A recurrences based technique for detecting genuine extremes

Davide Faranda

CNRS – SPHYNX/SPEC - CEA
davide.faranda@cea.fr
–
joint work with: Sandro Vaienti

November 5, 2013

Did we use the right approach to extremes?

What we want?

- **Define** when an event is **extreme**.
- **Compute** how likely is **its recurrence** in a certain time.

Did we use the right approach to extremes?

What we want?

- **Define** when an event is **extreme**.
- **Compute** how likely is **its recurrence** in a certain time.

What we haven't done carefully?

- **We used a statistical theory for** indipendent and identically distributed ***i.i.d*** **variables for not** ***i.i.d.*** **variables**.
- **We did not check** that our data are sufficient to sample **the asymptotic regime of the statistical theory implied**.

Did we use the right approach to extremes?

What we want?

- **Define** when an event is **extreme**.
- **Compute** how likely is **its recurrence** in a certain time.

What we haven't done carefully?

- **We used a statistical theory for** indipendent and identically distributed ***i.i.d*** **variables for not *i.i.d.* variables**.
- **We did not check** that our data are sufficient to sample **the asymptotic regime of the statistical theory implied**.

What we could do?

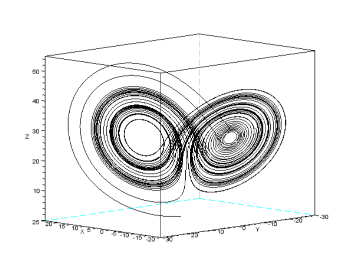
- **Link the results to asymptotic theories** verified for toy models.
- **Let the method ensure** that the **data are enough**.

Table of contents

Simple models and Complex phenomena

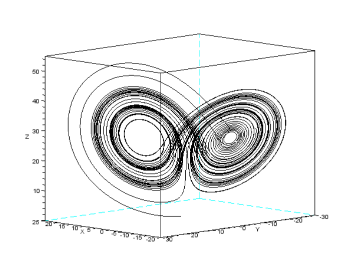
Simple models and Complex phenomena

Lorenz attractor



System of ODEs with well known Geometrical (attractor) and Dynamical properties.

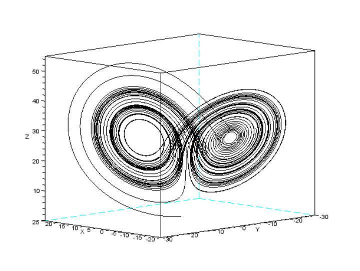
Lorenz attractor



System of ODEs with well known Geometrical (attractor) and Dynamical properties.

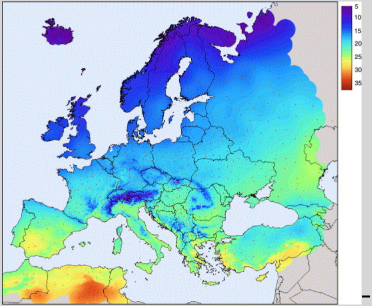
Simple models and Complex phenomena

Lorenz attractor



System of ODEs with well known Geometrical (attractor) and Dynamical properties.

Mean daily Temperatures



Complex system living on an **undetermined attractor**. Spatial-temporal chaos.

Common issues in matching models and data

For time series...

- The **attractor** is usually **unknown** i.e. we do not know the **dimension of the phase space**.
- Measurements are perturbed by **instrument errors**.
- Traditionally: Pure **statistical approach** for extracting information on the dynamics.

Common issues in matching models and data

For time series...

- The **attractor** is usually **unknown** i.e. we do not know the **dimension of the phase space**.
- Measurements are perturbed by **instrument errors**.
- **Traditionally**: Pure **statistical approach** for extracting information on the dynamics.

However...

- **Statistical properties** should match some **quantitative properties** of the systems.
- **Low dimensional models** usually provide interesting information about **the dynamics of much complicated systems.**

Goal and instruments

Goal

- **What?** Discriminate real extreme events from normal variability.
- **How?** By defining a rigorous approach to the study the recurrences in time series.
- **Why?** Because the classical definitions are not linked to the dynamics.

Goal and instruments

Goal

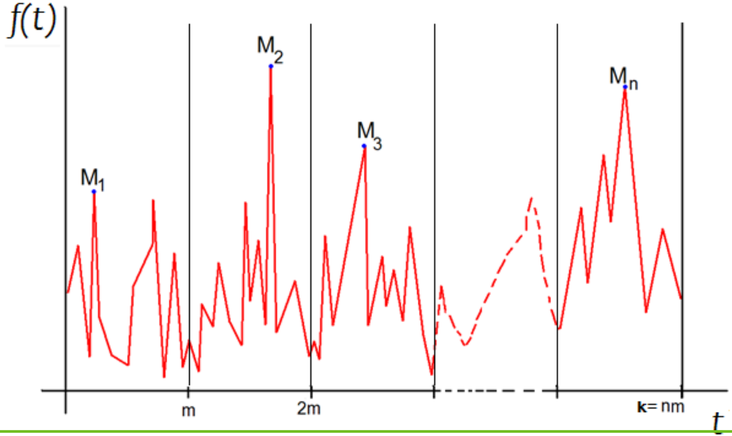
- **What?** Discriminate real extreme events from normal variability.
- **How?** By defining a rigorous approach to the study the recurrences in time series.
- **Why?** Because the classical definitions are not linked to the dynamics.

Theoretical instruments

- **What?** Extreme Value Theory (EVT) of recurrences for dynamical systems
- **How?** By studying the recurrences of time series like they are output of perturbed dynamical systems.
- **Why?** Because this technique provides a self-consistent check of the convergence to the expected values and does not depend on the phase space dimension.

How to select the Extremes

Time series whose maxima can be extracted at **fixed time intervals** e.g. annual maxima (minima) or daily maxima (minima).



Generalized Extreme Value (GEV) distribution

For **independent and identically distributed variables** (i.i.d.), the cumulative distribution of maxima M_n converges, asymptotically, to the **GEV distribution** (Gnedenko (1943)):

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

which holds for $1 + \xi(x - \mu)/\sigma > 0$.

- μ is the **location parameter**.

Generalized Extreme Value (GEV) distribution

For **independent and identically distributed variables** (i.i.d.), the cumulative distribution of maxima M_n converges, asymptotically, to the **GEV distribution** (Gnedenko (1943)):

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

which holds for $1+\xi(x-\mu)/\sigma > 0$.

- μ is the **location parameter**.
- σ is the **scale parameter**.

Generalized Extreme Value (GEV) distribution

For **independent and identically distributed variables** (i.i.d.), the cumulative distribution of maxima M_n converges, asymptotically, to the **GEV distribution** (Gnedenko (1943)):

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

which holds for $1 + \xi(x-\mu)/\sigma > 0$.

- μ is the **location parameter**.
- σ is the **scale parameter**.
- ξ discriminates the **type of distribution**:
  $\xi \to 0$: Type 1 (Exponential tails) - **Gumbel**
  $\xi > 0$: Type 2 (Power law tails) - **Frechét**
  $\xi < 0$: Type 3 (Tails bounded from above) - **Weibull**.

Consider a **discrete time dynamical systems**:

$x_{t+1} = T(x_t)$ with $t = 0, 1, 2, ..$ and initial condition x_0.

Mixing Stochastic processes

whose independence among maxima is guaranteed by existence of **mixing conditions**, allow for replacing i.i.d variables

Consider a **discrete time dynamical systems**:

$x_{t+1} = T(x_t)$ with $t = 0, 1, 2, ..$ and initial condition x_0.

Mixing Stochastic processes

whose independence among maxima is guaranteed by existence of **mixing conditions**, allow for replacing i.i.d variables

Observables of the distances among points of the trajectories

ensure convergences towards the **three different extreme value laws** through special functional forms.

Good observables are functions of the distances in the phase space (Collet, Freitas-Freitas, Todd), e.g.:

$$g_i(y,t) = -\log(\text{dist}(y,\zeta))$$

where $y = T^t(x_0)$ and the point ζ is chosen on the attractor.

- **The observable function must achieve a global maximum at ζ and be monotonically decreasing.**

Good observables are functions of the distances in the phase space (Collet, Freitas-Freitas, Todd), e.g.:

$$g_i(y, t) = -\log(\text{dist}(y, \zeta))$$

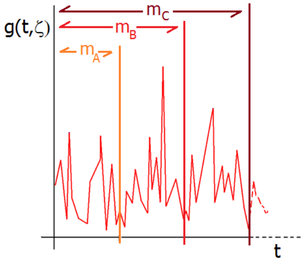
where $y = T^t(x_0)$ and the point ζ is chosen on the attractor.

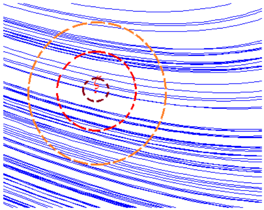
- **The observable function must achieve a global maximum at ζ and be monotonically decreasing.**
- By chosing the logarithm we **fix the asymptotic EVL to be type 1**:

$$\xi = 0 \qquad \sigma = 1/D$$

A geometric perspective

The link with the geometry is evident:
Maxima of $g_i(t, x_0, \zeta) = -\log(\mathrm{dist}(T^t(x_0), \zeta))$ are entrances in a ball centred in ζ.





GEV distributions are given in terms of the mass of the ball: this explains why the parameters depend on the **local dimension** (Lucarini et al (2012)).

Can we adapt this theory for time series?

- We can think to proceed as Wolf et al. did for the definition of the Lyapunov exponents.

Can we adapt this theory for time series?

- We can think to proceed as Wolf et al. did for the definition of the Lyapunov exponents.
- The theory must take into accounts two aspects arising in time series analysis:

Can we adapt this theory for time series?

- We can think to proceed as Wolf et al. did for the definition of the Lyapunov exponents.
- The theory must take into accounts two aspects arising in time series analysis:

Dynamical noise

- It is introduced by the **physical processes** at fast scales.
- In case of **temperature** daily data, it corresponds roughly to the **phenomena associated to the meteorology**.

Can we adapt this theory for time series?

- We can think to proceed as Wolf et al. did for the definition of the Lyapunov exponents.
- The theory must take into accounts two aspects arising in time series analysis:

Dynamical noise

- It is introduced by the **physical processes** at fast scales.
- In case of **temperature** daily data, it corresponds roughly to the **phenomena associated to the meteorology**.

Observational noise

- It is the effect of the combined **accuracy and precision** of an instrument.
- For **temperature data both effects** are relevant.

EVT for stochastic perturbations

- In the most general case:

$$x(\vec{\xi}, t) = f(x(t-1)) + \epsilon\xi(t) \quad \xi(t) \text{ random},$$
$$y(\vec{\xi}, \psi(t), t) = x(\vec{\xi}, t) + \eta\psi(t) \quad \psi(t) \text{ random}.$$

EVT for stochastic perturbations

- In the most general case:

$$x(\vec{\xi}, t) = f(x(t-1)) + \epsilon\xi(t) \quad \xi(t) \text{ random},$$
$$y(\vec{\xi}, \psi(t), t) = x(\vec{\xi}, t) + \eta\psi(t) \quad \psi(t) \text{ random}.$$

Dynamical noise

- EVLs found for mixing systems.
- For periodic or quasi-periodic, convergence towards the GEV model is found if the noise is *sufficiently* large.

EVT for stochastic perturbations

- In the most general case:

$$x(\vec{\xi},t) = f(x(t-1)) + \epsilon\xi(t) \quad \xi(t) \text{ random},$$
$$y(\vec{\xi},\psi(t),t) = x(\vec{\xi},t) + \eta\psi(t) \quad \psi(t) \text{ random}.$$

Dynamical noise

- EVLs found for mixing systems.
- For periodic or quasi-periodic, convergence towards the GEV model is found if the noise is *sufficiently* large.

Observational noise

- EVLs found for mixing and eriodic orbits.
- The properties of the **underlying attractor are preserved** with this kind of perturbation.

Finite size effects

The convergence is ensured only asymptotically and two limits must be ideally satisfied:

1. **The number of observations $\mathbf{m}$** in each bin should be large enough to sample a *real* extreme event.

Finite size effects

The convergence is ensured only asymptotically and two limits must be ideally satisfied:

1. **The number of observations $\mathbf{m}$** in each bin should be large enough to sample a *real* extreme event.
2. A sufficient **The number of maxima $\mathbf{n}$** is needed to get statistics.

Finite size effects

The convergence is ensured only asymptotically and two limits must be ideally satisfied:

1. **The number of observations $\mathbf{m}$** in each bin should be large enough to sample a *real* extreme event.
2. A sufficient **The number of maxima $\mathbf{n}$** is needed to get statistics.

Dataset dimension required for observing convergence to the EVLs

We have shown that in chaotic systems **we need at least $n > 100$ and $m > 300$** to achieve good agreement with respect to the theory.

The algorithm

1. Consider a given time series $x(t)$ $t = 1, 2, ..., s$.

The algorithm

1. Consider a given time series $x(t)$ $t = 1, 2, ..., s$.
2. **Fix the point** ζ to be a point of the series itself $x(t)$.

The algorithm

1. Consider a given time series $x(t)$ $t = 1, 2, ..., s$.
2. **Fix the point** ζ to be a point of the series itself $x(t)$.
3. Compute the series $y(t) = -\log(\text{dist}(\zeta, x(t)))$.

The algorithm

1. Consider a given time series $x(t)$ $t = 1, 2, ..., s$.
2. **Fix the point** ζ to be a point of the series itself $x(t)$.
3. Compute the series $y(t) = -\log(\text{dist}(\zeta, x(t)))$.
4. Once $y(t)$ in n bins each containing m data $(mn = s)$, **extract the maxima** M_j, $j = 1, ..., n$ for the series $y(t)$.

The algorithm

1. Consider a given time series $x(t)$ $t = 1, 2, ..., s$.
2. **Fix the point** ζ to be a point of the series itself $x(t)$.
3. Compute the series $y(t) = -\log(\text{dist}(\zeta, x(t)))$.
4. Once $y(t)$ in n bins each containing m data $(mn = s)$, **extract the maxima** M_j, $j = 1, ..., n$ for the series $y(t)$.
5. **Fit the maxima to the GEV model**, perform a test to check whether the fit succeeded or failed.

Our definition of extreme events

Our definition of extreme events

Our definition of extreme events

Fit succeeds

Our definition of extreme events

Fit succeeds

- There are enough recurrences around the point ζ: a recurrence of ζ is natural in **the bin lenght fixed.**

Our definition of extreme events

Fit succeeds

- There are enough recurrences around the point ζ: a recurrence of ζ is natural in **the bin lenght fixed.**
- One can **decrease the bin length** m and check the smallest m such that the fit converges.

Our definition of extreme events

Fit succeeds

- There are enough recurrences around the point ζ: a recurrence of ζ is natural in **the bin lenght fixed.**
- One can **decrease the bin length** m and check the smallest m such that the fit converges.

Our definition of extreme events

Fit succeeds

- There are enough recurrences around the point ζ: a recurrence of ζ is natural in **the bin lenght fixed.**
- One can **decrease the bin length** m and check the smallest m such that the fit converges.

Fit fails

Our definition of extreme events

Fit succeeds

- There are enough recurrences around the point ζ: a recurrence of ζ is natural in **the bin lenght fixed.**
- One can **decrease the bin length** m and check the smallest m such that the fit converges.

Fit fails

- **The recurrences of ζ can be considered as extreme events.**

Our definition of extreme events

Fit succeeds

- There are enough recurrences around the point ζ: a recurrence of ζ is natural in **the bin lenght fixed.**
- One can **decrease the bin length** m and check the smallest m such that the fit converges.

Fit fails

- **The recurrences of ζ can be considered as extreme events.**
- One can **increase the bin length** m and check the smallest m such that the fit converges. The definition of extreme events is therefore relative to m.

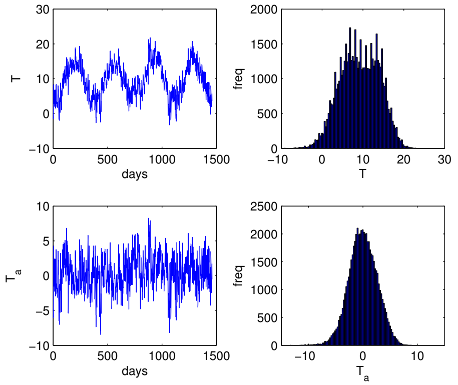


Figure : **Top**: An example of **temperature series and its histogram**. **Bottom**: the same for a **temperature anomalies** series. All the plots refer to Armagh (UK) weather station.

Results: Temperatures

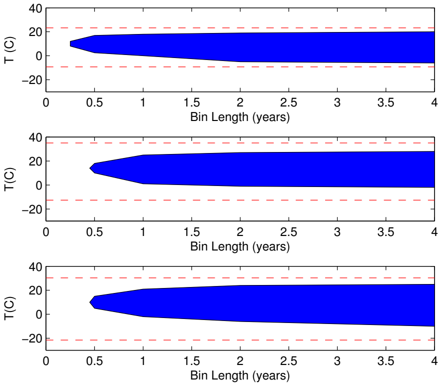


Figure : **Region of temperature with convergent Gumbel fit (blue area)** for different bin length. Red dotted lines: absolute extremes of the temperature series for Armagh (top), Milan (middle) and Vienna (bottom). **Extremes are located in the white area** of the plots between the two dotted lines.

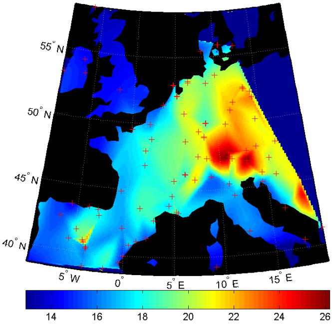


Figure : Map of the range of *admissible anomalies* for the European region, obtained considering the **interval of temperature anomalies** ζ **such that the fit passes the Kolmogorov Smirnov test**. The red crosses represent the location of the stations used for the analysis.

Conclusions

Remarks

- The method distinguishes between real extremes and natural climate variability.
- The method presents a built-in test of convergence to the predicted EVT.
- The method is easy to be implemented and fast.

Conclusions

Remarks

- The method distinguishes between real extremes and natural climate variability.
- The method presents a built-in test of convergence to the predicted EVT.
- The method is easy to be implemented and fast.

Open questions

- What happens to variables with non continuous supports?
- Can we define EVT on observable defined on networks?

References

- Faranda, Davide, et al. Journal of Statistical Physics 145.5 (2011): 1156-1180.

References

- Faranda, Davide, et al. Journal of Statistical Physics 145.5 (2011): 1156-1180.
- Lucarini, Valerio, et al. " Chaos: An Interdisciplinary Journal of Nonlinear Science 22.2 (2012): 023135-023135.

References

- Faranda, Davide, et al. Journal of Statistical Physics 145.5 (2011): 1156-1180.
- Lucarini, Valerio, et al. " Chaos: An Interdisciplinary Journal of Nonlinear Science 22.2 (2012): 023135-023135.
- Faranda, Davide, et al. Generalized extreme value distribution parameters as dynamical indicators of stability." International Journal of Bifurcation and Chaos 22.11 (2012).

References

- Faranda, Davide, et al. Journal of Statistical Physics 145.5 (2011): 1156-1180.
- Lucarini, Valerio, et al. " Chaos: An Interdisciplinary Journal of Nonlinear Science 22.2 (2012): 023135-023135.
- Faranda, Davide, et al. Generalized extreme value distribution parameters as dynamical indicators of stability." International Journal of Bifurcation and Chaos 22.11 (2012).
- Faranda, Davide, et al 2013 Nonlinearity 26 2597 doi:10.1088/0951-7715/26/9/2597