



## Extreme events: dynamics, statistics and prediction

M. Ghil, Pascal Yiou, Stéphane Hallegatte, B. D. Malamud, P. Naveau, A. Soloviev, P. Friederichs, V. Keilis-Borok, D. Kondrashov, V. Kossobokov, et al.

### ► To cite this version:

M. Ghil, Pascal Yiou, Stéphane Hallegatte, B. D. Malamud, P. Naveau, et al.. Extreme events: dynamics, statistics and prediction. *Nonlinear Processes in Geophysics*, 2011, 18 (3), pp.295-350. 10.5194/npg-18-295-2011 . hal-00716514

**HAL Id: hal-00716514**

**<https://enpc.hal.science/hal-00716514>**

Submitted on 4 Sep 2020

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

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



Distributed under a Creative Commons Attribution 4.0 International License

## Extreme events: dynamics, statistics and prediction

M. Ghil<sup>1,2</sup>, P. Yiou<sup>3</sup>, S. Hallegatte<sup>4,5</sup>, B. D. Malamud<sup>6</sup>, P. Naveau<sup>3</sup>, A. Soloviev<sup>7</sup>, P. Friederichs<sup>8</sup>, V. Keilis-Borok<sup>9</sup>, D. Kondrashov<sup>2</sup>, V. Kossobokov<sup>7</sup>, O. Mestre<sup>5</sup>, C. Nicolis<sup>10</sup>, H. W. Rust<sup>3</sup>, P. Shebalin<sup>7</sup>, M. Vrac<sup>3</sup>, A. Witt<sup>6,11</sup>, and I. Zaliapin<sup>12</sup>

<sup>1</sup>Environmental Research and Teaching Institute (CERES-ERTI), Geosciences Department and Laboratoire de Météorologie Dynamique (CNRS and IPSL), UMR8539, CNRS-Ecole Normale Supérieure, 75231 Paris Cedex 05, France

<sup>2</sup>Department of Atmospheric & Oceanic Sciences and Institute of Geophysics & Planetary Physics, University of California, Los Angeles, USA

<sup>3</sup>Laboratoire des Sciences du Climat et de l'Environnement, UMR8212, CEA-CNRS-UVSQ, CE-Saclay l'Orme des Merisiers, 91191 Gif-sur-Yvette Cedex, France

<sup>4</sup>Centre International pour la Recherche sur l'Environnement et le Développement, Nogent-sur-Marne, France

<sup>5</sup>Météo-France, Toulouse, France

<sup>6</sup>Department of Geography, King's College London, London, UK

<sup>7</sup>International Institute of Earthquake Prediction Theory and Mathematical Geophysics, Russian Academy of Sciences, Russia

<sup>8</sup>Meteorological Institute, University Bonn, Bonn, Germany

<sup>9</sup>Department of Earth & Space Sciences and Institute of Geophysics & Planetary Physics, University of California, Los Angeles, USA

<sup>10</sup>Institut Royal de Météorologie, Brussels, Belgium

<sup>11</sup>Department of Nonlinear Dynamics, Max-Planck Institute for Dynamics and Self-Organization, Göttingen, Germany

<sup>12</sup>Department of Mathematics and Statistics, University of Nevada, Reno, NV, USA

Received: 3 December 2010 – Revised: 15 April 2011 – Accepted: 18 April 2011 – Published: 18 May 2011

**Abstract.** We review work on extreme events, their causes and consequences, by a group of European and American researchers involved in a three-year project on these topics. The review covers theoretical aspects of time series analysis and of extreme value theory, as well as of the deterministic modeling of extreme events, via continuous and discrete dynamic models. The applications include climatic, seismic and socio-economic events, along with their prediction.

Two important results refer to (i) the complementarity of spectral analysis of a time series in terms of the continuous and the discrete part of its power spectrum; and (ii) the need for coupled modeling of natural and socio-economic systems. Both these results have implications for the study and prediction of natural hazards and their human impacts.

### Table of contents

<b>1</b>	<b>Introduction and motivation</b>	<b>297</b>
<b>2</b>	<b>Time series analysis</b>	<b>297</b>
2.1	Background . . . . .	297
2.2	Oscillatory phenomena and periodicities . . .	298
2.2.1	Singular spectrum analysis (SSA) . .	298
2.2.2	Spectral density and spectral peaks . .	299
2.2.3	Significance and reliability issues . .	300
2.3	Geoscientific records with heavy-tailed distributions . . . . .	301
2.3.1	Introduction . . . . .	301
2.3.2	Frequency-size distributions . . . . .	301
2.3.3	Heavy-tailed distributions: Practical issues . . . . .	301
2.4	Memory and long-range dependence (LRD) .	302
2.4.1	Introduction . . . . .	302
2.4.2	The auto-correlation function (ACF) and the memory of time series . . . .	303
2.4.3	Quantifying long-range dependence (LRD) . . . . .	303



Correspondence to: M. Ghil  
 (ghil@lmd.ens.fr)

2.4.4	Stochastic processes with the LRD property . . . . .	304	<b>6 Macroeconomic modeling and impacts of natural hazards</b>	<b>325</b>	
2.4.5	LRD: Practical issues . . . . .	305	6.1	Modeling endogenous economic dynamics . . . . .	326
2.4.6	Simulation of fractional-noise and LRD processes . . . . .	306	6.2	Endogenous dynamics and exogenous shocks . . . . .	327
			6.2.1	Modeling economic effects of natural disasters . . . . .	327
			6.2.2	Interaction between endogenous dynamics and exogenous shocks . . . . .	328
			6.2.3	Distribution of disasters and economic dynamics . . . . .	329
<b>3</b>	<b>Extreme value theory (EVT)</b>	<b>306</b>	<b>7 Prediction of extreme events</b>	<b>330</b>	
3.1	A few basic concepts . . . . .	306	7.1	Prediction of oscillatory phenomena . . . . .	330
3.2	Multivariate EVT . . . . .	308	7.2	Prediction of point processes . . . . .	331
3.2.1	Max-stable random vectors . . . . .	308	7.2.1	Earthquake prediction for the Vrancea region . . . . .	332
3.2.2	Multivariate varying random vectors . . . . .	309	7.2.2	Prediction of extreme events in socio-economic systems . . . . .	332
3.2.3	Models and inference for multivariate EVT . . . . .	309	7.2.3	Applications to economic recessions, unemployment and homicide surges . . . . .	333
3.3	Nonstationarity, covariates and parametric models . . . . .	309			
3.3.1	Parametric and semi-parametric approaches . . . . .	310	<b>8 Concluding remarks</b>	<b>335</b>	
3.3.2	Smoothing extremes . . . . .	310	<b>A Parameter estimation for GEV and GPD distributions</b>	<b>336</b>	
3.4	EVT and memory . . . . .	311	<b>B Bivariate point processes</b>	<b>337</b>	
3.4.1	Dependence and the block-maxima approach . . . . .	311	<b>C The M8 earthquake prediction algorithm</b>	<b>337</b>	
3.4.2	Clustering of extremes and the POT approach . . . . .	311	<b>D The pattern-based prediction algorithm</b>	<b>338</b>	
3.5	Statistical downscaling and weather regimes . . . . .	311	<b>E Table of acronyms</b>	<b>339</b>	
3.5.1	Bridging time and space scales . . . . .	311			
3.5.2	Downscaling methodologies . . . . .	312			
3.5.3	Two approaches to downscaling extremes . . . . .	312			
<b>4</b>	<b>Dynamical modeling</b>	<b>313</b>			
4.1	A quick overview of dynamical systems . . . . .	313			
4.2	Signatures of deterministic dynamics in extreme events . . . . .	314			
4.3	Boolean delay equations (BDEs) . . . . .	316			
4.3.1	Formulation . . . . .	317			
4.3.2	BDEs, cellular automata and Boolean networks . . . . .	317			
4.3.3	Towards hyperbolic and parabolic “partial” BDEs . . . . .	318			
<b>5</b>	<b>Applications</b>	<b>318</b>			
5.1	Nile River flow data . . . . .	318			
5.1.1	Spectral analysis of the Nile River water levels . . . . .	319			
5.1.2	Long-range persistence in Nile River water levels . . . . .	320			
5.2	BDE modeling in action . . . . .	321			
5.2.1	A BDE model for the El Niño/Southern Oscillation (ENSO) . . . . .	322			
5.2.2	A BDE model for seismicity . . . . .	323			

## 1 Introduction and motivation

Extreme events are a key manifestation of complex systems, in both the natural and human world. Their economic and social consequences are a matter of enormous concern. Much of science, though, has concentrated – until two or three decades ago – on understanding the mean behavior of physical, biological, environmental or social systems and their “normal” variability. Extreme events, due to their rarity, have been hard to study and even harder to predict.

Recently, a considerable literature has been devoted to these events; see, for instance, Katz et al. (2002), Smith (2004), Sornette (2004), Albeverio et al. (2005) and references therein. Still, much of this literature has relied heavily on classical extreme value theory (EVT) or on studying frequency-size distributions with a heavy-tailed character. Prediction has therefore been limited, by-and-large, to the expected time for the occurrence of an event above a certain size.

This review paper does not set out to cover the extensive knowledge accumulated recently on the theory and applications of extreme events: the task would have required at least a book, which would rapidly be overtaken by ongoing research. Instead, this paper is trying to summarize and examine critically the numerous results of a project on “Extreme Events: Causes and Consequences (E2-C2)” that brought together, over more than three years, 70–80 researchers belonging to 17 institutions in nine countries. The project produced well over 100 research papers in the refereed literature and providing some perspective on all this work might have therefore some merit.

We set out to develop methods for the description, understanding and prediction of extreme events across a range of natural and socio-economic phenomena. Good definitions for an object of study only get formulated as that study nears completion; hence there is no generally accepted definition of extreme events in the broad sense we wished to study them in the E2C2 project. Still, one can use the following operating definition, proposed by Kantz et al. (2005). Extreme events (i) are rare, (ii) they occur irregularly, (iii) they exhibit an observable that takes on an extreme value, and (iv) they are inherent to the system under study, rather than being due to external shocks.

General tools were developed to extract the distribution of these events from existing data sets. Models that are anchored in complex-systems concepts were constructed to incorporate a priori knowledge about the phenomena and to reproduce the data-derived distribution of events. These models were then used to predict the likelihood of extreme events in prescribed time intervals. The methodology was applied to three sets of problems: (i) natural disasters from the realms of hydrology, seismology, and geomorphology; (ii) socio-economic crises, including those associated with recessions, criminality, and unemployment; and (iii) rapid, potentially

catastrophic changes in the interaction between economic activity and climate variability.

The paper is organized as follows. In Sect. 2 we introduce several advanced methods of time series analysis, for the detection and description of spectral peaks, as well as for the study of the continuous spectrum. Section 3 addresses EVT, including multivariate EVT, nonstationarity and long-memory effects; some technical details appear in Appendices A and B.

Dynamical modeling of extreme events is described in Sect. 4. Both discrete-valued models, like cellular automata and Boolean delay equations (BDEs), and continuous-valued models, like maps and differential equations, are covered. Several applications are described in Sect. 5, for modeling as well as for time series analysis. It is here that we refer the reader to a few of the other papers in this special issue, for many more applications. Nonequilibrium macroeconomic models are outlined in Sect. 6, where we show how the use of such models affects the conclusions about the impact of natural hazards on the economy. Prediction of extreme events is addressed in Sect. 7. Here we distinguish between the prediction of extremes in continuous, typically differentiable functions, like temperatures, and point processes, like earthquakes and homicide surges. The paper concludes with Sect. 8 that provides a quick summary and some interesting directions of future development. Appendix E lists the numerous acronyms that appear in this fairly multidisciplinary review.

This review paper is part of a Special Issue on “Extreme Events: Nonlinear Dynamics and Time Series Analysis”, built around the results of the E2-C2 project, but not restricted to its participants. The Special Issue is introduced by the Guest Editors – B. Malamud, H. W. Rust and P. Yiou – and contains 14 research papers on the theory of extreme events and its applications to many phenomena in the geosciences. Theoretical developments are covered by Abaimov et al. (2007); Bernacchia and Naveau (2008); Blender et al. (2008); Schölzel and Friederichs (2008) and Serinaldi (2009), while the applications cover the fields of meteorology and climate dynamics (Vannitsem and Naveau, 2007; Vrac et al., 2007a; Bernacchia et al., 2008; Ghil et al., 2008b; Taricco et al., 2008; Yiou et al., 2008), seismology (Soloviev, 2008; Narteau et al., 2008) and geomagnetism (Anh et al., 2007). Many other peer-reviewed papers were published over the duration of the project and data sets and other unrefereed products can be found on the web sites of the participating institutions.

## 2 Time series analysis

### 2.1 Background

This section gives a quick perspective on the “nonlinear revolution” in time series analysis. In the 1960s and 1970s, the

scientific community realized that much of the irregularity in observed time series, which had traditionally been attributed to the random “pumping” of a linear system by infinitely many (independent) degrees of freedom (DOF), could be generated by the nonlinear interaction of a few DOF (Lorenz, 1963; Smale, 1967; Ruelle and Takens, 1971). This realization of the possibility of deterministic aperiodicity or “chaos” (Gleick, 1987) created quite a stir.

The purpose of this section is to describe briefly some of the implications of this change in outlook for time series analysis, with a special emphasis on environmental time series. Many general aspects of nonlinear time series analysis are reviewed by Drazin and King (1992), Ott et al. (1994), Abarbanel (1996), and Kantz and Schreiber (2004). We concentrate here on those aspects that deal with regularities and have proven most useful in studying environmental variability.

A connection between deterministically chaotic time series and the nonlinear dynamics generating them was attempted fairly early in the young history of “chaos theory”. The basic idea was to consider specifically a scalar, or univariate, time series with apparently irregular behavior, generated by a deterministic or stochastic system. This time series could be exploited, so the thinking went, in order to ascertain, first, whether the underlying system has a finite number of DOF. An upper bound on this number would imply that the system is deterministic, rather than stochastic, in nature. Next, we might be able to verify that the observed irregularity arises from the fractal nature of the deterministic system’s invariant set, which would yield a fractional, rather than integer, value of this set’s dimension. Finally, one could maybe reconstruct the invariant set or even the equations governing the dynamics from the data.

Environmental time series, as well as most other time series from nature or the laboratory, are likely to be generated by forced dissipative systems (Lorenz, 1963; Ghil and Childress, 1987, Chap. 5). The invariant sets associated with irregularity here are “strange attractors” (Ruelle and Takens, 1971), toward which all solutions tend asymptotically; long-term irregular behavior in such systems is associated with such attractors. Proving rigorously the existence of these objects and their fractal character, however, has been hard (Guckenheimer and Holmes, 1997; Lasota and Mackey, 1994; Tucker, 1999).

Under general assumptions, most physical systems and many biological and socio-economic ones can be described by a system of nonlinear differential equations:

$$\dot{X}_i = F_i(X_1, \dots, X_j, \dots, X_p), \quad 1 \leq i, j \leq p, \quad (1)$$

where  $X = (X_1, \dots, X_p)$  is a vector of variables and the  $F_i$  on the right-hand side are continuously differentiable functions. The phase space  $X$  of the system (1) lies in the Euclidean space  $\mathbb{R}^p$ .

In general, the  $F_i$ ’s are not well known, and an observer has only access to one time series  $\hat{X}(t) = X_{i_0}(t)$ , for some

$1 \leq i_0 \leq p$  and for a finite time. A challenge is thus to assess the properties of the underlying attractor of the system (1), given partial knowledge of only one of its components. The ambitious program to do so (Packard et al., 1980; Roux et al., 1980; Ruelle, 1981) relied essentially on the *method of delays*, based on the Whitney (1936) embedding lemma and the Mañé (1981) and Takens (1981) theorems.

First of all, the data  $\hat{X}(t)$  are typically given at discrete times  $t = n\Delta t$  only. Next, one has to admit that it is hard to actually get the right-hand sides  $F_i$ ; instead, one attempts to reconstruct the invariant set on which the solutions of Eq. (1) lie.

Mañé (1981), Ruelle (1981) and Takens (1981) had the idea, developed further by Sauer et al. (1991), that a single observed time series  $\hat{X} = X_{i_0}$  or, more generally, some scalar function of  $X$ ,

$$\hat{X}(t) = \phi(X_1(t), \dots, X_p(t)),$$

could be used to reconstruct the attractor of a forced dissipative system. The basis for this reconstruction idea is essentially the fact that, subject to certain technical conditions, such a solution covers the attractor densely; thus, as time increases, it will pass arbitrarily close to any point on the attractor. Time series observed in the natural environment, however, have finite length and sampling rate, as well as significant measurement noise.

The embedding idea has been applied therefore most successfully to time series generated numerically or by laboratory experiments in which sufficiently long series could be obtained and noise was controlled better than in nature. Broomhead and King (1986), for instance, successfully applied singular spectrum analysis (SSA) to the reconstruction of the Lorenz (1963) attractor. As we shall see below, for time series in the geosciences and other natural and socio-economic sciences, it might only be possible to attain a more modest goal: to describe merely a “skeleton” of the attractor, which is formed by a few robust periodic orbits. This motivates the next section on SSA. Applications are discussed in Sect. 5.1.

## 2.2 Oscillatory phenomena and periodicities

### 2.2.1 Singular spectrum analysis (SSA)

Given a discretely sampled time series  $X(t)$ , the goal of SSA is to determine properties of a reconstructed attractor – more precisely, to find its skeleton, formed by the least unstable periodic orbits on it – by the method of delays. SSA can also serve for data compression or smoothing, in preparation for the application of other spectral-analysis methods; see Sect. 2.2.2.

The main idea is to exploit the covariance matrix of the sequence of “delayed” vectors  $X(t) = (X_t, \dots, X_{t+M-1})$ , where  $1 \leq t \leq N - M + 1$ , in a well-chosen phase-space dimension  $M$ . The eigenelements of this covariance matrix

provide the directions and degree of extension of the attractor of the underlying system.

By analogy with the meteorological literature, the eigenvectors are generally called *empirical orthogonal functions* (EOFs). Each EOF  $\rho_k$  carries a fraction of the variance of  $X$  that is proportional to the corresponding eigenvalue  $\lambda_k$ . The projection  $A_k$  of the time series  $X$  on each EOF  $\rho_k$  is called a principal component (PC),

$$A_k(t) = \sum_{j=1}^M X(t+j-1)\rho_k(j). \quad (2)$$

If a set  $\mathcal{K}$  of meaningful modes is identified, it can provide a mode-wise filter of the series  $X$  with so-called reconstructed components (RCs):

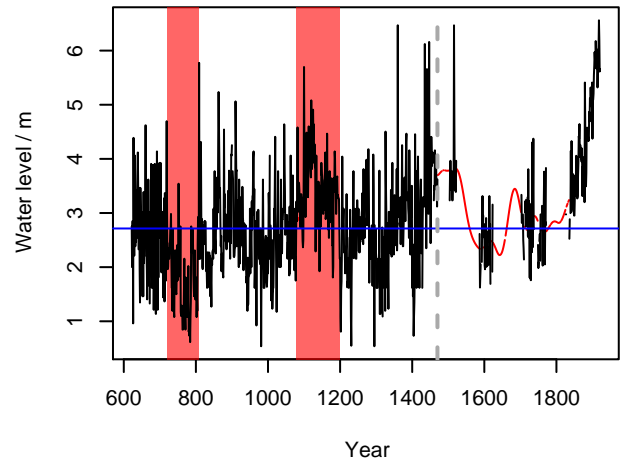
$$R_{\mathcal{K}}(t) = \frac{1}{M_t} \sum_{k \in \mathcal{K}} \sum_{j=L_t}^{U_t} A_k(t-j+1)\rho_k(j). \quad (3)$$

The normalization factor  $M_t$ , as well as the summation bounds  $L_t$  and  $U_t$ , differ between the central part of the time series and its end points; see Ghil et al. (2002) for the exact values.

Vautard and Ghil (1989) made the crucial observation that pairs of nearly equal eigenvalues may correspond to oscillatory modes, especially when the corresponding EOF and PC pairs are in phase quadrature. Moreover, the RCs have the property of capturing the phase of the time series in a well-defined least-squares sense, so that  $X(t)$  and  $R_{\mathcal{K}}(t)$  can be superimposed on the same time scale,  $1 \leq t \leq N$ . Hence, no information is lost in the reconstruction process, since the sum of all individual RCs gives back the original time series (Ghil and Vautard, 1991).

In the process of developing a methodology for applying SSA to climatic time series, a number of heuristic (Vautard and Ghil, 1989; Ghil and Mo, 1991; Unal and Ghil, 1995) or Monte Carlo-type (Ghil and Vautard, 1991; Vautard et al., 1992) methods have been devised for signal-to-noise separation and for the reliable identification of oscillatory pairs of eigenvalues. They are all essentially attempts to discriminate between the significant signal as a whole, or individual pairs, and white noise, which has a flat spectrum. A more stringent “null hypothesis” (Allen, 1992; Allen and Smith, 1996) is that of so-called red noise, since most geophysical and many other time series tend to have larger power at lower frequencies (Hasselmann, 1976; Mitchell, 1976; Ghil and Childress, 1987).

An important generalization of SSA is its application to multivariate time series, dubbed multi-channel SSA (Kepenne and Ghil, 1993; Plaut and Vautard, 1994; Ghil et al., 2002). Another significant extension of SSA is multi-scale SSA, which was devised to provide a data-adaptive method for analysing nonstationary, self-similar or heavy-tailed time series (Yiou et al., 2000).



**Fig. 1.** Annual minima of Nile River water levels for the 1300 yr between 622 AD (i.e., 1 AH) and 1921 AD (solid black line). The large gaps in the time series that occur after 1479 AD have been filled using SSA (cf. Sect. 2.2.1) (smooth red curve). The straight horizontal line (solid blue) shows the mean value for the 622–1479 AD interval, in which no large gaps occur; time intervals of large, persistent excursions from the mean are marked by red shading.

Kondrashov et al. (2005) proposed a data-adaptive algorithm to fill gaps in time series, based on the covariance structures identified by SSA. They applied this method to the 1300-yr-long time series (622–1921 AD) of Nile River records; see Fig. 1. We shall use this time series in Sect. 5.1 to illustrate the complementarity between the analysis of the continuous background, cf. Sects. 2.3 and 2.4, and that of the peaks rising above this background, as described in Sect. 2.2.2.

## 2.2.2 Spectral density and spectral peaks

Both deterministic (Eckmann and Ruelle, 1985) and stochastic (Hannan, 1960) processes can, in principle, be characterized by a function  $S$  of frequency  $f$ , rather than time  $t$ . This function  $S(f)$  is called the *power spectrum* in the engineering literature or the *spectral density* in the mathematical one. A very irregular time series – in the sense defined at the beginning of Sect. 2.1 – possesses a spectrum that is continuous and fairly smooth, indicating that all frequencies in a given band are excited by the process generating such a time series. On the other hand, a purely periodic or quasi-periodic process is described by a single sharp peak or by a finite number of such peaks in the frequency domain. Between these two end members, nonlinear deterministic but “chaotic” processes can have spectral peaks superimposed on a continuous and wiggly background (Ghil and Jiang, 1998; Ghil and Childress, 1987, Sect. 12.6).

In theory, for a spectral density  $S(f)$  to exist and be well defined, the dynamics generating the time series has to be ergodic and allow the definition of an invariant measure, with

respect to which the first and second moments of the generating process are computed as an ensemble average. In practice, the distinction between deterministically chaotic and truly random processes via spectral analysis can be as tricky as the attempted distinctions based on the dimension of the invariant set. In both cases, the difficulty is due to the shortness and noisiness of measured time series, whether coming from the natural or socio-economic sciences.

The spectral density  $S(f)$  is composed of a continuous part – often referred to as broad-band noise – and a discrete part. In theory, the latter is made up of Dirac  $\delta$  functions – also referred to as spectral lines: they correspond to purely periodic parts of the signal (Hannan, 1960; Priestley, 1992). In practice, spectral analysis methods attempt to estimate either the continuous part of the spectrum or the lines or both. The lines are often estimated from discrete and noisy data as more or less sharp “peaks.” The estimation and dynamical interpretation of the latter, when present, are at times more robust and easier to understand than the nature of the processes that might generate the broad-band background, whether deterministic or stochastic.

The numerical computation of the power spectrum of a random process is an ill-posed inverse problem (Jenkins and Watts, 1968; Thomson, 1982). For example, a straightforward calculation of the discrete Fourier transform of a random time series, which has a continuous spectrum, will provide a spectral estimate whose variance is equal to the estimate itself (Jenkins and Watts, 1968; Box and Jenkins, 1970). There are several approaches to reduce this variance.

The periodogram estimate of  $S(f)$  is the square Fourier transform of the time series  $X(t)$ .

$$\hat{S}(f) = \frac{1}{N} \left| \sum_{t=1}^N X(t) \exp(-2\pi i f t) \right|^2. \quad (4)$$

This estimate harbors two complementary problems: (i) its variance grows like the length  $N$  of the time series; and (ii) for a finite-length, discretely sampled time series, there is “power leakage” outside of any frequency band. The methods to overcome these problems (Percival and Walden, 1993; Ghil et al., 2002) differ in terms of precision, accuracy and robustness.

Among these, the multi-taper method (MTM: Thomson, 1982) offers a range of tests to check the statistical significance of the spectral features it identifies, with respect to various null hypotheses (Thomson, 1982; Mann and Lees, 1996). MTM is based on obtaining a set optimal weights, called the tapers, for  $X(t)$ ; these tapers provide a compromise between the two above-mentioned problems: each taper minimizes spectral leakage and optimizes the spectral resolution, while averaging over several spectral estimates reduces the variance of the final estimate.

A challenge in the spectral estimate of short and noisy time series is to determine the statistical significance of the peaks. Mann and Lees (1996) devised a heuristic test for

MTM spectral peaks with respect to a null hypothesis of red noise. Red noise is simply a Gaussian auto-regressive process of order 1, i.e., one that favours low frequencies and has no local maximum. Such a random process is a good candidate to test short and noisy time series, as already discussed in Sect. 2.2.1 above.

The key features of a few spectral-analysis methods are summarized in Table 3 of Ghil et al. (2002). The methods in that table include the Blackman-Tukey or correlogram method, the maximum-entropy method (MEM), MTM, SSA, and wavelets; the table indicates some of their strengths and weaknesses. Ghil et al. (2002) also describe precisely when and why SSA is advantageous as a data-adaptive pre-filter, and how it can improve the sharpness and robustness of subsequent peak detection by the correlogram method or MEM.

### 2.2.3 Significance and reliability issues

More generally, none of the spectral analysis methods mentioned above can provide entirely reliable results all by itself, since every statistical test is based on certain probabilistic assumptions about the nature of the physical process that generates the time series of interest. Such mathematical assumptions are rarely, if ever, met in practice.

To establish higher and higher confidence in a spectral result, such as the existence of an oscillatory mode, a number of steps can be taken. First, the mode’s manifestation is verified for a given data set by the best battery of tests available for a particular spectral method. Second, additional methods are brought to bear, along with their significance tests, on the given time series. Vautard et al. (1992) and Yiou et al. (1996), for instance, have illustrated this approach by applying it to a number of synthetic time series, as well as to climatic time series.

The application of the different univariate methods described here and of their respective batteries of significance tests to a given time series is facilitated by the SSA-MTM Toolkit, which was originally developed by Dettinger et al. (1995). This Toolkit has evolved as freeware over the last decade-and-a-half to become more effective, reliable, and versatile; its latest version is available at <http://www.atmos.ucla.edu/tcd/ssa>.

The next step in gaining confidence with respect to a tentatively detected oscillatory mode is to obtain additional time series produced by the phenomenon under study. The final and most difficult step on the road of confidence building is that of providing a convincing physical explanation for an oscillation, once we have full statistical confirmation of its existence. This step consists of building and validating a hierarchy of models for the oscillation of interest, cf. Ghil and Robertson (2000). The modeling step is distinct from and thus fairly independent of the statistical analysis steps

discussed up to this point. It can be carried out before, after, or in parallel with the other steps.

## 2.3 Geoscientific records with heavy-tailed distributions

### 2.3.1 Introduction

After obvious periodicities and trends have been removed from a time series  $X(t)$ , we are left with the component that corresponds to the continuous part of  $S(f)$ . We shall refer to this component as the “stochastic” one, although we have seen in Sect. 2.2 that deterministically chaotic processes can also produce continuous spectra.

We can broadly describe the properties of such a stochastic time series in two complementary ways: (i) the frequency-size distribution of values, i.e., how many values lie in a given size range; and (ii) the correlations among those values, i.e. how successive values cluster together. One also speaks of one-point and two-point (or many-point) properties; the latter reflect the “memory” of the process generating the time series and are discussed in Sect. 2.4. In this subsection, we briefly review frequency-size distributions. After discussing frequency-size distributions in general, we enter into some of the practicalities to consider when proposing heavy-tailed distributions as a model fit to data. We also refer the reader to Sect. 3 of this paper for an in-depth description of EVT. An example of such good practices is given by Rossi et al. (2010), who examined heavy-tailed frequency-size distributions for natural hazards.

### 2.3.2 Frequency-size distributions

Frequency-size distributions are fundamental for a better understanding of models, data and processes when considering extreme events. In particular, such distributions have been widely used for the study of risk. A standard approach has been to assume that the frequency-size distribution of all values in a given time series approaches a Gaussian, i.e., that the Central Limit Theorem applies. In a large number of cases, this approach is appropriate and provides correct statistical distributions. In many other cases, it is appropriate to consider a broader class of statistical distributions, such as log-normal, Lévy, Fréchet, Gumbel or Weibull (Stedinger et al., 1993; Evans et al., 2000; White et al., 2008).

Extreme events are characterized by the very largest (or smallest) values in a time series, namely those values that are larger (or smaller) than a given threshold. These are the “tails” of the probability distributions – the far right or far left of a univariate distribution, say – and can be broadly divided into two classes, thin tails and heavy tails: thin tails are those that fall off exponentially or faster, e.g., the tail of a normal, Gaussian distribution; heavy tails, also called fat or long tails, are those that fall off more slowly. There is still some confusion in the literature as to the use of the terms long, fat or heavy tails. Current usage, however, is tending

towards the idea that heavy-tailed distributions are those that have power-law frequency-size distributions, and where not all moments are finite; these distributions arise from scale-invariant processes.

In contrast to the Central Limit Theorem, where all the values in a time series are considered, EVT studies the convergence of the values in the tail of a distribution towards Fréchet, Gumbel or Weibull probability distributions; it is discussed in detail in Sect. 3. We now consider some practical issues in the study of heavy-tailed distributions.

### 2.3.3 Heavy-tailed distributions: practical issues

As part of the nonlinear revolution in time series analysis (see Sect. 2.1), there has been a growing understanding that many natural processes, as well as socio-economic ones, have heavy-tailed frequency-size distributions (Malamud, 2004; Sornette, 2004). The use of such distributions in the refereed literature – including Pareto and generalized Pareto or Zipf’s law – has increased dramatically over the last two decades, in particular in relationship to extreme events, risk, and natural or man-made hazards; e.g., Mitzenmacher (2003).

Practical issues that face researchers working with heavy-tailed distributions include:

*Finding the best power-law fit to data, when such a model is proposed.* Issues in preparing the data set and fitting frequency-size distributions to it:

1. use of cumulative vs. noncumulative data,
2. whether and how to “bin” data (linear vs. logarithmic bins, size of the bins),
3. proper normalization of the bin width,
4. use of advanced estimators for fitting frequency-size data, such as kernel density estimation or maximum likelihood estimation (MLE),
5. discrete vs. continuous data,
6. number of data points in the sample and, if bins are used, number of data points in each bin.

A common example of incorrectly fitting a model to data is to count the number of data values  $n$  that occur in a bin with width  $r$  to  $r + \Delta r$ , use bins with different sizes (e.g., logarithmically increasing) and not normalize by the size of the bin  $\Delta r$ , but rather just fit  $n$  as a function of  $r$ . If using logarithmic bins and  $n$ -counts in each bin that are not properly normalized, the power-law exponent will be off by 1.0. One should also carefully examine the data being used for any evidence of preliminary binning – e.g., wildfire sizes are preferentially reported in the United States to the nearest acre burned – and appropriately modify the subsequent treatment of the data.



White et al. (2008) compare different methods for fitting data and their implications. These authors conclude that MLE provides the most reliable estimators when fitting power laws to data. It stands to reason, of course, that if the data are sufficient in number and robustly follow a power-law distribution over many orders of magnitude, then binning (with proper normalization), kernel density estimation, and MLE should all provide broadly similar results: it is only when the number of data is small and when they cover but a limited range that the use of more advanced methods becomes important.

Another case in which extra caution is required occurs when – as is often the case for natural hazards – the data are reported in fixed and relatively large increments. An example is the number of hectares of forest burnt in a wildfire.

### Goodness of fit, and its appropriateness for the data

When fitting frequency counts or probability densities as a function of size, and assuming a power-law fit, a common method is to perform a least-squares linear regression fit on the log of the values, and “report” the correlation coefficient as a goodness of fit. However, many more robust techniques exist – e.g., the Kolmogorov-Smirnov test combined with MLE – for estimating how good a fit is, and if the fit is appropriate for the data. Clauset et al. (2009) have a broad discussion on techniques for estimating whether the model is appropriate for the data, in the context of MLE.

### Narrow-range vs. broad-range fits

Although many scientists identify power-law or “fractal” statistics in their data, often it is over a very narrow-range of orders. In a study by Avnir et al. (1998), they examined 96 articles in *Physical Review* journals, over a seven year period, with each study reporting power-law fits to their data. Of the studies, 45 % of them reported power-law fits based on just one order of magnitude in the size of their data, or even less, and only 8 % of the studies were basing their power-law fits on more than two orders of magnitude. More than a decade on, and many researchers are still satisfied to determine that their data support a power-law or heavy-tailed distribution based on a fit that is valid for quite a narrow range of behavior.

### Correlations in the data

When using a frequency-size distribution for risk analyses, or other interpretations, one is often making the assumption that the data themselves are independent and identically distributed (i.i.d.) in time. In other words, that there are no correlations or clustering of values. This assumption is often not true, and is discussed in the next subsection (Sect. 2.4) in the context of long-range memory or dependence.

## 2.4 Memory and long-range dependence (LRD)

### 2.4.1 Introduction

When periodicities, quasi-periodic modes or large-scale trends have been identified and removed – e.g., by using SSA (Sect. 2.2.1), MTM (Sect. 2.2.2) or other methods – the remaining component of the time series is often highly irregular but might still be of considerable interest. The increments of this residual time series are seldom totally uncorrelated – although even if this were the case, it would still be well worth knowing – but often shows some form of “memory”, which appears in the form of an auto-correlation (see below).

The characteristics of the noise processes that generate this irregular, residual part of the time series determine, for example, the significance levels associated with the separation of the original, raw time series into signals, such as harmonic oscillations or broader-peak modes, and noise (e.g., Ghil et al., 2002). More generally, the accuracy and reliability of parameter values estimated from the data are determined by these characteristics: roughly speaking, noise that has both low variance and short-range correlations only leads to small uncertainties, while high-variance or long-range-correlated noise renders parameter estimation more difficult, and leads to larger uncertainties (cf. Sect. 3.4). The noisy part of a time series, moreover, can still be useful in short-term prediction (Brockwell and Davis, 1991).

The linear part of the temporal structure in the noisy part of the time series is given by the auto-correlation function (ACF); this structure is also loosely referred to as persistence or memory. In particular, so-called *long memory* is suspected to occur in many natural processes; these include surface air temperature (e.g., Koscielny-Bunde et al., 1996; Caballero et al., 2002; Fraedrich and Blender, 2003), river run-off (e.g., Montanari et al., 2000; Kallache et al., 2005; Mudelsee, 2007), ozone concentration (Vyushin et al., 2007), among others, all of which are characterized by a slow decay of the ACF. This slow decay is often referred to as the *Hurst phenomenon* (Hurst, 1951); it is the opposite of the more common, rapid decay of the ACF in time, which can be approximated as exponential.

A typical example for the Hurst phenomenon occurs in the time series of Nile River annual minimum water levels, shown in Fig. 1; this figure will be further discussed in Sect. 5.1. The data set represents a very long climatic time series (e.g., Hurst, 1952) and covers 1300 yr, from 622 AD to 1921 AD; the fairly large gaps were filled by using SSA (Kondrashov et al., 2005, and references therein). This series exhibits the typical time-domain characteristics of long-memory processes with positive strength of persistence, namely the relatively long excursions from the mean (solid blue line, based on the gapless interval 622–1479 AD); such excursions are marked by the red shaded areas in the figure.

### 2.4.2 The auto-correlation function (ACF) and the memory of time series

The ACF  $\rho(\tau)$  describes the linear interdependence of two instances,  $X(t)$  and  $X(t+\tau)$ , of a time series separated by the lag  $\tau$ . As previously stated, the most frequently encountered case involves a rapid, exponential decay of the linear dependence of  $X(t+\tau)$  on  $X(t)$ , with  $\rho(\tau) \sim \exp(-\tau/T)$  when  $\tau$  is large, where  $T$  represents the decorrelation time. This behavior characterizes auto-regressive (AR) processes of order 1, also called AR[1] processes (Brockwell and Davis, 1991); we have already alluded to them in Sect. 2.2.1 under the colloquial name of “red noise”.

For a discretely sampled time series, with regular sampling at points  $t_n = n\Delta t$  (see Sect. 2.1), we get a discrete set of lags  $\tau_k = k\Delta t$ . The rapid decay of the ACF implies a finite sum  $\sum_{\tau_k=-\infty}^{\infty} \rho(\tau_k) = c < \infty$  that leads to the term finite-memory or *short-range dependence* (SRD). In contrast, a diverging sum (or infinite memory)

$$\sum_{\tau_k=-\infty}^{\infty} \rho(\tau_k) = \infty. \quad (5)$$

characterizes a *long-range dependent* (LRD) process, also called long-range correlated or long-memory process. The prototype for an ACF with such a diverging sum shows an algebraic decay for large time lags  $\tau$

$$\rho(\tau) \sim \tau^{-\gamma}, \quad \tau \rightarrow \infty, \quad (6)$$

with  $0 < \gamma < 1$  (Beran, 1994; Robinson, 2003).

By the Wiener-Khinchin theorem (e.g., Hannan, 1960), the spectral density  $S(f)$  of Sect. 2.2.2 and the ACF  $\rho(\tau)$  are Fourier transforms of each other. Hence the spectral density  $S(f)$  can also be used to describe the short- or long-term memory of a given time series (Priestley, 1992). In this representation, the LRD property manifests itself as a spectral density that diverges at zero frequency and decays slowly (i.e., like a power law) towards larger frequencies:  $S(f) \sim |f|^{-\beta}$  for  $|f| \rightarrow 0$ , with an exponent  $0 < \beta = 1 - \gamma < 1$  that is the conjugate of the behavior of  $\rho(\tau)$  at infinity. A result of this type is related to Paley-Wiener theory, which establishes a systematic connection between the behavior of a function at infinity and the properties of its Fourier transform (e.g., Yosida, 1968; Rudin, 1987).

The LRD property is also found in certain deterministically chaotic systems (Manneville, 1980; Procaccia and Schuster, 1983; Geisel et al., 1987), which form an important class of models in the geosciences (Lorenz, 1963; Ghil et al., 1985; Ghil and Childress, 1987; Dijkstra, 2005). Here, the strength of the long-range correlation is functionally dependent on the fractal dimension of the time series generated by the system (e.g., Voss, 1985).

The characterization of the LRD property in this subsection emphasizes the asymptotic behavior of the ACF, cf.

Eqs. (5) and (6), and is more commonly used in the mathematical literature. A complementary view, more popular in the physical literature, is based on the concept of a self-affine stochastic process, which generates time series that are similar to the one or ones under study. This is the view taken, for instance, in Sect. 2.4.4 and in the applications of Sect. 5.1.2. Clearly, for a stochastic process to describe well a time series with the LRD property, its ACF must satisfy Eqs. (5) and (6). The two approaches lead simply to different tools for describing and estimating the LRD parameters.

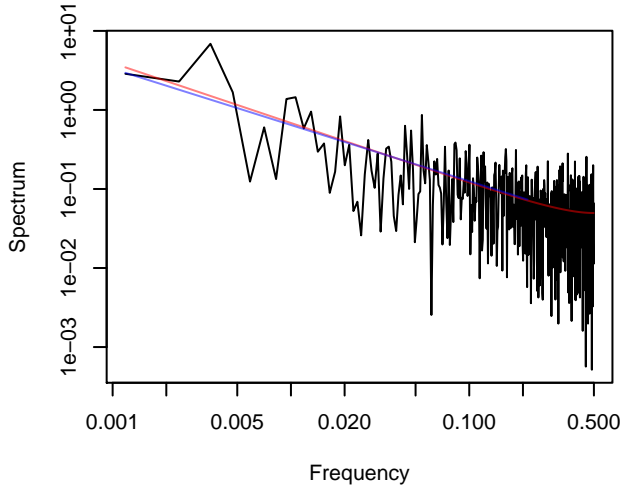
### 2.4.3 Quantifying long-range dependence (LRD)

The parameters  $\gamma$  or  $\beta$  are only two of several parameters used to quantify LRD properties; estimators for these parameters can be formulated in different ways. A straightforward way is to estimate the ACF from the time series (Brockwell and Davis, 1991) and fit a power-law, according to Eq. (6), to the largest time lags. Often this is done by least-square fitting a straight line to the ACF values plotted in log-log coordinates. The exponent found in this way is an estimate for  $\gamma$ . ACF estimates for large  $\tau$ , however, are in general noisy and hence this method is not very accurate nor very reliable.

One of the earliest techniques for estimating LRD parameters is the rescaled-range statistic (R/S). It was originally proposed by Hurst (1951) when studying the Nile River flow minima (Hurst, 1952; Kondrashov et al., 2005); see the time series shown here in Fig. 1. The estimated parameter is now called the Hurst exponent  $H$ ; it is related to the previous two by  $2H = 2 - \gamma = \beta + 1$ . While the symbol  $H$  is sometimes used in the nonlinear analysis of time series for the Hausdorff dimension, we will not need the latter in the present review, and thus no confusion is possible. The R/S technique has been extensively discussed in the LRD literature (e.g., Mandelbrot and Wallis, 1968b; Lo, 1991).

A method which has become popular in the geosciences is detrended fluctuation analysis (DFA; e.g., Kantelhardt et al., 2001). Like R/S, it yields a heuristic estimator for the Hurst exponent. We use here the term “heuristic” in the precise statistical sense of an estimator for which the limiting distribution is not known, and hence confidence intervals are not readily available, thus making statistical inference rather difficult.

The DFA estimator is simple to use, though, and believed to be robust against certain types of nonstationarities (Chen et al., 2002). It is affected, however, by jumps in the time series; such jumps are often present in temperature records (Rust et al., 2008). DFA has been applied to many climatic time series (e.g., Monetti et al., 2001; Eichner et al., 2003; Fraedrich and Blender, 2003; Bunde et al., 2004; Fraedrich and Blender, 2004; Király and János, 2004; Rybski et al., 2006; Fraedrich et al., 2009), and LRD behavior has been inferred to be present in them; the estimates obtained for the Hurst exponent have differed, however, fairly widely from study to study. Due to DFA’s heuristic nature and the lack



**Fig. 2.** Periodogram of the time series of Nile River minima (see Fig. 1), in log-log coordinates, including the straight line of the Geweke and Porter-Hudak (1983) (GPH) estimator (blue) and a fractionally differenced (FD) model (red). In both cases we obtain  $\hat{\beta} = 0.76$  and thus a Hurst exponent of  $\hat{H} = 0.88$ ; asymptotic standard deviations for  $\hat{\beta}$  are 0.09 for the GPH estimate and 0.05 for the FD model.

of confidence intervals, it has been difficult to choose among the distinct estimates, or to discriminate between them and the trivial value of  $H = 0.5$  expected for SRD processes.

Taqqu et al. (1995) have discussed and analyzed DFA and several other heuristic estimators in detail. Metzler (2003), Maraun et al. (2004) and Rust (2007) have also provided reviews of, and critical comments, on DFA.

As mentioned at the end of the previous subsection (Sect. 2.4.2), one can use also the periodogram  $\hat{S}(f_j)$  to evaluate LRD properties; it represents an estimator for the spectral density  $S(f)$ , cf. Eq. (4). The LRD character of a time series will manifest itself by the divergence of  $\hat{S}(f_j)$  at  $f = 0$ , according to  $S(f) \sim |f|^{-\beta}$ . More precisely, in a neighborhood of the origin,

$$\log \hat{S}(f_j) = \log c_f - \beta \log |f_j| + u_j; \quad (7)$$

here  $c_f$  is a positive constant,  $\{f_j = j/N : j = 1, 2, \dots, m\}$  is a suitable subset of the Fourier frequencies, with  $m < N$ , and  $u_j$  is an error term. The log-periodogram regression in Eq. (7) yields an estimator with a normal limit distribution that has a variance of  $\pi^2/(24m)$ . Asymptotic confidence intervals are therewith readily available and statistical inference about LRD parameters is thus greatly facilitated (Geweke and Porter-Hudak, 1983; Robinson, 1995). Figure 2 (blue line) shows the GPH estimator of Geweke and Porter-Hudak (1983), as applied to the low-frequency half of the periodogram for the time series in Fig. 1; the result, with  $m = 212$ , is that  $\hat{\beta} = 0.76 \pm 0.09$  and  $\hat{H} = 0.88 \pm 0.04$ , with the standard deviations indicated in the usual fashion.

For an application of log-periodogram regression to atmospheric-circulation data, see Vyushin and Kushner (2009); Vyushin et al. (2008) provide an R-package for estimates based on the GPH approach and on wavelets. The latter approach to assess long-memory parameters is described, for instance, by Percival and Walden (2000).

The methods discussed up to this point involve the relevant asymptotic behavior of the ACF at large lags or of the spectral density at small frequencies. The behavior of the ACF or of the spectral density at other time or frequency scales, respectively, is not pertinent per se. There are two caveats to this remark. First, given a finite set of data, one has only a limited amount of large-lag correlations and of low-frequency information. It is not clear, therefore, whether the limits of  $\tau \rightarrow \infty$  or of  $f \rightarrow 0$  are well captured by the available data or not.

Second, given the poor sampling of the asymptotic behavior, one often uses ACF behavior at smaller lags or spectral-density behavior at larger frequencies in computing LRD-parameter estimates. Doing so is clearly questionable a priori and only justified in special cases, where the asymptotic behavior extends beyond its normal range of validity. We return to this issue in the next subsection and in Sect. 8.

#### 2.4.4 Stochastic processes with the LRD property

As opposed to investigating the asymptotic behavior of the ACF or of the spectral density, as in the previous subsection (Sect. 2.4.3), one can aim for a description of the full ACF or spectral density by using stochastic processes as models. Thus, for instance, Fraedrich and Blender (2003) and ? have provided evidence of sea surface temperatures (SSTs) following an  $1/f$  spectrum, also called *flicker noise*, in mid-latitudes and for low frequencies, i.e. for time scales longer than a year, while the spectrum at shorter time scales follows the well-known red-noise or  $1/f^2$  spectrum. Fraedrich et al. (2004) then used a two-layer vertical diffusion model – with different diffusivities in the mixed layer and in the abyssal ocean – to simulate this spectrum, subject to spatially dependent but temporally uncorrelated atmospheric heat fluxes.

A popular class of stochastic processes used to model the continuous part of the spectrum (see Sect. 2.2.2) consists of the already mentioned AR[1] processes (see Sect. 2.4.2), along with auto-regressive processes of order  $p$ , denoted by AR[ $p$ ], and their extension to auto-regressive moving-average processes, denoted by ARMA[ $p, q$ ] (e.g., Percival and Walden, 1993). Roughly speaking,  $p$  indicates the number of lags involved in the regression, and  $q$  the number of lags involved in the averaging; both  $p$  and  $q$  here are integers.

It can be shown that the ACF of an AR[ $p$ ] process – or, more generally, of a finite-order ARMA[ $p, q$ ] process – decays exponentially; this leads to a converging sum  $\sum_{\tau_k=-\infty}^{\infty} \rho(\tau_k) = c < \infty$  and thus testifies to SRD behavior; see Brockwell and Davis (1991) and Sect. 2.4.2 above. In the first half of the 20th century, the statistical literature was

dominated by  $AR[p]$  processes and models to describe observed time series that exhibit slowly decaying ACFs (e.g., Smith, 1938; Hurst, 1951) were not available to practitioners.

In probability theory, though, important developments led to a study of so-called infinitely divisible distributions and of Lévy processes (e.g., Feller, 1968; Sato, 1999). In this context, also a class of stochastic processes associated to the concept of LRD were introduced into the physical literature by Kolmogorov (1941) and were widely popularized by Mandelbrot and Wallis (1968a) and by Mandelbrot (1982) in the statistics community as well, under the name of *self-similar* or, more generally, *self-affine* processes.

A process  $Y_t$  is called self-similar if its probability distribution changes according to a simple scaling law when changing the scale at which we resolve it. This can be written as  $Y_t \stackrel{d}{=} c^{-H} Y_{ct}$ , where  $X \stackrel{d}{=} Y$  means that the two random variables  $X$  and  $Y$  are equal in distribution;  $H$  is called the self-similarity parameter or Hurst exponent and was already encountered, under the latter name, in Sect. 2.4.3, while  $c$  is a positive constant. These processes have a time-dependent variance and are therefore not stationary. A stationary model can be obtained from the increment process  $X_t = Y_t - Y_{t-1}$ . In a self-affine processes, there are two or more dimensions involved, with separate scaling parameters,  $d_1, d_2, \dots$

When  $Y_t$  is normally distributed for each  $t$ ,  $Y_t$  is called *fractional Brownian motion*, while  $X_t$  is commonly referred to as *fractional Gaussian noise* (fGn); this was the first model to describe observed time series with algebraically, rather than exponentially decaying ACF at large  $\tau$  (Beran, 1994). For  $0.5 < H < 1$ , the ACF of the increment process  $X_t$  decays algebraically and the process is LRD.

For an fGn process with  $0 < H \leq 0.5$ , though, the ACF is summable, meaning that such a process has SRD behavior; in fact, its ACF even sums to zero. This “pathologic” situation is sometimes given the appealing name of “long-range anti-correlation” but is rarely encountered in practice. It is mostly the result of over-differencing the actual data (Beran, 1994). For a discussion on long-range anti-correlation in the Southern Oscillation Index, see Ausloos and Ivanova (2001) and the critical comment by Metzler (2003).

Self-similarity implies that all scales in the process are governed by the scaling exponent  $H$ . A very large literature exists on whether self-similarity – or, more generally, self-affinity – does indeed extend across many decades, or not, in a plethora of physical and other processes. Here is not the place to review this huge literature; see, for instance, Mandelbrot (1982) and Sornette (2004), among many others.

Fortunately, more flexible models for LRD are available too; these models exhibit algebraic decay of the ACF only for large time scales  $\tau$  and not for the full range of scales, like those that are based on self-similarity. They still satisfy Eq. (5) but do allow for unrelated behavior on smaller time scales. Most popular among these models are the fractional

ARIMA[ $p, d, q$ ] (or FARIMA[ $p, d, q$ ]) models (Brockwell and Davis, 1991; Beran, 1994). This generalization of the classical ARMA[ $p, q$ ] processes has been possible due to the concept of *fractional differencing* introduced by Granger and Joyeux (1980) and Hosking (1981). The spectral density of a fractionally differenced (FD) or fractionally integrated process with parameter  $d$  estimated from the Nile River time series of Fig. 1 is shown in Fig. 2 (red line). For small frequencies, it almost coincides with the GPH straight-line fit (blue).

If an FD process is used to drive a stationary AR[1] process, one obtains a *fractional* AR[1] or FAR[1] process. Such a process has the desired asymptotic behavior for the ACF, cf. Eq. (6), but is more flexible for small  $\tau$ , due to the AR[1] component. The more general framework that includes the moving-average part is then provided by the class of FARIMA[ $p, d, q$ ] processes (Beran, 1994). The parameter  $d$  is related to the Hurst exponent by  $H = 0.5 + d$ , so that  $0 \leq d = H - 0.5 < 0.5$  is LRD, in which  $0.5 \leq H < 1$ . These processes are LRD but do *not* show self-similar behavior, they do not show a power-law ACF across the whole range of lags or a power-law spectral density across the whole range of frequencies.

Model parameters of the fGn or the FARIMA[ $p, d, q$ ] models can be estimated using maximum-likelihood estimation (MLE) (Dahlhaus, 1989) or the Whittle approximation to the MLE (Beran, 1994). These estimates are implemented in the R-package *farima* at <http://wekuw.met.fu-berlin.de/~HenningRust/software/farima/>. Both estimation methods provide asymptotic confidence intervals.

Another interesting aspect of this fully parametric modeling approach is that the discrimination problem between SRD and LRD behavior can be formulated as a model selection problem. In this setting, one can revert to standard model selection procedure based on information criteria such as the Akaike information criterion (AIC) or similar ones (Beran et al., 1998) or to the likelihood-ratio test (Cox and Hinkley, 1994). A test setting that involves both SRD and LRD processes as possible models for the time series under investigation leads to a more rigorous discrimination because the LRD model has to compete against the most suitable SRD-type model (Rust, 2007).

#### 2.4.5 LRD: practical issues

A number of practical issues exist when examining long-range persistence. Many of these are discussed in detail by Malamud and Turcotte (1999) and include:

*Range of  $\beta$*  Some techniques for quantifying the strength of persistence – e.g., power-spectral analysis and DFA – are effective for time series that have any  $\beta$ -value, whereas others are effective only over a given range of  $\beta$ . Thus semivariograms are effective only over the range  $1 \leq \beta \leq 3$ , while R/S analysis is effective only over  $-1 \leq \beta \leq 1$ ; see Malamud and Turcotte (1999) for further details. *Time series*

*length* As the length of a time series is reduced, some techniques are much less robust than others in appropriately estimating the strength of the persistence (Gallant et al., 1994). *One-point probability distributions* As the distribution of the point values that make up a time series becomes increasingly non-Gaussian – e.g., log-normal or heavy-tailed – techniques such as R/S and DFA exhibit strong biases in their estimator of the persistence in the time series (Malamud and Turcotte, 1999). *Trends and periodicities* Any large-scale variability stemming from other sources than LRD will influence the persistence estimators. Rust et al. (2008), for instance, has discussed the deleterious effects of trends and jumps in the series on these estimators. Known periodicities, like the annual or diurnal cycle, should be explicitly removed in order to avoid biases (Markovic and Koch, 2005). More generally, one should separate the lines and peaks from the continuous spectrum before studying the latter (see Sect. 2.3.1). Witt et al. (2010) showed that it is important to examine whether temporal correlations are present in time series of natural hazards or not. These time series are, in general, unevenly spaced and may exhibit heavy-tailed frequency-size distributions. The combination of these two traits renders the correlation analysis more delicate.

#### 2.4.6 Simulation of fractional-noise and LRD processes

Several methods exist to generate time series from an LRD process. The most straightforward approach is probably to apply an AR linear filter to a white-noise sequence in the same way as usually done for an ARMA[ $p, q$ ] processes (Brockwell and Davis, 1991). Since a FARIMA[ $p, d, q$ ] process, however, requires an infinite-order of auto-regression, the filter has to be truncated at some point.

A different approach relies on spectral theory: the points in a periodogram are distributed according to  $S(f)Z$ , where  $Z$  denotes a  $\chi^2$ -distributed random variable with two degrees of freedom (Priestley, 1992). If the spectral density  $S(f)$  is known, this property can be exploited to straightforwardly generate a time series by inverse Fourier filtering (Timmer and König, 1995). Frequently used methods for generating a time series from an LRD process are reviewed by Bardet et al. (2003).

### 3 Extreme value theory (EVT)

#### 3.1 A few basic concepts

EVT provides a solid probabilistic foundation (e.g. Beirlant et al., 2004; De Haan and Ferreira, 2006) for studying the distribution of extreme events in hydrology (e.g., Katz et al., 2002), climate sciences, finance and insurance (e.g. Embrechts et al., 1997), and many other fields of applications. Examples include the study of annual maxima of temperatures, wind, or precipitation.

Probabilistic EVT theory is based on asymptotic arguments for sequences of i.i.d. random variables; it provides information about the distribution of the maximum value of such an i.i.d. sample as the sample size increases. A key result is often called the *first EVT theorem* or the Fisher-Tippett-Gnedenko theorem. The theorem states that suitably rescaled sample maxima follow asymptotically – subject to fairly general conditions – one of the three “classical” distributions, named after Gumbel, Fréchet or Weibull (e.g., Coles, 2001); these are also known as extreme value distributions of Type I, II and III.

For the sake of completeness, we list these three cumulative distribution functions below:

*Type I (Gumbel):*

$$G(z) = \exp \left\{ -\exp \left[ -\left( \frac{z-b}{a} \right) \right] \right\}, \quad -\infty < z < \infty; \quad (8)$$

*Type II (Fréchet):*

$$G(z) = \begin{cases} 0, & z \leq b, \\ \exp \left\{ -\left( \frac{z-b}{a} \right)^{-1/\eta} \right\}, & z > b; \end{cases} \quad (9)$$

*Type III (Weibull):*

$$G(z) = \begin{cases} \exp \left\{ -\left[ -\left( \frac{z-b}{a} \right)^{1/\eta} \right] \right\}, & z < b, \\ 1, & z \geq b. \end{cases} \quad (10)$$

The parameters  $b$ ,  $a > 0$  and  $\eta > 0$  are called, respectively, the location, scale and shape parameter. The Gumbel distribution is unbounded, while Fréchet is bounded below by 0 and Weibull has an upper bound; the latter is unknown a priori and needs to be determined from the data. These distributions are plotted, for instance, in Beirlant et al. (2004, p. 52).

Such a result is comparable to the Central Limit Theorem, which states that the sample's mean converges in distribution to a Gaussian variable, whenever the sample variance is finite. As is the case for many other results in probability theory, a *stability property* is the key element to understand the distributional convergence of maxima.

A random variable is said to be stable (or closed), or to have a stable distribution, if a linear combination of two independent copies of the variable has the same distribution, up to affine transformations, i.e. up to *location* and *scale* parameters. Suppose that we have drawn a sample  $(X_1, \dots, X_n)$  of size  $n$  from the variable  $X$ , and let  $M_n$  denote the maximum of this sample. We would like to know which type of distributions are stable for the maxima  $M_n$ , up to affinity, i.e., we wish to find the class of distributions that satisfies the equality  $\max(X_1, \dots, X_n) \stackrel{d}{=} a_n X_0 + b_n$  for the sample-size-dependent scale and location parameters  $a_n (> 0)$  and  $b_n$ ; here  $X_0$  is independent of  $X_i$  and has the same distribution as  $X_i$ , while the notation  $X \stackrel{d}{=} Y$  was defined in Sect. 2.4.4.

The solution of such a distributional equation is called the group of *max-stable distributions*. For example, a unit-Fréchet distribution is defined by  $P(X < x) = \exp(-1/x)$ . Its sample maximum satisfies

$$P(M_n < xn) = P(X_1 < xn, \dots, X_n < xn) \quad (11)$$

$$= P^n(X < xn) = \exp(-1/x);$$

consequently, it belongs to the max-stable category. Intuitively this implies that – if the rescaled sample maximum  $M_n$  converges in distribution to a limiting law – then this limiting distribution has to be max-stable.

Let us assume that, for all  $x$ , the distribution of the rescaled maximum  $P((M_n - b_n)/a_n \leq x)$  converges to a nondegenerate distribution for some suitable normalization sequences  $a_n > 0$  and  $b_n$ :

$$P((M_n - b_n)/a_n \leq x) = F^n(a_n x + b_n) \rightarrow G(x) \text{ as } n \rightarrow \infty, \quad (12)$$

where  $G(x)$  is a nondegenerate distribution. Then this limiting distribution has to be max-stable and can be written as a *Generalized Extreme Value* (GEV) distribution

$$G(x; \mu, \sigma, \xi) = \exp\left(-\left\{1 + \xi \frac{x - \mu}{\sigma}\right\}_+^{-1/\xi}\right), \quad (13)$$

where  $(\cdot)_+$  stands for the positive part of the quantity  $(\cdot)$ .

Here  $\mu$  replaces  $b$  as the mean or location parameter in Eqs. (8)–(10),  $\sigma$  replaces  $a$  as the scale parameter, while  $\xi$  replaces  $\eta$  as the shape parameter; note that  $\xi$  no longer need be positive. Thus, in Eq. (13),  $\mu \in \mathbb{R}$ ,  $\sigma > 0$  and  $\xi \neq 0$  are called the GEV location, scale and shape parameters, respectively, and they have to obey the constraint  $1 + \xi(x - \mu)/\sigma > 0$ . For  $\xi \rightarrow 0$  in Eq. (13), the function  $G(x; \mu, \sigma, 0)$  tends to  $\exp(-\exp\{-(x - \mu)/\sigma\})$ , i.e. to the Gumbel, or Type-I, distribution of Eq. (8). The sign of  $\xi$  is important because  $G(x; \mu, \sigma, \xi)$  corresponds to a heavy-tailed distribution (see also Sect. 2.3) when  $\xi > 0$  (Fréchet or Type II) and to a distribution with bounded tails when  $\xi < 0$  (Weibull or Type III).

Note that the result in Eq. (12) does not state that every rescaled maximum distribution converges to a GEV. It states only that *if* the rescaled maximum converges, then the limit  $H(x)$  is a GEV distribution. For example, discrete random variables do not satisfy this condition and other limiting results have to be derived for such variables. We shall return to this point in the application of Sect. 4.2, which involves deterministic, rather than random processes.

From a statistical point of view, it follows that the cumulative probability distribution function of the sample maximum is very likely to be correctly fitted by a GEV distribution  $G(x; \mu, \sigma, \xi)$  of the form given in Eq. (13): this form encapsulates all univariate max-stable distributions, in particular the three above-mentioned types (e.g., Coles, 2001).

In trying to assess and predict extreme events, one often works with so-called “block maxima”, i.e. with the maximum value of the data within a certain time interval. Although the choice of the block size – such as a year, a sea-

son or a month – can be justified in many cases by geophysical considerations, modeling block maxima is statistically wasteful. For example, working with annual maxima of precipitation implies that, for each year, all rainfall observations but one, the maximum, are disregarded in the inference procedure.

To remove this drawback, another approach consists in modeling exceedances above a pre-chosen threshold  $u$ . In the so-called “peaks-over-threshold” (POT) approach, the distributions of these exceedances are also characterized by asymptotic results: their intensities are approximated by the generalized Pareto distribution (GPD) and their frequencies by a Poisson point process.

The *survival function*  $\bar{H}_{\xi, \sigma}(y)$  of the GPD  $H_{\xi, \sigma}(y)$  is given by

$$\bar{H}_{\xi, \sigma}(y) := 1 - H_{\xi, \sigma}(y) = \begin{cases} \left(1 + \frac{\xi}{\sigma}y\right)^{-1/\xi} & \text{for } \xi \neq 0 \text{ and } 1 + \xi y/\sigma > 0, \\ e^{-y/\sigma} & \text{for } \xi = 0, \end{cases} \quad (14)$$

where  $\xi$  and  $\sigma = \sigma(u) > 0$  are the shape and scale parameters of this function, and  $y = x - u > 0$  corresponds to the exceedance above the fixed threshold  $u$ . The *second EVT theorem* or Pickands-Balkema-De Haan theorem (e.g., De Haan and Ferreira, 2006) then states that, for a large class of distribution functions  $F(x) = P(X \leq x)$ , we have

$$\lim_{u \rightarrow \tau_F} \sup_{0 < y < \tau_F - u} |F_u(y) - G_{\xi, \sigma(u)}(y)| = 0 \quad (15)$$

for some positive scaling function  $\sigma(u)$  that depends on the threshold  $u$ . While the first EVT theorem was developed in the 1920s, the second one only goes back to the 1970s. In Eq. (15),  $\tau_F$  is the upper bound of the distribution function and  $F_u = P(X - u < y | X - u > 0)$  for  $y > 0$ .

The result of Eq. (15) allows us to describe the POT approach. Given a threshold  $u_n$ , select the observations  $(X_{i_1}, \dots, X_{i_{N_{u_n}}})$  that exceed this threshold; the exceedances  $(Y_1, \dots, Y_{N_{u_n}})$  are given by  $Y_j = X_{i_j} - u_n > 0$ , and  $N_{u_n} = \text{card}\{X_i, i = 1, \dots, n : X_i > u_n\}$  is their number. According to the Pickands-Balkema-De Haan theorem, the distribution function  $F_{u_n}$  of these exceedances can be approximated by a GPD whose parameters  $\xi$  and  $\sigma_n = \sigma(u_n)$  have to be estimated. It follows that an estimator of the tail  $\bar{F}(u_n)$  of  $F_{u_n}$  is given by

$$\widehat{\bar{F}}(x) = \frac{N_{u_n}}{n} \bar{H}_{\widehat{\xi}, \widehat{\sigma}}(x - u_n), \quad (16)$$

where  $(\widehat{\xi}, \widehat{\sigma})$  are estimators of  $(\xi, \sigma)$  and  $N_{u_n}/n$  is the estimator of  $\bar{F}(u_n)$ . Finally, a POT estimator  $\widehat{x}_p(u_n)$  of the  $p$ -quantile  $x_p$  is obtained by inverting Eq. (16) to yield

$$\widehat{x}_p(u_n) = u_n + \frac{\widehat{\sigma}}{\widehat{\xi}} \left\{ \left( \frac{np}{N_{u_n}} \right)^{-\widehat{\xi}} - 1 \right\}. \quad (17)$$

In practice, the choice of the threshold  $u_n$  is difficult and the estimation of the parameters  $(\xi, \sigma)$  is clearly a question of

trade-off between bias and variance. As we lower  $u_n$  and thus increase the sample size  $N_{u_n}$ , the bias will grow since the tail satisfies less well the convergence criterion in Eq. (15), while if we increase the threshold, fewer observations are used and the variance increases. Dupuis (1999) selected a threshold by using a bias-robust scheme that gives larger weights to larger observations. In hydrology, a pragmatic and visual approach is to plot the estimates of  $\sigma$  and  $\xi$  as a function of different threshold values. The choice of  $u$  is made by identifying an interval in which these estimates are rather constant with respect to  $u$ . Then  $u$  is chosen to be the smallest value in this interval.

Recently, alternative methods based on mixture models have also been studied in order to avoid the problem of choosing a threshold. Frigessi et al. (2002) proposed a mixture model with two components and a dynamic weight function that is equivalent to using a smoothed threshold. One component of the mixture was a light-tailed parametric density – which they took to be a Weibull density – whereas the other component was a GPD with a positive shape parameter  $\xi > 0$ . Carreau and Bengio (2006) proposed a more flexible mixture model in the context of conditional density estimation. In contrast to the Frigessi et al. (2002) model, the number of components was not fixed a priori and each component was a hybrid Pareto distribution, defined as a Gaussian distribution whose upper-right tail was replaced by a heavy-tailed GPD. The threshold selection problem was replaced there-with by an estimation problem of additional parameters that was data-driven and “unsupervised”, i.e. fully automatic.

To estimate the three parameters of a GEV distribution or the two GPD parameters, several methods have been developed, studied and compared during the last two decades. These comparisons involve fairly technical points and are summarized in Appendix A.

### 3.2 Multivariate EVT

The probabilistic foundations for the statistical study of multivariate extremes are well developed. A classical work in this field is Resnick (1987), and the recent books by Beirlant et al. (2004), De Haan and Ferreira (2006), and Resnick (2007) pay considerable attention to the multivariate case. The aim of multivariate extreme value analysis is to characterize the joint upper tail of a multivariate distribution. As in the univariate case, two basic approaches exist, either analyzing the extremes extracted from prescribed-size blocks – e.g., annual componentwise maxima – or alternatively studying only the observations that exceed some threshold. In either case, one can use different representations of the dependence among block maxima or threshold exceedances, respectively.

For the univariate case, we saw in the previous subsection (Sect. 3.1) that parametric GEV and GPD models characterize the asymptotic behavior of extremes – such as sample maxima or exceedances above a high threshold – as the sample size increases. In the multivariate setting, the limiting

behavior of component-wise maxima has also been studied in terms of max-stable processes, e.g. De Haan and Resnick (1977); see Eq. (19) below for a definition of such processes. An important distinction between the univariate and the multivariate case, though, is that no parametric model can entirely represent max-stable vector processes. Multivariate inference techniques represent a very active field of research; see, for instance, Heffernan and Tawn (2004). For the bivariate case, a number of flexible parametric models have been proposed and studied; they include the Gaussian (Smith, 1990; Hüsler and Reiss, 1989), bilogistic (Joe et al., 1992), and polynomial (Nadarajah, 1999) models.

There has been a growing interest in the analysis of spatial extremes in recent years. For spatial data, when there are as many variables as locations, this number is too large and additional assumptions have to be made in order to work with manageable models. For example, De Haan and Pereira (2006) proposed two specific stationary models for extreme values: both of them depend on a single parameter that varies as a function of the distance between two points. Davis and Mikosch (2008) proposed space-time processes for heavy-tailed distributions by linearly filtering i.i.d. sequences of random fields at each location. Schlather (2002) and Schlather and Tawn (2003) simulated stationary max-stable random fields and studied the extremal coefficients for such fields. Several authors have also investigated Bayesian or latent processes for modeling extremes in space. In this case, the spatial structure was modeled by assuming that the extreme-value parameters were realizations of a smoothly varying process, typically a Gaussian process with spatial dependence (Coles and Casson, 1999).

#### 3.2.1 Max-stable random vectors

To understand dependencies among extremes, it is important to recall the definition of a max-stable vector process (Resnick, 1987; Smith, 2004). To define such a process, we first transform a given marginal distribution to unit Fréchet, cf. Sect. 3.1,

$$P(Z(x_i) \leq u_i) = \exp(-1/u_i), \text{ for any } u_i > 0.$$

The distribution  $Z(\mathbf{x})$  is said to be max-stable if all the finite-dimensional distributions are max-stable, i.e.,

$$\begin{aligned} P^t(Z(x_1) \leq tu_1, \dots, Z(x_r) \leq tu_r) \\ = P(Z(x_1) \leq u_1, \dots, Z(x_r) \leq u_r), \end{aligned} \quad (18)$$

for any  $t \geq 0$ ,  $r \geq 1$ ,  $x_i$ ,  $u_i > 0$  with  $i = 1, \dots, r$ . To motivate the definition in Eq. (18), one could see it as a multivariate version of Eq. (11). Such processes can be rewritten (Schlather, 2002; Schlather and Tawn, 2003; Davis and Resnick, 1993), for  $r = 2$ , say, as

$$\begin{aligned} P\left(Z(\mathbf{x}) \leq u(\mathbf{x}), \text{ for all } \mathbf{x} \in \mathbb{R}^2\right) \\ = \exp\left[-\int \max_{\mathbf{x} \in \mathbb{R}^2} \left\{ \frac{g(s, \mathbf{x})}{u(\mathbf{x})} \right\} \delta(ds)\right], \end{aligned} \quad (19)$$

where the function  $g(\cdot, \cdot)$  is a weighting function that is nonnegative, measurable in the first argument, upper semi-continuous in the second, and has total weight equal to unity,  $\int g(s, \mathbf{x}) \delta(ds) = 1$ .

To illustrate how such max-stable processes can be generated, we focus on two types of max-stable fields. A first class was developed by Smith (1990), where points in  $(0, \infty) \times \mathbb{R}^2$  are simulated according to a point process:

$$Z(x) = \sup_{(y,s) \in \Pi} [s f(x-y)], \text{ with } x \in \mathbb{R}, \quad (20)$$

where  $\Pi$  is a Poisson point process on  $\mathbb{R}^2 \times (0, \infty)$  with intensity  $s^{-2} dy ds$  and  $f$  is a nonnegative function such that  $\int f(x) dx = 1$ . For example,  $f$  can be taken to be a two-dimensional Gaussian density function with a covariance matrix equal to the two-by-two identity matrix.

Schlather (2002) proposed a second type, stemming from a Gaussian random field that is scaled by the realization of a point process on  $(0, \infty)$ :  $Z(x) = \max_{s \in \Pi} s Y_s(x)$ ; here the  $Y_s$  are independent, identically distributed, stationary Gaussian processes and  $\Pi$  is a Poisson point process on  $(0, \infty)$  with intensity  $\sqrt{2\pi} r^{-2} dr$ . An R code developed by Schlather (2006, <http://cran.r-project.org>, contributed package `RandomFields`) can be used to visualize these two types of max-stable processes. A few properties of these two examples are derived in Appendix B.

Additional contributions that deserve being mentioned at this point are the Brown-Resnick max-stable process of Kabluchko et al. (2009), the geometric max-stable model in Padoan et al. (2010), and the R-package `SpatialExtremes` of Ribatet (2008).

### 3.2.2 Multivariate varying random vectors

The concept of regular variation is fundamental to understand the concept of angular measures in EVT. Regular variation can be defined as follows. Let  $\mathbf{Z} = (Z_1, \dots, Z_p)^t$  be a nonnegative multivariate vector of finite dimension,  $\mathbf{A}$  be any set and  $t$  a positive scalar. Roughly speaking, the vector  $\mathbf{Z}$  is said to be multivariate varying (Resnick, 2007) if, given that the norm of  $\mathbf{Z}$  is greater than  $t$ , the probability that  $\mathbf{Z}/t$  in  $\mathbf{A}$  converges to  $\nu(\mathbf{A})$ , as  $t \rightarrow \infty$ , equals unity, where  $\nu(\cdot)$  is a measure. An essential point is that the measure  $\nu(\cdot)$  has to obey the scaling property  $\nu(s\mathbf{A}) = s^{-\alpha} \nu(\mathbf{A})$ , where the scalar  $\alpha > 0$  is called the tail index.

This definition implies three consequences: (i) the upper tail of multivariate varying vectors is always heavy tailed, with index  $\alpha$ . (ii) The scaling property implies that the norm  $\|\mathbf{Z}\|$  of  $\mathbf{Z}$  is asymptotically independent of the unit vector  $\mathbf{Z}/\|\mathbf{Z}\|$ ; in other words, the dependence among extreme events lives on the unit sphere and, independently, the intensity of a multivariate extreme event  $\mathbf{Z}$  is fully captured by its norm  $\|\mathbf{Z}\|$ . (iii) The probability of rare events, i.e., to be in a set  $\mathbf{A}$  far away from the core of the observations, can be

deduced by rescaling the set  $\mathbf{A}$  into a set closer to the observations; this property is essential in order to be able to go beyond the range of the observations.

For example, property (iii) allows one to compute hydrological return levels for return times that exceed the observational record length.

The multivariate varying assumption above may seem very restrictive. Most models used in times series analysis, however, satisfy this hypothesis. These models include infinite-variance stable processes, ARMA processes with i.i.d. regularly varying noise, GARCH processes with i.i.d. noise having infinite support (including normally and Student- $t$ -distributed noise), and stochastic volatility models with i.i.d. regularly varying noise (Davis and Mikosch, 2009).

### 3.2.3 Models and inference for multivariate EVT

Developing nonparametric and semi-parametric models for multivariate extremes is an active area. Einmahl and Segers (2009) have proposed a nonparametric, empirical MLE approach to fit an angular density to a bivariate data set. Using a semi-parametric approach, Boldi and Davison (2007) formulated a model for the angular density of a multivariate extreme value distribution of any dimension by applying mixtures of Dirichlet distributions that meet the required moment conditions.

The work of Ledford and Tawn (1997) spurred interest in describing and modeling dependence of extremes within the class of asymptotic independence. A bivariate couple  $(X_1, X_2)$  is termed asymptotically independent if the limiting probability of  $X_1 > x$ , given that  $X_2 > x$ , is null as  $x \rightarrow \infty$ . Heffernan and Tawn (2004) provided models for extremes that included the case of asymptotic independence, but the models were developed via bivariate conditional relationships, and higher-dimensional relationships were not made explicit. A recent paper by Ramos and Ledford (2009) proposes a parametric model that captures both asymptotic dependence and independence in the bivariate case. Several metrics have been suggested to quantify the levels of dependence found in multivariate extremes in general and in traditional max-stable random vectors in particular: the extremal coefficient of Schlather and Tawn (2009), the dependence measures of Coles et al. (1999) and of Davis and Resnick (1993), as well as the madogram, which is a first-order variogram, cf. Cooley et al. (2006).

### 3.3 Nonstationarity, covariates and parametric models

When studying climatological and hydrological data, for instance, it is not always possible to assume that the distribution of the maxima remains unchanged in time: this is a crucial problem in evaluating the effects of global warming on various extremes (Solomon et al., 2007). Trends are often present in the extreme value distribution of different climate-related time series (e.g., Yiou and Nogaj, 2004; Kharin et al.,



2007). The MLE approach, though, can easily integrate temporal covariates within the GEV parameters (e.g., Coles, 2001; Katz et al., 2002; El Adlouni et al., 2007) and, conceptually, the MLE procedure remains the same if the GEV parameters vary in time; see El Adlouni and Ouarda (2008) for a comparative study of different methods in nonstationary GEV models.

### 3.3.1 Parametric and semi-parametric approaches

The introduction of covariates is necessary in many cases when studying the behavior of environmental, hydrological or climatic extremes. The detection of temporal trends in extremes was already mentioned above, but other environmental or climate factors, such as large-scale circulation indices or weather regimes, may also affect the distribution of extremes (e.g., Yiou and Nogaj, 2004; Robertson and Ghil, 1999).

Existing results for asymptotic behavior of extremes, given prescribed forms of nonstationarity (Leadbetter et al., 1983), are often too specific to be used in modeling data whose temporal trend is not known a priori. A pragmatic approach, therefore, is to simply model the parameters of the GEV distribution for block maxima or of the GPD for POT models (see Sect. 3.1) as simple linear functions of time (e.g., Katz et al., 2002; Smith, 1989; Garca et al., 2007). Large-scale indices may also be used as covariates, in order to link extremes with these indices (Caires et al., 2006; Friederichs, 2010).

Since the MLE method is still straightforward to apply in the presence of covariates (Coles, 2001), it is often favored by practitioners (Smith, 1989; Zhang et al., 2004). For numerical stability reasons, link functions are widely used. For example, to ensure a positive scale parameter, one may model the logarithm of a variable as a linear function of covariates, instead of the variable itself.

Log-linear modeling of the GEV or GPD parameters can be viewed as a particular case of the Vector Generalized Linear Models (VGLMs) proposed by Yee and Hastie (2003). A Generalized Linear Model (GLM: Nelder and Wedderburn, 1972) is a linear regression model that supports dependent variables whose distribution belongs to the so-called “exponential family”: Poisson, Gamma, binomial, and normal. GLMs allow one to relate the linear predictor and the “natural parameter” of the distribution under consideration, where the natural parameter is a well-chosen function of the mean of the dependent variable. VGLMs, in turn, are generalizations of GLMs that extend modeling capabilities to distributions outside the exponential family and apply to the distributions of GEV and GPD parameters. The VGAM package of Yee (2006) provides a practical implementation of VGLMs in R software.

### 3.3.2 Smoothing extremes

When considering a nonstationary GEV or GPD distribution, one has to study joint variations of dependent parameters. Structural trend models are difficult to formulate in these circumstances, owing to the complex way in which different factors combine to influence the extremes. Moreover, it is not advisable to fit linear trend models without evidence of their suitability. The need for flexible exploratory tools led to the development of several approaches for smoothing sample extremes.

Davison and Ramesh (2000) fitted nonlinear trends by considering local likelihood in moving windows, with applications to hydrology. In Hall and Tajvidi (2000), the likelihood of each observation was weighted in such windows by using a kernel function. This approach allows one to give more importance to observations close to the center of the window, but results in destroying the asymptotic properties of MLE estimators.

An alternative to the local-likelihood approach is to use spline smoothers. Smooth modeling of position and scale parameters of the Gumbel distribution can be achieved by means of cubic splines; this leads to additive models adapted to extremes. Generalized Additive Models (GAMs) (Hastie and Tibshirani, 1990) are nonlinear extensions of GLMs (see Sect. 3.3.1), where the natural parameter of a distribution belonging to the exponential family is modeled as the sum of smooth, rather than linear functions of the covariates. Vector Generalized Additive Models (VGAMs: Yee and Wild, 1996) are generalizations of GAMs that, like GLMs and VGLMs, extend modeling capabilities to distributions outside the exponential family, and apply to the distributions of interest to us here, namely GEVs and GPDs. The GAM “backfitting” algorithm proposed by Hastie and Tibshirani (1990) was adapted to POT models by Chavez-Demoulin and Davison (2005), while Yee and Stephenson (2007) used “vector spline smoothers” for both GEV and GPD models; see Yee (2006) for practical implementation.

Common features of these semi-parametric techniques are that (i) few predictors should be used, say two or three at most; and (ii) they only provide approximate pointwise confidence intervals. If more precise inferences are required, an exploratory phase should be conducted using VGAMs, whose data-driven approach may suggest the proper linear model. Final results may then be obtained by applying VGLM techniques. Mestre and Halegatte (2009) illustrate this two-stage approach in connecting the distribution of extreme cyclones to large-scale climatic indices.

Finally, if time is the covariate of interest, specific extreme state-space models may be used (Davis and Mikosch, 2008). These approaches are rather new and new developments appear regularly.

### 3.4 EVT and memory

In many practical cases, one is confronted with a set of observations that do not seem to satisfy the standard assumption of i.i.d. random variables. The previous subsection (Sect. 3.3) was dedicated to the case of random variables that are not identically distributed. In the present subsection, we mention some issues that arise for dependent variables, in particular for observations from LRD processes (Sect. 2.4). We review first the application of GEV distributions to the block-maxima approach and then the one of GPD to the peaks-over-threshold (POT) approach (see Sect. 3.1).

The return periods that can be derived within these two approaches, GEV and GPD (Sects. 3.4.1 and 3.4.2, respectively), are the expected (or mean) time intervals between recurrent events. Recently, several authors focused on LRD processes in studying the distribution of return times of threshold exceedances. While mean return times do not change, their distribution differs for dependent and independent series (Bunde et al., 2003; Altmann and Kantz, 2005; Blender et al., 2008).

#### 3.4.1 Dependence and the block-maxima approach

In the block-maxima approach, one can relax the assumption of independence in a straightforward way. The convergence of the block-maxima distribution function towards the GEV holds, in fact, for a wide class of stationary processes, which satisfy a fairly general condition of near-independence (Leadbetter et al., 1983). Consequently, EVT supports the GEV distribution as a potentially suitable candidate for the model block-maxima distribution in this broader class of processes. This result also holds for a class of LRD processes, like the FARIMA[ $p, d, q$ ] processes (Sect. 2.4).

In practical applications with finite block sizes, however, two issues arise (Rust, 2009): (a) the convergence of the block-maxima distribution function towards the GEV distribution is slower; and (b) if the block maxima themselves cannot be considered as independent, the uncertainty of the estimated parameters increases. The first issue only stresses the need for the usual verification of the GEV distribution as a suitable model. This verification can be accomplished by checking quantiles, using probability plots or carrying out other statistical tests (e.g., Coles, 2001).

The second issue is more delicate, since standard procedures, such as the MLE approach, yield uncertainty estimates that are too small in the case of dependence and thus the confidence in the estimated parameters and all derived quantities, such as return levels, is too high. This problem arises basically only for LRD series (Rust, 2009) and, unfortunately, there is no straightforward way out. A possible solution is to obtain uncertainty estimates using Monte-Carlo or bootstrap approaches. Doing so may be somewhat tricky (Rust et al., 2011), but the uncertainty estimates thus obtained are more representative than using the false assumption of independence.

### 3.4.2 Clustering of extremes and the POT approach

For the POT approach (see Sect. 3.1), dependence implies a clustering of the exceedances that are to be modeled by using the GPD. An extreme that exceeds the threshold is usually accompanied by neighboring threshold exceedances, which then lead to a cluster of consecutive exceedances. A common approach to deal with these clustering effects is to consider only the largest event in any given cluster and disregard the others in further analysis (Coles, 2001; Ferro, 2003). Doing so, however, reduces the sample size for the subsequent GPD parameter estimation and thus leads to an increase in the uncertainty of the estimated parameter, compared to the case of independent observations. Unlike in the GEV case, though, the problem is treated naturally within the POT approach by declustering methods, available for example in the R package POT of Ribatet (2009). An alternative way to handle dependence in this approach is given by Markov chain models (Smith et al., 1997).

### 3.5 Statistical downscaling and weather regimes

#### 3.5.1 Bridging time and space scales

The spatial resolution of general circulation models (GCMs; more recently also referred to as “global climate models,” with the same acronym), corresponds to a grid size of about 100–250 km, while that of numerical weather prediction models is of about 25–50 km. Impact studies, on the other hand, often concentrate on an isolated location, such as a weather station or a city, or on a limited region, such as a river basin. The latter are of fundamental importance for hydrologists, climatologists, decision makers and risk managers. Downscaling methods aim at linking these different scales and they have been developed and tested for the last couple of decades. One of their main goals is to generate realistic local-scale climate or meteorological time series – e.g., temperature, winds or precipitation – based on large-scale atmospheric conditions.

Downscaling has essentially been applied within three different frameworks. In the *climatological and climate change* context, statistical downscaling methods (SDMs) are used to estimate the statistics of local-scale variables, given large-scale atmospheric GCM simulations (Zwiers and Kharin, 1998; Kharin and Zwiers, 2000; Busuioc and von Storch, 2003; Widmann et al., 2003; Haylock and Goodess, 2004; Sanchez-Gomez and Terray, 2005; Kharin et al., 2007; Min et al., 2009; Schmidli et al., 2007). The aim is to assess future local variations from large-scale circulations changes. Hence, an underlying assumption is that relationships between large-scale fields and local-scale variables remain valid in the future. This point is crucial and has to be tested on past changes before generating future time series (Vrac et al., 2007d).

In *weather forecasting*, downscaling has been understood as the transformation of large-scale model forecasts – or reanalyses that combine observational data sets with past model forecasts (Kalnay et al., 1996) – to the fine scale where observations are available. For instance, the large-scale state of the atmosphere may be described by the prevailing weather pattern for a specific day or week, over some region, either by a forecast or a (re)analysis (Vislocky and Young, 1989; Zorita and von Storch, 1998; Robertson and Ghil, 1999; Bremnes, 2004; Hamill and Whitaker, 2006; Friederichs and Hense, 2007). Downscaling provides various statistical outputs, e.g., the probability of occurrence or intensity of precipitation at one station location during that day or week.

Downscaling of weather forecasts can also help correct for systematic model errors. In this context, downscaling is very close to model post-processing or re-calibration methods – such as the perfect prog method (Klein, 1971), model output statistics (Glahn and Lowry, 1972) or the automated forecasting method (Ghil et al., 1979) – in the sense that these approaches apply a similar class of statistical (Klein, 1971; Glahn and Lowry, 1972) or physics-based (Ghil et al., 1979) models (see below) to describe relationships between large-scale situations and local-scale variables (Marzban et al., 2005; Friederichs and Hense, 2008).

A third approach to downscaling is based on *observations* only and focuses on *temporal-scale changes* alone. In this case, downscaling aims at estimating statistical characteristics of a variable at a small temporal scale (short accumulation time), given observations with a long accumulation time. This approach – sometimes called temporal downscaling – assumes, for instance, that accumulated rainfall has scaling properties, which imply that statistical moments computed at different accumulation times are related (Perica and Foufoula-Georgiou, 1996; Deidda, 2000).

### 3.5.2 Downscaling methodologies

The methods used in statistical downscaling for climatology or in post-processing for weather prediction fall into three categories (Wilby and Wigley, 1997): regression-based methods, stochastic weather generator methods, and weather typing approaches. In *regression-based models*, the relationships between large-scale variables and location-specific values are directly estimated (or “translated”) via parametric or nonparametric methods, which can in turn be linear or nonlinear. These approaches include multiple linear regression (Wigley et al., 1990; Huth, 2002), kriging (Biau et al., 1999), neural networks (Snell et al., 2000; Cannon and Whitfield, 2002; Hewitson and Crane, 2002), logistic regression or quantile regression (Koenker and Bassett, 1978). Friederichs et al. (2009) reviewed the performance of different regression approaches in modeling wind gust observations, while Vrac et al. (2007b) and Salameh et al. (2009) discussed a nonlinear extension of these approaches using GAM models.

A *stochastic weather generator* (SWG) is a statistical model that generates realizations of local-scale variables, driven by large-scale model outputs (Wilks, 1998, 1999; Wilks and Wilby, 1999; Busuioc and von Storch, 2003; Chen et al., 2006; Orłowsky and Fraedrich, 2009; Orłowsky et al., 2010). An SWG has two main stochastic aspects: first, it often relates, for example, the probability of rain occurrence on day  $d$  to the event “rain” or “no-rain” on day  $d - 1$ ; this 1-day lagged relationship is often modeled through Markov chains (Katz, 1977). The second stochastic aspect of SWGs refers to the fact that the local-scale realizations are randomly drawn from a statistical distribution that is adapted to the data set. Hence, unlike in regression-based methods, providing the same large-scale input twice to an SWG can result in two different outputs. SWGs are also of particular interest in assessing local climate changes (Semenov and Barrow, 1997; Semenov et al., 1998).

The *weather typing approach* encapsulates a wide range of methods that share an algorithmic step in which recurrent large-scale or regional atmospheric patterns are identified through clustering methods applied over a large spatial area (Ghil and Robertson, 2002). Downscaling models can then be developed conditionally on the weather patterns or “weather regimes” so obtained (Zorita et al., 1993; Schnur and Lettenmaier, 1998; Huth, 2001). Introducing the intermediate layer of weather regimes in a downscaling procedure provides much greater modeling flexibility.

The analog method of Lorenz (1969) can be adapted to yield a special case of the weather-typing approach. In this case, there are as many weather patterns as there are days available for the calibration, and one uses as a forecast the local observation associated with the closest large-scale situation, i.e. the closet analog, in the historical record. One drawback in the context of studying extremes is that this method does not produce new values, and hence is not appropriate for risk assessment, in which one needs to extrapolate the data set outside the recorded set of values. This analog method has been successfully used, however, in numerous prediction studies (Barnett and Preisendorfer, 1978; Zorita and von Storch, 1998; Wetterhall et al., 2005; Bliefernicht and Bårdossy, 2007).

Many statistical downscaling methods use combinations of the three categories mentioned above. Some methods are compared in Schmidli et al. (2007), while Vrac and Naveau (2007) and Vrac et al. (2007c) combined weather typing and an SWG. Finally, Maraun et al. (2010) provide a comprehensive review on downscaling methods for precipitation.

### 3.5.3 Two approaches to downscaling extremes

In the context of downscaling extreme values, two approaches exist. The first one consists in downscaling directly an index of extreme events, such as the number of days with precipitation higher than a given value or the number of consecutive dry days. These indices try to capture certain

characteristics of the multivariate and time-dependent phenomena associated with weather and climate extremes, particularly those characteristics that have large socio-economic impacts. The goal of the joint Expert Team on Climate Change Detection and Indices (ETCCDI; <http://www.clivar.org/organization/etccdi/etccdi.php>) is to develop such indices for climate change studies.

In the second approach, the daily time series are first downscaled and then the local extreme values are extracted; the associated indices can potentially be computed from the latter. Schmidli et al. (2007) compared the pros and cons of the two approaches through six SDMs, and found it difficult to choose a single, overall winner: the SDM performances vary by index, region, season, and station. This conclusion also holds when comparing different SDMs for other-than-extreme values (Wilby et al., 2004).

Most of the statistical downscaling methods presented in Sects. 3.5.2 and 3.5.1 above can be directly applied to either one of the two approaches. For example, weather typing was applied by Davis et al. (1993) to derive a climatology of severe storms in Virginia (USA), and by Plaut et al. (2001) to characterize heavy precipitation events in several Alpine sub-regions. Busuioc et al. (2008) used a regression-based approach to downscale indices of winter extreme precipitation in a climate context. In the same context, Haylock et al. (2006) concluded that SDMs based on artificial neural networks are best at representing the interannual variability of heavy precipitation indices but underestimate extremes.

In contrast to the downscaling of climate extremes, which aims at estimating their slowly changing statistics, downscaling of extremes in the context of weather forecasting intends to estimate the actual risk of a weather extreme to occur. All national weather services issue warnings of severe or extreme weather. In general, occurrence probabilities of extreme weather are based on global ensemble simulations, where the extreme forecast index is related to a comparison between the distribution of the model climate and that of the forecasts (Lalaurette, 2003).

In the weather forecasting context, downscaling is usually achieved using limited-area models (LAMs; e.g., Montani et al., 2003; Marsigli et al., 2005; Brankovic et al., 2008); this methodology is called *dynamical downscaling*. Statistical downscaling and model post-processing are often hampered by the scarcity of historical data and homogeneous model simulations to train the statistical model (Hamill et al., 2006). It would clearly be desirable to combine dynamical and statistical downscaling methods based on LAMs, but we are not aware so far of studies that would try to derive conditional probabilities for extremes on the basis of LAMs, while using EVT theory.

In downscaling time series according to the second approach of this subsection, some recent SDMs relied on EVT results presented in Sects. 3.1–3.4. Such methods take advantage of the GPD – for values larger than a given, sufficiently high threshold – or the GEV, for block-size maxima,

and connect their parameters with large-scale atmospheric conditions. For example, Pryor et al. (2005) related large-scale predictors to the parameters of a Weibull distribution via linear regression, in order to downscale wind speed probability distributions for the end of the 21st century, while Vrac and Naveau (2007) defined an inhomogeneous mixture of a Gamma distribution and a GPD to downscale low, medium and extreme values of precipitation. Friederichs (2010) employed a nonstationary GPD with variable threshold and a nonstationary point process of Poisson type to derive conditional local-scale extreme quantiles, while Furrer and Katz (2008) adapted the SWG approach to extreme precipitation by representing high-intensity precipitation as a nonstationary Poisson point process.

To conclude this review of downscaling, one should keep in mind that it is better to use a range of different “good” SDMs than to use a single SDM, even the best. As in the rapidly spreading practice of using ensembles of GCMs (Solomon et al., 2007), the integration of several SDM results allows one to capture a wider range of uncertainties.

## 4 Dynamical modeling

### 4.1 A quick overview of dynamical systems

As outlined in Sect. 1, an important motivation of the E2-C2 project was to go beyond the mere statistics of extreme events, and use dynamical and stochastic-dynamic models for their understanding and prediction. It is well known, for instance, that a simple, first-order AR process (see Sect. 2.4.2) perturbed by multiplicative noise,

$$x(t+1) = a(t)x(t) + w(t), \quad (21)$$

can generate a heavy-tailed distribution for  $x$  (Kesten, 1973; Takayasu et al., 1997; Sornette, 1998).

It suffices for  $x$  to be heavy-tailed that the noisy coefficient  $a$  (i) have an expected value of  $\log(a)$  that is negative – for the stochastic process defined by the equation to be stationary – but that  $a$  itself exceed 1 from time to time, for  $x$  to experience intermittent amplification; and (iii) that the additive term  $w(t)$  be present to ensure repulsion from the origin  $x = 0$ . In fact, while for a linear AR[1] process  $w$  needs to be white noise – or, in more precise terms, a Wiener process – it suffices for  $x$  given by Eq. (21) to have a power-law distribution that  $w = \text{const.} \neq 0$ , provided the coefficient  $a$  satisfies conditions (i) and (ii).

This extremely simple way of generating a heavy-tailed distribution (see Sect. 2.3) is not, however, terribly informative in more specific situations, in which various other types of mathematical models might be in wide use. We start, therefore, by a quick tour of dynamical systems used throughout the geosciences and elsewhere, in Fig. 3. This will be followed, in Sect. 4.2, by a glimpse of simple deterministic systems in which the distribution of extreme events

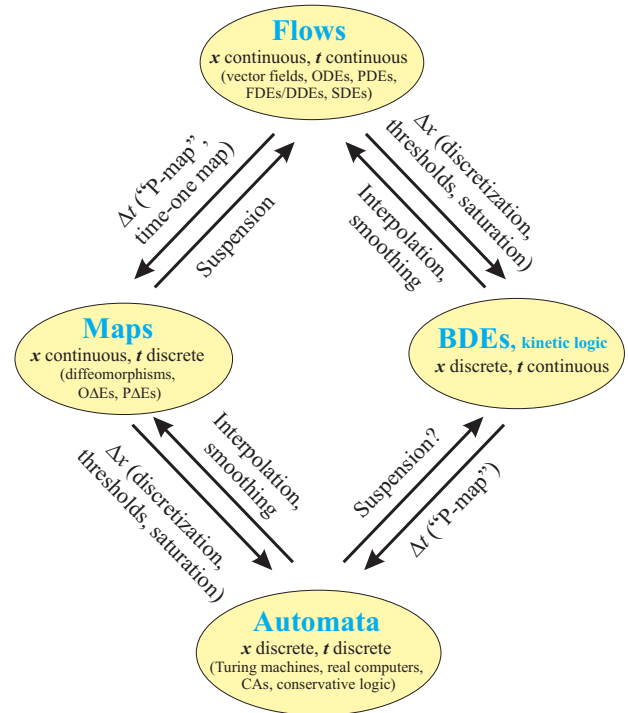
deviates from the one given by classical EVT theory. We concentrate next, in Sect. 4.3, on Boolean delay equations (BDEs: Dee and Ghil, 1984; Ghil and Mullhaupt, 1985; Ghil et al., 2008a), a fairly novel type of deterministic dynamical systems that facilitates the study of extreme events and clustering in complex systems for which the governing laws are known only very crudely.

Here goes our quick tour of Fig. 3: systems in which both variables and time are continuous are called *flows* (Arnol'd, 1983; Smale, 1967) (upper corner in the rhomboid of the figure). Vector fields, ordinary and partial differential equations (ODEs and PDEs), functional and delay-differential equations (FDEs and DDEs) – as well as stochastic differential equations (SDEs), like Eq. (21) – belong to this category. Ghil et al. (2008b) provide an example of studying extremes in a DDE model of the Tropical Pacific, namely warm (El Niño) and cold (La Niña) anomalies in its sea surface temperatures; see also Sect. 5.2.1 below. Systems with continuous variables and discrete time (middle left corner) are known as *maps* (Collet and Eckmann, 1980; Hénon, 1966) and include diffeomorphisms, as well as ordinary and partial difference equations (ODEs and PDEs).

In automata (lower corner) both the time and the variables are discrete; cellular automata (CAs) and all Turing machines (including real-world computers) are part of this group (Gutowitz, 1991; Wolfram, 1994; Von Neumann, 1966), and so are Boolean random networks (Kauffman, 1995, 1993) in their synchronous version. BDEs and their predecessors, kinetic (Thomas, 1979) and conservative logic, complete the rhomboid in the figure and occupy the remaining middle right corner.

In Fig. 3, we have outlined by labeled arrows the processes that can lead from the dynamical systems in one corner of the rhomboid to the systems in each one of the adjacent corners. The connections between flows and maps are fairly well understood, as they both fall in the broader category of *differentiable dynamical systems* (DDS: Smale, 1967; Collet and Eckmann, 1980; Arnol'd, 1983). Poincaré maps (“P-maps” in Fig. 3), which are obtained from flows by intersection with a plane (or, more generally, with a codimension-1 hyperplane) are standard tools in the study of DDS, since they are simpler to investigate, analytically or numerically, than the flows from which they were obtained. Their usefulness arises, to a great extent, from the fact that – under suitable regularity assumptions – the process of suspension allows one to obtain the original flow from its P-map, which also transfers its properties to the flow.

The systems that occupy the other two corners of the rhomboid – i.e., automata and BDEs – are referred to as *discrete dynamical systems* or dDS (Ghil et al., 2008a). Neither the processes that connect the two dDS corners, automata and BDEs, nor those that connect either type of dDS with the adjacent-corner DDS – maps and flows, respectively – are as well understood as the (P-map, suspension) pair of antiparallel arrows that connects the two DDS corners.



**Fig. 3.** Dynamical systems and connections between them. Note the links: The discretization of  $t$  can be achieved by the Poincaré map (P-map) or a time-one map, leading from *Flows* to *Maps*. The opposite connection is achieved by suspension. To go from *Maps* to *Automata* we use the discretization of  $x$ . Interpolation and smoothing can lead in the opposite direction. Similar connections lead from *BDEs* to *Automata* and to *Flows*, respectively. Modified after Mullhaupt (1984).

#### 4.2 Signatures of deterministic dynamics in extreme events

Complex deterministic dynamics encompasses models of abrupt transitions between multiple states and spatio-temporal chaos, along with many other types of irregular behavior; the latter are associated with a wide variety of phenomena of considerable concern, from the physical and environmental to the social sciences. It is therefore natural to inquire whether a theory of extremes can be built for such systems and, if so, does it reduce, for all practical intents and purposes to the classical EVT theory of Sect. 3 or, to the contrary, does it bear the imprint of the deterministic character of the underlying dynamics?

The answer is that both types of DDS behavior have been documented in the literature. A key difficulty is that – while interesting generalizations do exist – classical EVT theory deals with samples of i.i.d. variables, whereas successive points on the trajectory of a dynamical system are certainly not independent. Nor are they random, in the usual sense of the word, although the ergodic theory of deterministic DDS

(Eckmann and Ruelle, 1985) describes the circumstances under which such systems do possess invariant measures, as well as the properties of such measures.

One might suspect that – in the interesting case in which such a measure does exist, and the correlations among points on the trajectory decay rapidly – certain transformations of variables might reduce the problem of describing the extrema on a segment of trajectory to classical EVT theory. This idea has been successfully pursued by Freitas and Freitas (2008a,b); Freitas et al. (2010, 2011); Gupta et al. (2011) and Holland et al. (2011), among others. Their work did determine conditions under which EVT limit statistics holds for time series of observables – i.e., of scalar measurements – taken along trajectories of deterministic systems.

The opposite tack has been taken by a group of researchers centered in Brussels, and it is the latter work that we describe here briefly. Let  $X_t = f^t(X_0)$  be the formal solution of the evolution laws of a DDS. By the definition of determinism,  $f$  is a single-valued function of the initial state  $X_0$  and the observation time  $t$ . Our starting point is to realize that, in a system of this kind, the multivariate probability density to realize a sequence of states  $(X_0, \dots, X_{n-1})$ , sampled at equally spaced times  $t_0, t_0 + \tau, \dots, t_0 + (n-1)\tau$ , can be expressed as

$$\rho_n(X_0, \dots, X_{n-1}) = P(X = X_0 \text{ at } t = t_0) \times \prod_{k=1}^{n-1} P(X = X_{k-1} \text{ at } t = t^* | X = X_0 \text{ at } t = t^* - k\tau) \quad (22)$$

In either stochastic or deterministic systems, the first factor in Eq. (22) is given by the invariant probability density  $\rho(X_0)$ . This density is a smooth function of  $X_0$ , as long as the system possesses sufficiently strong ergodic properties. The two types of systems differ, however, by the nature of the conditional probabilities inside the  $n$ -fold product: in stochastic systems these quantities are typically smooth, while in DDS (see Sect. 4.1) they are Dirac delta functions,  $\delta(X_{k\tau} - f^{k\tau}(X_0))$ , where  $\tau$  is the sampling interval.

By definition, the cumulative probability distribution  $F_n(x)$  – a highly relevant quantity in a theory of extremes (see Sects. 2.3.3 and 3.1, for instance) – is the  $n$ -fold integral of  $\rho_n(X_0, \dots, X_{n-1})$  over its  $n$  arguments  $X_0, \dots, X_{n-1}$ , each taken from a finite lower bound  $a$  up to the level  $x$  of interest. This converts the delta-functions into Heaviside theta-functions, yielding:

$$F_n(x) = \int_a^x dX_0 \rho(X_0) \theta(x - f^\tau(X_0)) \dots \theta(x - f^{(n-1)\tau}(X_0)) \quad (23)$$

In other words,  $F_n(x)$  is obtained by integrating  $\rho(X_0)$  over those ranges of  $X_0$  in which  $\{x \geq f^\tau(X_0), \dots, x \geq f^{(n-1)\tau}(X_0)\}$ . As the threshold  $x$  is moved upwards, new integration ranges are thus added, since the slopes of the successive iterates  $f^{k\tau}$  with respect to  $X_0$  are, typically, different from each other, as well as  $X_0$ -dependent. Each of these ranges will open up past a threshold value where either the values of two different iterates will cross, or an iterate will

cross the manifold  $x = X_0$ . This latter type of crossing will occur at  $x$ -values that belong to the dynamical system's set of periodic orbits for all periods, up to and including  $n-1$ .

These properties of  $F_n(x)$  entail the following consequences: (i) Since a new integration range can only open up by increasing  $x$ , and the resulting contribution is necessarily nonnegative,  $F_n(x)$  is a monotonically increasing function of  $x$ . (ii) More unexpectedly, the slope of  $F_n(x)$  with respect to  $x$  will be subjected to abrupt changes at the discrete set of  $x$ -values that correspond to the successive threshold crossings. At these values the slope may increase or decrease, depending on the structure of the branches  $f^{k\tau}(X_0)$  involved in the particular crossing configuration considered. (iii) The probability density is the derivative of  $F_n(x)$  with respect to  $x$ , and will thus possess discontinuities at the points where  $F_n(x)$  is not differentiable; in general, it will not be monotonic.

While property (i) agrees, as expected, with statistical EVT theory, properties (ii) and (iii) are fundamentally different from the familiar ones in this theory: in EVT theory one deals with sequences of i.i.d. random variables, and the limiting distributions, if any, are smooth functions of  $x$  – see again Sect. 3.1 – while here we deal with the behavior of a time series generated by the evolution laws of a DDS. Balakrishnan et al. (1995) and Nicolis et al. (2006, 2008) have verified the predictions (i)–(iii) on prototypical DDS classes that show fully chaotic, intermittent and quasi-periodic behavior; in certain cases, moreover, they also derived the full analytic structure of  $F_n(x)$  and  $\rho_n(x)$ .

Figures 4a, b depict the cumulative distribution  $F(x)$  for two representative examples (solid lines):

- the cusp map (Fig. 4a):

$$X_{n+1} = 1 - 2|X_n|^{\frac{1}{2}}, \quad -1 \leq X_n \leq 1, \quad (24)$$

which gives rise to intermittent chaos; and

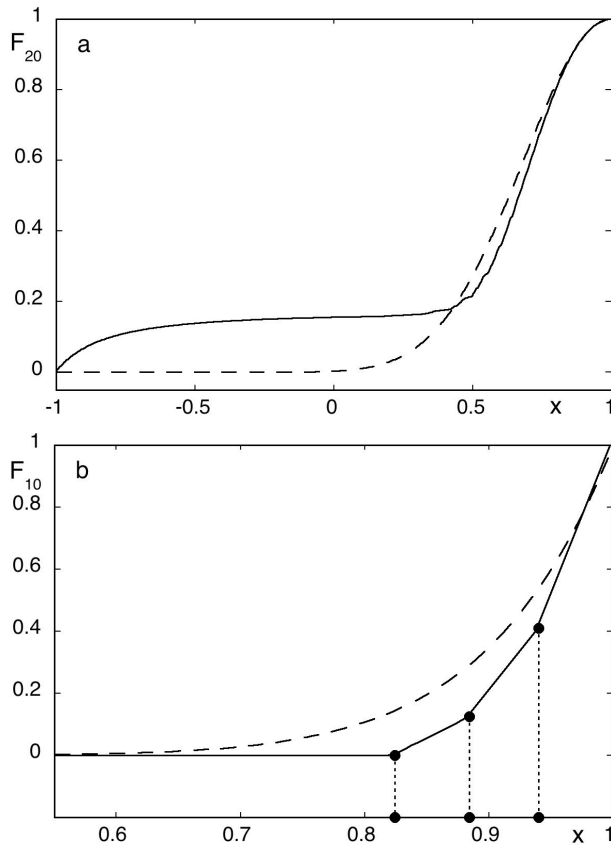
- the twist map (Fig. 4b):

$$X_{n+1} = a + X_n, \quad 0 \leq X_n \leq 1, \quad (25)$$

which, for  $a = (\sqrt{5} - 1)/3$ , gives rise to uniformly quasi-periodic behavior.

We turn next to the statistics of exceedance times of  $x$ , whose mean value is considered to be the principal predictor of extremes. Let  $r_k(x)$  be the probability density of the time  $k\tau$  of first exceedance of level  $x$ . Based on Eq. (22), it is straightforward to derive the formula

$$r_k(x) = \frac{1}{F(x)} \int_a^x dX_0 \rho(X_0) \int_a^x dX_\tau \delta(X_\tau - f^\tau(X_0)) \dots \int_a^b dX_k \delta(X_{k\tau} - f^{(k\tau)}(X_0)), \quad (26)$$



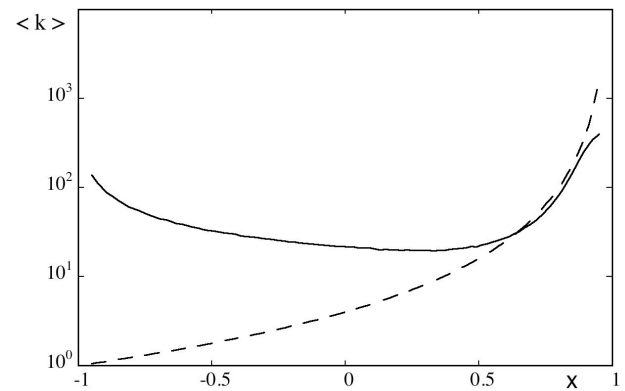
**Fig. 4.** Cumulative probability distribution  $F(x)$  of extremes (solid lines) for two deterministic dynamical systems. **(a)** For a discrete-time dynamical system that gives rise to intermittent behavior; and **(b)** for a uniform quasi-periodic motion described by two incommensurable frequencies. In panel **(b)**, the number of slope changes in  $F$  is indicated by vertical dotted lines and equals 3; dashed lines in both panels stand for the prediction of classical EVT theory. From Nicolis et al. (2006).

where  $a$  and  $b$  are the lower and upper bounds of  $X$  and  $F(x)$  is the cumulative distribution of the process. Converting, as before, the integrals of delta functions into theta functions and using Eq. (23), one arrives at a simple relation between  $r_k$  and  $F_k$ ,

$$r_k(x) = F_k(x) - F_{k+1}(x); \quad (27)$$

this relation suggests that  $r_k(x)$  may depend in an intricate way on  $x$  – much like  $F_k(x)$  in Fig. 4 – as well as on  $k$  (Nicolis and Nicolis, 2007).

Figure 5 depicts the mean exceedance time  $\langle k \rangle$  for the cusp map of Eq. (24). The difference between this result (solid line) and the predictions of EVT theory (dashed line) is even more striking than in Fig. 4. This discrepancy illustrates, once more, the importance of the differences between the assumptions, and hence the conclusions, of the two ap-



**Fig. 5.** Mean exceedance time  $\langle k \rangle$  for the deterministic system in Fig. 4a (solid line). The dashed line stands for the prediction of classical EVT theory. From Nicolis and Nicolis (2007).

proaches to EVT: classical, statistical vs. deterministic, at least in the absence of rapidly decaying correlations.

The results summarized in this subsection call for a closer look at the experimental data available on extreme events and at current practices prevailing in their prediction, from environmental science to insurance policies. As always, the most appropriate model selection for the process generating the extremes that we are interested in – deterministic or stochastic, discrete or continuous – is of utmost importance; see also Sects. 5.2.2 and 7.2.

### 4.3 Boolean delay equations (BDEs)

BDEs are a novel modeling framework especially tailored for the mathematical formulation of conceptual models of systems that exhibit threshold behavior, multiple feedbacks and distinct time delays (Dee and Ghil, 1984; Mullhaupt, 1984; Ghil and Mullhaupt, 1985; Ghil et al., 1987). The formulation of BDEs was originally inspired by advances in theoretical biology, following Jacob and Monod's discovery (Jacob and Monod, 1961) of on-off interactions between genes, which had prompted the formulation of “kinetic logic” (Thomas, 1973, 1978, 1979) and of Boolean regulatory networks (Kauffman, 1969, 1993, 1995). BDEs are intended as a heuristic first step on the way to understanding problems too complex to model using large systems of ODE, PDEs or FDEs at the present time. One hopes, of course, to be able to eventually write down and solve the exact equations that govern the most intricate phenomena. Still, in the geosciences as well as in the life sciences and elsewhere in the natural sciences, much of the preliminary discourse is often conceptual.

BDEs offer a formal mathematical language that may help bridge the gap between qualitative and quantitative reasoning. Besides, they are fun to play with and produce beautiful



fractals, by simple, purely deterministic rules (Ghil et al., 2008a).

In a hierarchical modeling framework (Ghil and Robertson, 2000), simple conceptual models are typically used to present hypotheses and capture isolated mechanisms, while more detailed models try to simulate the phenomena more realistically and test for the presence and effect of the suggested mechanisms by direct confrontation with observations. BDE modeling may be the simplest representation of the relevant physical concepts. At the same time, new results obtained with a BDE model often capture phenomena not yet found by using conventional tools (Saunders and Ghil, 2001; Zaliapin et al., 2003a,b). BDEs suggest possible mechanisms that may be investigated using more complex models once their “blueprint” was seen in a simple conceptual model. As the study of complex systems garners increasing attention and is applied to diverse areas – from microbiology to the evolution of civilizations, passing through economics and physics – related Boolean and other discrete models are being explored more and more (Gutowitz, 1991; Cowan et al., 1994; Wolfram, 1994; Kauffman, 1995; Gagneur and Casari, 2005).

BDEs are *semi-discrete dynamical systems*: the state variables are discrete – typically Boolean, i.e., taking the values 0 (“off”) or 1 (“on”) only – while time is allowed to be continuous. As such they occupy the previously “missing corner” in the rhomboid of Fig. 3, where dynamical systems are classified according to whether their time ( $t$ ) and state variables ( $x$ ) are continuous or discrete.

The development and applications of BDEs started only about two decades ago – a very short time span compared to ODEs, PDEs, maps, and even cellular automata. The results obtained so far, though, are sufficiently intriguing to warrant further exploration.

### 4.3.1 Formulation

Given a system with  $n$  continuous, real-valued state variables  $\mathbf{v} = (v_1, v_2, \dots, v_n) \in \mathbb{R}^n$  for which natural thresholds ( $q_i \in \mathbb{R} : 1 \leq q_i \leq n$ ) exist, one can associate with each variable  $v_i \in \mathbb{R}$  a Boolean-valued variable,  $x_i \in \mathbb{B} = \{0, 1\}$ , i.e., a variable that is either “on” or “off”, by letting

$$x_i = \begin{cases} 0, & v_i \leq q_i \\ 1, & v_i > q_i \end{cases}, \quad i = 1, \dots, n. \quad (28)$$

The equations that describe the evolution in time of the Boolean vector  $\mathbf{x} = (x_1, x_2, \dots, x_n) \in \mathbb{B}^n$ , due to the time-delayed interactions between the Boolean variables  $x_i \in \mathbb{B}$  are of the form:

$$\begin{cases} x_1(t) = f_1[t, x_1(t - \theta_{11}), \dots, x_n(t - \theta_{1n})], \\ x_2(t) = f_2[t, x_1(t - \theta_{21}), \dots, x_n(t - \theta_{2n})], \\ \vdots \\ x_n(t) = f_n[t, x_1(t - \theta_{n1}), \dots, x_n(t - \theta_{nn})]. \end{cases} \quad (29)$$

Each Boolean variable  $x_i : \mathbb{R} \rightarrow \mathbb{B}$  depends on time  $t$  and on the state of some or all the other variables  $x_j$  in the past. The functions  $f_i : \mathbb{B}^n \rightarrow \mathbb{B} \ 1 \leq i \leq n$ , are defined via Boolean equations that involve logical operators. Each delay value  $\theta_{ij} \in \mathbb{R}$ ,  $1 \leq i, j \leq n$ , is the length of time it takes for a change in variable  $x_j$  to affect the variable  $x_i$ .

*Asymptotic behavior.* In spite of the simplicity of their formulation, BDEs exhibit a considerable wealth of types of asymptotic behavior. Among these, the following were observed and analyzed in detail: (a) *fixed points* – the solution reaches one of a finite number of possible states and remains there; (b) *limit cycle* – the solution becomes periodic after a finite time elapses; (c) *quasi-periodicity* – the solution is a sum of several oscillatory “modes” with incommensurable periods; and (d) *growing complexity* – the number of jumps between the values 0 and 1, per unit time, increases with time  $t$ . This number grows like a positive, but fractional power of  $t$  (Dee and Ghil, 1984; Mullhaupt, 1984), with superimposed log-periodic oscillations (Ghil and Mullhaupt, 1985). The latter behavior is unique, so far, to BDEs, and seems to provide a metaphor for evolution in biological, historical, and other settings.

### 4.3.2 BDEs, cellular automata and Boolean networks

BDEs are naturally related to other dDS, to wit cellular automata and Boolean networks. We examine these relations very briefly herein.

A cellular automaton is defined as a set of  $N$  Boolean variables  $\{x_i : i = 1, \dots, N\}$  on the sites of a regular lattice, along with a set of logical rules for updating these variables at every time increment: the value of each variable  $x_j$  at epoch  $t$  is determined by the values of this and possibly some other variables  $\{x_i\}$  at the previous epoch  $t - 1$ . In the simplest case – of a one-dimensional lattice and first-neighbor interactions only – there are  $2^8$  possible updating rules  $f : \mathbb{B}^3 \rightarrow \mathbb{B}$ , which give 256 different *elementary cellular automata*, studied in detail by Wolfram (1983, 1994).

For a given  $f$ , these automata evolve according to:

$$x_i(t) = f[x_{i-1}(t-1), x_i(t-1), x_{i+1}(t-1)]. \quad (30)$$

For a finite  $N$ , Eq. (30) is a particular case of a BDE system (29) with a site-independent, but otherwise arbitrary connective  $f_i \equiv f$  and a unique delay  $\theta_{ij} \equiv 1$ . One speaks of *asynchronous cellular automata* when the variables  $x_i$  at different sites are updated at different discrete times, according to some deterministic scheme. Hence, for a finite  $N$ , the



latter dDS systems belong to the subset of BDEs that have integer delays  $\theta_{ij} \in \mathbb{N}$  only.

Boolean networks are a generalization of cellular automata in which the Boolean variables  $\{x_i\}$  are attached to the nodes (also called vertices) of a (possibly directed) graph and evolve synchronously, according to deterministic Boolean rules; these rules may vary from node to node. A further generalization is obtained by considering randomness, in the connections between the nodes, as well as in the choice of updating rules.

An extensively analyzed *random Boolean network*, the  $NK$  model, was introduced by Kauffman (1969, 1993). One considers a system of  $N$  Boolean variables such that each value  $x_i$  depends on  $K$  randomly chosen other variables through a Boolean function drawn at random and independently from  $2^{2^K}$  possibilities. The disorder is “quenched,” in the sense that the connections among the nodes and the updating functions are fixed for a given realization of the system’s evolution, and one looks for average properties at long times. Since the variables are updated synchronously, at the same discrete  $t$ -values, and  $N$  is finite, the evolution will ultimately reach a fixed point or a limit cycle for any network configuration and fixed dynamics (Ghil et al., 2008a).

#### 4.3.3 Towards hyperbolic and parabolic “partial” BDEs

In order to introduce an appropriate version of hyperbolic and parabolic BDEs, one looks for the Boolean-variable version of the typical hyperbolic and parabolic equations

$$\frac{\partial}{\partial t} v(z, t) = \frac{\partial}{\partial z} v(z, t), \quad \frac{\partial}{\partial t} v(z, t) = \frac{\partial^2}{\partial z^2} v(z, t), \quad (31)$$

respectively. In the scalar PDE framework and in one space dimension,  $v: \mathbb{R}^2 \rightarrow \mathbb{R}$ , and we wish to replace the real-valued function  $v$  with a Boolean function  $u: \mathbb{R}^2 \rightarrow \mathbb{B}$ , i.e.  $u(z, t) = 0, 1$ , with  $z \in \mathbb{R}$  and  $t \in \mathbb{R}^+$ , while conserving the same qualitative behavior of the solutions.

For Boolean-valued functions, one is led to use the “exclusive OR” (XOR) operator  $\nabla$  for evaluating differences,  $|u - w| = u \nabla w$ . Moreover, intuition dictates replacing the increments in the independent variables  $t$  and  $z$  by the *time* delay  $\theta_t$  and the *spatial* delay  $\theta_z$  when approximating the partial derivatives in Eq. (31). By doing so, first-order expansions in the differences yield the equations:

$$u(z, t + \theta_t) \nabla u(z, t) = u(z + \theta_z, t) \nabla u(z, t), \quad (32)$$

in the hyperbolic case, and

$$u(z, t + \theta_t) \nabla u(z, t) = u(z - \theta_z, t) \nabla u(z + \theta_z, t) \quad (33)$$

in the parabolic one, where we have used the associativity property of (mod 2) addition in  $\mathbb{B}$ .

Equation (32) has solutions of the form

$$u(z, t) = u(z + \theta_z, t - \theta_t). \quad (34)$$

This first step towards a partial BDE equivalent of a hyperbolic PDE displays therewith the expected behavior of a “wave” propagating in the  $(z, t)$  plane. The propagation is, respectively, from right to left for increasing times when using “forward differencing” in the right-hand side of Eq. (32), as we did, and it is in the opposite direction when using “backward differencing”. Notice that the solution exists and that it is unique for all  $(\theta_z, \theta_t) \in (0, 1]^2$ , given appropriate initial data  $u_0(z, t)$  with  $z \in (-\infty, \infty)$  and  $t \in [0, \theta_t)$ .

In continuous space and time, Eq. (32) under consideration is conservative and invertible. One can formulate for it either a pure Cauchy problem, when  $u_0(z, t)$  is given for  $z \in (-\infty, \infty)$  and is neither identically constant nor periodic, or a space-periodic initial boundary value problem, when  $u_0(z, t)$  is given for  $z \in [0, T_z)$  and  $u_0(z \pm T_z, t) = u_0(z, t)$ . In the latter case, the solution displays  $T_t$ -periodicity in time, as well as  $T_z$ -periodicity in space, with  $T_t = T_z \theta_t / \theta_z$ .

## 5 Applications

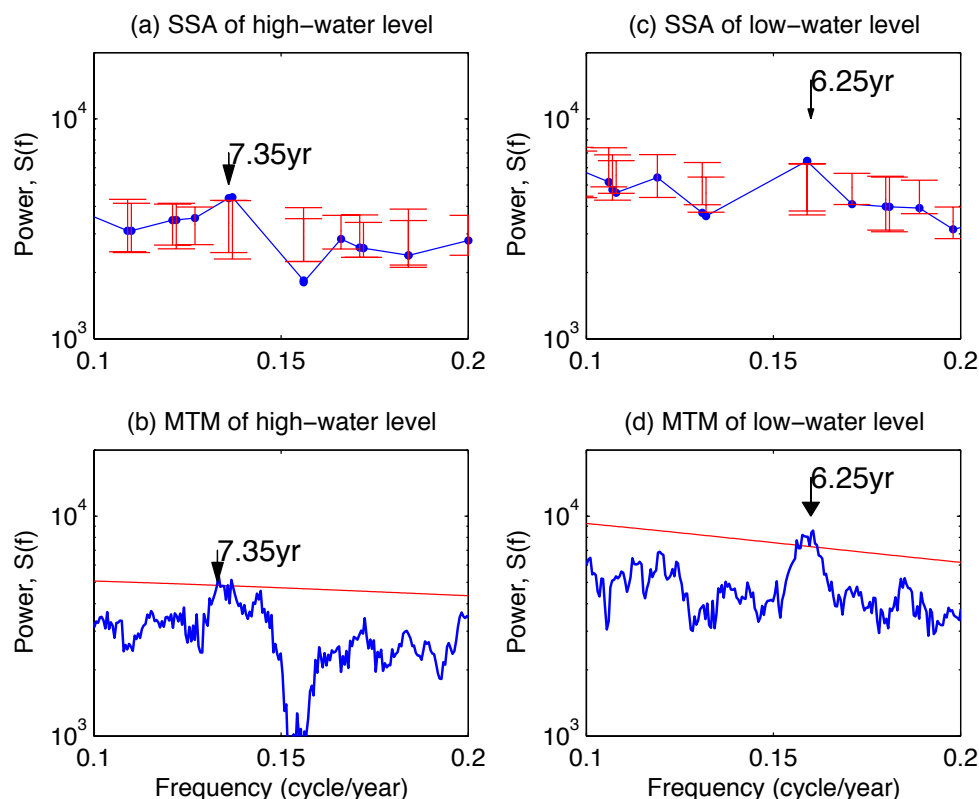
As indicated at the end of Sect. 1, many applications appear in the accompanying papers of this Special Issue. We select here only four applications. They illustrate first, in Sect. 5.1, the application of two distinct methodologies to the same problem, namely to the classical time series of Nile River water levels: in Sect. 5.1.1 we use SSA and MTM from Sect. 2.2, while in Sect. 5.1.2 we evaluate the LRD properties of these time series with the methods of Sect. 2.4.

Second, we apply in Sect. 5.2 the dDS modeling of Sect. 4.3 to two very different problems: the seasonal-to-interannual climate variability associated with the coupling between the Tropical Pacific and the global atmosphere, in Sect. 5.2.1, and the seismic activity associated with plate tectonics, in Sect. 5.2.2. The attentive reader can thus realize there is considerable flexibility for the methods-oriented, as well as the application-focused practitioner in the study of extreme events.

### 5.1 Nile River flow data

As an application of the theoretical concepts and tools of the statistical sections (Sects. 2 and 3), we consider the time series of Nile River flow data. These time series are the longest instrumental records of hydrological, climate-related phenomena, and extend over 1300 yr, from 622 AD – the initial year of the Muslim calendar (1 AH) – till 1921 AD, the year the first Aswan Dam was built. We use this time series in order to illustrate the complementarity between the analysis of the continuous background, cf. Sects. 2.3 and 2.4, and that of the narrow peaks that rise above this background, as described in Sect. 2.2.

The digital Nile River gauge data shown in Fig. 1 correspond to those used in Kondrashov et al. (2005) and are based on a compilation of primary sources (Toussoun, 1925;



**Fig. 6.** Spectral estimates (in blue) for the extended Nile River time series in the interannual frequency range of 5–10 yr. **(a, b)** Monte-Carlo SSA, and **(c, d)** MTM; **(a, c)** high-water levels, and **(b, d)** low-water levels. Arrows mark peaks that are highly significant against a null hypothesis of red noise; 95% error bars (for SSA) and 95% confidence level (for MTM), respectively, are in red. For SSA we used a 75-yr window and 1000 red-noise realizations, while for MTM, we took a bandwidth parameter of  $NW = 6$  and  $K = 11$  tapers. See information about other significant spectral peaks in Table 1.

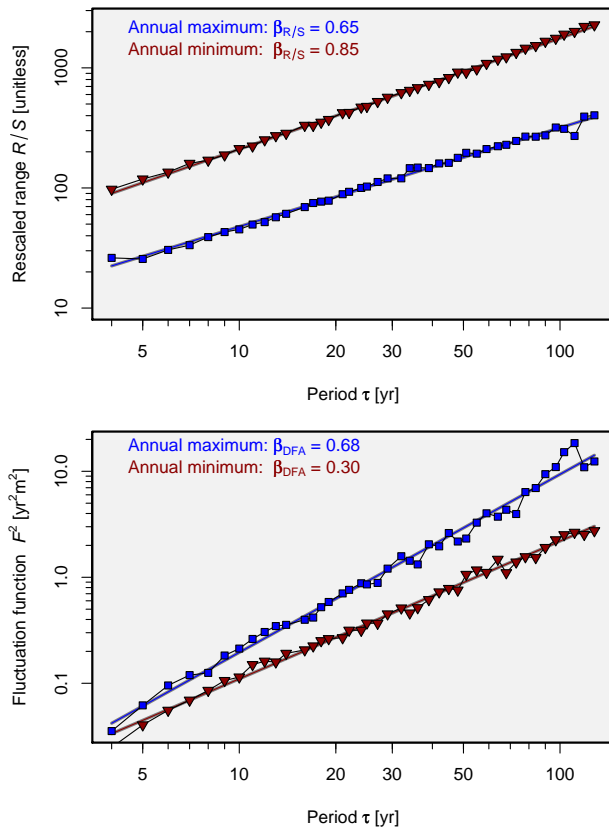
Ghaleb, 1951; Popper, 1951; De Putter et al., 1998; Whitcher et al., 2002). We consider first the oscillatory modes – as revealed by a combination of SSA (Sect. 2.2.1) with the maximum-entropy method and MTM (Sect. 2.2.2) – and then the LRD properties manifest in the continuous background, after subtracting the long-term trend and oscillatory modes, as described in Sects. 2.3 and 2.4.

### 5.1.1 Spectral analysis of the Nile River water levels

Following Kondrashov et al. (2005), we performed spectral analyses of the extended (622–1922) records of highest and lowest annual water level, with historic gaps filled by SSA (Kondrashov and Ghil, 2006). The two leading EOFs of an SSA analysis with a 75-yr window capture both a (nonlinear) trend and a very low-frequency component, with a period that is longer than the window. The SSA and MTM results for the detrended time series – obtained by removing this leading pair of RCs – are shown in Fig. 6; this figure focuses on interannual variability in the 5–10-yr range.

**Table 1.** Significant oscillatory modes in the extended Nile river records, 622–1922 AD. The main entries give the periods in years: bold entries without brackets indicate that the mode is significant at the 95 % level against a null hypothesis of red noise, in both SSA and MTM results; entries in square brackets are significant at this level only in the MTM analysis. Entries in parentheses provide the percentage of variance captured by the SSA mode with the given period.

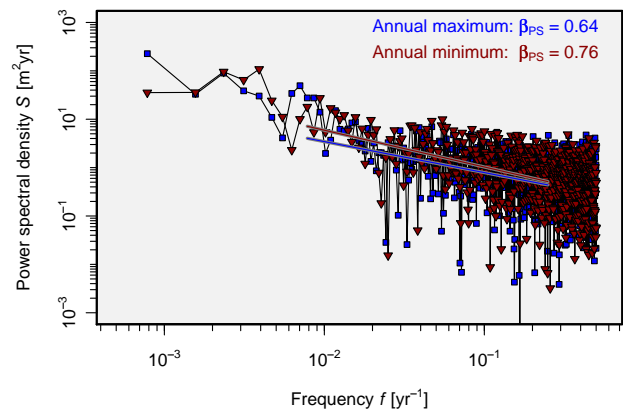
Period	Low	High
20–40 yr		23.2 (4.3 %)
10–20 yr	19.7 (6.9 %)	[15.5] (4.1 %)
	[12] (5.3 %)	
5–10 yr	6.25 (4.4 %)	7.35 (4.3 %)
0–5 yr	3.0 (4 %)	4.2 (3.4 %)
	2.2 (3.3 %)	2.9 (3.3 %)
		2.2 (3.5 %)



**Fig. 7.** LRD property estimates for Nile River time series in the 622–1900 AD interval. (a) Rescaled-range (R/S) analysis; and (b) detrended fluctuation analysis (DFA). Results for annual minimum water levels (red triangles; see also Fig. 2) and maximum levels (blue squares). Straight, blue and red lines are the best-fit power-law functions for window widths of  $4 \text{ yr} \leq \tau \leq 128 \text{ yr}$ . The estimated correlation strength values of  $\beta_{R/S}$  and  $\beta_{DFA}$  are given in the figure legend; the color of the estimates matches that of the corresponding line.

Both spectra show peaks that are significant with respect to an AR[1] background at the 95 % level:  $T = 7.35 \text{ yr}$  for the high-water and  $T = 6.25 \text{ yr}$  for the low-water levels. The higher-frequency part of the spectra have peaks at  $T \simeq 3 \text{ yr}$  and  $T \simeq 2.2 \text{ yr}$  (not shown in the figure); see detailed information about these and other significant spectral peaks in Table 1. The peaks in the 2–5-yr range were known from earlier work on these records and are commonly attributed to teleconnections with the El Niño/Southern Oscillation (ENSO) and with the monsoonal circulations over the Indian Ocean, especially over East Africa.

The 6–8 yr peak in the Nile river records was the main new finding of Kondrashov et al. (2005). These authors, followed by Feliks et al. (2010), attributed it to teleconnections with a similar spectral peak in the North Atlantic Oscillation. This spectral peak, in turn, could be of internal oceanic origin, arising through a gyre-mode instability of the wind-



**Fig. 8.** Spectral analysis (Lomb periodogram) results for the 622 to 1900 AD Nile River annual minimum (red triangles) and maximum (blue squares) water level time series (see Fig. 1). Shown are the power-spectral densities  $S$  plotted as a function of frequency  $f$  on log-log scale. Blue and red lines represent the best fitting power law functions in the frequency range  $(1/128) \text{ yr}^{-1} < f < (1/4) \text{ yr}^{-1}$ , the same frequency range (period range) as done for DFA and R/S (Fig. 7). The estimated long-range correlation strengths ( $\beta_{PS}$ ) are given in the color of the corresponding line.

driven, double-gyre circulation of the North Atlantic Ocean; it has been found across a hierarchy of ocean models (Dijkstra and Ghil, 2005; Simonnet et al., 2005). The results in Fig. 6 suggest that the climate of East Africa has been subject to influences from the North Atlantic, besides those already documented from the Tropical Pacific and Indian Ocean.

### 5.1.2 Long-range persistence in Nile River water levels

The Nile River records are a paradigmatic example of data sets with the LRD property (e.g., Hurst, 1951, 1952; Mandelbrot and Wallis, 1968a,b). Mandelbrot and Wallis (1968a), in particular, suggested that the “Joseph effect” of being able to predict good years for the Nile River floods – and hence for the crops in pharaonic Egypt – were due to this LRD property, while Mandelbrot and Wallis (1968b) verified that Hurst’s R/S estimate (Hurst, 1951, 1952) of what is now called the Hurst exponent  $H$  was essentially correct.

As we can see from the previous subsection, though, the attribution of the “Joseph effect” to LRD is not that conclusive: a 6–8-yr periodicity is present in the water levels, and is more likely to account for the biblical 14-yr cycle of fat and lean years than scaling effects. The latter would result in a much more irregular distribution of positive- and negative-sign excursions from the mean, and not allow for a precise prediction, like the one alluded to in the biblical story.

We reanalyzed, therefore, the high- and low-level records of Fig. 1, after subtracting the trend and main periodicities, as discussed in Sects. 2.3.1 and 2.4.1. Both time series were investigated to verify whether long-range correlations existed

**Table 2.** Results of long-range persistence analysis for the Nile River records (622–1900 AD) of low- and high-water levels in Fig. 7; the three different techniques are power-spectral analysis (PS), detrended fluctuation analysis (DFA), and rescaled-range analysis (R/S).

	$\beta_{\text{PS}}$	$\beta_{\text{DFA}}$	$\beta_{\text{R/S}}$
Nile River annual minimum water level	0.76	0.65	0.30
Nile River annual maximum water level	0.64	0.85	0.68

and, if so, to quantify their strength. Due to the gaps in the time series, the analysis techniques introduced in Sect. 2.4.3 needed slight modifications: The standard periodogram was replaced by the Lomb (1976) periodogram. For the R/S and DFA calculations, the time series was broken up into segments that cover time intervals of equal length, and only the segments without gaps were used in these calculations.

In Figs. 2 and 7, we show the results of several distinct estimation methods for self-affine LRD processes, cf. Sect. 2.4. Figure 2 refers exclusively to the low-level data and was already discussed in Sect. 2.4.3. The estimates there were based on the GPH estimator of Geweke and Porter-Hudak (1983), as applied to the low frequencies of the periodogram for the time series in Fig. 1; they yielded  $\hat{\beta} = 0.76 \pm 0.09$  and hence  $\hat{H} = 0.88 \pm 0.04$ . Applying the FD methodology to the same time series led to almost exactly the same estimates: the FD straight-line fit to the spectrum (red) is practically indistinguishable from the GPH slope (blue).

In Figs. 7 and 8, three LRD estimation methods were applied to the maximum, as well as the minimum water level series for the years 622–1900 AD only; the last two decades were taken out, since they have a strong trend, due to sil-tation. Figure 7 illustrates the results for the R/S and DFA methods, while those for the least-square fit to the Lomb periodogram are in Fig. 8.

All three evaluation methods exhibit power-law behavior in the respective ranges of frequencies or scales; they thus all indicate the presence of LRD properties. The estimated strengths, as measured by the  $\beta$ -exponent (see Sect. 2.4.2), are given in Table 2; they range between  $\beta = 0.30$  and  $\beta = 0.85$ . Further analyses were carried out in order to decide if this variation in the persistence strength is caused by finite-sample-size effects or whether nonstationarities or higher-order correlations are responsible for it.

In order to better understand the nature of both random and systematic errors in the strength estimators for long-range correlations, their performance was evaluated by generating 200 synthetic time series, with Gaussian one-point statistics: 100 for the annual maxima and 100 for the annual minima of the water levels. These time series were each generated so as to be self-affine with persistence strength  $\beta = 0.5$ . The synthetic time series were modeled to include the same data gaps as the original Nile River data for the 1300 yr of record

(622–1921 AD): for the minima there were 272 “missing” values, while for the maxima there were 190 missing values out of the 1300 yr. The resultant synthetic time series were thus unequally spaced in time, and they were treated as described above for the original “raw” Nile River data.

We applied all three estimators to these two sets of 100 time series with  $\beta = 0.5$ . The mean values and standard deviations of the measured persistence strengths for these synthetic time series are given in Table 3.

These results show that R/S analysis is strongly biased, i.e. the persistence strength is on average significantly underestimated, in agreement with the empirical LRD studies of Taqqu et al. (1995) and Malamud and Turcotte (1999). We restricted, therefore, our consideration for the actual records in Table 2 to power spectral analysis (PS in the table) and DFA, both of which have a negligible bias and relatively small standard deviations of approximately 0.10 for the synthetic data sets. Assuming that the uncertainties of the estimated persistence strength are of the same size for the synthetic data sets and the Nile River records, we can state that (i) the two Nile River time records both exhibit long-range persistence; (ii) the persistence strength of the two records is statistically indistinguishable; (iii) this strength is  $\hat{\beta} \approx 0.72$ , with a standard error of 0.10, such that the corresponding 95 % confidence interval is  $0.53 < \hat{\beta} < 0.91$ ; and (iv) PS and DFA lead to mutually consistent results.

## 5.2 BDE modeling in action

We now turn to an illustration of BDE modeling in action, first with a climatic example and then with one from lithospheric dynamics. Both of these applications introduce new and interesting extensions to and properties of BDEs. In particular, they illustrate the ability of this deterministic modeling framework to deal with the understanding and prediction of extreme events.

In Sect. 5.2.1 we show that a simple BDE model can mimic rather well the solution set of a much more detailed model, based on nonlinear PDEs, as well as produce new and previously unsuspected results, such as a Devil’s Staircase and a “bizarre” attractor in the parameter space. This model provides a minimal setting for the study of extreme warm and cold events in the Tropical Pacific.

The seismological BDE model in Sect. 5.2.2 introduces a much larger number of variables, organized in a directed graph, as well as random forcing and state-dependent delays. This BDE model also reproduces a regime diagram of seismic sequences resembling observational data, as well as the results of much more detailed models (Gabrielov et al., 2000a,b) that are based on a system of differential equations; furthermore, it allows the exploration of seismic prediction methods for large-magnitude events.

**Table 3.** Mean values and standard errors for the three estimators of the strength of long-range persistence for the two synthetic data sets with  $\beta = 0.5$ ; see text and caption of Fig. 7 for the construction of these  $2 \times 100$  time series. Symbols for the three analysis methods – PS, DFA and R/S – as in Table 2. The resultant distribution of persistence strengths is approximately Gaussian: the corresponding mean values and standard deviations are given in the table.

	$\beta_{PS}$	$\beta_{DFA}$	$\beta_{R/S}$
Synthetic time series with a time-sampling identical to the Nile River annual-minimum water level records	$0.49 \pm 0.10$	$0.46 \pm 0.11$	$0.41 \pm 0.17$
Synthetic time series with a time-sampling identical to the Nile River annual-maximum water level records	$0.49 \pm 0.10$	$0.48 \pm 0.08$	$0.43 \pm 0.11$

### 5.2.1 A BDE model for the El Niño/Southern Oscillation (ENSO)

#### Conceptual ingredients

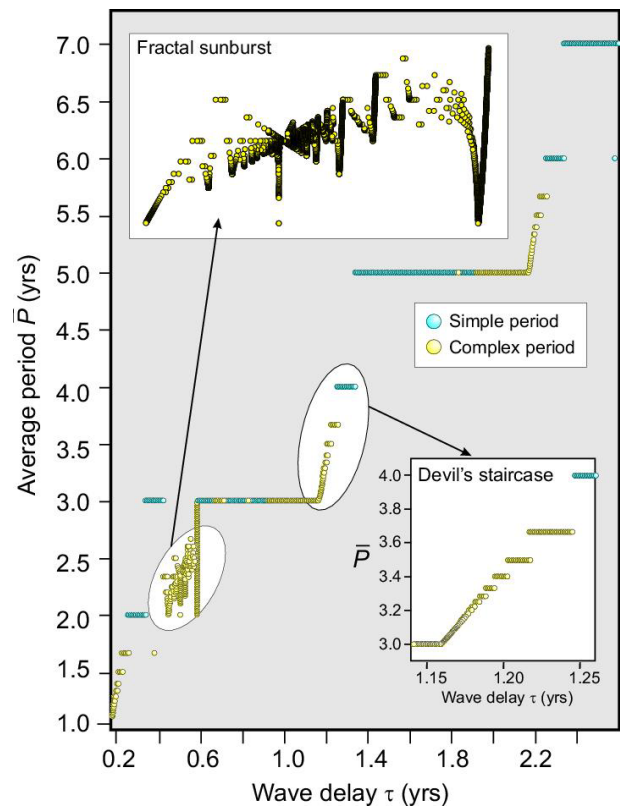
The ENSO phenomenon is the most prominent signal of seasonal-to-interannual climate variability (Diaz and Markgraf, 1993; Philander, 1990; Latif et al., 1994). The BDE model for ENSO variability of Saunders and Ghil (2001) incorporated the following three conceptual elements: (i) the Bjerknes hypothesis, which suggests a *positive feedback* as a mechanism for the growth of an internal instability that could produce large positive anomalies of sea surface temperatures (SSTs) in the eastern Tropical Pacific. (ii) A *negative feedback* via delayed oceanic wave adjustments that compensates Bjerknes's positive feedback and allows the system to return to colder conditions in the basin's eastern part. (iii) Seasonal forcing (Chang et al., 1994, 1995; Jin et al., 1994, 1996; Tziperman et al., 1994, 1995).

This model operates with five Boolean variables. The discretization of continuous-valued SSTs and surface winds into four discrete levels is justified by the pronounced multimodality of associated signals. The state of the *ocean* is depicted by SST anomalies, expressed via a combination of two Boolean variables,  $T_1$  and  $T_2$ . The anomalous *atmospheric* conditions in the Equatorial Pacific basin are described by the variables  $U_1$  and  $U_2$ . The latter express the state of the trade winds.

For both the atmosphere and the ocean, the first variable,  $T_1$  or  $U_1$ , describes the sign of the anomaly, positive or negative, while the second one,  $T_2$  or  $U_2$ , describes its amplitude, strong or weak. Thus, each one of the pairs  $(T_1, T_2)$  and  $(U_1, U_2)$  defines a four-level discrete variable that represents highly positive, slightly positive, slightly negative, and highly negative deviations from the climatological mean. The *seasonal cycle's* external forcing is represented by a two-level Boolean variable  $S$ .

#### BDE system

These conceptual ingredients lead to the following set of Boolean delay equations:



**Fig. 9.** Devil's staircase and fractal sunburst: A bifurcation diagram showing the average cycle length  $\bar{P}$  vs. the wave delay  $\tau$  for a fixed  $\beta = 0.17$ . Blue dots indicate purely periodic solutions; orange dots are for complex periodic solutions; small black dots denote aperiodic solutions. The two insets show a blow-up of the overall, approximate behavior between periodicities of two and three years ("fractal sunburst") and of three and four years ("Devil's staircase"). Modified after Saunders and Ghil (2001).

$$U_i(t) = T_i(t - \beta), \quad i = 1, 2,$$

$$S(t) = S(t - 1),$$

$$T_1(t) = \{[R \wedge \overline{U_1}](t - \tau)\} \vee \{\overline{R}(t - \tau) \wedge U_2(t - \beta)\},$$

$$T_2(t) = \{[S \Delta T_1](t - \beta)\} \vee \{[\overline{(S \Delta T_1)} \wedge T_2](t - \beta)\}.$$

Here the symbols  $\vee$  and  $\wedge$  represent the binary logical operators OR and AND, respectively.

The model's principal parameters are the two delays  $\beta$  and  $\tau$ ; they are associated with local adjustment processes and with basin-wide processes, respectively. The changes in wind conditions are assumed to lag the SST variables by a short delay  $\beta$ , of the order of days to weeks. The length of the delay  $\tau$  may vary from about one month to two years (Jin and Neelin, 1993a,b,c; Jin, 1996).

### Model solutions

The following solution types have been found in this model: (i) *Periodic solutions with a single cycle (simple period)*. Each succession of events, or *internal cycle*, is completely phase-locked here to the seasonal cycle, i.e., the warm events always peak at the same time of year. When keeping the local delay  $\beta$  fixed and increasing the wave-propagation delay  $\tau$ , intervals where the solution has a simple period equal to 2, 3, 4, 5, 6, and 7 yr arise consecutively.

(ii) *Periodic solutions with several cycles (complex period)*. We describe such sequences, in which several distinct cycles make up the full period, by the parameter  $\bar{P} = P/n$ ; here  $P$  is the length of the sequence and  $n$  is the number of cycles in the sequence. Notably, for a substantial range of  $\beta$ -values,  $\bar{P}$  becomes a nondecreasing step function of  $\tau$  that takes only rational values, arranged on a Devil's Staircase. This frequency-locking behavior is a signature of the universal quasi-periodic route to chaos. Its mathematical prototype is the Arnol'd (1983) circle map and it has been documented across a hierarchy of ENSO models, with some of its features being even present in certain time series of observed variables (Ghil and Robertson, 2000).

Aside from this Devil's Staircase, the BDE model also exhibits – between the two broad steps at the averaged periods of two and three years – a much more complex, and heretofore unsuspected, “fractal sunburst” structure; see the upper inset in Fig. 9. As the wave delay  $\tau$  is increased, mini-ladders build up, collapse or descend only to start climbing up again. The pattern's focal point is at a critical value of  $\tau \cong 0.5$  yr; near it these mini-ladders rapidly condense and the structure becomes self-similar, as each zoom reveals the pattern being repeated on a smaller scale. Saunders and Ghil (2001) called this a “bizarre” attractor because it is more than “strange”: strange attractors occur in a system's phase space, for fixed parameter values, while this fractal sunburst appears in the BDE model's phase–parameter space, like the Devil's Staircase. The structure in Fig. 9 is attracting, though, only in phase space; it is, therefore, a generalized attractor, and not just a bizarre one.

In Sect. 7.1, we shall return to the prediction of ENSO extrema. This prediction will be based on the dominant oscillatory patterns in the sequence of warm and cold events.

## 5.2.2 A BDE model for seismicity

### Conceptual ingredients

Lattice models of systems of interacting elements are widely applied for modeling seismicity, starting from the pioneering works of Burridge and Knopoff (1967), Allègre et al. (1982), and Bak et al. (1988). The state of the art is summarized in several review papers (Keilis-Borok and Shebalin, 1999; Newman et al., 1994; Rundle et al., 2000; Turcotte et al., 2000; Keilis-Borok, 2002). Recently, so-called *colliding cascade* models (Gabrielov et al., 2000a,b; Zaliapin et al., 2003a,b) have been able to reproduce a substantial set of observed characteristics of earthquake dynamics (Keilis-Borok, 1996; Turcotte, 1997; Scholz, 2002). These characteristics include (i) the seismic cycle; (ii) intermittency in the seismic regime; (iii) the size distribution of earthquakes, known as the Gutenberg-Richter relation; (iv) clustering of earthquakes in space and time; (v) long-range correlations in earthquake occurrence; and (vi) a variety of seismicity patterns that seem to foreshadow strong earthquakes.

Introducing the BDE concept into modeling of colliding cascades, Zaliapin et al. (2003a) replaced the simple interactions of elements in the system by their integral effect, represented by the delayed switching between the distinct states of each element: unloaded or loaded, and intact or failed. In this way, one can bypass the necessity of reconstructing the global behavior of the system from the numerous complex and diverse interactions whose details are only being understood very incompletely. Zaliapin et al. (2003a,b) have shown that this modeling framework does simplify the detailed study of the system's dynamics, while still capturing its essential features. Moreover, the BDE results provide additional insight into the system's range of possible behavior, as well as into its predictability.

Colliding cascade models (Gabrielov et al., 2000a,b; Zaliapin et al., 2003a,b) synthesize three processes that play an important role in lithosphere dynamics, as well as in many other complex systems: (i) the system has a hierarchical structure; (ii) it is continuously loaded (or driven) by external sources; and (iii) the elements of the system fail under the load, causing the redistribution of the load and strength throughout the system. Eventually the failed elements heal, thereby ensuring the continuous operation of the system.

The load is applied at the top of the hierarchy and transferred downwards, thus forming a *direct cascade of loading*. Failures are initiated at the lowest level of the hierarchy, and gradually propagate upwards, thereby forming an *inverse cascade of failures*, which is followed by healing. The interaction of direct and inverse cascades establishes the dynamics of the system: loading triggers the failures, and failures redistribute and release the load. In its applications to seismicity, the model's hierarchical structure represents a fault network, loading imitates the effect of tectonic forces, and failures imitate earthquakes.

**Table 4.** Long-term averaged GDP losses due to the distribution of natural disasters for different types of economic dynamics.

Calibration	Economic dynamics	Mean GDP losses due to natural disasters (% of baseline GDP)
No investment flexibility $\alpha_{\text{inv}} = 0.0$	Stable	0.15 %
Low investment flexibility $\alpha_{\text{inv}} = 1.0$	Stable	0.01 %
High investment flexibility $\alpha_{\text{inv}} = 2.5$	Endogenous business cycle	0.12 %

The model acts on a directed graph whose nodes, except the top one and the bottom ones, have connectivity six: each internal node has one parent, two siblings, and three children. Each element possesses a certain degree of *weakness* or *fatigue*. An element fails when its weakness exceeds a prescribed threshold.

The model runs in discrete time  $t = 0, 1, \dots$ . At each epoch  $t$ , a given element may be either *intact* or *failed (broken)*, and either *loaded* or *unloaded*. The state of an element  $e$  at a epoch  $t$  is defined by two Boolean functions:  $s_e(t) = 0$ , if an element is intact, and  $s_e(t) = 1$ , if it is in a failed state;  $l_e(t) = 0$ , if an element is unloaded, and  $l_e(t) = 1$ , if it is loaded.

An element may switch from a state  $(s, l)$  to another one under an impact from its nearest neighbors and external sources. The dynamics of the system is controlled by the time delays between the given impact and switching to another state. At the start,  $t = 0$ , all elements are in the state  $(0, 0)$ , intact and unloaded; failures are initiated randomly within the elements at the lowest level.

Most of the changes in the state of an element occur in the following cycle:

$$(0, 0) \rightarrow (0, 1) \rightarrow (1, 1) \rightarrow (1, 0) \rightarrow (0, 0) \dots$$

However, other sequences are also possible, but a failed and loaded element may switch only to a failed and unloaded state,  $(1, 1) \rightarrow (1, 0)$ . This mimics fast stress drop after a failure.

Zaliapin et al. (2003a,b) introduced four basic time delays:  $\Delta_L$ , between being impacted by the load and switching to the loaded state,  $(\cdot, 0) \rightarrow (\cdot, 1)$ ;  $\Delta_F$ , between an increase in weakness and switching to the failed state,  $(0, \cdot) \rightarrow (1, \cdot)$ ;  $\Delta_D$ , between failure and switching to the unloaded state,  $(\cdot, 1) \rightarrow (\cdot, 0)$ ; and  $\Delta_H$ , between the moment when healing conditions are established and switching to the intact state,  $(1, \cdot) \rightarrow (0, \cdot)$ . It was found, though, that the two most important delays are the loading time  $\Delta_L$  and the healing time  $\Delta_H$ .

## Model solutions

The output of this BDE model is a catalog of earthquakes – *i.e.*, of failures of its elements – similar to the simplest routine catalogs of observed earthquakes:

$$\mathcal{C} = (t_k, m_k, h_k), \quad k = 1, 2, \dots; \quad t_k \leq t_{k+1}. \quad (35)$$

In real-life catalogs,  $t_k$  is the starting time of the rupture;  $m_k$  is the magnitude, a logarithmic measure of energy released by the earthquake; and  $h_k$  is the vector that comprises the coordinates of the hypocenter. The latter is a point approximation of the area where the rupture started. In the BDE model, earthquakes correspond to failed elements,  $m_k$  is the highest level at which a failed element is situated within the model hierarchy, while its position within level  $m_k$  is a counterpart of  $h_k$ .

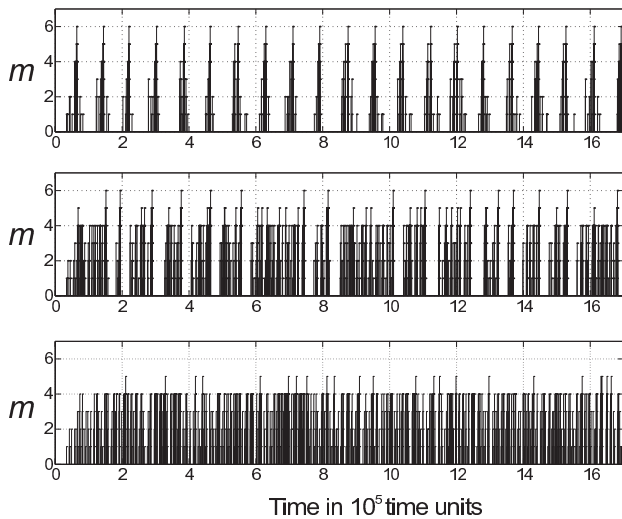
## Seismic regimes

A long-term pattern of seismicity within a given region is usually called a *seismic regime*. It is characterized by the frequency and irregularity of the strong earthquakes' occurrence, more specifically by (a) the Gutenberg-Richter relation, *i.e.* the time-and-space averaged magnitude–frequency distribution; (b) the variability of this relation with time; and (c) the maximal possible magnitude.

Our BDE model produces synthetic sequences that can be divided into three seismic regimes, illustrated in Fig. 10.

- Regime H: *High and nearly periodic seismicity* (top panel of Fig. 10). The fractures within each cycle reach the top level,  $m = L$ , where the underlying ternary graph has depth  $L = 6$ . The sequence is approximately periodic, in the statistical sense of cyclo-stationarity (Mertins, 1999).
- Regime I: *Intermittent seismicity* (middle panel of Fig. 10). The seismicity reaches the top level for some but not all cycles, and cycle length is very irregular.
- Regime L: *Low or medium seismicity* (lower panel of Fig. 10). No cycle reaches the top level and seismic





**Fig. 10.** Three seismic regimes: sample of earthquake sequences. Top panel – Regime H (High),  $\Delta_H = 0.5 \cdot 10^4$ ; middle panel – Regime I (Intermittent),  $\Delta_H = 10^3$ ; bottom panel – Regime L (Low),  $\Delta_H = 0.5 \times 10^3$ . Only a small fraction of each sequence is shown, to illustrate the differences between regimes. Reproduced from Zaliapin et al. (2003a), with kind permission of Springer Science and Business Media.

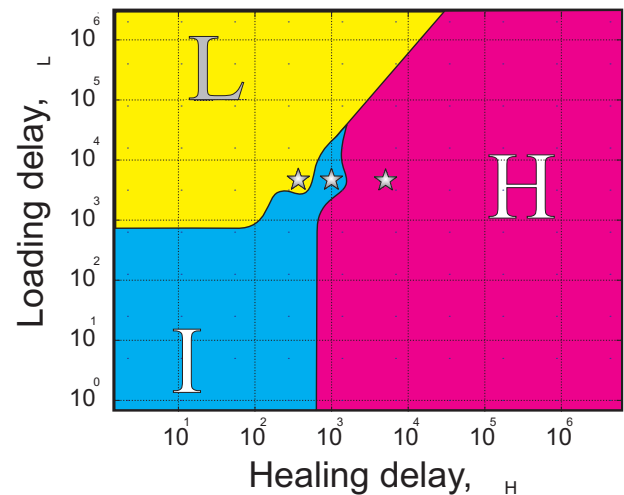
activity is much more constant at a low or medium level, without the long quiescent intervals present in Regimes H and I.

Figure 11 shows the location of these three regimes in the plane of the two key parameters  $(\Delta_L, \Delta_H)$ .

Zaliapin et al. (2003b) applied these results to the problem of earthquake prediction discussed in detail in Sect. 7.2.1. These authors used their model's simulated catalogs to study in greater detail the performance of pattern recognition methods tested already on observed catalogs and other models (Keilis-Borok and Malinovskaya, 1964; Molchan et al., 1990; Bufe and Varnes, 1993; Keilis-Borok, 1994; Pepke et al., 1994; Knopoff et al., 1996; Bowman et al., 1998; Jaume and Sykes, 1999; Keilis-Borok, 2002), as well as devising new methods. Most interestingly, they formulated optimization algorithms for combining different individual premonitory patterns into a collective prediction algorithm.

## 6 Macroeconomic modeling and impacts of natural hazards

A key issue in the study of extreme events is their impact on the economy. In the present section, we focus on whether economic shocks – such as those caused by climate-related damages or other natural hazards – will, or could, influence long-term growth pathways and generate a permanent and sizable loss of welfare. It turns out, though, that estimating this impact correctly depends on the type of macroeco-



**Fig. 11.** Regime diagram in the  $(\Delta_L, \Delta_H)$  plane of the loading and healing delays. Stars correspond to the sequences shown in Fig. 10. Reproduced from Zaliapin et al. (2003a), with kind permission of Springer Science and Business Media.

nomic model being used. Economists have so far used in such estimates, by-and-large, long-term growth models in the Solow (1956) tradition (Nordhaus and Boyer, 1998; Stern, 2006; Solomon et al., 2007), while relying on the idea that – over time scales of decades to centuries – the equilibrium paradigm is an acceptable metaphor. In such a setting, the crucial element of climate-change cost assessments, for instance, is the discount rate used for evaluating short-term costs of adaptation or mitigation policies vs. their long-term benefits.

Balanced-growth models, though, incorporate many sources of flexibility and they tend to suggest that the damages caused by disruptions of the natural – i.e., physical, chemical and ecological – system cannot entail more than fairly moderate losses in gross domestic product (GDP) over the 21st century (Solomon et al., 2007). Likewise, these models imply that the secondary, long-term effects of past events like the landfall of Hurricane Katrina near New Orleans in 2005 had only small economic consequences. These results have been heatedly debated, and it has been suggested that equilibrium models underestimate costs, because unbalanced economies may be more vulnerable to external shocks than the idealized, equilibrated ones that such models describe.

To investigate the role of short-term economic variability, such as business cycles, in analyzing the economic impacts of extreme events, we consider here a simple modeling framework that is able to reproduce endogenous economic dynamics arising from intrinsic factors. We then investigate how such an out-of-equilibrium economy would react to external shocks like natural disasters. The results reviewed here support the idea that endogenous dynamics and exogenous shocks interact in a nonlinear manner and that these



interactions can lead to a strong amplification of the long-term costs due to extreme events.

### 6.1 Modeling endogenous economic dynamics

One major debate in macroeconomics has been on the causes of business cycles and other short-term fluctuations in economic activity. This debate involves two main types of business cycle models: so-called *real business cycle (RBC)* models, in which all fluctuations of macroeconomic quantities are due to exogenous changes in productivity or to other exogenous shocks (e.g., Slutsky, 1927; Frisch, 1933; Kydland and Prescott, 1982; King and Rebelo, 2000) and *endogenous business cycle (EnBC)* models, in which it is primarily internal mechanisms that lead to the more-or-less cyclic fluctuations (e.g., Kalecki, 1937; Harrod, 1939; Samuelson, 1939; Chiarella et al., 2005).

Nobody would claim that exogenous forcing does not play a major role in business cycles. Thus, the strong economic expansion of the late 1990s was clearly driven by the rapid development of new technologies. But denying any role to endogenous fluctuations that are due to unstable and nonlinear feedbacks within the economic system itself seems also rather unrealistic.

The recent “great recession” has raised many questions about the neoclassical tradition that posits perfect markets and rational expectations (e.g., Soros, 2008). Even within this neoclassical framework, though, several authors (e.g., Gale, 1973; Benhabib and Nishimura, 1979; Day, 1982; Grandmont, 1985) proposed models in which endogenous fluctuations arise from savings behavior, wealth effects and interest rate movement or from interactions between overlapping generations and between different sectors. As soon as market frictions and imperfect rationality in expectations or aggregation biases are accounted for, strongly destabilizing processes can be identified in the economic system.

The existence of destabilizing processes has been proposed and their importance noted by numerous authors. Harrod (1939) already stated that the economy was unstable because of the absence of an adjustment mechanism between population growth and labor demand, even though Solow (1956) proposed later the choice of the labor–capital intensity by the producer as the missing adjustment process. Kalecki (1937) and Samuelson (1939) proposed simple business cycle models based on a Keynesian accelerator–multiplier and delayed investments. Later on, other authors (Kaldor, 1940; Hicks, 1950; Goodwin, 1951, 1967) developed business cycle models in which the destabilizing process was still the Keynesian accelerator–multiplier, while the stabilizing processes were financial constraints, distribution of income or the role of the reserve army of labor. In Hahn and Solow (1995, Ch. 6), fluctuations can arise from an imperfect goods market, frictions in the labor market and from the interplay of irreversible investment and monopolistic competition.

Exploration of EnBC theory was quite active in the middle of the 20th century but much less so over the last 30 yr. Still, numerous authors (e.g., Hillinger, 1992; Jarsulic, 1993; Flaschel et al., 1997; Nikaido, 1996; Chiarella and Flaschel, 2000; Chiarella et al., 2005, 2006) have proposed recent EnBC models. The business cycles in these models arise from nonlinear relationships between economic aggregates and the competing instabilities they generate; they are consistent with certain realistic features of actual business cycles. EnBC models have not been able, so far, to reproduce historical data as closely as RBC models do (see, e.g., King and Rebelo, 2000).

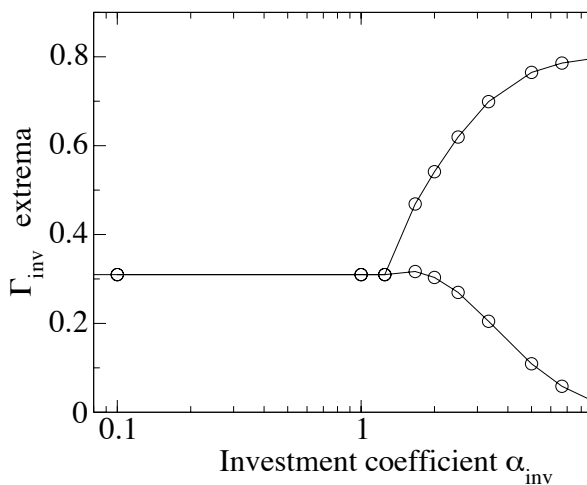
Even so, Chiarella et al. (2006) for instance show that their model is able to reproduce historical records when utilization data are taken as input. It is not surprising, moreover, that models with only a few state variables – typically less than a few dozen – are not able to reproduce the details of historical series that involve processes lying explicitly outside the scope of an economic model (e.g., geopolitical tensions). Furthermore, RBCs have gained from being fitted much more extensively to existing data sets.

Motivated in part by an interest in the role of economic instabilities in impact studies, Hallegatte et al. (2008) formulated a highly idealized Non-Equilibrium Dynamical Model (NEDyM) of an economy with a single goods and a single producer. NEDyM is a neoclassical model with myopic expectations, in which adjustment delays have been introduced in the clearing mechanisms for both the labor and goods market and in the investment response to profitability signals. In NEDyM, the equilibrium is neoclassical in nature, but the stability of this equilibrium is not assumed a priori.

Depending on the model’s parameter values, endogenous dynamics may arise via an exchange of stability between the model’s neoclassical equilibrium and a periodic solution. The business cycles in NEDyM thus originate from the instability of the profit–investment relationship, a relationship that resembles the Keynesian accelerator–multiplier effect. The interplay of three processes limits the amplitude of the model’s endogenous business cycles: (i) the increase of labor costs when the employment rate is high (a reserve army of labor effect); (ii) the inertia of production capacity and the consequent inflation in goods prices when demand increases too rapidly; and (iii) financial constraints on investment.

The model’s main control parameter is the investment flexibility  $\alpha_{\text{inv}}$ , which is inversely proportional to the adjustment time of investment in response to profitability signals. For slow adjustment, the model has a stable equilibrium, which was matched to the economic state of the European Union (EU-15) in 2001 (European Commission, 2002). As the adjustment time decreases, and  $\alpha_{\text{inv}}$  increases, this equilibrium loses its stability through a Hopf bifurcation, in which a periodic solution grows around the destabilized neoclassical equilibrium.

This stable periodic solution – called a “limit cycle” in DDS language – represents the model’s business cycle and it



**Fig. 12.** Bifurcation diagram for the Non-Equilibrium Dynamic Model (NEDyM) of Hallegatte et al. (2008). The abscissa displays the values of the control parameter  $\alpha_{\text{inv}}$ , while the ordinate displays the corresponding values of the investment ratio  $\Gamma_{\text{inv}}$ . The model has a unique, stable equilibrium for low  $\alpha_{\text{inv}}$  values,  $\alpha_{\text{inv}} \leq \alpha_0 = 1.39$ . For higher values, a Hopf bifurcation at  $\alpha_0$  leads to a limit cycle whose minimum and maximum values are shown in the figure. Transition to chaotic, irregular behavior occurs for even higher values (not shown). From Hallegatte et al. (2008).

is characterized by variables that oscillate around their equilibrium values. For instance, the investment ratio  $\Gamma_{\text{inv}}$ , i.e. the ratio of investment to the sum of investment and dividend distribution, oscillates and Fig. 12 shows the extrema of this oscillation as a function of  $\alpha_{\text{inv}}$ . For even larger values, the model exhibits chaotic behavior, with a positive Lyapunov exponent of about  $0.1 \text{ yr}^{-1}$ . In this case, therefore, no economic forecast would be able to provide an accurate and reliable prediction over more than 10 yr.

The model's business cycle has a fairly realistic period of 5–6 yr, as well as a sawtooth shape, with recessions that are considerably shorter than the expansions; see the upper panel of Fig. 13 in the next subsection. The latter feature, in particular, is not present in RBC models of the same degree of simplicity as NEDyM. Due to the present EnBC model's simplicity, though, the amplitude of its business cycle is too large.

## 6.2 Endogenous dynamics and exogenous shocks

It is obvious that historical economic series cannot be explained without taking into account exogenous shocks. Assuming that one part of the dynamics arises from intrinsic factors, it is therefore essential to better understand how the endogenous dynamics interacts with exogenous shocks.

Hallegatte and Ghil (2008), following up on the work described in the previous subsection, focused on the impact of

natural disasters on the economy for two reasons. First, natural disasters represent significant and fairly frequent shocks on many economies and their costs are increasing dramatically. It is important, therefore, to investigate their impacts, in order to better manage disaster aftermaths and inform preparedness and mitigation strategies. Second, natural disasters are interesting “natural experiments”, which provide information about how economies react to unexpected shocks. Our analysis aims, therefore, both at improving our understanding of the consequences of natural disaster and, more broadly, at investigating economic dynamics.

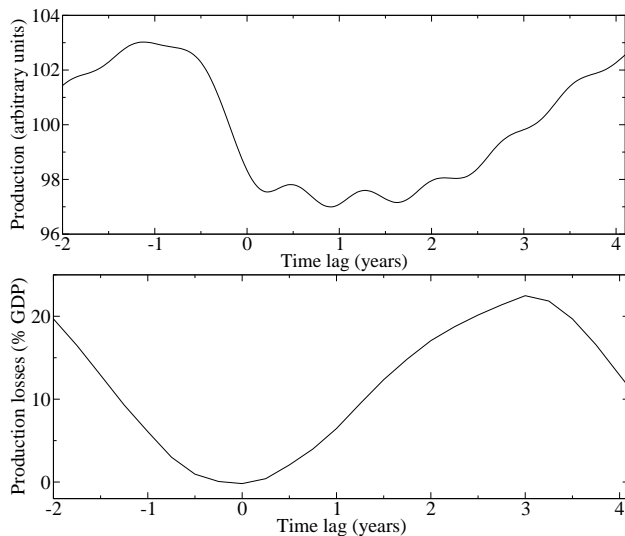
### 6.2.1 Modeling economic effects of natural disasters

Even in the framework of neoclassical, long-term growth models, Hallegatte et al. (2007) showed that natural disasters cannot be modeled without introducing short-term constraints into the pace of reconstruction. Otherwise, economic impact models do not reproduce the observed response to past disasters and reconstruction in the model is carried out in a couple of months, even for large-scale disasters, which is at odds with past experience. Several recent examples of much longer reconstruction times include the 1999 winter storms in Western Europe; the 2002 floods in Central Europe; and the 2004 hurricane season in Florida.

In most historical cases, reconstruction is carried out in 2 to 10 yr; the lower number applies to the 2002 floods in Central Europe, while the higher number is an estimate for Hurricane Katrina or the December 2004 tsunami in South Asia. The reconstruction delays arise from financial constraints, especially but not exclusively in developing countries, and from technical constraints, like the lack of qualified workers and construction equipment; substantial empirical evidence for the existence of these constraints exists. An additional effect of the latter is the so-called demand surge, i.e. the inflation in the price of reconstruction-related goods and services in disaster aftermaths, as shown by Benson and Clay (2004) or Risk Management Solutions (2005), among others.

These constraints can increase dramatically the cost of a single disaster by extending the duration of the reconstruction period. Basically, if a disaster destroys a plant, which can be instantaneously rebuilt, the cost of the disaster is only the replacement cost of the plant. But if the plant is destroyed and can be replaced only one year later, the total cost is its replacement cost plus the value of one year of lost production. For housing, the destruction of a house with a one-year delay before reconstruction has a total cost equal to the replacement cost of the house plus the value attributed to inhabiting the house during one year.

The value of such production losses, in a broad sense, can be quite high in certain sectors, especially when basic needs – such as housing, health or employment – are at stake. Applied to the whole economic system, this additional cost can be very substantial for large-scale disasters. For instance, if we assume that Katrina destroyed about \$100 billion of



**Fig. 13.** The effect of one natural disaster on the endogenous business cycle (EnBC) of the NEDyM model in Fig. 12. Top panel: the business cycle in terms of annual growth rate (in arbitrary units), plotted as a function of the time lag with respect to the cycle minimum; bottom panel: total GDP losses due to one disaster, as a function of when they occur, measured as in the top panel. A disaster occurring at the cycle minimum causes a limited production loss, while a disaster occurring during the expansion lead to a much larger production loss. After Hallegatte and Ghil (2008).

productive and housing capital, that this capital will be replaced over 5 to 10 yr, and we use a mean capital productivity of 23 %, we can assess Katrina-related production losses to lie between \$58 and \$117 billion. Hence production losses will increase the cost of this event by 58 % to 117 %.

Hallegatte et al. (2007) have justified using the mean capital productivity of 23 % as follows. Since one is considering the destruction of an undetermined portion of capital and not the addition or removal of a chosen unit of capital, it makes sense to consider the mean productivity and not the marginal one. Taking into account the specific role of infrastructures in the economy, and the positive returns they often involve, even this assumption is rather optimistic. Assuming a Cobb-Douglas production function with a distribution of 30 %-to-70 % between capital income and labor income yields a mean productivity of capital  $Y/K$  of  $1/0.30$  times the marginal productivity of capital  $dY/dK$ . Since the marginal productivity of capital in the US is approximately 7 %, it follows that the mean productivity of capital can, therefore, be estimated at about 23 %.

The fact that the cost of a natural disaster depends on the duration of the reconstruction period is important because this duration depends in turn on the capacity of the economy to mobilize resources in carrying out the reconstruction. The problem is that the poorest countries, which have lesser re-

sources, cannot reconstruct as rapidly and efficiently as rich countries. This fact may help explain the existence of poverty traps in which some developing countries seem to be stuck.

Because of this low reconstruction capacity and a large exposure to dangerous events – such as tropical cyclones, monsoon floods or persistent droughts – a series of disasters can prevent such countries from accumulating infrastructure and capital, and, therefore, from developing their economy and improving their capacity to reconstruct after a disaster. This deleterious feedback effect may maintain some countries at a reduced-development stage in the long run. The effect of repeated disasters in some Central American countries in the late 1990s and early 2000s (e.g., Guatemala and Honduras) illustrates the consequences of such a series of catastrophes on economic development.

In fact, these days, development agencies consider the management of catastrophic risks as an important component of development policies. This is but one example of the way in which extreme events – in the same domain or in different ones, e. g. combining exposure to climatic disasters with the effects of recurrent warfare – can lead to much greater damage than the mere superposition of separate extremes.

## 6.2.2 Interaction between endogenous dynamics and exogenous shocks

We are now ready to combine the results of the economic modeling in Sect. 6.1 with those of modeling the economic effect of natural disasters in Sect. 6.2.1. To evaluate how the cost of a disaster depends on the pre-existing economic situation, we apply the same loss of productive capital at different points in time along NEDyM's business cycle, and we assess the total GDP losses, summed up over time and without any discounting.

Figure 13 shows in the top panel the model's business cycle, with respect to the time lag relative to the end of the recession. The bottom panel shows the overall cost of a major disaster that causes destruction amounting to 3 % of GDP, with respect to the time the disaster occurs, also expressed as a time lag relative to the end of the recession. We find that the total GDP losses caused by such a disaster depend strongly on the phase of the business cycle in which the disaster occurs: the cost is minimal if the event occurs at the end of the recession and it is maximal if the disaster occurs in the middle of the expansion phase, when the growth rate is largest.

The presence of endogenous dynamics seems, therefore, to induce a “vulnerability paradox”:

- A disaster that occurs when the economy is depressed results in lower damages, thanks to the recovery effect of the reconstruction, which activates resources that have been going unused. In this situation, since employment is low, additional hiring for reconstruction purposes will not drive wages upward too rapidly. Also, the stock of goods is, during a recession, larger than

its equilibrium value. A disruption of production can, therefore, be damped by inventories. Finally, the investment ratio is low and the financial constraint on production is not active: the producer can easily increase investment. In this case, the economic response brings the disaster cost back to zero, according to the EnBC model.

- A disaster that occurs during a high-growth period results in larger damages, as it enhances pre-existing disequilibria. First, inventories are below their equilibrium value during a high-growth phase, so they cannot compensate for the reduced production. Second, employment is at a very high level and hiring more employees induces wage inflation. Finally, because the investment ratio is high, the producer lacks financial resources to increase requisite investments. In fact, the maximum production loss given by the NEDyM model reaches 20 % of GDP. This loss is unrealistically high, a fact that can be related to the model's business cycle having too large an amplitude (see Sect. 6.1).

The crudeness of the model and its imperfect ability to reproduce realistic business cycle make it impossible to evaluate, let alone predict, more precisely the importance of the mechanisms involved in this vulnerability paradox. Tentative confirmation is provided, though, by the relatively low cost of production losses after major disasters that occurred during recessions, in agreement with the suggestion of Benson and Clay (2004).

As an example, the Marmara earthquake in 1999 caused destructions that amounted to between 1.5 and 3 % of Turkey's GDP; the latter value is the one we used in generating Fig. 13. The cost of this earthquake in terms of production loss, however, is believed to have been kept at a relatively low level by the fact that the country was experiencing a strong recession of  $-7\%$  of GDP in the year before the disaster (World Bank, 1999).

The intriguing theoretical and potentially important practical aspects of this vulnerability paradox call, therewith, for more research in the modeling of business cycle and the understanding of the consequences of exogenous shocks impacting out-of-equilibrium economies (Dumas et al., 2011).

### 6.2.3 Distribution of disasters and economic dynamics

The ultimate cost of a single supply-side shock thus depends on the phase of the business cycle. It is likely, therewith, that the cost of a given distribution in time of natural disasters that affect an economy depends on the characteristics of this economy's dynamics. For instance, two economies that exhibit fluctuations of different amplitudes would probably cope differently with the same distribution of natural disasters.

To investigate this issue, Hallegatte and Ghil (2008) used NEDyM to consider different economies of the same size and

at the same development stage, which differ only by their investment flexibility  $\alpha_{\text{inv}}$ . As shown in Fig. 12 here, if  $\alpha_{\text{inv}}$  is lower than a fixed bifurcation value, the model possesses a single stable equilibrium. Above this value, the model solutions exhibit oscillations of increasing amplitude, as the investment flexibility increases.

Hallegatte et al. (2007) calibrated their natural-disaster distribution on the disasters that impacted the European Union in the previous 30 yr. This distribution of events was used in Hallegatte et al. (2007) to assess the mean production losses due to natural disasters for a stable economy at equilibrium, while Hallegatte and Ghil (2008) did so in NEDyM.

The latter results are reproduced in Table 4 and highlight the very substantial, but complex role of investment flexibility. If this flexibility is null (first row of the table) or very low (not shown), the economy is incapable of responding to the natural disasters through investment increases aimed at reconstruction. Total production losses, therefore, are very large, amounting to 0.15 % of GDP when  $\alpha_{\text{inv}} = 0$ . Such an economy behaves according to a pure Solow growth model, where the savings, and therefore the investment, ratio is constant.

When investment can respond to profitability signals without destabilizing the economy, i.e. when  $\alpha_{\text{inv}}$  is non-null but still lower than the bifurcation value, the economy has a new degree of freedom to improve its situation and respond to productive-capital shocks. Such an economy is much more resilient to disasters, because it can adjust its investment level in the disaster's aftermath: for  $\alpha_{\text{inv}} = 1.0$  (second row of the table), GDP losses are as low as 0.01 %, i.e. a decrease by a factor of 15 with respect to a constant investment ratio, thanks to the added investment flexibility.

If investment flexibility is larger than the bifurcation value (third row of the table), the economy undergoes an endogenous business cycle and, along this cycle, it crosses phases of high vulnerability, as shown in Sect. 6.2.2. Production losses due to disasters that occur during the model's expansion phases are thus amplified, while they are reduced when the shocks occur during the recession phases. On average, though, (i) expansions last much longer than recessions, in both the NEDyM model and in reality; and (ii) amplification effects are larger than damping effects. It follows that the net effect of the cycle is strongly negative, with an average production loss of 0.12 % of GDP.

The results in this section suggest the existence of an optimal investment flexibility; such a flexibility will allow the economy to react in an efficient manner to exogenous shocks, without provoking endogenous fluctuations that would make it too vulnerable to such shocks. According to the NEDyM model, therefore, stabilization policies may not only help prevent recessions from being too strong and costly; they may also help control expansion phases, and thus prevent the economy from becoming too vulnerable to unexpected shocks, like natural disasters or other supply-side shocks. Examples of the latter are energy-price shocks, like the oil

shock of the 1970s, and production bottlenecks, for instance when electricity production cannot satisfy the demand from a growing industrial sector.

These results – while still preliminary – support the idea that assessing extreme event consequences with long-term models that disregard short-term fluctuations and business cycles can be misleading. Indeed, our results suggest that nonlinear interactions between endogenous dynamics and exogenous shocks play a crucial role in macroeconomic dynamics and may amplify significantly the long-term cost of natural disasters and extreme events.

## 7 Prediction of extreme events

Prediction tries to address a fundamental need of the individual and the species in adapting to its environment. The crucial difficulty lies not so much in finding a method to predict, but in finding ways to trust such predictions, all the way from Nostradamus to numerical weather prediction. Predicting extreme events poses a particularly difficult quandary, because of their two-edged threat: small number and large impact.

In this section, we will distinguish between the prediction of real-valued functions of continuous time  $X(t)$ ,  $t \in \mathbb{R}$ , and that of so-called point processes  $X(t_i)$  – for which  $X \neq 0$  only at discrete times  $t_i$  that are irregularly distributed; these will be addressed in Sects. 7.1 and 7.2, respectively. As usual, the decision to model a particular phenomenon of interest in one way or the other is a matter of convenience. Still, it is customary to use the former approach for phenomena that can be well approximated by the solutions of differential equations, the latter for phenomena that are less easily so modeled. Examples of the former are large-scale average temperatures in meteorology or oceanography, while the latter include earthquakes, floods or riots.

### 7.1 Prediction of oscillatory phenomena

In this subsection, we focus on phenomena that exhibit a significant oscillatory component: as hinted already in the overall introduction to this section, repetition increases understanding and hence confidence in a prediction method that is closely connected with such understanding. In Sect. 2.2, we reviewed a number of methods for the study of such phenomena.

Among these methods, singular spectrum analysis (SSA) and the maximum entropy method (MEM) have been combined to predict a variety of phenomena in meteorology, oceanography and climate dynamics (Ghil et al., 2002, and references therein). First, the “noise” is filtered out by projecting the time series onto a subset of leading EOFs obtained by SSA (Sect. 2.2.1); the selected subset should include statistically significant, oscillatory modes. Experience shows that this approach works best when the partial variance asso-

ciated with the pairs of RCs that capture these modes is large (Ghil, 1997).

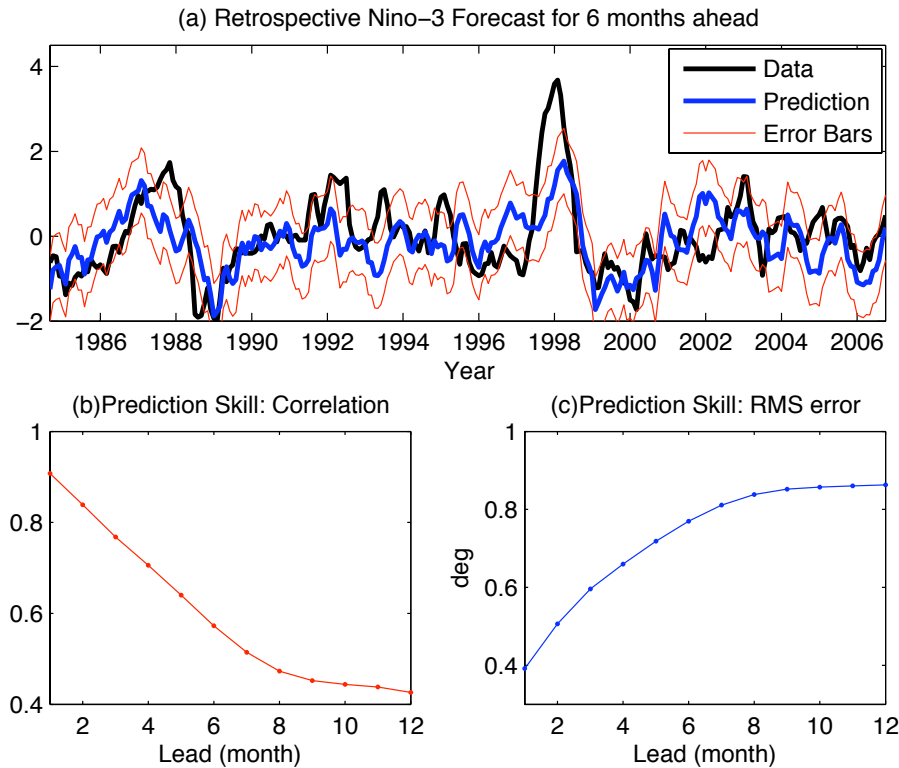
The prefiltered RCs are then extrapolated by least-square fitting to an autoregressive model  $AR[p]$ , whose coefficients give the MEM spectrum of the remaining “signal”. Finally, the extended RCs are used in the SSA reconstruction process to produce the forecast values. The reason why this approach – via SSA prefiltering, AR extrapolation of the RCs, and SSA reconstruction – works better than the customary AR-based prediction is explained by the fact that the individual RCs are narrow-band signals, unlike the original, noisy time series  $X(t)$  (Penland et al., 1991; Keppenne and Ghil, 1993). In fact, the optimal order  $p$  obtained for the individual RCs is considerably lower than the one given by the standard Akaike information criterion (AIC) or similar ones.

The SSA-MEM prediction example chosen here relies on routine prediction of SST anomalies, averaged over the so-called Niño-3 area ( $5^\circ \text{S}$ – $5^\circ \text{N}$ ,  $150^\circ \text{W}$ – $90^\circ \text{W}$ ), and of the Southern Oscillation Index (SOI), carried out by a research group at UCLA since 1992. The SOI is the suitably normalized sea surface pressure difference between Tahiti and Darwin, and it is very well anti-correlated with the Niño-3 index; see, for instance, Saunders and Ghil (2001) and references therein. Thus a strong warm event in the eastern Tropical Pacific (i.e., a strong El Niño) is always associated with a highly positive Niño-3 index and a highly negative SOI, while a strong cold event there (La Niña) is always manifested by the opposite signs of these two indices.

As shown by Ghil and Jiang (1998), these predictions are quite competitive with, or even better than other statistical methods in operational use, in the standard range of 6–12 months ahead. The same is true when comparing them to methods based on the use of deterministic models, including highly sophisticated ones, up to and including coupled GCMs. This competitiveness is due to the relatively large partial variance associated with the leading SSA pairs that capture the quasi-quadrennial (QQ) and the quasi-biennial (QB) modes of ENSO variability (Keppenne and Ghil, 1992a,b).

Figure 14 shows the SSA-MEM method’s Niño-3 SSTA retrospective forecasts for a lead time of 6 months, from January 1984 to February 2006. The forecast for each point (blue curve) utilizes only the appropriate part of the record, namely the one that precedes the initial forecast time: there is no “look ahead” involved in these retrospective forecasts. The prediction skill is uneven over the verification period, and is best during the 1984–1996 interval (Ghil and Jiang, 1998): while the huge 1997–1998 El-Niño event was forecast as early as December 1996 by this SSA-MEM method, the amplitude of the event was not. In particular, this method – like most others – does not capture the skewness of the records of SOI or Niño-3 index, with stronger but fewer warm than cold events.

A different approach is required to capture such non-Gaussian features of the ENSO records, and thus to achieve



**Fig. 14.** ENSO forecasts by the SSA-MTM method. **(a)** Retrospective forecasts of the Niño-3 SST anomalies for a lead time of 6 months. The forecasts (blue curve) for each point in the interval January 1984–February 2006 utilize only the appropriate part of the record (black curve) that precedes the initial forecast time; data from the International Research Institute (IRI) for Climate and Society, <http://iridl.ldeo.columbia.edu/SOURCES/KAPLAN/EXTENDED/>. An SSA window of  $M = 100$  months and 20 leading modes have been used. **(b, c)** Prediction skill for validation times of 1–12 months: **(b)** anomaly correlation and **(c)** root-mean-square (RMS) error, respectively. The one-standard deviation confidence levels (red curve) in panel **(a)** correspond to the RMS error for a lead time of 6 months given in panel **(c)**.

better amplitude, as well as phase prediction. This approach is based on constructing an Empirical Model Reduction (EMR) model of ENSO (Kondrashov et al., 2005b), which combines a nonlinear deterministic with a linear, but state-dependent stochastic component; the latter is often referred to as multiplicative noise. This EMR ENSO model has also been quite competitive in real-time ENSO forecasts.

## 7.2 Prediction of point processes

In this subsection, we treat the prediction of phenomena that can be more adequately modeled as point processes. A radically different approach to the analysis and prediction of time series is in order for such processes: here the series is assumed to be the result of individual events, rather than merely a discretization, due to sampling, of a continuous process (e.g., Brémaud, 1981; Brillinger, 1981; Brillinger et al., 2002; Guttorp et al., 2002). Obvious examples are earthquakes, hurricanes, floods, landslides (Rossi et al., 2010; Witt et al., 2010) and riots. Such processes have

been discussed theoretically in Sect. 3.2.1 and the associated Appendix B, and applied to the downscaling problem in Sect. 3.5.2.

A point process has two aspects: the counting process, which follows the number of events (of equal or different sizes) in a fixed time interval, and the interval process, which deals with the length of the intervals between events. The counterpart of the Central Limit Theorem for point processes states that the sum of an increasing number of mutually independent, ergodic point processes converges to a Poisson process.

Not surprisingly, the assumptions of this theorem, too, are violated by complex systems. In the case of earthquakes, for instance, this violation is associated with deviations from the Gutenberg-Richter power law for the size distribution of earthquakes. V. I. Keilis-Borok and his colleagues (Keilis-Borok et al., 1980, 1988; Keilis-Borok, 1996; Keilis-Borok and Soloviev, 2003) have applied a deterministic pattern-recognition approach, based on the mathematical work of I. M. Gelfand, to the point-process realizations given by

earthquake catalogues (Gelfand et al., 1976). This group has recently extended their predictions from the intermediate-range prediction of earthquakes (Keilis-Borok et al., 1988; Keilis-Borok, 1996) to short-range predictions, on the one hand, and to the prediction of various socio-economic crises – such as US recessions or surges of unemployment and of homicides – on the other.

Keeping in mind the dynamics of a complex system, such as Earth's lithosphere, we formulate the problem of prediction of extreme events generated by the system as follows. Let  $t$  be the current moment of time. The problem is *to decide whether an extreme event* – i.e. an event that exceeds a certain preset threshold – *will or will not occur during a subsequent time interval*  $(t, t + \Delta)$ , using the information on the behavior of the system prior to  $t$ . In other words, we have to reach a yes/no decision, rather than extrapolating a full set of values of the process over the interval  $(t, t + \Delta)$ , as in the previous subsection.

The information on the behavior of the system is extracted from *raw data* that include observable background activity, the so-called “static,” as well as from appropriate external factors that affect the system. Numerous small earthquakes, small fluctuations of macroeconomics indicators or varying rates of occurrence of misdemeanors are typical examples of such static that characterizes the systems of interest.

*Pattern recognition* has proved to be a natural, efficient tool in addressing the kind of problems stated above. Specifically, the methodology of pattern recognition of infrequent events developed by the school of I. M. Gelfand did demonstrate useful levels of skill in several studies of rare phenomena of complex origin, whether of physical (Gelfand et al., 1976; Kossobokov and Shebalin, 2003) or socio-economic (Lichtman and Keilis-Borok, 1989) origin. The M8 algorithm for earthquake prediction is reviewed in Appendix C and applied in Sect. 7.2.1 to predicting earthquakes in Romania's Vrancea region; Zaliapin et al. (2003b) have also studied an approach to optimizing such prediction algorithms for premonitory seismic patterns (PSPs). The overall methodology for predicting such phenomena is presented in Appendix D and is applied in Sect. 7.2.2 to socio-economic phenomena.

### 7.2.1 Earthquake prediction for the Vrancea region

We consider the lithosphere as a complex system where strong earthquakes, with magnitude  $M \geq M_0$ , are the extreme events of interest. The problem is to decide on the basis of the behavior of the seismic activity prior to  $t$ , whether a strong earthquake will or will not occur in a specified region during a subsequent interval  $(t, t + \Delta)$ .

We present here the adaptation of the intermediate-range earthquake prediction algorithm M8 of Keilis-Borok and Kossobokov (1990) for application in the Vrancea region. This algorithm was documented in detail by Kossobokov (1997) and has been tested in real-time applications on both

global and regional scales (Kossobokov et al., 1999a,b; Kossobokov and Shebalin, 2003; Kossobokov, 2006). The M8 algorithm is summarized here in Appendix C. The Vrancea region in the elbow of the Carpathian mountain chain of Romania produces some of the most violent earthquakes in Europe, and thus it was selected in the E2-C2 project for further tests of the pattern-recognition methodology of this group.

A data set on seismic activity in the Vrancea region was first compiled from several sources. This data set is intended to be used for the prediction of strong and possibly moderate earthquakes in the region. Two catalogues have been used for the time interval 1900–2005: (i) the Global Hypocenter Data Base catalogue, in the National Earthquake Information Center (NEIC) archive of the US Geological Survey; and (ii) a local Vrancea seismic catalogue. The local catalogue is complete for magnitudes  $M \geq 3.0$  since 1962 and it is complete for  $M \geq 2.5$  since 1980. The data set so obtained is being continued by the RomPlus earthquake catalogue compiled at the National Institute of Earth Physics (Magurele, Romania).

Three values have been chosen for the magnitude threshold of strong earthquakes:  $M_0 = 6.0, 6.5$ , and  $7.0$ . We have specified  $\Delta M = 0.5$  (see Appendix C) for all three values of the magnitude threshold.

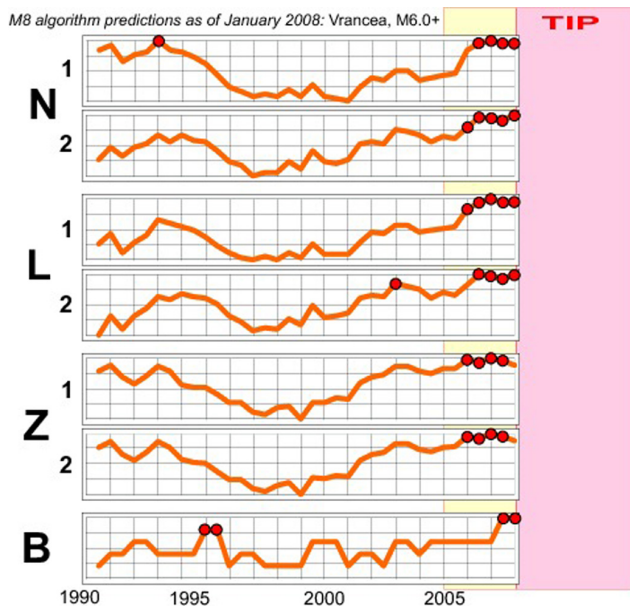
The retrospective simulation of seismic monitoring by means of the M8 algorithm is encouraging: the methodology allows one to identify, though retrospectively, three out of the last four earthquakes with magnitude 6.0 or larger that occurred in the region. A real-time prediction experiment in the Vrancea region has been launched starting in January 2007. Note that the M8 algorithm focuses specifically on earthquakes in a range  $M_0 \leq M < M_0 + \Delta M$  with given  $\Delta M$ , rather than on earthquakes with any magnitude  $M \geq M_0$ .

The semi-annual updates of the RomPlus catalogue on January 2007, July 2007, and so on, up till and including July 2010, confirm that the region has entered a time of increased probability (TIP) for earthquakes of magnitude 6.5+ or 7.0+ in the middle of 2006 (not shown); this TIP is to expire by June 2011. Figure 15 presents the 2008 Update A (on January 2008) of the M8 test: at that time, the Vrancea region entered a TIP for earthquakes of magnitude 6.0+.

### 7.2.2 Prediction of extreme events in socio-economic systems

The key idea in applying pattern recognition algorithms to socio-economic systems is that in complex systems where many agents and actions contribute to generate small, moderate and large effects, premonitory patterns that are qualitatively similar to the PSPs used in the previous subsection might be capable of raising alarms. This hypothesis is sometimes called the “broken-window effect” in criminology: many broken windows in a neighborhood can foretell acts of major vandalism, many of these lead to robberies, and many of those lead eventually to homicides, while the





**Fig. 15.** Application of the standard version of the M8 algorithm (Kossobokov, 1997) to the RomPlus catalogue for magnitude  $M = 6.0+$ . Plotted are normalized graphs of seven measures of seismic activity depicting the dynamical state of the Vrancea region in terms of seismic rate  $N$  (two measures,  $N_1$  and  $N_2$ ), rate differential  $L$  (likewise two measures,  $L_1$  and  $L_2$ ), dimensionless concentration  $Z$  (also  $Z_1$  and  $Z_2$ ), and a characteristic measure of earthquake clustering  $B$ . The extreme values of these quantities in the extended phase space of the system are outlined by red circles. The M8 algorithm raises an alarm – called *Time of Increased Probability* (TIP) – in a given geographic region, when the state vector  $(N_1, N_2, L_1, L_2, Z_1, Z_2, B)$  enters a pre-defined subset of this phase space: previous pattern recognition has determined that, in this subset, the probability of an above-threshold event increases to a level that justifies the alarm; see Sect. 7.2.1 and Appendix C here, as well as Kossobokov (2006) and Zaliapin et al. (2003b).

multiplication of the latter can even lead to riots (Kelling and Wilson, 1982; Bui Trong, 2003; Keizer et al., 2008). Such a chain of events resembles phenomenologically the clustering of small, increasingly large, and finally major earthquakes. It is clear that this approach does not address in any way the issues of causality, and hence of remediation, only those of prediction and law enforcement.

The pattern-recognition approach was thus extended, based on the above reasoning, to the start of economic recessions (Keilis-Borok et al., 2000), to episodes of sharp increase in the unemployment rate, called fast acceleration of unemployment (FAU; Keilis-Borok et al., 2005), and to homicide surges in megacities (Keilis-Borok et al., 2005). Based on these applications to complex economic and social systems, we try to formulate here a universal algorithm that is applied to monthly series of several relevant indices of system activity, including the appropriate definition of pa-

rameter values for the prediction of the extreme events of interest in the given system.

Examples of relevant indices of system activity are the industrial production total (IPT) and the index of help wanted advertising (LHELL) for the US economy or the grand total of all actual offences (GTAAO) for the socio-economic urban system of Los Angeles and New York City. In particular, IPT is an index of real (constant dollars, dimensionless) output for the entire economy of a country or region that represents mainly manufacturing; this index has been used, rather than GDP, because of the difficulties in measuring, on a monthly basis, the quantity of output in services, where the latter includes travel agencies, banking, etc. The index LHELL measures the amount of job advertising (column-inches) in a number of major newspapers, while GTAAO below is the number of offences registered by the Los Angeles Police Department or the New York Police Department during a month. The generalized prediction algorithm is presented in Appendix D.

### 7.2.3 Applications to economic recessions, unemployment and homicide surges

*Economic recessions in the United States.* In this application we take the activity index  $f$  of Appendix D to equal the IPT for the time interval January 1959–May 2010. The index is available on the web site of the Economic Research Section of the Federal Reserve Bank of St. Louis (FRBSL, <http://research.stlouisfed.org/fred2/series/INDPRO/>).

The seven parameters  $(s, u, \mu, v, L_1, L_2, \tau)$  of the prediction algorithm are defined in Appendix D and in Figs. 19 and 20 there. Their values are taken here, after extensive experimentation, as follows:

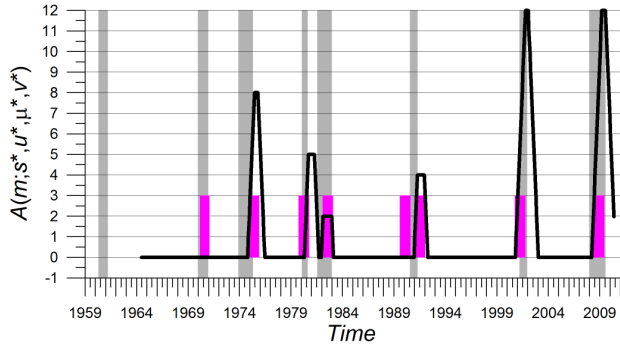
$$s^* = 6, u^* = 48, \mu^* = 0.35, v^* = 12, L_1^* = 0.0, L_2^* = 1.0, \tau^* = 12,$$

i.e.,  $(s^*, u^*, \mu^*, v^*, L_1^*, L_2^*, \tau^*) = (6, 48, 0.35, 12, 0.0, 1.0, 12)$ .

The resulting graph of  $A(m; s^*, u^*, \mu^*, v^*)$  is given in Fig. 16 (solid black line), along with the time intervals of economic recession (grey vertical strips) in the United States according to the National Bureau of Economic Research (NBER) and the alarms declared by the algorithm (magenta boxes). The 1960 recession falls outside the range of the definition of our activity-level function  $A(m; s^*, u^*, \mu^*, v^*)$  – and could thus not contribute to testing the algorithm – since  $A(m)$  can only be defined starting at  $m = m_0 + s + u$  (see Appendix D and Fig. 19c), while  $m_0$  = July 1964.

The 1980, 1991, and 2001 recessions confirm the algorithm predictions, while the other three recessions – in 1970, 1973, and 1981 – appear before the three alarms. The delay of prediction is small for the recession in 1970 (2 months) but larger for the recessions in 1973 (12 months) and 1981 (6 months). The delay in the alarm that corresponds to the 1981 recession might be due to the rather short time interval since the preceding recession in 1980: there was thus not enough time for the activity function  $A(m; s^*, u^*, \mu^*, v^*)$ ,





**Fig. 16.** The graph of the activity-level function  $A(m; s^*, u^*, \mu^*, v^*)$  for the prediction of US economic recessions. Parameter values are  $(s^*, u^*, \mu^*, v^*) = (6, 48, 0.4, 12)$  and they were calculated from the Federal Reserve Bank of St. Louis (FRBSL) monthly series of industrial production total (IPT; black curve). Official economic recessions in the US are shown as grey strips, and alarms declared by the algorithm as magenta bars of duration  $\tau = 12$  months; see Sect. 7.2.3 and Appendix D for details.

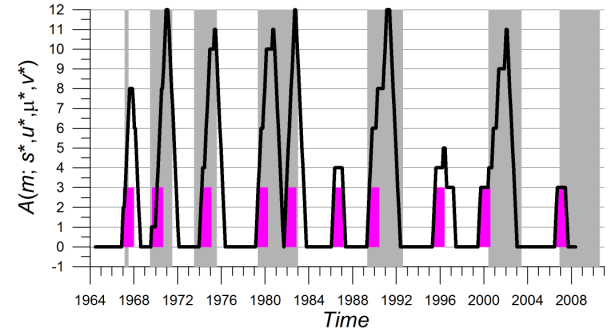
with  $s^* = 6$  months and  $u^* = 48$  months, to detect the second of the two recessions. Note that declaring an alarm 2 months after the start of a recession may, nevertheless, be of interest for potential end users because the NBER announcements of a recession are usually delayed by 5 months or more.

The algorithm declared, in real time, a current alarm for a recession to start in June 2008 (Fig. 16). According to the NBER, the recession started in January 2008, but the announcement was only made in December 2008.

**Fast acceleration of unemployment (FAU).** In this application, we consider the data on monthly unemployment rates for the US civilian labor force in the time interval from January 1959 to July 2010 as given by the US Department of Labor (USDL, <http://stats.bls.gov/>). The index  $f(m)$  is now LHELL, as defined and explained at the end of the previous subsection, and we use the values from January 1959 to May 2008 that are available from the Economic Research Section of the FRBSL, <http://research.stlouisfed.org/fred2/series/INDPRO/>. The values of this index are not available after May 2008.

In the present case, the parameter values  $(s^*, u^*, \mu^*, v^*, L_1^*, L_2^*, \tau^*)$  are the same as above, except that now  $\mu^* = 1$ . The resulting graph of  $A(m; s^*, u^*, \mu^*, v^*)$  is given in Fig. 17 (solid black line), along with the time intervals of FAU (grey vertical strips) and the alarms declared by the algorithm (magenta boxes). Each FAU interval here is determined, following Keilis-Borok et al. (2005), as an interval between the local minimum and the next local maximum of the smoothed monthly unemployment rate.

The FAU intervals that start in 1967, 1979 and 2006 are well predicted by the algorithm. The alarm in 1989 is identified in the same month as the start of the FAU. The other two



**Fig. 17.** Same as Fig. 16, but for the prediction of periods of fast acceleration of unemployment (FAU). Parameter values were calculated from the FRBSL monthly series of the index of help wanted advertising (LHELL); they are the same as for the economic recession prediction, except for  $\mu^* = 1.0$ . Same conventions as in the previous figure.

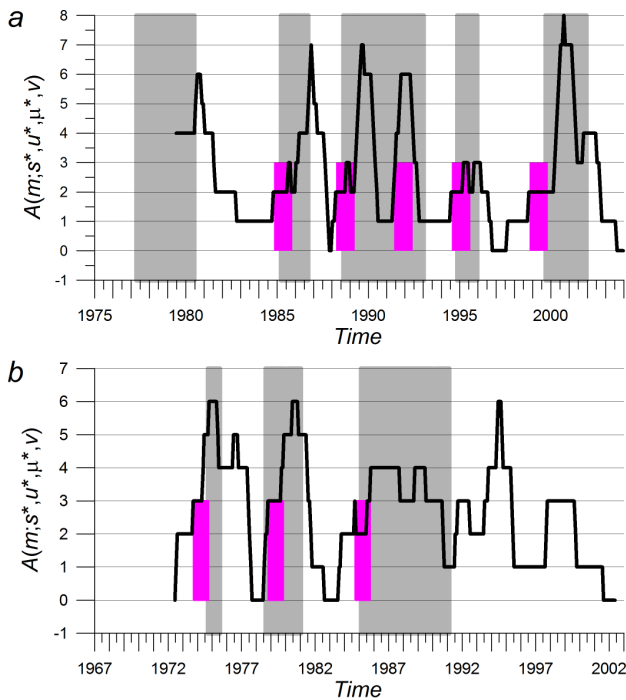
alarms appear after the beginning of the corresponding FAU: specifically, the alarms are delayed by 2 months in 1969 and by 7 months in 1973. The 1981 alarm is an immediate continuation of the longest FAU (1979–1983) that occurred during the time of data availability, while those in 1986 and 1995 are evidently false alarms. The overall statistics of this application suggests that the algorithm does have some predictive capacity, although the two false alarms raise legitimate doubts. At the same time, the last FAU interval, which started in December 2006, has been predicted in advance by means of this algorithm (Fig. 17), as well as by the algorithm already suggested by Keilis-Borok et al. (2005).

**Homicides surges in megacities.** We describe here the application for the City of Los Angeles, with its 3.8 million inhabitants (in 2003), and for New York City, with its 8.1 million inhabitants (likewise in 2003). The index  $f$  in both cases is GTAAO, as defined and explained at the end of Sect. 7.2.2. For the City of Los Angeles, the data used are for January 1974 to December 2003; they are available from the website of the National Archive of Criminal Justice Data (NACJD), <http://www.icpsr.umich.edu/NACJD/>. Based on this dataset, the following parameter values were chosen:

$$(s^*, u^*, \mu^*, v^*, L_1^*, L_2^*, \tau^*) = (6, 48, 184, 12, 1.0, 2.0, 12).$$

The resulting graph of  $A(m; s^*, u^*, \mu^*, v^*)$  is given in Fig. 18a (solid black line) along with the periods of the homicide surge (grey vertical strips) and the alarms declared by the algorithm (magenta boxes). The periods of homicide surge are determined, following Keilis-Borok et al. (2003), as periods of a lasting homicide raise after smoothing seasonal variations.

The homicide surge in 1977 falls outside the range of the  $A(m; s^*, u^*, \mu^*, v^*)$  definition; as explained for the economic recessions, the range in which  $A(m)$  is defined starts only in



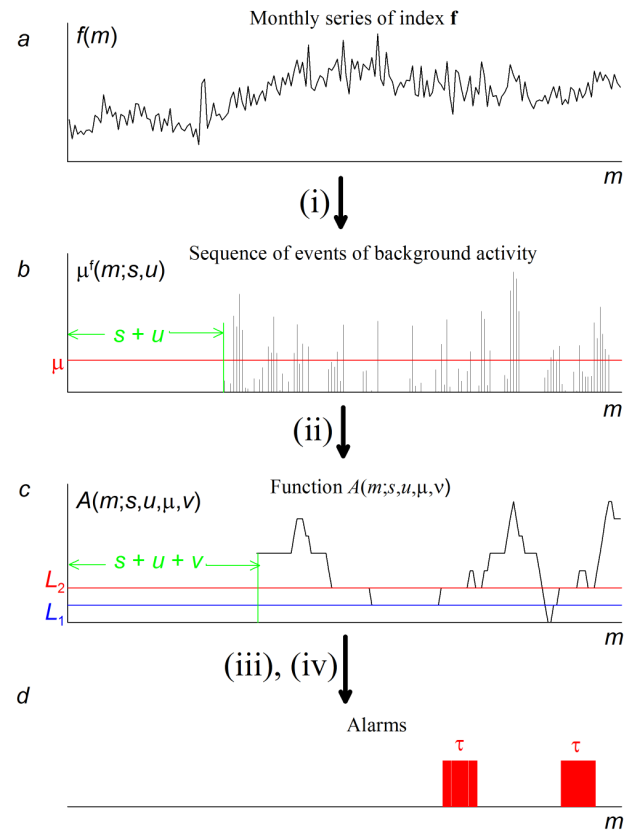
**Fig. 18.** Same as Fig. 16, but for the prediction of homicide surges in (a) the City of Los Angeles, and (b) New York City. In both cases, data on the grand total of all actual offences (GTAO) from the National Archive of Criminal Justice Data (NACJD) were used. The NACJD data on crime statistics in (a) the City of Los Angeles were available for January 1974 to December 2003, and in (b) New York City for January 1967 to May 2002. Based on these data, the parameter values chosen for both megacities were  $(s^*, u^*, v^*) = (6, 48, 12)$ , while  $\mu = 184$  in panel (a) and  $\mu = 1000$  in panel (b). Same conventions as in Figs. 16 and 17.

July 1979. Each of the other four periods of homicide surge is predicted. The alarm in 1991 appears inside the longest homicide surge period for the city, which extended from mid-1988 till early 1993.

The NACJD data set on crime statistics in New York City contains the GTAEO index from January 1967 to May 2002. It clearly makes sense to exclude the victims of September 11, 2001. The resulting parameter values for New York City are

$$(s^*, u^*, \mu^*, v^*, L_1^*, L_2^*, \tau^*) = (6, 48, 1000, 12, 0.0, 3.0, 12).$$

The results for New York City are summarized in Fig. 18b. The two homicide surges in 1974 and 1985 are predicted by the algorithm, while the alarm associated with the remaining one in 1978 is delayed by 4 months.

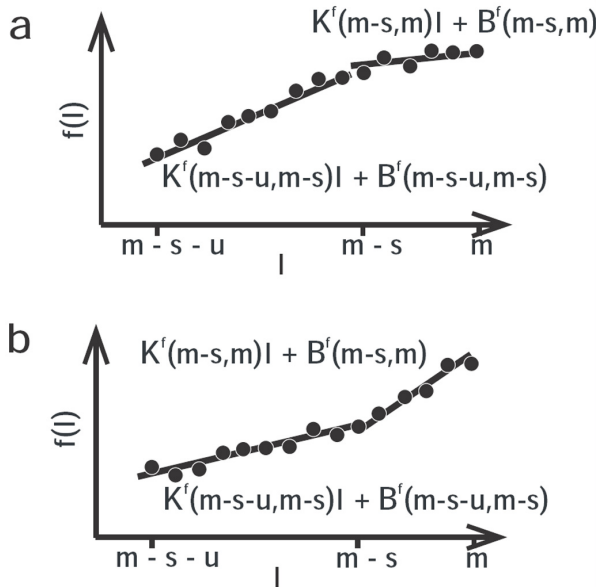


**Fig. 19.** Schematic diagram of the prediction algorithm. The four panels illustrate the steps in the transition from the original time series  $f(m)$  to the sequence of background-activity events of size  $\mu^f(m; s, u)$  to the event-counting, activity-level function  $A(m; s, u, \mu, v)$  and on to the alarms of length  $\tau$  at times of significant change in the activity-level function  $A(m)$ . The parameters  $s, u, \mu$  are introduced in panel (b),  $L_1, L_2$  in panel (c) and  $\tau$  in panel (d). See Appendix D for details.

## 8 Concluding remarks

The results of the investigations reviewed herein have clearly made significant contributions to the various aspects of the overall E2-C2 project and to the study of extreme events in general. Given the length of this review, though, it seems most appropriate to conclude with a number of questions, rather than with a summary and with definitive conclusions, if any.

The key question for the description, understanding and prediction of extreme events can be formulated as a paraphrase to F. Scott Fitzgerald's assertion in "The Great Gatsby": "The rich are like you and me, only richer". Is this true of extreme events, i.e., are they just other events in the life of a given system, only bigger? If so, one can – as one often does – extrapolate from the numerous small ones to the few large ones. This approach allows one to jump from



**Fig. 20.** Definition of changes in background activity for the index  $f(m)$ . The two panels illustrate the case when extreme events are associated either (a) with a decline or (b) with an increase in the general trend  $K^f(m-s, m)$  of  $\{f(l) : m-s \leq l \leq m\}$ ; in either case, the change in trend  $\mu^f(m; s, u)$  is defined as  $\mu^f(m; s, u) = |K^f(m-s, m) - K^f(m-s-u, m-s)|$ . Dots show monthly values  $f(l)$ , while straight lines are the linear least-square regressions of  $f(l)$  on  $l$  in the two adjacent intervals  $(m-s-u, m-s)$  and  $(m-s, m)$ . See Appendix D for details.

the description of the many to the prediction of the few. It is essentially this idea that underlies the use of power laws for many classes of phenomena and the design of skyscrapers that have to withstand the “50-year wind burst” or of bridges that have to survive the “100-year flood”.

The modest opinion of the authors – after dealing with many kinds of extreme events, in the physical as well as the socio-economic realm – is that the yes-or-no answer to the “Great-Gatsby” question is a definite “maybe”, i.e. “it depends”. To conclude with confidence that the large are like the small, only larger, requires a deep understanding of the system. This understanding passes not only through careful description, but also through modeling. Some systems are better known than others, and can be modeled by using fairly sophisticated tools, like sets of differential equations and other modeling frameworks, whether deterministic, stochastic or both. Validating such models on existing data can certainly enhance our confidence in predictions of the large events, along with the smaller ones.

There are maybe just two striking conclusions from the work reviewed herein that we’ll select. The first one is the following: The interaction between the project’s researchers has elucidated a highly intriguing aspect of the description of time series that include extreme events (Sect. 2). This as-

pect has to do with a well-known saying in time series analysis, but in a somewhat different sense than the one in which it is usually taken: “one person’s signal is another person’s noise”. Namely both the study of the continuous background – the absolutely continuous spectral measure, in mathematical parlance – and that of the line spectrum or (more precisely), in practice, of the peaks rising above this background are referred to by the fans of either as spectral analysis. In fact, a given spectral density (for mathematicians) or power spectrum (for engineers and physical scientists) can be, and often is, the sum (or superposition, if you will) of the two kinds of spectra. There is much to learn, and much that is useful for prediction, in both.

The second striking conclusion pertains to the interaction of extrema-generating processes, whether from the same realm or from different ones. We have found a “vulnerability paradox” in the fact that natural catastrophes produce whatever direct damage they do, but that the secondary impact on the economy, via subsequent production loss, is smaller during a recession than during an expansion (Sect. 6). This suggests giving greater attention to the interaction between natural and man-made hazards in a truly coupled-modeling context. This context should take into account internal variability of both the natural and the socio-economic dynamics of interest.

Some further musings are in order. Clearly, the entire field of extreme events and the related risks and hazards is in full swing. As clearly stated in Sect. 1, we were only able to cover a limited subset of all that’s happening and that is interesting, promising or both. In terms of statistics (Sect. 3), there is a lot going on in modeling multivariate and spatially dependent processes, long-range dependence and downscaling. In dynamics, Boolean delay equations (BDEs, Sect. 4.3) can capture extremal properties of both continuous (Sect. 5.2.1) and point processes (Sect. 5.2.2), while interesting ideas are being developed on these properties for maps and other DDS (Sect. 4.2).

Finally, prediction – in the sense of forecasting future events in a given system – is a severe test for any theory of evolutive phenomena. Clearly, all the means at our disposal have to be brought to bear on it, especially when the number and accuracy of the observations leaves much to be desired: advanced statistical methods (Sect. 7.1) and pattern recognition (Sect. 7.2), as well as explicit dynamical models (Sect. 4). A lot was done, but even more is yet to do.

## Appendix A

### Parameter estimation for GEV and GPD distributions

In 1979, Greenwood et al. (1979) and Landwehr et al. (1979) introduced the so-called probability-weighted moments (PWM) approach. This method-of-moments approach has been very popular in hydrology (Hosking, 1985) and

the climate sciences because of its conceptual simplicity, its easy implementation and its good performance for most distributions encountered in the geosciences. Smith (1985) studied and implemented the maximum-likelihood estimation (MLE) method for the GEV density. According to Hosking et al. (1985), the PWM approach is superior to the MLE for small GEV-distributed samples.

Coles and Dixon (1999) argued that the PWM method makes a priori assumptions about the shape parameter that are equivalent to assuming a finite mean for the distribution under investigation. To integrate a similar condition in the MLE approach, these authors proposed a penalized MLE scheme with the constraint  $\xi < 1$ . If this condition is satisfied, then the MLE approach is as competitive as the PWM one, even for small samples. Still, the debate over the advantages and drawbacks of these two estimation methods is not closed and new estimators are still being proposed.

Diebolt et al. (2008) introduced the concept of generalized probability-weighted moments (GPWM) for the GEV. It broadens the domain of validity of the PWM approach, as it allows heavier tails to be fitted. Despite this advantage, the MLE method has kept one strong advantage over the PWM approach, namely its inherent flexibility in a non-stationary context. To complete this appendix, we would like to emphasize that, besides the three aforementioned estimation methods – MLE, PWM and GPWM – there exist other variants and extensions. For example, Hosking (1990) proposed and studied the L-moments, Zhang (2007) derived a new and interesting method-of-moments, and Bayesian estimation has also generated a lot of interest in extreme value analysis (e.g., Lye et al., 1993; Coles, 2001; Cooley et al., 2007; Coles and Powell, 1996).

## Appendix B

### Bivariate point processes

Equation (19) yields

$$P(Z(x) \leq u \text{ and } Z(x+h) \leq u) = \exp[-\theta(h)/u],$$

where

$$\theta(h) = \int \max\{g(s, x), g(s, x+h)\} \delta(ds)$$

is called the *extremal coefficient* and has been used in many dependence studies for bivariate vectors (Fougères, 2004; Ledford and Tawn, 1997). The coefficient  $\theta(h)$  lies in the interval  $1 \leq \theta \leq 2$  and it gives partial information about the dependence structure of the bivariate pair  $(Z(x), Z(x+h))$ . If the two variables of this pair are independent, then  $\theta(h) = 2$ ; at the other end of the spectrum of possibilities, the variables are equal in probability, and  $\theta(h) = 1$ . Hence, the extremal coefficient value  $\theta(h)$  provides some dependence information as a function of the distance between two locations. For

example, the bivariate distribution for the Schlather model,  $P(Z(x) \leq s, Z(x+h) \leq t)$ , corresponds to

$$\exp \left\{ -\frac{1}{2} \left( \frac{1}{t} + \frac{1}{s} \right) \left( 1 + \left[ 1 - 2(\rho(h) + 1) \frac{st}{(s+t)^2} \right]^{1/2} \right) \right\},$$

where  $\rho(h)$  is the covariance function of the underlying Gaussian process. This yields an extremal coefficient of  $\theta(h) = 1 + \{1 - \frac{1}{2}(\rho(h) + 1)\}^{1/2}$ . Naveau et al. (2009) proposed nonparametric techniques to estimate such bivariate structures. As an application, Vannitsem and Naveau (2007) analyzed Belgium precipitation maxima. Again, to work with exceedances – rather than modeling maxima, as done here – requires one to develop another strategy.

## Appendix C

### The M8 earthquake prediction algorithm

This intermediate-term earthquake prediction method was originally designed by the retrospective analysis of the dynamics associated with seismic activity that preceded the great earthquakes – i.e., with magnitude  $M \simeq 8.0$  – worldwide, hence its name. Since the announcement of its prototype in 1984 and the original version in 1986, the M8 algorithm has been tested in retrospective prediction, as well as in ongoing real-time applications; see, for instance, Kossobokov and Shebalin (2003), Kossobokov (2006), and references therein.

Algorithm M8 is based on a simple pattern-recognition scheme of prediction aimed at earthquakes in a magnitude range designated by  $M_0+$ , where  $M_0+ = [M_0, M_0 + \Delta M)$  and  $\Delta M$  is a prescribed constant. Overlapping *circles of investigation* with diameters  $D(M_0)$ , starting with  $D$  from 5 to 10 times larger than those of the target earthquakes, are used to scan the territory under study. Within each circle, one considers the sequence of earthquakes that occur, with aftershocks removed.

Denote this sequence by  $\{(t_i, m_i, h_i, b_i(e)) : i = 1, 2, \dots\}$ , where  $t_i$  are the times of occurrence, with  $t_i \leq t_{i+1}$ ;  $m_i$  is the magnitude;  $h_i$  is focal depth; and  $b_i(e)$  is the number of aftershocks of magnitude  $M_{\text{aft}}$  or greater during the first  $e$  days after  $t_i$ . Several running counts are computed for this sequence in the trailing time window  $(t-s, t)$  and magnitude range  $M \leq m_i < M_0$ . These counts correspond to different measures of seismic flux, its deviations from the long-term trend, and clustering of earthquakes. The averages include:

- $N(t) = N(t|M, s)$ , the number of main shocks of magnitude  $M$  or larger in  $(t-s, t)$ ;
- $L(t) = L(t|M, s, s^*)$ , the deviation of  $N(t)$  from the longer term trend, i.e.,  

$$L(t) = N(t) - N(t-s|M, s^*)(t-s^*)/(t-s-s^*);$$

- $Z(t) = Z(t|M, M_0 - g, s, \alpha, \beta)$ , the linear concentration of main shocks  $\{i\}$  that lie in the magnitude range  $[M, M_0 - g]$  and the time interval  $(t - s, t)$ , while the parameters  $(\alpha, \beta)$  are used in determining the average source diameter for earthquakes in this magnitude range; and
- $B(t) = B(t|M_0 - p, M_0 - q, s, M_{\text{aft}}, e) = \max_i \{b_i\}$ , the maximum number of aftershocks – i.e., a measure of earthquake clustering – with the sequence  $\{i\}$  considered in the trailing time window  $(t - s, t)$  and in the magnitude range  $[M_0 - p, M_0 - q]$ .

Each of the functions  $N$ ,  $L$ , and  $Z$  is calculated twice, with  $M$  providing the long-term average  $\nu$  of the annual number of earthquakes in the sequence, with  $\nu = 20$  and with  $\nu = 10$ , respectively. As a result, the earthquake sequence is given a robust description by seven functions:  $N$ ,  $L$ ,  $Z$  (twice each), and  $B$  (once only). Threshold values are identified for each function from the condition that they exceed the  $Q$ th percentile, i.e., that they exceed  $Q\%$  of the encountered values.

An alarm or *time of increased probability (TIP)* is declared for  $\Delta = 5$  yr when at least six out of the seven functions, including  $B$ , are larger than the selected threshold within a narrow time window  $(t - u, t)$ . To stabilize prediction, this criterion is checked at two consecutive moments, namely at  $t$  and at  $t + 0.5$  yr, and the TIP is declared only after the second test agrees with the first. In the course of a real-time application, the alarm can extend beyond or terminate in less than 5 yr when updating of the catalogue causes changes in the magnitude cutoffs or the percentiles.

The following standard values of the parameters listed above are used in the M8 algorithm:  $D(M_0) = \exp(M_0 - 5.6) + 1$  in degrees of meridian,  $s = 6$  yr,  $s^* = 30$  yr,  $s = 1$  yr,  $g = 0.5$ ,  $p = 2$ ,  $q = 0.2$ , and  $u = 3$  yr, while  $Q = 75\%$  for  $B$  and  $90\%$  for the other six functions. The linear concentration is estimated as the ratio of the average source diameter  $l$  to the average distance  $r$  between sources. Usually,  $l = n^{-1} \sum_i 10^{\beta(M_i - \alpha)}$ , where  $n$  is the number of main shocks in the catalogue  $\{i\}$ , and one takes  $r \sim n^{-1/3}$ , while  $\beta = 0.46$  is chosen to represent the linear dimension of the source, and  $\alpha = 0$  without loss of generality,

Running averages are defined in a robust way, so that a reasonable variation of parameters does not affect predictions. At the same time, the point-process character of seismic data and the strict usage of the preset thresholds result in a certain discreteness of the alarms.

The M8 algorithm uses a fairly traditional description of a rather complex dynamical system, by merely adding dimensionless concentration  $Z$  and a characteristic measure of clustering  $B$  to the more commonly used phase space coordinates of rate  $N$  and rate differential  $L$ . The algorithm defines certain thresholds in these phase space coordinates to capture the proximity of a singularity in the system's dynamics. When a trajectory enters the region defined by these threshold values, the probability of an extreme event increases to

a level sufficient to predict its actual occurrence. The choice of the M8 thresholds focuses on a specific intermediate-term rise of seismic activity.

## Appendix D

### The pattern-based prediction algorithm

We consider monthly time series of indices that describe the aggregated activity of a given complex system. Let  $f(m)$  be a sample series of the index  $f$  and  $W^f(l; q, p)$  be the local linear least-squares regression of  $f(m)$  estimated on an interval  $[q, p]$ :

$$W^f(l; q, p) = K^f(q, p)l + B^f(q, p), \quad (\text{D1})$$

where  $q \leq l \leq p$ ,  $K^f(q, p)$  and  $B^f(q, p)$  are the regression coefficients.

We introduce  $S^f(m; s, u) = K^f(m - s, m) - K^f(m - s - u, m - s)$ , where  $s$  and  $u$  are positive integers. This function reflects the difference between the two trends of  $f(m)$  in the  $s$  months after  $m - s$  and in the  $u$  months before  $m - s$ . The value of  $S^f(m; s, u)$  is referred to month  $m$  so that to avoid using information from the future.

Let us assume that extreme events in the system considered are associated with a decline in the trend of the  $f$ -index. Mark an event that reflects background activity ("broken windows") of magnitude  $\mu^f(m; s, u) = |S^f(m; s, u)|$  as happening at month  $m$  if  $S^f(m; s, u) < 0$ , as shown in Fig. 20a. When, to the contrary, the extreme events are associated with a rise of the trend in  $f$ , mark an event at  $m$  if  $S^f(m; s, u) > 0$ , as shown in Fig. 20b.

Given the values of the parameters  $s$  and  $u$ , the procedure described above transforms a monthly series of  $f(m)$  into a unique sequence of events at  $m_1, m_2, \dots$  with the corresponding magnitudes  $\mu^f(m_1; s, u)$ ,  $\mu^f(m_2; s, u)$ ,  $\dots$ . Note that, if a series of  $f(m)$  is given on the interval  $[m_0, m_T]$ , then the occurrences of events fall on the interval  $[m_0 + s + u, m_T]$ .

Consider the events of background activity determined as above, on the basis of the index  $f(m)$  and parameters  $s$  and  $u$ . Let  $A(m; s, u, \mu, v)$  be the number of events with magnitudes  $\mu^f(m_i; s, u) \geq \mu$  that occurred within the preceding  $v$  months, i.e.,  $m - v + 1 \leq m_i \leq m$ . Evidently, the increase of  $A(m; s, u, \mu, v)$  characterizes a rise of activity. To give a formal definition, we introduce the two thresholds  $L_1 < L_2$  and identify a significant rise in background activity at  $m$ , provided  $A(m - 1; s, u, \mu, v) \leq L_1$  and  $L_2 \leq A(m; s, u, \mu, v)$ .

The following general scheme for developing a prediction algorithm adapted to a given data set  $\{f(m) : m_0 \leq m \leq m_T\}$  appears to be rather natural. The scheme below and Fig. 19 that illustrates it are given on the assumption that the rise of background activity is a precursory pattern for the occurrence of extreme events. The changes to the scheme and figure in the case it is a decrease that is premonitory are obvious.

1. Given  $s$  and  $u$ , transform a monthly series  $f(m)$ , as shown in Fig. 20b, into a sequence of events  $\{\mu^f(m_j; s, u) : 0 \leq j \leq J\}$ , with  $J \leq m_T - m_0$ .
2. Given  $\mu$  and  $v$ , evaluate  $A(m; s, u, \mu, v)$ .
3. Given  $L_1$  and  $L_2$ , identify from  $A(m; s, u, \mu, v)$  the months  $\{m_k : 0 \leq k \leq K\}$  of background activity rise, with  $K \leq J$ .
4. Given  $\tau$ , declare  $\tau$  months of alarm for an extreme event each time the rise of background activity is identified.
5. Iterate the selection of the parameters  $s, u, \mu, v, L_1, L_2, \tau$  to optimize the algorithm's performance and robustness.

The phrase “With three parameters I can fit an elephant, and with four I can have it wag its tail” has been attributed to a number of famous mathematicians and physicists: it is supposed to highlight the problems generated by using too large a number of parameters to fit a given data set. The pattern recognition approach allows considerable freedom in the retrospective development of a prediction algorithm, like the one outlined above. To avoid the overfitting problems highlighted by the “elephant-fitting” quote, such an algorithm has to be insensitive to the variations in its adjustable parameters, choice of premonitory phenomena, definition of alarms, etc. The sensitivity tests have to comprise an exhaustive set of numerical experiments, which represent a major portion of the effort in the algorithm development.

Molchan (1997) introduced a particular type of error diagrams to evaluate earthquake prediction algorithms. The definition of such an error diagram is the following: Consider the outcomes of prediction during a time interval of length  $T$ . During that time interval,  $N$  strong events occurred and  $N_F$  of them were not predicted. The number of declared alarms was  $A$ , with  $A_F$  of them being false alarms, while the total duration of alarms was  $D$ .

Obviously if  $D = T$  there will be no strong events missed, while if  $D = 0$  there will be no false alarms. The error diagram of Molchan (1997) shows the trade-off between the relative duration of alarms  $\tau = D/T$ , the rate of failures to predict  $n = N_F/N$ , and the rate of false alarms  $f = A_F/A$ . In the  $(n, \tau)$ -plane, the straight line  $n + \tau = 1$  corresponds to a random binomial prediction – at each epoch an alarm is declared with probability  $\tau$  and is not declared with probability  $1 - \tau$ . Outcomes have to improve upon this straight line in order for the algorithm to be competitive; see also Zaliapin et al. (2003b) for the use of this diagram in the idealized world of a BDE model that can generate an arbitrarily long catalog of earthquakes. Similar considerations apply when adapting the algorithm above to other complex systems.

## Appendix E

### Table of acronyms.

ACF	Auto-correlation function
AIC	Akaike information criterion
AR	Auto regressive
ARIMA	Autoregressive integrated moving average
ARMA	Autoregressive moving average
BDE	Boolean delay equation
CA	Cellular automaton
DDE	Delay differential equation
dDS	Discrete dynamical system
DDS	Differentiable dynamical system
DFA	Detrended fluctuation analysis
DOF	Degrees of freedom
E2-C2	Extreme Events: Causes and Consequences
EnBC	Endogenous business cycle
ENSO	El Niño–Southern Oscillation
EOF	Empirical orthogonal function
ETCCDI	Expert Team on Climate Change Detection and Indices
EVT	Extreme value theory
FARIMA	Fractional Autoregressive integrated moving average
FAU	Fast acceleration of unemployment
FDE	Functional differential equation
fGn	Fractional Gaussian noise
FRBSL	Federal Reserve Bank of St. Louis
GAM	Generalized additive model
GARCH	Generalized autoregressive conditional heteroskedasticity
GCM	General circulation model
GDP	Gross domestic product
GEV	Generalized Extreme Value (distribution)
GLM	Generalized Linear Model
GPD	Generalized Pareto distribution
GPH	Geweke and Porter-Hudak (estimator)
GPWM	Generalized probability-weighted moments
GTAAO	Grand total of all actual offences
IPT	Industrial production total
i.i.d.	independent and identically distributed (random variables)
LAM	Limited-area model
LHELL	Index of help wanted advertising
LRD	Long-range dependence
MLE	Maximum likelihood estimation
MTM	Multi-taper method
NACJD	National Archive of Criminal Justice Data
NBER	National Bureau of Economic Research
NEIC	National Earthquake Information Center
NEDyM	Non-Equilibrium Dynamical model
ODE	Ordinary differential equation
OΔE	Ordinary difference equation
PC	Principal component
PDE	Partial differential equation
PΔE	Partial difference equation
PSP	Premonitory seismic pattern
POT	Peak over threshold
PWM	Probability-weighted moments
QBO	Quasi-biennial oscillation
QOQ	Quasi-quadrennial oscillation
RBC	Real business cycle
RC	Reconstructed component
R/S	Rescaled-range statistic
SDM	Statistical dynamical model
SOI	Southern Oscillation Index
SRD	Short-range dependence
SSA	Singular spectrum analysis
SST	Sea surface temperature
SWG	Stochastic weather generator
TIP	Time of increased probability
VGAM	Vector Generalized Additive Model
VGLM	Vector Generalized Linear Model



**Acknowledgements.** It is a pleasure to thank all our collaborators in the E2-C2 project and all the participants in the project's several open conferences, special sessions of professional society meetings (American Geophysical Union and European Geosciences Union), and smaller workshops. R. Blender and two anonymous reviewers read carefully the paper's original version and provided useful suggestions, while V. Lucarini, S. Vaienti and R. Vitolo suggested improvements to Sect. 4.2. M.-C. Lanceau provided precious help in editing the merged list of references. This review paper reflects work carried out under the European Commission's NEST project "Extreme Events: Causes and Consequences (E2-C2)."

Edited by: J. Kurths

Reviewed by: R. Blender and two anonymous referees



The publication of this article is financed by CNRS-INSU.

## References

- Abaimov, S. G., Turcotte, D. L., Shcherbakov, R., and Rundle, J. B.: Recurrence and interoccurrence behavior of self-organized complex phenomena, *Nonlin. Processes Geophys.*, 14, 455–464, doi:10.5194/npg-14-455-2007, 2007.
- Abarbanel, H. D. I.: *Analysis of Observed Chaotic Data*, Springer-Verlag, New York, 272 pp., 1996.
- Albeverio, S., Jentsch, V., and Kantz, H.: *Extreme Events in Nature and Society*, Birkhäuser, Basel/Boston, 352 pp., 2005.
- Allègre, C. J., Lemouel, J. L., and Provost, A.: Scaling rules in rock fracture and possible implications for earthquake prediction, *Nature*, 297, 47–49, 1982.
- Allen, M.: *Interactions Between the Atmosphere and Oceans on Time Scales of Weeks to Years*, Ph.D. thesis, St. John's College, Oxford, 352 pp., 1992.
- Allen, M. and Smith, L.: Monte Carlo SSA: Detecting irregular oscillations in the presence of colored noise, *J. Climate*, 9, 3373–3404, 1996.
- Altmann, E. G. and Kantz, H.: Recurrence time analysis, long-term correlations, and extreme events, *Phys. Rev. E*, 71, 2005.
- Anh, V., Yu, Z.-G., and Wanliss, J. A.: Analysis of global geomagnetic variability, *Nonlin. Processes Geophys.*, 14, 701–708, doi:10.5194/npg-14-701-2007, 2007.
- Arnol'd, V. I.: *Geometrical Methods in the Theory of Ordinary Differential Equations*, Springer-Verlag, New York, 386 pp., 1983.
- Ausloos, M. and Ivanova, K.: Power-law correlations in the Southern-Oscillation-Index fluctuations characterizing El Niño, *Phys. Rev. E*, 63, doi:10.1103/PhysRevE.63.047201, 2001.
- Avnir, D., Biham, O., Lidar, D., and Malcai, O.: Is the geometry of nature fractal?, *Science*, 279, 39–40, 1998.
- Bak, P., Tang, C., and Wiesenfeld, K.: Self-organized criticality, *Phys. Rev. A*, 38, 364–374, 1988.
- Balakrishnan, V., Nicolis, C., and Nicolis, G.: Extreme-value distributions in chaotic dynamics, *J. Stat. Phys.*, 80, 307–336, 1995.
- Bardet, J.-M., Lang, G., Oppenheim, G., Philippe, A., and Taqqu, M. S.: Generators of Long-Range Dependent Processes: A Survey, in: *Theory and Applications of Long-Range Dependence*, edited by: Doukhan, P., Oppenheim, G., and Taqqu, M. S., 579–623, Birkhäuser, Boston, 750 pp., 2003.
- Barnett, T. and Preisendorfer, R.: Multifield analog prediction of short-term climate fluctuations using a climate state vector, *J. Atmos. Sci.*, 35, 1771–1787, 1978.
- Beirlant, J., Goegebeur, Y., Segers, J., and Teugels, J.: *Statistics of Extremes: Theory and Applications*, Probability and Statistics, Wiley, 522 pp., 2004.
- Benhabib, J. and Nishimura, K.: The Hopf-bifurcation and the existence of closed orbits in multi-sectoral models of optimal economic growth, *J. Econ. Theory*, 21, 421–444, 1979.
- Benson, C. and Clay, E.: *Understanding the Economic and Financial Impact of Natural Disasters*, The International Bank for Reconstruction and Development, The World Bank, Washington, D.C., 2004.
- Beran, J.: *Statistics for Long-Memory Processes*, Monographs on Statistics and Applied Probability, Chapman & Hall, 315 pp., 1994.
- Beran, J., Bhansali, R. J., and Ocker, D.: On unified model selection for stationary and nonstationary short- and long-memory autoregressive processes, *Biometrika*, 85, 921–934, 1998.
- Bernacchia, A. and Naveau, P.: Detecting spatial patterns with the cumulant function – Part 1: The theory, *Nonlin. Processes Geophys.*, 15, 159–167, doi:10.5194/npg-15-159-2008, 2008.
- Bernacchia, A., Naveau, P., Vrac, M., and Yiou, P.: Detecting spatial patterns with the cumulant function – Part 2: An application to El Niño, *Nonlin. Processes Geophys.*, 15, 169–177, doi:10.5194/npg-15-169-2008, 2008.
- Biau, G., Zorita, E., von Storch, H., and Wackernagel, H.: Estimation of precipitation by kriging in the EOF space of the sea level pressure field, *J. Climate*, 12, 1070–1085, 1999.
- Blender, R., Fraedrich, K., and Sienn, F.: Extreme event return times in long-term memory processes near  $1/f$ , *Nonlin. Processes Geophys.*, 15, 557–565, doi:10.5194/npg-15-557-2008, 2008.
- Bliefernicht, J. and Bárdossy, A.: Probabilistic forecast of daily areal precipitation focusing on extreme events, *Nat. Hazards Earth Syst. Sci.*, 7, 263–269, doi:10.5194/nhess-7-263-2007, 2007.
- Boldi, M. and Davison, A. C.: Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution, *J. Roy. Stat. Soc.: Series B*, 69, 217–229, 2007.
- Bowman, D. D., Ouillon, G., Sammis, C. G., Sornette, A., and Sornette, D.: An observational test of the critical earthquake concept, *J. Geophys. Res. Solid Earth*, 103, 24359–24372, 1998.
- Box, G. E. P. and Jenkins, G. M.: *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco, 575 pp., 1970.
- Brankovic, C., Matijacic, B., Ivatek-Sahdan, S., and Buizza, R.: Dynamical downscaling of ECMWF ensemble forecasts for cases of severe weather: ensemble statistics and cluster analysis, *Mon. Weather Rev.*, 136, 3323–3342, 2008.
- Brémaud, P.: *Point Processes and Queues*, Springer-Verlag, New York, 373 pp., 1981.
- Bremnes, J. B.: Probabilistic forecasts of precipitation in terms of quantiles using NWP model output, *Mon. Weather Rev.*, 132,



- 338–347, 2004.
- Brillinger, D. R.: *Time Series: Data Analysis and Theory*, Holden-Day, San Francisco, 540 pp., 1981.
- Brillinger, D. R., Guttorp, P. M., and Schoenberg, F. P.: Point processes, temporal, in: *Encyclopedia of Environmetrics*, 3, edited by: El-Shaarawi, A. H. and Piegorsch, W. W., 1577–1581, J. Wiley & Sons, Chichester, 2002.
- Brockwell, P. J. and Davis, R. A.: *Time Series: Theory and Methods*, Springer Series in Statistics, Springer, Berlin, 584 pp., 1991.
- Broomhead, D. S. and King, G. P.: Extracting qualitative dynamics from experimental data, *Physica*, D20, 217–236, 1986.
- Bufe, C. G. and Varnes, D. J.: Predictive Modeling of the Seismic Cycle of the Greater San-Francisco Bay-Region, *J. Geophys. Res. Solid Earth*, 98, 9871–9883, 1993.
- Bui Trong, L.: *Les racines de la violence, De l'émeute au communautarisme*, Louis Audibert, 221 pp., 2003.
- Bunde, A., Eichner, J. F., Havlin, S., and Kantelhart, J. W.: The effect of long-term correlations on the return periods of rare events, *Physica A*, 330, 1–7, 2003.
- Bunde, A., Eichner, J. F., Havlin, S., Koscielny-Bunde, E., Schellnhuber, H. J., and Vyushin, D.: Comment on “Scaling of Atmosphere and Ocean Temperature Correlations in Observations and Climate Models”, *Phys. Rev. Lett.*, 92, 039801, doi:10.1103/PhysRevLett.92.039801, 2004.
- Burridge, R. and Knopoff, L.: Model and Theoretical Seismicity, *Bull. Seism. Soc. Am.*, 57, 341–371, 1967.
- Busuioc, A. and von Storch, H.: Conditional stochastic model for generating daily precipitation time series, *Clim. Res.*, 24, 181–195, 2003.
- Busuioc, A., Tomozeiu, R., and Cacciamani, C.: Statistical downscaling model based on canonical correlation analysis for winter extreme precipitation events in the Emilia-Romagna region, *Int. J. Climatol.*, 28, 449–464, 2008.
- Caballero, R., Jewson, S., and Brix, A.: Long memory in surface air temperature: detection, modeling, and application to weather derivative valuation, *Clim. Res.*, 21, 127–140, 2002.
- Caires, S., Swail, V., and Wang, X.: Projection and analysis of extreme wave climate, *J. Climate*, 19, 5581–5605, 2006.
- Cannon, A. J. and Whitfield, P. H.: Downscaling recent streamflow conditions in British Columbia, Canada using ensemble neural network models, *J. Hydrol.*, 259, 136–151, 2002.
- Carreau, J. and Bengio, Y.: A Hybrid Pareto Model for Asymmetric Fat-Tail, Tech. rep., Dept. IRO, Université de Montréal, 2006.
- Chang, P., Wang, B., Li, T., and Ji, L.: Interactions between the seasonal cycle and the Southern Oscillation: Frequency entrainment and chaos in intermediate coupled ocean-atmosphere model, *Geophys. Res. Lett.*, 21, 2817–2820, 1994.
- Chang, P., Ji, L., Wang, B., and Li, T.: Interactions between the seasonal cycle and El Niño – Southern Oscillation in an intermediate coupled ocean-atmosphere model, *J. Atmos. Sci.*, 52, 2353–2372, 1995.
- Chavez-Demoulin, V. and Davison, A. C.: Generalized additive models for sample extremes, *Appl. Stat.*, 54, 207–222, 2005.
- Chen, Y. D., Chen, X., Xu, C.-Y., and Shao, Q.: Downscaling of daily precipitation with a stochastic weather generator for the subtropical region in South China, *Hydrol. Earth Syst. Sci. Discuss.*, 3, 1145–1183, doi:10.5194/hessd-3-1145-2006, 2006.
- Chen, Z., Ivanov, P. C., Hu, K., and Stanley, H. E.: Effects of non-stationarities on detrended fluctuation analysis, *Phys. Rev. E*, 65, 041107, doi:10.1103/PhysRevE.65.041107, 2002.
- Chiarella, C. and Flaschel: *The Dynamics of Keynesian Monetary Growth*, Cambridge University Press, 434 pp., 2000.
- Chiarella, C., Flaschel, P., and Franke, R.: *Foundations for a Disequilibrium Theory of the Business Cycle*, Cambridge University Press, 550 pp., 2005.
- Chiarella, C., Franke, R., Flaschel, P., and Semmler, W.: *Quantitative and Empirical Analysis of Nonlinear Dynamic Macromodels*, Elsevier, 564 pp., 2006.
- Clauset, A., Shalizi, C., and Newman, M.: Power-law distributions in empirical data, *SIAM Rev.*, 49, doi:10.1137/070710111, 2009.
- Coles, S.: *An Introduction to Statistical Modeling of Extreme Values*, Springer series in statistics, Springer, London, New York, 224 pp., 2001.
- Coles, S. and Casson, E.: Spatial regression for extremes, *Extremes*, 1, 339–365, 1999.
- Coles, S. and Dixon, M.: Likelihood-based inference of extreme value models, *Extremes*, 2, 2–53, 1999.
- Coles, S., Heffernan, J., and Tawn, J.: Dependence measures for extreme value analysis, *Extremes*, 2, 339–365, 1999.
- Coles, S. G. and Powell, E. A.: Bayesian methods in extreme value modelling: A review and new developments, *Intl. Statist. Rev.*, 64, 119–136, 1996.
- Collet, P. and Eckmann, J.-P.: *Iterated Maps on the Interval as Dynamical Systems*, Birkhauser Verlag, Boston, 248 pp., 1980.
- Cooley, D., Naveau, P., and Poncet, P.: Variograms for spatial max-stable random fields, in: *Statistics for dependent data, Lect Notes Stat.*, Springer, 373–390, 2006.
- Cooley, D., Nychka, D., and Naveau, P.: Bayesian spatial modeling of extreme precipitation return levels, *J. Amer. Stat. Assoc.*, 102, 824–840, 2007.
- Cowan, G. A., Pines, D., and Melzer, D.: *Complexity: Metaphors, Models and Reality*, Addison-Wesley, Reading, Mass, 731 pp., 1994.
- Cox, D. R. and Hinkley, D. V.: *Theoretical Statistics*, Chapman & Hall, London, 528 pp., 1994.
- Dahlhaus, R.: Efficient parameter estimation for self-similar processes, *Ann. Stat.*, 17, 1749–1766, 1989.
- Davis, R. and Mikosch, T.: *Extreme Value Theory for Space-Time Processes with Heavy-Tailed Distributions*, *Stoch. Proc. Appl.*, 118, 560–584, 2008.
- Davis, R. and Mikosch, T.: *Extreme Value Theory for GARCH Processes*, in: *Handbook of Financial Time Series*, Springer, New York, 187–200, 2009.
- Davis, R. and Resnick, S. I.: Prediction of stationary max-stable processes, *Ann. Applied Probab.*, 3, 497–525, 1993.
- Davis, R., Dolan, R., and Demme, G.: Synoptic climatology of Atlantic coast northeasters, *Int. J. Climatol.*, 13, 171–189, 1993.
- Davison, A. C. and Ramesh, N. I.: Local likelihood smoothing of sample extremes, *J. Roy. Statist. Soc. B*, 62, 191–208, 2000.
- Day, R.: Irregular growth cycles, *Amer. Econ. Rev.*, 72, 406–414, 1982.
- De Haan, L. and Ferreira, A.: *Extreme Value Theory: An Introduction*, Springer, 436 pp., New York, 2006.
- De Haan, L. and Pereira, T. T.: Spatial extremes: Models for the stationary case, *Ann. Statist.*, 34, 146–168, 2006.
- De Haan, L. and Resnick, S. I.: Limit theory for multivariate sample extremes, *Z. Wahrscheinlichkeitstheorie u. verwandte Gebiete*, 40, 317–337, 1977.

- De Putter, T., Loutre, M. F., and Wansard, G.: Decadal periodicities of Nile River historical discharge (AD 622–1470) and climatic implications, *Geophys. Res. Lett.*, 25, 3193–3196, 1998.
- Dee, D. and Ghil, M.: Boolean difference equations, I: Formulation and dynamic behavior, *SIAM J. Appl. Math.*, 1984.
- Deidda, R.: Rainfall downscaling in a space-time multifractal framework, *Water Resour. Res.*, 36, 1779–1794, 2000.
- Dettinger, M. D., Ghil, M., Strong, C. M., Weibel, W., and Yiou, P.: Software expedites singular-spectrum analysis of noisy time series, *Eos Trans. AGU*, 76, 12, 20, 21, 1995.
- Diaz, H. and Markgraf, V.: *El Niño: Historical and Paleoclimatic Aspects of the Southern Oscillation*, Academic Press, Cambridge University Press, New York, 476 pp., 1993.
- Diebolt, J., Guillo, A., Naveau, P., and Ribereau, P.: Improving probability-weighted moment methods for the generalized extreme value distribution, *REVSTAT – Statistical Journal*, 6, 33–50, 2008.
- Dijkstra, H. A.: *Nonlinear Physical Oceanography: A Dynamical Systems Approach to the Large-Scale Ocean Circulation and El Niño*, Springer, Dordrecht, The Netherlands, 534 pp., 2005.
- Dijkstra, H. A. and Ghil, M.: Low-frequency variability of the large-scale ocean circulation: A dynamical systems approach, *Rev. Geophys.*, 43, RG3002, doi:10.1029/2002RG000122, 2005.
- Drazin, P. G. and King, G. P., eds.: *Interpretation of Time Series from Nonlinear Systems (Proc. of the IUTAM Symposium and NATO Advanced Research Workshop on the Interpretation of Time Series from Nonlinear Mechanical Systems)*, University of Warwick, England, North-Holland, 1992.
- Dumas, P., Ghil, M., Groth, A., and Hallegatte, S.: Dynamic coupling of the climate and macroeconomic systems, *Math. Social Sci.*, in press, 2011.
- Dupuis, D.: Exceedances over high thresholds: A guide to threshold selection, *Extremes*, 1, 251–261, 1999.
- Eckmann, J. and Ruelle, D.: Ergodic theory of chaos and strange attractors, *Rev. Mod. Phys.*, 57, 617–656, and 57, 1115, 1985.
- Eichner, J. F., Koscielny-Bunde, E., Bunde, A., Havlin, S., and Schellnhuber, H.-J.: Power law persistence in the atmosphere: A detailed study of long temperature records, *Phys. Rev. E*, 68, 2003.
- Einmahl, J. and Segers, J.: Maximum empirical likelihood estimation of the spectral measure of an extreme-value distribution, *Ann. Statist.*, 37, 2953–2989, 2009.
- El Adlouni, S. and Ouarda, T. B. M. J.: Comparaison des méthodes d'estimation des paramètres du modèle GEV non-stationnaire, *Revue des Sciences de l'Eau*, 21, 35–50, 2008.
- El Adlouni, S., Ouarda, T. B. M. J., Zhang, X., Roy, R., and Bobee, B.: Generalized maximum likelihood estimators for the nonstationary generalized extreme value model, *Water Resour. Res.*, 43, doi:10.1029/2005WR004545, 2007.
- Embrechts, P., Klüppelberger, C., and Mikosch, T.: *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin/New York, 648 pp., 1997.
- European Commission: *Economic Portrait of the European Union 2001*, European Commission, Brussels, *Panorama of the European Union*, 161 pp., 2002.
- Evans, M., Hastings, N., and Peacock, B.: *Statistical Distributions*, Wiley-Interscience, New York, 170 pp., 2000.
- Feliks, Y., Ghil, M., and Robertson, A. W.: Oscillatory climate modes in the Eastern Mediterranean and their synchronization with the North Atlantic Oscillation, *J. Clim.*, 23, 4060–4079, 2010.
- Feller, W.: *An Introduction to Probability Theory and its Applications* (3rd ed.), I, Wiley, New York, 509 pp., 1968.
- Ferro, C. A. T.: Inference for clusters of extreme values, *J. Roy. Statist. Soc. B*, 65, 545–556, 2003.
- Flaschel, P., Franke, R., and Semmler, W.: *Dynamic Macroeconomics: Instability, Fluctuations and Growth in Monetary Economies*, MIT Press, 455 pp., 1997.
- Fougères, A.: Multivariate extremes, in: *Extreme Values in Finance, Telecommunication and the Environment*, edited by: Finkenstadt, B. and Rootzen, B., Chapman and Hall CRC Press, London, 373–388, 2004.
- Fraedrich, K. and Blender, R.: Scaling of atmosphere and ocean temperature correlations in observations and climate models, *Phys. Rev. Lett.*, 90, 108501, doi:10.1103/PhysRevLett.90.108501, 2003.
- Fraedrich, K. and Blender, R.: Reply to Comment on “Scaling of atmosphere and ocean temperature correlations in observations and climate models”, *Phys. Rev. Lett.*, 92, 039802–1, 2004.
- Fraedrich, K., Luksch, U., and Blender, R.:  $1/f$ -model for long-time memory of the ocean surface temperature, *Phys. Rev. E*, 70, 037301(1–4), 2004.
- Fraedrich, K., Blender, R., and Zhu, X.: Continuum climate variability: long-term memory, extremes, and predictability, *Intl. J. Modern Phys. B*, 23, 5403–5416, 2009.
- Freitas, A. and Freitas, J.: Extreme values for Benedicks-Carleson quadratic maps, *Ergod. Theor. Dyn. Syst.*, 28(4), 1117–1133, 2008b.
- Freitas, A., Freitas, J., and Todd, M.: Extreme value laws in dynamical systems for non-smooth observations, *J. Stat. Phys.*, 142(1), 108–126, 2011.
- Freitas, A. C. M. and Freitas, J. M.: On the link between dependence and independence in extreme value theory for dynamical systems, *Stat. Prob. Lett.*, 78, 1088–1093, 2008a.
- Freitas, J., Freitas, A., and Todd, M.: Hitting time statistics and extreme value theory, *Probab. Theory Related Fields*, 147(3–4), 675–710, 2010.
- Friederichs, P.: Statistical downscaling of extreme precipitation using extreme value theory, *Extremes*, 13, 109–132, 2010.
- Friederichs, P. and Hense, A.: Statistical downscaling of extreme precipitation events using censored quantile regression, *Mon. Weather Rev.*, 135, 2365–2378, 2007.
- Friederichs, P. and Hense, A.: A Probabilistic forecast approach for daily precipitation totals, *Weather Forecast.*, 23, 659–673, 2008.
- Friederichs, P., Goeber, M., Bentzien, S., Lenz, A., and Krampitz, R.: A probabilistic analysis of wind gusts using extreme value statistics, *Meteorol. Z.*, 18, 615–629, 2009.
- Frigessi, A., Haug, O., and Rue, H.: A dynamic mixture model for unsupervised tail estimation without threshold selection, *Extremes*, 5, 219–235, 2002.
- Frisch, R.: Propagation Problems and Impulse Problems in Dynamic Economics, in: *Economic Essay in honor of Gustav Cassel*, 171–206, George Allen and Unwin, London, 1933.
- Furrer, E. and Katz, R. W.: Improving the simulation of extreme precipitation events by stochastic weather generators, *Water Resour. Res.*, 44, W12439, doi:10.1029/2008WR007316, 2008.
- Gabrielov, A., Keilis-Borok, V., Zaliapin, I., and Newman, W. I.: Critical transitions in colliding cascades, *Phys. Rev. E*, 62, 237–

- 249, 2000a.
- Gabrielov, A., Zaliapin, I., Keilis-Borok, and Newman, W. I.: Colliding cascades model for earthquake prediction, *J. Geophys. Intl.*, 143, 427–437, 2000b.
- Gagneur, J. and Casari, G.: From molecular networks to qualitative cell behavior, *Febs Letters*, 579, 1867–1871, 2005.
- Gale, D.: Pure exchange equilibrium of dynamic economic models, *J. Econ. Theory*, 6, 12–36, 1973.
- Gallant, J. C., Moore, I. D., Hutchinson, M. F., and Gessler, P.: Estimating fractal dimension of profiles, *Math. Geol.*, 26, 455–481, 1994.
- Garca, J. A., Gallego, M. C., Serrano, A., and Vaquero, J. M.: Dependence estimation and visualization in multivariate the Iberian Peninsula in the second half of the Twentieth Century, *J. Climate*, 20, 113–130, 2007.
- Geisel, T., Zacherl, A., and Radons, G.: Generic  $1/f$ -noise in chaotic Hamiltonian dynamics, *Phys. Rev. Lett.*, 59, 2503–2506, 1987.
- Gelfand, I. M., Guberman, S. A., Keilis-Borok, V. I., Knopoff, L., Press, F., Ranzman, E. Y., Rotwain, I. M., and Sadovsky, A. M.: Pattern recognition applied to earthquake epicenters in California, *Phys. Earth Planet. Inter.*, 11, 227–283, 1976.
- Geweke, J. and Porter-Hudak, S.: The estimation and application of long memory time series models, *J. Time Ser. Anal.*, 4, 221–237, 1983.
- Ghaleb, K. O.: Le Mikiyas ou Nilomètre de l'île de Rodah, *Mem. Inst. Egypte*, 54, 182 pp., 1951.
- Ghil, M.: The SSA-MTM Toolkit: Applications to analysis and prediction of time series, *Proc. SPIE*, 3165, 216–230, 1997.
- Ghil, M. and Childress, S.: *Topics in Geophysical Fluid Dynamics: Atmospheric Dynamics, Dynamo Theory, and Climate Dynamics*, Springer-Verlag, New York, 485 pp., 1987.
- Ghil, M. and Jiang, N.: Recent forecast skill for the El Niño/Southern Oscillation, *Geophys. Res. Lett.*, 25, 171–174, 1998.
- Ghil, M. and Mo, K. C.: Intraseasonal oscillations in the global atmosphere. Part I: Northern hemisphere and tropics, *J. Atmos. Sci.*, 48, 752–779, 1991.
- Ghil, M. and Mullhaupt, A.: Boolean delay equations. 2. Periodic and aperiodic solutions, *J. Stat. Phys.*, 41, 125–173, 1985.
- Ghil, M. and Robertson, A.: General circulation models and their role in the climate modeling hierarchy, in: *General Circulation Model Developments: Past, Present and Future*, edited by: Randall, D., Academic Press, San Diego, 285–325, 2000.
- Ghil, M. and Robertson, A. W.: “Waves” vs. “particles” in the atmosphere’s phase space: A pathway to long-range forecasting?, *Proc. Natl. Acad. Sci. USA*, 99 (Suppl. 1), 2493–2500, 2002.
- Ghil, M. and Vautard, R.: Interdecadal oscillations and the warming trend in global temperature time series, *Nature*, 350, 324–327, 1991.
- Ghil, M., Halem, M., and Atlas, R.: Time-continuous assimilation of remote-sounding data and its effect on weather forecasting, *Mon. Weather Rev.*, 107, 140–171, 1979.
- Ghil, M., Benzi, R., and Parisi, G.: Turbulence and Predictability in Geophysical Fluid Dynamics and Climate Dynamics, North-Holland, Amsterdam, 1985.
- Ghil, M., Mullhaupt, A., and Pestiaux, P.: Deep water formation and Quaternary glaciations, *Clim. Dyn.*, 2, 1–10, 1987.
- Ghil, M., Allen, M. R., Dettinger, M. D., Ide, K., Kondrashov, D., Mann, M. E., Robertson, A. W., Saunders, A., Tian, Y., Varadi, F., and Yiou, P.: Advanced spectral methods for climatic time series, *Rev. Geophys.*, 40, 3.1–3.41, doi:10.1029/2000RG000092, 2002.
- Ghil, M., Zaliapin, I., and Coluzzi, B.: Boolean delay equations: A simple way of looking at complex systems, *Physica D*, 237, 2967–2986, 2008a.
- Ghil, M., Zaliapin, I., and Thompson, S.: A delay differential model of ENSO variability: parametric instability and the distribution of extremes, *Nonlin. Processes Geophys.*, 15, 417–433, doi:10.5194/npg-15-417-2008, 2008b.
- Glahn, H. R. and Lowry, D. A.: The use of model output statistics (MOS) in objective weather forecasting, *J. Appl. Meteor.*, 11, 1203–1211, 1972.
- Gleick, J.: *Chaos: Making a New Science*, Viking, New York, 352 pp., 1987.
- Goodwin, R.: The non-linear accelerator and the persistence of business cycles, *Econometrica*, 19, 1–17, 1951.
- Goodwin, R.: A growth cycle, in: *Socialism, Capitalism and Economic Growth*, edited by: Feinstein, C., Cambridge University Press, Cambridge, 54–58, 1967.
- Grandmont, J.-M.: On endogenous competitive business cycles, *Econometrica*, 5, 995–1045, 1985.
- Granger, C. W. and Joyeux, R.: An Introduction to Long-Memory Time Series Models and Fractional Differences, *J. Time Ser. Anal.*, 1, 15–29, 1980.
- Greenwood, J. A., Landwehr, J. M., Matalas, N. C., and Wallis, J. R.: Probability Weighted Moments - Definition and Relation to Parameters of Several Distributions Expressable in Inverse Form, *Water Resour. Res.*, 15, 1049–1054, 1979.
- Guckenheimer, J. and Holmes, P.: *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*, Springer, New York, corr. 5th print. edn., 480 pp., 1997.
- Gupta, C., Holland, M., and Nicol, M.: Extreme value theory and return time statistics for dispersing billiard maps and flows, *Lozi maps and Lorenz-like maps, Ergodic Theory Dynam. Systems*, in press, <http://sites.google.com/site/9p9emrak/B111785S>, 2011.
- Gutowitz, H.: *Cellular Automata: Theory and Experiment*, MIT Press, Cambridge, MA, 1991.
- Guttorp, P., Brillinger, D. R., and Schoenberg, F. P.: Point processes, spatial, in: *Encyclopedia of Environmetrics*, vol. 3, edited by: El-Shaarawi, A. H. and Piegorsch, W. W., 1571–1573, J. Wiley & Sons, Chichester, 2002.
- Hahn, F. and Solow, R.: *A Critical Essay on Modern Macroeconomic Theory*, The MIT Press, Cambridge, Massachusetts and London, England, 208 pp., 1995.
- Hall, P. and Tajvidi, N.: Distribution and dependence-function estimation for bivariate extreme-value distributions, *Bernoulli*, 6, 835–844, 2000.
- Hallegatte, S. and Ghil, M.: Natural disasters impacting a macroeconomic model with endogenous dynamics, *Ecol. Econ.*, 68, 582–592, doi:10.1016/j.ecolecon.2008.05.022, 2008.
- Hallegatte, S., Hourcade, J.-C., and Dumas, P.: Why economic dynamics matter in assessing climate change damages: illustration on extreme events, *Ecol. Econ.*, 62, 330–340, doi:10.1016/j.ecolecon.2006.06.006, 2007.
- Hallegatte, S., Ghil, M., Dumas, P., and Hourcade, J.-C.: Business cycles, bifurcations and chaos in a neo-classical model with investment dynamics, *J. Economic Behavior & Organization*, 67,

- 57–77, doi:10.1016/j.jebo.2007.05.001, 2008.
- Hamill, T. M. and Whitaker, J. S.: Probabilistic quantitative precipitation forecasts based on reforecast analogs: theory and application, *Mon. Weather Rev.*, 134, 3209–3229, 2006.
- Hamill, T. M., Whitaker, J. S., and Mullen, S. L.: Reforecasts: An important dataset for improving weather predictions, *B. Am. Meteorol. Soc.*, 87, 33–46, 2006.
- Hannan, E. J.: *Time Series Analysis*, Methuen, New York, 152 pp., 1960.
- Harrod, R.: An essay on dynamic economic theory, *Economic J.*, 49, 1433, 1939.
- Hasselmann, K.: Stochastic climate models, *Tellus*, 6, 473–485, 1976.
- Hastie, T. and Tibshirani, R.: *Generalized Additive Models*, Chapman and Hall, 335 pp., 1990.
- Haylock, M., Cawley, G., Harpham, C., Wilby, R., and Goodess, C.: Downscaling heavy precipitation over the United Kingdom: a comparison of dynamical and statistical methods and their future scenarios, *Int. J. Climatol.*, 26, 1397–1415, 2006.
- Haylock, M. R. and Goodess, C. M.: Interannual variability of European extreme winter rainfall and links with mean large-scale circulation, *Int. J. Climatol.*, 24, 759–776, 2004.
- Heffernan, J. E. and Tawn, J.: A conditional approach for multivariate extreme values, *J. Roy. Statist. Soc. B*, 66, 497–546, 2004.
- Hénon, M.: Sur La Topologie Des Lignes De Courant Dans Un Cas Particulier, *C. R. Acad. Sci. Paris*, 262, 312–414, 1966.
- Hewitson, B. and Crane, R.: Self organizing maps: applications to synoptic climatology, *Clim. Res.*, 26, 1315–1337, 2002.
- Hicks, J.: *The Cycle in Outline*, in: *A Contribution to the Theory of the Trade Cycle*, Oxford University Press, Oxford, 8, 95–107, 1950.
- Hillinger, C.: *Cyclical Growth in Market and Planned Economies*, Oxford University Press, 224 pp., 1992.
- Holland, M. P., Nicol, M., and Török, A.: Extreme value theory for non-uniformly expanding dynamical systems, *Trans. Amer. Math. Soc.*, in press, 2011.
- Hosking, J. R. M.: Fractional differencing, *Biometrika*, 68, 165–176, 1981.
- Hosking, J. R. M.: Maximum-likelihood estimation of the parameters of the generalized extreme-value distribution, *Appl. Stat.*, 34, 301–310, 1985.
- Hosking, J. R. M.: L-moment: Analysis and estimation of distributions using linear combinations of order statistics, *J. Roy. Statist. Soc. B*, 52, 105–124, 1990.
- Hosking, J. R. M., Wallis, J. R., and Wood, E.: Estimation of the generalized extreme-value distribution by the method of probability-weighted moments, *Technometrics*, 27, 251–261, 1985.
- Hurst, H. E.: Long-term storage capacity of reservoirs, *Trans. Am. Soc. Civil Eng.*, 116, 770–799, 1951.
- Hurst, H. E.: *The Nile*, Constable, London, 326 pp., 1952.
- Hüsler, J. and Reiss, R.-D.: Maxima of normal random vectors: Between independence and complete dependence, *Stat. Probabil. Lett.*, 7, 283–286, 1989.
- Huth, R.: Disaggregating climatic trends by classification of circulation patterns, *Int. J. Climatol.*, 21, 135–153, 2001.
- Huth, R.: Statistical downscaling of daily temperature in central Europe, *J. Climate*, 15, 1731–1742, 2002.
- Jacob, F. and Monod, J.: Genetic regulatory mechanisms in synthesis of proteins, *J. Molecular Biol.*, 3, 318–356, 1961.
- Jarsulic, M.: A nonlinear model of the pure growth cycle, *J. Econ. Behav. Organ.*, 22, 133–151, 1993.
- Jaume, S. C. and Sykes, L. R.: Evolving towards a critical point: A review of accelerating seismic moment/energy release prior to large and great earthquakes, *Pure Appl. Geophys.*, 155, 279–305, 1999.
- Jenkins, G. M. and Watts, D. G.: *Spectral Analysis and its Applications*, Holden-Day, San Francisco, 525 pp., 1968.
- Jin, F.: Tropical ocean-atmosphere interaction, the Pacific cold tongue, and the El Niño-Southern Oscillation, *Science*, 274, 1996.
- Jin, F., Neelin, J., and Ghil, M.: El Niño on the Devil's staircase: Annual subharmonic steps to chaos, *Science*, 264, 70–72, 1994.
- Jin, F., Neelin, J., and Ghil, M.: El Niño/Southern Oscillation and the annual cycle: Subharmonic frequency locking and aperiodicity, *Physica D*, 98, 1996.
- Jin, F. F. and Neelin, J. D.: Modes of Interannual Tropical Ocean-Atmosphere Interaction – a Unified View.1. Numerical Results, *J. Atmos. Sci.*, 50, 3477–3503, 1993a.
- Jin, F. F. and Neelin, J. D.: Modes of Interannual Tropical Ocean-Atmosphere Interaction – a Unified View.2. Analytical results in fully coupled cases, *J. Atmos. Sci.*, 50, 3504–3522, 1993b.
- Jin, F. F. and Neelin, J. D.: Modes of Interannual Tropical Ocean-Atmosphere Interaction – a Unified View.3. Analytical Results in Fully Coupled Cases, *J. Atmos. Sci.*, 50, 3523–3540, 1993c.
- Joe, H., Smith, R., and Weissmann, I.: Bivariate threshold methods for extremes, *Royal Statistic Society, Series B*, 54, 171–183, 1992.
- Kabluchko, Z., Schlather, M., and de Haan, L.: Stationary max-stable fields associated to negative definite functions, 37, 2042–2065, 2009.
- Kaldor, N.: A model of the trade cycle, *Econ. J.*, 50, 78–92, 1940.
- Kalecki, M.: A theory of the business cycle, *Review of Economic Studies*, 4, 77–97, 1937.
- Kallache, M., Rust, H. W., and Kropp, J.: Trend assessment: applications for hydrology and climate research, *Nonlin. Processes Geophys.*, 12, 201–210, doi:10.5194/npg-12-201-2005, 2005.
- Kalnay, E., Kanamitsu, M., Kistler, R., Collins, W., Deaven, D., Gandin, L., Iredell, M., Saha, S., White, G., Woollen, J., Zhu, Y., Chelliah, M., Ebisuzaki, W., Higgins, W., Janowiak, J., Mo, K., Ropelewski, C., Wang, J., Leetmaa, A., Reynolds, R., Jenne, R., and Joseph, D.: The NCEP/NCAR 40-year reanalysis project, *Bull. Amer. Meteorol. Soc.*, 77, 437–471, 1996.
- Kantelhardt, J. W., Koscielny-Bunde, E., Rego, H. H. A., Havlin, S., and Bunde, A.: Detecting Long-range Correlations with Detrended Fluctuation Analysis, *Physica A*, 295, 441–454, 2001.
- Kantz, H. and Schreiber, T.: *Nonlinear Time Series Analysis*, 2nd Edition, Cambridge Univ. Press, Cambridge, 388 pp., 2004.
- Kantz, H., Altman, E. G., Hallerberg, S., Holstein, D., and Riegert, A.: Dynamical interpretation of extreme events: Predictability and prediction, in: *Extreme Events in Nature and Society*, edited by: Albeverio, S., Jentsch, V., and Kantz, H., 69–94, Birkhäuser, Basel/Boston, 2005.
- Katz, R.: Precipitation as a chain-dependent process, *J. Appl. Meteorol.*, 16, 671–676, 1977.
- Katz, R. W., Parlange, M. B., and Naveau, P.: Statistics of extremes in hydrology, *Adv. Water Resour.*, 25, 1287–1304, 2002.
- Kauffman, S.: *The Origins of Order: Self-Organization and Selec-*

- tion in Evolution, Oxford University Press, 709 pp., 1993.
- Kauffman, S. A.: Metabolic stability and epigenesis in randomly constructed genetic nets, *J. Theor. Biol.*, 22, 437–467, 1969.
- Kauffman, S. A.: *At Home in the Universe: The Search for Laws of Self-Organization and Complexity*, Oxford University Press, New York, 336 pp., 1995.
- Keilis-Borok, V. I.: Symptoms of instability in a system of earthquake-prone faults, *Physica D*, 77, 193–199, 1994.
- Keilis-Borok, V. I.: Intermediate-term earthquake prediction, *Proc. Natl. Acad. Sci. USA*, 93, 3748–3755, 1996.
- Keilis-Borok, V. I.: Earthquake prediction: State-of-the-art and emerging possibilities, *Ann. Rev. Earth Planet. Sci.*, 30, 1–33, 2002.
- Keilis-Borok, V. I. and Kossobokov, V.: Premonitory activation of earthquake flow: algorithm M8, *Phys. Earth Planet. Inter.*, 61, 73–83, 1990.
- Keilis-Borok, V. I. and Malinovskaya, L.: One regularity in the occurrence of strong earthquakes, *J. Geophys. Res.*, 69, 3019–3024, 1964.
- Keilis-Borok, V. I. and Shebalin, P.: Dynamics of lithosphere and earthquake prediction, *Phys. Earth Planet. Int.*, 111, 179–330, 1999.
- Keilis-Borok, V. I. and Soloviev, A. A.: *Nonlinear Dynamics of the Lithosphere and Earthquake Prediction*, Springer-Verlag, Berlin-Heidelberg, 338 pp., 2003.
- Keilis-Borok, V. I., Knopoff, L., and Rotwain, I. M.: Bursts of aftershocks, long-term precursors of strong earthquakes, *Nature*, 283, 258–263, 1980.
- Keilis-Borok, V. I., Knopoff, L., Rotwain, I. M., and Allen, C. R.: Intermediate-term prediction of occurrence times of strong earthquakes, *Nature*, 335, 690–694, 1988.
- Keilis-Borok, V. I., Stock, J. H., Soloviev, A., and Mikhalev, P.: Pre-recession pattern of six economic indicators in the USA, *J. Forecast.*, 19, 65–80, 2000.
- Keilis-Borok, V. I., Gascon, D. J., Soloviev, A. A., Intriligator, M. D., Pichardo, R., and Winberg, F. E.: On predictability of homicide surges in megacities, in: *Risk Science and Sustainability*, edited by: Beer, T. and Ismail-Zadeh, A., vol. 112 of NATO Science Series. II. Mathematics, Physics and Chemistry, 91–110, Kluwer Academic Publishers, Dordrecht-Boston-London, 2003.
- Keilis-Borok, V. I., Soloviev, A. A., Allègre, C. B., Sobolevskii, A. N., and Intriligator, M. D.: Patterns of macroeconomic indicators preceding the unemployment rise in Western Europe and the USA, *Pattern Recognition*, 38, 423–435, 2005.
- Keizer, K., Lindenberg, S., and Steg, L.: The spreading of disorder, *Science*, 322, 1681–1685, doi:10.1126/science.1161405, 2008.
- Kelling, G. L. and Wilson, J. Q.: Broken windows, *Atlantic Monthly*, 1982, 29 pp., March 1982.
- Keppen, C. L. and Ghil, M.: Extreme weather events, *Nature*, 358, 547, 1992a.
- Keppen, C. L. and Ghil, M.: Adaptive filtering and prediction of the Southern Oscillation Index, *J. Geophys. Res.*, 97, 20449–20454, 1992b.
- Keppen, C. L. and Ghil, M.: Adaptive filtering and prediction of noisy multivariate signals: An application to subannual variability in atmospheric angular momentum, *Intl. J. Bifurcation & Chaos*, 3, 625–634, 1993.
- Kesten, H.: Random difference equations and renewal theory for products of random matrices, *Acta Mathematica*, 131, 207–248, 1973.
- Kharin, V. V. and Zwiers, F.: Changes in the extremes in an ensemble of transient climate simulation with a coupled atmosphere-ocean GCM, *J. Climate*, 13, 3760–3788, 2000.
- Kharin, V. V., Zwiers, F. W., Zhang, X., and Hegerl, G. C.: Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations, *J. Climate*, 20, 1419–1444, 2007.
- King, R. and Rebelo, S.: Resuscitating real business cycles, in: *Handbook of Macroeconomics*, edited by: Taylor, J. and Woodford, M., North-Holland, Amsterdam, 927–1007, 2000.
- Király, A. and Jánosi, I. M.: Detrended fluctuation analysis of daily temperature records: Geographic dependence over Australia, *Meteorol. Atmos. Phys.*, 88, 119–128, 2004.
- Klein, W. H.: Computer Prediction of Precipitation Probability in the United States., *J. Appl. Meteorol.*, 10, 903–915, 1971.
- Knopoff, L., Levshina, T., Keilis-Borok, V. I., and Mattoni, C.: Increased long-range intermediate-magnitude earthquake activity prior to strong earthquakes in California, *J. Geophys. Res. Solid Earth*, 101, 5779–5796, 1996.
- Koenker, R. and Bassett, B.: Regression quantiles, *Econometrica*, 46, 33–49, 1978.
- Kolmogorov, A. N.: Local structures of turbulence in fluid for very large Reynolds numbers, in: *Transl. in Turbulence*, edited by: Friedländer, S. K. and Topper, L., Interscience Publisher, New York, 1961, 151–155, 1941.
- Kondrashov, D. and Ghil, M.: Spatio-temporal filling of missing points in geophysical data sets, *Nonlin. Processes Geophys.*, 13, 151–159, doi:10.5194/npg-13-151-2006, 2006.
- Kondrashov, D., Feliks, Y., and Ghil, M.: Oscillatory modes of extended Nile River records (AD 622–1922), *Geophys. Res. Lett.*, 32, L10702, doi:10.1029/2004GL022156, 2005.
- Kondrashov, D., Kravtsov, S., Robertson, A. W., and Ghil, M.: A hierarchy of data-based ENSO models, *J. Climate*, 18, 4425–4444, 2005b.
- Koscielny-Bunde, E., Bunde, A., Havlin, S., and Goldreich, Y.: Analysis of daily temperature fluctuations, *Physica A*, 231, 393–396, 1996.
- Kossobokov, V.: User Manual for M8, in: *Algorithms for Earthquake Statistics and Prediction*, edited by: Healy, J. H., Keilis-Borok, V. I., and Lee, W. H. K., 6 of IASPEI Software Library, Seismological Society of America, El Cerrito, CA, 167–222, 1997.
- Kossobokov, V.: Quantitative Earthquake Prediction on Global and Regional Scales, in: *Recent Geodynamics, Georisk and Sustainable Development in the Black Sea to Caspian Sea Region*, edited by: Ismail-Zadeh, A., American Institute of Physics Conference Proceedings, Melville, New York, 825, 32–50, 2006.
- Kossobokov, V. and Shebalin, P.: Earthquake Prediction, in: *Nonlinear Dynamics of the Lithosphere and Earthquake Prediction*, edited by: Keilis-Borok, V. I. and Soloviev, A. A., Springer-Verlag, Berlin-Heidelberg, 141–207, 2003.
- Kossobokov, V. G., Maeda, K., and Uyeda, S.: Precursory activation of seismicity in advance of Kobe, 1995, M = 7.2 earthquake, *Pure Appl. Geophys.*, 155, 409–423, 1999a.
- Kossobokov, V. G., Romashkova, L. L., Keilis-Borok, V. I., and Healy, J. H.: Testing earthquake prediction algorithms: Statistically significant real-time prediction of the largest earthquakes in the Circum-Pacific, 1992–1997, *Phys. Earth Planet. Inter.*, 111, 1973.

- 187–196, 1999b.
- Kydland, F. and Prescott, E.: Time to build and aggregate fluctuations, *Econometrica*, 50, 1345–1370, 1982.
- Lalaurette, F.: Early detection of abnormal weather using a probabilistic extreme forecast index, *Q. J. R. Meteorol. Soc.*, 129, 3037–3057, 2003.
- Landwehr, J. M., Matalas, N. C., and Wallis, J. R.: Probability Weighted Moments Compared with Some Traditional Techniques in Estimating Gumbel Parameters and Quantiles, *Water Resour. Res.*, 15, 1055–1064, 1979.
- Lasota, A. and Mackey, M. C.: *Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics*, 97, Appl. Mathematical Sci., Springer-Verlag, New York, 2 edn., 472 pp., 1994.
- Latif, M., Barnett, T., Cane, M., Flügel, M., Graham, N., Von Storch, H., Xu, J., and Zebiak, S.: A review of ENSO prediction studies, *Clim. Dyn.*, 9, 167–179, 1994.
- Leadbetter, M. R., Lindgren, G., and Rootzén, H.: *Extremes and Related Properties of Random Sequences and Processes*, Springer Series in Statistics, Springer-Verlag, New York, 336 pp., 1983.
- Ledford, A. W. and Tawn, J. A.: Modelling dependence within joint tail regions, *J. Roy. Statist. Soc. B*, 59, 475–499, 1997.
- Lichtman, A. J. and Keilis-Borok, V. I.: Aggregate-level analysis and prediction of midterm senatorial elections in the United States, *Proc. Natl. Acad. Sci. USA*, 86, 10176–10180, 1989.
- Lo, A. W.: Long-term memory in stock market prices, *Econometrica*, 59, 1279–1313, 1991.
- Lomb, N.: Least-squares frequency analysis of unequally spaced data, *Astrophys. Space Sci*, 39, 447–462, 1976.
- Lorenz, E.: Deterministic nonperiodic flow, *J. Atmos. Sci.*, 20, 130–141, 1963.
- Lorenz, E. N.: Atmospheric predictability as revealed by naturally occurring analogues, *J. Atmos. Sci.*, 26, 636–646, 1969.
- Lye, L. M., Hapuarachchi, K. P., and Ryan, S.: Bayes estimation of the extreme-value reliability function, *IEEE Trans. Reliab.*, 42, 641–644, 1993.
- Malamud, B.: Tails of natural hazards, *Physics World*, 17, 31–35, 2004.
- Malamud, B. and Turcotte, D.: Self-affine time series: measures of weak and strong persistence, *J. Statistical Planning Inference*, 80, 173–196, 1999.
- Mandelbrot, B.: *The Fractal Geometry of Nature*, W. H. Freeman, New York, 460 pp., 1982.
- Mandelbrot, B. B. and Wallis, J. R.: Noah, Joseph and operational hydrology, *Water Resour. Res.*, 4, 909–918, 1968a.
- Mandelbrot, B. B. and Wallis, J. R.: Robustness of the rescaled range R/S and measurement of non-cyclic long-run statistical dependence, *Water Resour. Res.*, 5, 967–988, 1968b.
- Mañé, R.: On the dimension of the compact invariant sets of certain non-linear maps, in: *Dynamical Systems and Turbulence*, edited by: Rand, D. A. and Young, L.-S., 898, Lect. Notes Math., Springer, Berlin, 230–242, 1981.
- Mann, M. E. and Lees, J. M.: Robust estimation of background noise and signal detection in climatic time series, *Clim. Change*, 33, 409–445, 1996.
- Manneville, P.: Intermittency, self-similarity and 1/f spectrum in dissipative dynamical systems, *J. Physique*, 41, 1235–1243, 1980.
- Maraun, D., Rust, H. W., and Timmer, J.: Tempting long-memory - on the interpretation of DFA results, *Nonlin. Processes Geophys.*, 11, 495–503, doi:10.5194/npg-11-495-2004, 2004.
- Maraun, D., Wetterhall, F., Ireson, A., Chandler, R., Kendon, E., Widmann, M., Brien, S., Rust, H., Sauter, T., Themessl, M., Venema, V., Chun, K., Goodess, C., Jones, R., Onof, C., Vrac, M., and Thiele-Eich, I.: Precipitation Downscaling under climate change. Recent developments to bridge the gap between dynamical models and the end user, *Rev. Geophys.*, 48, RG3003, doi: 10.1029/2009RG000314, 2010.
- Markovic, D. and Koch, M.: Sensitivity of Hurst parameter estimation to periodic signals in time series and filtering approaches, *Geophys. Res. Lett.*, 32, L17401, doi:10.1029/2005GL024069, 2005.
- Marsigli, C., Boccanera, F., Montani, A., and Paccagnella, T.: The COSMO-LEPS mesoscale ensemble system: validation of the methodology and verification, *Nonlin. Processes Geophys.*, 12, 527–536, doi:10.5194/npg-12-527-2005, 2005.
- Marzban, C., Sandgathe, S., and Kalnay, E.: MOS, Perfect Prog. and Reanalysis Data, *Mon. Weather Rev.*, 134, 657–663, 2005.
- Mertins, A.: *Signal Analysis: Wavelets, Filter Banks, Time-Frequency Transforms and Applications*, 330 pp., 1999.
- Mestre, O. and Halegatte, S.: Predictors of tropical cyclone numbers and extreme hurricane intensities over the North Atlantic using generalized additive and linear models, *J. Climate*, 22, 633–648, doi:10.1175/2008JCLI2318.1, 2009.
- Metzler, R.: Comment on “Power-Law correlations in the southern-oscillation-index fluctuations characterizing El Niño”, *Phys. Rev. E*, 67, doi:10.1103/Physreve.67.018201, 2003.
- Min, S.-K., Zhang, X., Zwiers, F. W., Friederichs, P., and Hense, A.: Signal detectability in extreme precipitation changes assessed from twentieth century climate simulations., *Clim. Dynam.*, 32, 95–111, 2009.
- Mitchell, J. M.: An overview of climatic variability and its causal mechanisms, *Quatern. Res.*, 6, 481–493, 1976.
- Mitzenmacher, M.: A Brief History of Generative Models for Power Law and Lognormal Distributions, *Internet Mathematics*, 1, 226–251, 2003.
- Molchan, G. M.: Earthquake prediction as a decision-making problem, *Pure Appl. Geophys.*, 149, 233–247, 1997.
- Molchan, G. M., Dmitrieva, O., Rotwain, I., and Dewey, J.: Statistical analysis of the results of earthquake prediction, based on burst of aftershocks, *Phys. Earth Planet. Int.*, 61, 128–139, 1990.
- Monetti, R. A., Havlin, S., and Bunde, A.: Long term persistence in the sea surface temperature fluctuations, *Physica A*, 320, 581–589, 2001.
- Montanari, A., Rosso, R., and Taqqu, M. S.: A seasonal fractional ARIMA model applied to the Nile River monthly flows at Aswan, *Water Resour. Res.*, 36, 1249–1259, 2000.
- Montani, A., Capaldo, M., Cesari, D., Marsigli, C., Modigliani, U., Nerozzi, F., Paccagnella, T., Patruno, P., and Tibaldi, S.: Operational limited-area ensemble forecasts based on the Lokal Model, *ECMWF Newsletter*, 98, 2–7, 2003.
- Mudelsee, M.: Long memory of rivers from spatial aggregation, *Water Resour. Res.*, 43, doi:10.129/2006WR005721, 2007.
- Mullhaupt, A. P.: *Boolean Delay Equations: A Class of Semi-Discrete Dynamical Systems*, Ph.D. thesis, New York University, 193 pp., 1984.
- Nadarajah, S.: A polynomial model for bivariate extreme value distributions, *Stat. Probabil. Lett.*, 42, 15–25, 1999.

- Narteau, C., Shebalin, P., and Holschneider, M.: Loading rates in California inferred from aftershocks, *Nonlin. Processes Geophys.*, 15, 245–263, doi:10.5194/npg-15-245-2008, 2008.
- Naveau, P., Guillou, A., Cooley, D., and Diebolt, J.: Modelling pairwise dependence of maxima in space, *Biometrika*, 96, 1–17, 2009.
- Nelder, J. A. and Wedderburn, R. W.: Generalized Linear Models, *J. Roy. Statist. Soc. A*, 135, 370–384, 1972.
- Newman, W. I., Gabrielov, A., and Turcotte, D.: Nonlinear Dynamics and Predictability of Geophysical Phenomena, American Geophysical Union, Washington, DC, 107 pp., 1994.
- Nicolis, C. and Nicolis, S. C.: Return time statistics of extreme events in deterministic dynamical systems, *Europ. Phys. Lett.*, 80, doi:10.1209/0295-5075/80/40003, 2007.
- Nicolis, C., Balakrishnan, V., and Nicolis, G.: Extreme events in deterministic dynamical systems, *Phys. Rev. Lett.*, 97, doi:10.1103/PhysRevLett.97.210602, 2006.
- Nicolis, G., Balakrishnan, V., and Nicolis, C.: Probabilistic aspects of extreme events generated by periodic and quasiperiodic deterministic dynamics, *Stoch. Dynam.*, 8, 115–125, 2008.
- Nikaido, H.: Prices, Cycles, and Growth, MIT Press, Cambridge, MA, 285 pp., 1996.
- Nordhaus, W. D. and Boyer, J.: Roll the DICE Again: The Economics of Global Warming, Tech. rep., Yale University, <http://www.econ.yale.edu/~nordhaus/homepage/dicemodels.htm>, 1998.
- Orlowsky, B. and Fraedrich, K.: Upscaling European surface temperatures to North Atlantic circulation-pattern statistics, *Intl. J. Climatology*, 29, 839–849, doi:10.1002/joc.1744, 2009.
- Orlowsky, B., Bothe, O., Fraedrich, K., Gerstengarbe, F.-W., and Zhu, X.: Future climates from bias-bootstrapped weather analogues: an application to the Yangtze river basin, *J. Climate*, 23, 3509–3524, 2010.
- Ott, E., Sauer, T., and Yorke, J. A., eds.: Coping with Chaos, Analysis of Chaotic Data and the Exploitation of Chaotic Systems, J. Wiley & Sons, New York, 418 pp., 1994.
- Packard, N. H., Crutchfield, J. P., Farmer, J. D., and Shaw, R. S.: Geometry from a time series, *Phys. Rev. Lett.*, 45, 712–716, 1980.
- Padoan, S. A., Ribatet, M., and Sisson, S. A.: Likelihood-Based Inference for Max-Stable Processes, *Journal of the American Statistical Association*, 105, 263–277, doi:10.1198/jasa.2009.tm08577, 2010.
- Penland, C., Ghil, M., and Weickmann, K. M.: Adaptive filtering and maximum entropy spectra, with application to changes in atmospheric angular momentum, *J. Geophys. Res.*, 96, 22659–22671, 1991.
- Pepke, G., Carlson, J., and Shaw, B.: Prediction of large events on a dynamical model of fault, *J. Geophys. Res.*, 99, 6769–6788, 1994.
- Percival, D. B. and Walden, A. T.: Spectral Analysis for Physical Applications, Cambridge University Press, Cambridge, UK, 583 pp., 1993.
- Percival, D. B. and Walden, A. T.: Wavelet Methods for Time Series Analysis, Cambridge University Press, Cambridge, UK, 600 pp., 2000.
- Perica, S. and Foufoula-Georgiou, E.: Linkage of scaling and thermodynamic parameters of rainfall: Results from midlatitude mesoscale convective systems, *J. Geophys. Res.*, 101, 7431–7448, 1996.
- Philander, S. G. H.: El Niño, La Niña, and the Southern Oscillation, Academic Press, San Diego, 289 pp., 1990.
- Plaut, G. and Vautard, R.: Spells of low-frequency oscillations and weather regimes in the Northern Hemisphere, *J. Atmos. Sci.*, 51, 210–236, 1994.
- Plaut, G., Schuepbach, E., and Doctor, M.: Heavy precipitation events over a few Alpine sub-regions and the links with large-scale circulation, *Clim. Res.*, 17, 285–302, 2001.
- Popper, W.: The Cairo Nilometer, Univ. of Calif. Press, Berkeley, 269 pp., 1951.
- Priestley, M. B.: Spectral Analysis and Time Series, Academic Press, London, 890 pp., 1992.
- Procaccia, I. and Schuster, H.: Functional renormalization-group theory of universal  $1/f$  noise in dynamical systems, *Phys. Rev. A*, 28, 1210–1212, 1983.
- Pryor, S., Schoof, J. T., and Barthelmie, R.: Empirical downscaling of wind speed probability distributions, *J. Geophys. Res.*, 110, doi:10.1029/2005JD005899, 2005.
- Ramos, A. and Ledford, A.: A new class of models for bivariate joint tails, *J. Roy. Stat. Soc. B*, 71, 219–241, 2009.
- Resnick, S. I.: Extreme Values, Regular Variation, and Point Processes, 4 of Applied Probability, Springer-Verlag, New York, 334 pp., 1987.
- Resnick, S. I.: Heavy-Tail Phenomena: Probabilistic and Statistical Modeling, Operations Research and Financial Engineering, Springer, New York, 410 pp., 2007.
- Ribatet, M.: SpatialExtremes: A R package for Modelling Spatial Extremes, <http://cran.r-project.org/>, 2008.
- Ribatet, M.: POT: Generalized Pareto Distribution and Peaks Over Threshold, <http://people.epfl.ch/mathieu.ribatet>, <http://r-forge.r-project.org/projects/pot/>, R package version 1.0-9, 2009.
- Risk Management Solutions: Hurricane Katrina: Profile of a Super Cat. Lessons and Implications for Catastrophe Risk Management., Tech. rep., Risk Management Solutions, Newark, CA, 30 pp., 2005.
- Robertson, A. W. and Ghil, M.: Large-scale weather regimes and local climate over the Western United States, *J. Climate*, 12, 1796–1813, 1999.
- Robinson, P. M.: Log-periodogram regression of time series with long-range dependence, *Ann. Statist.*, 23, 1048–1072, 1995.
- Robinson, P. M.: Time Series with Long Memory, Advanced Texts in Econometrics, Oxford University Press, 392 pp., 2003.
- Rossi, M., Witt, A., Guzzetti, F., Malamud, B. D., and Peruccacci, S.: Analysis of historical landslide time series in the Emilia-Romagna region, northern Italy, *Earth Surf. Proc. Land.*, 35, 1123–1137, 2010.
- Roux, J. C., Rossi, A., Bachelart, S., and Vidal, C.: Representation of a strange attractor from an experimental study of chemical turbulence, *Phys. Lett. A*, 77, 391–393, 1980.
- Rudin, W.: Real and Complex Analysis (3rd ed.), McGraw-Hill, New York, 483 pp., 1987.
- Ruelle, D.: Small random perturbations of dynamical systems and the definition of attractors, *Commun. Math. Phys.*, 82, 137–151, 1981.
- Ruelle, D. and Takens, F.: On the nature of turbulence, *Commun. Math. Phys.*, 20, 167–192, and 23, 343–344, 1971.
- Rundle, J., Turcotte, D., and Klein, W., eds.: Geocomplexity and the



- Physics of Earthquakes, American Geophysical Union, Washington, DC, 2000.
- Rust, H. W.: Detection of Long-Range Dependence – Applications in Climatology and Hydrology, Ph.D. thesis, Potsdam University, Potsdam, <http://nbn-resolving.de/urn:nbn:de:kobv:517-opus-13347>, 2007.
- Rust, H. W.: The effect of long-range dependence on modelling extremes with the generalised extreme value distribution, *Europ. Phys. J. Special Topics*, 174, 1, 91–97, doi:10.1140/epjst/e2009-01092-8, 2009.
- Rust, H. W., Mestre, O., and Venema, V. K. C.: Fewer jumps, less memory: Homogenized temperature records and long memory, *J. Geophys. Res.*, 113, doi:10.1029/2008JD009919, 2008.
- Rust, H. W., Kallache, M., Schellnhuber, H.-J., and Kropp, J.: Confidence intervals for flood return level estimates using a bootstrap approach, in: *In Extremis*, edited by: Kropp, J. and Schellnhuber, H.-J., Springer, 61–81, 2011.
- Rybski, D., Bunde, A., Havlin, S., and von Storch, H.: Long-term persistence in the climate and the detection problem, *Geophys. Res. Lett.*, 33, doi:10.1029/2005GL025591, 2006.
- Salameh, T., Drobinski, P., Vrac, M., and Naveau, P.: Statistical downscaling of near-surface wind over complex terrain in southern France, *Meteorol. Atmos. Phys.*, 103, 243–256, 2009.
- Sanchez-Gomez, E. and Terray, L.: Large-scale atmospheric dynamics and local intense precipitation episodes, *Geophys. Res. Lett.*, 32, L24711, doi:10.1029/2005GL023990, 2005.
- Samuelson, P.: A Synthesis of the Principle of Acceleration and the Multiplier, *J. Political Econ.*, 47, 786–797, 1939.
- Sato, K.-I.: *Lévy Processes and Infinitely Divisible Distributions*, Cambridge University Press, London, 486 pp., 1999.
- Sauer, T., Yorke, J. A., and Casdagli, M.: Embedology, *J. Stat. Phys.*, 65, 579–616, 1991.
- Saunders, A. and Ghil, M.: A Boolean delay equation model of ENSO variability, *Physica D*, 160, 54–78, 2001.
- Schlather, M.: Models for stationary max-stable random fields, *Extremes*, 5, 33–44, 2002.
- Schlather, M.: *RandomFields: Simulations and Analysis of Random Fields*, R package version 1.3.28, 2006.
- Schlather, M. and Tawn, J.: Inequalities for the extremal coefficients of multivariate extreme value distributions, *Extremes*, 1, 87–102, 2009.
- Schlather, M. and Tawn, J. A.: A dependence measure for multivariate and spatial extreme values: Properties and inference, *Biometrika*, 90, 139–156, 2003.
- Schmidli, J., Goodess, C., Frei, C., Haylock, M., Hurrell, Y., Ribalaygua, J., and Schmith, T.: Statistical and dynamical downscaling of precipitation: An evaluation and comparison of scenarios for the European Alps, *J. Geophys. Res.-Atmos.*, D04105, doi:10.1029/2005JD007026, 2007.
- Schnur, R. and Lettenmaier, D.: A case study of statistical downscaling in Australia using weather classification by recursive partitioning, *J. Hydrol.*, 212–213, 362–379, 1998.
- Scholz, C. H.: *The mechanics of earthquakes and faulting*, Cambridge University Press, Cambridge, U.K.; New York, 2nd edn., 471 pp., 2002.
- Schölzel, C. and Friederichs, P.: Multivariate non-normally distributed random variables in climate research – introduction to the copula approach, *Nonlin. Processes Geophys.*, 15, 761–772, doi:10.5194/npg-15-761-2008, 2008.
- Semenov, M. and Barrow, E.: Use of a stochastic weather generator in the development of climate change scenarios, *Clim. Res.*, 35, 397–414, 1997.
- Semenov, M., Brooks, R., Barrow, E., and Richardson, C.: Comparison of the WGEN and the LARS-WG stochastic weather generators in diverse climates, *Clim. Res.*, 10, 95–107, 1998.
- Serinaldi, F.: Comment on “Multivariate non-normally distributed random variables in climate research – introduction to the copula approach” by C. Schölzel and P. Friederichs, *Nonlin. Processes Geophys.*, 15, 761–772, 2008, *Nonlin. Processes Geophys.*, 16, 331–332, doi:10.5194/npg-16-331-2009, 2009.
- Simonnet, E., Ghil, M., and Dijkstra, H. A.: Homoclinic bifurcations in the quasi-geostrophic double-gyre circulation, *J. Mar. Res.*, 63, 931–956, 2005.
- Slutsky, E.: The summation of random causes as a source of cyclic processes, III(1), *Conjuncture Institute, Moscow, Econometrica*, 5, 105–146, 1927.
- Smale, S.: Differentiable dynamical systems, *Bull. Amer. Math. Soc.*, 73, 747–817, 1967.
- Smith, H. F.: An empirical law describing heterogeneity in the yields of agricultural crops, *J. Agric. Sci.*, 28, 1–23, 1938.
- Smith, R. L.: Maximum-likelihood estimation in a class of nonregular cases, *Biometrika*, 72, 67–90, 1985.
- Smith, R. L.: Extreme value analysis of environmental time series: An application to trend detection in ground-level ozone, *Stat. Sci.*, 4, 367–393, 1989.
- Smith, R. L.: Max-stable processes and spatial extremes, <http://www.stat.unc.edu/postscript/rs/spatex.pdf>, 1990.
- Smith, R. L.: Statistics of extremes, with applications in environment, insurance and finance., in: *Extreme Values in Finance, Telecommunication and the Environment*, edited by: B. Finkenstadt, and Rootzen, H., 1–78, Chapman and Hall CRC Press, London, 1–78, 2004.
- Smith, R. L., Tawn, J., and Coles, S. G.: Markov chain models for threshold exceedances, *Biometrika*, 84, 249–268, doi:10.1093/biomet/84.2.249, 1997.
- Snell, S., Gopal, S., and Kaufmann, R.: Spatial interpolation of surface air temperatures using artificial neural networks: Evaluating their use for downscaling GCMs, *J. Climate*, 13, 886–895, 2000.
- Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K., Manning, M., and Miller, H., eds.: *Climate Change 2007: The Physical Science Basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, Cambridge; New York, 2007.
- Soloviev, A.: Transformation of frequency-magnitude relation prior to large events in the model of block structure dynamics, *Nonlin. Processes Geophys.*, 15, 209–220, doi:10.5194/npg-15-209-2008, 2008.
- Solow, R. M.: A contribution to the theory of economic growth, *Q. J. Econ.*, 70, 65–94, 1956.
- Sornette, D.: Multiplicative processes and power laws, *Phys. Rev. E*, 57, 4811–4813, 1998.
- Sornette, D.: *Critical Phenomena in Natural Sciences, Chaos, Fractals, Self-organization and Disorder: Concepts and Tools*, 2nd edn., Springer, Heidelberg, 528 pp., 2004.
- Soros, G.: *The New Paradigm for Financial Markets: The Credit Crisis of 2008 and What It Means*, BBS, PublicAffairs, New York, 162 pp., 2008.

- Stedinger, J. R., Vogel, R. M., Foufoula-Georgiou, E., and Maidment, D. R.: Frequency analysis of extreme events, in: *Handbook of Hydrology*, McGraw-Hill, 18–23, 1993.
- Stern, N.: *The Economics of Climate Change*. The Stern Review, Cambridge University Press, 712 pp., 2006.
- Takayasu, H., Sato, A.-H., and Takayasu, M.: Stable infinite variance fluctuations in randomly amplified Langevin systems, *Phys. Rev. Lett.*, 79, 966–969, 1997.
- Takens, F.: Detecting strange attractors in turbulence, in: *Dynamical Systems and Turbulence*, edited by: Rand, D. A. and Young, L.-S., 898 *Lec. Notes Math.*, 366–381, Springer, Berlin, 1981.
- Taqq, M. S., Teverovsky, V., and Willinger, W.: Estimators for long-range dependence: An empirical study, *Fractals*, 3, 785–798, 1995.
- Taricco, C., Alessio, S., and Vivaldo, G.: Sequence of eruptive events in the Vesuvio area recorded in shallow-water Ionian Sea sediments, *Nonlin. Processes Geophys.*, 15, 25–32, doi:10.5194/npg-15-25-2008, 2008.
- Thomas, R.: Boolean formalization of genetic-control circuits, *J. Theor. Biol.*, 42, 563–585, 1973.
- Thomas, R.: Logical analysis of systems comprising feedback loops, *J. Theor. Biol.*, 73, 631–656, 1978.
- Thomas, R.: *Kinetic Logic: A Boolean Approach to the Analysis of Complex Regulatory Systems*, Springer-Verlag, Berlin, Heidelberg, New York, 507 pp., 1979.
- Thomson, D. J.: Spectrum estimation and harmonic analysis, *Proc. IEEE*, 70, 1055–1096, 1982.
- Timmer, J. and König, M.: On generating power law noise, *Astron. Astrophys.*, 300, 707–710, 1995.
- Toussoun, O.: Mémoire sur l'histoire du Nil, *Mem. Inst. Egypte*, 18, 366–404, 1925.
- Tucker, W.: The Lorenz attractor exists, *C. R. Acad. Sciences–Math.*, 328, 1197–1202, 1999.
- Turcotte, D.: *Fractals and Chaos in Geology and Geophysics*, Cambridge University Press, Cambridge, 2nd edition edn., 231 pp., 1997.
- Turcotte, D., Newman, W., and Gabrielov, A.: A statistical physics approach to earthquakes, in: *Geocomplexity and the Physics of Earthquakes*, edited by: Rundle, J., Turcotte, D., and Klein, W., American Geophysical Union, Washington, DC, 2000.
- Tziperman, E., Stone, L., Cane, M. A., and Jarosh, H.: El Niño Chaos - Overlapping of Resonances between the Seasonal Cycle and the Pacific Ocean-Atmosphere Oscillator, *Science*, 264, 72–74, 1994.
- Tziperman, E., Cane, M., and Zebiak, S.: Irregularity and locking to the seasonal cycle in an ENSO prediction model as explained by the quasi-periodicity route to chaos, *J. Atmos. Sci.*, 50, 293–306, 1995.
- Unal, Y. S. and Ghil, M.: Interannual and Interdecadal Oscillation Patterns in Sea-Level, *Clim. Dyn.*, 11, 255–278, 1995.
- Vannitsem, S. and Naveau, P.: Spatial dependences among precipitation maxima over Belgium, *Nonlin. Processes Geophys.*, 14, 621–630, doi:10.5194/npg-14-621-2007, 2007.
- Vautard, R. and Ghil, M.: Singular spectrum analysis in nonlinear dynamics, with applications to paleoclimatic time series, *Physica D*, 35, 395–424, 1989.
- Vautard, R., Yiou, P., and Ghil, M.: Singular spectrum analysis: a toolkit for short noisy chaotic signals, *Physica D*, 58, 95–126, 1992.
- Vislocky, R. L. and Young, G. S.: The use of perfect prog forecasts to improve model output statistics forecasts of precipitation probability, *Weather Forecast.*, 4, 202–209, 1989.
- Von Neumann, J.: *Theory of Self-Reproducing Automata*, University of Illinois Press, Urbana, Illinois, 488 pp., 1966.
- Voss, R. F.: Random fractal forgeries, in: *Fundamental Algorithms for Computer Graphics*, edited by: Earnshaw, R., Springer, Berlin, 805–835, 1985.
- Vrac, M. and Naveau, P.: Stochastic downscaling of precipitation: From dry events to heavy rainfalls, *Water Resour. Res.*, 43, doi:10.1029/2006WR005308, 2007.
- Vrac, M., Naveau, P., and Drobinski, P.: Modeling pairwise dependencies in precipitation intensities, *Nonlin. Processes Geophys.*, 14, 789–797, doi:10.5194/npg-14-789-2007, 2007a.
- Vrac, M., Marbaix, P., Paillard, D., and Naveau, P.: Non-linear statistical downscaling of present and LGM precipitation and temperatures over Europe, *Clim. Past*, 3, 669–682, doi:10.5194/cp-3-669-2007, 2007b.
- Vrac, M., Stein, M., and Hayhoe, K.: Statistical downscaling of precipitation through nonhomogeneous stochastic weather typing, *Clim. Res.*, 34, 169–184, doi:10.3354/cr00696, 2007c.
- Vrac, M., Stein, M., Hayhoe, K., and Liang, X.-Z.: A general method for validating statistical downscaling methods under future climate change, *Geophys. Res. Lett.*, 34, doi:10.1029/2007GL030295, 2007d.
- Vyushin, D., Mayer, J., and Kushner, P.: *PowerSpectrum: Spectral Analysis of Time Series*, r-package version 0.3, <http://www.atmosph.physics.utoronto.ca/people/vyushin/mysoftware.html>, 2008.
- Vyushin, D. I. and Kushner, P. J.: Power-law and long-memory characteristics of the atmospheric general circulation, *J. Climate*, 22, 2890–2904, doi:10.1175/2008JCLI2528.1, 2009.
- Vyushin, D. I., Fioletov, V. E., and Shepherd, T. G.: Impact of long-range correlations on trend detection in total ozone, *J. Geophys. Res.*, 112, D14307, doi:10.1029/2006JD008168, 2007.
- Wetterhall, F., Halldin, S., and Xu, C.-Y.: Statistical precipitation downscaling in central Sweden with the analogue method, *J. Hydrol.*, 306, 174–190, 2005.
- Whitcher, B., Byers, S., Guttorp, P., and Percival, D.: Testing for homogeneity of variance in time series: Long memory, wavelets, and the Nile River, *Water Resour. Res.*, 38, 5, 1054, doi:10.1029/2001WR000509, 2002.
- White, E., Enquist, B., and Green, J.: On estimating the exponent of power-law frequency distributions, *Ecology*, 89, 905–912, 2008.
- Whitney, B.: Differentiable manifolds, *Ann. Math.*, 37, 645–680, 1936.
- Widmann, M., Bretherton, C., and Salathé, E.: Statistical precipitation downscaling over the Northwestern United States using numerically simulated precipitation as a predictor, *J. Climate*, 16, 799–816, 2003.
- Wigley, T. M., Jones, P., Briffa, K., and Smith, G.: Obtaining sub-grid scale information from coarse resolution general circulation output, *J. Geophys. Res.*, 95, 1943–1953, 1990.
- Wilby, R. and Wigley, T.: Downscaling general circulation model output: a review of methods and limitations, *Prog. Phys. Geogr.*, 21, 530–548, doi:10.1177/030913339702100403, 1997.
- Wilby, R., Charles, S., Zorita, E., Timbal, B., Whetton, P., and Mearns, L.: *Guidelines for use of Climate Scenarios Developed from Statistical Downscaling Methods*, IPCC Task Group on

- Data and Scenario Support for Impact and Climate Analysis (TG-ICA), <http://ipcc-ddc.cru.uea.ac.uk/guidelines/StatDownGuide.pdf>, 2004.
- Wilks, D. S.: Multi-site generalization of a daily stochastic precipitation model, *J. Hydrol.*, 210, 178–191, 1998.
- Wilks, D. S.: Multi-site downscaling of daily precipitation with a stochastic weather generator, *Clim. Res.*, 11, 125–136, 1999.
- Wilks, D. S. and Wilby, R. L.: The weather generation game: a review of stochastic weather models, *Prog. Phys. Geogr.*, 23, 329–357, 1999.
- Witt, A., Malamud, B. D., Rossi, M., Guzzetti, F., and Peruccacci, S.: Temporal correlations and clustering of landslides, *Earth Surf. Proc. Land.*, 35, 1138–1156, 2010.
- Wolfram, S.: Statistical mechanics of cellular automata, *Rev. Mod. Phys.*, 55, 601–644, 1983.
- Wolfram, S., ed.: *Cellular Automata and Complexity: Collected Papers*, Addison-Wesley, Reading, Mass., 1994.
- World Bank: Turkey: Marmara Earthquake Assessment, The World Bank, Washington, D.C., world Bank Working Paper, 1999.
- Yee, T. W.: A User's Guide to the VGAM Package, Tech. rep., Department of Statistics, University of Auckland, <http://www.stat.auckland.ac.nz/~yee/VGAM>, 2006.
- Yee, T. W. and Hastie, T. J.: Reduced-rank vector generalized linear models, *Stat. Modell.*, 3, 15–41, 2003.
- Yee, T. W. and Stephenson, A.: Vector Generalized Linear and Additive Extreme Value Models, *Extremes*, 10, 1386–1999, 2007.
- Yee, T. W. and Wild, C. J.: Vector generalized additive models, *J. Roy. Statist. Soc. B*, 58, 481–493, 1996.
- Yiou, P. and Nogaj, M.: Extreme climatic events and weather regimes over the North Atlantic: When and where?, *Geophys. Res. Lett.*, 31, L07202, doi:10.1029/2003GL019119, 2004.
- Yiou, P., Loutre, M. F., and Baert, E.: Spectral analysis of climate data, *Surveys Geophys.*, 17, 619–663, 1996.
- Yiou, P., Sornette, D., and Ghil, M.: Data-adaptive wavelets and multi-scale singular-spectrum analysis, *Physica D*, 142, 254–290, 2000.
- Yiou, P., Goubanova, K., Li, Z. X., and Nogaj, M.: Weather regime dependence of extreme value statistics for summer temperature and precipitation, *Nonlin. Processes Geophys.*, 15, 365–378, doi:10.5194/npg-15-365-2008, 2008.
- Yosida, K.: *Functional Analysis*, Academic Press, San Diego, 513 pp., 1968.
- Zaliapin, I., Keilis-Borok, and Ghil, M.: A Boolean delay equation model of colliding cascades, Part I: Multiple seismic regimes., *J. Stat. Phys.*, 111, 2003a.
- Zaliapin, I., Keilis-Borok, V., and Ghil, M.: A Boolean delay equation model of colliding cascades. Part II: Prediction of critical transitions., *J. Stat. Phys.*, 111, 2003b.
- Zhang, J.: Likelihood moment estimation for the generalized pareto distribution, *Aust Nz J. Stat.*, 49, 69–77, 2007.
- Zhang, X., Zwiers, F., and Li, G.: Monte Carlo experiments on the detection of trends in extreme values, *J. Climate*, 17, 1945–1952, 2004.
- Zorita, E. and von Storch, H.: The analog method as a simple statistical downscaling technique: Comparison with more complicated methods, *J. Climate*, 12, 2474–2489, 1998.
- Zorita, E., Hughes, J., Lettenmaier, D., and von Storch, H.: Stochastic characterization of regional circulation patterns for climate model diagnosis and estimation of local precipitation, Report of the Max-Planck-Institut für Meteorologie, Hamburg, 109, 1993.
- Zwiers, F. and Kharin, V. V.: Changes in the extremes of the climate simulated by CCC GCM2 under CO<sub>2</sub> doubling, *J. Climate*, 11, 2200–2222, 1998.