of the underlying dynamical system. By using the link between extreme value theory and Poincaré recurrences, it is possible to compute this quantity from time series of high-dimensional systems without embedding the data. In general $d < n$, where n is the dimension of the full phase-space, as the dynamics freezes some of the available degrees of freedom. This is equivalent to constraining trajectories on a compact object in phase space, namely the attractor. Information theory shows that the equality $d = n$ holds for random systems. However, applying extreme value theory, we show

*Corresponding author
email: `flavio.pons@lsce.ipsl.fr`

that this result cannot be recovered and that $d < n$. We attribute this effect to the curse of dimensionality, and in particular to the phenomenon of concentration of the norm observed in high-dimensional systems. We derive a theoretical expression for $d(n)$ for Gaussian random vectors, and we show numerically that similar curse of dimensionality effects are found for random systems characterized by non-Gaussian distributions. Finally, we show that the effect of the curse of dimensionality can be quantified using the extreme value theory, thus enabling to retrieve the degree of non-randomness of a system. We provide examples issued from real-world climate and financial datasets.

1 Introduction

Ever-increasing computational capabilities have made large, high-dimensional datasets tractable. In a similar way, theoretical and technological advances have overcome many of the algorithmic limitations to inferring the statistical and dynamical properties of the systems underlying those data Mayer-Kress (2012). The most outstanding example of such technological progress is the widespread use of machine learning and data mining techniques since the 80s Samuel (1988). Dynamical system algorithms have benefited from the same progress: tasks such as the computation of dynamical indicators (e.g. Lyapunov exponents, generalized dimensions, entropies) for large spatio-temporal datasets are now feasible, at least from a technical point of view Schubert and Lucarini (2015); Vannitsem and Lucarini (2016).

However, such impressive development has yet to solve the so-called curse of dimensionality [1]; on the contrary, given the large availability of high-dimensional datasets, the curse has become ever more ubiquitous. The difficulty in analyzing high-dimensional data results from the fact that algorithms usually designed and tested in 2 or 3 dimensions are then ported to high dimensional spaces, whose geometrical properties are different and often counterintuitive Verleysen and François (2005). The effects of the curse of dimensionality can be mitigated whenever it is possible

[1] For a definition of the curse of dimensionality we quote Cabestany et al. (2005) which defines it as: "the expression of all phenomena that appear with high-dimensional data, and that have most often unfortunate consequences on the behavior and performances of learning algorithms"

to select a subset of features, or to project the data onto a lower dimensional space (e.g. EOFs), as usually happens in data mining problems. On the contrary, algorithms aiming at measuring properties of the system underlying the data unfortunately depend on the phase-space topology. This is the case of the estimation of the Lyapunov exponents, the entropy and the dimensions of attractors in complex systems.

In this paper, we analyse the properties of an estimator of the Hausdorff dimension d. The algorithms used to measure such a quantity were first devised in the 80s and their developers were well-aware of the curse of dimensionality. The most widespread algorithm to estimate the Hausdorff dimension was, at that time, based on the nearest-neighbors search Grassberger (1988) in a space with dimension $k \ll n$, where n indicates the number of variables (or, equivalently, the dimension of the full phase space). This technique was very successful in determining the Hausdorff dimension of low-dimensional attractors, such as the Henon map, the Lorenz and the Rossler flows, but it showed severe drawbacks when applied to high dimensional systems. It was a common belief that the limitations in measuring d in a reliable way were mostly due to the choice of k, known as embedding dimension. The key observation, unveiled by a series of studies aimed at determining the dimension of the climate attractor Grassberger and Procaccia (1984); Nicolis and Nicolis (1984); Grassberger (1986); Lorenz (1991), was that for $k > 6$, the quality of the neighbors in phase space drops significantly, and estimates of d are flawed and trivially equal to k.

A recently developed technique, based on the application of extreme value theory (EVT) to dynamical systems, allows to estimate d without introducing the embedding dimension. The founding idea is that, under suitable rescaling, the recurrences around a state ζ in the phase space are distributed according to the generalized extreme value distribution or the generalized Pareto distribution. This technique does not measure d directly but, by sampling recurrences around ζ, it estimates the so-called local dimension $d_l(\zeta)$, namely the scaling of the hypersphere centered around the point ζ in phase space. When a sufficient number of states ζ is considered, those hyperspheres densely cover the attractor and, by averaging d_l over ζ one obtains an estimate for

d.

This EVT approach has been found as reliable as nearest-neighbors algorithms for low dimensional systems Lucarini et al. (2016), and has also been applied to high-dimensional systems. In Faranda et al. (2017), it has been shown that for atmospheric data with $n \simeq 1000$, $d \simeq 12$ with $5 < d_l(\zeta) < 25$ depending on ζ. In that study, the main goal was determining whether differences between $d_l(\zeta)$ were physically meaningful, and little attention was given to the actual value of d. However, it is natural to ask how much we can trust the estimates of d provided by the EVT technique. This question was first addressed in Buschow and Friederichs (2018). There, the authors found that estimates of d provided in Faranda et al. (2017) were consistent with those obtained in much longer climate model runs, but that the value of d slowly increased with increasing length of the simulation, instead of settling on a constant value. This provided an indication that the estimates of d provided by EVT are also affected by the curse of dimensionality.

In this paper, we investigate this problem in a systematic way, both theoretically and numerically. First, we consider the estimation of d in the case of random vectors, for which Rényi et al. (1961) proved that $d = n$. Through numerical simulation, we show that the estimates $\hat{d}$ obtained using EVT are $\hat{d} < n$, and that the functional form of $\hat{d}(n)$ is the same among different distributions, thus providing a universal scaling. We derive an analytic expression for such scaling in the case of Gaussian random vectors. Moreover, we estimate d for selected real dynamical systems, showing that the scaling is preserved, even though observed values are lower than in the case of a random vector with the same marginal distribution. Finally, we argue that our results enable us to retrieve the degree of non-randomness of a dynamical system.

2 Theoretical framework

Recent theoretical advances in our understanding of the limiting distribution of Poincaré recurrences enable us to compute both mean and local (in phase space) dynamical properties of complex systems. The key idea is that for Axiom A systems, under suitable rescaling, the probability p of entering a hypersphere centred on a state ζ with a radius r in phase space obeys a generalized

Pareto distribution whose parameters are linked to dynamical properties, such as d_l and the Lyapunov exponents Freitas et al. (2010); Faranda et al. (2011); Lucarini et al. (2012, 2016).

Here we recall the basic procedure to obtain such scaling laws and we refer to Lucarini et al. (2016) for further details. Let as consider an Axiom A dynamical system $\dot{x} = F(x)$ on a compact manifold $X \subset \mathbb{R}^n$ (phase space), where $x(t) = f^t(x_{in})$, with $x(t=0) = x_{in} \in X$ initial condition and f^t evolution operator, defined for all $t \in \mathbb{R}\geq 0$. Let us define Ω as the attracting invariant set of the dynamical system, so that μ is the associated SRB measure supported in $\Omega = \mathrm{supp}(\mu)$. We consider distance observables that can be expressed as functions $g : X \to \mathbb{R} \cup \{+\infty\}$ written as $g(\mathrm{dist}(x(t), \zeta)) \geq 0$, where $\zeta \in \Omega$ is a reference state. In order to compute the probability p of entering a hypersphere in phase space centred on a state ζ, we first calculate the series of distances $\mathrm{dist}(\zeta, x(t))$ between the state ζ and all other states $x(t)$ on the trajectory. We then introduce a logarithmic weight on the time series of the distances to increase the discrimination of small values of $\mathrm{dist}(\zeta, x(t))$, which correspond to large values of $g(x(t))$:

$$g(x(t)) = -\log(\mathrm{dist}(\zeta, x(t))).$$

The probability of entering a hypersphere of radius r centred on ζ can now be expressed as the probability p of exceeding a threshold corresponding to a high quantile g_q of the distribution of $g(x(t))$. In the limit of $t \to \infty$, such probability is the exponential member of the generalized Pareto distribution:

$$p = \Pr(g(x(t)) > g_q, \zeta) \simeq \exp(-[g(x(t)) - \nu(\zeta)]/\sigma(\zeta)),$$

whose parameters ν and σ are a function of the state ζ chosen on the attractor. Remarkably, $\sigma = 1/d_l(\zeta)$, where $d_l(\zeta)$ is the local dimension around the state ζ. By averaging d_l over a sufficiently large ensemble of states ζ on the attractor, one then obtains the attractor dimension d. The universality of the convergence law implies that the above is akin to a central limit theorem of Poincaré recurrences.

The step based on the computation of $\mathrm{dist}(\zeta, x(t))$ is the origin of the curse of dimensionality. Problems arise due to the so-called concentration of the norm, i.e. the fact that, in sufficiently high-dimensional spaces, the ratio between L^p norms of the nearest and farthest points from the reference state ζ tends to 1, as stressed by Verleysen and François (2005). This reflects on the shape of the distribution of the computed distances, and then on the functional $g(x(t))$. In the following section, we will discuss how the shape of this distribution can be linked to estimates of d, and in particular we will show that its standard deviation modulates the loss of degrees of freedom with respect to the theoretical value $d = n$ in the case of random vectors.

In deterministic chaotic systems, the asymptotic results ($t \to \infty$) predict that $d_l(\zeta) = d$ for all ζs except a measure zero set, including the unstable fixed points of the dynamics. Note that, for t finite, a broader distribution of d_l is recovered, as the dynamics around unstable fixed points affects finite, yet potentially extensive, regions of the phase space Faranda et al. (2017); Caby et al. (2018). In Faranda et al. (2013), the case of stochastically perturbed dynamical systems has been analyzed, showing that, with the addition of noise and in the $t \to \infty$ limit, $d_l(\zeta) = d$ everywhere. This implies that the dependence on ζ for random vectors can be dropped as any ζ is statistically equivalent. We can therefore apply this framework to the random vectors considered in this study.

3 Asymptotic results for Gaussian random vectors

As first step, we compare estimates of d to the theoretical result provided by Rényi et al. (1961) stating that, for n-dimensional random vectors, $d = n$. To this purpose, we consider random vectors such that all finite dimensional distributions $F_{t_1,\ldots,t_s}$ of the recurrences are Gaussian $\forall\ s$. In the following, the symbol $\hat{z}$ will denote the estimator for the generic quantity z, and $E()$ will denote statistical expectation.

Let $j = 1, \ldots, T$ denote time and $i = 1, \ldots, n$ the sampling points. Let $X_{j,i}$ be T i.i.d.

(independent and identically distributed) realizations in time of the n-dimensional random vector (or field) and, consistently with the notation introduced above, let us denote the reference state $\zeta = \{\zeta_i\}_{i=1}^n$.

Let us now assume that, at any time j, the vector of recurrences Z_j is an n-dimensional Gaussian random vector with zero mean and covariance matrix $\Sigma = I$. This implies that $Z_{j,i} = X_{j,i} - \zeta_i$ are n i.i.d. standard Gaussian random variables. The result for variables with non-zero mean and non-unit variances follows applying standardized instead of simple Euclidean distance. We define $\delta^2(n) = \{\delta_j^2(n)\}_{j=1}^T$ as the T-dimensional vector of the squared Euclidean distances between each realized field $X_{j,i}$ and the reference field ζ_i. It is straightforward to see that $\delta_j^2(n)$ has a Chi-squared distribution with n degrees of freedom:

$$\delta_j^2(n) = Z_j^T Z_j = \sum_{i=1}^{n} (X_{j,i} - \zeta_i)^2 \simeq \chi^2(n). \tag{1}$$

We now consider the functional of the Euclidean distance (L^2 norm) $g(n) = -\log\sqrt{\delta^2(n)}$ Freitas et al. (2010), implying that all of our results will be valid with respect to this metric. Then, the probability density function f_g of $g(n)$ reads:

$$f_g = f_{\delta^2}(h^{-1}(\delta^2)) \left| \frac{d\delta^2}{dg} \right| = \frac{2^{1-\frac{n}{2}}}{\Gamma\left(\frac{n}{2}\right)} \exp\left\{ -ng - \frac{1}{2} e^{-2g} \right\}, \tag{2}$$

where $h(\delta^2(n)) = -\frac{1}{2}\log\delta^2(n)$, $\left|\frac{d\delta^2}{dg}\right| = |-2e^{-2g}|$, and $\Gamma(\cdot)$ is the Euler's Gamma function. Let q be an extreme percentile, corresponding to a threshold value g_q. We consider the Peaks Over Threshold (POT) extremes in the $(1-q)$ right tail of the distribution, denoted $Y = \{Y_\ell\}_{\ell=1}^m$, where $\ell = 1, \ldots, m$ is the number of exceedances, shifted so that $\liminf\{Y\} = 0$, with expected sample size $m = (1-q)T$. From EVT Leadbetter et al. (2012), we know that $Y \sim Exp(\lambda)$, so that $f_Y(y) = \lambda^{-1} e^{-y/\lambda}$. The mean and standard deviation of the Exponential distribution are $\mu_Y = \sigma_Y = \lambda$, and it is knownFreitas et al. (2010) that the local dimension around state ζ is given by $d_l = \lambda^{-1}$. Notice that both μ_Y and σ_Y are functions of n, as will be detailed later in this section. The unbiased maximum likelihood estimator for λ is the sample mean, $\hat{\lambda} = \frac{1}{m}\sum_{\ell=1}^m y_\ell = \overline{y}$, so that the local dimension can be obtained as $\hat{d}_l = \overline{y}^{-1}$.

As a second step, we determine the link between λ and n. Since the cumulative distribution function of g is not invertible, we cannot rely on the quantile function for this purpose. Instead, we approximate λ as a function of known moments of g. From the properties of the Exponential distribution, we know that the mean value λ corresponds to the percentile $(1-e^{-1})$. On the support of g, the mean value of the extremes is given by g_λ and corresponds to the $q + (1-q)(1-e^{-1})$ percentile of the distribution. When applying POT to the sample, the threshold g_q (and then also g_λ) can be seen as the point on the support of g located in the right tail of the distribution, at a distance from the mean corresponding to a certain number k of standard deviations:

$$g_\lambda(n) = \mu_g(n) + k(q,n)\sigma_g(n). \tag{3}$$

Therefore, we can infer the functional dependence of λ on n considering the first two moments of g. The moment generating function of g reads: $M_g(t) = E[e^{tg}] = \frac{2^{-t/2}\Gamma(\frac{n-t}{2})}{\Gamma(\frac{n}{2})}$, and the first two moments can be obtained as $E(g) = M_g'(t=0)$ and $E(g^2) = M_g''(t=0)$, so that:

$$\mu_g(n) = E(g) = -\frac{1}{2}\left[\log 2 + \psi\left(\frac{n}{2}\right)\right] \tag{4}$$

$$\sigma_g(n) = \sqrt{E(g^2) - E(g)^2} = \frac{1}{2}\sqrt{\psi^1\left(\frac{n}{2}\right)} \tag{5}$$

where apexes denote differentiation, $\psi(z) = \Gamma'(z)/\Gamma(z) = \frac{d}{dz}\log\Gamma(z)$ is the digamma function and $\psi^1(z) = \frac{d}{dz}\psi(z) = \frac{d^2}{dz^2}\log\Gamma(z)$ is the trigamma function.

Plugging Equations 4 and 5 into Equation 3, we obtain an expression for $E(d(n))$ for Gaussian random vectors as a function of n, up to the knowledge of $k(q,n)$. The latter, as the expected value of the extremes itself, could be known exactly only given the quantile function of g. However, we can proceed heuristically in order to investigate the dependence of k on q and n. In particular, we move from the consideration that any percentile q on a distribution has its sample counterpart in the r-th order statistics. Thus $r/T \simeq q$, where T is the sample size. Hence, we can exploit

results provided in Balakrishnan et al. (2003) about tight bounds for the expected value of the r-th sample order statistics $g_{r:T}$, drawn from a distribution with mean μ_g and standard deviation σ_g. Since we are dealing with POTs in the right tail of the distribution, we consider the upper bound, which reads $g_{r:T} \leq \mu_g + \sigma_g \sqrt{\frac{r-1}{T-r-1}}$. As $T \to \infty$, $(r-1)/(T-r-1) \to q/(1-q)$, so that we expect that, given n, k is a monotonic function of $\sqrt{(q)/(1-q)}$. The linear relation between the expected value of the extremes and the standard deviation σ_g is preserved also in the asymptotic case $q \to 1$: in fact Arnold et al. (1979) proved that, for the sample maximum, $g_{T:T} \leq \sqrt{T-1}$.

In summary, for the Gaussian case, we expect to observe two empirical facts. First, that the dimension d must scale as $\psi^1(n/2)$; second, that the scaling factor must be a monotonic function of $\sqrt{(q)/(1-q)}$. In the next section, we investigate numerically the robustness of these results, and the extent to which they are valid for non-Gaussian random vectors and for datasets issued from real-world dynamical systems.

4 Numerical Results for Random Vectors

4.1 Gaussian random vectors

In order to test the statistical robustness of the results derived in the previous section, we perform a Monte Carlo (MC) experiment. We generate $Z_j \sim N(0,1)$ with topological dimension $n \in [2, 500]$ and an interval of threshold percentiles $q \in [0.980, 0.999]$, partitioned with a grid step $\delta q = 0.0005$. We fix the number of MC replicates for the extremes to $m = 10000$, so that the number of generated samples depends on the percentile. All the p-values and coefficients of determination R^2 discussed in the following refer to ordinary least squares (OLS) estimates of the corresponding regression equation.

In the left panel of Fig. 1, we show a comparison between estimates $\hat{d}$ (black dots) and the corresponding expected values $E(\hat{d})$ derived from the combination of Eq.s 3, 4 and 5 (grey surface), as a function of q and n. We then obtain OLS estimates $\hat{k}(q,n)$ from Eq. 3-5, with $R^2 > 0.989$

for all cases. The plot of $\hat{k}(q,n)$ against q for each n suggests a natural logarithmic link between $k(q,n)$ and $\sqrt{q/(1-q)}$, so that we can introduce two parameters $\alpha_0(n)$ and $\beta_0(n)$ and write:

$$k(q,n) \simeq \alpha_0(n) + \beta_0(n) \log \sqrt{(q)/(1-q)}. \tag{6}$$

In the right panel of Fig. 1, we compare observed values of $\hat{k}(q,n)$ (black dots) to the predictions provided by Eq. 6 (grey surface). The accuracy of Eq. 6 is remarkably high, with $R^2 > 0.9982$ for every n. Therefore, we can write:

$$g_\lambda(n) = \mu_g(n) + \left[\alpha_0(n) + \beta_0(n) \log\left(\sqrt{(q)/(1-q)}\right)\right] \sigma_g, \tag{7}$$

and we thus obtain an expression for the expected value of $\hat{d}$ as:

$$E(\hat{d}) = \left\{ -\frac{1}{2}\left[\log 2 + \psi\left(\frac{n}{2}\right)\right] + \frac{1}{2}\left[\alpha_0(n) + \beta_0(n) \log\left(\sqrt{\frac{q}{1-q}}\right)\right] \sqrt{\psi^1\left(\frac{n}{2}\right)} - g_q(n) \right\}^{-1}. \tag{8}$$

Eq. 8 constitutes an upper bound for the expected dimension (interpretable also as the number of active degrees of freedom) of any random vector with Gaussian but not i.i.d. Z. This covers both the case of random vectors with non-diagonal covariance matrix, and of data issued from real dynamical systems featuring nonlinear dependencies among the variables. We also observe that the two parameters $\alpha_0(n)$ and $\beta_0(n)$ are fast-decaying functions of n, and both appear to reach a constant value for $n \gtrsim 100$.

An inspection of the differences between computed and predicted $k(n,q)$, shown in Fig. 2, suggests the existence of an unexplained residual effect producing small mean absolute relative differences ($\simeq 3 \cdot 10^{-3}$). Since $k(n,q)$ multiplies σ_g, such error has a non-negligible effect for very large values of n, when also σ_g becomes large. This results in a negative bias on $E(\hat{d})$ at large values of n when using Eq. 8 instead of simply combining Eq.s 3, 4 and 5. Interestingly, we find that the effect of n on $k(q,n)$ can be neglected by specifying an alternative to Eq. 8 as a linear model in $[\psi^1(n/2)]^{-\frac{1}{2}}$, with an intercept depending on the threshold percentile q:

$$E(\hat{d}) \simeq \alpha_1(q) + \beta_1(q) \left[\psi^1\left(\frac{n}{2}\right)\right]^{-\frac{1}{2}}. \tag{9}$$

For all the considered percentiles in [0.980,0.999], we estimate the regression in Eq. 9, which provides a very accurate prediction of the number of active degrees of freedom, with linear determination coefficients $R^2 > 0.9984$ for all q (p-values $< 10^{-23}$ associated to β_1 for every q).

Concerning the functional dependence of β_1 on q, we find that a model $\beta_1(q) = \beta_1^{(0)} + \beta_1^{(1)} \log\left(\sqrt{(q)/(1-q)}\right)$ results in a high goodness-of-fit ($R^2 = 0.996$), indicating that β_1 in Eq. 9 is characterized by the same functional link to the threshold percentile as $k(q, n)$. The intercept $\alpha(q)$ lacks a physical meaning and appears to be a quadratic function of q: $\alpha_1(q) = \alpha_1^{(0)} + \alpha_1^{(1)} q + \alpha_1^{(2)} q^2$, even though the goodness of fit is lower ($R^2 = 0.887$). The estimated coefficients are $\alpha_1^{(0)} = -7172$, $\alpha_1^{(1)} = 14583$, $\alpha_1^{(2)} = -7417$, $\beta_1^{(0)} = 1.62$, $\beta_1^{(1)} = 0.277$. In Fig. 3 we display the results referred to the estimation of Eq. 9 for all the considered q. The in-set plot highlights the lowest (red) and the highest (green) curves representing $q = 0.98$ and $q = 0.999$, respectively; black circles mark the corresponding observed values.

4.2 Generalization to symmetric non-Gaussian random vectors

We next consider random systems characterized by symmetric non-Gaussian distributions with different values of the kurtosis κ, which determines the flatness of the probability density function. The Freitas-Freitas-Todd theorem specifies that $g(\cdot)$ must have a global maximum at 0 Freitas et al. (2010). There are therefore two extreme cases: i) systems with uniformly distributed δ^2 in a fixed interval, so that the global maximum is at all the points of the interval and ii) systems with δ^2 distributions characterized by a sharp central maximum, degenerating in a Dirac delta. The results for systems with different probability density functions are displayed in Fig. 4, where we show curves representing $\hat{d}$ as a function of a set of values of n. When the maximum in zero is sharper than for the Gaussian distribution (e.g. Student's t distribution), the curve is lower than

the analytic Gaussian curse of dimensionality curve. Conversely, when the maximum is flatter (e.g. Beta distribution, with the uniform as a particular case), the curve is higher than the reference Gaussian curve. The uniform limiting case displays the highest values, as predicted by information theory (see Rényi et al. (1961)). In other words, flatter distributions (with very light tails) are more efficient in exploring the phase space with respect to systems with very peaked δ^2, thus being less affected by the curse of dimensionality. To verify this claim, we devise a MC experiment testing whether for a fixed quantile q, β_1 is a function of kurtosis. We repeat the MC experiment with the same parameters as for the Gaussian case discussed in the previous section, but fixing $q = 0.98$. We explore symmetric distributions (Student's t and symmetric Beta distributions) with different values of the excess kurtosis κ. The results (Figure 5) suggest that there is indeed an exponential relation between β_1 and κ namely $\beta_1(\kappa|q) \simeq 2.12 + 0.58 \exp(-\kappa)$ with $R^2 = 0.993$.

5 Applications to Auto-Regressive processes, dynamical systems and real world data

So far, we have shown that the introduction of the euclidean norm in the computation of the recurrences yields an underestimation of the attractor dimension for random vectors in high dimensional phase spaces. We now show that we can still obtain important information about the dynamical behavior of a system using the extreme value theory framework. To provide these insights we consider different processes and datasets where the underlying dynamical systems introduces some sort of space-time dependence among the variables.

We begin with the case of uncoupled variables with time autocorrelation. In particular, we consider a Markovian stationary auto-regressive process $x_{n+1}^{(i)} = \phi x_n^{(i)} + \eta_n^{(i)}$, with $\phi = 0.8$ and $\eta_n^{(i)} \sim N(0,1)$ i.i.d. $\forall\ i$. Different points i are uncoupled, so that there is no spatial (cross-sectional) dependence. It can be seen from Fig. 4 (red curve) that the values of $E(\hat{d})$ follow the same analytic curve as the i.i.d. Gaussian random vector. This is due to the fact that time correlation does not affect the computation of d, which is insensitive to time reshuffling because

of the POT procedure. Indeed, by sorting the data in a different way, we will anyway have the same sample of exceedances. This is not true when the recurrences selection is made via a block maxima approach (see Lucarini et al. (2012) for further discussion).

We next turn to analysing the effect of spatial dependence on the estimated value of d. We consider the Bernoulli Shift coupled map:

$$x_{n+1}^{(i)} = (1-\epsilon) f(x_n^{(i)}) + \frac{\epsilon}{n} \sum_j^n f(x_n^{(i \neq j)}) \tag{10}$$

where

$$f(x) = 3x \mod 1$$

, and ϵ is the strength of the coupling. As shown in Figure 4 (turquoise lines), by increasing ϵ the dimension decreases, as more spatial degrees of freedom are progressively frozen. In the case of a completely coupled system ($\epsilon = 0.7$), the dimension is independent on n and equal to $d = 1$. In other words, ϵ is beyond a critical value such that all the variables have the same values when iterating the map, i.e. there is only one single degree of freedom left in the system. The estimate $\hat{d}$ is therefore correct and not affected by the curse of dimensionality. Thus, the latter appears to depend on the real value of d, rather than on the phase space dimension n.

The above consideration paves the way for checking the degree of non-randomness of a series of multidimensional observations. Both analytic results and evidence drawn from numerical experiments that $\hat{d}$ is always biased imply that an accurate estimate of the true value of d is currently impossible. However, from our results it follows that, for a random system, the difference $n - E(\hat{d})$ quantifies the bias due to the curse of dimensionality. Moreover, and even more importantly, the difference $E(\hat{d}) - \hat{d}$ quantifies the loss of attractor dimension, or in other words the number of frozen degrees of freedom, due to spatial correlations and dynamics in the system, since $E(\hat{d})$ is the expected dimension for the random vector case. While this quantity is biased itself, one may argue for a relative measure such as $\Delta = (\hat{d} - E(\hat{d}))/E(\hat{d})$ as indicator of the degree of non-randomness

in a dynamical system.

We verify this idea on datasets issued from finance and climate science. The details of the 22 financial time series analysed here are provided in table 1. As climate variable, we use 22 sea-level pressure (SLP) time series chosen from the roughly 10000 grid-points constituting the NCEP/NCAR 1948-2015 reanalysis dataset Kistler et al. (2001). We choose these data since both the SLP anomalies and the log-return of financial datasets have approximately Gaussian marginal distributions, so that we can compare the observed curves to our previous results. For each dataset, we extract randomly 30 sequences of 22 timeseries. The extraction determines the order in which we add a new time series to construct different databases for $1 < n < 22$. We fix the threshold quantile to be $q = 0.98$.

The existence of an underlying dynamics, or of spatial/cross-sectional dependence, should reduce d to a smaller value with respect to the one predicted by Eq. 9, i.e. we can still use the extreme value theory methodology to assess whether the behavior of a set of variables is compatible with a random system. This is indeed what we observe and show in Fig. 6: when $n = 1$, we obtain $d = 1$ for all the choices, but as soon as $n > 5$ the correlations in the data result in an effective dimension d well below both the phase space dimension and the analytic result obtained for Gaussian random vectors. For the financial time series, we have $d \simeq 10$ for $n = 22$, slightly higher than for the SLP. We also remark that the values of d obtained by varying n in the datasets follow a similar curve as in Eq. 9, with $R^2 = 0.994$ for the financial dataset, and $R^2 = 0.9996$ for the SLP data.

The finance and climate data thus have a non-random behavior and dependence among variables which reduces the effective number of degrees of freedom to a value lower than the one predicted by the curse of dimensionality. Such non-randomness can be quantified by the measure Δ suggested above. In fig. 7, we show the value of Δ for the two time series. Besides instabilities at very small phase space dimensions, this quantity is clearly decreasing and dependent on n. For the case of the financial data, we can assess its magnitude for the full phase space: with $n = 22$, $\Delta \simeq -0.47$, indicating almost 50% of *randomness loss*.

6 Discussion and Conclusions

We have investigated the properties and performance of an estimator of the attractor dimension, or number of active degrees of freedom, d for dynamical and random systems with n-dimensional phase spaces, based on extreme value theory. In particular, we determined the analytic relationship between d and n, and assessed the accuracy of the estimator for high-dimensional systems, i.e. when n is large. Even though the method does not rely on embedding, we observe a systematic underestimation of d suggesting that our approach suffers from a similar curse of dimensionality to embedding techniques. This effect can be explained by the phenomenon known as concentration of the norm, which causes the ratio of L^p-norm distances of nearest and farthest neighbours from a given point to tend to 1 as the topological space dimension becomes larger than 3.

We derive an analytic expression for expected value $E(\hat{d})$ of the Hausdorff dimension induced by the curse of dimensionality as a function of the phase space dimension n for Gaussian random vectors and, consequently, for the bias $n - E(\hat{d})$. Then, we show through numerical simulation that this result holds also for non-Gaussian random vectors. Flatter distributions, characterized by very light tails, yield estimated values of d larger than the one obtained for Gaussian random vectors for all considered phase space dimensions n, while the converse holds for more peaked distributions. This implies that the former are more efficient in exploring the phase space than the latter, as predicted by information theory. Thus, tail heaviness of the distribution has a pivotal impact on the quality of recurrences and therefore on the curse of dimensionality.

Additional investigation is required to establish more precisely the properties of the extreme value theory estimator: in particular, the role of skewness should be studied in addition to the one of kurtosis, to assess whether not only tail heaviness, but also asymmetry plays a role in freezing degrees of freedom. Furthermore, distributions with fractal support should also be considered. Moreover, we plan to study the performance of a modified version of the estimator, based on the Mahalanobis distance, instead of the Euclidean norm. In fact, ours and previous results have only shown that degrees of freedom are frozen in the general case of dependence among

the variables in the system. However, no indication is given about whether such dependence is linear, i.e. arising from cross-sectional/spatial correlation, or nonlinear. The Mahalanobis norm provides a way to compute the distance between two vectors or fields normalised with respect to the presence of linear correlation among variables, and reduces to the Euclidean norm in case of mutually uncorrelated random vectors when the covariance matrix is diagonal. Comparing values of d estimated using Euclidean and Mahalanobis distances will allow not only to assess the level of non-randomness in a covariance-stationary systems, but also the relative contribution of linear and nonlinear dependence to the dimensionality reduction. Finally, extreme value theory is a useful tool to derive a certain number of useful quantities in dynamical systems theory: the correlation dimension and the Lyapunov exponents Faranda and Vaienti (2018), the generalized dimensions Caby et al. (2018) and the synchronization of coupled lattice map Faranda et al. (2018). It will then be interesting to evaluate the curse of dimensionality effects on those quantities.

Our findings point to an intrinsic difficulty in estimating the attractor dimension d of dynamical systems accurately, in particular when d itself is large. Indeed, we have shown that the curse of dimensionality is not affected directly by the dimension of the phase space n, but is rather dependent on the real value of d. For highly coupled systems with a small number of active degrees of freedom, we can retrieve an accurate estimate of d independently of the phase space dimension.This suggests that the use of the extreme value theory to extract information about the dimensionality in real datasets is justified whenever $d \leq O(10)$ as in the studies Faranda et al. (2017); Messori et al. (2017); Rodrigues et al. (2018); Buschow and Friederichs (2018); Faranda et al. (2019). Notwithstanding the curse of dimensionality, our estimator can further provide important information about complex dynamical systems with large d and with variables having a temporal or spatial interdependence. In fact, for any dynamical system with n-dimensional phase space, it is possible to produce a random system with the same marginal probability distribution, and compute the corresponding value of the estimator. Then, the difference between the result for the random vector and for the dynamical system, $(\hat{d} - E(\hat{d}))/E(\hat{d})$, measures the level of non-randomness of the system itself.

7 Acknowledgments

We thank P Yiou, B Dubrulle and S Vaienti for useful discussion. DF, MCAC, GM acknowledge the support of the ERC grant A2C2 n No. 338965. MCAC was further supported by Swedish Research Council grant No. C0629701 and GM was further supported by a grant from the Department of Meteorology of Stockholm University and by Swedish Research Council grant No. 2016-03724.

Table 1: International codes of the financial time series used in this study (daily data for the period 01-Jan-1998 to 03-May-2017).

IEUCT01	MOVEIDX	USEURSP
USDOLLR	JAPAYE$	S&PCOMP
DJEURST	FTSEMIB	FRCAC40
DAXINDX	JAPDOWA	ER00(ML:OAS)
HE00(ML:OAS)	C0A0(ML:OAS)	H0A0(ML:OAS)
CBOEVIX	BMUS02Y(RY)	BMBD02Y(RY)
BMIT02Y(RY)	BMFR02Y(RY)	BMUK02Y(RY)
BMJP02Y(RY)		

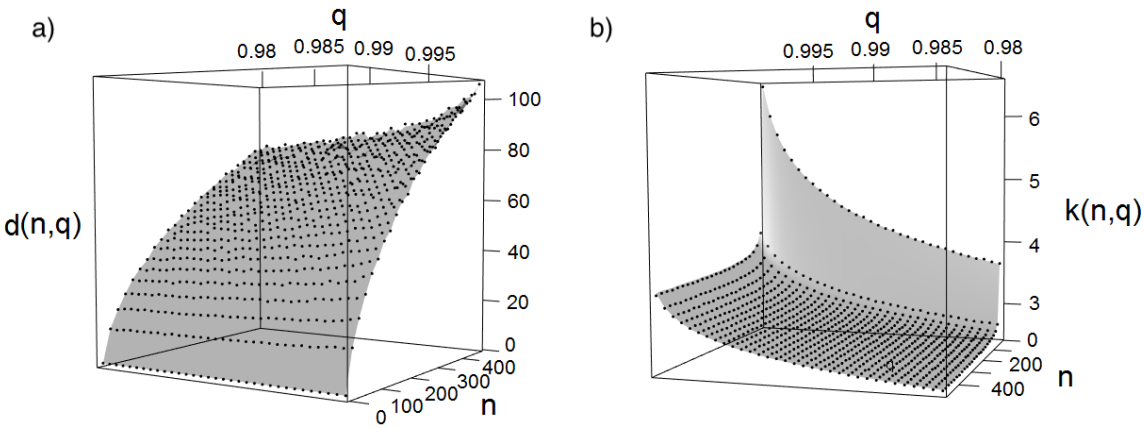

Figure 1: a) Dependence of the dimension d and b) $k(n, q)$ on the threshold percentile q and topological dimension n. Plots have been rotated differently for greater clarity. The grey surfaces represent the predictions given by Eq. 8 (a)) and Eq. 6 (b)) respectively. Black dots mark the results of Monte Carlo simulations.

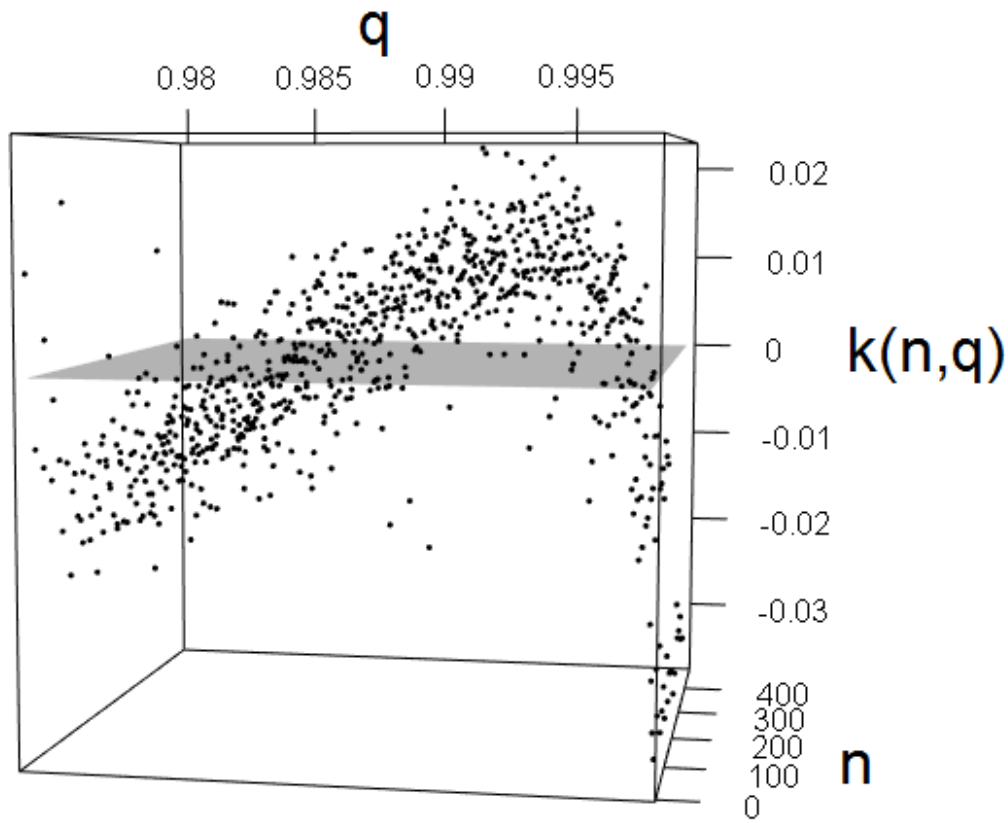

Figure 2: Prediction error on $k(q, n)$ (black dots); the zero plane (grey surface) is shown for reference. The error is generally small compared to the dimension estimates and mostly depends on the considered threshold quantile.

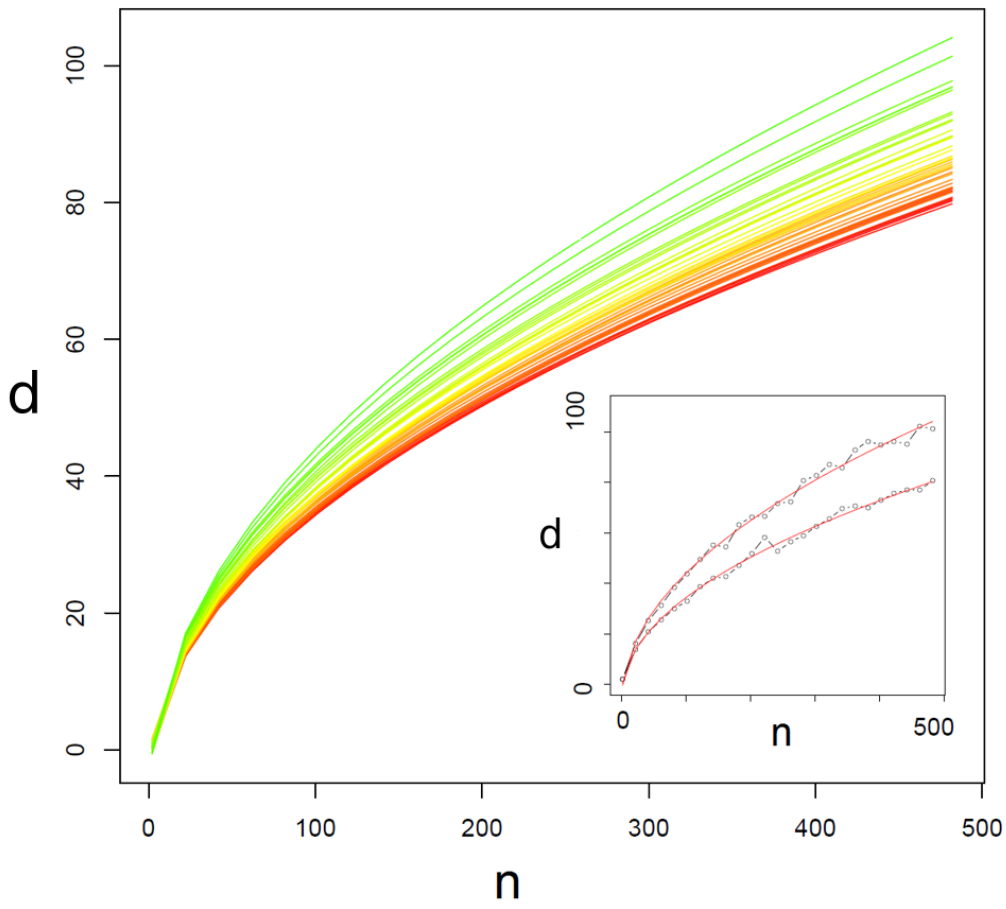

Figure 3: Main panel: expected values of d as a function of the phase space dimension n for Gaussian random vectors. Curves represent values predicted using Eq. 9 for different thresholds, ranging from q=0.98 (lower red line) to $q = 0.999$ (upper green line). Inset: comparison between observed values of d (black circles) and prediction based on Eq. 6 (red line)

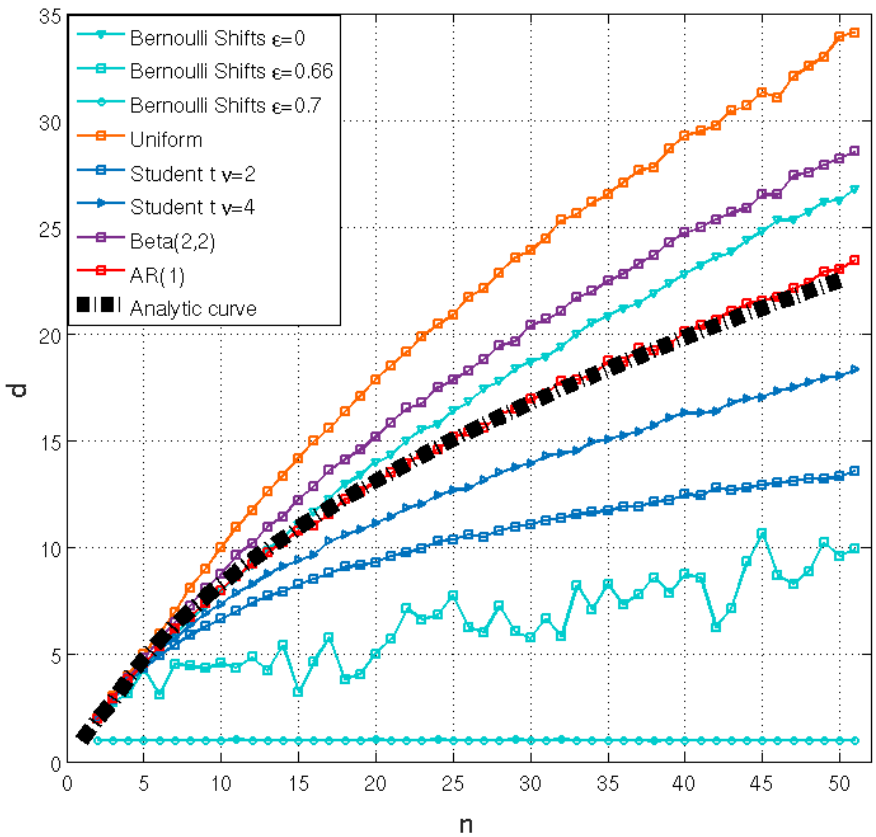

Figure 4: Attractor dimension d as a function of n for random vectors with different symmetric but non-Gaussian underlying distributions and for the Bernoulli shift map with three different values of the coupling parameter ϵ. All displayed points are averages over 300 realizations for series of length $T = 10^6$.

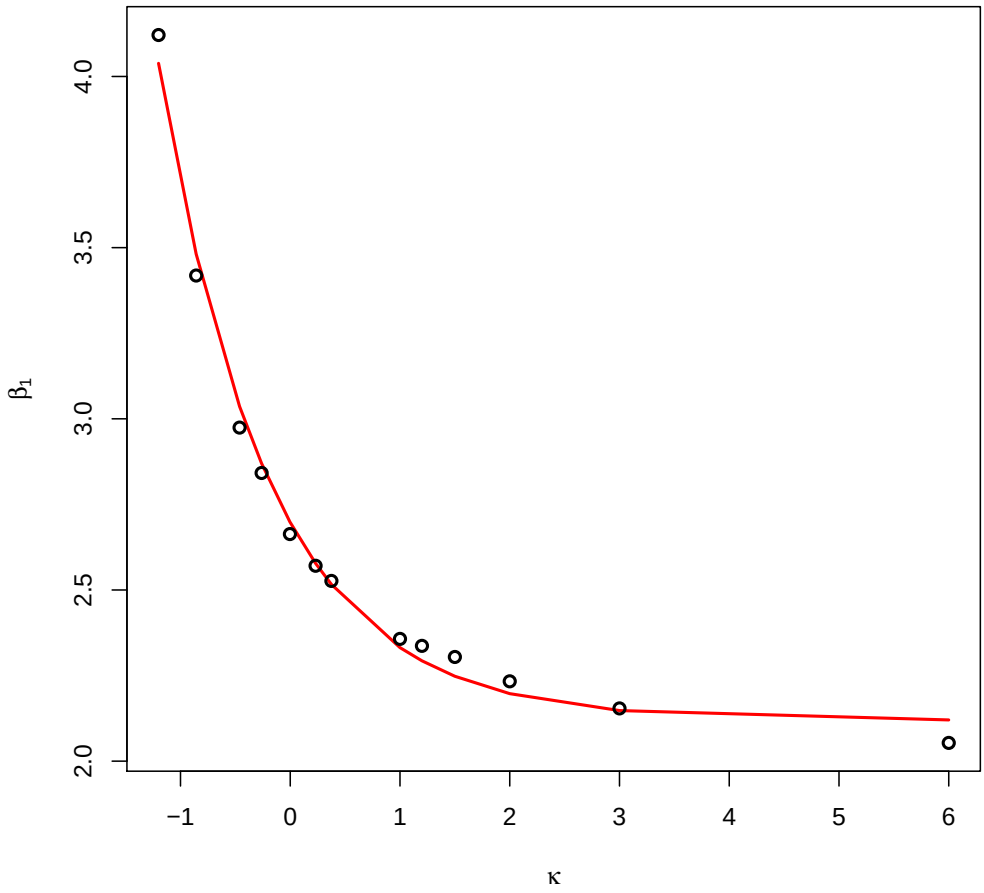

Figure 5: Relationship between the scaling coefficient β_1 and the kurtosis κ for all the distributions considered in Fig. 4. Black dots correspond to OLS estimates of β_1 from Eq. 9, the red line is a negative exponential fit.

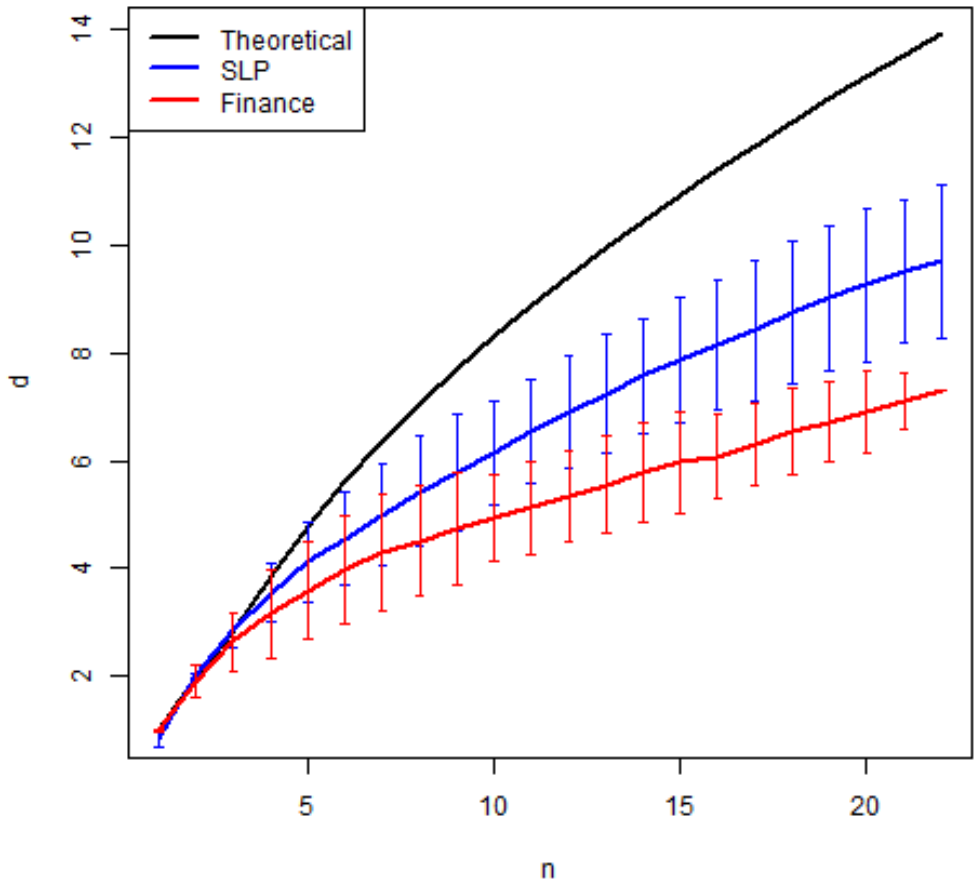

Figure 6: Comparison between the analytic result for Gaussian random vectors from Eq. 9 (black) vs observed values of d for 30 random combinations of 22 daily sea level pressure (blue) and financial (red) time series as a function of n. The error bars represent 1.96 standard deviations of the mean among the 30 combinations.

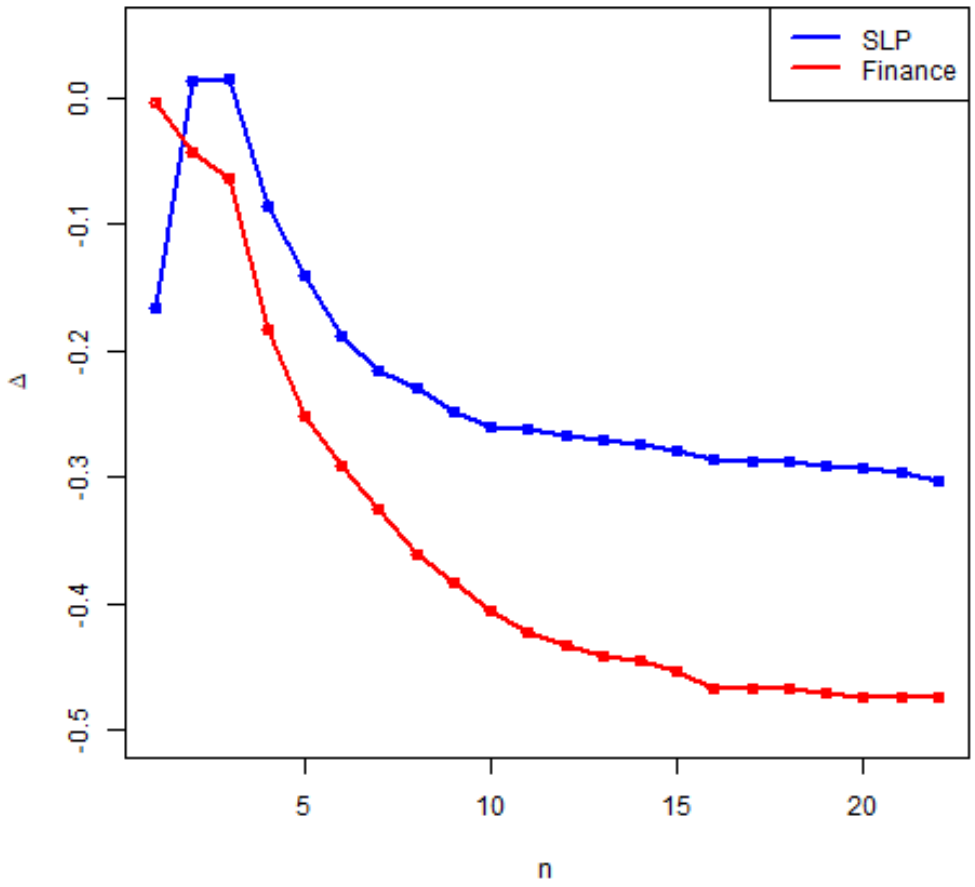

Figure 7: Values of the measure Δ for the financial time series (red line) and the sea level pressure dataset (blue line).

References

Arnold, B. C., Groeneveld, R. A., et al. (1979). Bounds on expectations of linear systematic statistics based on dependent samples. *The Annals of Statistics*, 7(1):220–223.

Balakrishnan, N., Charalambides, C., and Papadatos, N. (2003). Bounds on expectation of order statistics from a finite population. *Journal of statistical planning and inference*, 113(2):569–588.

Buschow, S. and Friederichs, P. (2018). Local dimension and recurrent circulation patterns in long-term climate simulations. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(8):083124.

Cabestany, J., Prieto, A., and Sandoval, F. (2005). *Computational Intelligence and Bioinspired Systems: 8th International Work-Conference on Artificial Neural Networks, IWANN 2005, Vilanova i la Geltrú, Barcelona, Spain, June 8-10, 2005, Proceedings*, volume 3512. Springer.

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

Faranda, D., Alvarez-Castro, M. C., Messori, G., Rodrigues, D., and Yiou, P. (2019). The hammam effect or how a warm ocean enhances large scale atmospheric predictability. *Nature communications*, 10(1):1316.

Faranda, D., Freitas, J. M., Lucarini, V., Turchetti, G., and Vaienti, S. (2013). Extreme value statistics for dynamical systems with noise. *Nonlinearity*, 26(9):2597.

Faranda, D., Ghoudi, H., Guiraud, P., and Vaienti, S. (2018). Extreme value theory for synchronization of coupled map lattices. *Nonlinearity*, 31(7):3326.

Faranda, D., Lucarini, V., Turchetti, G., and Vaienti, S. (2011). Numerical convergence of the block-maxima approach to the generalized extreme value distribution. *J. Stat. Phys.*, 145(5):1156–1180.

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

Faranda, D. and Vaienti, S. (2018). Correlation dimension and phase space contraction via extreme value theory. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 28(4):041103.

Freitas, A. C. M., Freitas, J. M., and Todd, M. (2010). Hitting time statistics and extreme value theory. *Probab. Theory Rel.*, 147(3-4):675–710.

Grassberger, P. (1986). Do climatic attractors exist? *Nature*, 323(6089):609–612.

Grassberger, P. (1988). Finite sample corrections to entropy and dimension estimates. *Physics Letters A*, 128(6-7):369–373.

Grassberger, P. and Procaccia, I. (1984). Dimensions and entropies of strange attractors from a fluctuating dynamics approach. *Physica D*, 13(1):34–54.

Kistler, R., Collins, W., Saha, S., White, G., Woollen, J., Kalnay, E., Chelliah, M., Ebisuzaki, W., Kanamitsu, M., Kousky, V., et al. (2001). The ncep–ncar 50–year reanalysis: Monthly means cd–rom and documentation. *Bulletin of the American Meteorological society*, 82(2):247–267.

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (2012). *Extremes and related properties of random sequences and processes*. Springer Science & Business Media.

Lorenz, E. N. (1991). Dimension of weather and climate attractors. *Nature*, 353(6341):241–244.

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

Lucarini, V., Faranda, D., and Wouters, J. (2012). Universal behaviour of extreme value statistics for selected observables of dynamical systems. *J. Stat. Phys.*, 147(1):63–73.

Mayer-Kress, G. (2012). *Dimensions and Entropies in Chaotic Systems: Quantification of Complex Behavior Proceeding of an International Workshop at the Pecos River Ranch, New Mexico, September 11–16, 1985*, volume 32. Springer Science & Business Media.

Messori, G., Caballero, R., and Faranda, D. (2017). A dynamical systems approach to studying midlatitude weather extremes. *Geophysical Research Letters*, 44(7):3346–3354.

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

Rényi, A. et al. (1961). On measures of entropy and information. In *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*. The Regents of the University of California.

Rodrigues, D., Alvarez-Castro, M. C., Messori, G., Yiou, P., Robin, Y., and Faranda, D. (2018). Dynamical properties of the north atlantic atmospheric circulation in the past 150 years in cmip5 models and the 20crv2c reanalysis. *Journal of Climate*, 31(15):6097–6111.

Samuel, A. L. (1988). Some studies in machine learning using the game of checkers. ii—recent progress. In *Computer Games I*, pages 366–400. Springer.

Schubert, S. and Lucarini, V. (2015). Covariant lyapunov vectors of a quasi-geostrophic baroclinic model: analysis of instabilities and feedbacks. *Quart. J. Roy. Meteor. Soc.*, 141(693):3040–3055.

Vannitsem, S. and Lucarini, V. (2016). Statistical and dynamical properties of covariant lyapunov vectors in a coupled atmosphere-ocean model—multiscale effects, geometric degeneracy, and error dynamics. *Journal of Physics A: Mathematical and Theoretical*, 49(22):224001.

Verleysen, M. and François, D. (2005). The curse of dimensionality in data mining and time series prediction. In *International Work-Conference on Artificial Neural Networks*, pages 758–770. Springer.