Recurrences of Extreme Events

Eduardo G. Altmann

Max Planck Institute
for the Physics of Complex Systems
Dresden -Germany

http://www.pks.mpg.de/~edugalt/

Eötvös University Budapest November 14 2007.

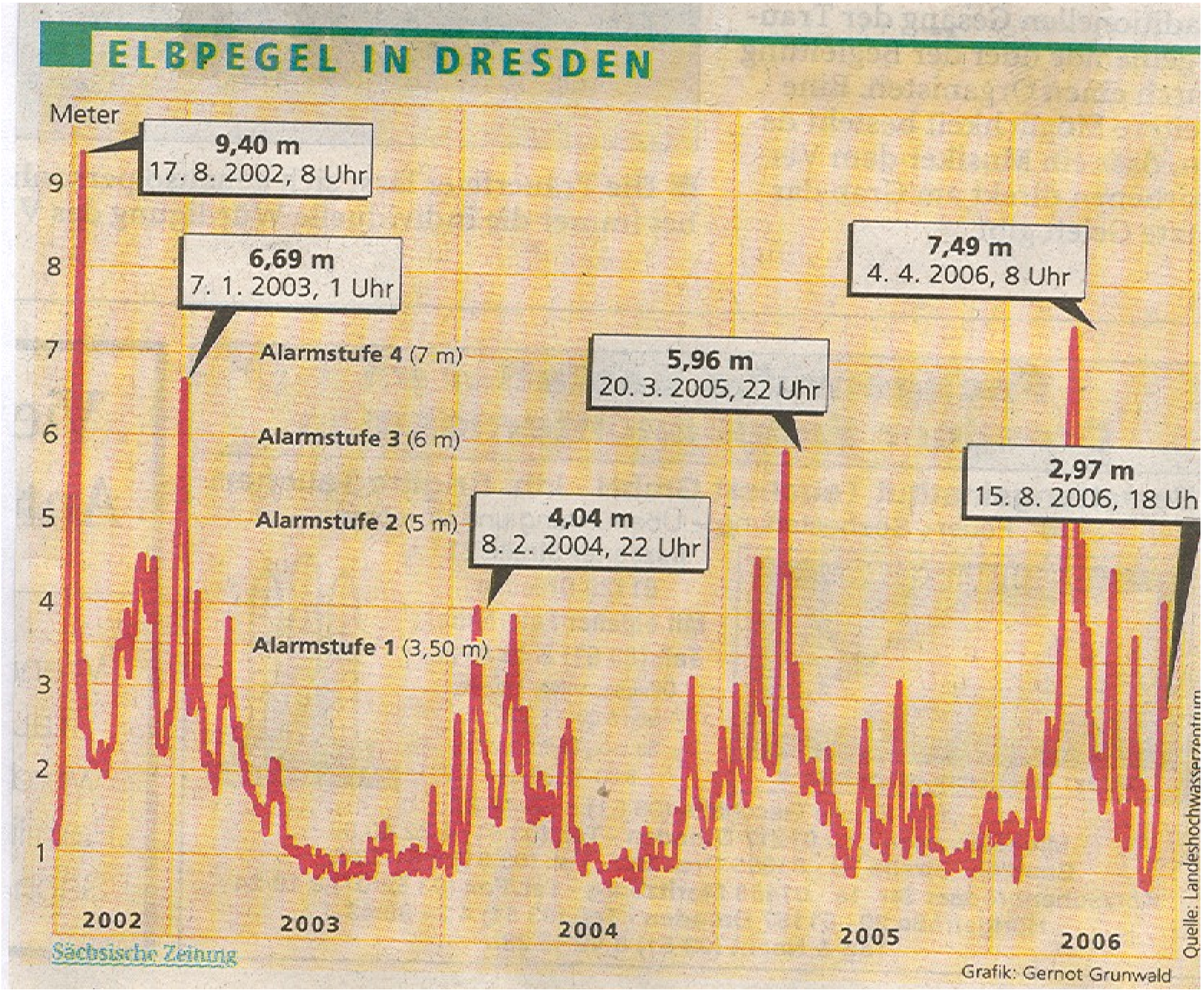
Contents

I - Introduction

What are extreme events?

I – Example 1: Floods

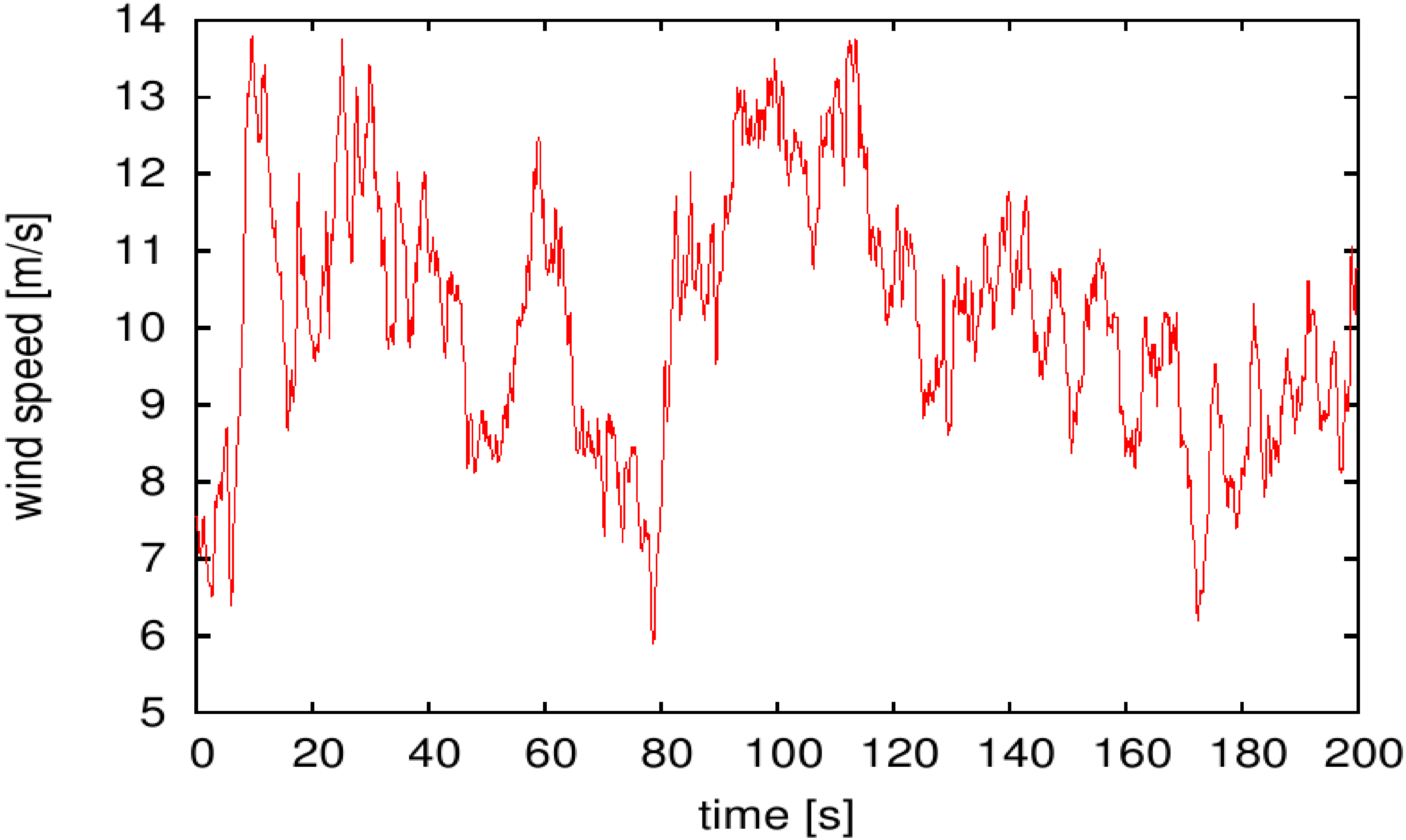
I – Example 1: Floods



I – Example 2: Wind Gusts

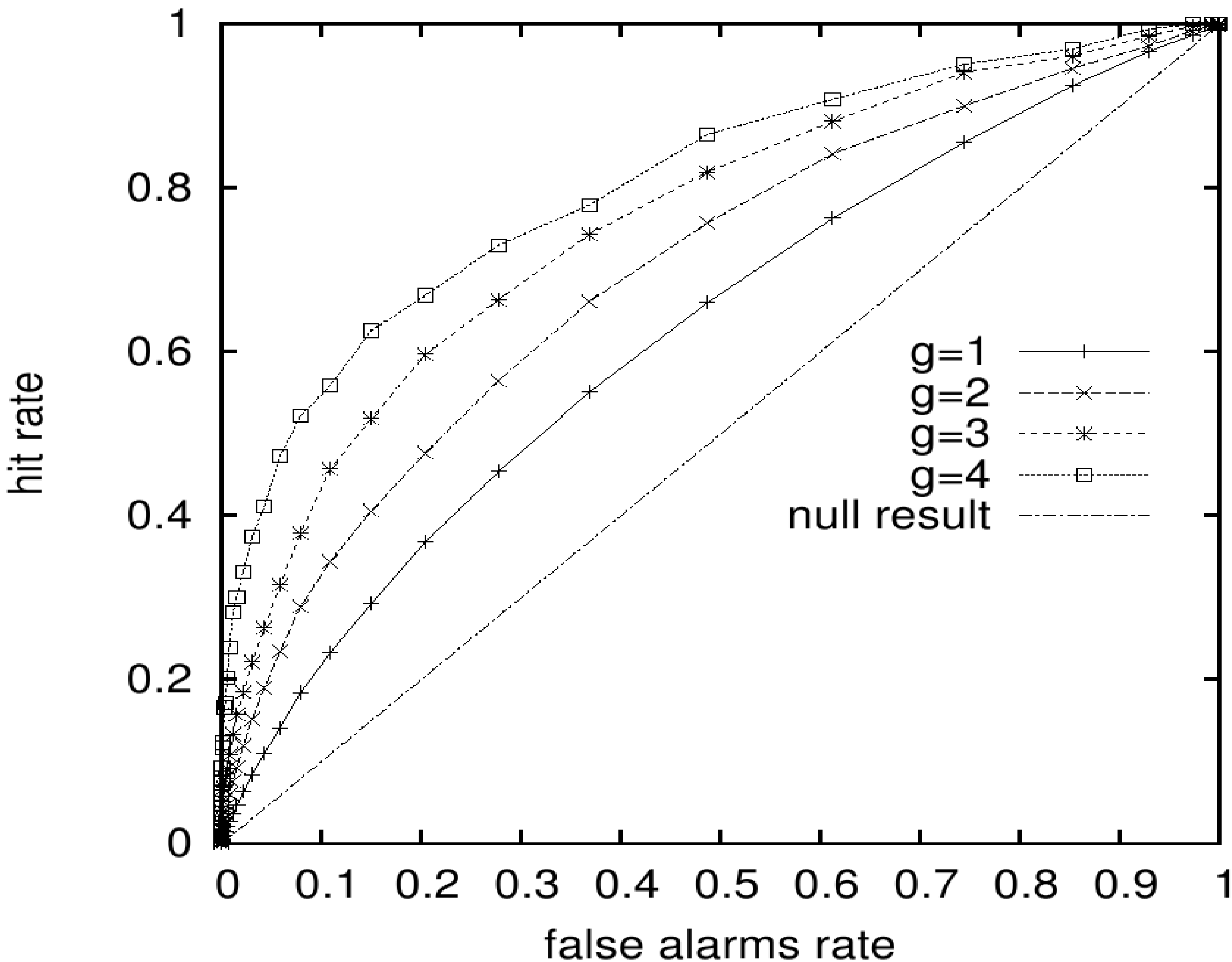
Photos from http://members.aol.com/fswemedien/ZZUnfalldatei.htm

I – Example 2: Wind Gusts



Kantz et al., Springer-Verlag (2006).

I – Example 2: Wind Gusts



Kantz et al., Springer-Verlag (2006).

Different prediction strategies

ξ: observation.

Ξ: extreme event.

•**<u>Strategy I</u>:**

maximize the *a posteriori* PDF $\rho(\xi \mid \Xi)$

(search Ξ look how the ξ were before)

•**<u>Strategy II:</u>**

maximize the likelihood PDF $\rho(\Xi \mid \xi)$

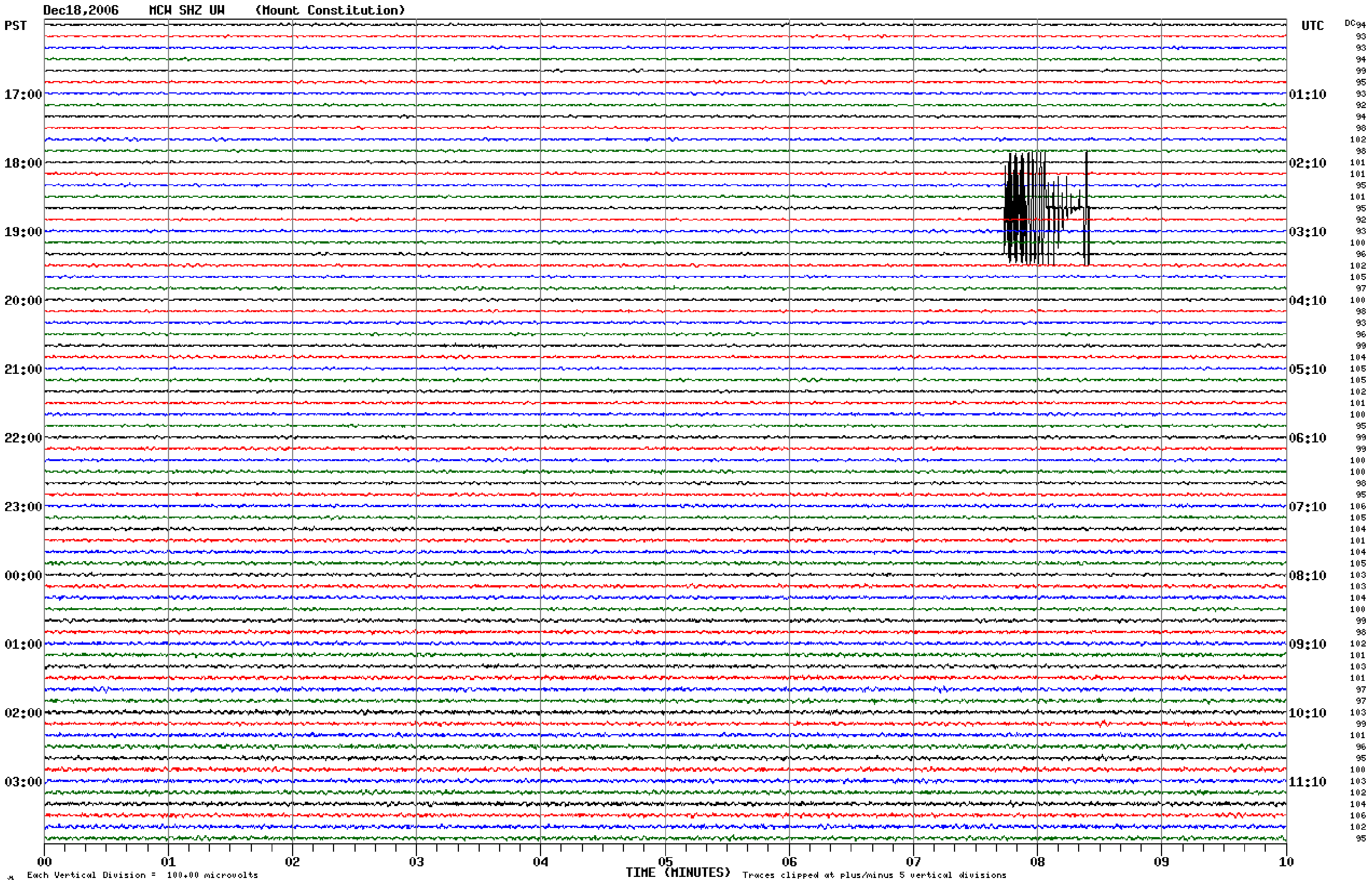
(for a given x look whether a Ξ follows)

Bayes theorem: $\rho(\xi, \Xi) = \rho(\xi \mid \Xi)\, \rho(\Xi) = \rho(\Xi \mid \xi)\, \rho(\xi)$

For wind speed and ARMA process <u>Strategy II</u> is superior.

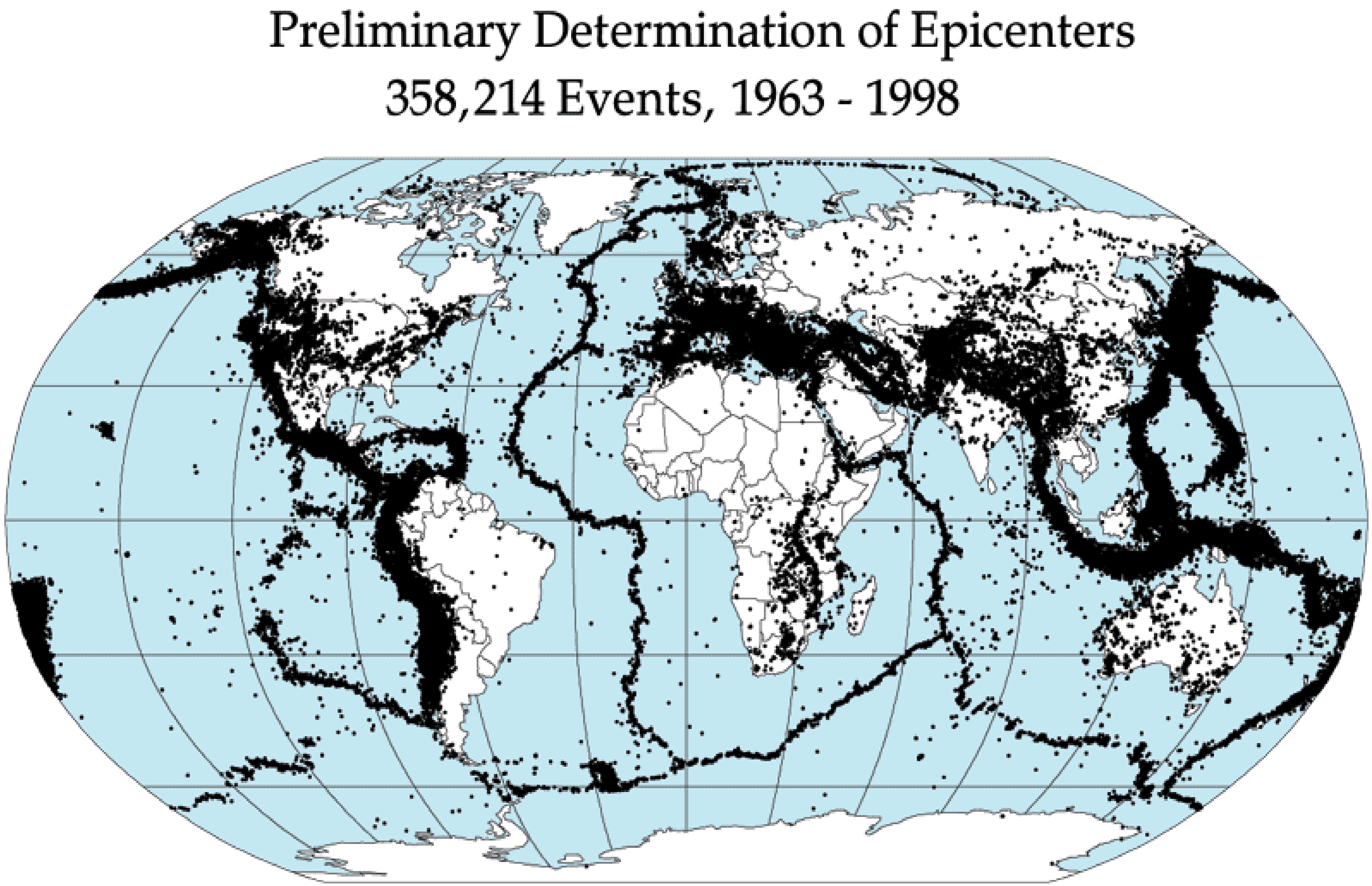
Phys. Rev. E 75, 016706 (2007).

I – Example 3: Earthquakes



http://www.pnsn.org

I – Example 3: Earthquakes



http://www.wikipedia.org

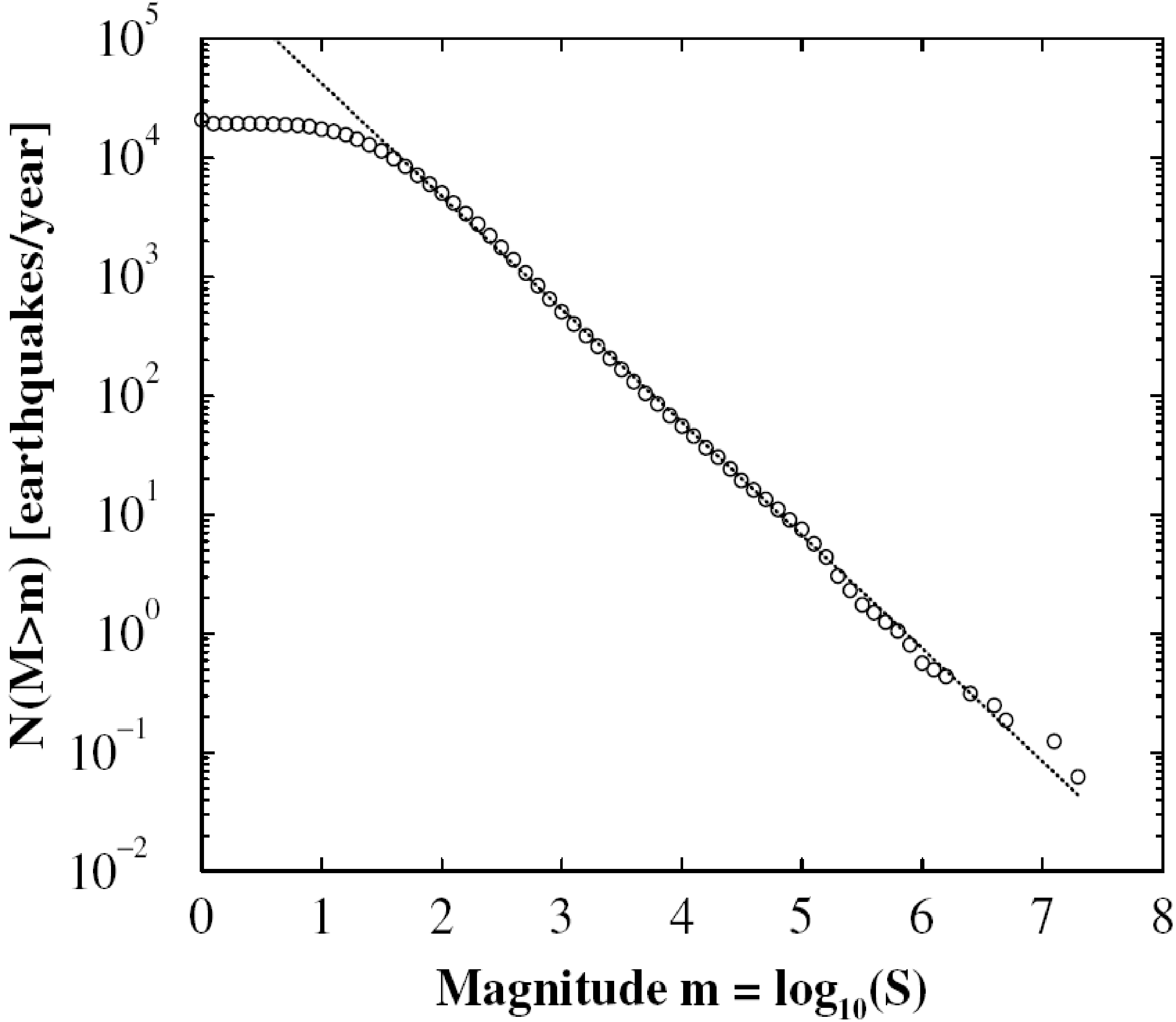


FIG. 1. The number of earthquakes $N(M > m)$ with a magnitude larger than m per year (open circles). The dashed line is the Gutenberg-Richter law $\log_{10} N(M > m) \propto -bm, b = 0.95$.

Bak et al PRL (2002).

I - Introduction

- Reasons for an unified treatment of extreme events:
  - **statistical characterization of the phenomena.**
  - prediction procedures.
  - risk estimations and precautions.

  See Book: "*Extreme events in nature and society*", Albeverio, Jentsch, Kantz (Eds.) Springer, (2006).

I - Introduction

- Reasons for an unified treatment of extreme events:
  - **statistical characterization of the phenomena.**
  - prediction procedures.
  - risk estimations and precautions.

  See Book: "*Extreme events in nature and society*", Albeverio, Jentsch, Kantz (Eds.) Springer, (2006).

- Definition of extreme events?
  Some necessary conditions:

Large	(in some relevant observable)
Rare	(otherwise adaptation)
Unexpected	(otherwise precaution)
Harmful	(otherwise irrelevant)

I - Introduction

- Reasons for an unified treatment of extreme events:
    - **statistical characterization of the phenomena.**
    - prediction procedures.
    - risk estimations and precautions.

  See Book: "*Extreme events in nature and society*", Albeverio, Jentsch, Kantz (Eds.) Springer, (2006).

- Definition of extreme events?
  Some necessary conditions:

Large	(in some relevant observable)
Rare	(otherwise adaptation)
Unexpected	(otherwise precaution)
Harmful	(otherwise irrelevant)

- Classification of extreme events:
  extrinsic X intrinsic origin (own dynamics).
  recurrent X non-recurrent events.

II – Recurrences in Physics

A fundamental concept and a statistical tool

II – Recurrences in Physics

- Poincaré recurrence theorem (1892): all trajectories of closed Hamiltonian systems return an infinite number of times to their initial conditions.

II – Recurrences in Physics

- Poincaré recurrence theorem (1892): all trajectories of closed Hamiltonian systems return an infinite number of times to their initial conditions.
- Zermelo Paradox: the theorem above applied to the systems studied in statistical mechanics implies a violation of 2^{nd} law of Thermodynamics.

II – Recurrences in Physics

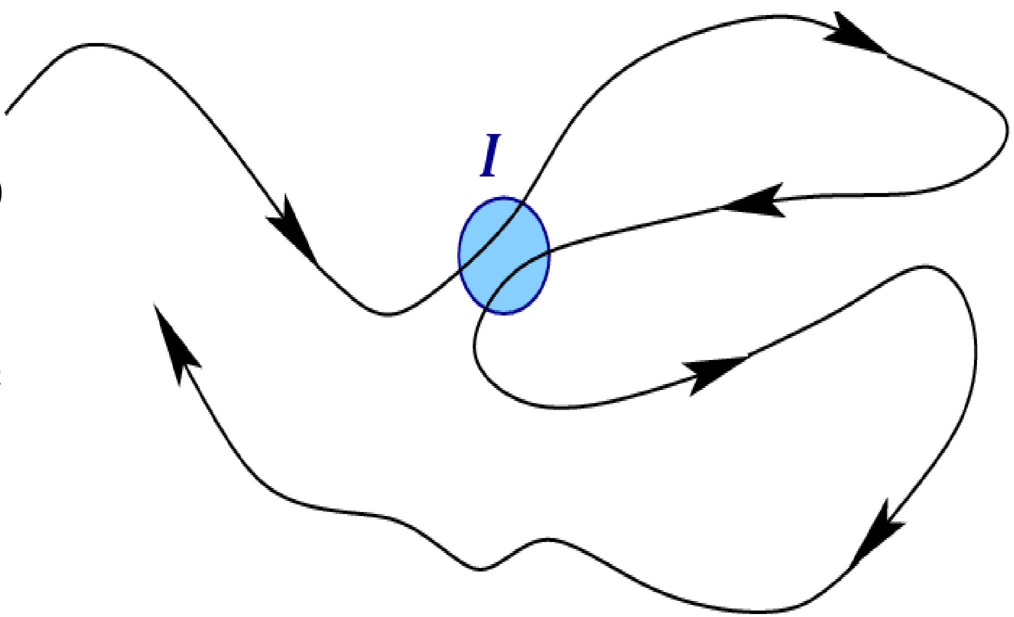
- Poincaré recurrence theorem (1892): all trajectories of closed Hamiltonian systems return an infinite number of times to their initial conditions.
- Zermelo Paradox: the theorem above applied to the systems studied in statistical mechanics implies a violation of 2^{nd} law of Thermodynamics.
- Boltzmann's solution: the mean recurrence time of a typical thermodynamical system is huge (statistical interpretation of the 2^{nd} law).

II – Recurrences in Physics

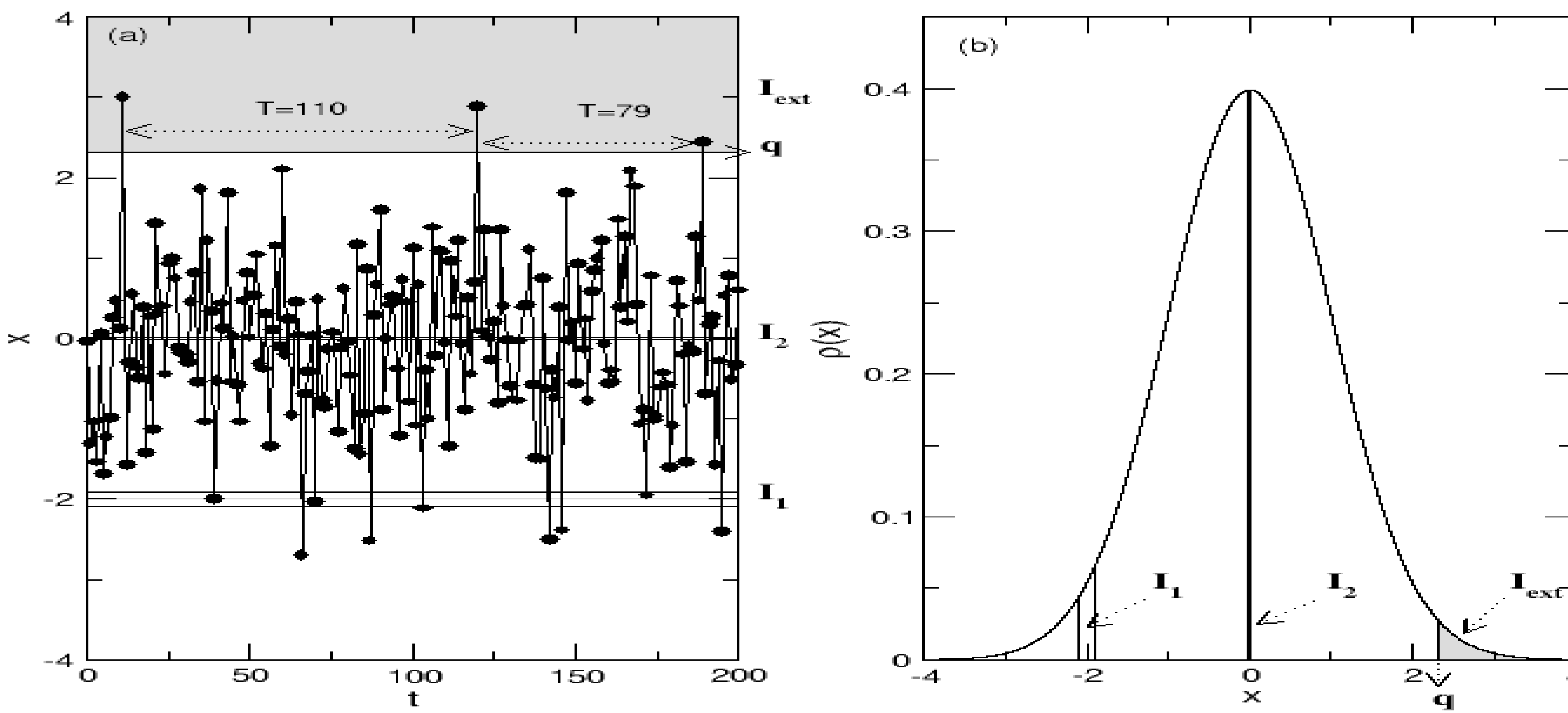
- Poincaré recurrence theorem (1892): all trajectories of closed Hamiltonian systems return an infinite number of times to their initial conditions.
- Zermelo Paradox: the theorem above applied to the systems studied in statistical mechanics implies a violation of 2nd law of Thermodynamics.
- Boltzmann's solution: the mean recurrence time of a typical thermodynamical system is huge (statistical interpretation of the 2nd law).

Poincaré recurrence (phase space)
X
recurrence in time series (1-dimension)

Projection of the high-dimensional phase space to a specific observable (e.g., the entropy).



II – Recurrences in time series



Where extreme and non-extreme recurrence intervals are:

$$I_{ext}(q) = [q, \infty[\qquad I(X_c, \delta) = [X_c - \delta, X_c + \delta].$$

And lead to sequences of recurrence times: $\{T_i : i = 1, 2, \ldots, \infty\}$

II – Statistical analysis of recurrences

Analysis of the sequence of recurrence times: $\{T_i : i = 1, 2, \ldots, \infty\}$

II – Statistical analysis of recurrences

Analysis of the sequence of recurrence times: $\{T_i : i = 1, 2, \ldots, \infty\}$

(1) The recurrence time distribution (RTD) denoted as P(T) is the probability of finding a recurrence time $T_i = T$ (or between T and T+dT).

(2) The mean recurrence time: $$\langle T \rangle \equiv \lim_{N_e \to \infty} \frac{1}{N_e} \sum_{1}^{N_e} T_i,$$

II – Statistical analysis of recurrences

Analysis of the sequence of recurrence times: $\{T_i : i = 1, 2, \ldots, \infty\}$

(1) The recurrence time distribution (RTD) denoted as P(T) is the probability of finding a recurrence time $T_i = T$ (or between T and T+dT).

(2) The mean recurrence time: $$\langle T \rangle \equiv \lim_{N_e \to \infty} \frac{1}{N_e} \sum_{1}^{N_e} T_i,$$

Total observation time: $$t = N_{total} \Delta t = N_{events} \langle T \rangle$$

The mean is determined by the recurrence interval: $$\langle T \rangle = \Delta t \frac{N_{total}}{N_{events}} \equiv \frac{\Delta t}{\mu(I)} = \frac{\Delta t}{\int_I \rho(x) dx}$$

II – Statistical analysis of recurrences

Two basic statistical properties of the recurrence times:

1. The mean recurrence time depends only on the PDF (x) and not on the temporal properties of the time series (Kac's Lemma).

2. The RTD depends only on the temporal properties of the time series and not on the PDF (x).

II – Statistical analysis of recurrences

Two basic statistical properties of the recurrence times:

1. The mean recurrence time depends only on the PDF (x) and not on the temporal properties of the time series (Kac's Lemma).

2. The RTD depends only on the temporal properties of the time series and not on the PDF (x).

For **{*x*}** being a sequence of uncorrelated random numbers (i.i.d.) (fixed probability of recurrence μ) a binomial/Poisson distribution is obtained:

$$P(T) = \mu(1-\mu)^{T-1} \Rightarrow P(T) = \mu e^{-\mu T} \text{ for } \mu \to 0$$

Usually exponential decay of correlations **=>** the RTD ***P(T)*** decays exponentially.

II – Examples

There are many different examples of application of such analysis:

- Climate data (e.g., temperature, water height).
- Turbulent data (e.g., solar flares): RT is called laminar phase between irregular bursts.
- Time series from neurons: RT is called interspike intervals.
- Stock market data.
- Earthquakes: RT is called interocurrence time.

For references see:

E.G.A. and H. Kantz, Phys. Rev. E **71**, 056106.

II – Example: earthquakes

The events are distributed according to the Gutenberg-Richter law:

$$\rho(M) \propto e^{-b \ln(10M})$$

By the previous results, we obtain as mean recurrence time between two earthquakes:

$$\langle T \rangle (M) = T_0 e^{b \ln(10) M_c}$$

$$T_0 \propto b \ln(10)/(1 - e^{-b \ln(10)\delta})$$

$$I(X_c, \delta) = [X_c - \delta, X_c + \delta].$$

II – Example: earthquakes

The events are distributed according to the Gutenberg-Richter law:

$$\rho(M) \propto e^{-b\ln(10M})$$

By the previous results, we obtain as mean recurrence time between two earthquakes:

$$\langle T\rangle(M) = T_0 e^{b\ln(10)M_c}$$

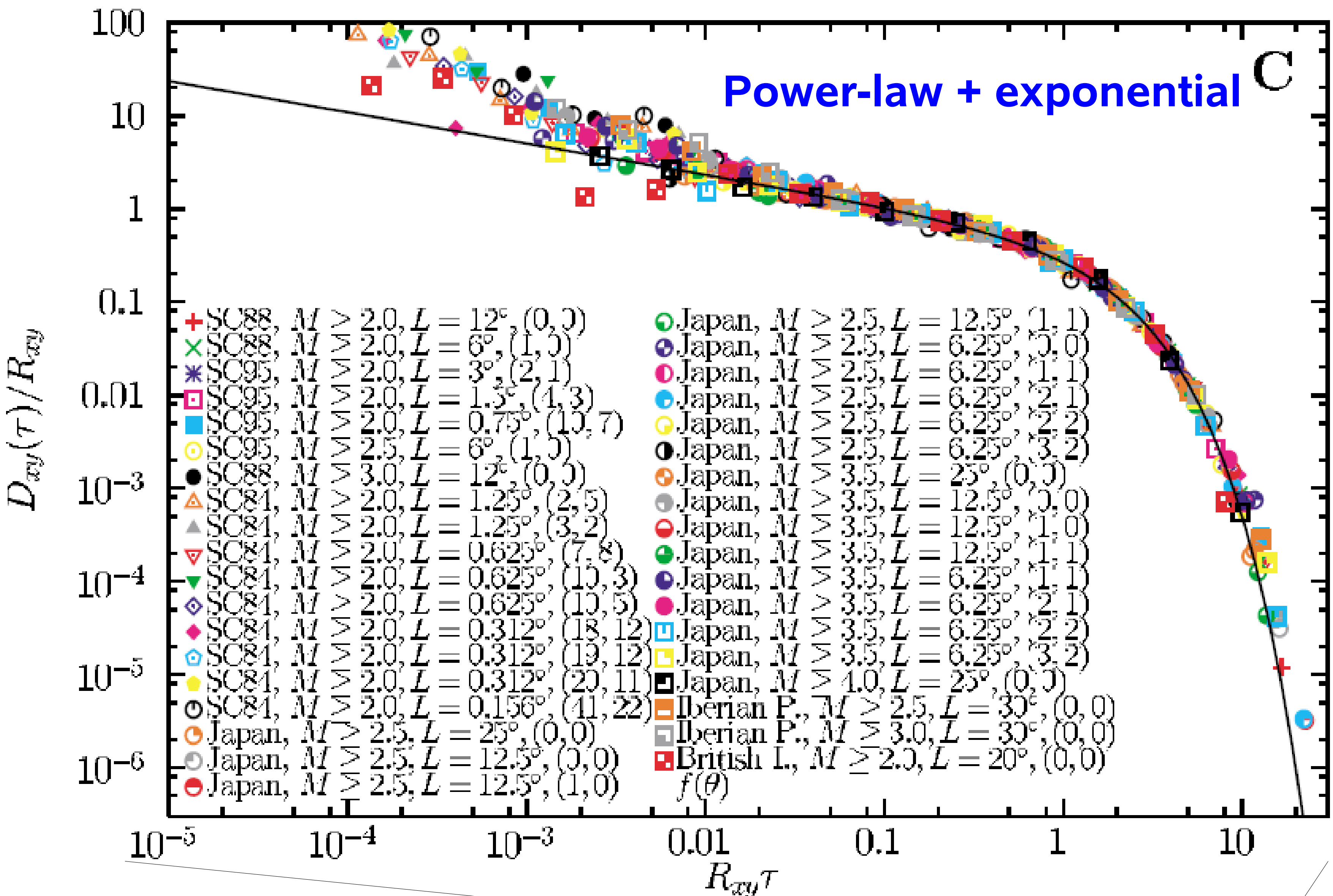
$$T_0 \propto b\ln(10)/(1 - e^{-b\ln(10)\delta})$$

$$I(X_c, \delta) = [X_c - \delta, X_c + \delta].$$

Comparing with previous results:

1. An equivalent "*remarkable*" result was obtained previously through different mean-field approximations in *[A. Sornette and D. Sornette, EPL (1989)].*

2. More confusing is the analysis of SOC models and earthquakes present in [Yang et al., PRL (2004)], where the PDF is incorrectly associated to the PDF of the sequence of earthquakes.



P. Bak *et al.*, Phys. Rev. Lett. **88**, 178501 (2002); A. Corral, *ibid.* **92**, 108501 (2004); N. Scafetta and B. J. West, *ibid.* **92**, 138501 (2004); J. Davidsen and C. Goltz, Geophys. Res. Lett. **31**, L21612 (2004).

III – Long-range correlated time series

Auto-correlation and the recurrence time distribution.

Reference: E.G. Altmann and H. Kantz, "*Recurrence time analysis, long-term correlation, and extreme events*", Phys. Rev. E 71, 056106 (2005).

III- Long-range correlated time series

We are interested in time series presenting long-range correlation:

$$C_x(s) = \langle x_i x_{i+s} \rangle = \frac{1}{N-s} \sum_{i=1}^{N-s} x_i x_{i+s} \sim s^{-\gamma_c}. \qquad 0 < \gamma_c < 1$$

What corresponds to a power-sepctrum: $S(\omega) \sim \omega^{-\beta}$, $\beta = 1 - \gamma_c$.

Observed, e.g., in financial data, meteorological and climatological records, turbulence data, physicological records, and DNA sequences.

III- Long-range correlated time series

We are interested in time series presenting long-range correlation:

$$C_x(s) = \langle x_i x_{i+s} \rangle = \frac{1}{N-s} \sum_{i=1}^{N-s} x_i x_{i+s} \sim s^{-\gamma_c}. \qquad 0 < \gamma_c < 1$$

What corresponds to a power-sepctrum: $S(\omega) \sim \omega^{-\beta},\ \beta = 1 - \gamma_c.$

Observed, e.g., in financial data, meteorological and climatological records, turbulence data, physicological records, and DNA sequences.

Gaussian distributed random variables with long-range correlation are generated through a Fourier transform technique [Prakash et al. PRA 1992].

Notice that defining the autocorrelation function of a Gaussian distributed stochastic process uniquely defines it: I will call it **linear process**.

III- Long-range correlated time series

We are interested in time series presenting long-range correlation:

$$C_x(s) = \langle x_i x_{i+s} \rangle = \frac{1}{N-s} \sum_{i=1}^{N-s} x_i x_{i+s} \sim s^{-\gamma_c}. \qquad 0 < \gamma_c < 1$$

What corresponds to a power-sepctrum: $S(\omega) \sim \omega^{-\beta}$, $\beta = 1 - \gamma_c$.

Observed, e.g., in financial data, meteorological and climatological records, turbulence data, physicological records, and DNA sequences.

Gaussian distributed random variables with long-range correlation are generated through a Fourier transform technique [Prakash et al. PRA 1992].

Notice that defining the autocorrelation function of a Gaussian distributed stochastic process uniquely defines it: I will call it **linear process**.

Exponential decay of correlations => exponential decay of the RTD.

Should we expect power-law decay of the recurrence time distribution (RTD)?

The effect of long-term correlations on the return periods of rare events

Armin Bunde[a], Jan F. Eichner[a], Shlomo Havlin[b], Jan W. Kantelhardt[a,*]

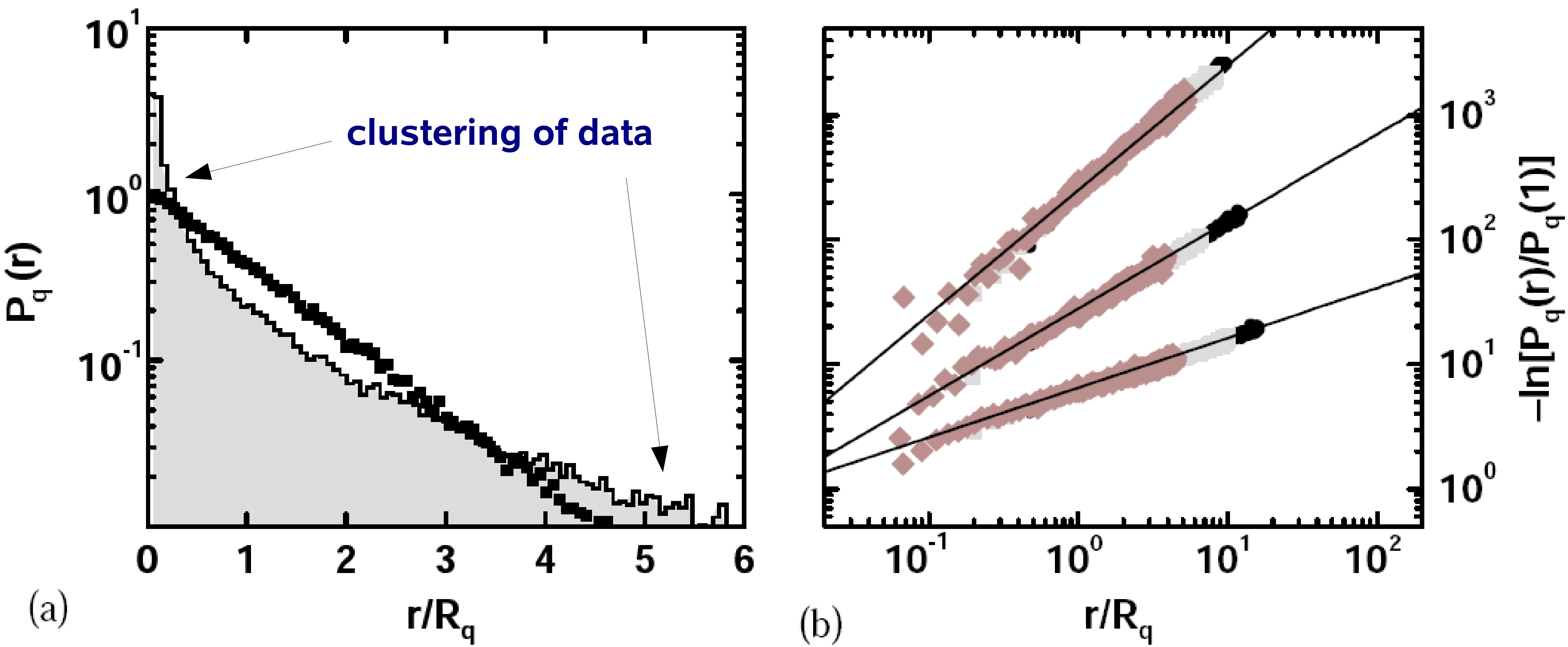
[a]Institut für Theoretische Physik III, Justus-Liebig-Universität, Giessen, Heinrich-Buff-Ring 16, Giessen 35392, Germany

[b]Department of Physics, Minerva Center, Bar-Ilan University, Ramat-Gan, Israel

Physica A 330 (2003) 1–7

Abstract

The basic assumption of common extreme value statistics is that different events in a time record are uncorrelated. In this case, the return intervals r_q of events above a given threshold size q are uncorrelated and follow the Poisson distribution. In recent years there is growing evidence that several hydro-meteorological and physiological records of interest (e.g. river flows, temperatures, heartbeat intervals) exhibit long-term correlations where the autocorrelation function decays as $C_x(s) \sim s^{-\gamma}$, with γ between 0 and 1. Here we study how the presence of long-term correlations changes the statistics of the return intervals r_q. We find that (a) the mean return intervals $R_q = \langle r_q \rangle$ are independent of γ, (b) the distribution of the r_q follows a stretched exponential, $\ln P_q(r) \sim -(r/R_q)^\gamma$, and (c) the return intervals are long-term correlated with an exponent γ' close to γ.



Similar results were observed in different data:

- Maximum daily temperature (Bunde et al. 2003).
- Wind speed velocity (Santhanam and Kantz 2004).
- Water height of the Nile (Bunde et al. 2005).
- Systems with intermediate correlations (Penetta 2006).
- Data with different non-Gaussian PDFs (Eichner et al. 2007).

III- The stretched exponential distribution

The stretched exponential distribution proposed by Bunde et al. :

$$P_{\gamma}(T) = ae^{-(bT)^{\gamma}}$$

III- The stretched exponential distribution

The stretched exponential distribution proposed by Bunde et al. :

$$P_\gamma(T) = ae^{-(bT)^\gamma}$$

Assuming that this distribution is valid for all times $T \in]0,\infty[$

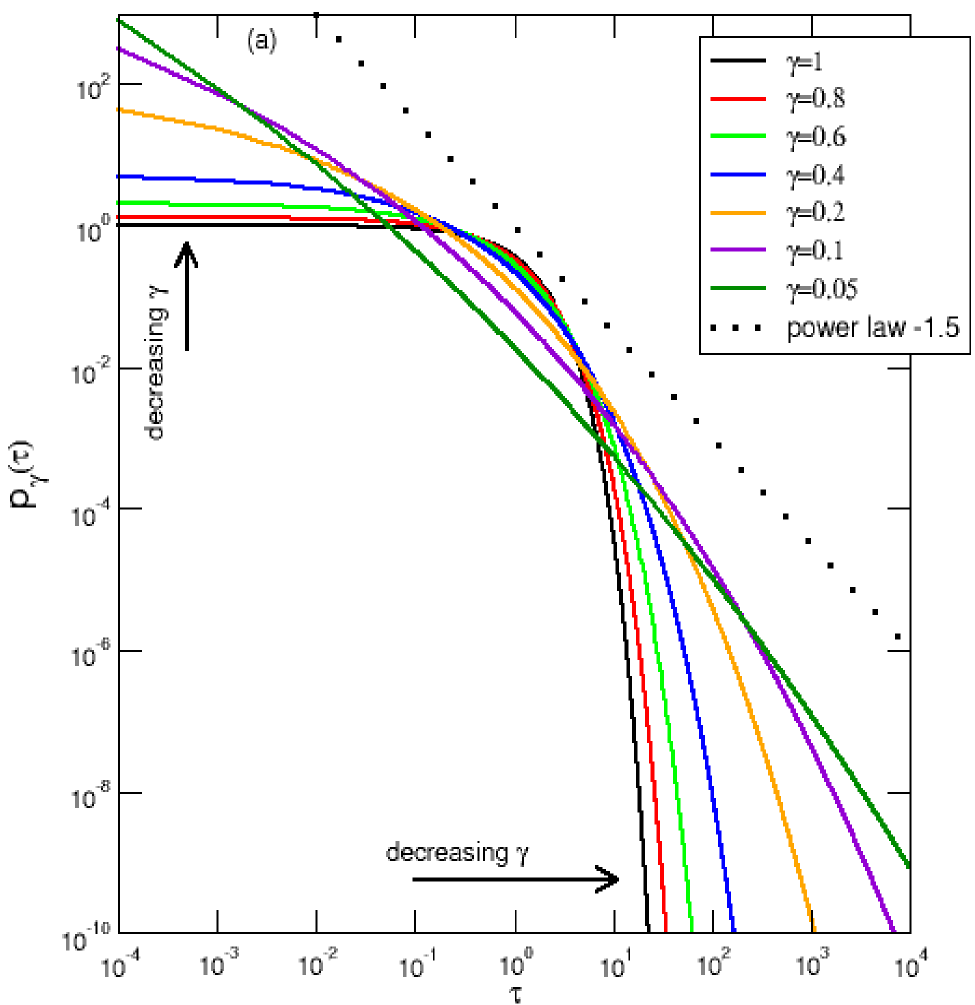
$$\int_0^\infty P(T)\,dT = 1, \qquad \langle T\rangle \equiv \int_0^\infty T P(T)\,dT = \frac{1}{\mu(I)}.$$

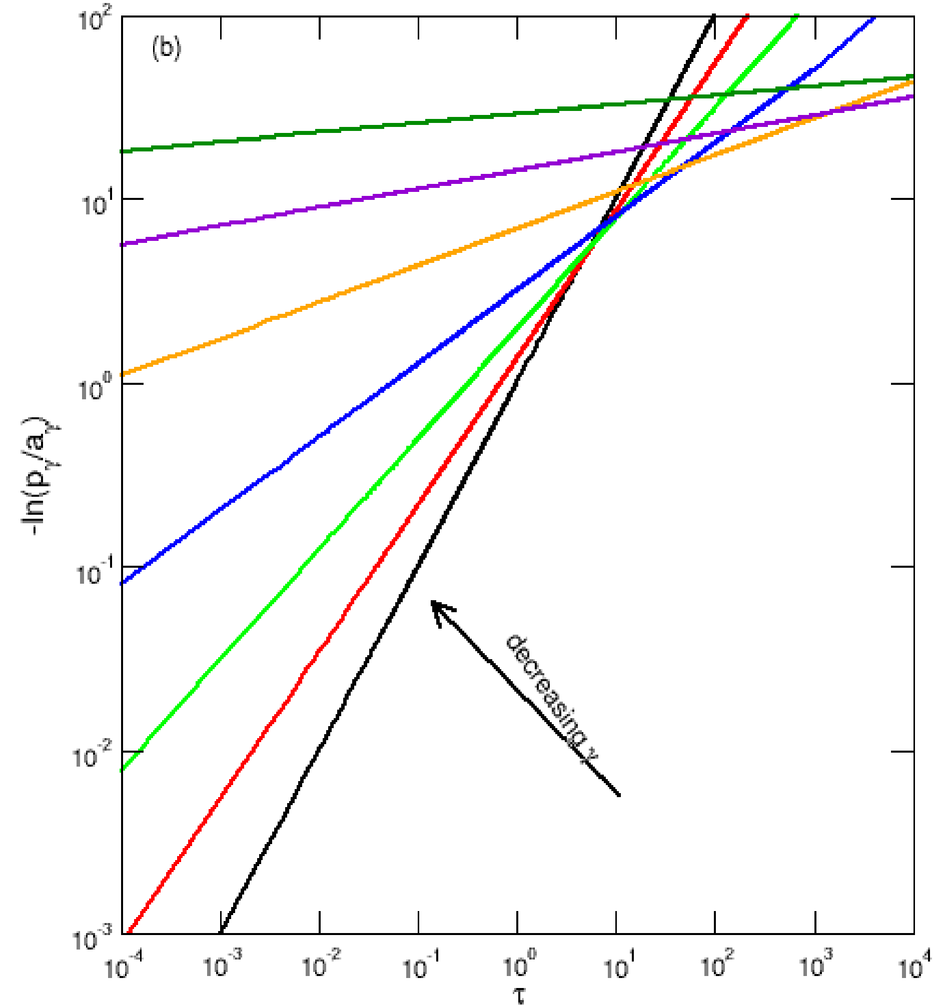
Counting the time as units of the average recurrence time $\tau = T/\langle T\rangle = \mu(I)\,T$,

$$p_\gamma(\tau) = a_\gamma e^{-(b_\gamma \tau)^\gamma}, \quad \text{with} \begin{cases} a_\gamma = b_\gamma \dfrac{\gamma}{\Gamma(1/\gamma)}, \\ b_\gamma = \dfrac{(2^{1/\gamma})^2 \Gamma\left(\dfrac{2+\gamma}{2\gamma}\right)}{2\sqrt{\pi}}, \end{cases}$$

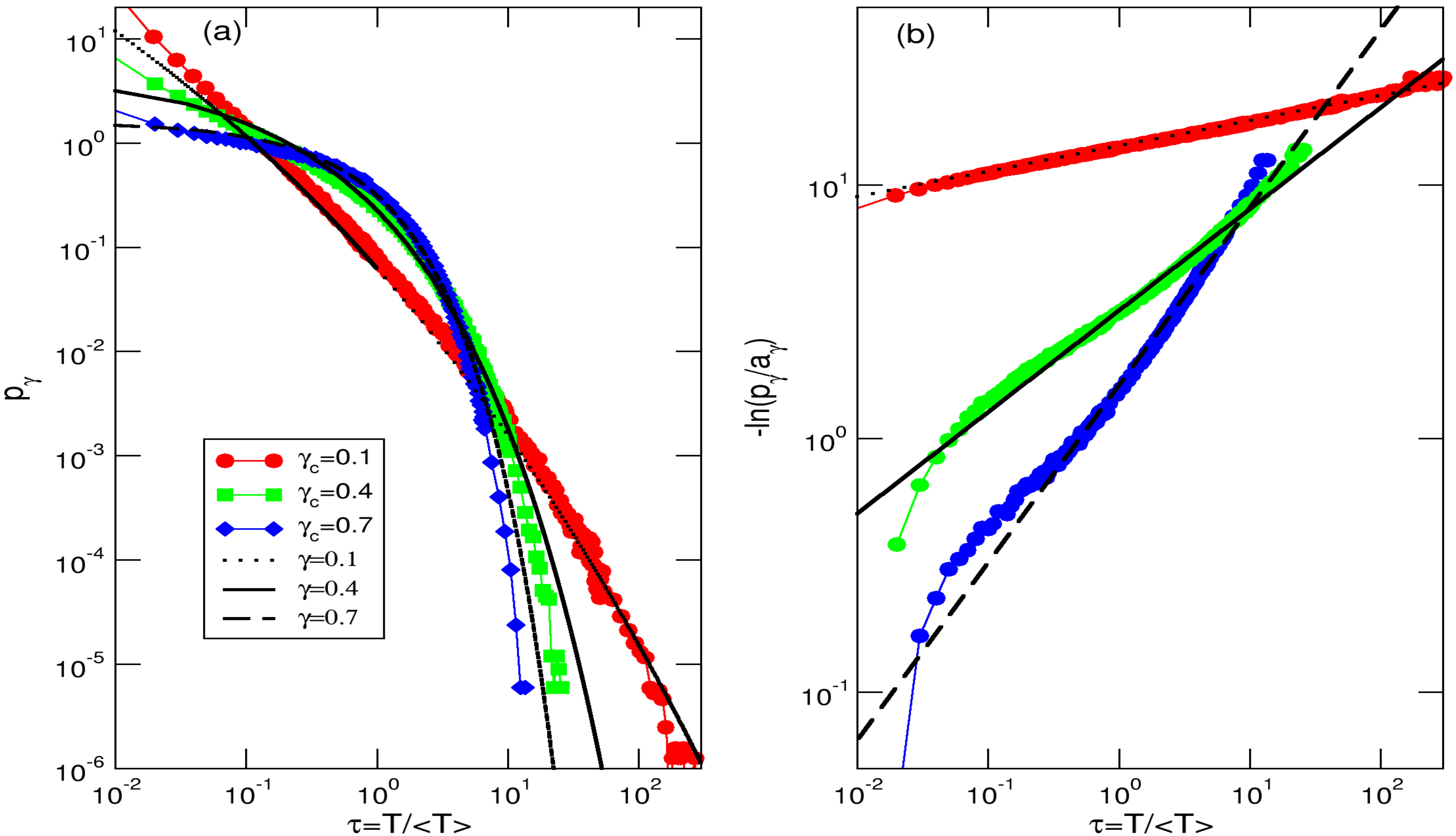
III- The stretched exponential distribution

$$p_\gamma(\tau) = a_\gamma e^{-(b_\gamma \tau)^\gamma}, \quad \text{with} \begin{cases} a_\gamma = b_\gamma \dfrac{\gamma}{\Gamma(1/\gamma)}, \\ b_\gamma = \dfrac{(2^{1/\gamma})^2 \Gamma\left(\dfrac{2+\gamma}{2\gamma}\right)}{2\sqrt{\pi}}, \end{cases}$$

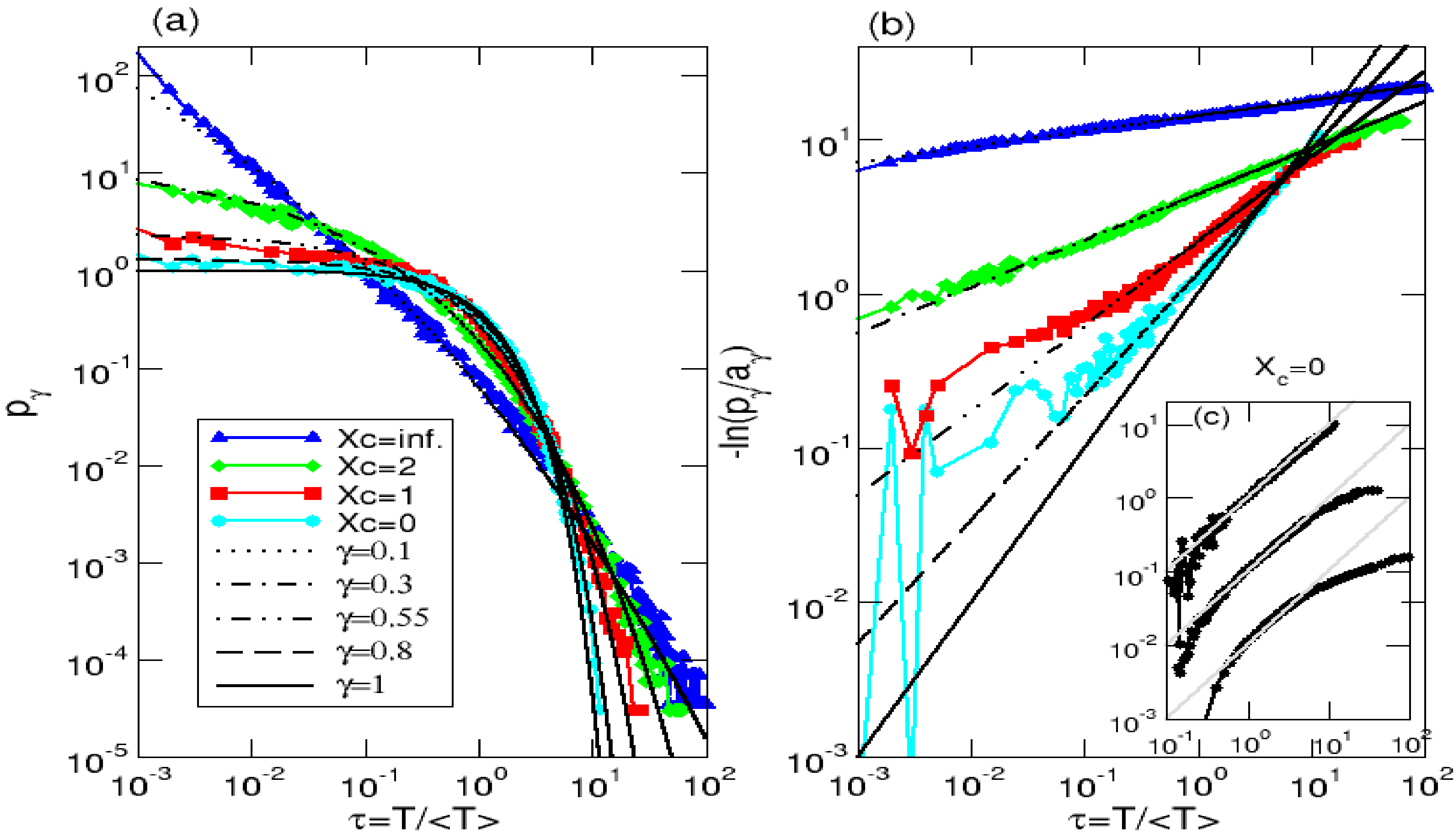




III- The stretched exponential distribution



III- The stretched exponential distribution



III- The stretched exponential distribution

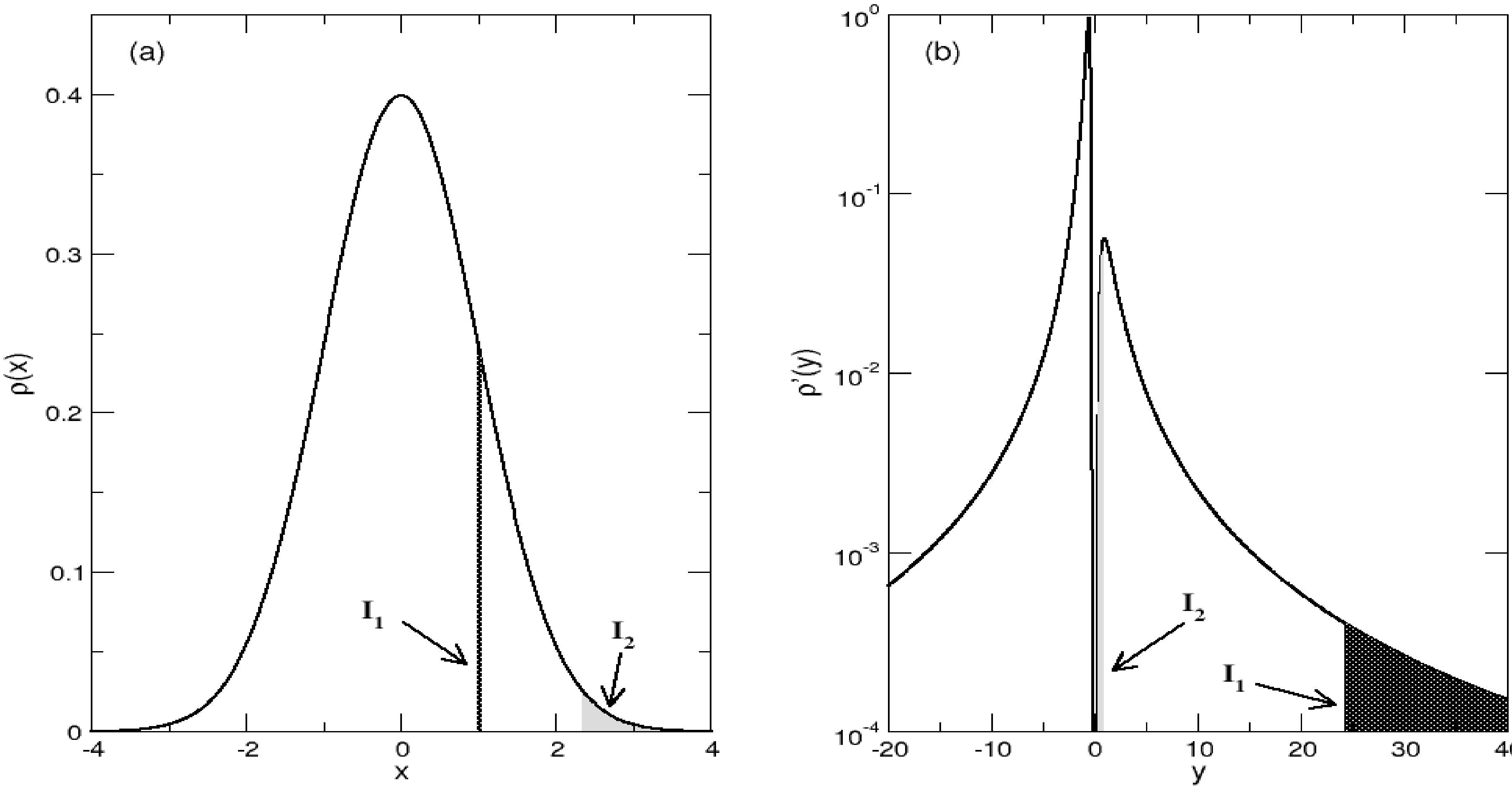
In summary good agreement for both kinds of recurrence intervals:

$$I_{ext}(q) = [q, \infty[\qquad\qquad I(X_c, \delta) = [X_c - \delta, X_c + \delta].$$

The single free parameter is given by:

$$\gamma = \begin{cases} \gamma_c, & \text{when } X_c \to \infty (\text{extreme}), \\ 1, & \text{when } X_c = 0. \end{cases}$$

III – Dependence on the observable



III – Dependence on the observable

- Long-range correlation concentrates on the extreme events.
- Different observables have different recurrence and correlation properties: they are **<u>not</u>** measures of the "system".
- Analogously:

  Climate:
  daily maximum temperature (long range).
  X
  rainfall (short range).

  Stock-market:
  fluctuation of prices (short range).
  X
  volatility (long-range).

IV – Effect of (human) reactions

What we do and what we should do

E.G. Altmann, S. Hallerberg, and H. Kantz, "*Reactions to extreme events: moving threshold model*", Physica A **364** p. 435 (2006).

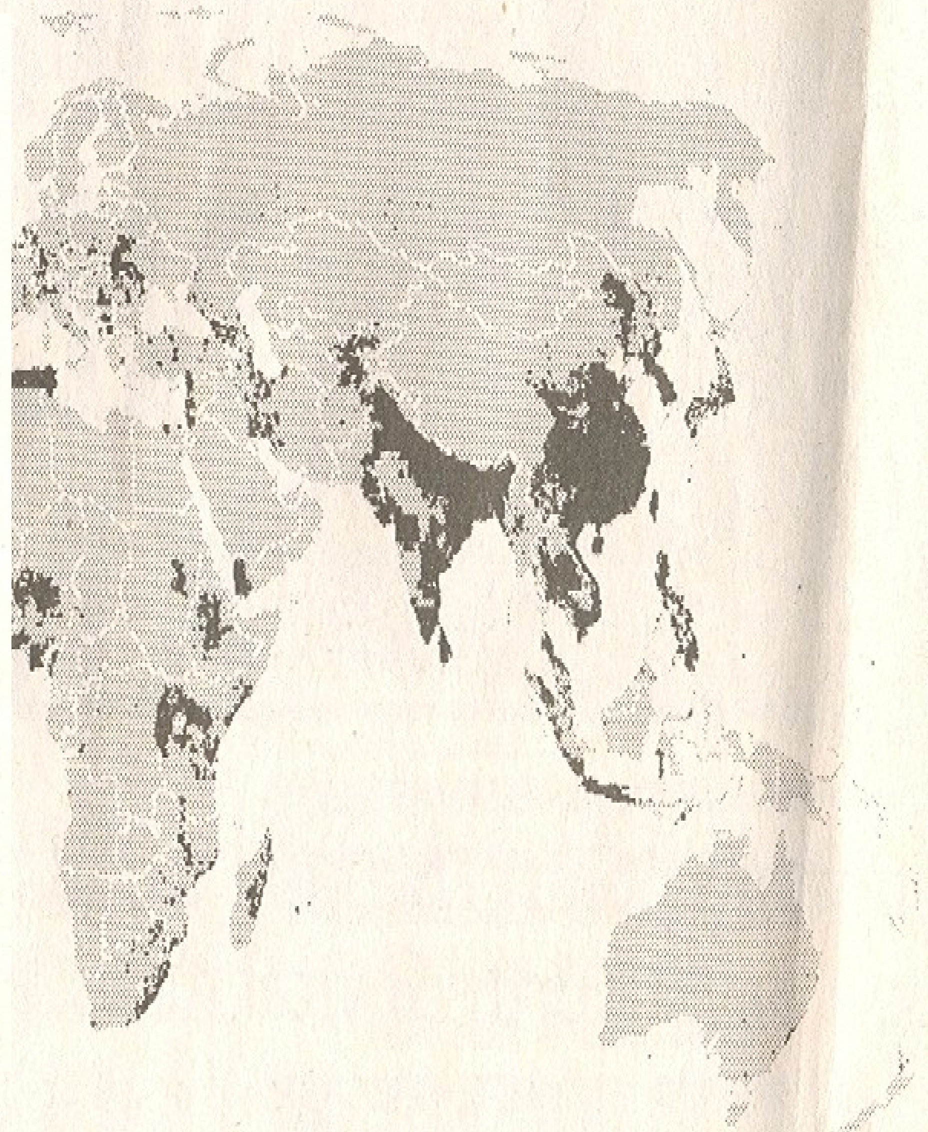
IV – Effect of (human) reactions

NY Times Jan. 2, 2005

IV – Effect of (human) reactions

10 Worst Floods

China	1931	**3,700,000**
China	1959	**2,000,000**
China	1939	**500,000**
China	1935	**142,000**
China	1911	**100,000**
China	1949	**57,000**
Guatemala	1949	**40,000**
China	1954	**30,000**
Venezuela	1999	**30,000**
Bangladesh	1974	**28,700**



NY Times Jan. 2, 2005

IV – Effect of (human) reactions

Usual human reactions to the occurrence/absence of extreme events:

- after an extreme event protection barriers are increased.
- if no extreme event happens for a long time protections are reduced.

IV – Effect of (human) reactions

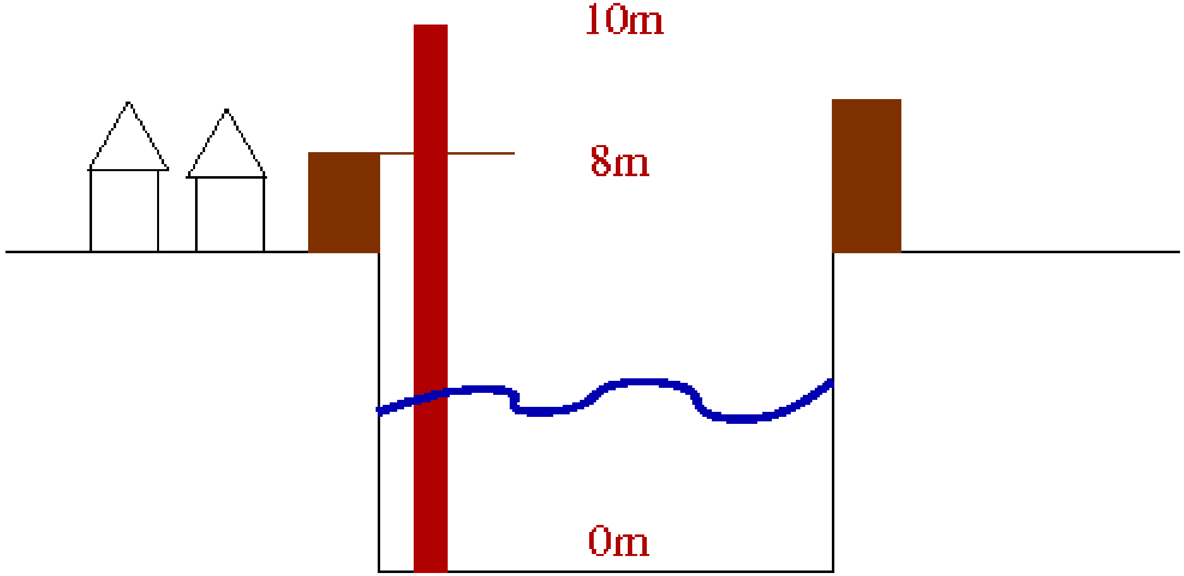
Usual human reactions to the occurrence/absence of extreme events:

- after an extreme event protection barriers are increased.
- if no extreme event happens for a long time protections are reduced.

Consider the example of **floods** in river:

observable: height of the water.

extreme event: if the water overpass the levees (protection barriers).



IV – Effect of (human) reactions

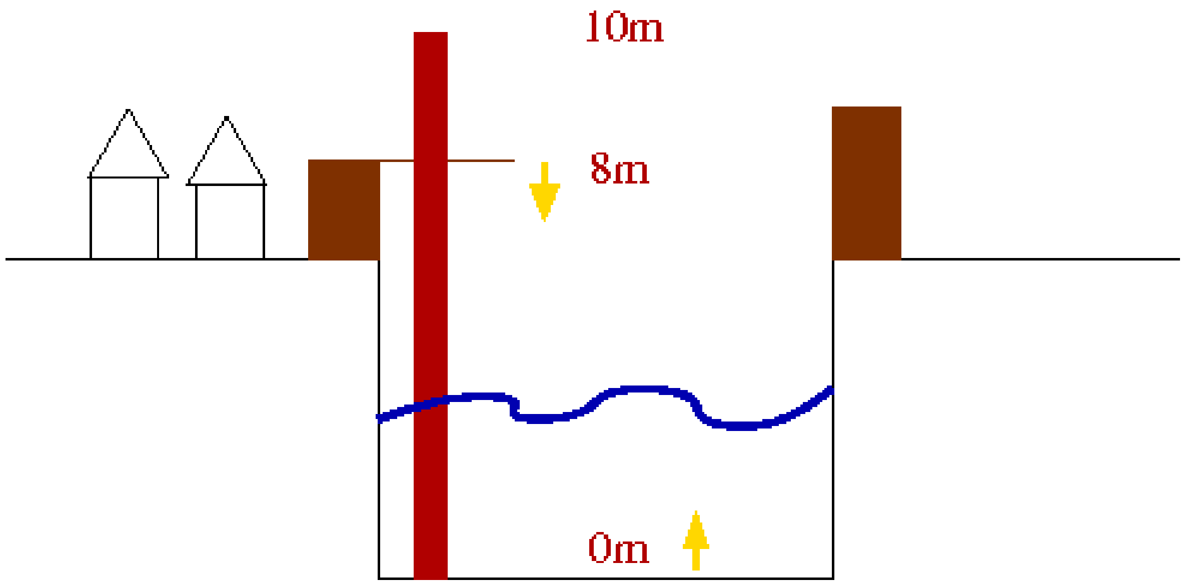
Usual human reactions to the occurrence/absence of extreme events:

- after an extreme event protection barriers are increased.
- if no extreme event happens for a long time protections are reduced.

Consider the example of **floods** in river:

observable: height of the water.

extreme event: if the water overpass the levees (protection barriers).



IV – Effect of (human) reactions

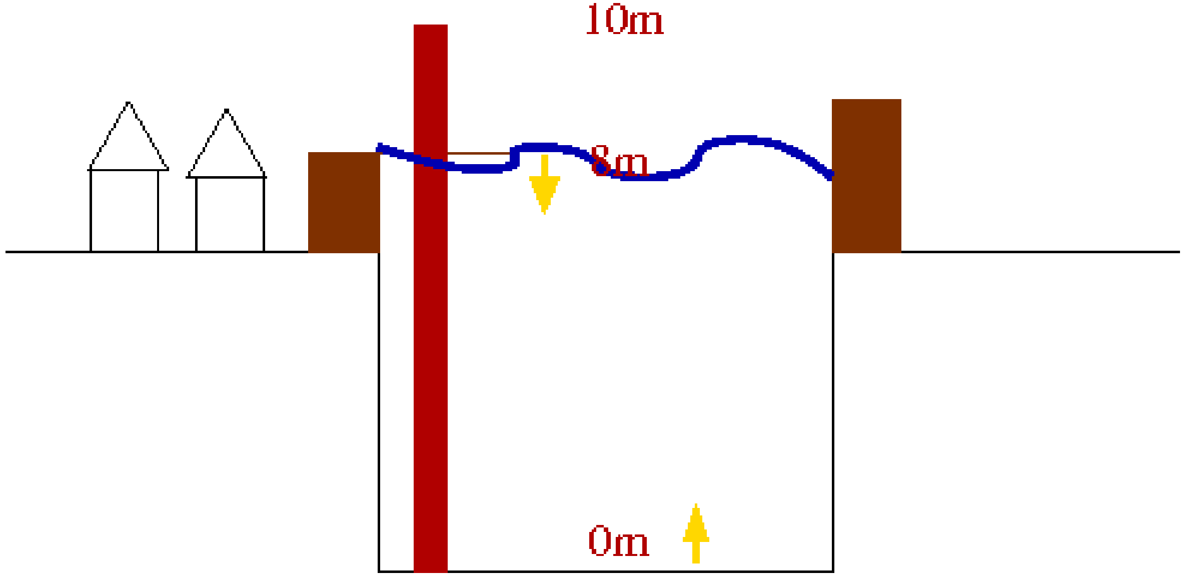
Usual human reactions to the occurrence/absence of extreme events:

- after an extreme event protection barriers are increased.
- if no extreme event happens for a long time protections are reduced.

Consider the example of **floods** in river:

observable: height of the water.

extreme event: if the water overpass the levees (protection barriers).



IV – Effect of (human) reactions

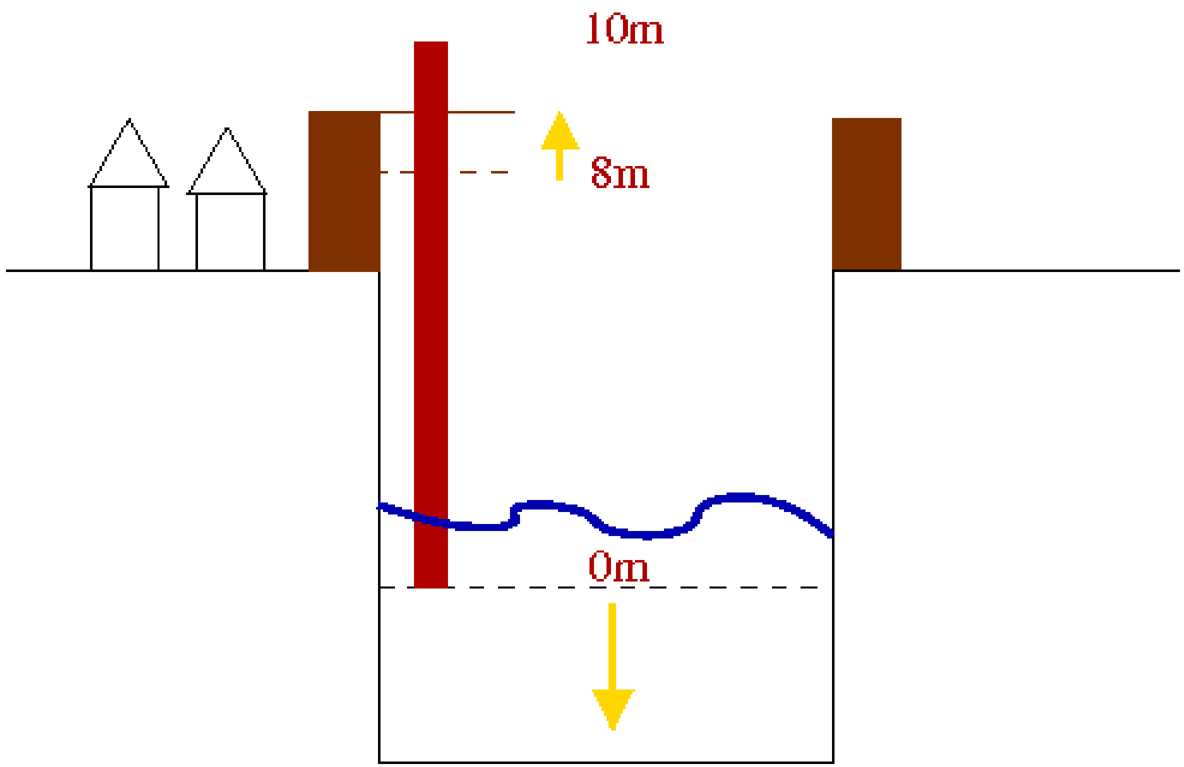
Usual human reactions to the occurrence/absence of extreme events:

- after an extreme event protection barriers are increased.
- if no extreme event happens for a long time protections are reduced.

Consider the example of **floods** in river:

observable: height of the water.

extreme event: if the water overpass the levees (protection barriers).



IV – Effect of (human) reactions

Questions:

Q1: Which is the influence of the reactions on our perception and on the occurrence of extreme events?

Q2: Which is the best method in order to reduce the number of extreme events?

IV – Effect of (human) reactions

Questions:

Q1: Which is the influence of the reactions on our perception and on the occurrence of extreme events?

Q2: Which is the best method in order to reduce the number of extreme events?

Class of systems we are interested in:

- Observable that compose a time series ξ (e.g., the height of the river).
- Time dependent size of the preventive barrier $\boldsymbol{q}$ (e.g., the height of the levees).

Condition for extreme event: if at a given time n^* we have $\xi > q$ we say that an extreme event of size $y = \xi - q$ occurred. (e.g., flood).

IV – Very simplified model that mimics the effects described before

$$q_{n+1} = \begin{cases} \max\{\alpha\xi_n, \beta q_n\} & \text{if } \xi_n > q_n, \text{ extreme event => barrier increase.} \\ \beta q_n & \text{if } \xi_n \leqslant q_n, \text{ normal event => barrier decrease.} \end{cases}$$

Control parameters are α and β:

0< β <1 (usually β close to 1).
0< α (usually 1< α, but close to 1).

For simplicity ξ is white noise.

IV - Answer to question 1:

Recurrence time (time between two extreme events)

Probability of having an extreme event at time *t*

$$r(t) = \int_{q_t}^{\infty} \rho(\xi) d\xi = \int_{q_t}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\xi^2}{2}} d\xi = \frac{1}{2} \mathrm{erfc}(\frac{\sqrt{2}}{2} q_t).$$

The recurrence time statistics (first extreme event at time T) is given by:

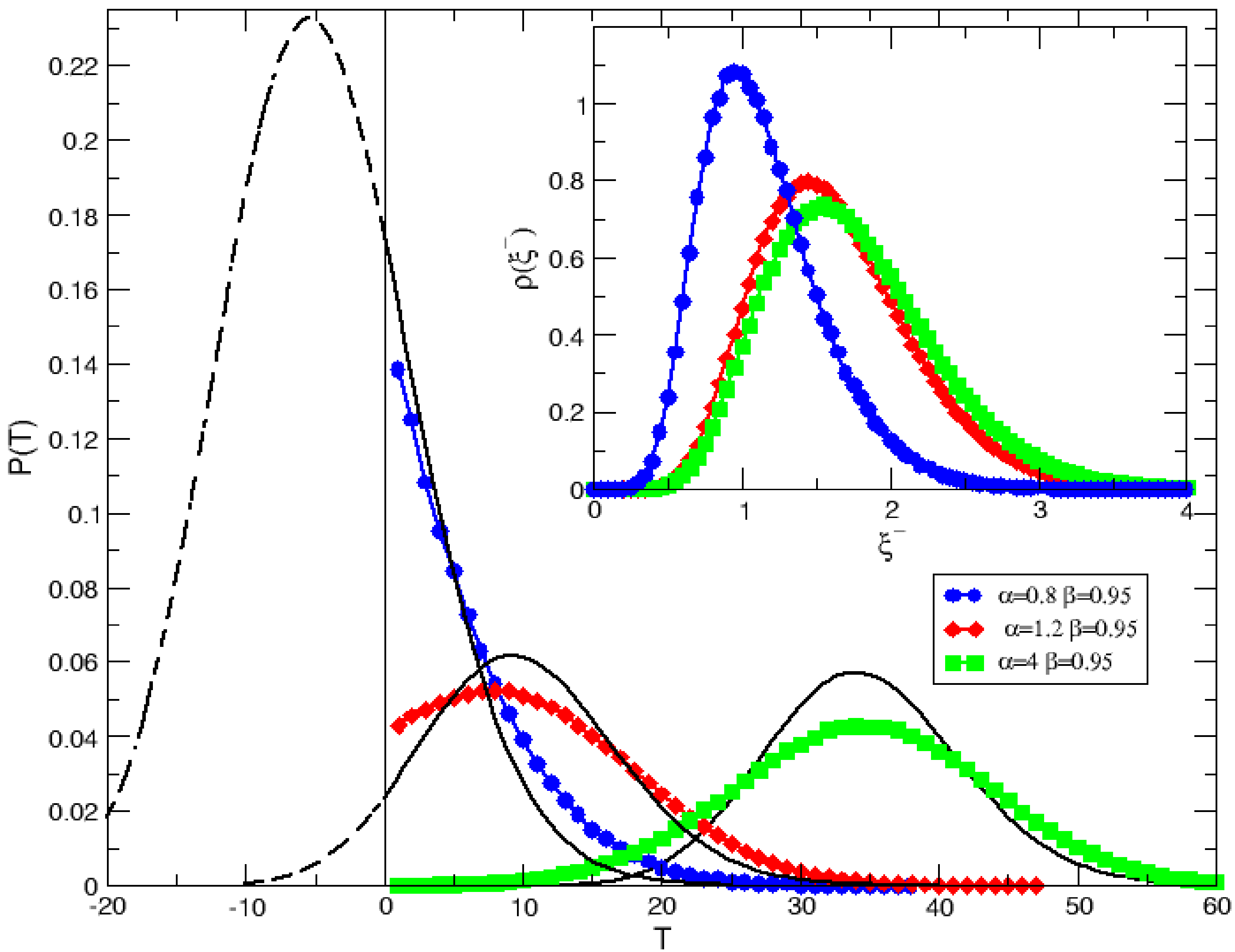
$$P(T) = C r(T) e^{-\int_0^T r(s) ds}.$$

$$P(T;\xi^-) = \frac{C}{2} \mathrm{erfc}(\frac{\sqrt{2}}{2} \alpha \xi^- \beta^T) \exp[-\frac{T}{2} - \frac{\sqrt{2} \alpha \xi^- \beta^T}{2 \ln(\beta) \sqrt{\pi}} \, {}_2F_2(\frac{1}{2}, \frac{1}{2}; \frac{3}{2}, \frac{3}{2}; -\frac{1}{2} (\alpha \xi^-)^2 \beta^{2T})].$$

$$P(T) = \int_0^{\infty} P(T;\xi^-) \rho(\xi^-) d\xi^-.$$

IV - Answer to Q1:

Recurrence time (time between two extreme events)



The maximum of *P(T)* scales as:

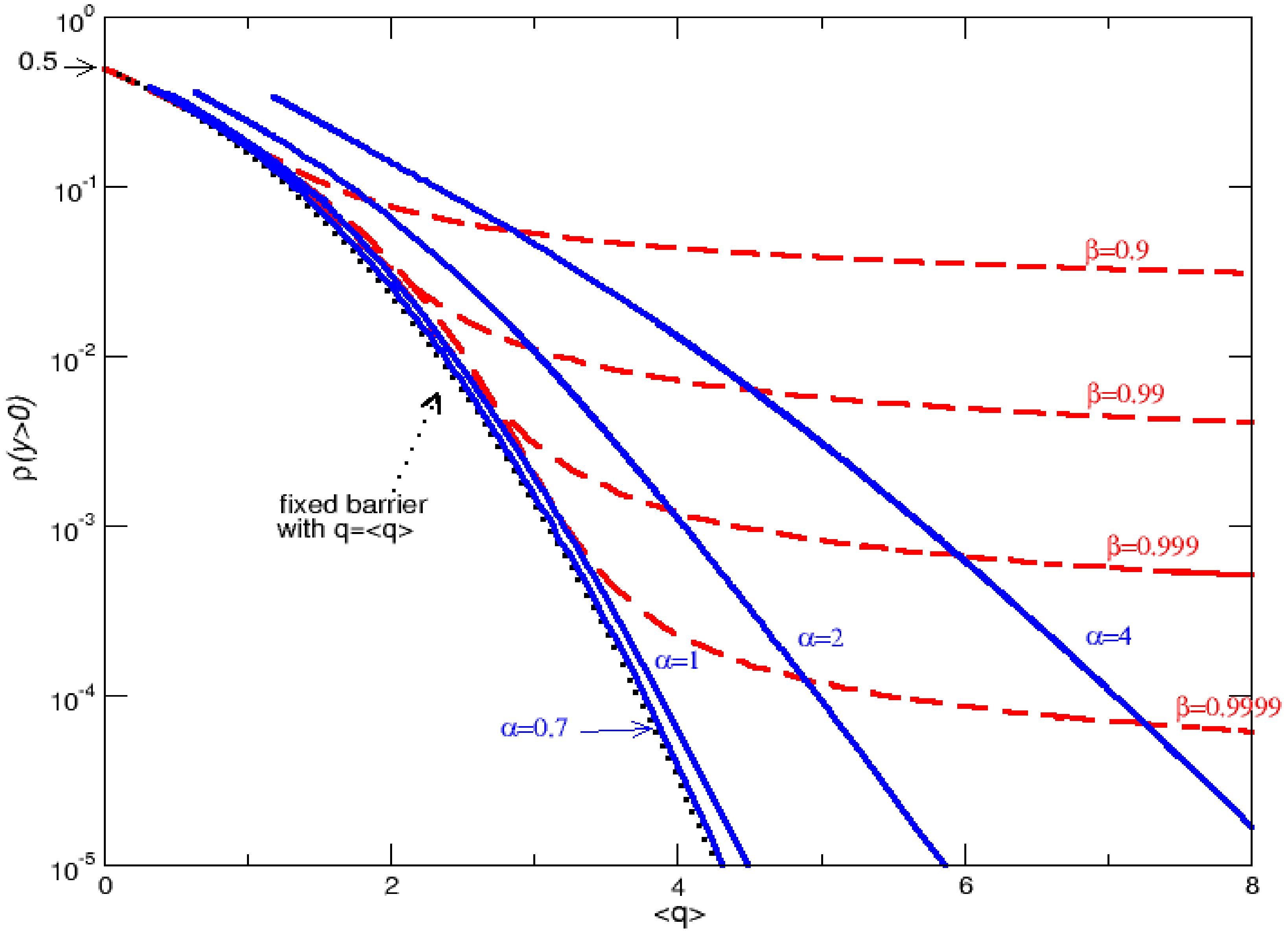
$$T^* = \ln(a)/\ln(\beta)$$

Human reactions introduce a periodicity on the occurrence of extreme events.

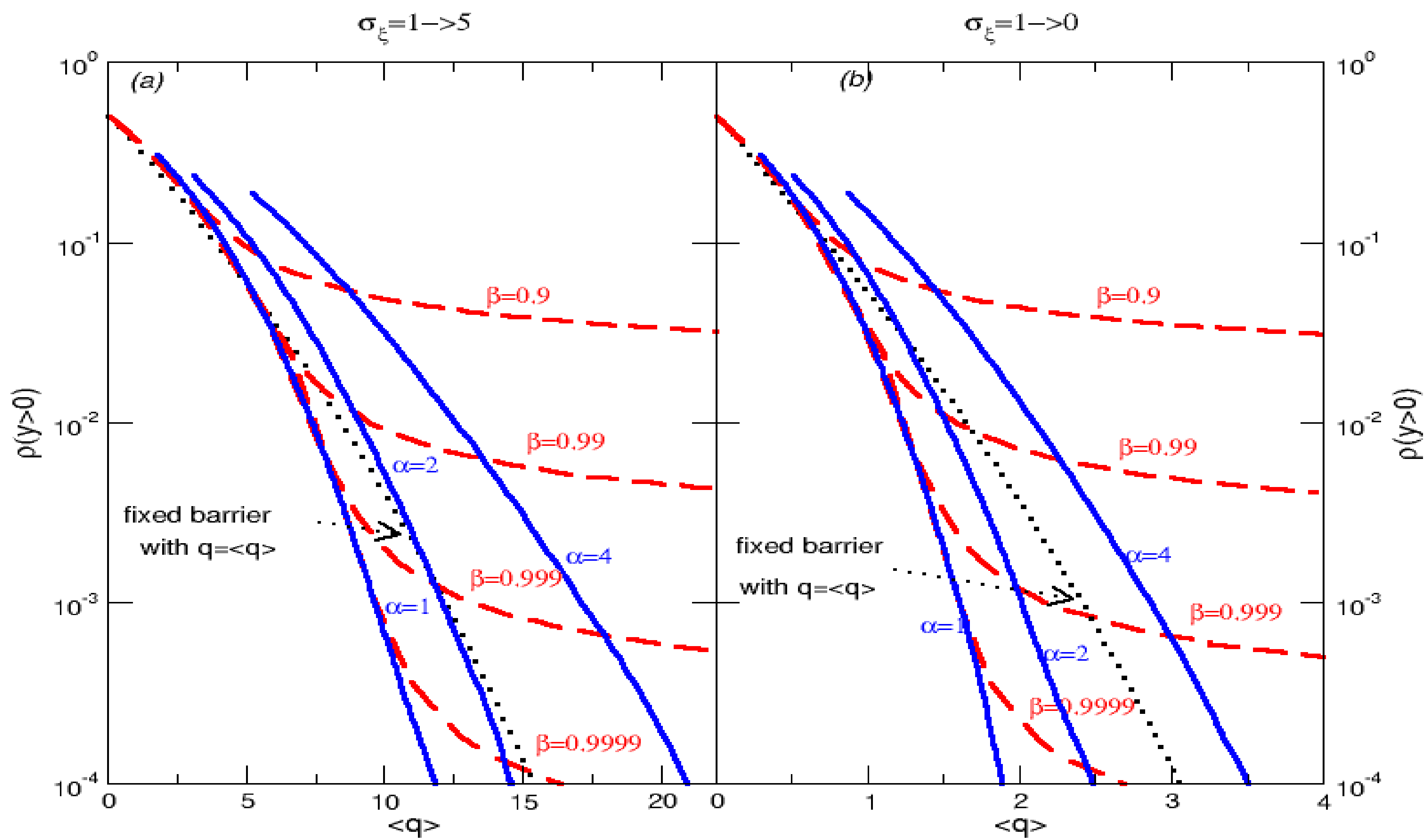
IV - Answer to Q2: Efficiency of the method

- Aim: Reduce the number of extreme events [$\boldsymbol{\rho(y>0)}$].
- Costs: mean value of the barrier <q>.
- Which are the optimal values of (α,β)?

IV - Answer to Q-2: Efficiency of the method



IV - Answer to Q-2: Non-stationary time series



V - Conclusions

V – Conclusions

1. Statistical characterization of the recurrences of extreme events is a general procedure.

   -$\langle T \rangle = 1/\mu(x)$ depends only on the PDF-(x) and on the rec. interval I.

   -The recurrence time distribution (RTD) is independent of the PDF-(x).

   - Uncorrelated and short-time correlated data show exponential decay of the RTD.

V – Conclusions

1. Statistical characterization of the recurrences of extreme events is a general procedure.
2. We have explored different properties of the statistics of recurrence times for long-range correlated time series.

- No unique correspondence between auto-correlation and RTD.

- A closed expression of the stretched exponential distribution shows good agreement with the numerical results for linear long-range correlated time series.

- The single free parameter *g* varies as

$$\gamma = \begin{cases} \gamma_c, & \text{when } X_c \to \infty (\text{extreme}), \\ 1, & \text{when } X_c = 0. \end{cases}$$

- RTD and auto-correlation function depend (independently) on the observable.

V – Conclusions

1. Statistical characterization of the recurrences of extreme events is a general procedure.
2. We have explored different properties of the statistics of recurrence times for linear long-range correlated time series.
3. The feed-back reaction to extreme events was investigated in a simple model of moving threshold were it was shown that:

   - **<u>periodicity</u>** in the occurrence of extreme events.

   - it is **<u>less efficient</u>** than maintaining the barrier constant: avoid the reduction of barriers in the quiet time (*sustainability*).

List of References

http://www.pks.mpg.de/~edugalt/

Presented in more detail in this seminar:

- E.G. Altmann and H. Kantz, "*Recurrence time analysis, long-term correlation, and extreme events*", Phys. Rev. E 71, 056106 (2005).
- E.G. Altmann, S. Hallerberg, and H. Kantz, "*Reactions to extreme events: moving threshold model*" Physica A 364 p. 435 (2006).

Prediction of extreme events (wind gusts):

- S. Hallerberg, E. G. Altmann, D. Holstein, and H. Kantz "*Precursors of extreme increments*", Phys. Rev. E 75, 016706 (2007).
- H. Kantz, E. G. Altmann, S. Hallerberg, D. Holstein, A. Riegert, "*Dynamical interpretation of extreme events: predictability and predictions*", chapter in the book "*Extreme Events in Nature and Society*", Springer (2006).