Extreme value analysis in dynamical systems: Two case studies

Tamás Bódai

Meteorological Institute, University of Hamburg

October 9, 2016

Abstract

We give here a brief summary of classical Extreme Value Theory for random variables, followed by that for deterministic dynamical systems, which is a rapidly developing area of research. Here we would like to contribute to that by conducting a numerical analysis designed to show particular features of extreme value statistics in dynamical systems, and also to explore the validity of the theory. We find that formulae that link the extreme value statistics with geometrical properties of the attractor hold typically for high-dimensional systems – whether a so-called geometric distance observable or a physical observable is concerned. In very low-dimensional settings, however, the fractality of the attractor prevents the system from having an extreme value law, which might well render the evaluation of extreme value statistics meaningless and so ill-suited for application.

0.1 Introduction

The climate system is a multiscale nonlinear dynamical system. The application of mathematical theories to it that approximate observables by simple random variables cannot be expected to always work. The example that will be examined here is classical Extreme Value Theory (EVT). Its main result, the Extreme Value Theorem due to Fisher and Tippett (1928) and Gnedenko (1943), states that uncorrelated random variables, no matter how they are distributed, share common features in their statistics at high quantiles: there is a family of distribution functions, that of the Generalized Extreme Value Distribution (GEVD), that universally describe the statistics in the tails. It is a limit theorem for extremes, similar to the Central Limit Theorem for means of random variables. The great value in this universality would be the possibility of predicting occurrence frequencies of events that have never been observed.

In the last century a scientific awakening took place, recognizing that random-like unpredictable phenomenon can be down to not only 'chance' but 'dynamical instabilities' and 'uncertainties', errors in observations, too. On this paradigm was built chaos theory, which describes in great detail aspects of the chaotic behaviour of deterministic nonlinear dynamical systems governed by differential equations devoid of any notion of chance. However, when it comes to observation, it can be argued that whether the underlying principles that govern the phenomena are stochastic or deterministic are just philosophical issues, for practical purposes we just want a framework that describes the statistics. The issue is philosophical, however, only if the different mathematical constructions yield identical result for the same enquiry. Whether this is the case, is not so intuitive. – The development of a new EVT for deterministic dynamical systems has seen some disagreement by different schools of thought. The earliest enquiries on the extreme value statistics (EVS) in dynamical systems go back to the 1990s (Balakrishnan et al., 1995), and at the turn of the millennium the field has received a powerful new impetus, starting with the landmark paper by Collet (2001). Fundamental achievements to date are: the formulation of conditions under which the EVS of certain types of observables can be described by a GEVD (Collet, 2001; Haiman, 2003; Freitas and Freitas, 2008; Freitas et al., 2009), and the linking of its presumably all-important parameter the *tail index* or *shape parameter*, characterizing the decay of the tail, with some fractal dimensions of the chaotic attractor (Collet, 2001; Freitas et al., 2009; Lucarini et al., 2012; Holland et al., 2012; Lucarini et al., 2014). The monograph by Lucarini et al. (2016) gives a comprehensive overview of the field.

We set ourselves the objective here to provide, following a concise and hopefully tangible exposition of the theoretical background, a detailed numerical investigation testing the relevance and power of the said recent theoretical results. Our main line of enquiry addresses the question: to what extent can we rely on universality in 'predicting events of previously unobserved magnitudes'?

EVT has been applied to environmental phenomena perhaps more than any other type of natural or man-made process. Storms and floods are only too common threats to property and life. Therefore there is a persistent demand on deepening the understanding of such extremes and on the development of new tools to quantify them. In practice one ubiquitous problem is the shortage of data. For the characterization of such data limit theorems might not apply even approximately; or, the determinism may be felt, resulting in markedly different characteristics from those of random variables. In nonautonomous systems that slowly change over time, data shortage arises in the sense that for predictions data from the too remote past cannot be used if it is not representative (is 'counterfactual') of the present, let alone the future. In a changing climate that will possibly bring about more extremes of some magnitude (Field et al., 2012), it is indeed a timely endeavour to revisit our assumptions of EVT, and acknowledge the dynamical nature of the climate system. This might prove to be necessary for the better service of society.

0.2 Classical Extreme Value Theory

The following exposition relies mostly on Leadbetter et al. (1983) and Coles (2001).

0.2.1 The block maxima approach

Consider the random process given by a sequence of *independent identically distributed* (iid) *random variables* (rv) $X_1, X_2, \ldots$ of *parent* cumulative distribution function (CDF): $F(x) = \mathbb{P}(X \leq x)$. We shall call a running sequence of m consecutive data points a *block*, whereby a long time series of length N may be broken down into blocks. A *block maximum* (BM; pl. block maxima, shortened in the following as BMa) can be formally expressed as $M_m = \max\{X_1, X_2, \ldots, X_m\}$. In search of a limit theorem for extreme values, one could take the limit of the block size $m \to \infty$. The analogy with the Central Limit Theorem comes from replacing the BM by the average $A_m = \langle X_i \rangle$. To help tackle the problem, take some *normalizing sequences* a_m, b_m, with which the CDF of the normalized block maxima:

$$F_{u,m}(u_m) = \mathbb{P}((M_m - b_m)/a_m \leq u_m) = \mathbb{P}(M_m \leq x(u_m)) = F_{x,m}(x(u_m)),$$

where $x = a_m u_m + b_m$. The main interest here is whether for any choice of a_m and b_m the normalized distribution $F_{u,m}(u_m)$ converges to a member of the family of some distributions as $m \longrightarrow \infty$. We say that we have an *extreme value law* (EVL), if for every $\tau^{-1} \in [0, \infty)$ there exists a sequence $x_m = x_m(\tau^{-1})$ such that (i)

$$m\mathbb{P}(X \geq x_m) = m(1 - F(x_m)) \xrightarrow[m \to \infty]{} \tau^{-1},$$

and therefore:

$$\mathbb{P}(M_m \leq x_m) = F_{x,m}(x_m(\tau^{-1})) = F^m(x_m(\tau^{-1})) \sim (1 - (m\tau)^{-1})^m \xrightarrow[m \to \infty]{} \exp(-\tau^{-1});$$

and (ii) the limit $u_m = (x_m(\tau^{-1}) - b_m)/a_m \xrightarrow[m \to \infty]{} u_\infty = u$ exists for every τ, and therefore:

$$\mathbb{P}((M_m - b_m)/a_m \leq u_m) = F_{u,m}(u_m) \xrightarrow[m \to \infty]{} G(u).$$

Extreme Value Theorem (Fisher and Tippett, 1928; Gnedenko, 1943): If we have an EVL with suitable sequences $a_m, b_m, \tau(u)$ will belong to one of the following family of distributions:

For a Type I or **Gumbel** *extreme value distribution* (EVD) $G(u; \mu', \sigma', \xi')$:

$$\tau(u) = \exp((u - \mu')/\sigma'), \ -\infty < u < \infty$$

For a Type II or **Fréchet** EVD:

$$\tau(u) = \begin{cases} 0, & u \leq \mu' \\ ((u - \mu')/\sigma')^{\xi'}, & u > \mu', \xi' > 0 \end{cases}$$

For a Type III or **Weibull** EVD:

$$\tau(u) = \begin{cases} (-(u - \mu')/\sigma')^{\xi'}, & u \leq \mu', \xi' < 0 \\ \infty, & u > \mu' \end{cases}$$

Type	a_m		b_m	
I	$H(\gamma_m)$	(stagnates)	γ_m	(diverges)
II	γ_m	(diverges)	0	(stagnates)
III	$x_F - \gamma_m$	(decays)	x_F	(stagnates)

Table 1: Suitable normalizing sequences for the three different types of extreme value distributions and their typical behaviour. For x_F, γ_m, $H(\cdot)$ see eq. (1).
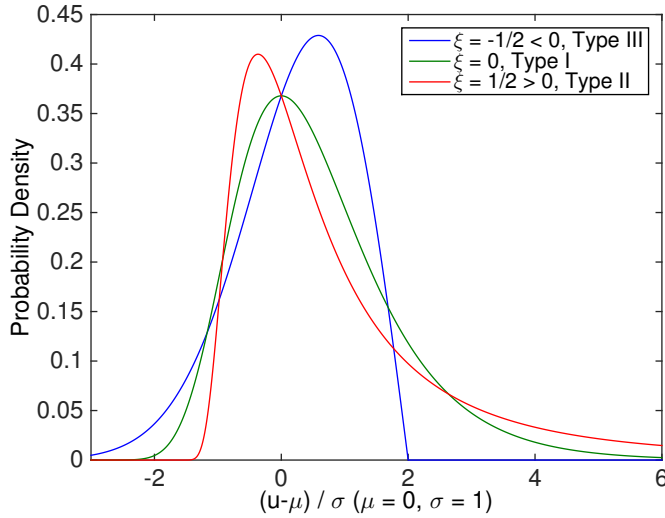

Figure 1: The three different types of extreme value distributions.

Suitable normalizing sequences that yield a *location parameter* $\mu' = 0$ and *scale parameter* $\sigma' = 1$ (Gnedenko, 1943) can be seen in Table 1, where

$$x_F = \sup\{x : F(x) < 1\},\ \gamma_m = F^{-1}(1 - 1/m),\ H(x) = \int_x^{x_F} \frac{1 - F(y)}{1 - F(x)} dy,\ x < x_F. \quad (1)$$

The above three Types can be combined into the single form of the **Generalized Extreme Value Distribution** (GEVD):

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

where $\xi = 1/\xi' \in \mathbb{R}$, $1 + \xi(u - \mu)/\sigma > 0$, $\mu \in \mathbb{R}$, $\sigma > 0$. The three different Types, I, II, III, are distinguished by the sign of the *shape parameter* ξ. Examples of these Types are shown in Fig. 1. The slow decay of probability in the case of the Fréchet distribution versus the Gumbel distribution, according to a power-law versus an exponential law, constitutes a qualitative difference between the tail behaviours. Furthermore, in contrast with both of the latter, the Weibull distribution is bounded from above, x_F being finite.

Note that $\mathbb{P}(M_m \leq x) = F_{x,m}(x) \approx G(x; \mu_{x,m}, \sigma_{x,m}, \xi) = G((x - b_m)/a_m; \mu, \sigma, \xi)$, i.e. the *unnormalised* distribution of BMa, is also approximated by a GEVD for large m's. This is due to the *max-stable* property of the GEVD. (In fact no other type of distribution bears this property, which is exploited in the proof of the Extreme Value Theorem.) For

$$\mathbb{P}((M_m - b_m)/a_m \leq u_m) = F_{u,m}(u_m) \to G(u; 0, 1, \xi),$$

3

one can apply the sequences (Faranda et al., 2011):

$$a_m = \sigma_{x,m} = [M_m^2 - \langle M_m \rangle^2]^{1/2}, \quad b_m = \mu_{x,m} = \langle M_m \rangle.$$

Although we will have 0 and 1 only approximately, as $\mu \neq \mathrm{E}[X]$ and $\sigma \neq \sqrt{\mathrm{E}[X^2] - (\mathrm{E}[X])^2}$ for a rv X of a GEVD. Otherwise, a normalization is not done in practice, as all three parameters of the GEVD are estimated, as described in the next section.

Analytic examples can provide some insight. The following three examples, one for each Type, are adopted from Coles (2001), with a correction.

Example 1.: standard exponential rv
$F(x) = 1 - \exp(-x)$
$a_m = 1,\ b_m = \ln m$

$$\begin{aligned}
\mathbb{P}((M_m - b_m)/a_m \leq u) &= F^m(u + \ln m) \\
&= [1 - \exp(-(u + \ln m))]^m \\
&= [1 - m^{-1}\exp(-u)]^m \\
&\to \exp(-\exp(-u)) \longrightarrow \xi = 0 \quad \text{(Type I Gumbel)}
\end{aligned}$$

Example 2.: standard Fréchet rv
$F(x) = \exp(-1/x)$
$a_m = m,\ b_m = 0$

$$\begin{aligned}
\mathbb{P}((M_m - b_m)/a_m \leq u) &= F^m(mu) \\
&= [\exp(-1/(mu))]^m \\
&\to \exp(-1/u) \longrightarrow \xi = 1 \quad \text{(Type II Fréchet)}
\end{aligned}$$

Example 3.: uniformly distributed rv
$F(x) = x,\ 0 \leq x < 1$
$a_m = 1/m,\ b_m = 1$

$$\begin{aligned}
\mathbb{P}((M_m - b_m)/a_m \leq u) &= F^m(m^{-1}u + 1) \\
&= (1 + u/m)^m \\
&\to \exp(u) \longrightarrow \xi = -1 \quad \text{(Type III Weibull)}
\end{aligned}$$

These examples represent typical behaviours of the normalizing sequences suitable for the different Types. Table 1 lists in which case they stagnate, decay or diverge.

0.2.2 Maximum likelihood estimation

In practice we have to estimate the extreme value distribution from a finite amount of data. This amounts to determining the parameters of the GEVD that best represents the data in some sense. The single most common way to do that is the *maximum likelihood estimation* (MLE) as follows.

In a general setting assume $x_1, \ldots, x_n$ are realizations of a sequence of continuous iidrv's of common probability density function (PDF): $\rho(x; \theta) = dF(y; \theta)/dy$. Here θ is the parameter to be estimated. The **likelihood** function is defined as:

$$L(\theta) = \prod_{i=1}^{n} \rho(x_i, \theta),$$

assigning a 'probability' to data. Although the value of L is *not* maximized in 1, and it is clearly not a probability of a possible outcome of an experiment. Nevertheless for values of θ further off the true value, clearly, smaller likelihoods are assigned to the data. Therefore, by way of the estimation we can look for a $\theta = \hat{\theta}$ that maximizes the likelihood. It yields the same result if the *log-likelihood*

$$l(\theta) = \ln L(\theta) = \sum_{i=1}^{n} \ln \rho(x_i, \theta)$$

is maximized. The estimation boils down to a numerical root-finding of $\frac{dl}{d\theta}$. The estimate is sometimes called the 'fit'.

For the GEV distribution, with $x_1, \ldots, x_n$ being realizations of the BMa $M_{m,i} = x_i$:

$$l(\mu, \sigma, \xi) = -n \ln \sigma - (1 + 1/\xi) \sum_{i=1}^{n} \ln(1 + \xi(x_i - \mu)/\sigma) - \sum_{i=1}^{n} (1 + \xi(x_i - \mu)/\sigma)^{-1/\xi}$$

if $\xi \neq 0$ and $1 + \xi(x_i - \mu)/\sigma > 0$ for any i; and if $\xi = 0$,

$$l(\mu, \sigma, \xi) = -n \ln \sigma - \sum_{i=1}^{n} (x_i - \mu)/\sigma - \sum_{i=1}^{n} \exp(-(x_i - \mu)/\sigma)).$$

0.2.3 Diagnostics

Instead of first checking the applicability of the Extreme Value Theorem and then performing MLE, one can also proceed by performing MLE to data straight away, and then check the result of that. One diagnostic may be a comparison of the empirical distribution function (EDF)

$$\tilde{G}(x_i) = i/(n + 1)$$

to the estimated GEVD

$$\hat{G}(x_i) = \exp\left\{ -\left[1 + \hat{\xi} \left(\frac{x_i - \hat{\mu}}{\hat{\sigma}} \right) \right]^{-1/\hat{\xi}} \right\}.$$

at the data points x_i. For a $\hat{G}$ that is a good model of the sample distribution: $\hat{G}(x_i) \approx \tilde{G}(x_i)$; and so when $\hat{G}(x_i)$ is plotted against $\tilde{G}(x_i)$, which is called the *probability plot*, the data points in the plot should lie tightly around the straight line of slope 1. In the case of long tails a better visualization of the goodness of estimation or fit is given by the *quantile plot*, whereby $\hat{G}^{-1}(\tilde{G}(x_i))$ is plotted against x_i, and a divergence from linearity would again indicate the lack of usefulness of the fit.

A **hypothesis test** can provide a more quantitative alternative to the above visual diagnostics. Here we seek to *reject* the hypothesis that the data x_i is drawn from the statistical population distributed by $\hat{G}(x)$. The confidence level of rejection is a matter of choice; usually high confidence is required. The so-called *p-value* of the test gives the probability that the observed deviation from the hypothetical distribution is purely by chance, i.e., a finite-sample statistical error. If this probability is very low, then we can reject the hypothesis with high confidence. A critical p-value of 0.05 is commonly adopted. A test that can take a generic $\hat{G}(x)$ (i.e. not only a GEVD) is *Pearson's chi-squared test* (Zeidler, 2004). The procedure starts out with binning the n data points into B bins. Then the following *test statistic* is calculated:

$$\chi^2 = \sum_{b=1}^{B} \frac{(O_b - E_b)^2}{E_b} = n \sum_{b=1}^{B} p_b \left(\frac{O_b/n - p_b}{p_b} \right)^2, \tag{2}$$

where O_b is the bin count of observation, and $E_b = p_b n$ is the expected bin count under the hypothesis. For large B and under the hypothesis the test statistic χ^2 is distributed approximately according to a χ^2-distribution, $F_{\chi^2}(\chi^2)$, of *degrees-of-freedom* of $B - 1 - n_\theta$, where n_θ is the number of estimated parameters used to determine E_b. In our case of the GEVD $n_\theta = 3$, and for the GPD (to be introduced in Sec. 0.2.5) $n_\theta = 2$. (That is, the actual hypothesis is not $\hat{G}(x, \hat{\mu}, \hat{\sigma}, \hat{\xi})$ with fixed parameters, but most generically $G(x)$.) Note that the distribution $F_{\chi^2}(\chi^2)$ is independent of n. The p-value is then the probability that a larger test statistic than the presently observed one, χ_*^2, will be observed in another identical experiment: $1 - F_{\chi^2}(\chi_*^2)$.

When the data does not actually conform to the hypothesis, in Eq. (2) $\sum_{b=1}^B (O_b/n - p_b)^2$ does not decay with increasing n, but converges to a finite value, and so χ_*^2 is expected to increase with n, and so therefore the p-value is expected to decay. That is, an insufficient amount of data (too small n) cannot reject the hypothesis even if it is actually false. This intrinsic characteristic should be eliminated in any application, meaning that two p-values are meaningful to compare only if the same n and (approximately) the same B ($1 \ll B \ll n$) are used. Taking the BM approach in evaluating EVS, with increasing block size m we have fewer and fewer data points n to fit the GEVD to. This way the p-value is potentially an increasing function because of a methodological error, masking the only effect of interest: the convergence to a GEVD, if any, with increasing m. This problem can be eliminated by fixing the number n_{min} of data points used in cases of all m's, assigning some acceptable minimum value to it. This way a bootstrapping is possible in the case of all but the maximal $m_{max} = N/n_{min}$. To utilize the same amount N of data for any m, and enforce $n = n_{min}$, n_{min} should be chosen as a factor of N; to this end N should be possibly truncated.

We note that the value of the likelihood function $L(\hat{\theta})$ itself cannot serve as a diagnostic quantity, which is related to the problem with the interpretation of it.

0.2.4 Correlated sequences

In a sequence determinism is indicated when the decay of the autocorrelation function happens over time. The present state depends partly on previous states. The extent of correlation of events long times apart is limited by the following $\mathcal{D}(x_m)$ **condition**: for all $i_1 < \cdots < i_p < j_1 < \cdots < j_q$ with $j_1 - i_p > l$

$$|\mathbb{P}\{X_{i_1} \leq x_m, \ldots, X_{i_p} \leq x_m, X_{j_1} \leq x_m, \ldots, X_{j_q} \leq x_m\} -$$
$$\mathbb{P}\{X_{i_1} \leq x_m, \ldots, X_{i_p} \leq x_m\} \mathbb{P}\{X_{j_1} \leq x_m, \ldots, X_{j_q} \leq x_m\}| \leq \alpha(m, l),$$

where $\alpha(m, l_m) \to 0$ for some sequence l_m such that $l_m/m \to 0$ as $m \to \infty$. This requires that at extreme levels, well-separated in time, events are independent to the extent that the correlation does not affect the limit law.

Theorem (Leadbetter et al., 1983): Given appropriate regularity conditions including $\mathcal{D}(x_m)$, we have an EVL for the $X_1, X_2, \ldots$ stationary process if and only if we have an EVL for the $X_1^*, X_2^*, \ldots$ sequence of iidrv's of the same parent distribution:

$$\mathbb{P}\{(M_m^* - b_m)/a_m \leq u\} \to G_1(u) \quad \text{if and only if}$$
$$\mathbb{P}\{(M_m - b_m)/a_m \leq u\} \to G_2(u) \quad \text{and}$$
$$G_2(u) = G_1^\theta(u),$$

where the constant $0 < \theta \leq 1$ is called the **extremal index**. ($G_2(u)$ is also a GEVD because of the max-stable property.) Therefore:

$$\mu^* = \mu - \sigma(1 - \theta^{-\xi})/\xi \quad \text{and} \quad \sigma^* = \sigma\theta^\xi \quad \text{when} \quad \xi \neq 0$$
$$\mu^* = \mu + \sigma \ln\theta \quad \text{and} \quad \sigma^* = \sigma \quad \text{when} \quad \xi = 0$$

Note that θ can be calculated by fitting the GEVD to the original correlated sequence and, in turn, to the the shuffled surrogate sequence. However, it is not practised. In fact $\theta < 1$ indicates a propensity of exceedances of high thresholds to group or cluster. The extremal index can be interpreted as the reciprocal of the mean cluster size at limiting threshold levels. In practice θ is estimated as the ratio of the number of clusters and the number of exceedances above the threshold level. There are various approaches to estimating the number of clusters.

0.2.5 The peak-over-threshold approach

Consider again a process given by iidrv's X of common parent distribution $F(x)$. We look for the distribution of *threshold excesses* $y = x - T$:

$$\bar{H}(y; T) = \mathbb{P}(X > T + y | X > T) = \frac{1 - F(T + y)}{1 - F(T)}$$

with a high threshold T. Suppose we have an EVL. Then, approximating (i) by a Taylor expansion for large T's and (ii) by the GEVD fitted to the unscaled BMa M_m:

$$1 - F(T) \approx -\ln F(T) \approx -\frac{1}{m} \ln G_m = \frac{1}{m} \left[1 + \xi_m \left(\frac{T - \mu_m}{\sigma_m} \right) \right]^{-1/\xi_m}.$$

This leads to

$$\bar{H}(y; T) \approx \left[1 + \frac{\xi_m y}{\tilde{\sigma}_m} \right]^{-1/\xi_m}, \quad \tilde{\sigma}_m = \sigma_m + \xi_m(T - \mu_m),$$

defined on $\{y : 1 + \frac{\xi_m y}{\tilde{\sigma}_m} > 0, \ y > 0\}$. The CDF H what $\bar{H}$ is a complimentary (cumulative) distribution function (CCDF) of is a member of the **Generalized Pareto Distributions** (GPD).

We can relate the GEVD and GPD such that (supressing the m-index)

$$\bar{H}(y - T; \tilde{\sigma}, \xi) = -\ln(G(y; T, \tilde{\sigma}, \xi)) \quad (y > T),$$

and so for large y:

$$H = 1 - \bar{H}(y - T; \tilde{\sigma}, \xi) \sim G(y; T, \tilde{\sigma}, \xi).$$

Since the shape parameter ξ of the GPD is identical to that of the GEVD, its sign distinguishes the different types of tail behaviour the same way. Examples for the three Types are shown in Fig. 2. A comparison of Figs. 1 and 2 provides a visual check of the asymptotic equivalence of the corresponding GEVDs and GPDs.

0.3 Extreme Value Theory for dynamical systems

0.3.1 Geometric distance observables

Consider a process generated by a *mapping rule* $f(x)$, i.e. a discrete *deterministic* dynamical system $x_{i+1} = f(x_i)$, $x \in \mathbb{R}^d$:

$$X_i = g(\text{dist}(f^{(i)}(x_0), \zeta)),$$

where $f^{(i)}$ is the i-fold application of the mapping rule, dist$(\cdot, \cdot)$ is the *Euclidean distance* between two points in the phase space or $\mathbb{R}^d$, and ζ is an arbitrary point on the (chaotic) *attractor*. The *geometrical distance observable* $g(\text{dist}(x, \zeta))$ is assumed to be such that it reaches a maximum at almost any ζ. Such an observable was first considered by Collet (2001). Note that here we will adhere to the term 'geometric distance observable' commonly used in the literature of this theory, even if the distance is not necessarily – and typically not – directly, experimentally observable.
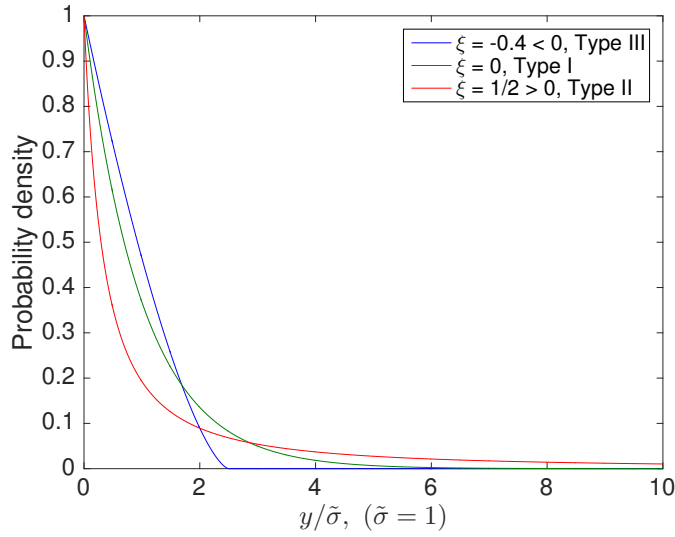
Figure 2: The three different types of tail behaviour in view of the Generalised Pareto Distribution.

The following results are due to Collet (2001) and Freitas and Freitas (2008): We have an EVL for the sequence $X_0, X_1, X_2, \ldots$ if two conditions are satisfied: (i) condition $\mathcal{D}_2(x_m)$ on mixing (Freitas and Freitas, 2008), or, equivalently, on the rate of decay of correlations (Collet, 2001), and (ii) condition $\mathcal{D}'(x_m)$ of anti-clustering. These two conditions are analogues of those in Leadbetter's (1983) theorem (Sec. 0.2.4). Later Freitas et al. (2009) identified a condition that is easier in practice to check: We have an EVL if and only if the (first) *return time distribution* is asymptotically *exponential* concerning infinitesimally small neighbourhoods of the point ζ (just like in the case of a rv).

It is insightful to distinguish the following three geometrical observables, first considered by Freitas et al. (2009):

$$g_1(\text{dist}(x, \zeta)) = -\ln(\text{dist}(x, \zeta)),$$
$$g_2(\text{dist}(x, \zeta)) = \text{dist}(x, \zeta)^{-1/\alpha},$$
$$g_3(\text{dist}(x, \zeta)) = C - \text{dist}(x, \zeta)^{1/\alpha},$$

$\alpha > 0$. Via the POT approach, considering an 'equivalent' rv of parent distribution identical to the one produced by the deterministic chaotic dynamical system, Lucarini et al. (2012) showed that for the normalization constants the Gnedenko formulae (1) lead to:

$$g_1: \ \gamma_m \sim \frac{\ln m}{D}, \quad H(\gamma_m) \sim D, \quad g_2: \ \gamma_m \sim m^{1/(\alpha D)}, \quad g_3: \ \gamma_m \sim C - m^{-1/(\alpha D)}, \quad (3)$$

where

$$D = \lim_{r \to 0} \frac{\ln \nu(B_r(\zeta))}{\ln r} \quad (4)$$

is clearly a *scaling exponent* for the *natural* or *SRB measure* ν in balls B_r of radius r around ζ, and hence it can be considered a *dimension*, which can be noninteger, i.e. a *fractal dimension*. (Note that the parent distribution $F(r) = \mathbb{P}(-\text{dist}(x, \zeta) \leq -r) = 1 - \nu(B_r(\zeta))$.) As we will see below, the results of (3) and (4) are exact only if the measure is smooth (absolutely continuous wrt. Lebesgue), just as the theorems of Collet (2001) and Freitas and Freitas (2008) require. In the literature D is referred to as the *local* (Faranda et al., 2011; Lucarini et al., 2012; Holland et al., 2012) or *pointwise* (Ott, 2002) dimension, and is known to be independent of ζ in ergodic systems. By substituting eqs. (3) into the formulae in Table 1, earlier results by Faranda et al. (2011) can be recovered for the

Observable	$a_m = \sigma_{x,m}$	$b_m = \mu_{x,m}$	ξ	Type
g_1	$\sim D$	$\sim \frac{\ln m}{D}$	0	I (Gumbel)
g_2	$\sim m^{1/(\alpha D)}$	~ 0 or $\sim m^{-1/(\alpha D)}$	$(\alpha D)^{-1} > 0$	II (Fréchet)
g_3	$\sim m^{-1/(\alpha D)}$	$\sim C$	$-(\alpha D)^{-1} < 0$	III (Weibull)

Table 2: Suitable normalization sequences and the resulting shape parameter ξ for three distinguished geometrical observables. The symbol $\sim$ means that an asymptotic scaling is sufficient – it does not matter what a_m and b_m are for finite m.

different geometrical observables, as listed in Table 2. We note that a dependence on m is identified only for a_m and b_m in some cases, but not for ξ within this framework. That is, an analytic law of convergence of the estimates of ξ is not available. Also, the scalings should hold only asymptotically, and a law of convergence to these scalings is not available either. Presumably, no universality applies to such convergence. We note that for any analytic parent distribution analytic expressions for the convergence of ξ are provided by the so-called Penultimate Extreme Value Theory; see e.g. (Gomes, 1984).

Clearly, the observable g_3 (third row in Table 2) can be distinguished by the fact that it preserves the Weibullian characteristics (negative shape parameter) of the EVS of the smallest values of merely the distance $\mathrm{dist}(x, \zeta)$, and the (modulus of the) shape parameter is most simply the reciprocal of the dimension (with the trivial choices $\alpha = 1$, $C = 0$):

$$\xi = -1/D. \tag{5}$$

That is, a geometrical property of the chaotic attractor predicts the EVS. Because of this importance of the 'local' dimension to EVS, we qualify it now in more detail.

For a singular measure bearing multifractal properties a dimension spectrum can be defined as follows (Ott, 2002). Consider first, for a time series of length N, the number of data points within a radius r of y_l:

$$\hat{n}(y_l, r) = \sum_{k=1}^{N} \theta(r - |y_k - y_l|),$$

where $\theta(\cdot)$ is the Heaviside theta (step) function. A *correlation sum* (an unbiased estimator of a correlation integral) can be obtained by geometric averaging:

$$C_1(r) = \exp[N^{-1} \sum_{l=1}^{N} \ln(\hat{n}(y_l, r)/(N-1))]. \tag{6}$$

Or, in general, by applying nonlinear averages we have:

$$C_q(r) = [N^{-1} \sum_{l=1}^{N} [\hat{n}(y_l, r)/(N-1)]^{q-1}]^{1/(q-1)}, \tag{7}$$

each of which scales (with $N \to \infty$) as:

$$C_q(r) \sim r^{D_q}. \tag{8}$$

For a (true) multifractal $D_0 < D_1 < D_2 < \ldots$. The inequalities are clearly due to the fact that the variance $\mathrm{var}_l[\lim_{N \to \infty} \ln(\hat{n}(y_l, r)/(N-1))] = \mathrm{var}_\zeta[\ln \nu(\zeta, r)]$ increases with $\ln r$ beyond any limit. This seems to be in contradiction with the statement that D in eq. (4) exists and is independent of ζ. It is not so in fact, because more precisely $D(\zeta) = D_1$ in all ζ that belongs to the *core region* of
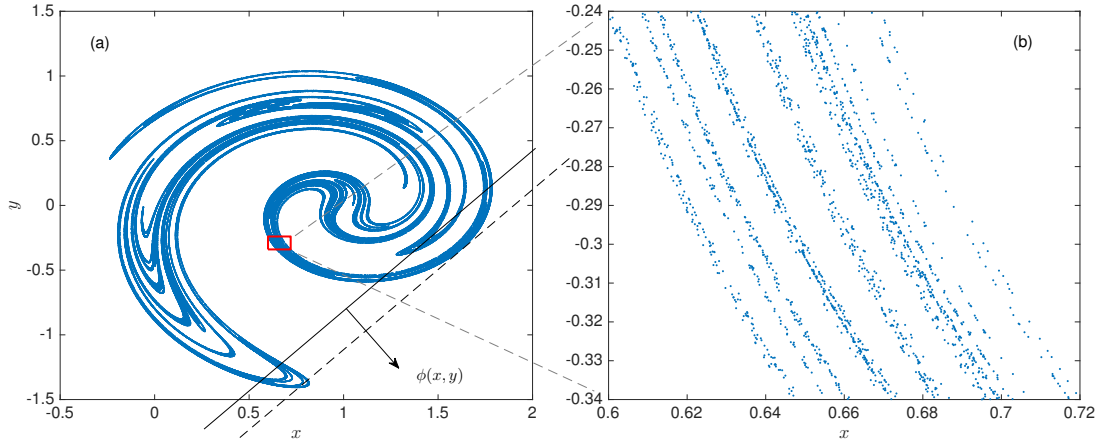
Figure 3: Geometric structure of the Ikeda attractor. (a) Full attractor and (b) a magnified part of it. The geometrical object of the attractor can be constructed locally as a Cartesian product of sets living in the stable and unstable dimensions, and so the measure supported by the attractor can be factorized wrt. these dimensions (also in high-dimensional settings). The dimension of the attractor restricted to the stable and unstable dimensions are therefore additive: $D_q = D_{q,u} + D_{q,s} + D_{q,n}$, where $D_{q,n}$ is the neutral dimension, equal to the number of zero Lyapunov exponents (LE). (There is at least one zero LE in a time-continuous system.) $D_{q,u}$ is the number of positive LEs, which is of course an integer. For $q = 1$ the Kaplan-Yorke relation (Ott, 2002) establishes a link between the geometry of the chaotic attractor and the dynamical stability of the trajectories in different directions quantified by the LEs $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_d$:

$$D_1 = D_{KY} = K + \frac{\sum_{k=1}^{K} \lambda_k}{|\lambda_{K+1}|}, \qquad (9)$$

where $\sum_{k=1}^{K} \lambda_k > 0$ but $\sum_{k=1}^{K+1} \lambda_k < 0$.

the attractor, and could be different elsewhere. Note that the core region can be covered by *typical boxes* of *uniform* probability that give the *maximum* probability out of all boxes of that size. In the limit of vanishing box sizes the typical boxes in fact contain the total probability, and the Hausdorff dimension D_0 of the core is the information dimension D_1 of the complete attractor.

Two sources of non-GEVD characteristics. After averaging, as the notation of (8) suggests, $\ln C_q(r) - D_q \ln r$ becomes a monotonically vanishing function for some large enough r. However, in practice the naturally occurring situation is such that one particular observable is of concern, specified by a particular ζ; that is, no averaging is done or achieved in any way. In this case $\epsilon = \ln \nu(\zeta, r) - D_1 \ln r$ is typically nonmonotonic and nonvanishing, but characterized by asymptotically stationary fluctuation properties wrt. $\ln r$ as the independent variable. We give the following argument to support this. Figure 3 shows the filaments of a chaotic attractor of a 2D map, which are continuous along the unstable direction and layered along the stable direction. This layering in the stable direction is created by the folding of phase space volumes, inevitable beside the stretching in the unstable direction, under dissipative chaotic dynamics. The layering is dictated typically by a fractal set, namely, a Cantor set, in which there are 'holes' of any finite size. When in the course of reducing the radius of the ball such a hole is swept, and so the ball is nowhere tangential to a filament, we should have something like $\nu(r)/\nu_f \approx (r/r_f)^{D_u}$. This feature is recurring at any scale. Therefore, to have a scaling of $\nu(r) \approx r^{D_u + D_s}$ on average, there must be – in a recurring manner – a scaling $\nu(r)/\nu_f \approx (r/r_f)^{D_u + \hat{D}_s}$ where $\hat{D}_s > D_s$. This gives rise to the fluctuations in ϵ mentioned above. Therefore, $\nu(r)$ does not conform to a GPD distribution for any small r.

Translating this result of the POT approach into terms of the BM approach we can say that the

10

deviations of the distributions of BMa $\rho_{u,m}$ from a GEVD are nonvanishing by increasing the BS, m, beyond any limit. This is simply because at any $x = r$ of irregularity in $\nu(r)$, there is an irregularity in $\rho_{x,m}$, which remains a prominent feature of the rescaled distribution $\rho_{u,m}$. We would like to argue for this on the basis that the fluctuations in ϵ, which translate into the said deviations, are governed by the *selfsimilarity* of the fractal set that supports the measure. For the argument let us clarify the concept of selfsimilarity on the example of a particular Cantor set. It is constructed in a way that the middle of the unit interval is removed leaving two (equal) segments of length $a < 1/2$. This configuration constitutes the *basic shape*. Next from each of these two segments remove a piece from the middle so that in total four segments of length a^2 are remaining. This process is continued ad infinitum in order to yield the actual Cantor set. Selfsimilarity means here that when we have a 'magnified view' of the set resolving a scale of $\varepsilon = a^k$, we will see the same basic shape for any k. The capacity or Hausdorff dimension D_0 is the power exponent of the scaling of the number $N_b(\varepsilon)$ of 'boxes' of edge length ε needed to cover the fractal: $N_b(\varepsilon) \sim \varepsilon^{-D_0}$. For the Cantor set $N_b = 2^k$, and so $D_0 = \ln 2 / \ln(1/a)$. Let us now associate what we called 'magnification' above with the rescaling $\rho_m((x - b_m)/a_m)$ by a_m taken from Table 2, and so 'equate' $m^{-1/(\lambda D_0)} = a_m \approx \varepsilon = a^k$, where $D_1 = \lambda D_0$, $\lambda \lessapprox 1$. Expressing m from this we have $m = \lceil (2^k)^\lambda \rceil$. This dictates (in this example) the (approximate) doubling of the block size so that we see a sequence of selfsimilar rescaled density distributions $\rho_{u,m}(u)$. Taking the limit $m \to \infty$ in this fashion is a 'valid choice', and this way it is clear that the irregularities in $\rho_{u,m}(u)$ and so $F_{u,m}(u)$ will not vanish. Therefore, $F_{u,m}(u)$ does not converge to some $G(u)$, and so **no EVL exists**. This deviation from classical properties can be measured by e.g. the p-value of a χ^2-test (see Sec. 0.2.3), whereby in lower dimensions the hypothesis of the data coming from a GEVD can be rejected based on smaller data sets.

When $D_0 < d$ is an integer, there are no holes in the layering. However, there might still be fluctuations of $\epsilon(\ln r)$ that prevent the existence of an EVL. We will analyse such situations numerically at least.

A non-GEVD or non-GPD character can have another source, which would occur even if the attractor has an integer dimension and the measure is smooth. Consider first the situation with a uniform measure over an area-filling convex support in 2D. Concerning a particular reference point, a GPD would describe $\nu(r)$ for $r < r_0$ when the ball is completely filled by the support. The convergence is typically achieved for some finite r_0. Next, we modify this situation by introducing a gap of width Δr that parts two (not necessarily completely disconnected) parts of the support; and we consider a reference point on the side of the gap. If $\Delta r \ll r_0$, then $\nu(r)$ is well-approximated by a GPD for $\Delta r \ll r < r_0$, but an irregularity will be emphasized by $\Delta r \lessapprox r$, and would disappear with $r < \Delta r$. Without limiting generality, since Δr can be arbitrarily small, our example prompts that the ultimate convergence can be delayed arbitrarily. Such an irregularity would persist even if one averages over all possible reference points. This might be a fairly typical situation in dynamical systems as the attractor is naturally folded. In practice, however, when no averaging is done, it could be very difficult to tell apart an 'irregular fold' from a regular fold or hole among the filaments, because there could be an intricate structure of many irregular folds featured by the attractor. In other words, it is hard to identify that a convergence to a stationary $\epsilon(\ln r)$ has been achieved at some r.

Finally, we argue that an EVL can exist only in the not so natural situation when averaging is achieved in the sense of eq. (6) or (7), thanks to the ensuing decay of $\epsilon(\ln r)$. We point out that in practice it is straightforward to fit a GEVD to lumped BMa data belonging to different ζ's. However, this corresponds to an algebraic averaging, i.e. to $q = 2$ in eq. (7), and so this will yield, via a similar algorithm to the Grassberger-Procaccia algorithm (Ott, 2002), the correlation dimension $D_2 = -1/\xi$. We must perform a geometrical averaging, as in eq. (6), if we want to estimate D_1 from EVS. This is not so straightforward, and we will describe a suitable procedure in Sec 0.4.1.

0.3.2 Physical observables

A physical observable $\phi(x) : \mathbb{R}^d \to \mathbb{R}$ is required to be a smooth function, e.g.:

$$\phi_E(x) = x^T E x \text{ (energy)}$$
$$\phi_W(x) = \|Wx\| \text{ (wind speed)}$$
$$\phi_V(x) = Vx \text{ (vorticity)}$$

These are generally *not* in the form of the geometrical distance observables $g_{1,2,3}(\text{dist}(x, \zeta))$. Therefore, the results of Table 2 do not necessarily apply. However, it is still a valid objective to link some parameter of the EVS, like ξ, of a class of observables with 'global' characteristic numbers of the chaotic attractor, like some fractal dimension of it.

Holland et al. (2012) obtained the following formula in the case of a few specific discrete-time systems for generic physical observables:

$$-\frac{1}{\xi} = \frac{D_u}{2} + D_s, \tag{10}$$

where $D_s = D_{1,s}$ is meant again, and conjectured that it holds for any well-behaved systems. Clearly, this would result in a *negative* shape parameter, which is consistent with the fact that the attractor of a finite-dimensional system is *compact*. Subsequently, Lucarini et al. (2014) derived this formula via a POT approach in a generic setting. These authors included the neutral dimension in their formula, such that: $-1/\xi = (D_u + D_n)/2 + D_s$. We note that for a discrete-time map generically $D_n = 0$. The following intuitive commentary can be given to the formula.

The observable function defines level surfaces $\phi(x) = constant$ in the phase space. One particular level surface, say with $\phi(x) = \phi_1$, is tangential to the attractor. The maximum of the observable quantity occurs in the point of the attractor where the level surface is tangential with it. The level surface is certainly tangential with the unstable directions, and so 'perpendicular' to the stable directions. The question is then: how do the measure factors confined to these dimensions above the level surface vary with moving to other levels? Because of the perpendicularity we have $\nu(\phi > \phi^*) \sim (\phi_1 - \phi^*)^{D_s}$; and because of the so-called *quadratic tangency* we have $\nu(\phi > \phi^*) \sim (\phi_1 - \phi^*)^{D_u/2}$.

As for an individual physical observable the holes and (irregular) folds play the same roles as described for geometrical observables above: they introduce a deviation of the extreme value distribution from the GEVD/GPD. In addition, the particular choice of the scalar physical observable might imply a folded 'view' of the higher-dimensional attractor. An example of this is shown in Fig. 3 (a). Close to the maximal value ϕ_1 of ϕ, there is a slightly smaller value ϕ_2 belonging to which the level surface (dashed black line in Fig. 3 (a)) is tangential to another part of the attractor. That is, the views of the two different parts of the attractor are folded into one. For $\phi_2 < \phi < \phi_1$ only one part shapes the tail, and for $\phi < \phi_2$ both parts do. Hence, there will be a crossover in the scaling $\nu(\phi > \phi_*)$ at about $\phi_* = \phi_2$; the scaling exponent remains the same after the crossover, but the prefactor will be different. Therefore, with a threshold $T \lesssim \phi_2$ the deviation of the distribution of excesses from a GPD will be the most emphasized. The same way, for a block size when the distribution of (unscaled) BMa has an expected value around ϕ_2, it will deviate the most from a GEVD. Just like the gap size Δr for a distance observable in our example earlier, here $\phi_1 - \phi_2$ can be any small, with which the convergence to a GEVD/GPD can be delayed arbitrarily.

We saw for the distance observables that wrt. the holes an averaging over the attractor fixes the problem, at least in principle. In case of physical observables, however, an averaging over different observables is not obviously possible. It would take a function space and a measure over it, but a pertinent choice is not dictated, i.e. any choice seems arbitrary.

No such problem arises for invariant sets with nonsingular measures when, beside $D_{1,u}$, $D_{1,s}$ too is an integer. This is the case with conservative dynamical systems. All such systems studied by Holland et al. (2012) were found to satisfy eq. (10); but none of the dissipative systems studied

convincingly satisfied that equation. No convergence could be confidently registered, which should be due to the holes (exposed also in their Fig. 13) or folds.

Beside discrete-time systems, these authors also treated continuous-time systems, but applied the same formula (10) without involving D_n. This is indeed the correct thing to do. To see why, we point out that a block maximum for any observable is the level that a turning point or apex of the continuous time series reaches, where the derivative vanishes: $\dot{\phi} = (\partial\phi/\partial x)^T \cdot \dot{x} = 0$. Making use of the differential equations, the latter expression designates a surface in phase space of dimension $d-1$. That is, the turning points fall on a kind of Poincaré section by which the time-continuous dynamics can be reduced to a discrete-time dynamics. (This property has a bearing also on the predictability of threshold exceedances of such turning points; see (Bódai, 2015).) We note furthermore that the dimension of any appropriate Poincaré section of the attractor is the same. This argument applies also to the distance observables: the minima of $\mathrm{dist}(x, \zeta)$ fall also on a Poincaré section. Note that the BM and POT approaches yield the same result for time-continuous systems when the *discrete* time series of the apexes are considered when adopting the POT approach.

When a level surface grazes the attractor without a generic quadratic tangency, a completely different (and to the situation possibly unique) approach is needed in order to work out the EVS of the system (Holland et al., 2012). This is the situation with the butterfly attractor of the classical Lorenz 63 system, taking any of its components as an observable. It features a *cusp*, where the maxima of the components are situated.

0.3.3 Nonclassical signatures of extreme value distributions

Signatures of extreme value statistics in deterministic sequences were first reported by Balakrishnan et al. (1995), and later by Nicolis et al. (2006) including also the case of quasi-periodic dynamics. The salient points of their study are summarized next, followed by our assessment. For the *deterministically* dependent sequence given by a 1D map

$$F_{x,m}(x) = \mathbb{P}(M_m \leq x) = \mathbb{P}(X_0 \leq x, \ldots, X_{m-1} \leq x) \neq F^m(x)$$

is a *joint probability* ($F_{x,m}(a) = 0$, $F_{x,m}(b) = 1$). The joint PDF describing the process:

$$\rho_m(X_0, \ldots, X_{m-1}) = \mathbb{P}(\text{to be in } X_0 \text{ in the first place}) \times$$
$$\Pi_{i=1}^{m-1}\mathbb{P}(\text{to be in } X_{i-1} \text{ given that one was in } X_0 \text{ } i \text{ steps earlier})$$
$$= \rho(X_0)\Pi_{i=1}^{m-1}\delta(X_i - f^{(i)}(X_0)),$$

with which

$$F_{x,m}(x) = \int_a^x dX_0 \rho(X_0)\theta(x - f(X_0))\ldots\theta(x - f^{(m-1)}(X_0)).$$

The Heaviside step functions $\theta(\cdot)$ provide the generally distinct *ranges of integrating* $\rho(X_0)$ between a and x, namely, where $x \geq \{f(X_0), \ldots, f^{(m-1)}(X_0)\}$. New integration ranges are *added* as x increases. This occurs e.g. past a *periodic point* of *period* up to $m - 1$, assuming chaotic dynamics. This characterizes the CDF $F_{x,m}(x)$ and the PDF $\rho_{x,m}(x)$ as follows:

(i) $F_{x,m}(x)$ is *monotonically increasing* (as expected),

(ii) and at the periodic points its slope *changes abruptly*, increases or decreases.

(iii) Accordingly, $\rho_{x,m}(x)$ has *discontinuities*.

Properties (ii) and (iii) are unlike those of the GEVD, and so 'nonclassical'. The question is whether the limit function as $m \to \infty$ belongs to the GEVD family.

Regarding the latter question the following two facts have a bearing. First, the periodic points are *exponentially proliferating* with increasing m, which gives rise to the topological entropy $h = \limsup_{m\to\infty}(\ln N_m)/m$, where N_m is the number of periodic points of up to period m. Second, Balakrishnan et al. (1995) proved for the *tent map* on the unit interval that the weight of $\rho_{x,m}(x)$ in $x \in [1 - 1/m, 1]$ is of $o(1)$. This is in agreement with the third row of Table 2, namely, that the width or standard deviation of the EVD is $1/m$ for this map. From these it follows that for abrupt changes of the slope of the *rescaled* limit EVD we have to look in a dense set of points. An interesting question that can be asked is if the changes of the slope will remain finite in the limit or vanish. Balakrishnan et al. (1995) provided the following formula [their eq. (3.18)] for the changes of slope, or the increase of the density, at R-type (slope steepens) intersections of $f^{(m-l)}$ with the bisectrix ($x = X_0$) at ξ_R:

$$\rho_{x,m}(x = \xi_R + 0) - \rho_{x,m}(x = \xi_R - 0) = l(1 + 2^{l-m}).$$

With $l = \lfloor m - \log_2 m \rfloor$ there is just a single such ξ_R, and it is at about $1 - 1/m$. This is a good enough place for us to look at. For the rescaled $\rho_{u,m}(u)$ we find the discontinuous jump to be of $o(1)$ in the limit. That is, it is not vanishing. This is indeed a different feature from that of random variables. However, considering that these jumps of the density occur in a dense set of points, it is possible that the series of distribution functions $F_{u,m}(u)$ converge to a member of the GEVD family, whereby the limit function have different properties from any of the finite (but arbitrarily large) m approximants. That is, even without any monotonic decay of the slope changes, there could still be a *convergence in distribution*, giving rise to an EVL, consistently with the theorems by Collet (2001), Freitas and Freitas (2008) and Haiman (2003). In fact for e.g. the Bernoulli map ($x_{m+1} = 2x_m \bmod 1$) that produces a uniform invariant measure over the unit interval similarly to the tent map:

$$F_{x,m}(x) \to x^{m/2}.$$

This is indicated by the diagram in Fig. 4 showing the root-mean-square-errors (RMSE) between the EDF $i/(n+1)$ vs x_i (see Sec. 0.2.3) of the actual distribution $F_{x,m}(x)$ and the asymptotic form $x^{m/2}$. The diagram suggests a scaling law $RMSE \sim (BS/BS_0)^\gamma$, $\gamma < 0$, that is, that the RMSE decays, i.e. vanishes in the limit. (It follows that the RMSE vanishes much slower than the reciprocal of the number of periodic points.) This suggests indeed that **an EVL exists**. This is similar to the situation with the sequence of X_m random variables with densities $\rho_m(x) = 1 - \cos(2\pi m x)$. These variables converge in distribution to a uniformly distributed variable, whereas the densities do not converge at all (Romano and Siegel, 1985).

We emphasize that the limit is $x^{m/2}$ and not x^m which would correspond to a uniformly distributed random variable on the unit interval. That is, the deterministic nature of the process does not matter wrt. having an EVL, but the limit is not reached in the same way as with a random variable. In the case of the tent map studied in great detail by Balakrishnan et al. (1995) there is not even such a difference, but $F_{x,m}(x) \to x^m$. Nevertheless, the rescaled $F_{u,m}(u)$ of the Bernoulli map converges in distribution to the same GEVD when we use a normalizing sequence $a_m = 2/m$ (instead of $1/m$), to the tune of the examples in Sec. 0.2.1. And therefore the all-important shape parameter ξ is unchanged by determinism.

We note that what we call here *nonclassical signatures* refer to differences from the case of random variables of smooth parent distributions, but the term does *not* concern the *applicability* of classical EVT. This is a strong difference from the signatures due to fractality discussed in Secs. 0.3.1 and 0.3.2, in which case no EVL exists.

0.4 Applicability of theory

In the rest of this chapter we will conduct two numerical case studies. Their objective is first to numerically confirm the theoretical results linking the shape parameter of the Weibull EVD to the
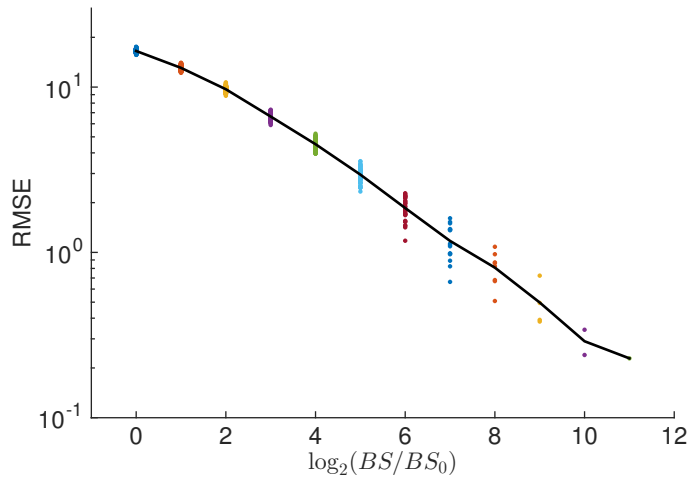
Figure 4: Convergence of $F_{x,m}(x)$ for the Bernoulli map: decaying deviation between $F_{x,m}(x)$ and the form $x^{m/2}$ that it asymptotically approximates. The deviation is given in terms of the root-mean-square-errors (RMSE) between the empirical distribution wrt. $F_{x,m}(x)$ and the asymptotic form. The results have been obtained from a time series of length $N = 2^{24}$. Here $BS_0 = 1$ simply, and so $BS = m$. For each RMSE figure $n = 2^{12}$ data points were taken, and so the maximal block size to be used was $BS_{max} = 2^{24-12}$. The black line indicates the average over realisations for all BS's except BS_{max}, as explained in Sec. 0.2.3. Note that with a limited $n = 2^{12}$ a scaling behaviour can be traced numerically only up to a limited BS. That is, a scaling exponent should be estimated if desired in a range of BS's bounded from above and below, in which the log-log plot features a straight line.

15

dimensions of the chaotic attractor given by eqs. (5) and (10). Second, we will demonstrate that it is crucial to aid our computations with appropriate diagnostics, and that good results can be obtained in rather limited situations.

0.4.1 Geometric distance observables of the Lorenz 84 model

The main aim here is to verify numerically eq. (5) for the minima of distances from reference points on the attractor. We shall do that in a chaotic system with the lowest possible dimensionality, in order to demonstrate that wrt. a single reference point it is impossible to evaluate the EVS, but with averaging over the attractor we recover classical properties accurately. Furthermore, we choose to consider a time-continuous dynamical system to back up our theoretical argument that when adopting the BM approach the neutral dimension of these systems has to be ignored. The low dimensionality will be favourable for the demonstrability of the latter; but also for the demonstrability that it is the Kaplan-Yorke/information dimension D_1, not the correlation dimension $D_2 < D_1$ or the Hausdorff dimension $D_0 > D_1$, that satisfies the theoretical formula (Brandstater and Swinney, 1987). These requirements are satisfied by Lorenz's model of global atmospheric circulation (L84), whose equations with standard parameter settings read as follows (Lorenz, 1984; Provenzale and Balmford, 1999; Tél and Gruiz, 2006; Drótos et al., 2015):

$$
\begin{aligned}
\dot{x} &= -y^2 - z^2 - x/4 + F/4, \\
\dot{y} &= xy - 4xz - y + 1, \\
\dot{z} &= xz + 4xy - z.
\end{aligned}
\tag{11}
$$

The model describes – in a very coarse manner (van Veen, 2003) – the meridional heat transport via eddies, represented by principal mode amplitudes y and z, given rise by the baroclininc instability of the midlatitude jet, represented by its average speed x. The instability occurs for appropriate conditions defined by the large-scale meridional temperature gradient, represented in the model by F, due to differential heating between the equator and the poles. The equations are nondimensionalised with respect to time by the average damping time of eddies, being about 5 days. We will examine the autonomous dynamics in perpetual winter conditions realized by, say, $F = 8$, since this gives rise to chaotic dynamics.

First we show the severity of the problem with evaluating any EVS in case of low dimensionality and when the reference point ζ is fixed. Figure. 5 shows the (unscaled) density distribution of BMa $\rho_{2^2 BS_0}(-\text{dist}(x, \zeta))$ for five different randomly chosen reference points ζ on the attractor of L84. (Note that the basic block size is chosen to be $BS_0 = 40$ [model time unit (MTU)], which implies that we are endowed by a time series for a data set that contains $o(10^7)$ turning points.) They look nothing like a GEVD in that they are very 'spiky', unsmooth. Therefore fitting of a GEVD to these data sets looks rather meaningless, even if possible, beside that the estimates of the shape parameter ξ would vary strongly. We proposed that the overlay/averaging of even such irregular distributions would result in a smooth GEVD, and therefore the existence of an EVL.

Note that in a different context we have found that the overlay of $o(100)$ ensemble-based snapshot attractors of a nonautonomous system driven, say, by a stationary stochastic process would reproduce fairly well the fuzzy attractor traced out by a single long trajectory (Bódai et al., 2011). Therefore we attempt the overlay of EVDs belonging to $R = 2^6 = 64$ different reference points. (These reference points can be taken from a long trajectory, well separated in time.) The result of estimating ξ is shown in Fig. 6 (a). In the diagram two reference values of ξ are marked by dashed lines calculated by eq. (5) from estimates of $D_1 = 1.46$ and $D_2 = 1.25$. The former has been estimated from Lyapunov exponents via the Kaplan-Yorke relation (9). The latter has been calculated directly by the Grassberger-Procaccia algorithm. For this a Poincaré section of the attractor was taken by the surface $z = 0$, as done by Bódai et al. (2013). The two dimension values are fairly distinct. The p-values of χ^2-tests are also given in Fig. 6. For the block size, or $m = BS/BS_0$, where we have the

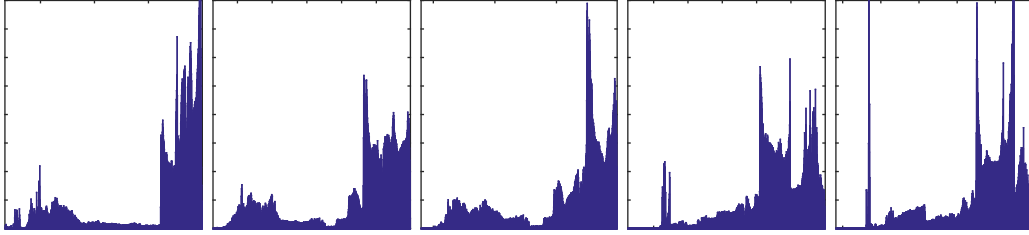Figure 5: Histograms of $\rho_{2^2 BS_0}(-\text{dist}(x, \zeta))$ for five different randomly chosen reference points ζ on the attractor of L84.

largest p-value, and in a range around that sample value, the ξ estimate is about midway between the reference values. This is unfortunate because we want to back up our argument that D_2 determines ξ with a simple lumping of BMa data belonging to different ζ's. To this end we propose to increase the number R of ζ's.

But before we get down to that, we note that as the block size m increases, the ξ estimates depart from the theoretical prediction. What is more, the p-values improve such that in some cases it cannot be rejected that the data comes from a GEVD. The sequence of density distributions $\rho_{x,m}$ of BMa shown in Fig. 7 reflects both the ξ estimates and the p-values. The reason for the said departure should be due to the following numerical detail. The lower turning points of $\text{dist}(x, \zeta)$ were searched for by fitting a cubic spline to a few data points around the point where a turning point was detected in the discrete numerical time series. With this analytic approximation of the continuous time series, a root-finding procedure is used to locate the apex. However, this technique does not work; the turning point is always found where we have the locally minimal discrete value. That is, applying this technique makes no difference from considering the discrete time series only. This way the minima of $\text{dist}(x, \zeta)$ are not restricted to fall on a 2D Poincaré surface, and the dimension of the whole attractor needs to be considered, which is that on the Poincaré section plus one. A pair of dashed reference lines based on these dimensions are also included in the diagrams in Fig. 6. – The ξ estimates approximate this prediction. Although the approximations are not very accurate, neither do the p-values exceed – or not by far – our confidence level. It might be because the amount of data is not enough for convergence to set in. We note that the location/range of the crossover wrt. m from one theoretical prediction to the other should depend on the time step size of numerical integration: the smaller the step size, the 'later' the crossover.

Quadrupling the number of reference points finally vindicates our argument for D_2 versus D_1; the results with this increase are displayed in panel (b) of Fig. 6. The better estimation of the theoretical prediction goes with larger p-values.

A further improvement is attempted in such a way that the apexes of the continuous time series are estimated by a root-finding procedure implemented in Matlab's event handler, as an option for the solver `ode45`. The result with this is displayed in panel (c) of Fig. 6. There is a very modest improvement in that the tail of the distribution is now concave ($\xi < -0.5$), see panel (n) of Fig. 7, but there is still a departure from the theoretical value. This situation illuminates the value of the diagnostics by the p-value at least. The reason for this departure is not known. It might be due to what was called an 'irregular fold' at a particular small scale in Sec. 0.3.1, something that the averaging cannot repair.

Finally, we would like to make the point that, as eqs. (6) and (7) suggest, it may be another dimension than D_2 that governs the EVS, depending on the type of average applied. Conversely, to the tune of the proposal of Lucarini et al. (2014, 2016), different dimensions can be estimated in principle via evaluating EVS. Here we would like to demonstrate this on the example of D_1. It is most natural to take the POT approach for this, as dictated by eq. (6). Therefore, it is not the GEVD
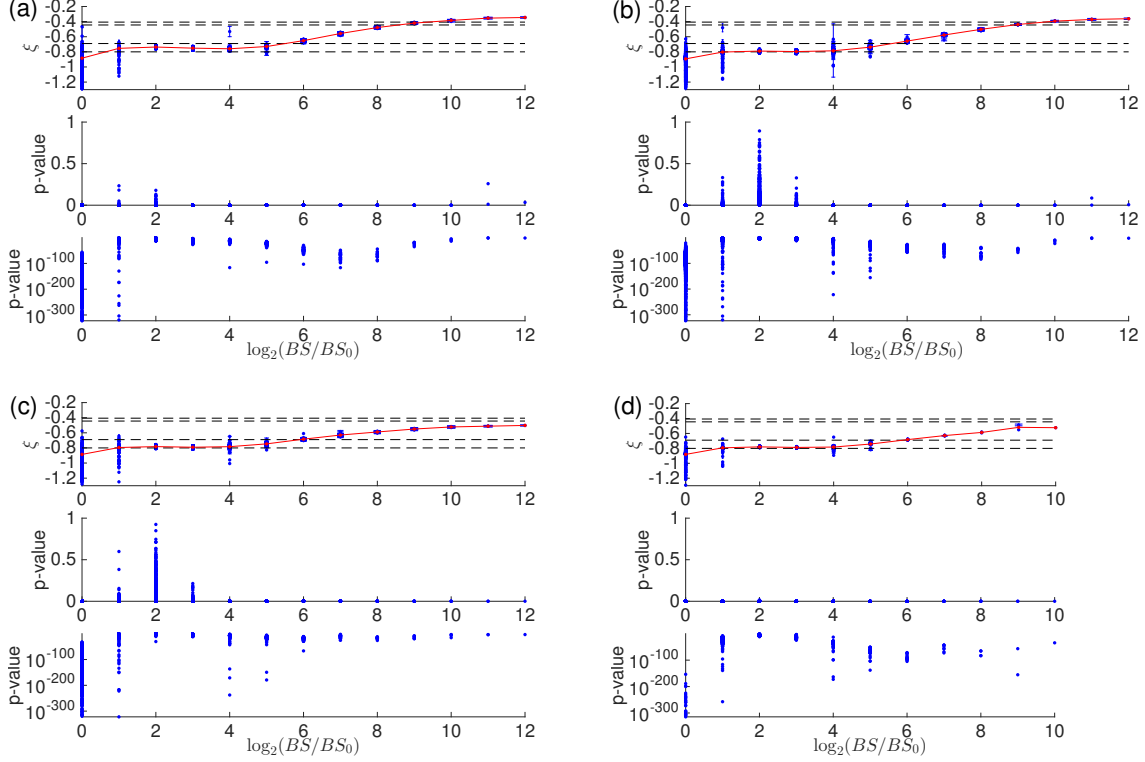
17

Figure 6: Estimates of the shape parameter ξ for $-\mathrm{dist}(x, \zeta)$ of L84 with averaging over the reference points ζ, along with estimation diagnostics: the p-values of χ^2-tests. (The p-values were obtained by `chigof` of Matlab, adopting its default setting of $B = 10$ for all results in this chapter.) The BM approach and MLE method are used. (We used `gevfit` of Matlab that implements the MLE of GEVDs.) (a) $n_{min} = 2^{12}$, $R = 2^6 = 64$, interpolation by cubic spline to find the turning point. (b) $n_{min} = 2^{12}$, $R = 2^8 = 256$, interpolation by cubic spline. (c) $n_{min} = 2^{12}$, $R = 2^8 = 256$, Matlab's event handler is used to find the turning point with very high accuracy. (d) $n_{min} = 2^{14}$, $R = 2^8 = 256$, Matlab's event handler is used. Red lines connect the bootstrapping means. Errorbars (hardly ever visible) indicate the 95% confidence interval of the MLE for all individual estimates. The p-values are given in lin-lin as well as log-lin diagrams. Horizontal dashed lines mark theoretical predictions $\xi = -1/D_1 = -0.69$ and $-1/D_2 = -0.8$ from estimated values of $D_1 = 1.46$ from Lyapunov exponents based on the Kaplan-Yorke formula (9), and $D_2 = 1.25$ from direct calculation by the Grassberger-Procaccia algorithm (Kantz and Schreiber, 2003). The Lyapunov exponents were computed by a standard method implemented in the user-defined Matlab function `lyapunov.m` downloadable from (Govorukhin, 2004). Another pair of dashed lines correspond to $D_1 = 1.46 + 1$ and $D_2 = 1.25 + 1$ to do with no event handling.
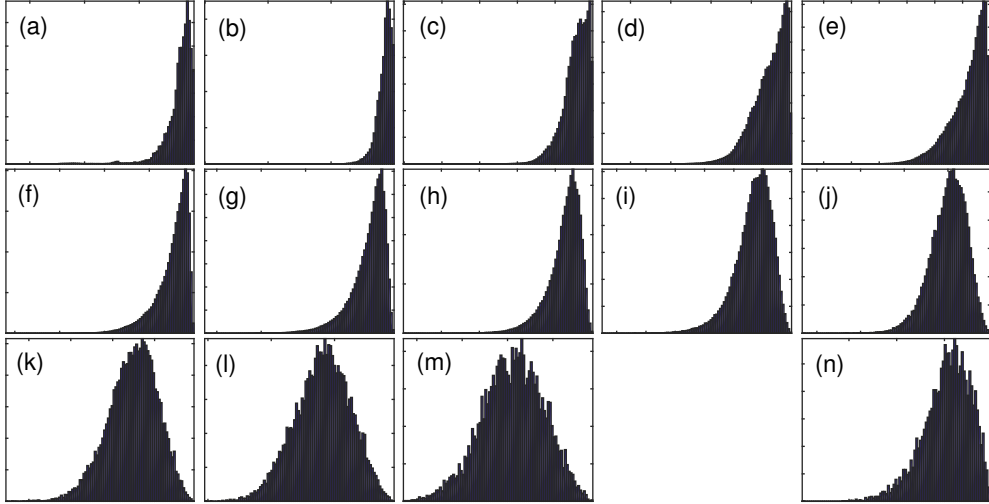
Figure 7: Density distributions $\rho_{x,m}$ of block maxima BMa for the observable $-\mathrm{dist}(x, \zeta)$ of L84 averaged/overlaid wrt. the different reference points ζ's. (a)-(m) $R = 2^6 = 64$ corresponding to panel (a) of Fig. 6, $m = 2^k$, $k = 0, \ldots, 12$ corresponding to the alphabetically-ordered panel labels; (n) $R = 2^8 = 256$ corresponding to panel (c) of Fig. 6, $m = 2^{12}$.

but the GPD that we 'fit' by the MLE (for D_2) or some other method (for D_1, to be described next). We also note that it is the CCDF of the apexes that are to be averaged, and then we take the CCDF of the average (as a reversal) to be fitted by the GPD. In the numerics we approximate the CDFs by their EDF. The 'average EDF' is obtained by averaging the different EDFs belonging to the different ζ's where their plateaus are overlapping. This average EDF will then have as many plateaus as all the EDFs have combined. We then look for a GPD that fits the average EDF the closest in some sense. We measure the closeness by the RMSE of the diagnostic probability plot from the diagonal line, in a similar fashion as done for Fig. 4. For the 'fit' we look for the minimum of the RMSE by a standard root-finding algorithm (called commonly a *least-squares fitting*). An initializing 'guess' value is provided as the parameter estimates by the MLE method, which should be rather close to the solution of this problem. One problem is that the geometric averaging does not yield a result if any of the terms is zero. The last plateaus of the EDFs start at a value of the distance less than zero. Therefore, we have a bottle neck problem. The only way to alleviate this problem is to increase n for the individual EDFs. Given an unchanged amount of total data, this requires us to take a smaller m_{max}. Still, in calculating the RMSE a finite small tail of the GPD is disregarded. The results are shown in Fig. 8. The MLE in conjunction with algebraic averaging finds very accurately the value $-1/D_2$, which occurs for the largest p-values. These p-values are, however, very small; the GPD hypothesis is easily rejected. This is just due to the fact that by trying to remedy the bottleneck problem mentioned above, the number $n = 2^{16}$ of datapoints is enough to reveal that they are certainly not drawn from a GPD given finite m and R. Showing this was exactly the objective of the exercise for panel (d) of Fig. 6 already: the p-values are smaller (as compared to those in panel (c)), being below the confidence level, nevertheless, the ξ estimates are hardly different. Back to Fig. 8, the green curve shows that our new method to estimate D_1 from EVS works. No diagnostics has been worked out for this estimation as yet, but noticeably the accurate estimates occur for the largest p-values to do with the MLE.

Although the method works, it is an outstanding question if it is an efficient one. Given a scalar time series of an observable of a system, by time-delay embedding one can reconstruct qualitative properties of the attractor, and even quantitative properties such as the correlation dimension or a spectrum of Lyapunov exponents (see Chapter 11 of Kantz and Schreiber (2003)) that can feed into the Kaplan-Yorke relation for D_1. It would be worthwhile to compare the bias (accuracy) and variance
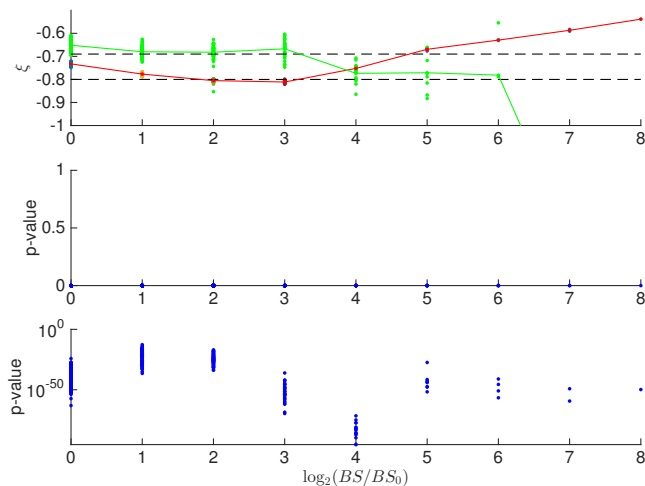
Figure 8: Estimates of the shape parameter ξ for $-\text{dist}(x, \zeta)$ of L84 with averaging over the reference points ζ, along with estimation diagnostics. The POT approach and the MLE (red) (utilizing Matlab's `gpfit`) and another ad-hoc diagnostics-based fitting method (described in the main text) are used. The red (green) curve corresponds to linear (nonlinear, geometric) averaging over ζ, and so $\xi = -1/D_2$ ($\xi = -1/D_1$). Horizontal dashed lines mark theoretically calculated values $\xi = -1/D_1 = -0.69$ and $-1/D_2 = -0.8$ from estimated values of $D_1 = 1.46$ based on the Kaplan-Yorke formula (9), and $D_2 = 1.25$ from direct calculation by the Grassberger-Procaccia algorithm (Kantz and Schreiber, 2003).

(precision) of the different D_1 estimators depending on the data set size. Furthermore, we should note that the usefulness of this method is restricted to very low dimensionality, because $D_1 - D_2$ can quickly get smaller than the absolute error made in the estimation of the dimensions, as also remarked by Brandstater and Swinney (1987)

0.4.2 Physical observables of the Lorenz 96 model

Similarly as with the distance observable, we seek here to confirm the theoretical result eq. (10). So far such attempts have only been made concerning 2D mappings (Holland et al., 2012; Lucarini et al., 2014) and 3D time-continuous dynamical systems (Holland et al., 2012). These attempts have seen limited success because of the problems with singular measures detailed in Sec. 0.3.1 and 0.3.2. Here we go beyond this and consider higher dimensional systems. Given preliminary indication that higher-dimensional systems produce – perhaps typically – fairly regular EVS (Vannitsem, 2007; Gálfi, 2015), it is an interesting question how high does the dimension need to be for regularity; what is the transition like from low to high dimensions? We will also investigate if integer-dimensional systems produce regular EVS. In this regard we consider dissipative systems, as we are aware of the study of low-dimensional conservative systems only.

A model that suits these purposes particularly well is Lorenz's 1996 model (Lorenz, 1996) that models wave dynamics on a longitudinal circle due to baroclinic instability. Its equations read as follows:

$$\dot{x}_j = -x_{j-2}x_{j-1} + x_{j-1}x_{j+1} - x_j + F, \ j = 1, ..., \hat{N}. \tag{12}$$

The system state on site j of the longitudinal circle is denoted by x_j, and because of the periodicity of the configuration: $x_{\hat{N}+1} = x_1$. The external forcing F can be thought of as the meridional thermal gradient, similarly to L84 (11). Since the law that governs the behaviour on each site is identical,

the behaviour should be identical too, at least in the long term. The 'field' of the L96 is *statistically homogeneous*, i.e. each site variable has the same distribution density. The instantaneous states are obviously not identical in general, except that they are more correlated the closer the sites are. We will consider fairly strongly chaotic solutions, when even neighbouring sites may not be very much correlated.

The dimensionality of the system is controlled clearly by the number $\hat{N}$ of sites, and the forcing F. The maximal dimension, say $D_1 = D_{KY}$, is obviously $\hat{N}$, the size of phase space. From statistical mechanics we expect that for some large enough $\hat{N}$ the system becomes fully 'turbulent' (Cross and Hohenberg, 1993; Gallavotti and Lucarini, 2014), and some scaling laws will be largely independent of any further increase of $\hat{N}$. The dimension can be argued (Cross and Hohenberg, 1993; Karimi and Paul, 2010) to be an extensive quantity, which allows for introducing the concept of the *dimension density* of the system:

$$\delta(F) = (D_{KY}(F) - 1)/(\hat{N} - 1).$$
(13)

which turns out to be (approximately) the same function of F for any large enough $\hat{N}$. This is demonstrated in Fig. 9, where curves belonging to as low values as $\hat{N} = 16$ and 32 overlap very closely. Even a third curve belonging to $\hat{N} = 10$ fits closely; although a blowup in an inset shows more deviation.

We observe that at about some finite $F = F_0$ chaos sets in, and then the dimension grows 'explosively' before it starts to level. The significance of this is that we would like to examine systems of fairly low dimensions, $D_{KY} < 10$, and so we have to find samples on this steep slope. We wish to examine integer dimensions $D_{KY} = i$, $i = 2, \ldots, 9$, in order to study the effects of what we called 'folds' in Sec. 0.3.2 alone, or perhaps other nonsmooth features of the measure in the stable direction too; and fractional dimensions, say, $0.4 + i$, in order to study the effects of 'holes' too. We need to find sample values of F that produce these D_{KY} values. As for the integer dimensions we have to be particularly careful, because we do not want to treat a situation where D_{KY} is in actual fact just slightly larger than some integer, because the holes thereby arising might have a very strong effect, just what we are trying to eliminate on the first place. Rather we prefer values which are just slightly smaller than an integer. To spare resources our strategy is the following. 0) We choose as few sites as possible; $\hat{N} = 10$ would satisfy our need as stated above. 1) First we estimate D_{KY} for a few sample values of F that reasonably populate the steep part, and some of the leveling bit up to $\delta \approx 0.9$. These are shown by the circle markers in Fig. 9. 2) Then we attempt to fit this data by a functional form. We observe that for large F, $1 - \delta(F)$ is decaying by a power law, as shown in the lower inset of Fig. 9, and the initial explosively increasing part can be modeled by something like a square-root function. These features can be combined in the following functional form:

$$\delta(F) = \left[1 - \left(\frac{F_0}{F}\right)^{\beta}\right]^{\alpha}, \quad 0 < \alpha < 1, \quad 0 < \beta.$$
(14)

$1 - \delta(F)$ scales asymptotically as $F^{-\beta}$. At F_0 $\alpha < 1$ ensures a vertical slope. For $\hat{N} = 32$, a least-squares fit of data $(\ln F_i, \ln(1 - \delta_i))$ yields the following parameter values: $F_0 = 4.94$, $\alpha = 0.3$, $\beta = 0.6$. The fit is rather accurate. Although for larger $\hat{N}$'s the parameter values should prevail, for the low $\hat{N} = 10$ we should do another fitting. The result of this is shown by the red curves in Fig. 9. The fit is good but not perfect. Therefore we proceed as follows. 3) Using the invertible functional relationship (14) we produce a series of sample values of F corresponding to uniformly spaced values of $\delta = 0.009i$, $i = 0, \ldots, 100$. 4) With these F values we calculate the actual D_{KY} values – results shown in panel (a) of Fig. 10 by coloured dot markers – and we pick a subset of F values that belong to our desired D_{KY} values, making sure that none of the 10 estimates for some fixed F exceed an integer value. Our picks are marked by green circles in the same figure. The blowup shows that for $F \approx F_0$ chaotic and periodic regimes alternate, and even when regular dynamics does not return by
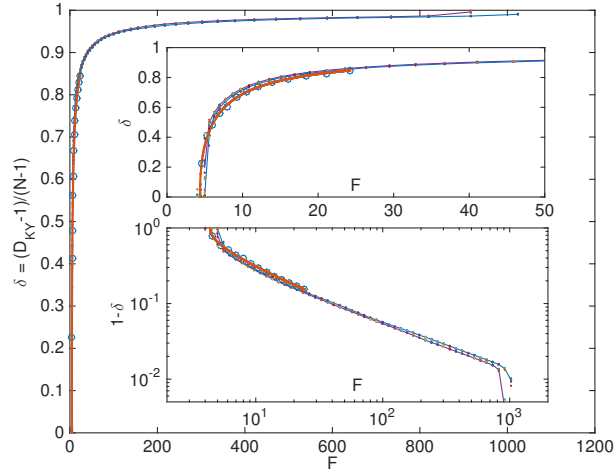
Figure 9: Dimension density as a function of forcing for the Lorenz 96 model. For computations the codes in the Appendix of (Gallavotti and Lucarini, 2014) are made use of. For the blue, magenta and red curves (and hollow circle markers) $\hat{N} = 32$, 16 and 10 (10), in that order. The blue and magenta curves connect the mean values for $R = 10$ different estimates of D_{KY} per sample value of $F = 2^{(2+i(10-2)/51)}$, $i = 0, \ldots, 51$ (represented by dot markers scattered only slightly). Each D_{KY} estimate is based on a simulation time span of $T = 500$ [MTU]. The circle markers each represent a single estimate with $T = 2000$ [MTU] at sample values $F = 2^{(1+i(5-1)/21)}$, $i = 0, \ldots, 19$. The red curve gives a least-squares fit of the form (14) to the circle markers. The upper inset shows a magnified view of the diagram, and the lower inset uses a log-log diagram to indicate an asymptotic scaling. For the largest few F's the scaling is not captured because of numerical inaccuracies.

increasing F, $D_{KY}(F)$ does not increase monotonically in the beginning. Nevertheless, we do wish to pick a few useful samples from this irregular structure.

We note in passing that $\xi(D_{KY})$ has discontinuous jumps, and so therefore does $\xi(F)$, since $D_{KY}(F)$ is (presumably) continuous here. When a Lyapunov exponent changes sign by changing F smoothly, $D_{KY,u}$ and $D_{KY,s}$ both have equal but opposite discontinuous jumps, and therefore they cancel in determining D_{KY} according to eq. (9). However, because in eq. (10) the same changes of $D_{1,u}$ and $D_{1,s}$ have different contributions to $-1/\xi$, their jumps cannot cancel each other. This effect is well visible in panel (b) of Fig. 10.

The outcome of evaluating EVS for the described scenarios of interest is collected in Figs. 11 and 12. We estimated ξ with and without bootstrapping. Without it, the p-values show an increasing tendency, which is rather unhelpful; and this problem is clearly fixed by bootstrapping, as nonmonotonic variation of p with m (or block size BS) is typical. The two types of estimates of ξ (magenta and blue curves) agree crudely for low dimensions, and the agreement tends to be better for larger m as well as larger dimensions. However, a third estimation with the shuffling of the time series when no bootstrapping is done (curves not plotted) reveals that up to $D_1 \approx 5$ the mismatch of the blue and magenta curves is dominantly due to autocorrelation, and for higher dimensions, with an abrupt transition, the effect due to bootstrapping becomes dominant.

For the quantitative p-values we provide the visual reference of the distribution densities of BMa displayed in Fig. 12. This illuminates the irregularities for low dimensions, and that they are gradually vanishing with increasing dimensions. Let us bear in mind that only every other row corresponds to noninteger fractal dimensions. Nevertheless, the transition seems to be gradual. This can allow us to think that also the folds (and perhaps remaining nonsmoothness of the measure) – not only the holes
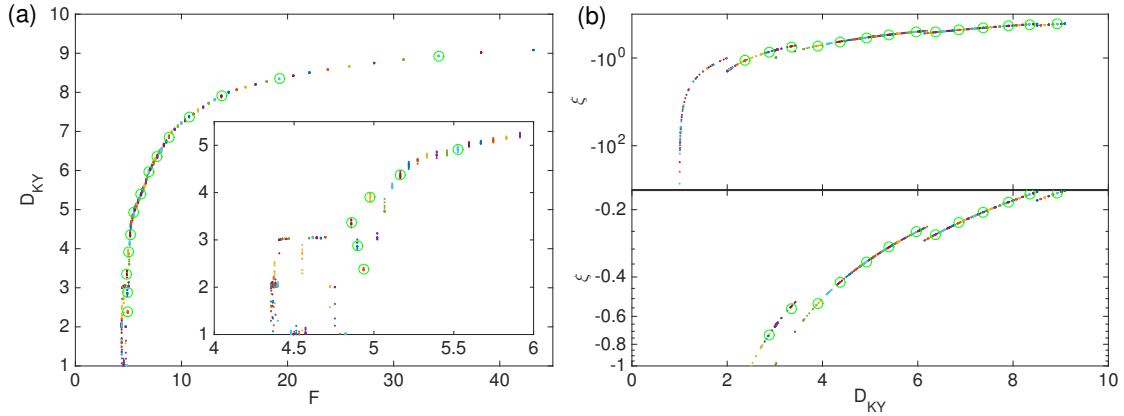
Figure 10: (a) Dimension density as a function of forcing for the Lorenz 96 model, (b) and associated theoretical predictions of the shape parameter by eq. (10) for physical observables as a function of the attractor dimension. Green circle markers highlight scenarios for which EVS is evaluated in Fig. 11. The inset in panel (a) shows a blowup. In panel (b) the diagram is shown for different ranges of ξ, in order to resolve discontinuities of $\xi(D_{KY})$.

– can have an intricate structure. Or, perhaps determinism, as discussed in Sec. 0.3.3, is still felt at a BS given by $m = 2^6 = 64$. However, in higher dimensions such irregularities, beside the holes, would be also 'masked' as argued by Nicolis and Nicolis (2012). The mechanism for this is a kind of averaging as the natural measure living in a high-dimensional phase space is projected onto a single dimension belonging to the scalar observable.

A very striking observation is that none of the ξ estimates agree well with the theoretical prediction. Evidently, we cannot rely even on high p-values (marker coloured in blue) as to accepting (without knowing the theoretical value) a direct MLE estimate. Typically ξ varies with m nonmonotonically/monotonically for very low/higher dimensions. But even in monotonic cases, and endowed with plenty of data, no convergence or leveling of estimates is typically registered. But rather there is a divergence from the theoretical prediction, towards smaller (in modulus) values.

We mention that we decided not to locate the turning points accurately, unlike in some cases with the L84. The reason is the following. When the distribution has a longer tail (ξ smaller in modulus), smaller errors are made by approximating the upper turning point by the maximum value in a discrete time series. Besides, one extra dimension changes the theoretical prediction modestly relative to the typical errors obtained.

Another idea for explaining the divergence may be that spatial correlations introduce a bias. This is tested on one particular scenario in Fig. 13. In it we show ξ estimates for the different sites separately. The variance of the estimates is rather small relative to the error, therefore each of them represent the errors due only to a finite BS.

It cannot be excluded either that the systems possess some degeneracy, similarly to the cusp in the L63 system, which requires a new approach to work out the EVS (Holland et al., 2012).

The reason for the error in our opinion, however, is rather that some folds even at this point are not masked by the averaging mechanism of projecting the measure onto one dimension. In Fig. 12 for lower-dimensional scenarios we can observe clearly a very extended thin tail. This tail seems to survive with the increase of D_{KY}, except the steep slope or 'shoulder' of the distribution becomes less pronounced. With larger BS's the irregular tails gain more weight in the distribution, and so they bias ξ to smaller values in modulus more. This tail might be a feature of the type of equation, because the x variable, but not y and z, of L84 features a similar very weakly populated tail protruding from the main probability mass.
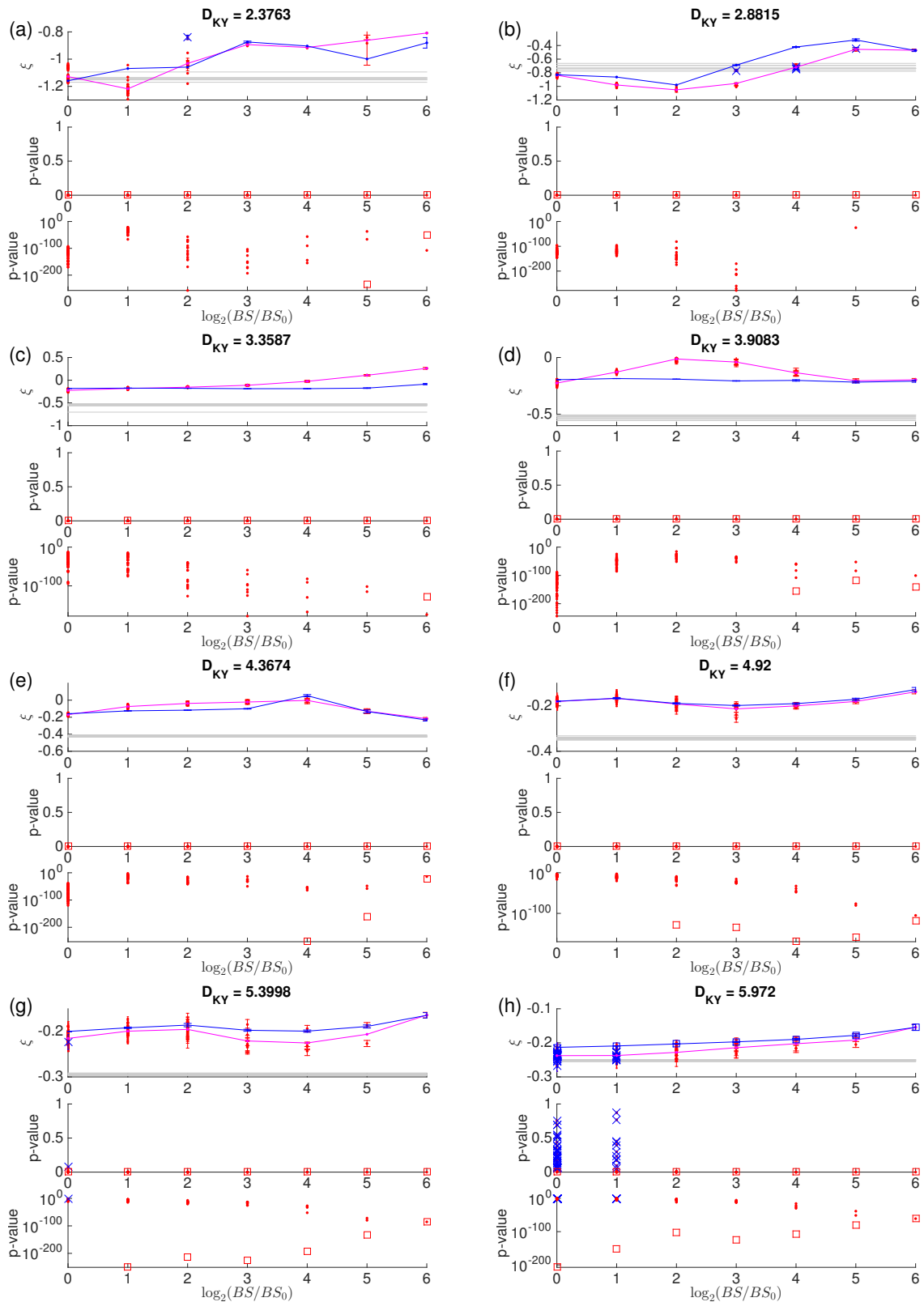
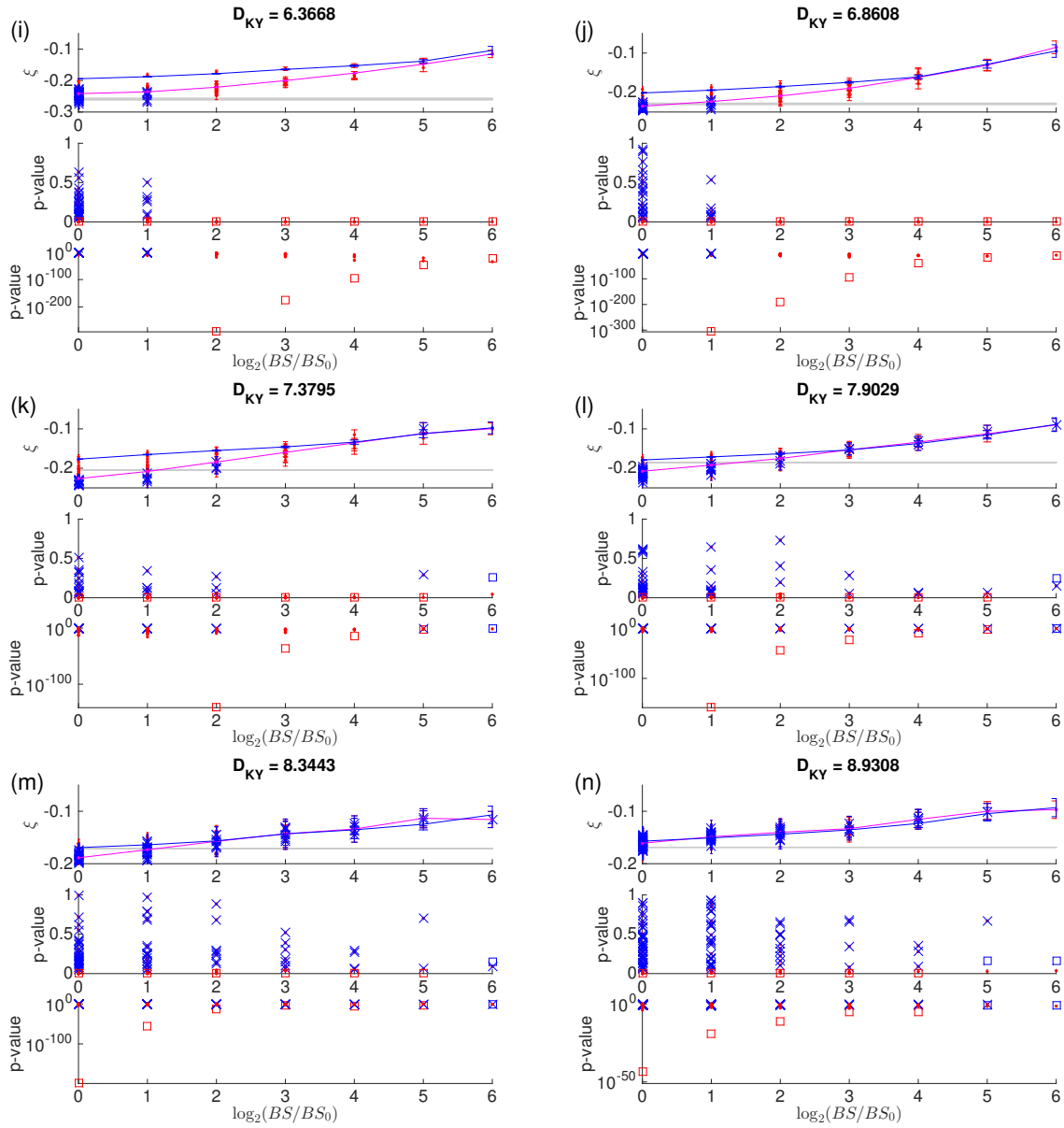Figure 11: Estimates of the shape parameter ξ, continued. . .

24

Figure 11: Estimates of the shape parameter ξ for x_j's of L96, along with estimation diagnostics. We take the same $BS_0 = 40$ [MTU] as for L84. Gray lines mark ten theoretical predictions $\xi \approx -1/D_1$. Time series data for x_j, $j = 1, \ldots, \hat{N}$, $\hat{N} = 10$, shuffled into one (magenta) or concatenated (blue) are used. A (no) bootstrapping is used to obtain the magenta (blue) curve (providing ξ). The magenta curve belongs to red dot markers ($p < 0.05$) and blue $\times$ markers ($p > 0.05$); and the blue curve belongs to red ($p < 0.05$) and blue ($p > 0.05$) square $\square$ markers. The sample values of F used, as indicated by the green circle markers in Fig. 10, are: (a) 4.9354 (b) 4.8958 (c) 4. 8580 (d) 4.9771 (e) 5.1654 (f) 5.5250 (g) 6.1991 (h) 6.9176 (i) 7.7255 (j) 8.8154 (k) 10.7137 (l) 13.7581 (m) 19.2049 (n) 34.2371. These F values yield D_{KY} values sufficiently close to and below the desired integer dimensions $D_{KY} = i$, $i = 3, \ldots, 9$, and fractional dimensions $0.4 + i$, $i = 2, \ldots, 8$; the actual values are written on top of each panel.
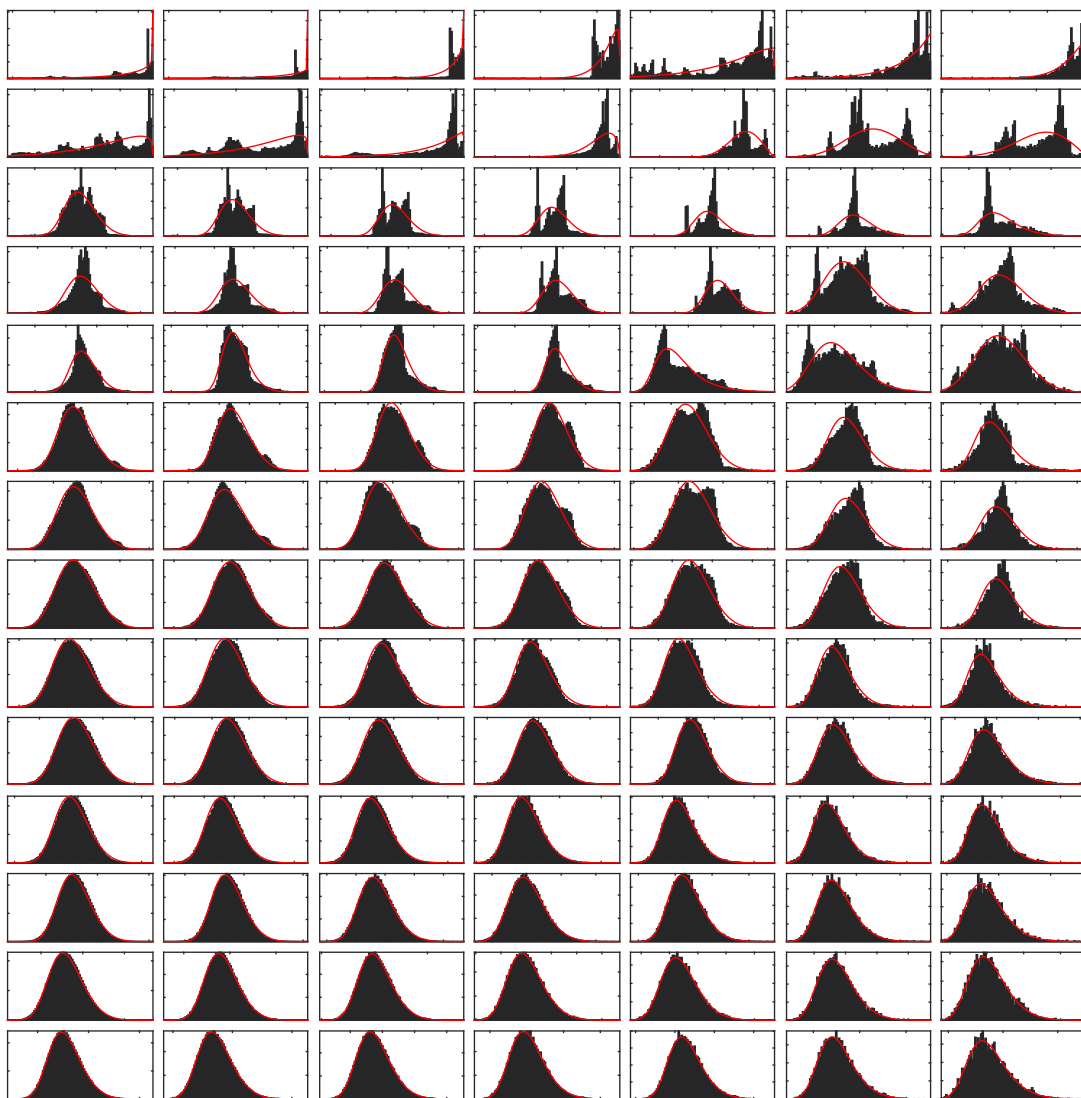
Figure 12: Stamp diagram of distribution densities ρ_m of block maxima. No bootstrapping is done. Red curves give the GEVD fits by MLE. This is a companion diagram to Fig. 11; the consecutive rows correspond to consecutive panels of the said figure, and consecutive columns correspond to the increasing sample values of m.
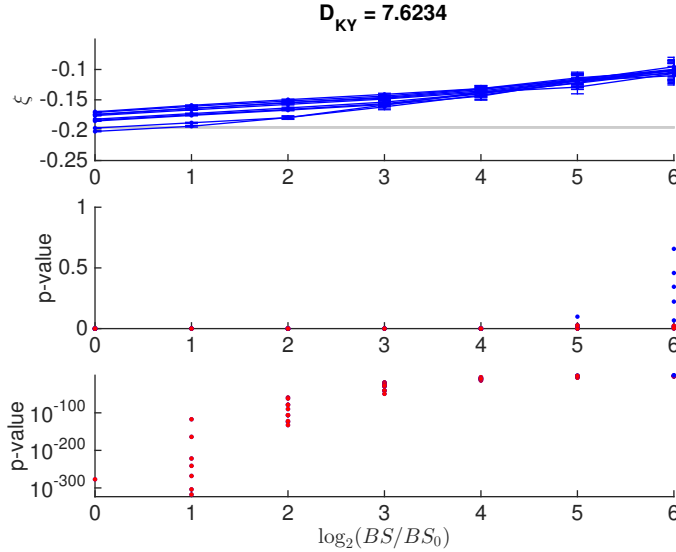
Figure 13: Estimates of the shape parameter ξ for the different x_j's of L96 separately with $F = 12$, $\hat{N} = 10$. No bootstrapping is done.

This argument leads us to examine a scenario of much higher dimension, for which the averaging by projection is expected to be 'complete' enough to yield the right GEVD. This seems to be true for the scenario investigated in Fig. 14. The theoretical value is finally approximated rather well. The problem is still, however, that when the estimate is still very inaccurate (the bias being much larger than the variance of bootstrapping estimates) using small BS's, the p-value can be very high, providing false diagnostics.

0.5 Conclusions

We argued in this chapter that singular measures on chaotic attractors prevent having an extreme value law (EVL), both for geometric distance (Fig. 5) and physical observables (Fig. 12). For distance observables at least, an EVL can be found when block maxima (BMa) of distances from all points on the attractor lumped together are considered (Figs. 6 and 7). This is essentially an averaging of distributions of BMa.

For physical observables such averaging over different observables is not defined. Perhaps self-similar (but not identical) distributions of BMa can be averaged, increasing the block size (BS) infinitely. Otherwise, the projection of the measure onto the single dimension of the scalar physical observable does achieve some sort of averaging (Fig. 12). In practice for some high-dimensional system one might be unable to reject a hypotheses that distributions of BMa come from a Generalized Extreme Value Distribution (GEVD). In our experience, on the Lorenz 96 model, this can be the case even if the estimates of the shape parameters are still very inaccurate (Figs. 11 and 14).

In another example studied by Gálfi (2015), a quasi-geosthropic two-layer model of midlatitude atmospheric dynamics, estimates of the shape parameter seem to converge from below to the true value (say, $\xi \approx -0.002$), but very slowly. This is believed to be because the system is very high-dimensional ($D_{KY} \approx 600$), and so its observables are approximately normally distributed, entailing an (approximately) Gumbel EVD ($\xi \approx 0$). The slow convergence of the ξ estimates should have to do with the logarithmic rescaling of the distribution of BMa, as indicated by $b_m = \ln m$ for Example 1 in Sec. 0.2.1. However, it is still open to question whether this poor ability of estimating ξ accurately would precipitate in poor estimations of *average return times* for events of some unobserved
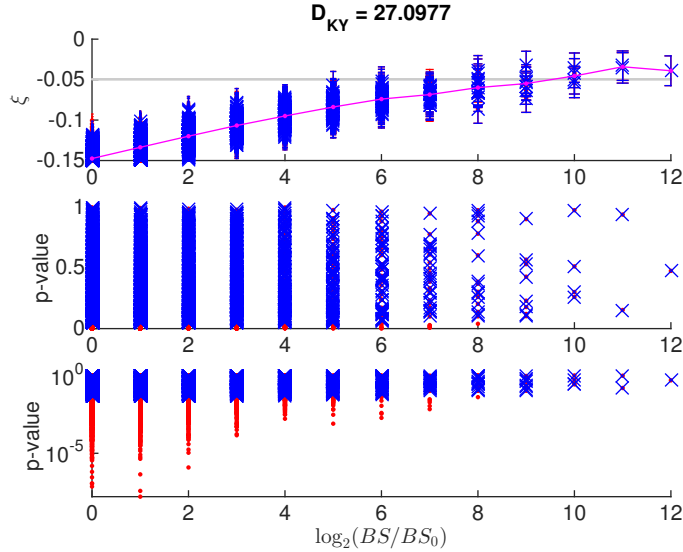
Figure 14: Estimates of the shape parameter ξ for the x_j's of L96, with $\hat{N} = 32$ and $F = 20$, in the spirit of Fig. 11. Bootstrapping is done. The p-value for each MLE is indicated by red dot markers ($p < 0.05$) and blue $\times$ markers ($p > 0.05$).

magnitude. One thing seems to be certain though: the magnitude of the events of interest cannot be just any large when the predicted tail is shorter than the actual one.

It is rather ironic that the theoretical prediction of the shape parameter (10) applies only to rather high-dimensional systems (Fig. 14), in which case it is difficult to estimate this value from limited data because of a slow convergence, and perhaps not even needed, because it is known to be very close to zero.

The problem of low-versus-high dimensionality is relevant to climate research. The climate system encompassing the fluids of the atmosphere and ocean which feature turbulent motion is clearly a high dimensional system. Nevertheless, it does feature scale separations and partial decoupling of subsystems, such that some large length scale quantities are governed dominantly by rather low-order dynamics, and global influences can be modeled as noise. An example of this is the El Niño Southern Oscillation (ENSO) (Dijkstra, 2005).

We attempted to demonstrate that a diagnostics of modeling extreme value data by the GEVD is necessary, even if not sufficient. We preferred to calculate the p-value of the χ^2-test. We showed on the example of the Lorenz 84 model that estimates of ξ best approximate the theoretical prediction for some intermediate (in the range the data set size allows for) BS (Figs. 6 and 8), and this 'optimal' BS was prompted by large p-values ($p > 0.05$). A seemingly asymptotic convergence of $\xi(BS)$ was misleading, however, 'approaching' a value that the low p-values prompted to be spurious, not belonging to a GEVD. In (Bódai et al., 2013) the authors relied on such an apparent convergence of $\xi(BS)$, and concluded that ξ depends sensitively on the (relative) time scale τ of the driving process. We have now evaluated the theoretical predictions for ξ as a function of τ (see Fig. 15), and although some sensitivities are revealed, the range of variation is far less than that suggested by their Fig. 11 (b). Still, the bottom line is that in this low-dimensional setting the theoretical prediction does not even apply – classical EVT fails this situation, distributions of BMa of the 'unaveraged/unaverageable' physical observable do not conform to a GEVD.

However, dissimilarly to the problem with physical observables, in our example in Fig. 6 for a distance observable, in spite of averaging over the attractor, we encounter the said non-GEVD characteristics. This was attributed to what we called 'folds' – with moderate confidence though.
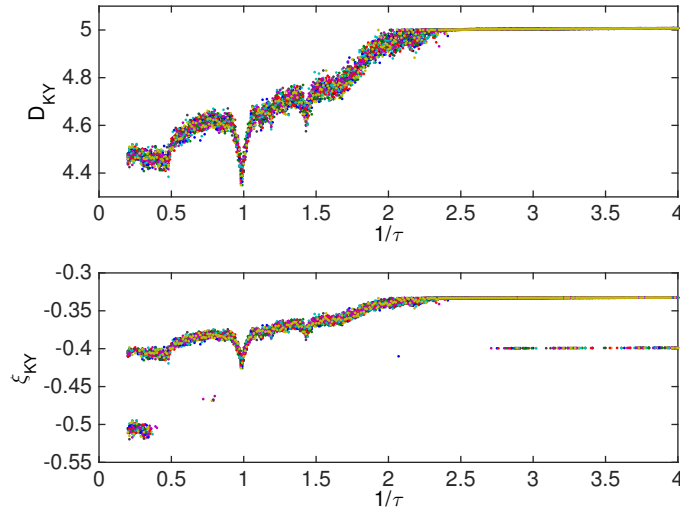
Figure 15: Theoretical predictions of ξ for the variable $r = \sqrt{y^2 + z^2}$ in L84 driven by the first component of the Rössler system, as defined by Bódai et al. (2013).

Such folds can occur at any small finite scale, and so crossovers in $\xi(BS)$ can happen 'unexpectedly'. This means that even if an EVL exists (and if so, instead of a 'crossover' we should say 'anomaly'), the convergence of the estimates of parameters of the GEVD (and even the goodness of this estimation reflected by e.g. the p-value) do not obey universal laws. Needless to say, this could hinder predictions of unobserved events.

It might well add to the practical problem that at finite BS's allowed by the data set size, determinism could still be felt, as the decorrelation between values separated by time is not instantaneous.

With the addition of noise of some sort to data generated by a deterministic system possessing an attractor of a fractional dimension, an EVL could possibly be established. This is also the case with deterministic systems that show regular (periodic, quasi-periodic (Nicolis et al., 2006)) behaviour (Faranda et al., 2013), which did not in themselves allow for an EVL either. However, the effects of noise can be detected at scales depending on the noise strength. It is typical to still see fractality at larger scales (Ben-Mizrachi et al., 1984). Therefore, it could be expected in some situations that estimates $\xi(BS)$ in a noisy dynamical system have a crossover, at some predictable BS depending on the noise strength, from spurious values due to fractality towards the true value.

The latter has implications for stochastic parametrization schemes intended for a next generation of climate/Earth system models (Franzke et al., 2015). It should be appreciated that at some quantiles the tails of the distributions in this model are dominantly characterized by the parametrization. The true (asymptotic) EVS in a high-resolution deterministic physics-resolving model is an emergent property of the model, whereas in a lower-resolution model that represents the sub-grid scale physics by a stochastic parametrization scheme, the EVS should be considered an emergent property (dominantly) of the parametrization scheme when it had been constructed to respect other features than the EVS.

Acknowledgements

Bibliography

Balakrishnan, V., Nicolis, C., and Nicolis, G. (1995). Extreme value distributions in chaotic dynamics. *Journal of Statistical Physics*, 80(1):307–336.

Ben-Mizrachi, A., Procaccia, I., and Grassberger, P. (1984). Characterization of experimental (noisy) strange attractors. *Phys. Rev. A*, 29:975–977.

Bódai, T. (2015). Predictability of threshold exceedances in dynamical systems. *Physica D: Nonlinear Phenomena*, 313:37 – 50.

Bódai, T., Károlyi, G., and Tél, T. (2011). Fractal snapshot components in chaos induced by strong noise. *Phys. Rev. E*, 83:046201.

Bódai, T., Károlyi, G., and Tél, T. (2013). Driving a conceptual model climate by different processes: Snapshot attractors and extreme events. *Phys. Rev. E*, 87:022822.

Brandstater, A. and Swinney, H. L. (1987). Strange attractors in weakly turbulent Couette-Taylor flow. *Phys. Rev. A*, 35:2207–2220.

Coles, S. (2001). *An Introduction to Statistical Modeling of Extreme Values*. Springer.

Collet, P. (2001). Statistics of closest return for some non-uniformly hyperbolic systems. *Ergod. Theory Dyn. Syst.*, 21:401–420.

Cross, M. C. and Hohenberg, P. C. (1993). Pattern formation outside of equilibrium. *Rev. Mod. Phys.*, 65:851–1112.

Dijkstra, H. A. (2005). *Nonlinear Physical Oceanography*. Springer, Dordrecht.

Drótos, G., Bódai, T., and Tél, T. (2015). Probabilistic concepts in a changing climate: a snapshot attractor picture. *Journal of Climate*, 28:3275–3288.

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., Lucarini, V., Turchetti, G., and Vaienti, S. (2011). Numerical convergence of the block-maxima approach to the Generalized extreme value distribution. *Journal of Statistical Physics*, 145(5):1156–1180.

Field, C., Barros, V., and Stocker, T. (2012). *Managing the risks of extreme events and disasters to advance climate change adaptation. Special report of the Intergovernmental Panel on Climate Change (IPCC)*. Cambridge University Press, New York, NY (United States).

Fisher, R. A. and Tippett, L. H. C. (1928). Limiting forms of the frequency distribution of the largest and smallest member of a sample. *Proc. Cambridge Phil. Soc.*, 24:180–190.

Franzke, C. L. E., O'Kane, T. J., Berner, J., Williams, P. D., and Lucarini, V. (2015). Stochastic climate theory and modeling. *Wiley Interdisciplinary Reviews: Climate Change*, 6(1):63–78.

Freitas, A. C. M. and Freitas, J. M. (2008). On the link between dependence and independence in extreme value theory for dynamical systems. *Statistics & Probability Letters*, 78(9):1088 – 1093.

Freitas, A. C. M., Freitas, J. M., and Todd, M. (2009). Hitting time statistics and extreme value theory. *Probability Theory and Related Fields*, 147(3):675–710.

Gálfi, V. M. (2015). Extreme events in a two-layer quasi-geostrophic atmospheric model. Master's thesis, University of Hamburg.

Gallavotti, G. and Lucarini, V. (2014). Equivalence of non-equilibrium ensembles and representation of friction in turbulent flows: The Lorenz 96 model. *Journal of Statistical Physics*, 156(6):1027–1065.

Gnedenko, B. V. (1943). Sur la distribution limite du terme maximum d'une serie aleatoire. *Annals of Mathematics*, 44:423–453.

Gomes, M. I. (1984). Penultimate limiting forms in extreme value theory. *Annals of the Institute of Statistical Mathematics*, 36(1):71–85.

Govorukhin, V. (2004). Calculation Lyapunov exponents for ODE. http://de.mathworks.com/matlabcentral/fileexchange/4628-calculation-lyapunov-exponents-for-ode/content/lyapunov.m.

Haiman, G. (2003). Extreme values of the tent map process. *Statistics & Probability Letters*, 65(4):451 – 456.

Holland, M. P., Vitolo, R., Rabassa, P., Sterk, A. E., and Broer, H. W. (2012). Extreme value laws in dynamical systems under physical observables. *Physica D: Nonlinear Phenomena*, 241(5):497 – 513.

Kantz, H. and Schreiber, T. (2003). *Nonlinear Time Series Analysis*. Cambridge University Press.

Karimi, A. and Paul, M. R. (2010). Extensive chaos in the Lorenz-96 model. *Chaos*, 20(4).

Leadbetter, M. R., Lindgren, G., and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*. Springer-Verlag.

Lorenz, E. (1996). Predictability – a problem partly solved. In *Seminar on Predictability*. ECMWF.

Lorenz, E. N. (1984). Irregularity: a fundamental property of the atmosphere. *Tellus*, 36A:98–110.

Lucarini, V., Faranda, D., de Freitas, A. C. G. M. M., de Freitas, J. M. M., Holland, M., Kuna, T., Nicol, M., Todd, M., and Vaienti, S. (2016). *Extremes and Recurrence in Dynamical Systems*. Wiley.

Lucarini, V., Faranda, D., Turchetti, G., and Vaienti, S. (2012). Extreme value theory for singular measures. *Chaos*, 22(2).

Lucarini, V., Faranda, D., Wouters, J., and Kuna, T. (2014). Towards a general theory of extremes for observables of chaotic dynamical systems. *Journal of Statistical Physics*, 154(3):723–750.

Nicolis, C., Balakrishnan, V., and Nicolis, G. (2006). Extreme events in deterministic dynamical systems. *Phys. Rev. Lett.*, 97:210602.

Nicolis, C. and Nicolis, G. (2012). Extreme events in multivariate deterministic systems. *Phys. Rev. E*, 85:056217.

Ott, E. (2002). *Chaos in Dynamical Systems*. Cambridge University Press.

Provenzale, A. and Balmford, N. J. (1999). Chaos and structures in geophysics and astrophysics (Woods Hole lecture notes).

Romano, J. P. and Siegel, A. F. (1985). *Counterexamples in probability and statistics*. Chapman & Hall, Great Britain.

Tél, T. and Gruiz, M. (2006). *Chaotic Dynamics*. Cambridge University Press, Cambridge.

van Veen, L. (2003). Baroclinic flow and the Lorenz-84 model. *International Journal of Bifurcation and Chaos*, 13(08):2117–2139.

Vannitsem, S. (2007). Statistical properties of the temperature maxima in an intermediate order quasi-geostrophic model. *Tellus A*, 59(1).

Zeidler, E. (2004). *User's Guide to Mathematics*. Oxford University Press.