

**Fabio D'Andrea**

LMD – 4<sup>e</sup> étage “dans les serres”

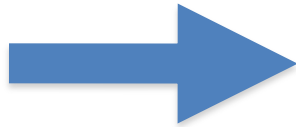
01 44 32 22 31

[dandrea@lmd.ens.fr](mailto:dandrea@lmd.ens.fr)

<http://www.lmd.ens.fr/dandrea/TEACH>

## ***Program***

17/1 Elementary statistics – 1  
23/1 Elementary statistics - 2  
14/2 Exercises – Computer room



14/2 Fourier Analysis -1  
3/3 Fourier Analysis -2, stochastic processes  
6/3 Exercises – Computer room  
13/3 Exercises – Computer room

13/3 Principal component analysis -1  
20/3 Principal component analysis -2  
27/3 Exercises – Computer room  
3/4 Exercises – Computer room

17/4 Cluster analysis – Machine learning  
24/4 Exercises – Computer room

15/5 Principal component analysis: Complements

22/5 Exam

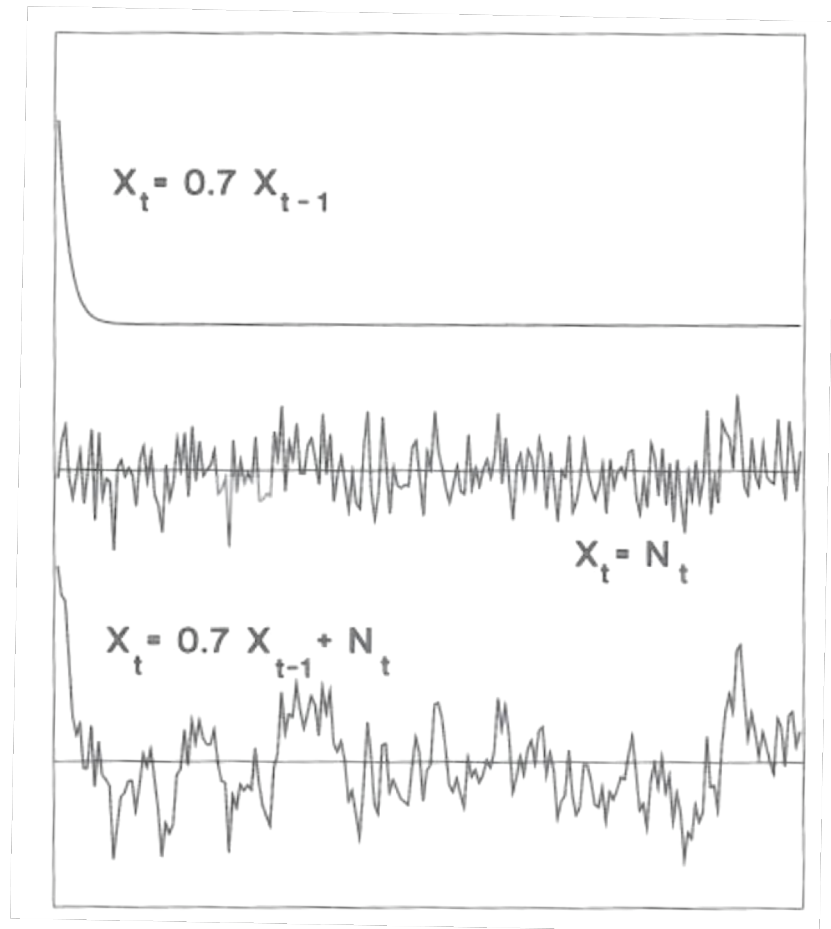
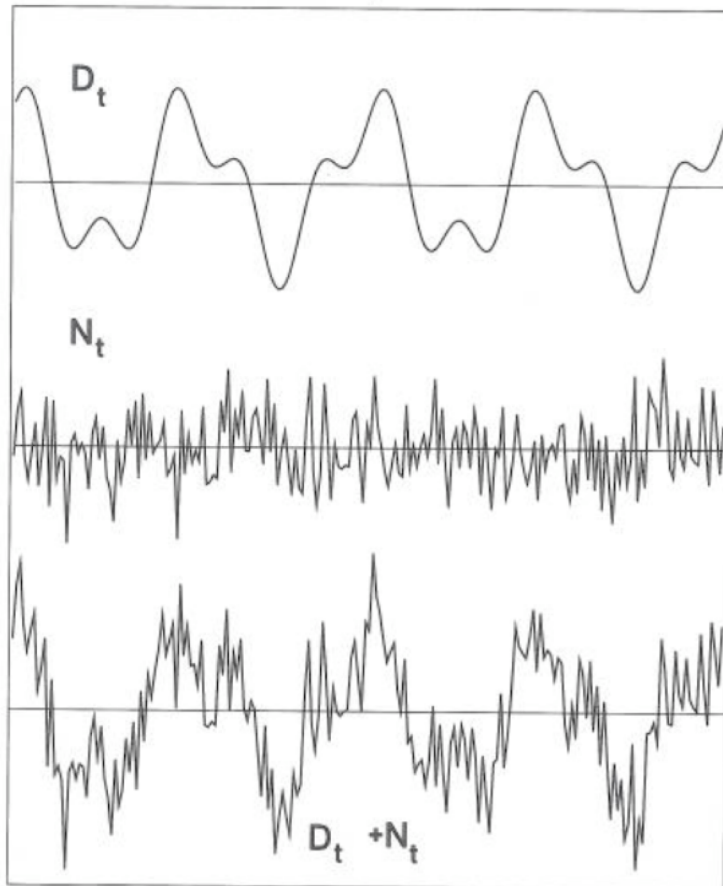
**. Lesson 3.**

**Stochastic Processes, the role of Noise**

# Noise

A time series can be seen as the sum of a deterministic part and a noise part

$$x_t = D_t + N_t$$



A time series can be seen as a finite sample of a stochastic process.

The state of a stochastic process  $x_t$  at time  $t$  depends on the state at all other times.

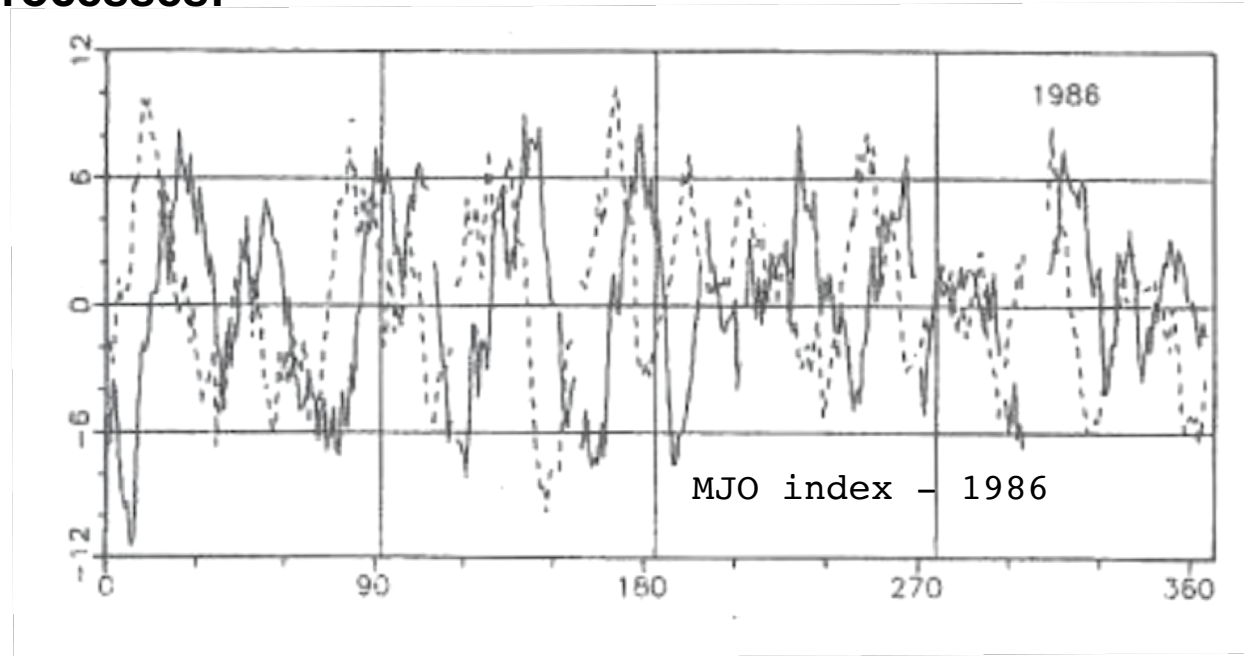
## Properties of stochastic processes:

### **Characteristic time.**

The series is  
Autocorrelated:

$$P(x_{t+\tau} > 0 \mid x_t > 0) > 0.5.$$

The time  $\tau$  for which the  
Probability becomes 0.5  
Is the characteristic time.



### **Stationarity**

The statistical properties do not depend on time:  $x_t$  has the same PDF for all  $t$ , and the joint PDF of  $x_t$  and  $x_s$  depend only on  $|t - s|$ .

### **Weak Stationarity**

Same but only for mean and variance.

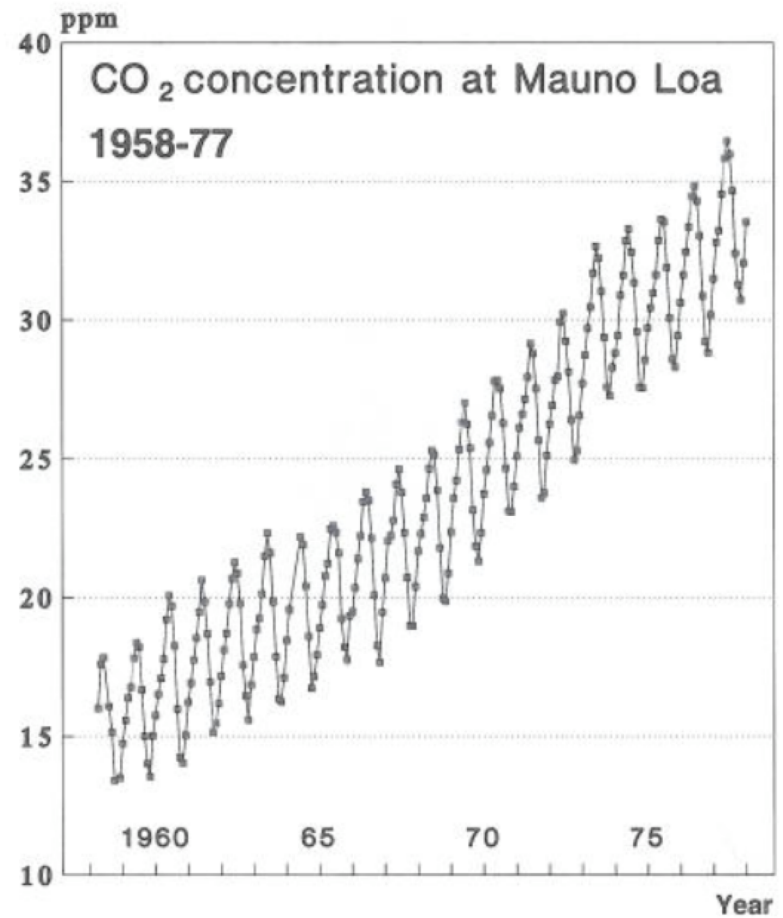
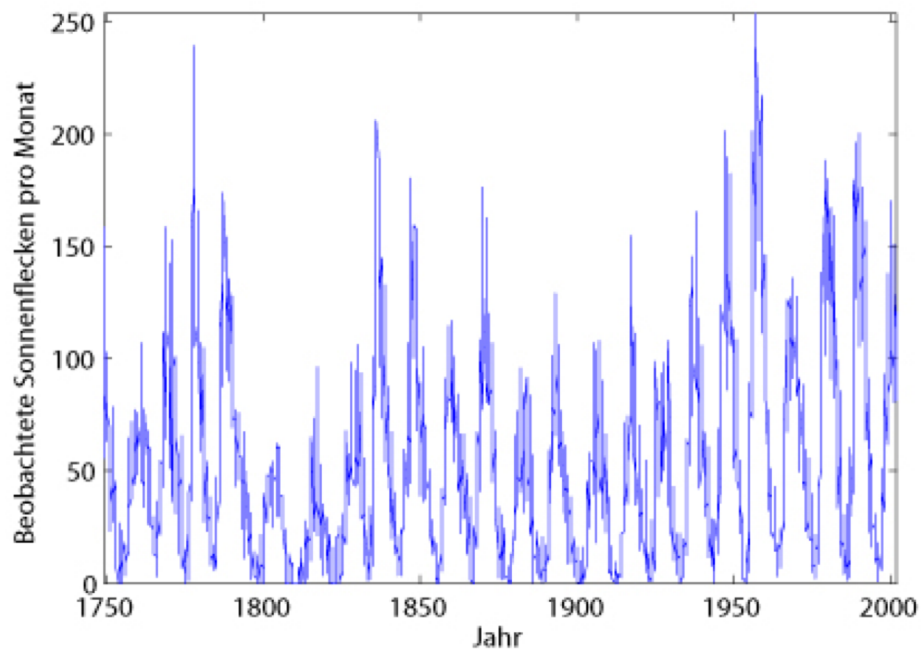
**stationarity with trend**

**Cyclo-stationarity**

**Cyclo-stationarity with trend**

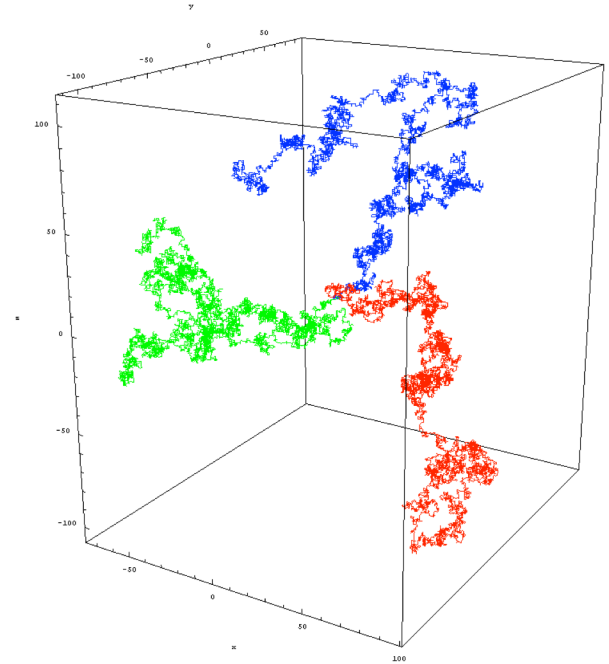
**Ergodicity**

Solar radiation



# Random Walk

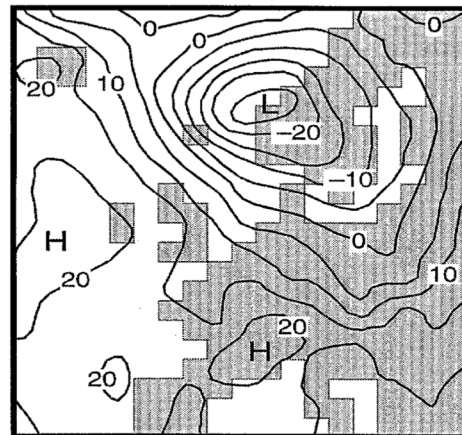
$$\mathbf{x}_t = \sum_{i=1}^t \mathbf{z}_i, \text{ where } \mathbf{z}_t \text{ is a white noise}$$



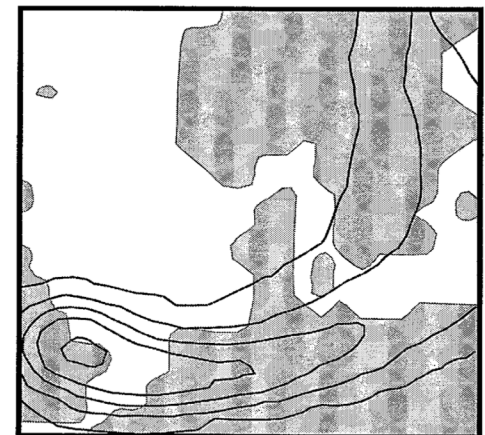
Example, pollutant transport:

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \mathbf{U} + \mathbf{z}_t$$

*Example of a simulation of long-range transport of air pollutants. Left: Simulated 1000 hPa height field 24 hours after model initialization. Right: Distribution of pollutant continuously emitted in east England after 24 hours. From Lehmhaus et al.*



1000 hPa height 24 hours after initialization



Concentration after 24 hours of emissions

## Autoregressive processes

It's a sub-class of stochastic processes. But one can show that well behaved stochastic processes (stationary and ergodic) can be approximated by an autoregressive process.

$$x_t = a_0 + \sum_{k=1}^p a_k x_{t-k} + w_t$$

$p$  is the *order*. Common writing is AR(p): AR(1) is an order 1, AR(2) etc etc. If a process has mean zero it is  $a_0 = 0$ .

So, an order 1, zero mean process can be written (red noise):

$$x_t = a_1 x_{t-1} + w_t .$$

An order 0 (white noise):

$$x_t = w_t$$



Autoregressive processes are cool because:

1) Given any weakly stationary ergodic process it is possible to find an AR-process that approximates it arbitrarily closely

2) They can be seen as discretizations of ordinary differential equations. For example, order 1:

$$\frac{dx(t)}{dt} + \alpha x(t) = f(t)$$

$$\frac{x_t - x_{t-dt}}{dt} = \alpha x_t + f_t$$

$$x_t = ax_{t-1} + w_t \qquad a = \frac{1}{1-\alpha}, w_t = \frac{f_t}{1-\alpha}$$

And order 2? Exercise!

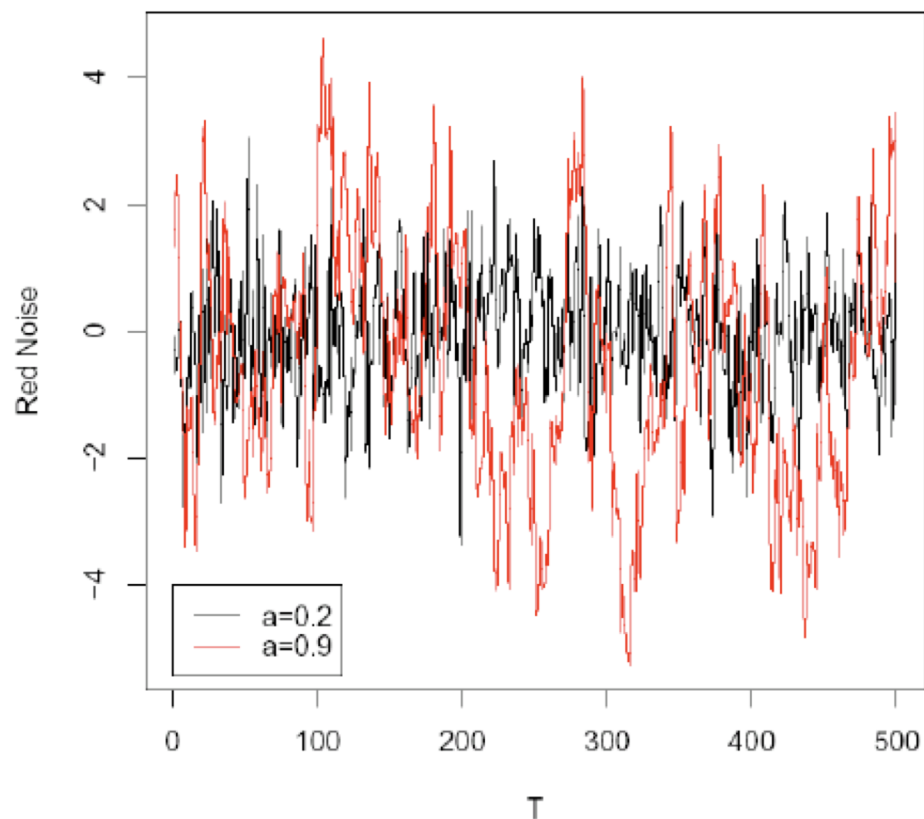
**AR(1) processes**  $x_{t+1} = a_1 x_t + w_t$

If we know the parameters  $a$  and  $\sigma_w$  we can compute the variance and the autocorrelation of the AR(1) process. The autocorrelation is simply given by  $a$  and

$$\sigma_x^2 = \frac{\sigma_w^2}{1 - a^2}.$$

One can construct a “rednoise model” of a given timeseries.

Equalling the rednoise variance and autocorrelation and the timeseries variance and autocorrelation. That gives two equations for  $a$  and  $\sigma_w^2$ .

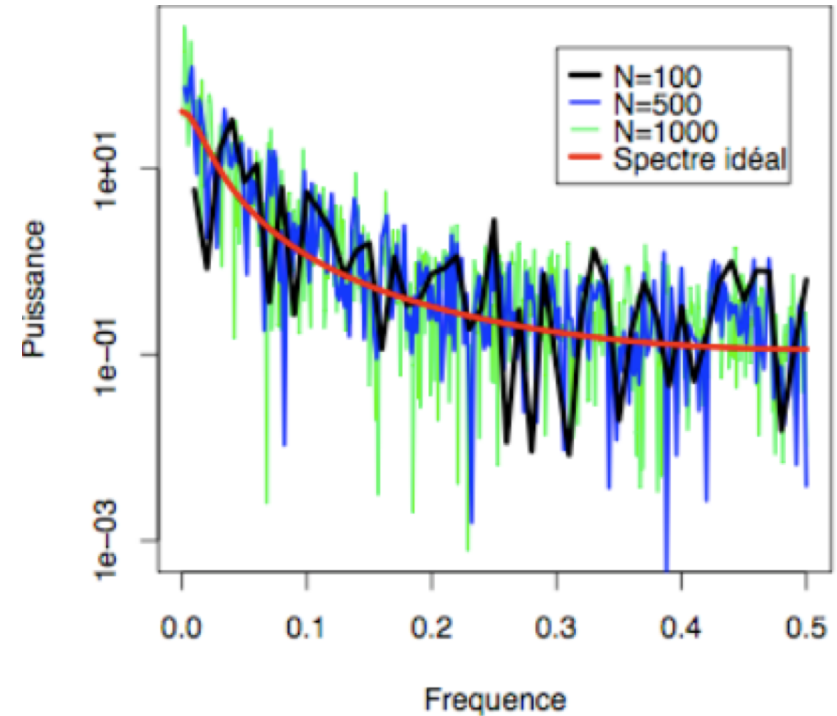


AR(p) spectrum. It can be shown that:

$$P(f) = \frac{\sigma_w^2}{1 - \left| 1 - \sum_{k=1}^p a_k e^{i2\pi f k} \right|^2}.$$

Which in the case of an AR(1) becomes:

$$P(f) = \sigma_x^2 \frac{1 - a^2}{1 - 2a \cos(2\pi f) + a^2}.$$



Although there is no “memory” beyond lag 1, there is power at very low frequencies. Nonzero autocorrelation at short time lags can create high power at low frequency!

This is a feature of climatic – and geophysical – timeseries.

The rednoise process is used as a null hypothesis in spectral estimation.