UNIVERSITY OF TWENTE.

BOX-JENKINS APPROACH FOR TIME SERIES ANALYSIS

MARGARITA HUESCA MARTINEZ

Presentation overview

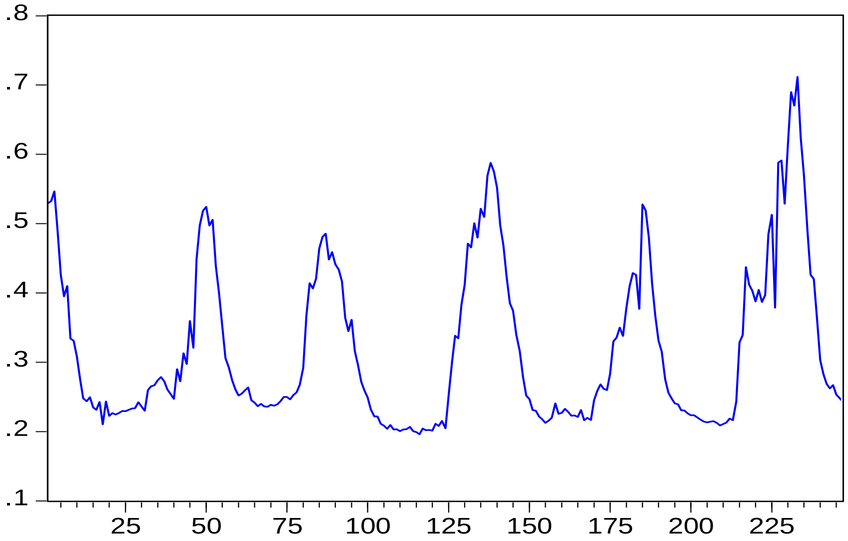
- Learning objectives
- Time series definition & Remote Sensing time series
- Box and Jenkins approach
- Stationary process
- Seasonal time series
- Autoregressive models
- Examples
- Conclusions

Learning objectives

- To characterize a time series
- To assess the seasonal components of a time series
- To define autoregressive models using the & Box and Jenkins approach

Introduction: Time series

Time series: sequence of observations of a variable in the increasing order of time



Time series provided reliable and quantitative information about the history of the variable.

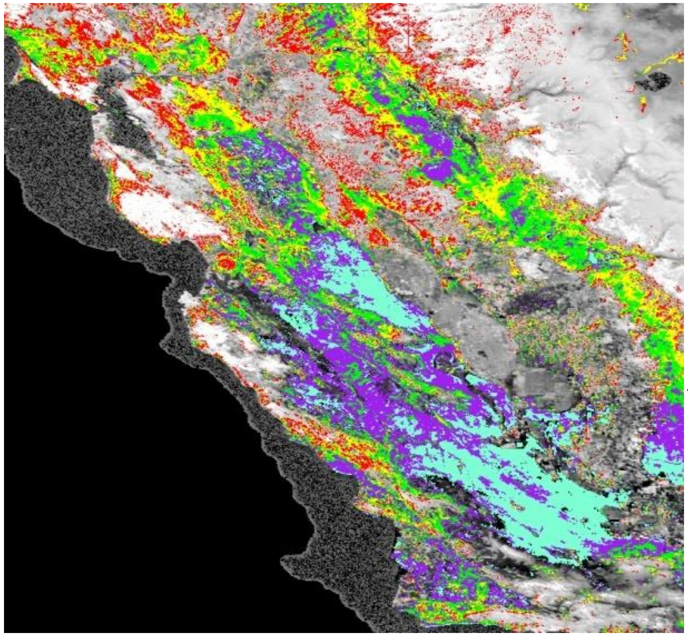
Remote sensing time series

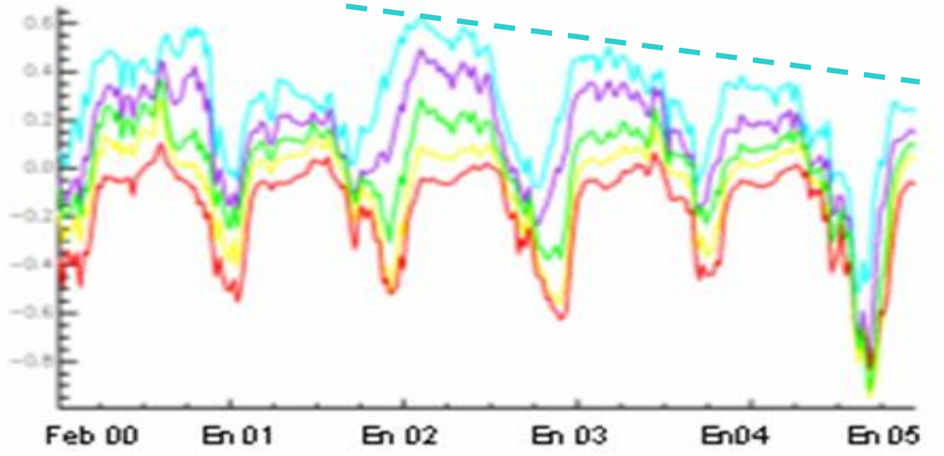
- **Trends** at medium term, and less in the long term.
- **Seasonality**: very frequent in agricultural and environmental related variables. Irregular periodicity inside of a year.
- **Cycles**: uncommon, the dynamics is repeated after several years. Sun related series.
- **Structural changes**: the level and variance of a variable change at some point and can evolve with different patterns.
- **Outliers**: Influential, extreme or aberrant observations.
- **Not available data**: due to problems in data acquisition
- **Irregular variance**: variance that evolves over time.

Remote sensing time series: Trends

Trends at medium term

Grasslands in California





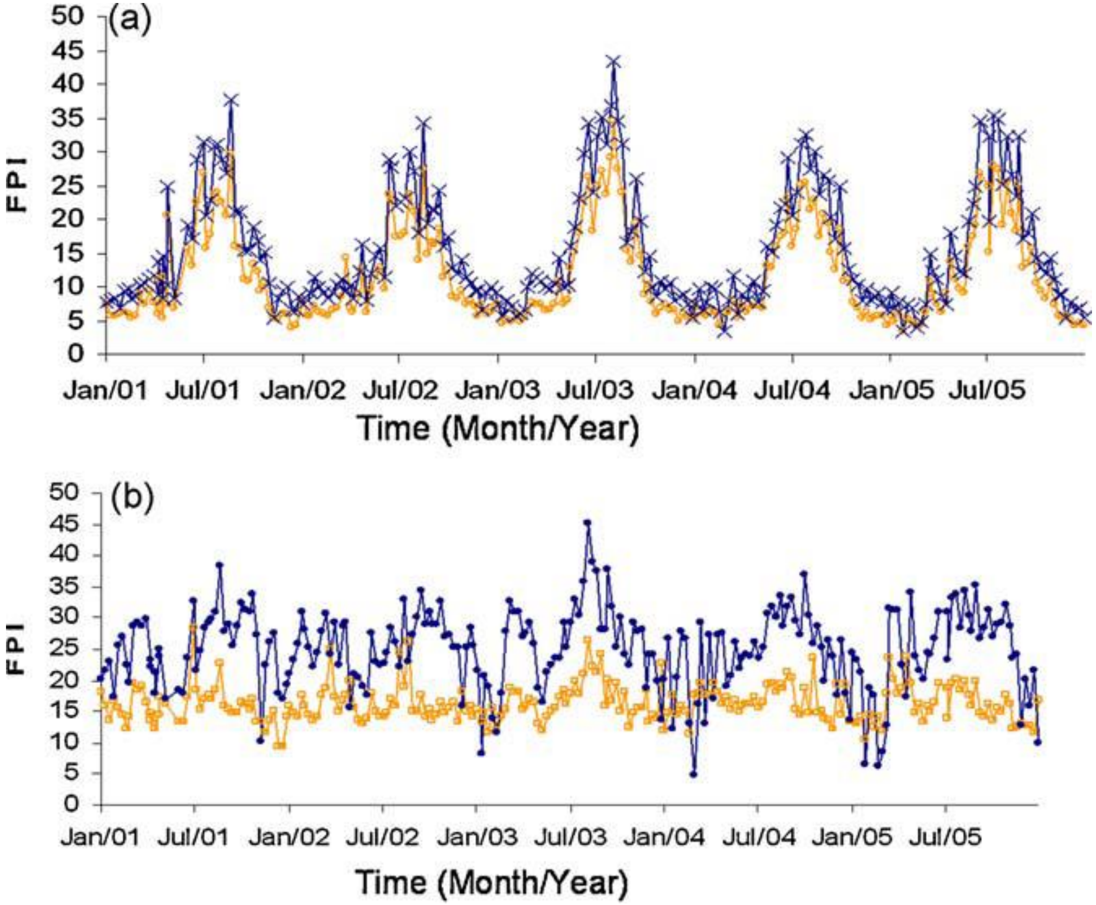
decrease of rainfall

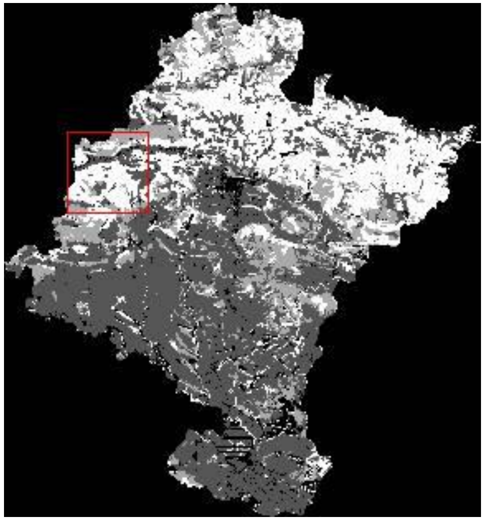
Average rainfall (mm)

261
366
439
500
532

Remote sensing time series: Seasonality

Seasonality: periodicity inside a year, frequent in most vegetation variables





Remote sensing: Structural Changes

Type of disturbance

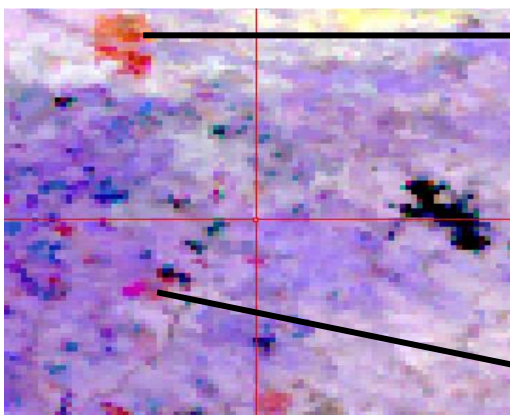
Temporal scale		Climate changes	Human impact
Fast response	Permanent changes	Fires	Land conversion
	Temporary changes	Decrease vegetation growth period	Fires
Long term response	Permanent changes	Increase evapotranspiration	Decrease land cover
	Temporary changes	Decrease primary productivity	Crop rotations

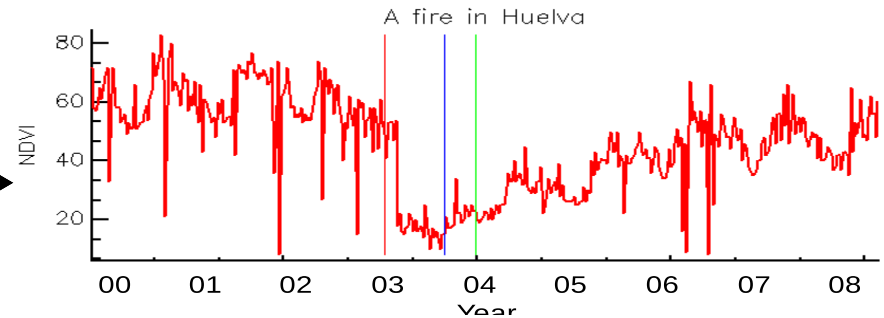
- Need for monitoring and early warning systems
- To provide quantitative and standardizes information for decision maker.
- TS of ecosystem properties will show altered ecosystems

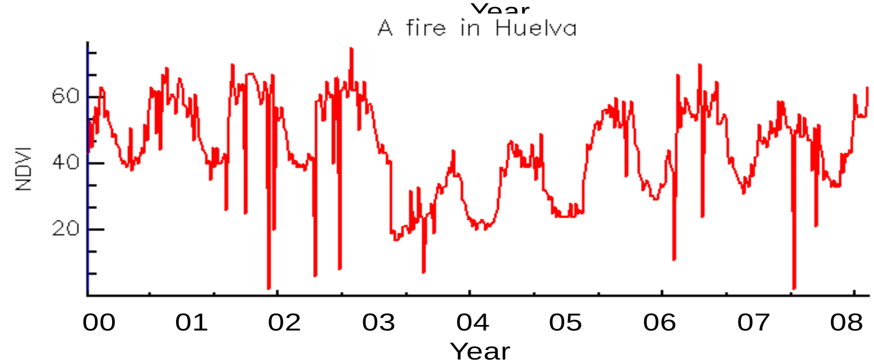
Remote sensing: Structural Changes

Structural changes: land conversion or fires

Fires in South Spain



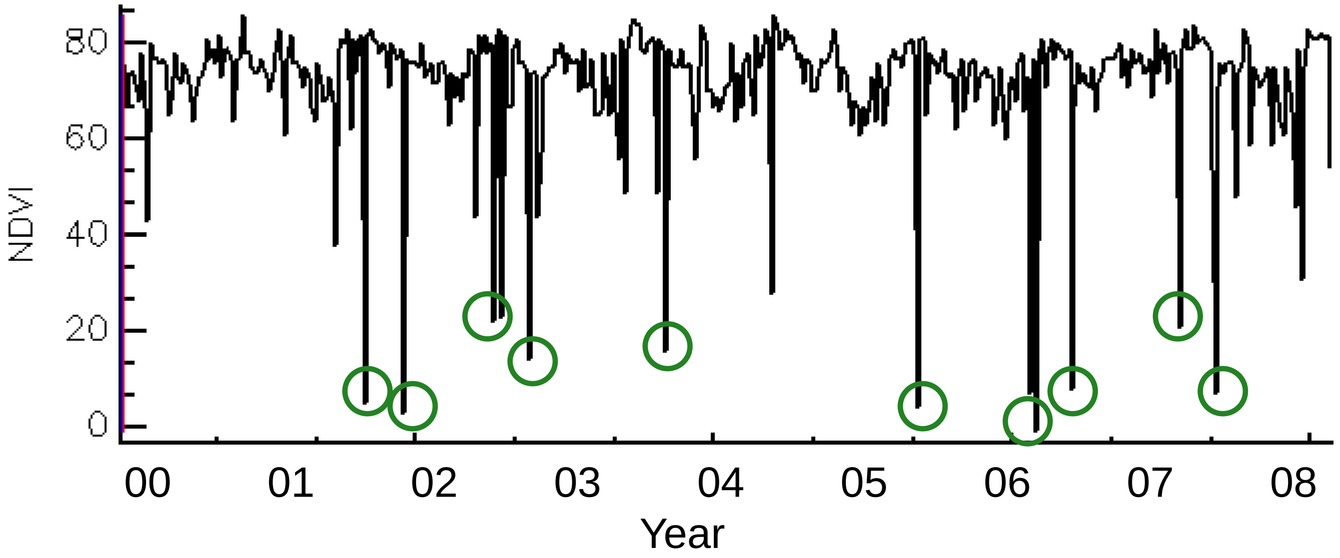




Remote sensing: Outliers or not available data

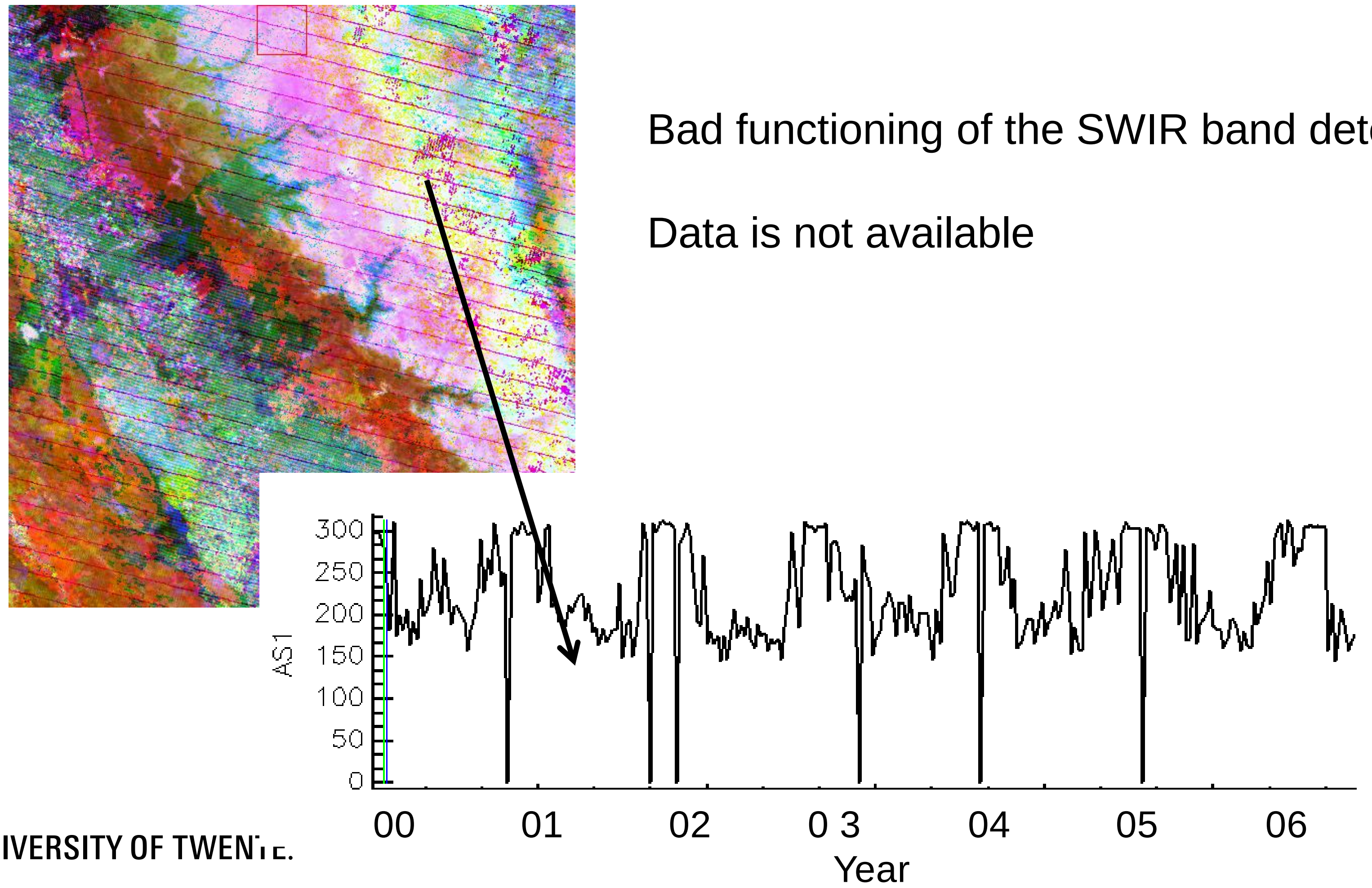
Not available data: common in remote sensing due to external factors such as clouds

Clouds in Cádiz



Remote sensing: Outliers or not available data

Not available data: due to problems in data acquisition

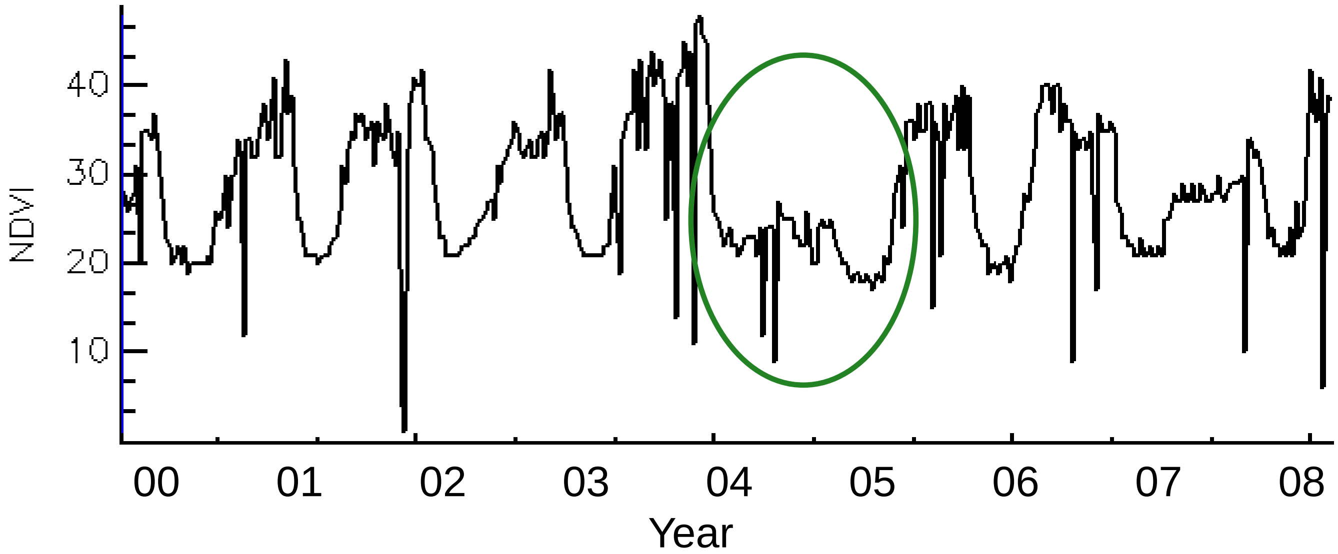


Bad functioning of the SWIR band detector.

Data is not available

Remote sensing: Irregular variance

Irregular variance: Drought in 2005



Time Series Approaches

Process history → Future evolution

To project on the future the knowledge we have of the present and past

Objectives:
- Describe
- Forecast
- Control

Approaches:
- Univariate
- Multivariate
- Deterministic
- Stochastic

Stochastic process → Time series

(Stochastic process: A set of time-dependent random variables)

Objective: Identify, estimate and forecast the stochastic process.

Univariate time series models

Linear processes frequently used to represent a variable y_t in time

1.- White noise ε_t (<u>uncorrelated</u> random variable)

$$y_t = \varepsilon_t$$

where ε_t satisfy: $E\left\{\varepsilon_t\right\} = 0 \quad \forall t$

$$E\left\{\varepsilon^2{}_t\right\} = \sigma^2 \quad \forall t$$

$$E\left\{\varepsilon_t \varepsilon_{t'}\right\} = 0 \quad \forall t \neq t'$$

2.- Autoregressive process, AR(p) (<u>time-dependent</u> variable + white noise)

$$y_t = \phi y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_n y_{t-p} + \varepsilon_t$$

Univariate time series models

3.- Moving average, MA(q) (*sum of weighted white noise*)

$$y_t = \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q} + \varepsilon_t$$

where: ε_t ~ white noise

4.- Autoregressive moving average, ARMA (p, q) (*time-dependent variable* + *white noise*)

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \ldots + \theta_q \varepsilon_{t-q}$$

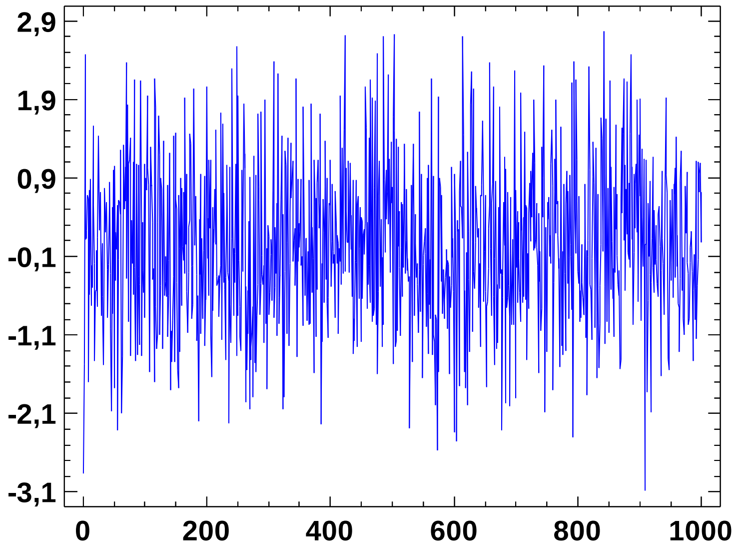
5.- Random walk (*the first differences of y is white noise)*

$$y_t - y_{t-1} = \nabla y_{t-1} = \varepsilon_t$$

White noise and Random walk

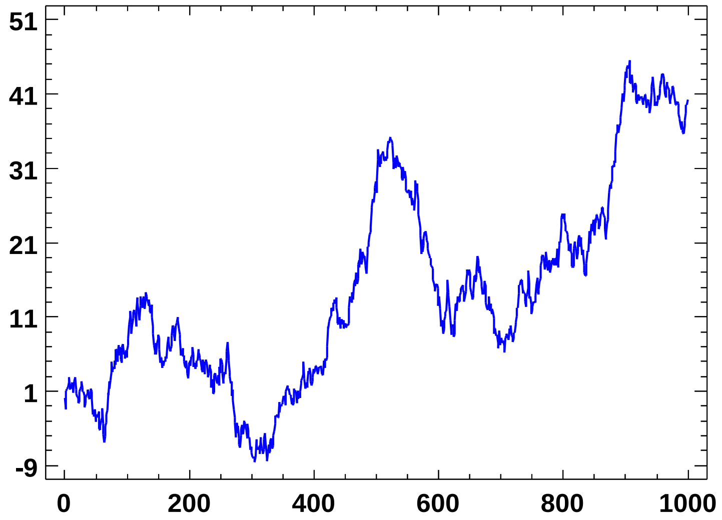
White noise process

$$Y_t = \varepsilon_t$$



Random Walk

$$Y_t = Y_{t-1} + \varepsilon_t$$



Box & Jenkins Approach

Identification

↓

Estimation

↓

Validation

↓

Forecasting

↓

Evaluation

Box & Jenkins Approach

Identification

series data

↓ ← series transformation

series statistics calculation

↓ ↑

Stationarity of the series:

Identification of the order of integration ***d*** → No

↓

Yes

↓

Identification of the orders, ***p*** and ***q*,**
of tentative AR and MA models

Box & Jenkins Approach

Estimation

estimation of model parameters, statistics and residuals

Validation

Diagnostic check of model adequacy

No

Yes

Forecasting

forecasts horizon and calculation

Evaluation

statistical evaluation of forecast accuracy

No

Yes

Identification

Stationary process

Stationary process:

- It has a constant mean in the long term
- It has a finite and constant variance over time
- Its autocorrelations (Acf) decreases rapidly with time

Non-stationary process:

- It has not a constant mean in the long term
- Its variance depends on time, growing to infinity
- Its autocorrelations (Acf) do not decay, or they do it slowly

Stationary process

Causes of non-stationarity:

- The series contains a unit root (stochastic trend) ***
- The series contains seasonal, cyclical or periodic variations
- The series contains a break or a structural change

***If in the AR representation of a series the parameter $\phi_i = 1$

the series has a unit root and consequently it is non-stationary

Stationary process

Differencing: a common transformation to attain stationarity

A series that needs to be differenced ‘d’ times (y_t - y_{t-d}) to attain stationarity is said to be Integrated of order d, I (d)

$y_t - y_{t-d} = \nabla y_t = w_t$ now w_t will be non-integrated or I(0)

If a series y_t contains a unit root, it is non-stationary and it needs to be differenced once to attain stationarity $y_t \sim I(1) : y_t - y_{t-1}$

Most of the series are integrated of order d = 1 or 2, they are I (1) or I (2)

Identification objective: determine the orders ‘p, d, q’ of the ARIMA process that suitably and parsimoniously represents the series

Unit root tests

A method to determine if a series contains one or more unit roots

The Dickey & Fuller (1979, 1981) unit root test (*DF*)

AR(1): $y_t = \mu + \rho y_{t-1} + \varepsilon_t$

- If: $|\rho| < 1$ $\longrightarrow$ y_t is stationary
- If: $\rho = 1$ $\longrightarrow$ y_t is not stationary
- If: $|\rho| > 1$ $\longrightarrow$ y_t is explosive

Seasonal Analysis: Harmonic pattern

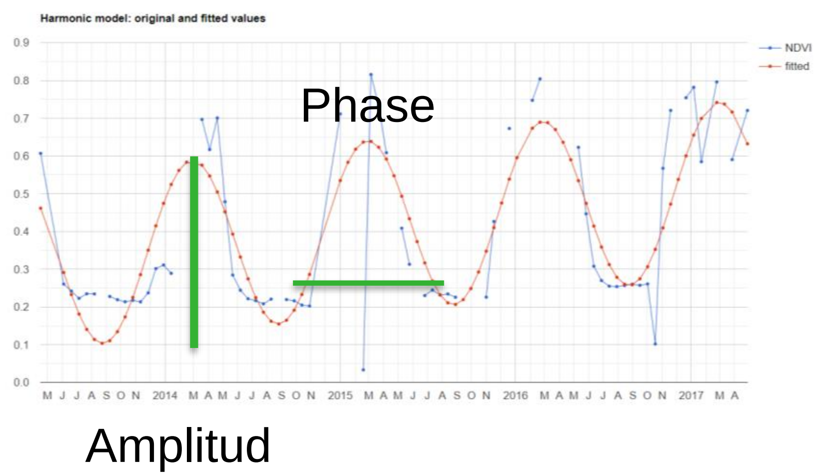
Harmonic pattern

From the trigonometric identity: $\alpha\, sen(\omega t + \delta) = A\, sen(\omega t) + B \cos(\omega t)$

where the amplitude α and the phase δ verify:

$$\alpha^2 = A^2 + B^2$$
$$\delta = \arctan(-B / A)$$

$$Y_t = \mu + A\, sen(\omega t) + B \cos(\omega t) + e_t$$



Seasonal Analysis: Autocorrelation Function

Autocovariance function: γ_k

$$\gamma_k = E\{(y_{t+k} - \bar{y})(y_t - \bar{y})\}$$

where $\bar{y}$ is the mean of y

Autocorrelation function (Acf): ρ_k (Stationary process)

$$\rho_k = \frac{\gamma_k}{\gamma_0}$$

$$\hat{r}_k = \frac{\sum_{t=k+1}^{N} (y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{N} (y_t - \bar{y})^2}$$

Seasonal Analysis: Partial Autocorrelation Function

Partial autocorrelation function (Pacf): ϕ_k

$$\text{AR}(2) \qquad y_t = \phi_{21} y_{t-1} + \phi_{22} y_{t-2} + \varepsilon_t \qquad \phi_{22} \equiv \text{partial autocorrelation}$$

$$\phi_k = r_1 \qquad \textit{for } k = 1$$

$$\phi_k = \frac{r_k - \sum_{j=1}^{k-1} \phi_{k-1,j}\, r_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j}\, r_j} \qquad \textit{for } k > 1$$

where: $\phi_{k,j} = \phi_{k-1,j} - \phi_{kk}\phi_{k-1,k-j}$; for $j = 1, 2, \ldots, k-1$

Box & Jenkins Approach: Identification

Identification of the orders p and q

↓

ACF and PACF: regular and partial autocorrelation functions

Identification of the orders p and q

- The Acf and Pacf of a stationary process decreases rapidly toward zero as the number of lags increase.

<u>Acf</u>: k = nº of lags

If $r_1 \neq 0$ $\longrightarrow$ process with first-order correlation

If r_k decline geometrically when k increases $\longrightarrow$ Low order AR

If $r_k = 0$ after a low number k of periods $\longrightarrow$ Low order MA (k).

Identification of the orders p and q

<u>Pacf</u>:

- If $\phi_{kk} = 0$ for a certain *k*   *order of AR ≤ k*

- The Pacf of a pure AR(p) = 0 for k = p

- The Pacf of a pure MA(q) decreases gradually and asymptotically tending to zero.

$$k > p$$

Identification of the orders p and q

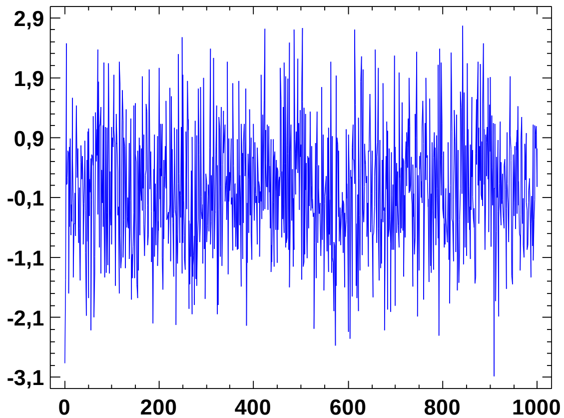
Process	Acf	Pacf
White noise	All the $r_k = 0$	All the $\phi_{kk} = 0$
AR(1) : $a_1>0$ $[y_t = a_1\, y_{t-1} + \varepsilon_t]$	Direct exponential decrease: $r_k = a_1^k$	$\phi_{11} = r_1$; $\phi_{kk} = 0, \forall k \geq 2$
AR(1) : $a_1<0$	Oscillatory decrease: $r_k = a_1^k$	$\phi_{11} = r_1$; $\phi_{kk} = 0, \forall k \geq 2$
AR(p)	Decrease towards zero. The coefficients can oscillate.	Peaks up to lag p, all the $\phi_{kk} = 0, \forall k \geq p$
MA(1); $\beta>0$ $[y_t = \varepsilon_t - \beta \varepsilon_{t-1}]$	Peak (+) in k = 1, $r_k = 0 \quad \forall k \geq 2$	Oscillatory decrease: $\phi_{11} > 0$

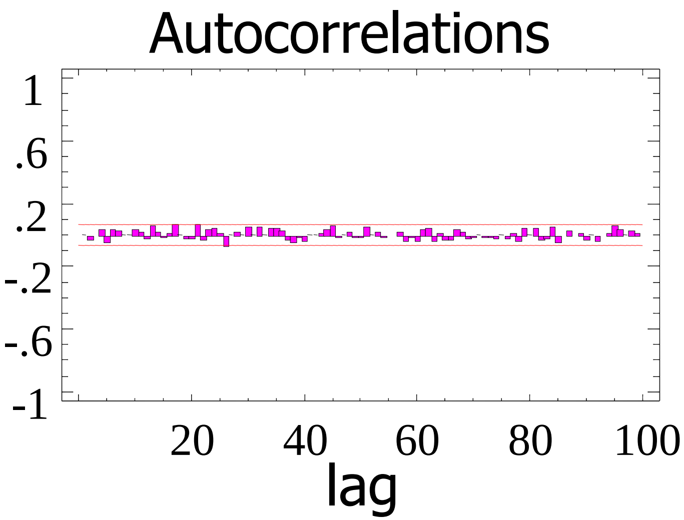
Identification of the orders p and q

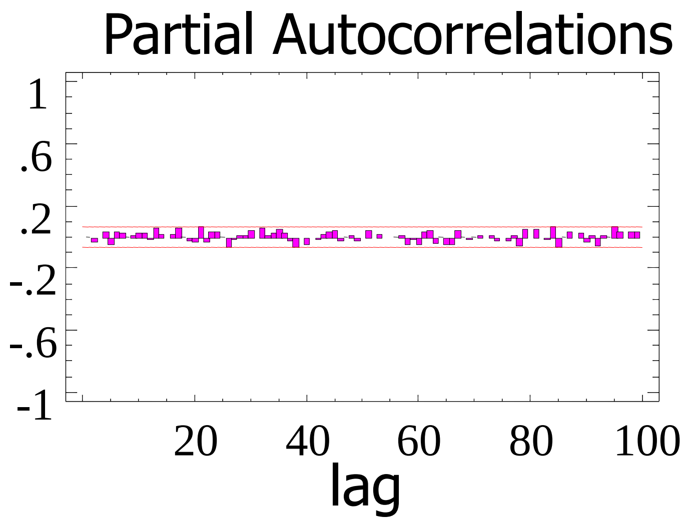
Process	Acf	Pacf
MA(1); $\beta<0$	Peak (-) in k = 1, $r_k = 0 \;\; \forall k \geq 2$	Decrease: $\phi_{11}<0$
ARMA(1,1) A1>0	Exponential decrease from k = 1, sign r_1 = sign($a_1+\beta$)	Oscillatory decrease from k = 1, $\phi_{11} = r_1$
ARMA(1,1) A1<0	Oscillatory decrease from k = 1, sign r_1 = sign($a_1+\beta$)	Exponential decrease from k = 1, $\phi_{11} = r_1$ sign(ϕ_{kk}) = sign(ϕ_{11})
ARMA(p, q) a1<0	Direct or oscillatory decrease from k=p	Direct or oscillatory decrease from k=p

Identification of the order of integration. Examples

White noise process $Y_t = \varepsilon_t$

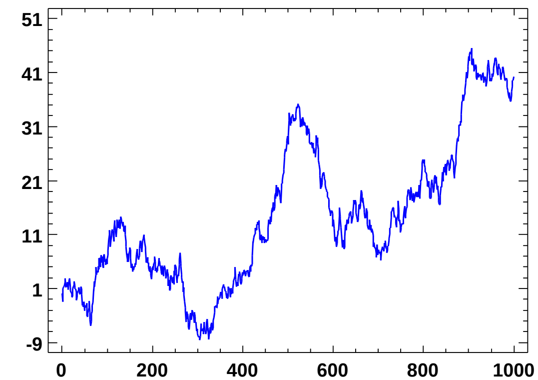




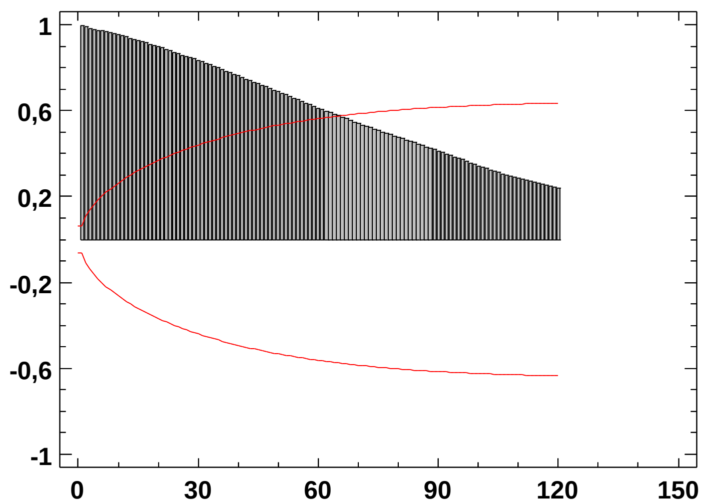


Identification of the order of integration. Examples

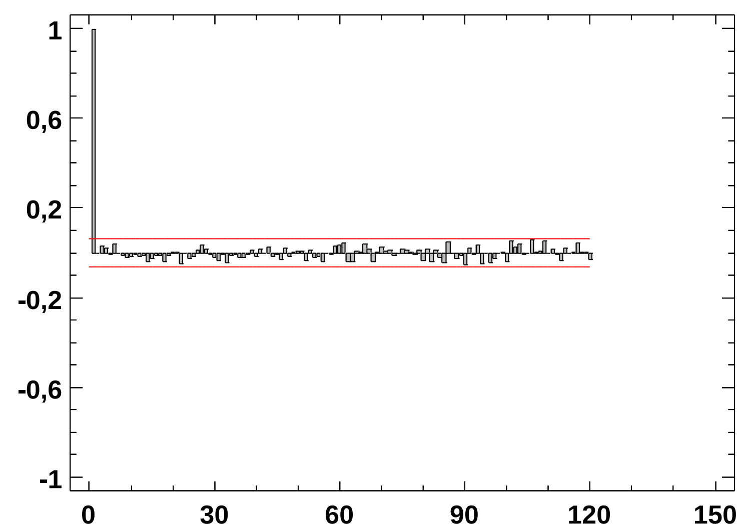
Random Walk $Y_t = Y_{t-1} + \varepsilon_t$



Autocorrelations



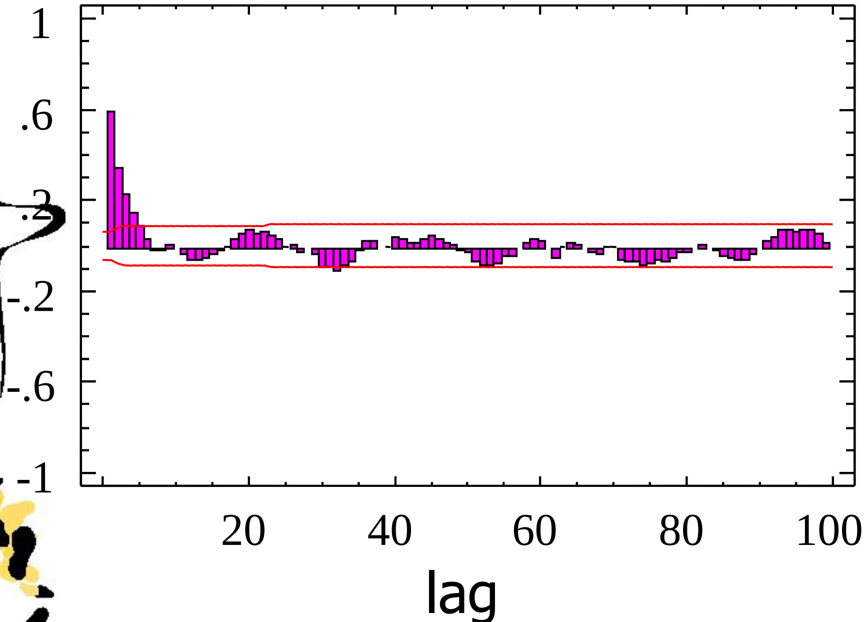
Partial Autocorrelations



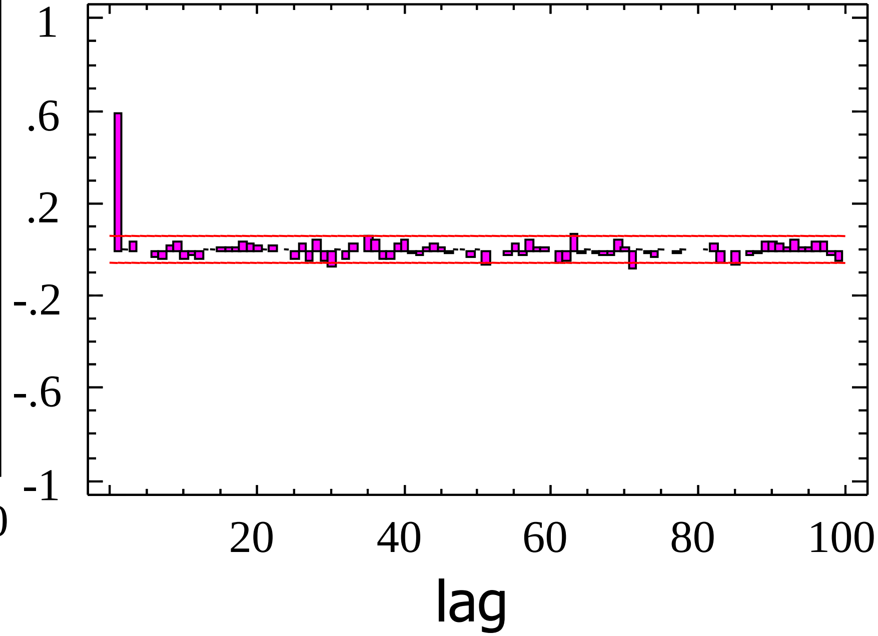
Identification of the order of integration. Examples

First-order Autoregressive process AR(1) $Y_t = 0.6Y_{t-1} + \varepsilon_t$

Autocorrelations for AR(1)



Partial Autocorrelations for AR(1)



Seasonal time series

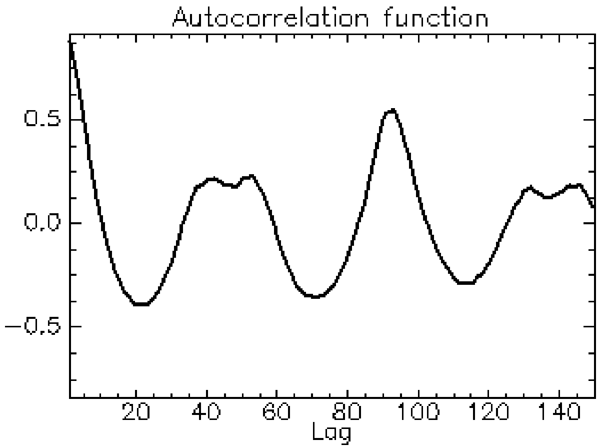
- Time series presents behaviors that are repeated with any frequency (daily, weekly, monthly, quarterly, yearly, 8-days in MODIS data, …).

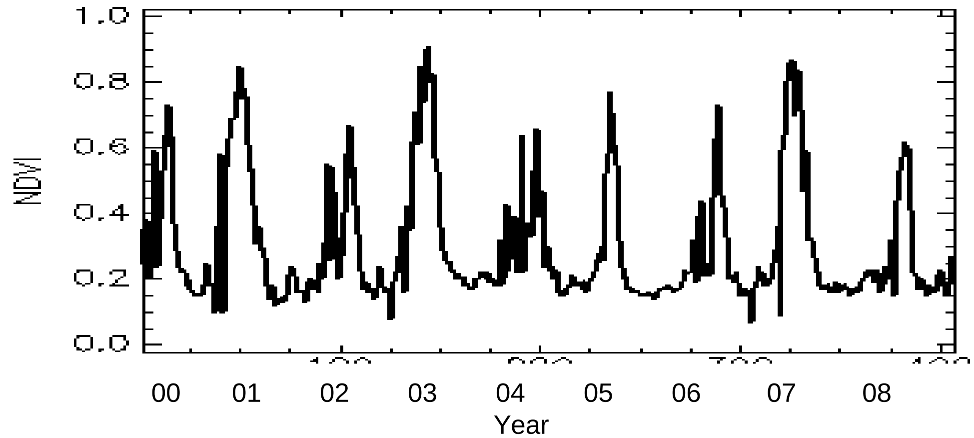
- The order of this periodicity is usually induced by the data collection Frequency.

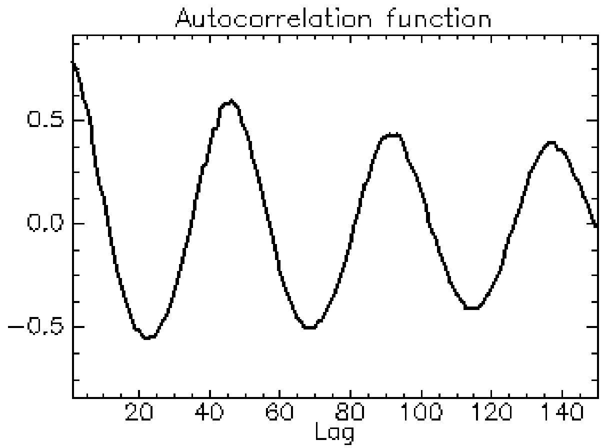
- The Box-Jenkins methodology takes into account the seasonal factor and incorporates it in the development of the models.

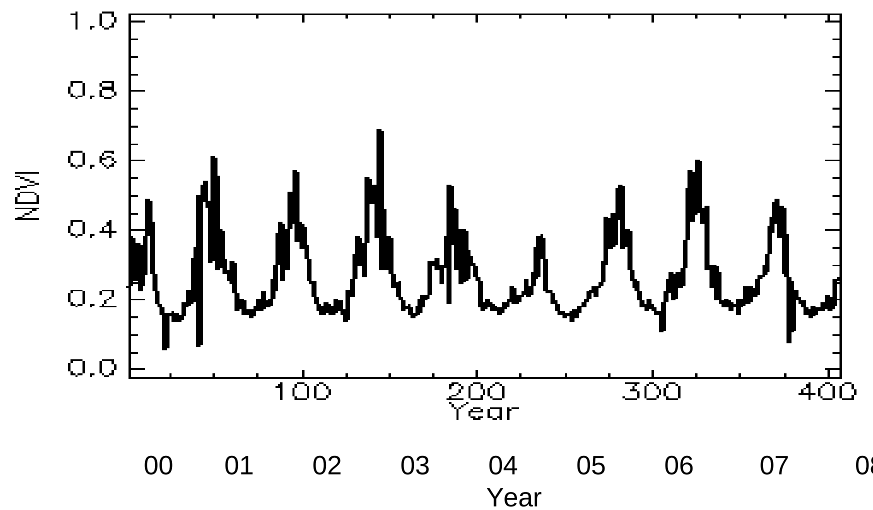
Seasonal time series

Crops





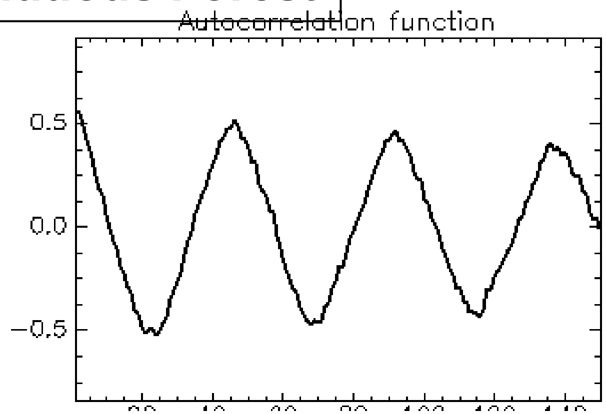


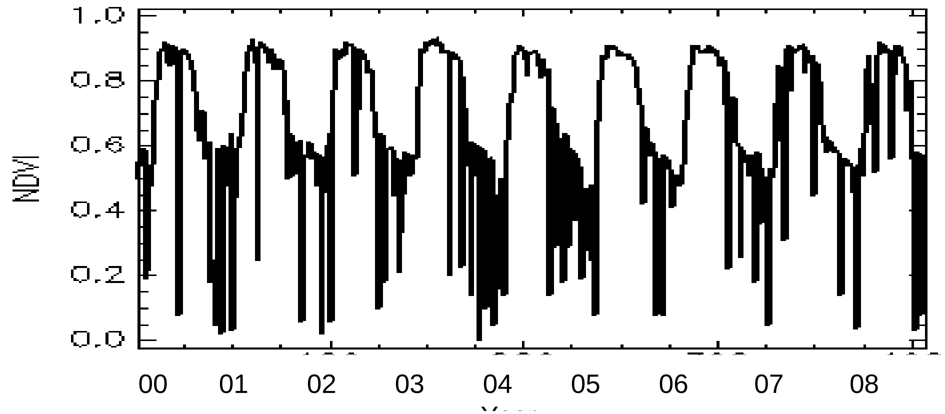


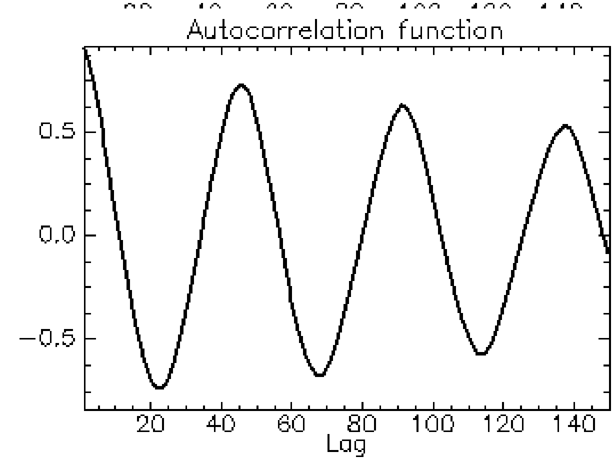
lag

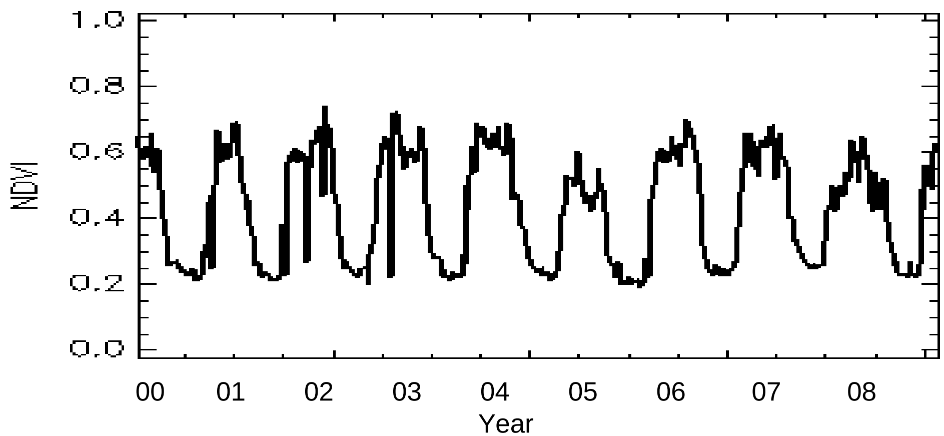
Seasonal time series

Deciduous Forest









Dehesa

Seasonal time series

The specific aspects are:

1. Identify the stationarity of the seasonal component. If the series is not stationary, take the differences of the seasonal order *s*:

   if $s = 4$, take the 1st dif. of order 4:

$$\nabla^4 y_t = (1 - B^4) y_t = y_t - y_{t-4}$$

   it includes 4 unit roots: ± 1 , $\pm i$:

$$\nabla^4 = (1 - B)(1 + B)(1 - iB)(1 + iB)$$

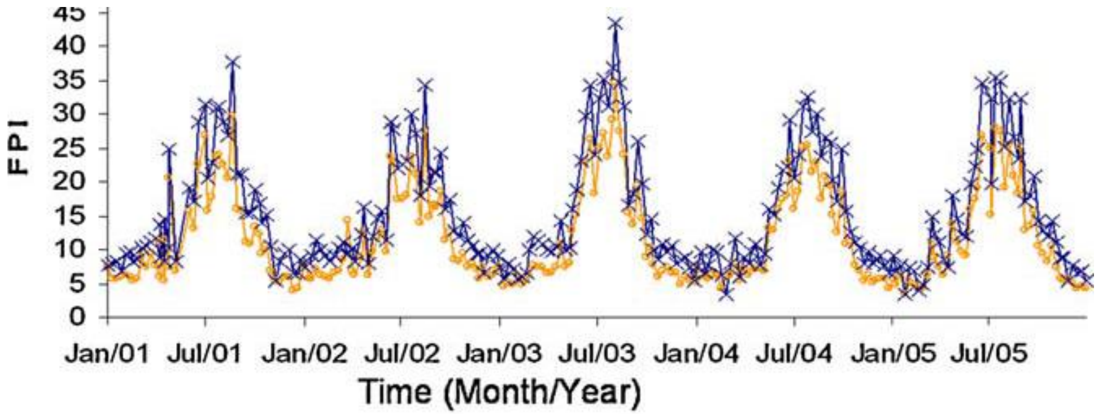
   if $s = 12$ take first differences of order 12:

$$\nabla^{12} y_t = (1 - B^{12}) y_t = y_t - y_{t-12}$$

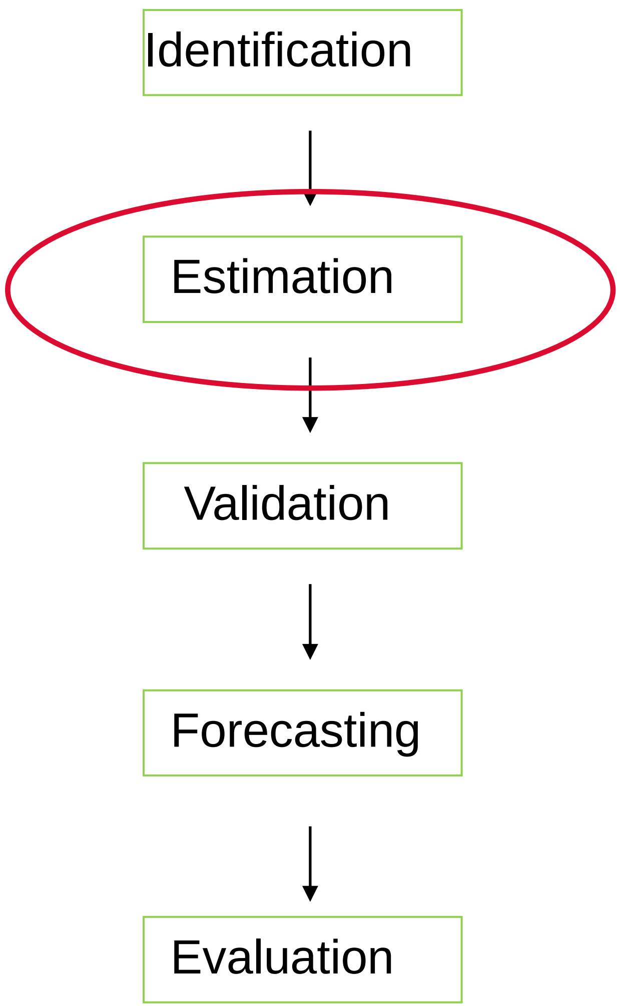
Seasonal time series

2.- Specify the seasonal parameters in the model and estimate them by the general estimation methods.

3.- Validate the model using the general methods checking the presence of seasonal autocorrelation in the residuals.



Box & Jenkins Approach



Box & Jenkins Approach: Estimation

$$y_t = \phi y_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_n y_{t-p} + \varepsilon_t$$

The *Student t*-statistics = coefficient / standard error

P-value= critical significance level: level of admissibility of H_0

$$\bar{R}^2 = 1 - (1 - R^2)\frac{T-1}{T-k}$$

$$F = \frac{R^2 / k - 1}{(1 - R^2) / T - k}$$

$$DW = \sum_{i=2}^{T} (\hat{\varepsilon}_i - \hat{\varepsilon}_{i-1})^2 / \sum_{i=1}^{T} \hat{\varepsilon}_i^2$$

Box & Jenkins Approach: Estimation

$$AIC = \frac{-2\ell}{T + 2kT}$$

$$SC = -2\ell / T + (k \log T) / T$$

where: $\ell = -T \; / \; 2(1 + \log(2\pi) + \log(\widehat{\varepsilon}'\widehat{\varepsilon} \; / \; T)$

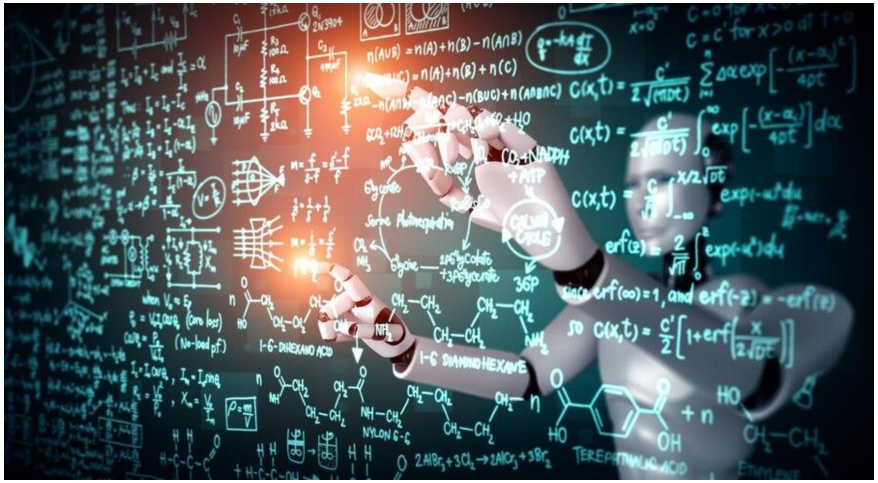
Box & Jenkins Approach: Validation

<u>ARIMA models:</u>

Find a model that represents the dynamics of the studied variable

<u>Validation procedure:</u>

Extract the residual series of the estimated model and examine their dynamics by the analysis of its ACF and PACF.



Box & Jenkins Approach: Validation

Requirements for model adequacy:

a) The residual must be white noise.

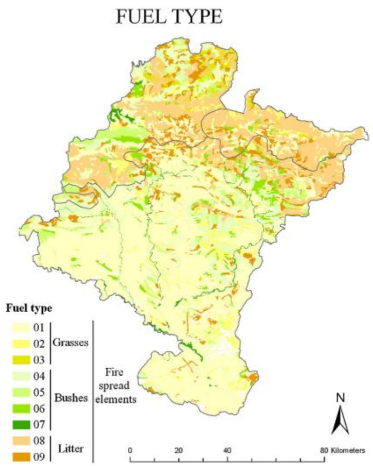
b) The estimated model must be stationary.

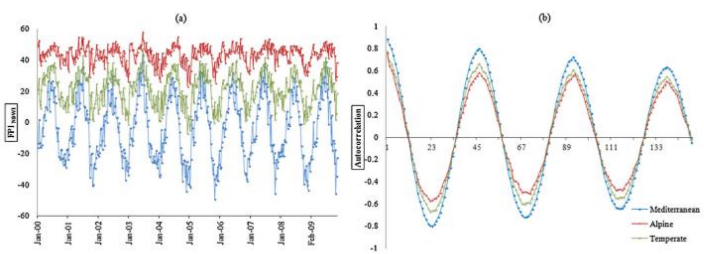
c) The coefficients must be:

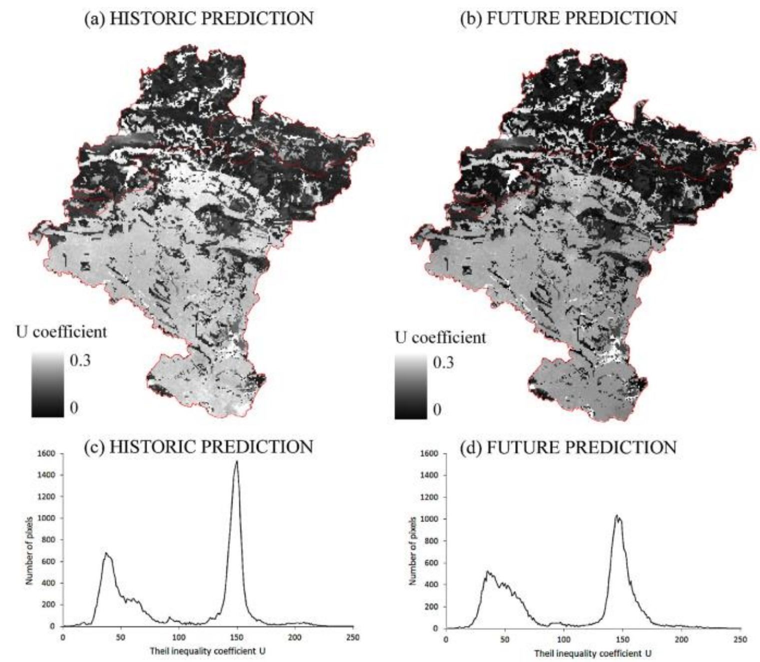
- Statistically significant.
- Non or weakly correlated
- Sufficiently representative of the series.

d) Select the model with the greatest degree of adjustment.

Example







Modelling and forecasting MODIS-based Fire Potential Index on a pixel basis using time series models. Huesca et al. 2014

Conclusions

- Time series are characterized by trends, seasonality, cycles, structural changes and outliers
- Box and Jenkins approach has to be applied to stochastic process.
- Box and Jenkins approach's steps: identification, estimation, validation, forecasting and evaluation
- The autocorrelation and partial autocorrelation function help you to identify periodic components and the orders p and q of the ARMA model
- The series with a marked seasonal component must be differenced

UNIVERSITY OF TWENTE.

THANK YOU!!

MARGARITA HUESCA MARTINEZ