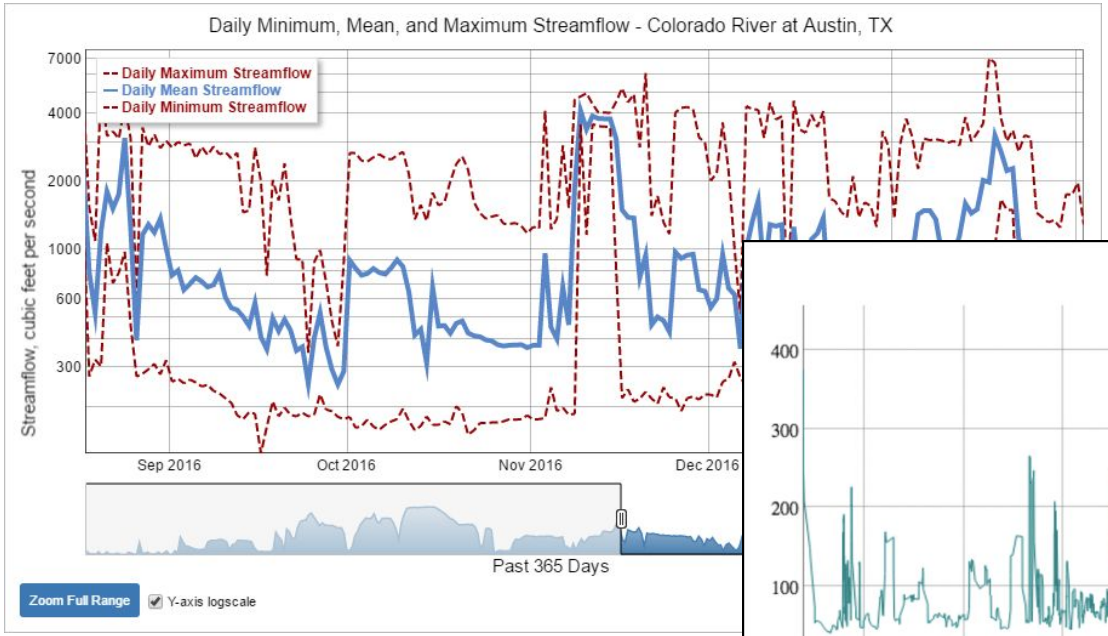
Datascience

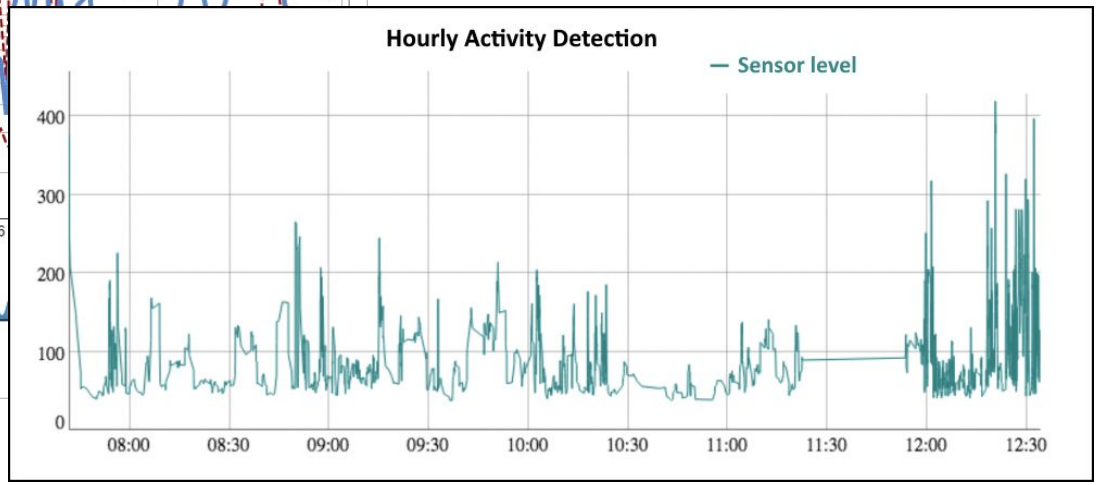
02 - Timeseries analysis

Dávid Visontai
ELTE, Physics of Complex Systems Department
2021.02.17.

What is timeseries?

- A series of **data points** indexed (or listed or graphed) in time order.
  Successive, equally spaced points in time -> **discrete-time data**.





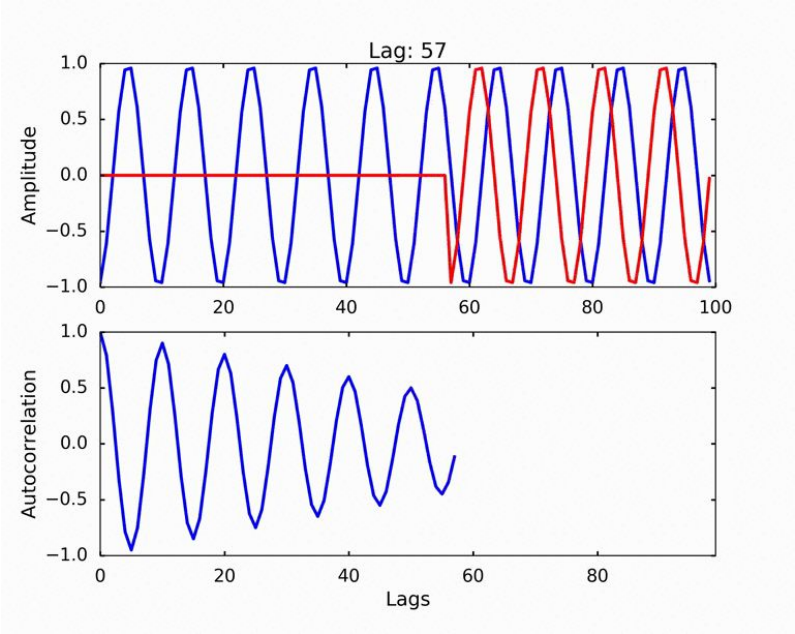
Explore timeseries data - Autocorrelation

- **Autocorrelation** analysis to examine serial dependence

  a. **Degree of similarity** between a given time series and **a lagged version of itself**

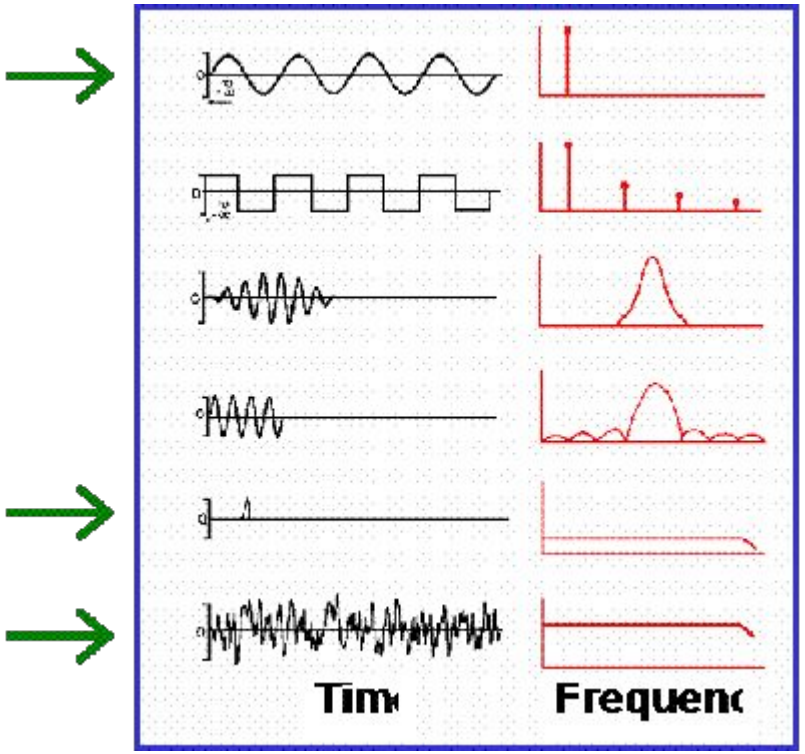
  b. **Relationship between** a variable's **current value and its past** values.

  c. It can be **positive** or **negative**



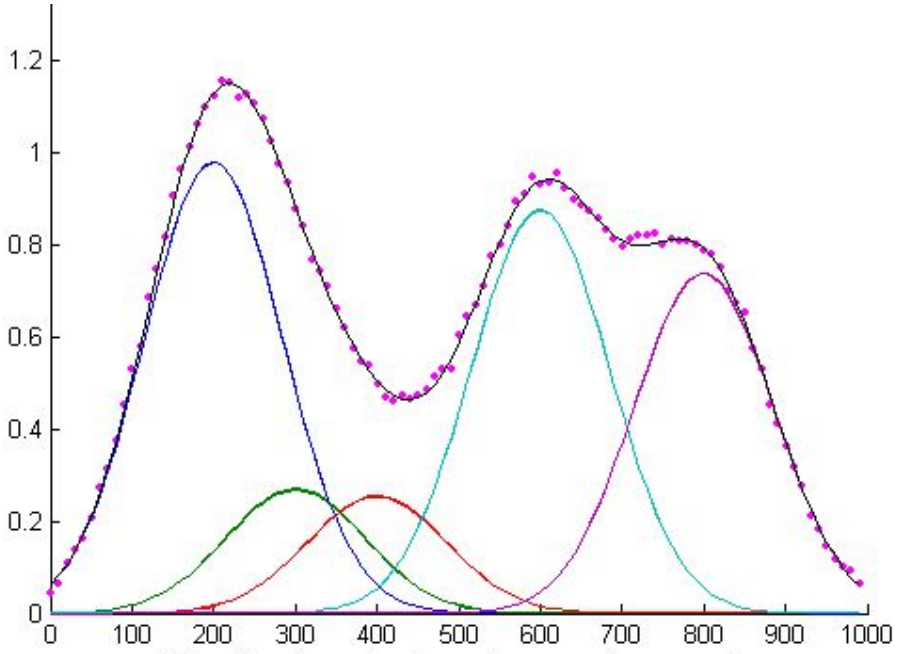
Explore timeseries data - Spectral analysis

- **Spectral analysis** to examine **cyclic behavior.**
  a. E.g. sun spot activity varies over 11 year cycles.
  b. Celestial phenomena
  c. Weather patterns
  d. Neural activity
  e. Commodity prices
  f. Economic activity.



Explore timeseries data - Component analysis

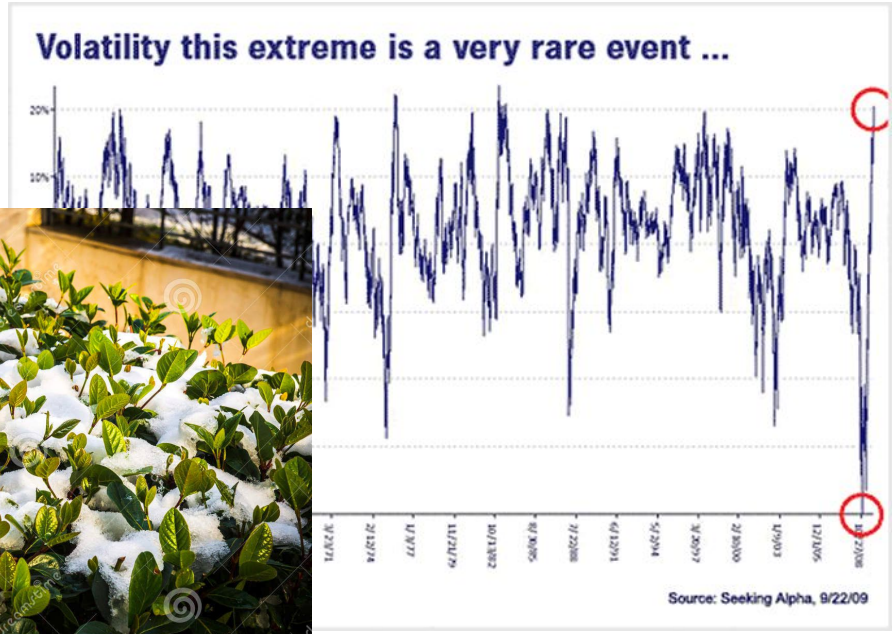
- **Separation into components** representing:
  a. **Trend**
  b. **Seasonality**
  c. **Speed of variation**
  d. **Cyclical irregularity**



Timeseries analysis to predict extreme events

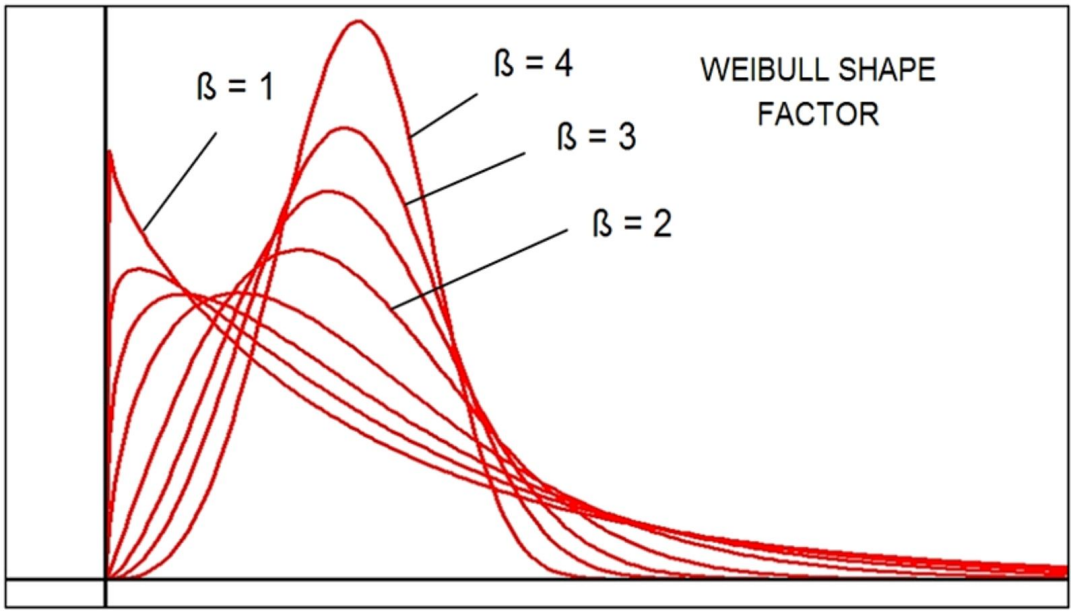
Extreme events, though rare, can have an enormous negative impact on individuals, society, business, economies and infrastructure.

- River or coastal flooding, droughts and heatwaves, and stock market crashes.
- Prediction can help to **prepare for the possible measure** of such events, **design the appropriate defence mechanisms** and **estimate the cost** of it.



Timeseries analysis to predict extreme events

- Extreme events are **rare** by definition, prediction of future events relies on extrapolation from a suitable model fitted to historical data.
- **Extreme value analysis** provides a statistical framework for this kind of analysis.
- **Standard statistical methods** are designed to **characterise the mean behaviour** of a process or data sample and are therefore not generally useful for capturing this tail behaviour ⟺ Subsequently, **methods which focus specifically on tail events are required**.



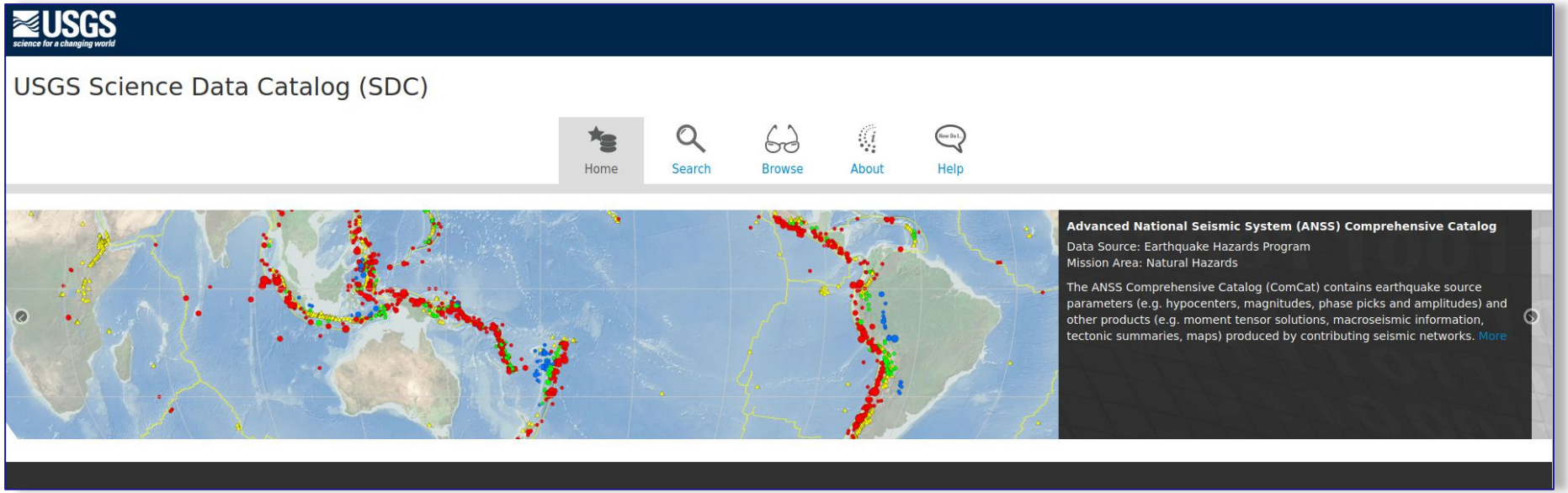
Further readings

- Rare events, LSTM, autoencoder
  - Extreme Rare Event Classification using Autoencoders in Keras
  - LSTM Autoencoder for Extreme Rare Event Classification in Keras
  - Step-by-step understanding LSTM Autoencoder layers | by Chitta Ranjan

- Trend estimation and decomposition of time series Link1, Link2
- Fourier transform explained: https://youtu.be/spUNpyF58BY

Statistical analysis of water discharge of surface

- Large amounts of historical surface water data are available from the United States Geological Survey (USGS) at https://waterdata.usgs.gov/nwis
- The goal of the project is to
  - **retrieve sample data** using the webpage manually, and then later automate the process by calling the web service as described at https://help.waterdata.usgs.gov/faq/automated-retrievals.
  - Discover rainy events, analyse their occurrences and draw conclusion from various aspects



Statistical analysis of water discharge of surface

- Large amounts of historical surface water data are available from the United States Geological Survey (USGS) at https://waterdata.usgs.gov/nwis