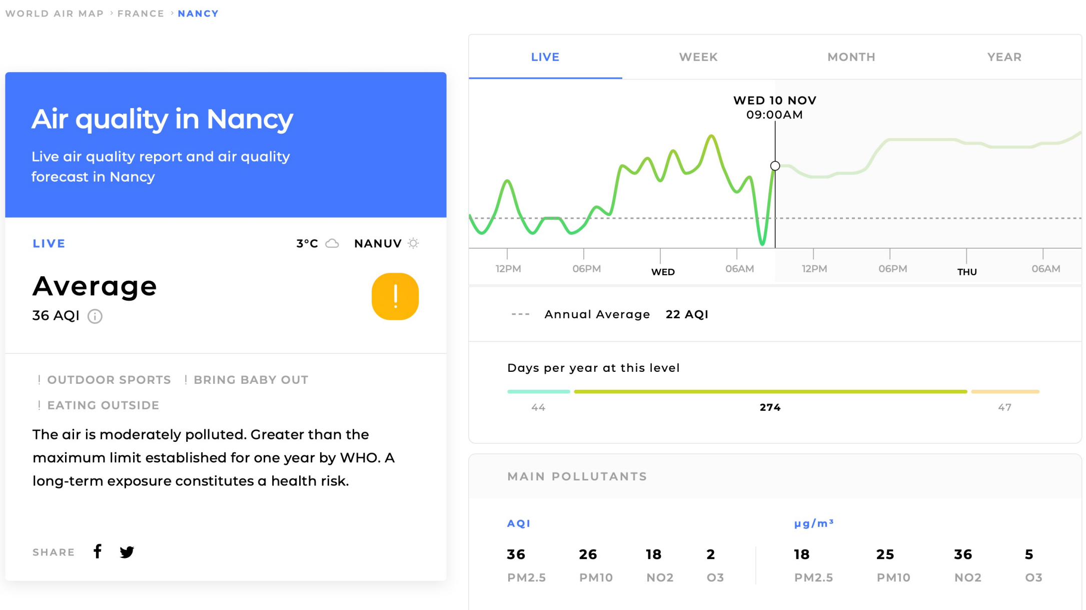
Introduction to time series analysis and forecasting

Hugues GARNIER

hugues.garnier@univ-lorraine.fr

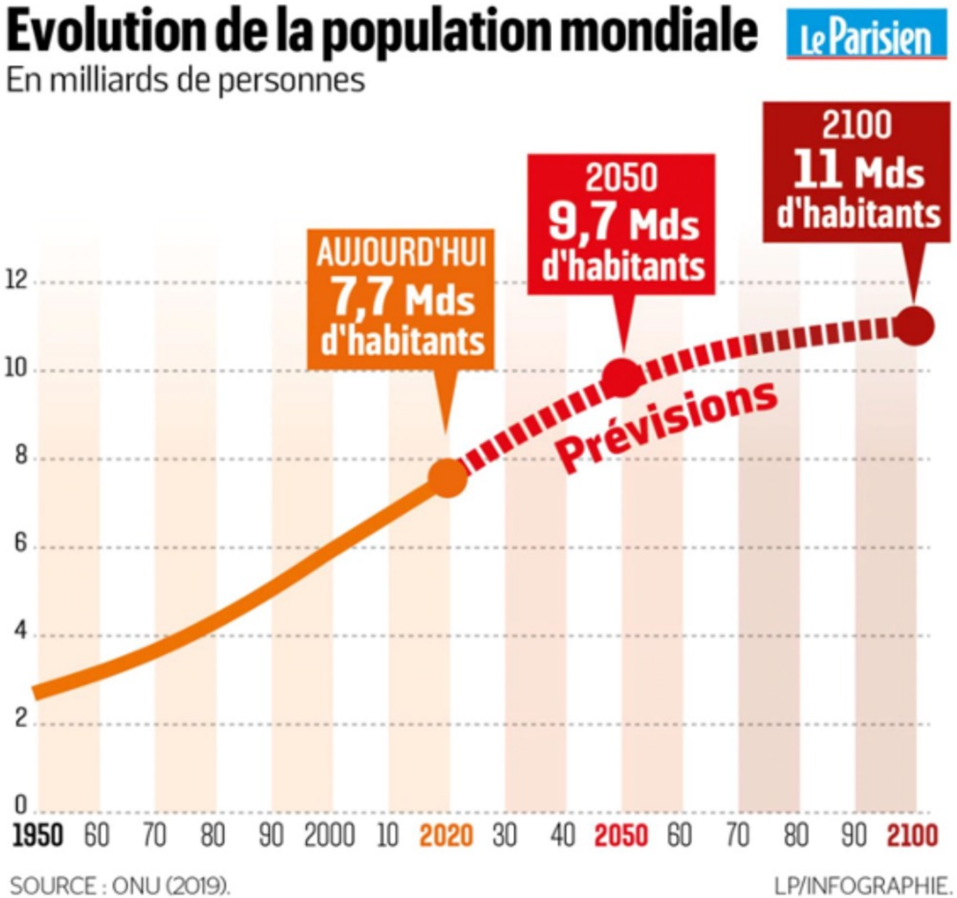
We rely on forecasts in our daily life



plumelabs.com/en/

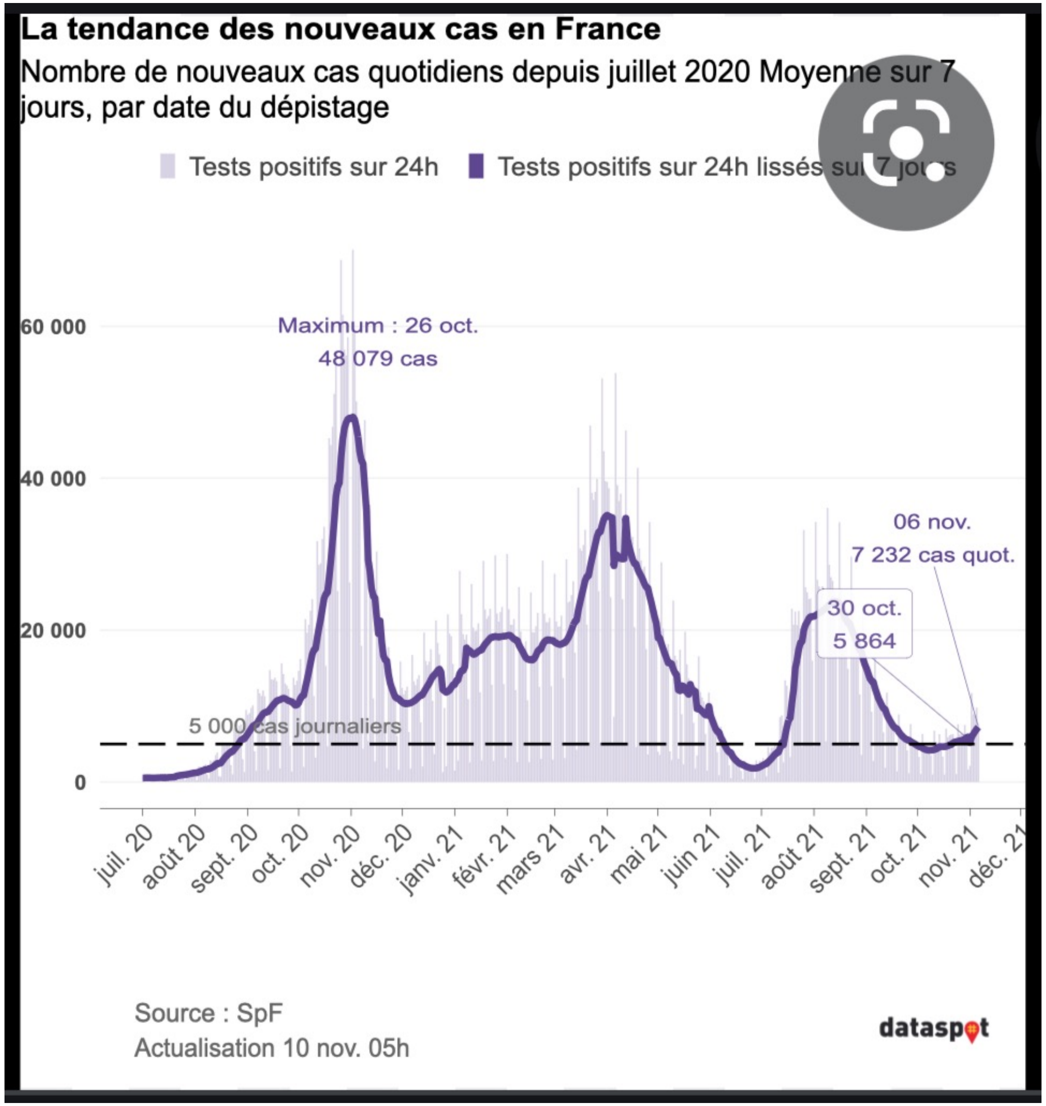
- *Weather forecasts* help you decide if you should bring an umbrella before leaving home
- *Pollution forecasts* help you decide to better plan your sport activities or take adequate measures to reduce exposure (*Air pollution costs every human an average 2 years life expectancy…)*

We rely on forecasts in our daily life
Forecasting is a natural part of human behaviour



Source : Le Parisien (20/06/19)

Time series modelling & forecasting methods: A major decision-making tool



Course organization and prerequisites

- Organization
  - 6h00 of lecture
  - 10h00 of tutorials
- Skill assessments
  - Team project where you will work on a forecasting problem using real-life data
  - Oral presentation of your time series analysis and forecasting
- Prerequisites
  - A sound knowledge about probability and statistics
  - Regression analysis
  - Basic programming proficiency in Matlab

Time series: definition

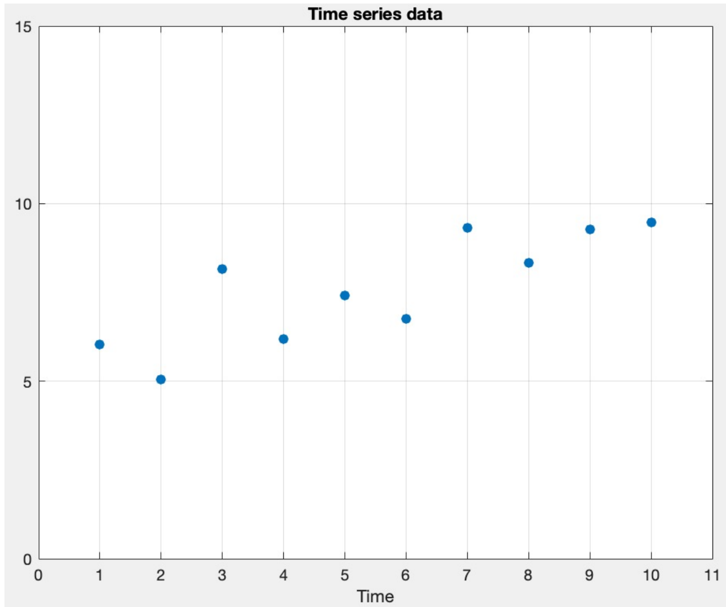
- A time series is
  - a series of data points indexed in time order
  - a sequence taken at successive equally spaced points in time
  - it is a sequence of discrete time data

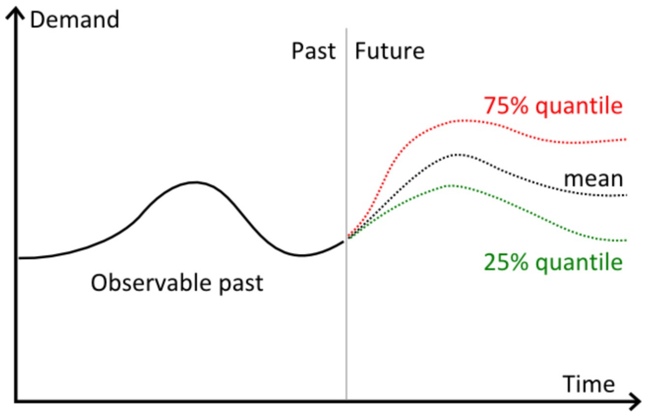
$$x_t = (x_1, \dots, x_N)$$

where t represents time in second, hour, day, month, quarter, year,…

- The main goal is to forecast the future values of the time series

$$x_{N+1}, x_{N+2} \dots$$

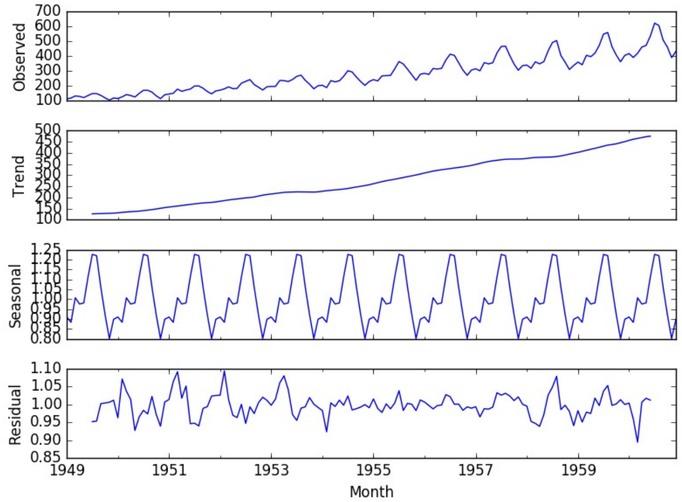




H. Garnier

Time series analysis

- Time series analysis is concerned with:
  - Identifying patterns



  - Modeling patterns

$$\left(1-\sum_{i=1}^{p}\phi_i L^i\right)(1-L)^d X_t = \left(1+\sum_{i=1}^{q}\theta_i L^i\right)\varepsilon_t$$

  - Forecasting values with uncertainty intervals

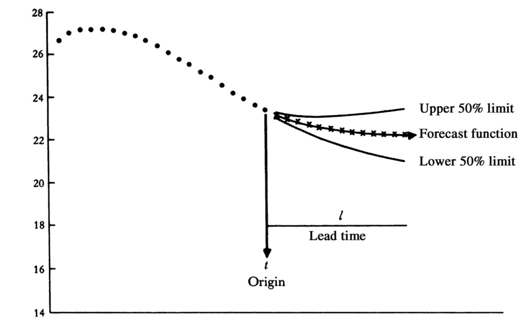


FIGURE 1.1 Values of a time series with forecast function and 50% probability limits.

Time series analysis: fields of application

- Until recently, the use of time series data was mainly related to fields of science, such as
  - economics
  - finance
  - astronomy
  - industry

- However, in recent years, as the ability to collect data improved with the use of digital devices such as computers, mobiles, sensors, or satellites

- Time series data analysis is now exploited everywhere !

Historical developments for time series analysis

- Tools available find their sources in four historical phases:
  - *Graphic representation method*: diagrams appeared in Astronomy, the oldest known dating back to the 10th century
  - *Deterministic methods*: they appeared in the 18th and 19th centuries, there are two very important ways:
    - frequency analysis of a time series (Fourier analysis)
    - decomposition of a time series into trend, cyclical, seasonal and accidental components
  - *Model-based stochastic methods*: they emerge in the middle of the 20th century, as for example
    - the ARIMA model method
  - *Data-based methods*: they emerge in the beginning of the 21st century with the deluge of data collected every day. It goes beyond our ability to observe, analyze, and exploit them
    - the machine learning and deep learning methods

An overview of 8 different time series analysis methods

www.youtube.com/watch?v=d4Sn6ny_5LI

11 mn

Introduction to time series analysis and forecasting
Course outline

Before exploring recent machine learning and deep learning methods, it is good idea to ensure you have tried classical and statistical time series forecasting methods. These methods are still performing well on a wide range of problems when the number of data is relatively limited

Course outline

I. Main characteristics of time series data

II. Time series decomposition

III. Basic time series modelling and forecasting methods

IV. Stochastic time series modelling and forecasting: ARIMA method

Software requirements for the course

We will make extensive use of Matlab and of the Econometrics Toolbox

All Examples Functions Apps

Econometrics Toolbox

Model and analyze financial and economic systems using statistical methods

Econometrics Toolbox™ provides functions for analyzing and modeling time series data. It offers a wide range of visualizations and diagnostics for model selection, including tests for autocorrelation and heteroscedasticity, unit roots and stationarity, cointegration, causality, and structural change. You can estimate, simulate, and forecast economic systems using a variety of modeling frameworks. These frameworks include regression, ARIMA, state-space, GARCH, multivariate VAR and VEC, and switching models. The toolbox also provides Bayesian tools for developing time-varying models that learn from new data.

Get Started
Learn the basics of Econometrics Toolbox

Data Preprocessing
Format, plot, and transform time series data

Model Selection
Specification testing and model assessment

Time Series Regression Models
Bayesian linear regression models and regression models with nonspherical disturbances

Conditional Mean Models
Autoregressive (AR), moving average (MA), ARMA, ARIMA, ARIMAX, and seasonal models

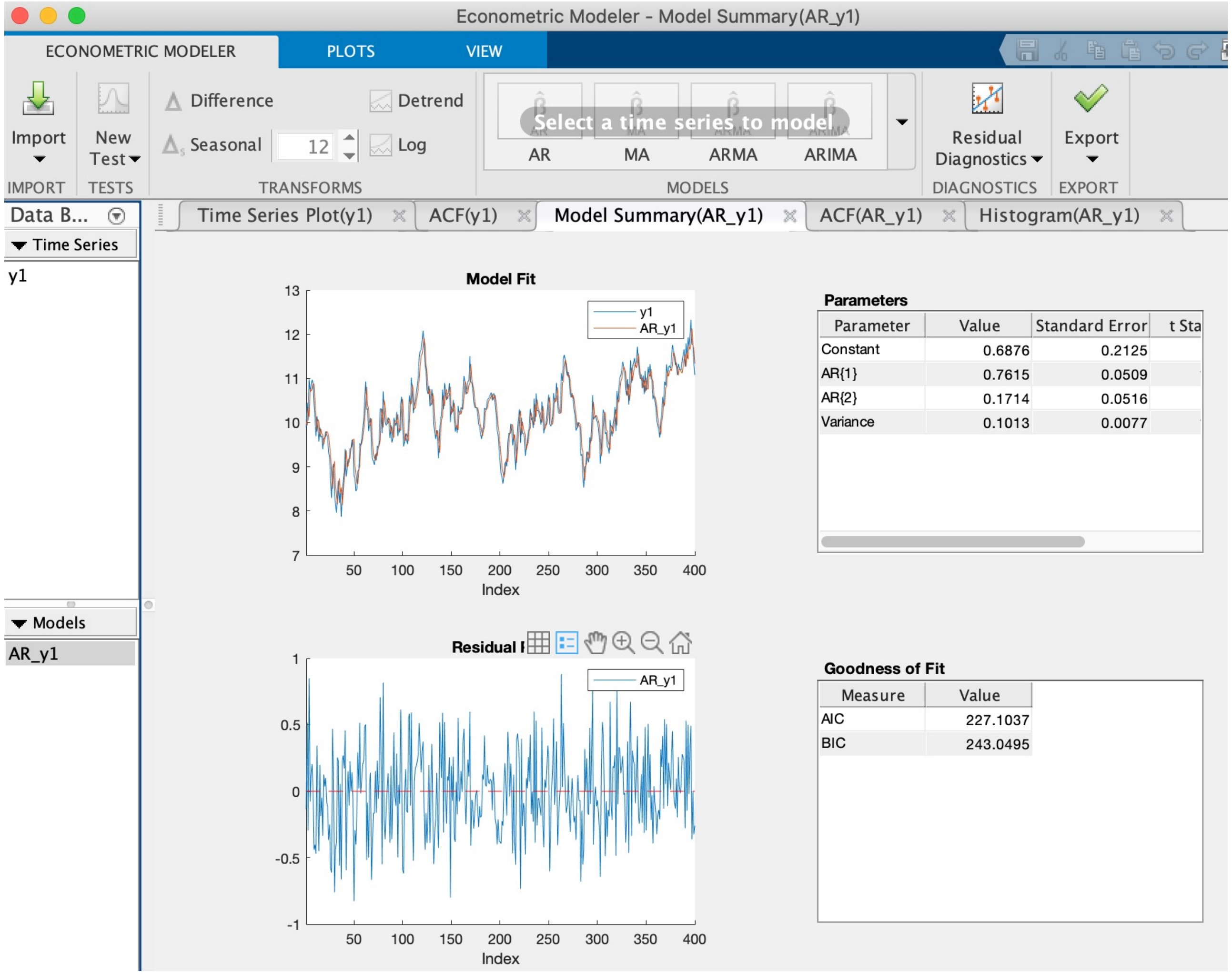
Conditional Variance Models
GARCH, exponential GARCH (EGARCH), and GJR models

Multivariate Models
Cointegration analysis, vector autoregression (VAR), vector error-correction (VEC), and Bayesian VAR models

Markov Models
Discrete-time Markov chains, Markov-switching autoregression, and state-space models

Software requirements for the course

We will also make use of the recent *Econometrics Modeller* App



H. Garnier

UNIVERSITÉ DE LORRAINE

POLYTECH NANCY

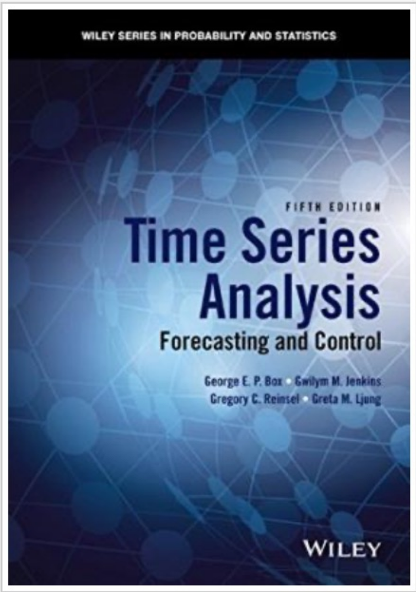
There is a wealth of books on time series !!!

(most are focused on using the R platform)

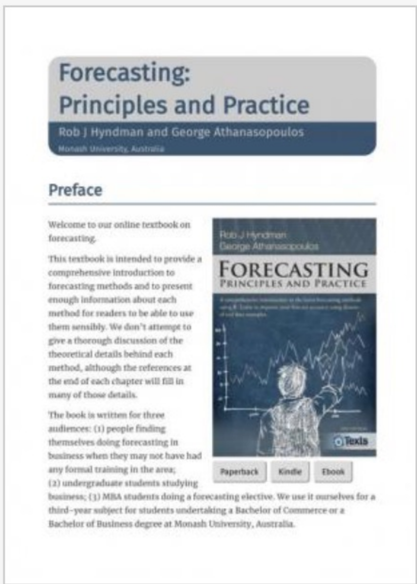
Course website & recommended textbooks

- Website of the course
  - w3.cran.univ-lorraine.fr/hugues.garnier/?q=content/teaching

- Recommended textbooks

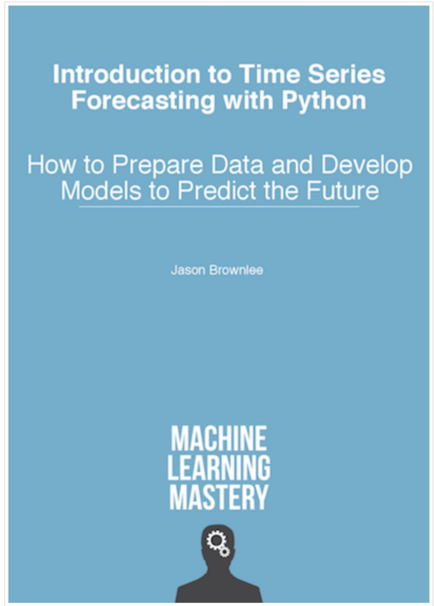


The bible !



Free online textbook
otexts.com/fpp2/

Free online videos
bit.ly/2qM9eHL



A masterpiece
for Python fans

Key takeaways from the course

- Introduction to classical time series analysis methods along with implementation on real-life data examples
  - Understand the importance of forecasting for planning and decision making
  - Have a basic knowledge of the main fields where forecasting is used
  - Be familiar with the difference between descriptive and forecasting goals
  - Know how to visualize time series data for discovering their main components
  - Be familiar with the concepts of stationarity and autocorrelation
  - Know ARIMA methods and be able to choose adequate methods for different types of data
  - Understand how different models and methods can be used for forecasting
  - Know how to evaluate and compare the performance of forecasting methods

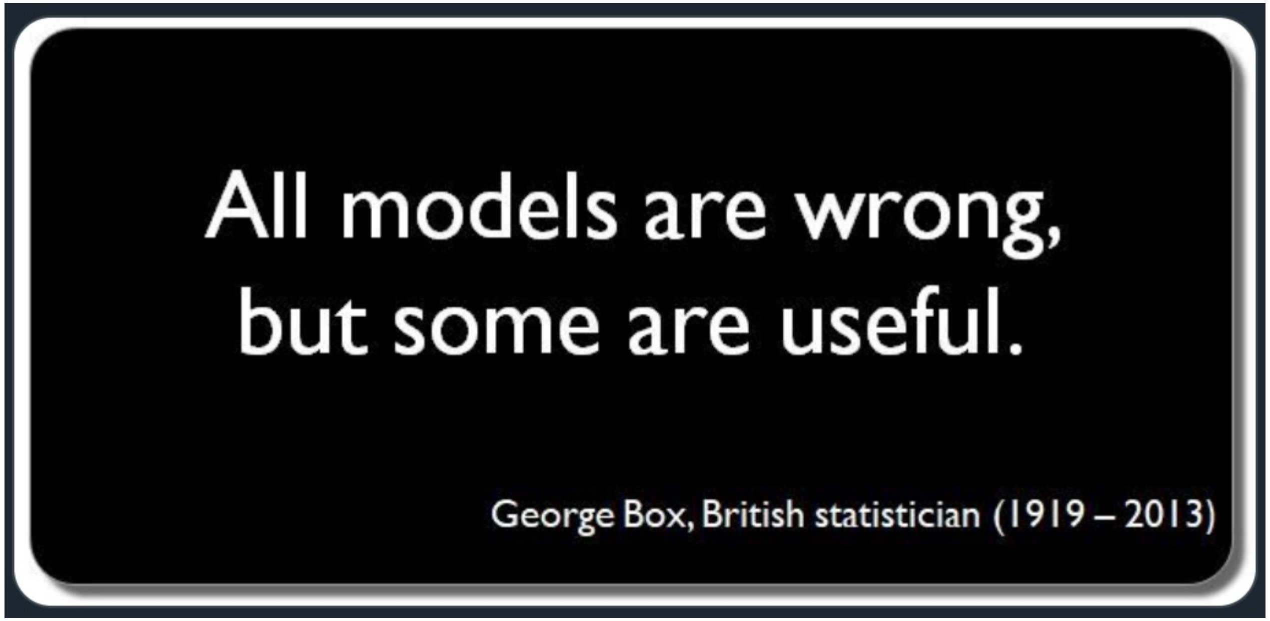
Time series modelling & forecasting

is a discipline of Data Science

that requires practical skills and experience

Do not forget this quote from

George E. P. Box



Course outline

Introduction to time series analysis and forecasting

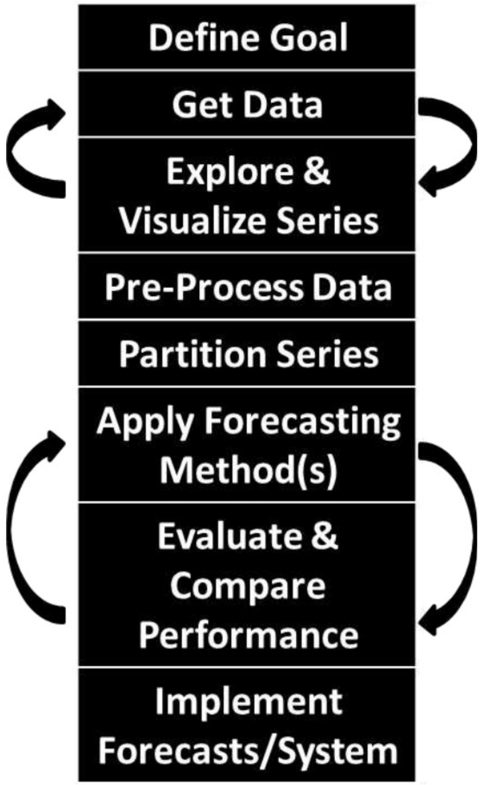
I. Main characteristics of time series data

II. Time series decomposition

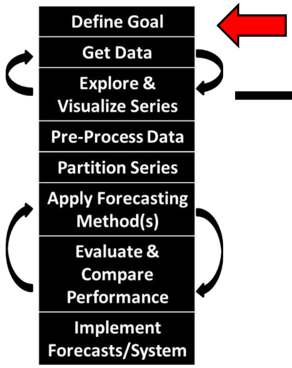
III. Basic time series modelling and forecasting methods

IV. Stochastic time series modelling and forecasting:
The Box-Jenkins method for ARIMA models

Steps in the statistical iterative modeling process



From Galit Schmueli
Practical time series forecasting with R



Time series analysis: Goals

Description

- Determine the main features of the series: trend, seasonal, cyclical patterns

Explanation

- Understand the mechanism generating the series. Find a model to describe the time dependence in data, assess the impact of an event

Forecasting

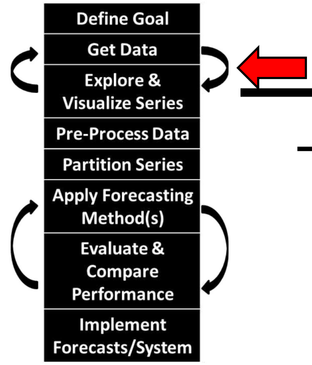
- Forecast the future value(s) based on the past

Control

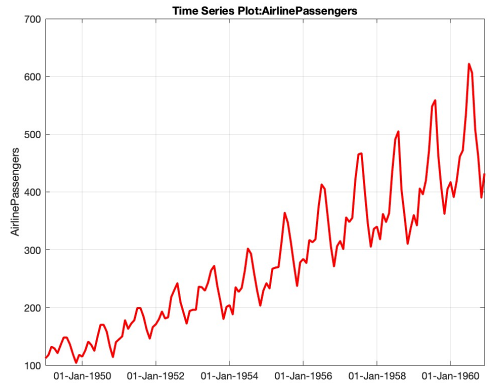
- of the process producing the time series

Predictive maintenance

- to predict when equipment failure might occur and to prevent its occurrence by performing maintenance



Time series: Visualization

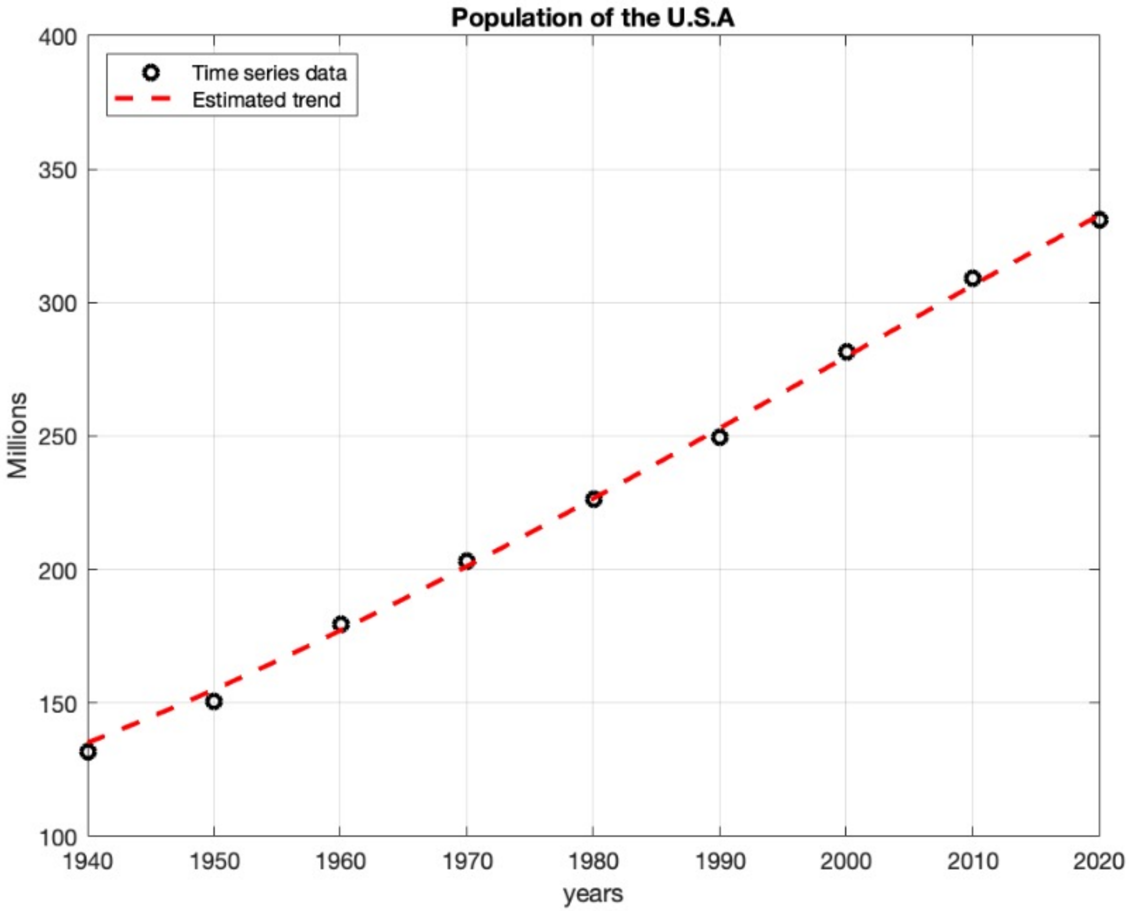


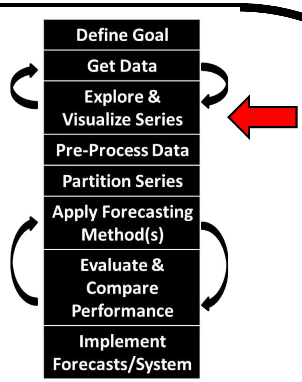
- The first thing to do is to make a time plot and look for patterns

- The following may be observed:
  - a ***trend pattern***, which is a long term increase or decrease in the variable of interest
  - a ***seasonal/periodic pattern*** appears when a time series is affected by seasonal factors such as time of the year or the day of the week
  - a ***cyclical pattern***, which is one where there are rises and falls but not of regular period, generally thought of as longer in time, e.g., several years
  - ***no special or random pattern***, the irregular variation seems to be stochastic

Combinations of the above first three types of pattern occur frequently

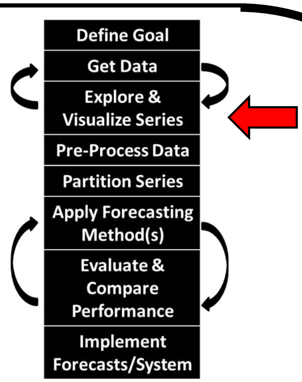
Polynomial pattern: example

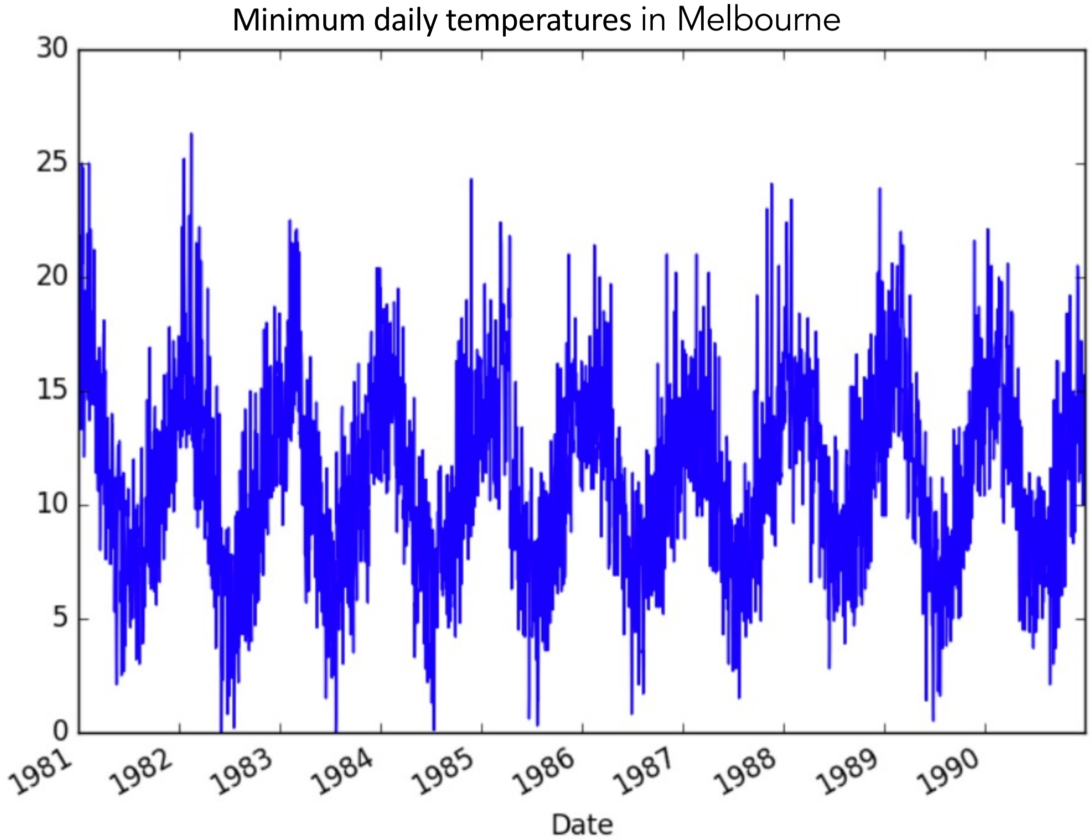




It seems to be a simple linear pattern. It could be quadratic or cubic, …

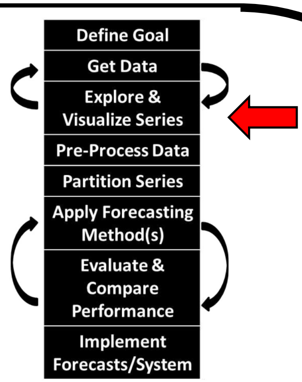
Seasonal pattern: example

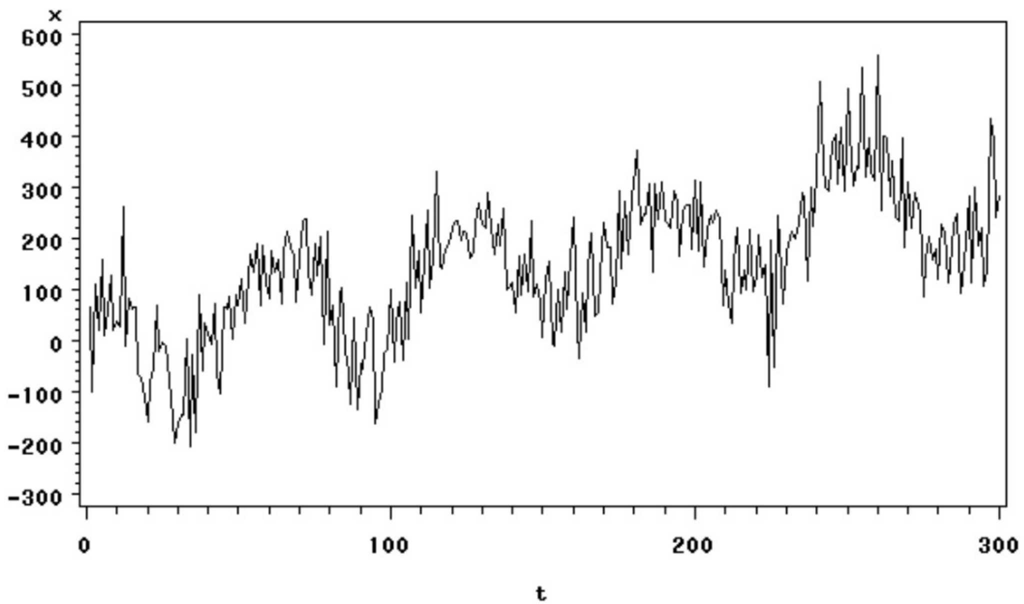




It seems to be a clear annual periodicity: we talk about seasonal pattern, which occurs when time series are affected by seasonal factor (day of the week, month of the year. . .). The period is fixed and known

Combination of patterns: example



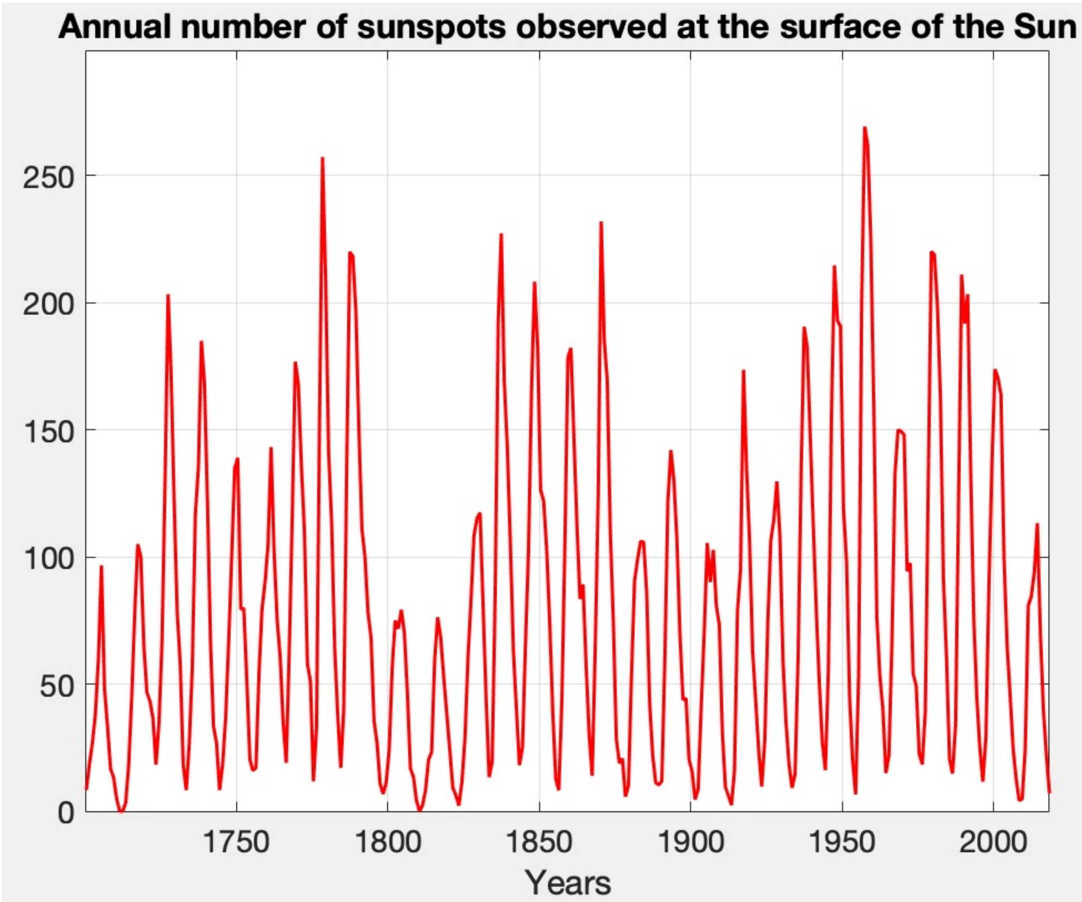


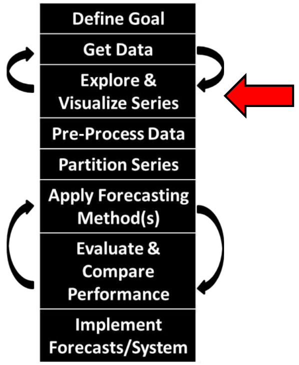
It seems to be a linear pattern and a seasonal pattern

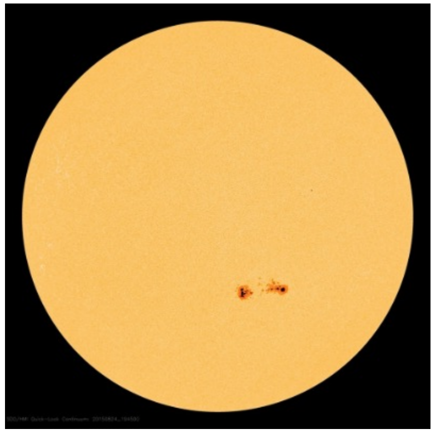
UNIVERSITÉ DE LORRAINE

POLYTECH NANCY

Cyclical pattern: example

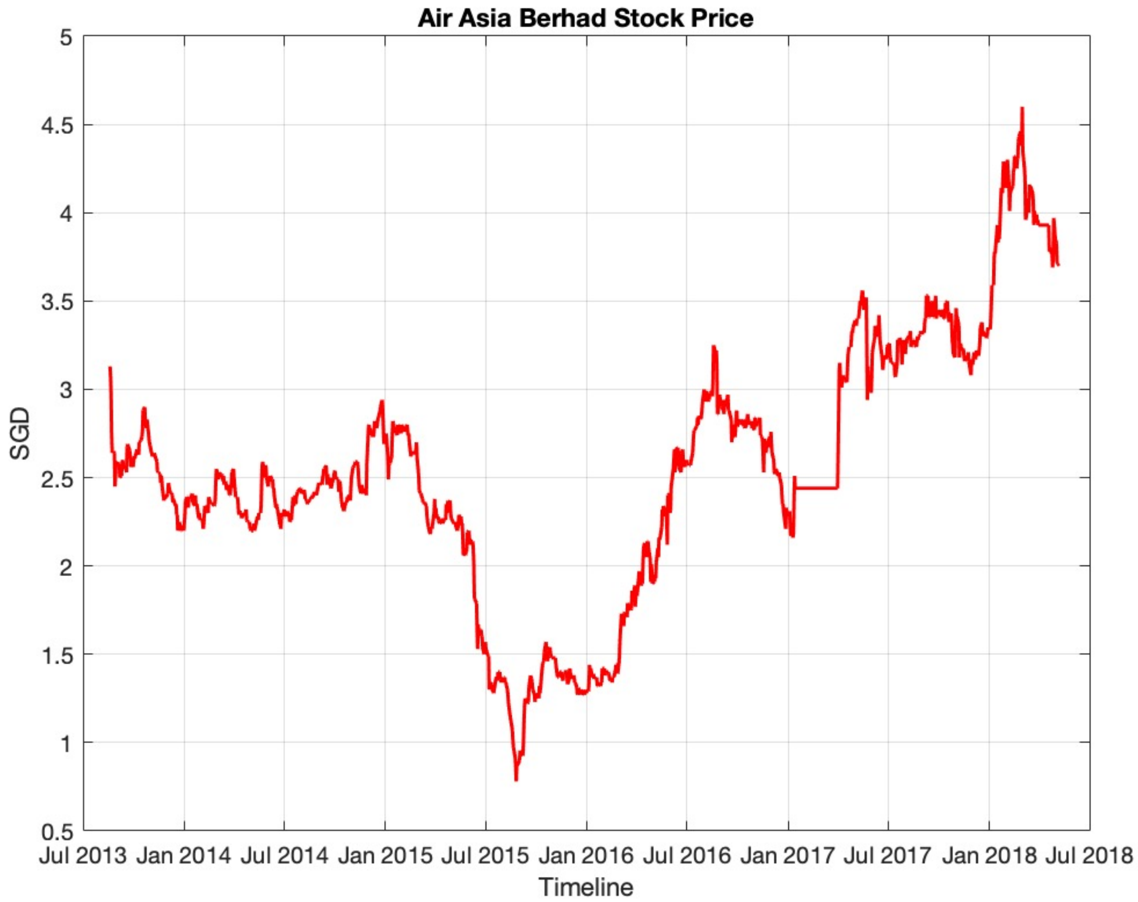


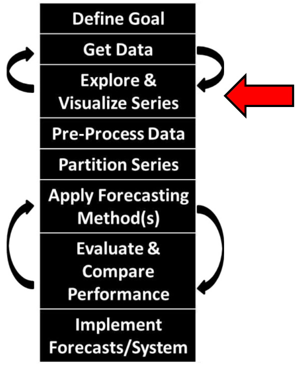




It seems to be a cyclical pattern: rises and falls are every 11 years

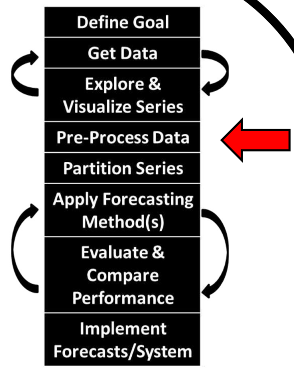
No special pattern: example





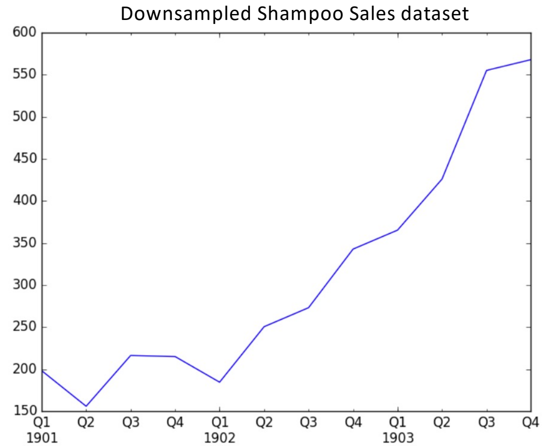
It seems to be no special pattern as it is often the case for daily stock market time series

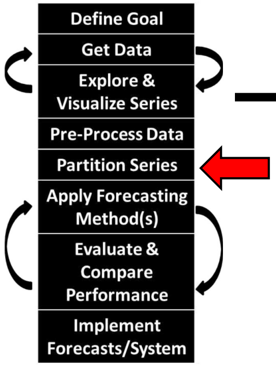
Pre-process data



Time series data often requires cleaning, scaling, and even transformation. For example:

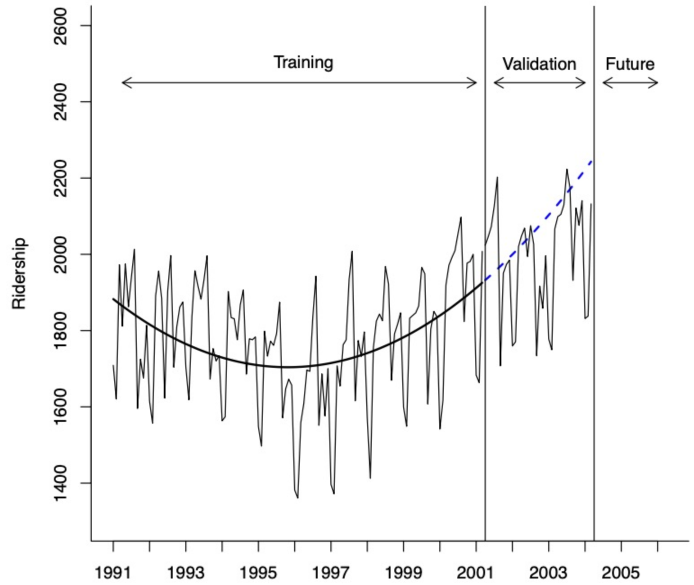
- Outliers
  - data can be corrupted or extreme outlier values that need to be detected and handled
- Missing
  - Data can have gaps or missing data that need to be interpolated or imputed
- Baseline (offset) removing
- Pattern removing
- Smoothing
- Filtering
- Resampling
  - Data can be provided at a frequency that is too high to model or is unevenly spaced through time requiring downsampling for use in some models or methods





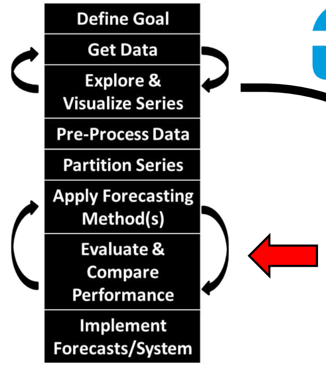
Partition the time series data

- To address the problem of overfitting, an important preliminary step before applying any forecasting method is data partitioning where the series is split into two periods



- The model is learned using the training dataset only
- The estimated model is used to make predictions on the validation dataset and see how it performs
- The evaluation of these predictions will provide a good indication for how the model will perform when it will be used operationally (in the future)

Common predictive accuracy measures



- Given $\{y_1, \ldots, y_N\}$ actual observations of a time series $\{y_t\}$, and let $\hat{y}_t$ be the forecast value at time t
- We can calculate the *residuals* or *forecast errors* and $e_t = y_t - \hat{y}_t$
- Common predictive accuracy measures based on the residuals are:
  – Root Mean Square Error (RMSE)

$$\text{RMSE}=\sqrt{\frac{1}{N}\sum_{t=1}^{N}(y_t-\hat{y}_t)^2}$$

  – Mean Absolute Percentage Error (MAPE)

$$\text{MAPE}=\frac{100}{N}\sum_{t=1}^{N}\left|\frac{y_t-\hat{y}_t}{y_t}\right|$$

Compute some function of the forecast error:	Then, average across all records:
e_t	*Average Error*
$\lvert e_t \rvert$	*Mean absolute error* (MAE)
$(e_t)^2$	*Mean Squared Error* (MSE) or take a root (RMSE)
$\lvert e_t / y_t \rvert$x100%	*Mean absolute % error* (MAPE)

- Video from Galit Schmueli: *bit.ly/2qM9eHL*
  – Performance 4: Predictive metrics & charts

Course outline

Introduction to time series analysis and forecasting

I. Main characteristics of time series data

II. Time series decomposition

III. Basic time series modelling and forecasting methods

IV. Stochastic time series modelling and forecasting: The Box-Jenkins method for ARIMA models

Time series: basic characteristics

- **Trend component**
  - long-term increase or decrease in the data over time
- **Seasonal component**
  - influenced by seasonal factors (*e.g.* quarter of the year, month, or day of the week)
  - exact repetition in regular pattern (seasonal series often called periodic, although they do not exactly repeat themselves)
- **Cyclical component**
  - data exhibit rises and falls that are not of a fixed period
- **Random or stochastic component**
  - irregular variation data without any special pattern
- **Correlation** between the series and its past value
  - we need to build a model that is able to deal with such dependencies

Time series: standard decomposition model

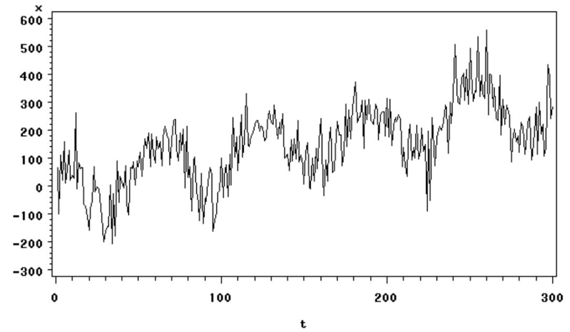
- **Data:** y_t where t indexes time, *e.g.* hour, day, month, year
- **Standard model:** $y_t = \mathrm{f}(m_t, s_t, x_t)$
  - m_t is a trend-cycle component
  - s_t is a seasonality component
  - x_t is a stationary random component
- Standard functional forms for f
  - Additive (linear): $y_t = m_t + s_t + x_t$
  - Multiplicative (non linear): $y_t = m_t \times s_t \times x_t$
  - Mixed (non linear): $y_t = m_t \times s_t + x_t$

Time series: standard decomposition model

- Additive model

$$y_t = m_t + s_t + x_t$$

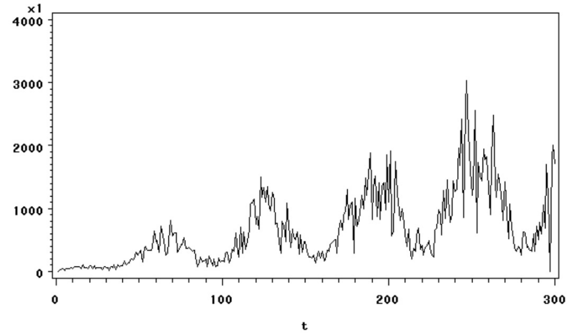
  – assumes constant variability of the time series



- Multiplicative model

$$y_t = m_t \times s_t \times x_t$$

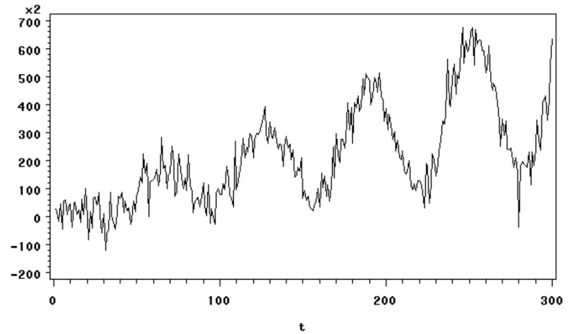
  – assumes that variability (seasonal and random) is amplified with trend



- Mixed model

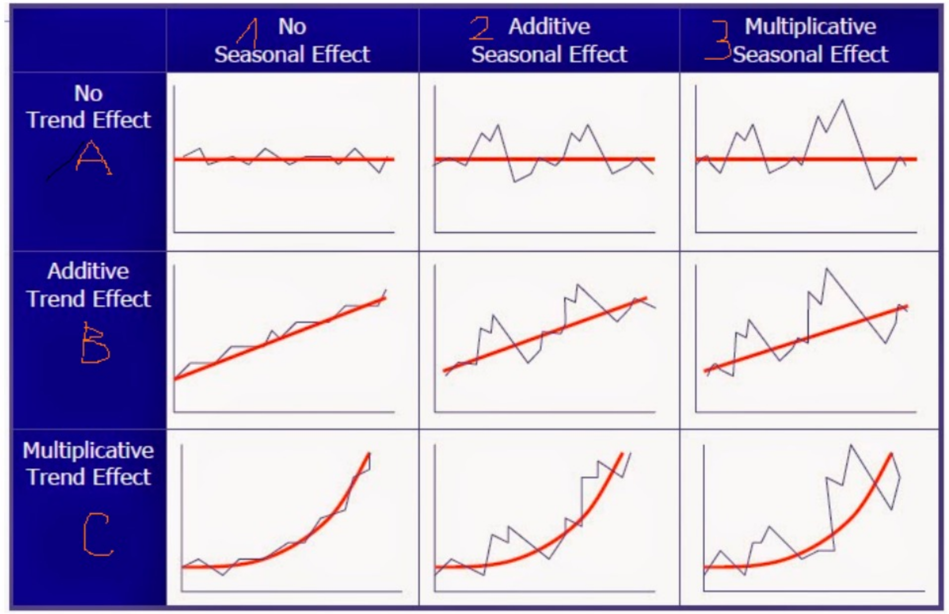
$$y_t = m_t \times s_t + x_t$$

  – assumes that variability is amplified with trend but the random component remains constant over time



Guide to choose the model form

- Visual inspection can help to decide whether the model should be additive or multiplicative is given by the patterns below, suggested first by Pegels in 1969



  - The additive model assumes constant variability of the time series
  - The multiplicative model assumes that variability is amplified with trend

H. Garnier

Time series analysis: The standard decomposition procedure

- The primary goal of time series decomposition is to provide the analyst with a better understanding of the underlying behavior and patterns of the time series
- Usual assumption: additive model

$$y_t = m_t + s_t + x_t$$

  - If the model is multiplicative, apply first a log transformation on the data

$$Y_t = \log(y_t)$$

- **Standard parametric decomposition procedure**
  - m_t and s_t are first estimated
  - they are subtracted from y_t to have left the stationary process x_t
  - x_t can be further analyzed and modelled using time series modeling approaches if necessary

Modelling the trend component

$y_t = \mathrm{f}(m_t, s_t, x_t)$

- Polynomial model *(in t)*

$$m_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \cdots + \beta_p t^p$$

- Which polynomial order to select ?
  - Use model selection to select among the predicting variables (t, t^2,…, t^p)
  - Cautious! Strong correlation among the predicting variables
- Commonly used small order polynomial ($p=1$ or 2)
  - linear: $m_t = \beta_0 + \beta_1 t$
  - quadratic: $m_t = \beta_0 + \beta_1 t + \beta_2 t^2$
- Parameters
  - Estimated by using linear regression where the predicting variables are (t, t^2,…, t^p)

- Exponential model *(in t)*

$$m_t = m_0 e^{at}$$

- Parameters
  - Estimated by using linear regression where the predicting variables are (t, t^2,…, t^p) after the use of the log for the exponential case

Modelling the seasonal component

$y_t = \mathrm{f}(m_t, s_t, x_t)$

- Single periodic component

$$s_t = s_{t+d}$$

- How to determine the period *d* ?
  - Visual inspection
  - Fourier analysis *(see course on digital signal processing from last year)*

- Harmonic seasonal model
  - uses of sine and cosine functions to describe the pattern of fluctuations seen across period

$$s_t = \sum_{i=1}^{S/2} a_i \cos\left(\frac{2\pi i t}{S}\right) + b_i \sin\left(\frac{2\pi i t}{S}\right)$$

  - *S* is the number of seasons, a_i and b_i are the parameters to be estimated

Modelling the random component
Stationarity assumption

$y_t = \mathrm{f}(m_t, s_t, x_t)$

- x_t is modelled as a realization of a *stochastic process* X_t

- A *stochastic process* is a collection of random variables $\{X_t, t \in T\}$ defined on a probability space (Ω, F, P)

- The stochastic process X_t is assumed to be stationary
  - Its probability distribution does not change when shifted in time
  - Realizations of a stationary stochastic process, vary over time in a stable manner about a fixed mean
  - It is *(weakly) stationary* if *it* can be described by its first two moments only
    - Mean, variance
    - AutoCorrelation Function (ACF)
      - ACF is also very useful for describing/testing the stationary property of a random component

Some intuitions for (weak) stationary time series

- The properties of one section of a data are much like the properties of the other sections of the data
  - no systematic change in the mean *i.e.*, no trend
  - no systematic change in variation
  - no periodic fluctuations
- For non-stationary time series, some transformations (such as differencing, decomposition or logarithm, …) can be applied to get stationary time series

Moments of a probability distribution
A brief review

- Moments of a random variable X with density $f_X(x)$:
  - *l-th moment*

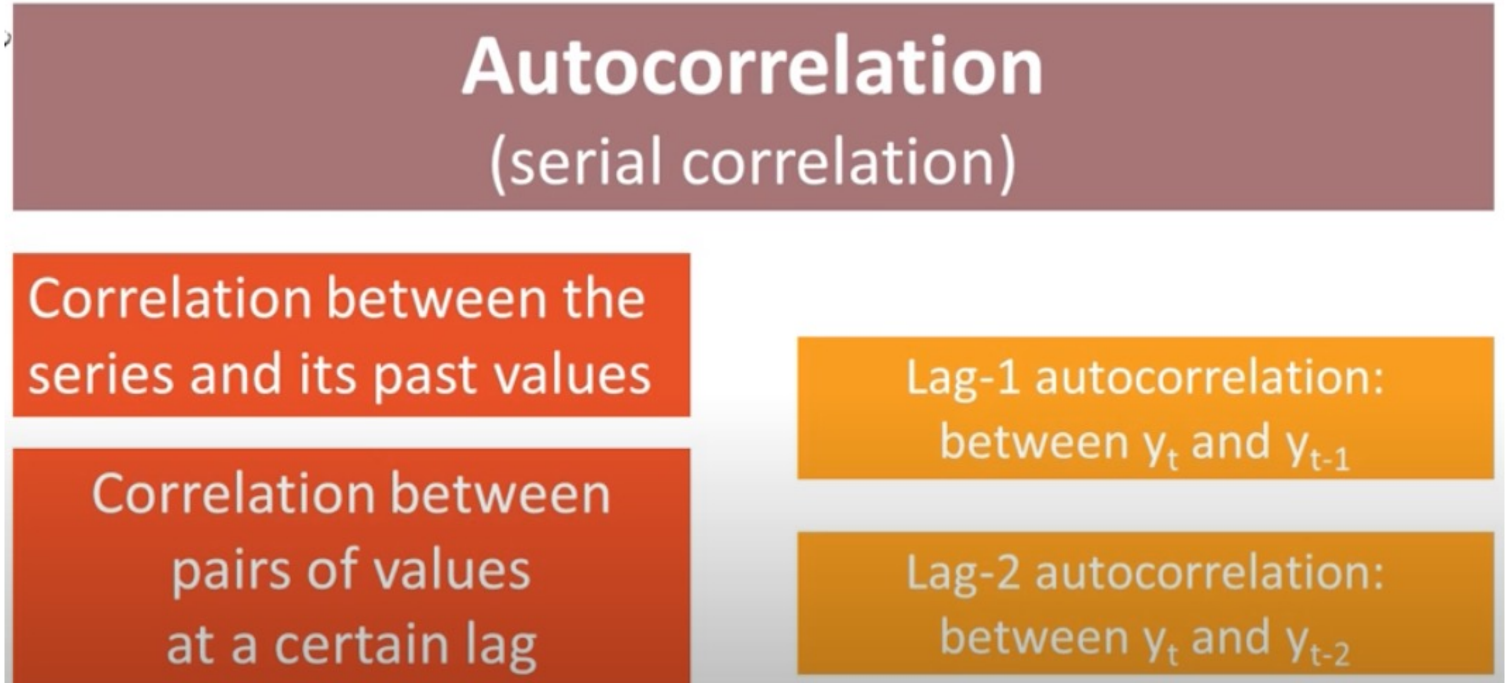
$$m_l' = \mathrm{E}[X^l] = \int_{-\infty}^{\infty} x^l f_X(x)\, dx$$

  - *l-th central moment*

$$m_l = \mathrm{E}[(X-\mu)^l] = \int_{-\infty}^{\infty} (x-\mu)^l f_X(x)\, dx$$

- Examples of low-order moments
  - Expectation: $m_1 = \mu = \mathrm{E}[X]$
  - Variance: $m_2 = \sigma^2 = \mathrm{E}[(X-\mu)^2]$

Autocorrelation function (or correlogram)



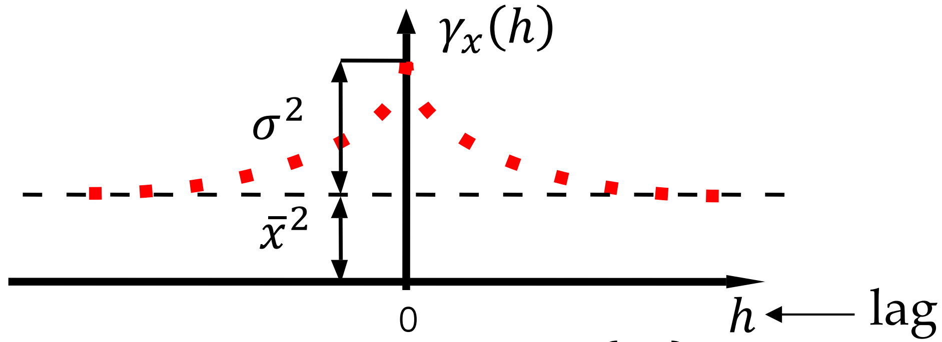
Autocovariance function: stationarity

- Autocovariance function of a stationary time series $\{x_t\}$

$$\gamma_x(h)=\text{Cov}(x_{t+h},x_t)=\text{E}[(x_{t+h}-\mu)(x_t-\mu)] \quad |h|<N$$

with the following 3 properties

1. $\gamma_x(0) \geq 0,$
2. $|\gamma_x(h)| \leq \gamma_x(0)$
3. $\gamma_x(h) = \gamma_x(-h)$



- Autocorrelation function of a stationary time series $\{x_t\}$

$$\rho_x(h) = \frac{\gamma_x(h)}{\gamma_x(0)} \qquad 0 \leq h < N$$

with all the properties of the autocovariance function, except $\rho_x(0) = 1$

- They measure the linear correlation between x_t and x_{t+h}
  - It is of most interest in statistical time series analysis

Sample or empirical statistics

- Given $\{x_1, ..., x_N\}$ observations of a stationary time series $\{x_t\}$, estimate the sample mean, variance and autocovariance
  - Sample mean

$$\hat{\mu} = \bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

  - Sample variance

$$\hat{\sigma}^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \hat{\mu})^2$$

  - Sample autocovariance function

$$\hat{\gamma}_x(h) = \frac{1}{N}\sum_{j=1}^{N-h}\left(x_{j+h} - \bar{x}\right)\left(x_j - \bar{x}\right), \qquad 0 \leq h < N,$$

with $\hat{\gamma}_x(h) = \hat{\gamma}_x(-h), \qquad -N < h \leq 0$

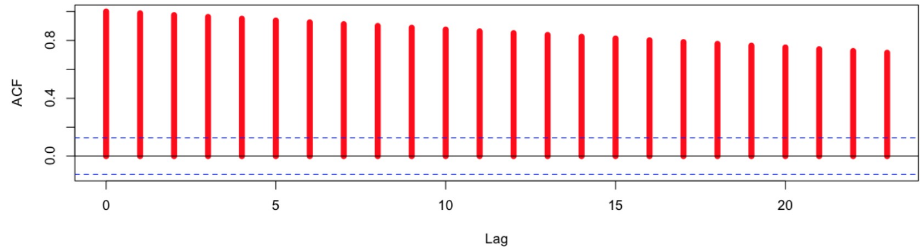
  - Sample autocorrelation function

$$\hat{\rho}_x(h) = \frac{\hat{\gamma}_x(h)}{\hat{\gamma}_x(0)}, \qquad |h| < N$$

Autocorrelation function properties of basic trend and seasonal patterns

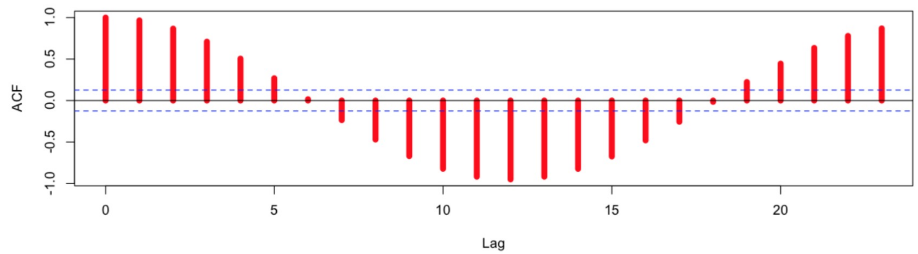
- If the time series $\{x_t\}$, $1 \leq t \leq N$, is a deterministic pure linear trend
  - $x_t = \beta_0 + \beta_1 t$, then for all h

    $$\hat{\rho}_x(h) \xrightarrow[N \to +\infty]{} 1$$



- If the time series $\{x_t\}$, $1 \leq t \leq N$, is a deterministic pure seasonal pattern, as for example
  - $x_t = \cos\left(\frac{2\pi t}{T}\right)$, then for all h

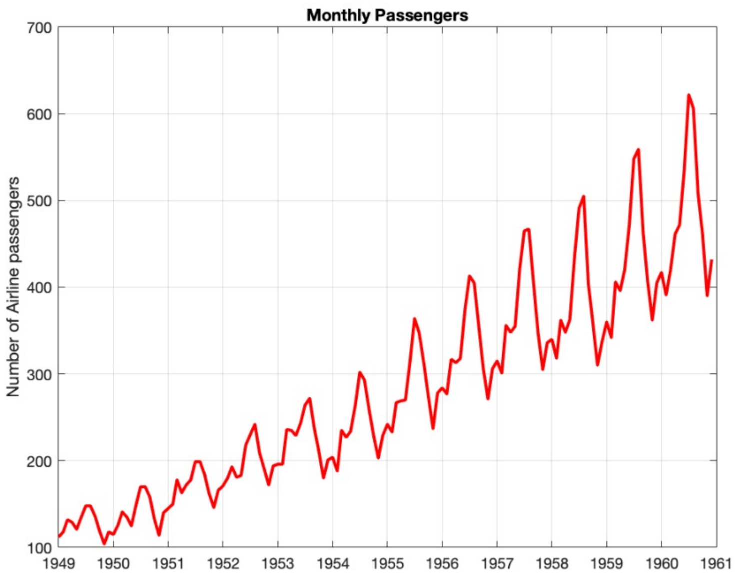
    $$\hat{\rho}_x(h) \xrightarrow[N \to +\infty]{} \cos\left(\frac{2\pi h}{T}\right)$$

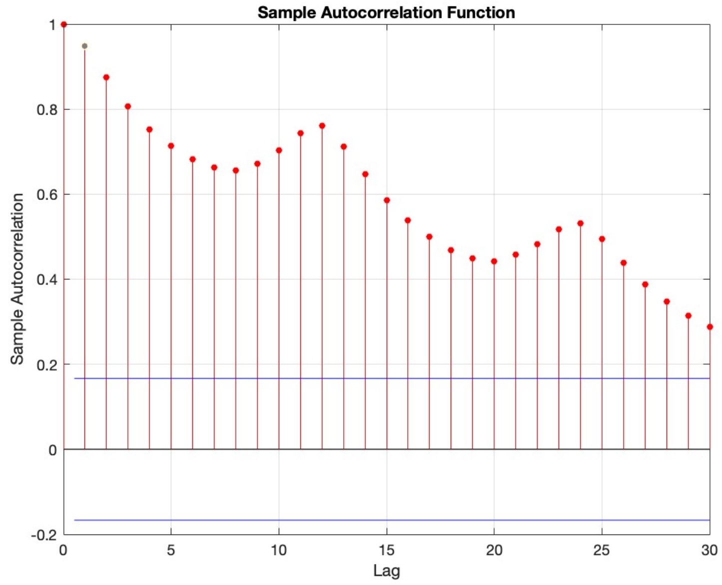


- The presence of basic trend and seasonal patterns is easily observable in the autocorrelation plot
  - It can also help to measure the value of the period of the seasonality

Testing for seasonality in a time series from the ACF plot

- If there is seasonality, the ACF at the seasonal lag will be large and positive
  - Annual seasonality in monthly data can be observed from the ACF where a large value will be seen at lag 12 and possibly also at lags 24, 36, . . .

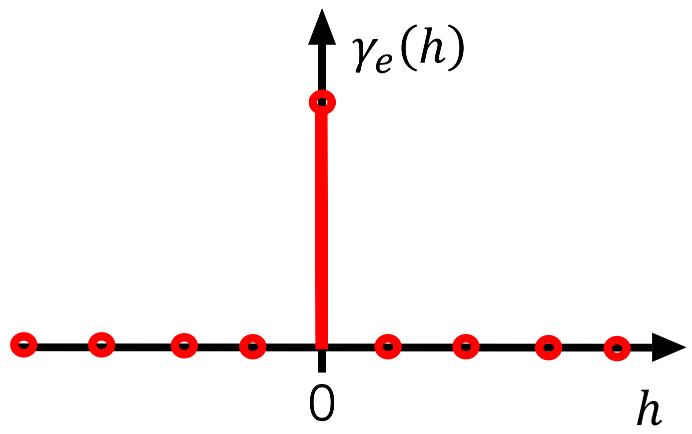




H. Garnier

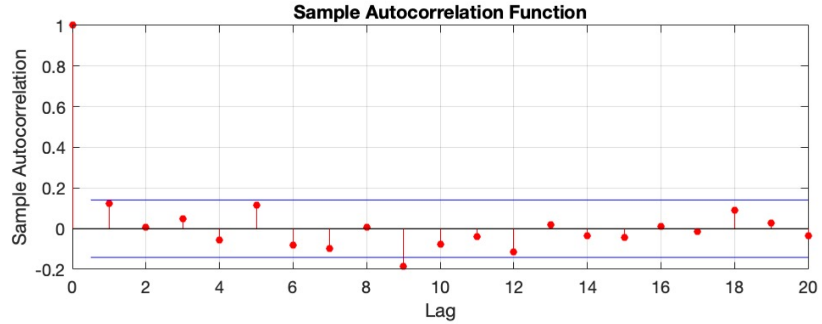
White noise

- The most fundamental example of a (weakly) stationary process is a sequence of independent and identically distributed random variables
  - Such a process may also be referred to as a white noise process
  - Its probability function can be uniform, Gaussian, …
  - They are *uncorrelated*, have mean zero, and common variance
  - Because independence implies that its variables are uncorrelated at different times, its autocovariance function is simply a Kronecker impulse

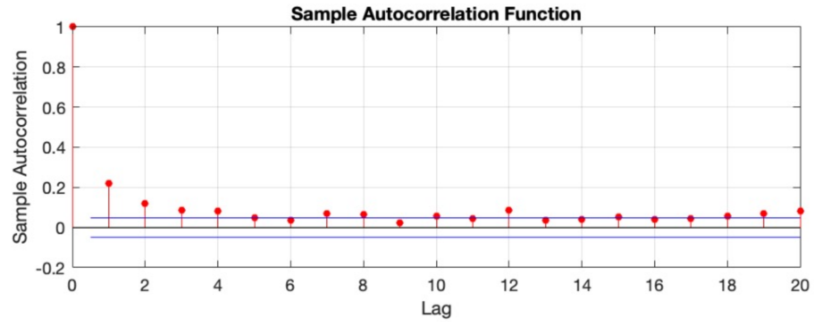


Testing for white noise from ACF plots

- If a time series is white noise, it is unpredictable. The best forecast is zero



- If the series of forecast errors (=residuals) are not white noise, it suggests improvements could be made to the predictive model



  - ACF shows some significant autocorrelation at lags 1, 2, 3,...
  - ACF at lag 12 may indicate some slight seasonality

    These show the time series is not a white noise. There is information left in the residuals that should be used in computing forecasts

- See also the Ljung-Box test for testing white noise

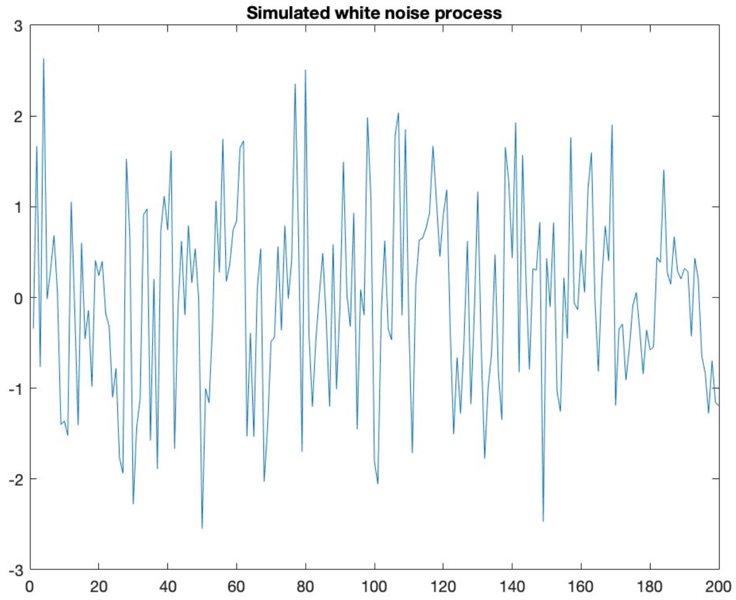
Gaussian white noise

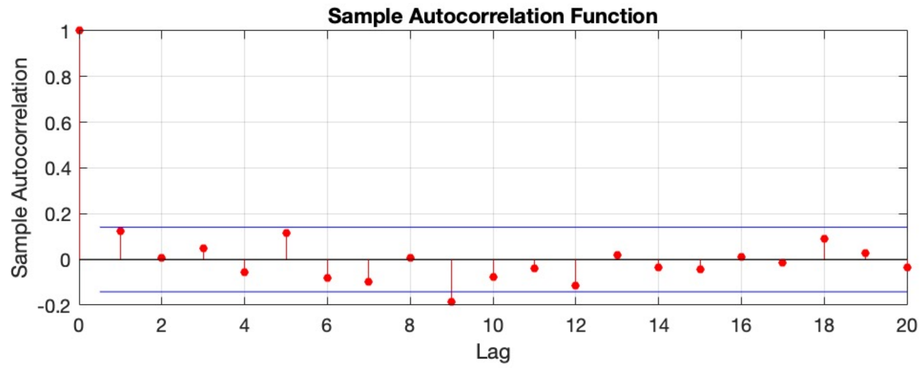
- It is a white noise whose probability function is Gaussian
- In Matlab

>> e=randn(200,1);

>> plot(e)

>> autocorr(e)





Gaussian probability function
A brief review

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$$

m : mean

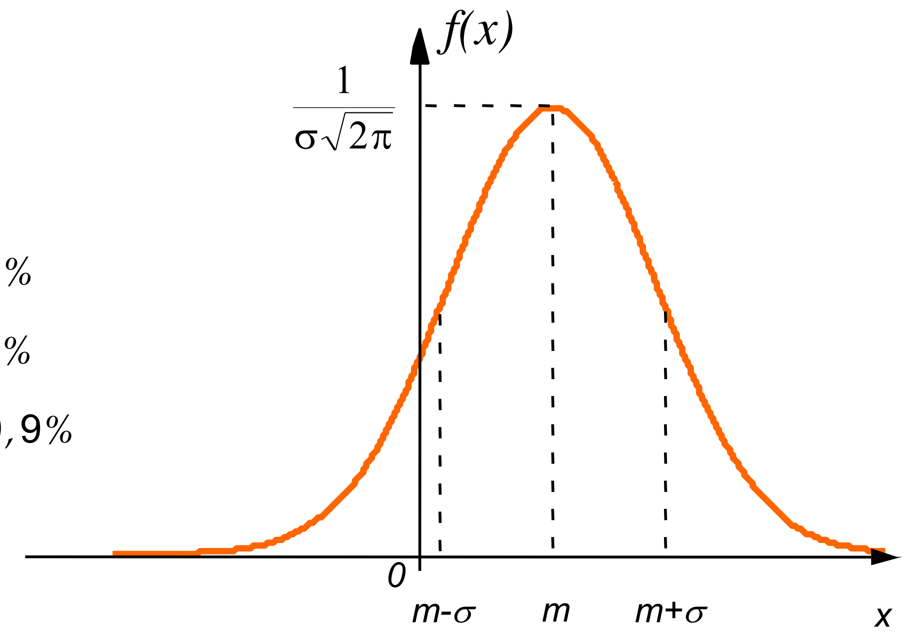
σ : standard-deviation

$$P(m - \sigma \le x \le m + \sigma) = 67\%$$

$$P(m - 2\sigma \le x \le m + 2\sigma) = 95\%$$

$$P(m - 3\sigma \le x \le m + 3\sigma) = 99\%$$

$$P(m - 4\sigma \le x \le m + 4\sigma) = 99,9\%$$

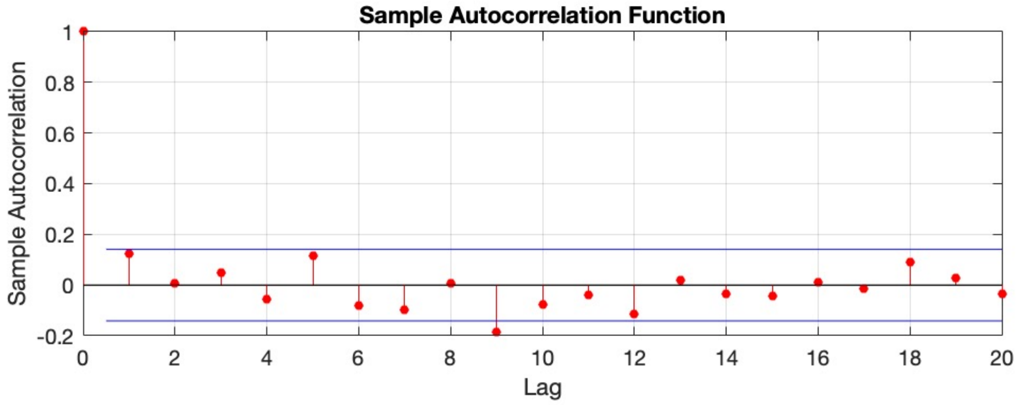


Gaussian probability function
Review

- Important properties
  - Two Gaussian random signals x_k and x_l for $k \neq l$ are uncorrelated (property of white noise) and therefore independent (property of Gaussian probability density)
  - The Gaussian probability density is the only law for which there is equivalence between non-correlation and independence
  - Gaussian laws preserve their Gaussian character in any linear operation: derivation, integration, convolution, filtering

Sampling distribution of autocorrelations
White noise case

- Sampling distribution of $\gamma_e(h)$ for a white noise is asymptotically Gaussian $\mathrm{N}\left(0, \frac{1}{N}\right)$
  - 95% of all $\gamma_e(h)$ must lie within $\pm\frac{1.96}{\sqrt{N}}$
  - It is common to plot limit lines at $\pm\frac{1.96}{\sqrt{N}}$ when plotting the ACF $\gamma_e(h)$
  - If this is not the case, the series is probably not WN
- Example: If N = 125, critical values at $\pm\frac{1.96}{\sqrt{125}} = \pm 0.175$



  - All ACF coefficients lie within these limits, confirming that the data are white noise (*more precisely, the data cannot be distinguished from white noise)*

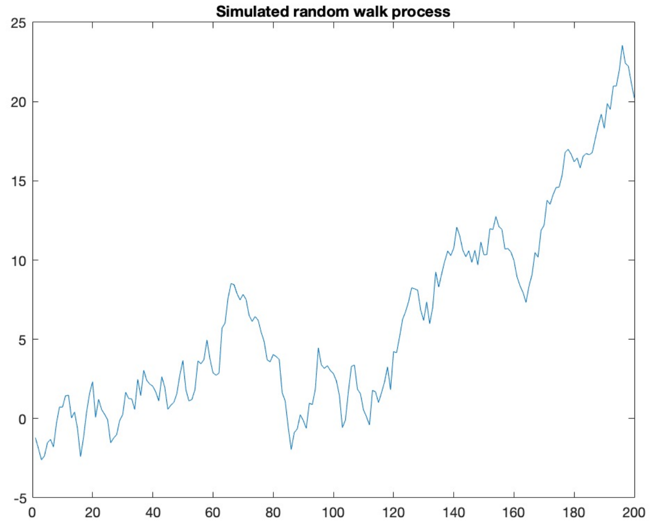
Example of non-stationary time series: The random walk

- A random walk $\{z_t\}$ is the cumulative sum of a white noise $\{x_t\}$ with mean μ and variance σ^2

$$z_t = z_{t-1} + x_t$$

or

$$\text{If } x_0 = 0, \quad z_t = \sum_{j=1}^{t} x_j$$



- We cannot see any trend in the time plot !

- Its mean and variance vary with time
  - Mean: $\mathrm{E}[z_t] = \mu t$
  - Variance: $\mathrm{Var}[z_t] = \sigma^2 t$

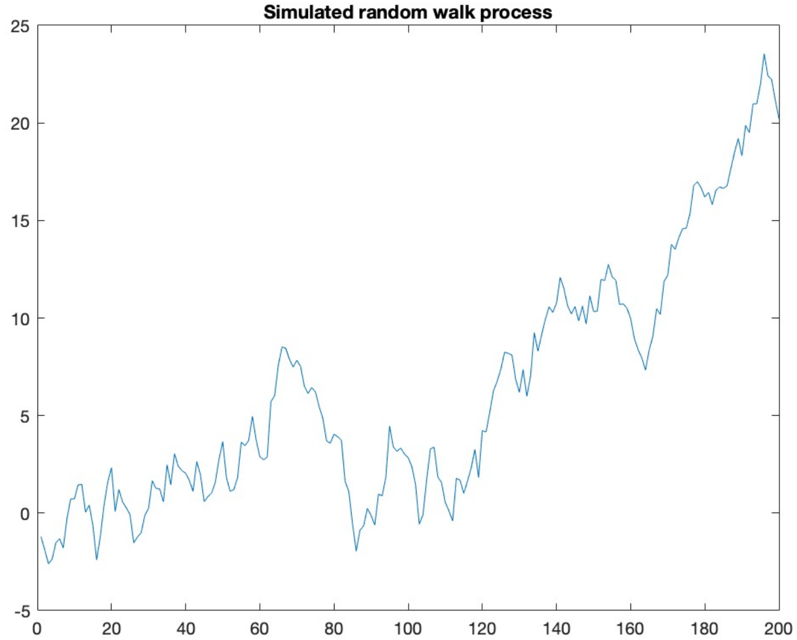
- A random walk is a non-stationary process

H. Garnier

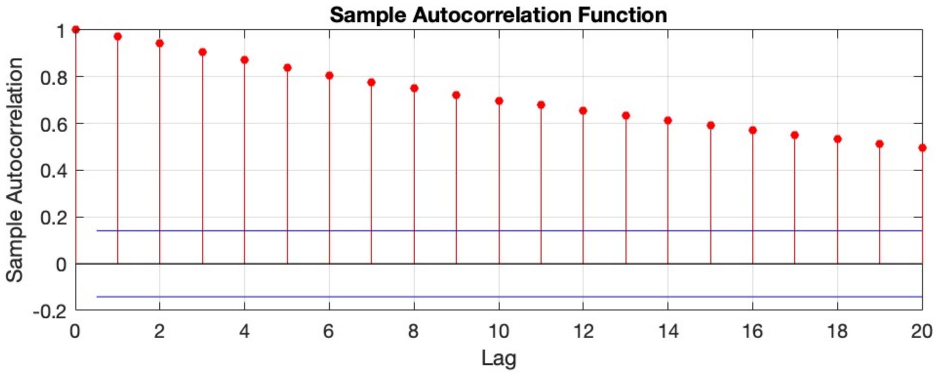
Random walk - *Example*

- In Matlab

```
>> z=cumsum(randn(200,1);
>> plot(z)
```




```
>> autocorr(z)
```



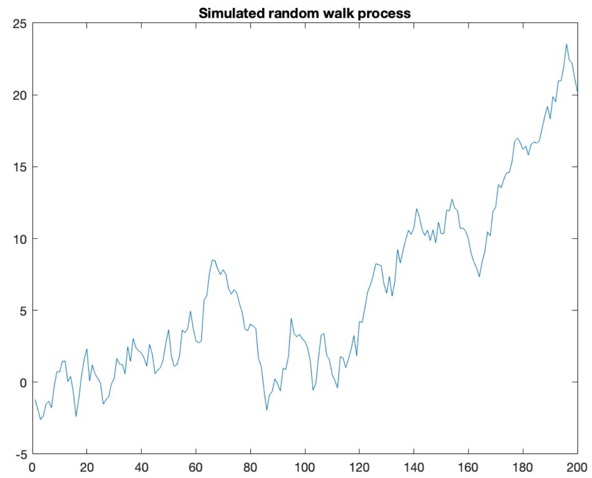
Random walks are often highly correlated

First difference of a random walk

- As seen before, a non-stationary random walk $\{z_t\}$ is defined as

$$z_t = z_{t-1} + x_t$$

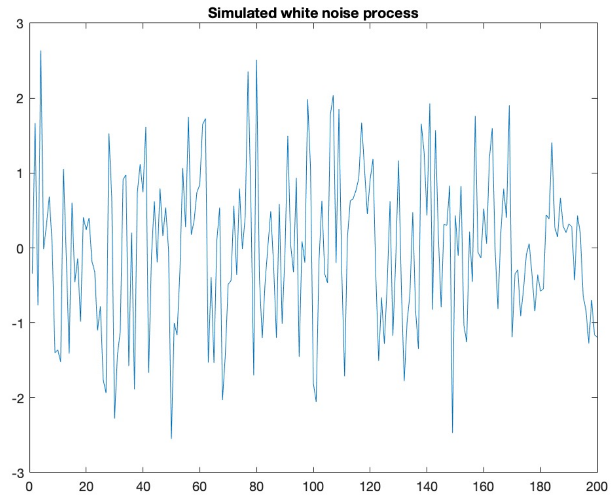
where $\{x_t\}$ is white noise



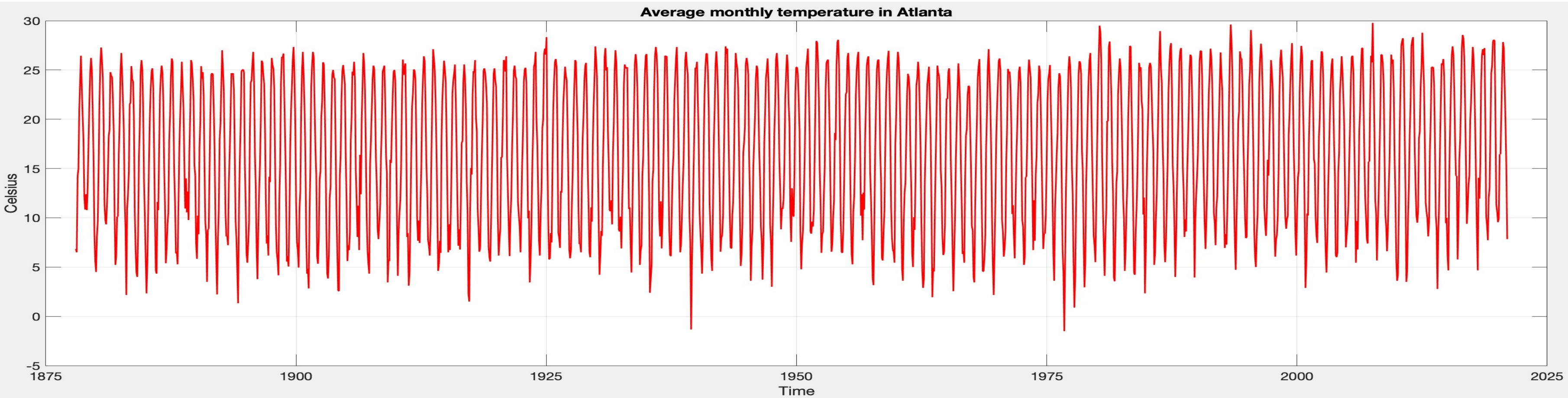
- The first difference of a random walk given by

$$z_t - z_{t-1} = x_t$$

form a purely random process, which is stationary !



Example of a time series exploratory analysis: Average monthly Temperatures in Atlanta, USA



Time series: Average monthly temperature starting from 1878 until 2020

- Available from iWearherNet.com
- The National Weather Service began keeping weather records for Atlanta 142 years ago on October 1, 1878

- Investigate the structure of the dataset. Is an additive model acceptable ?
- Perform a decomposition of the time series
- After removing trend and seasonality, is the residual stationary?

Time series exploratory analysis: general approach

1. Plot the series and check for a trend-cycle pattern, a seasonal component, any apparent sharp changes in behavior (accidental component), any outliers

2. Estimate the trend component and plot the detrended time series

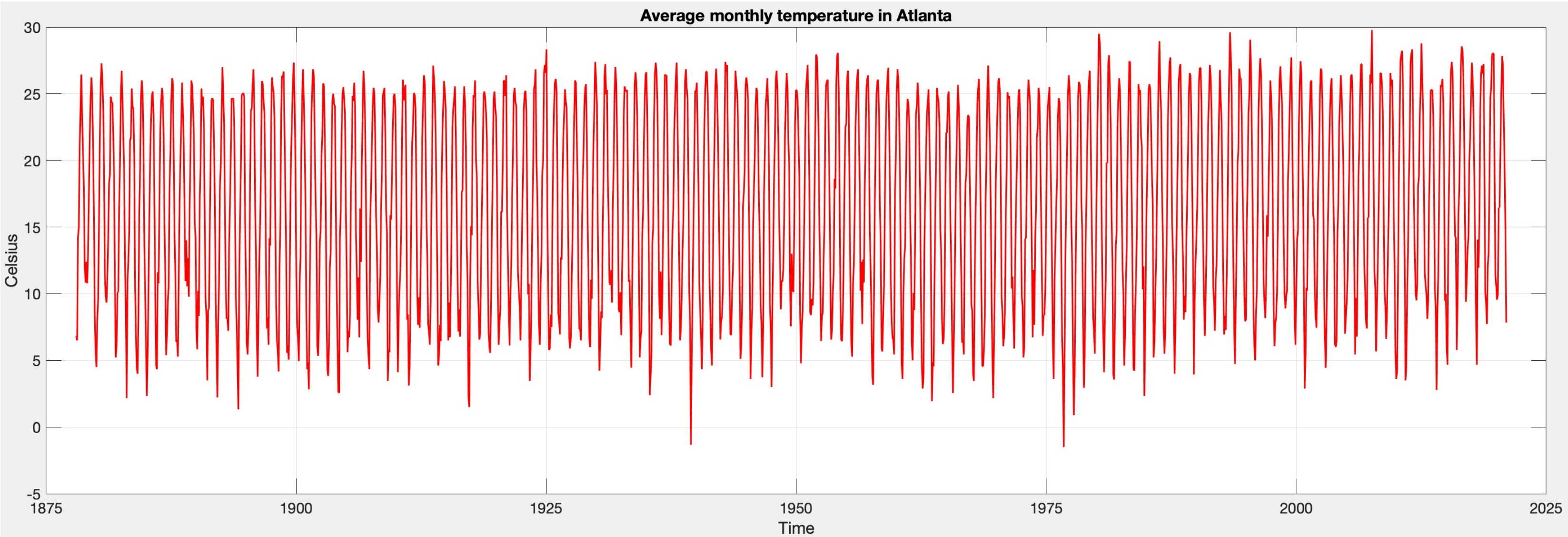
3. Plot the sample autocorrelation function of the detrended time series and check for a seasonal and cyclical pattern

4. Estimate the seasonal component from the detrended time series. Deseasonalize the time series by substracting the seasonality component from the original time series

5. Compute and plot the estimated random component or residuals

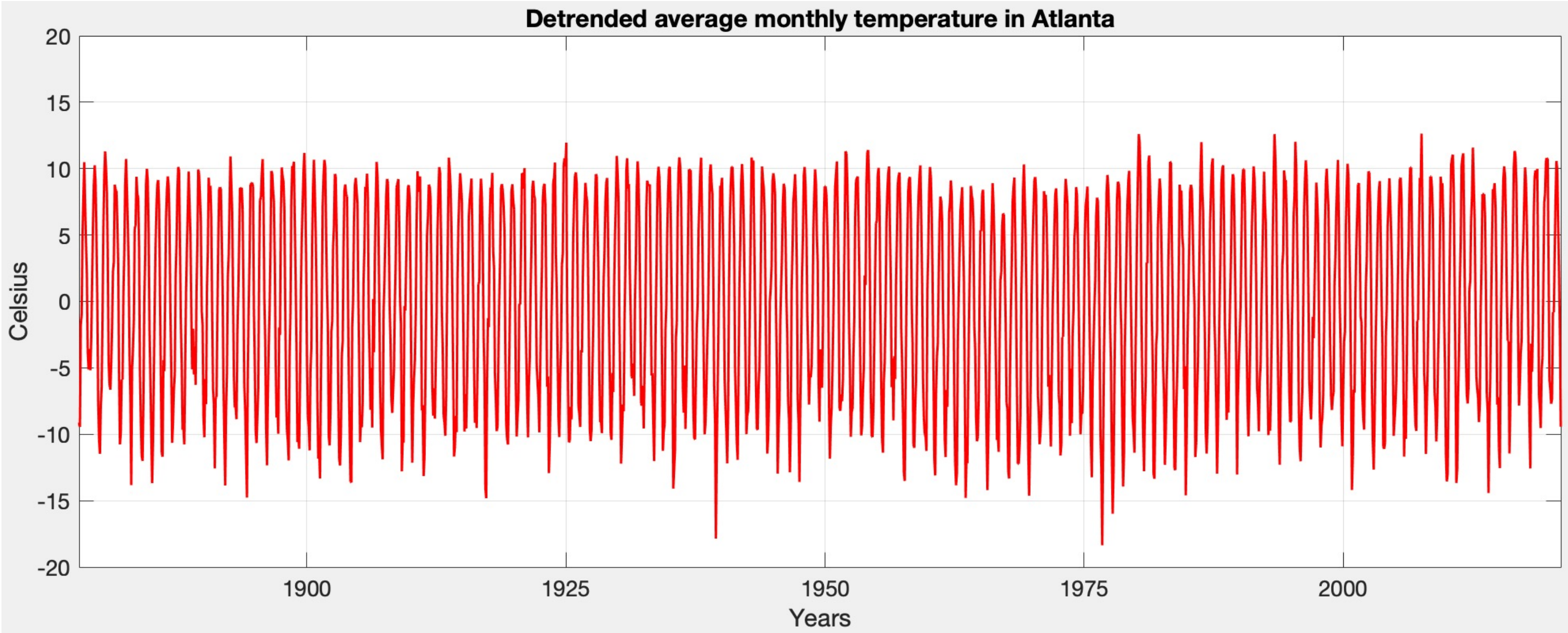
6. Plot the sample autocorrelation function of the residuals and check for stationary residuals close to a white noise

1. Plot of the original time series



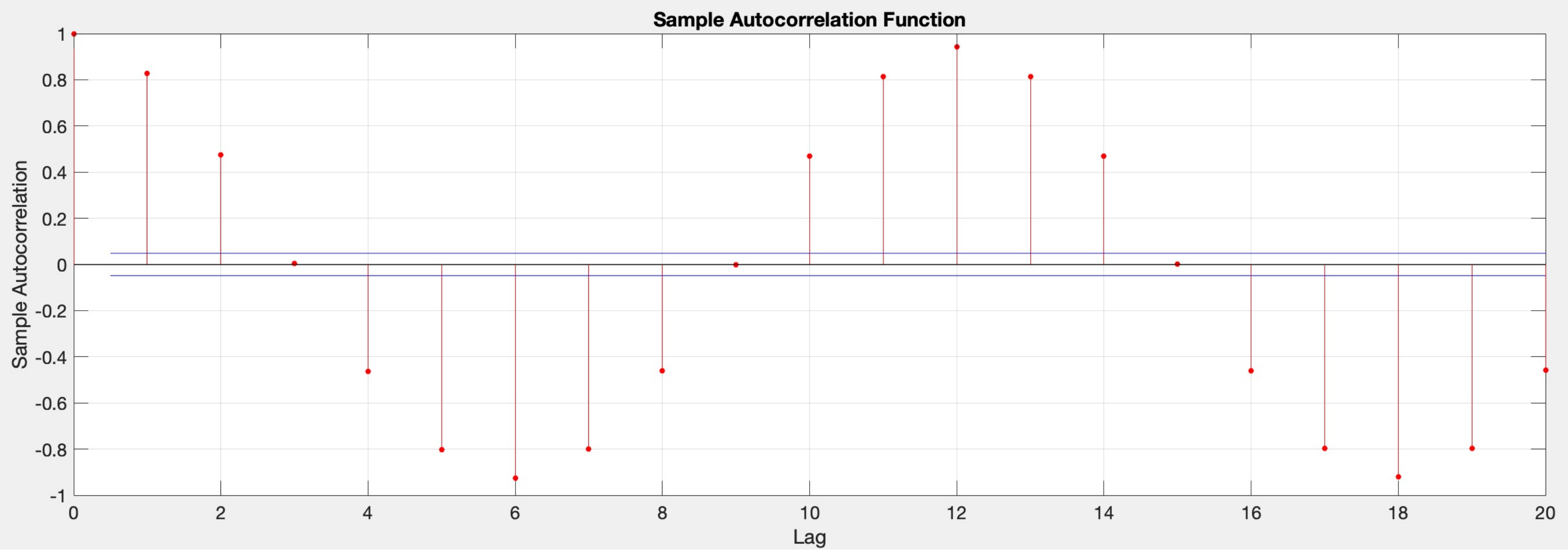
- Although temperature is slowly changing over years, global warming effects can be observed as the maximum temperatures are slightly higher since 1975 than before 1925
  - ⇨ Needs to be modelled by a slight linear increase trend

2. Plot of the detrended time series



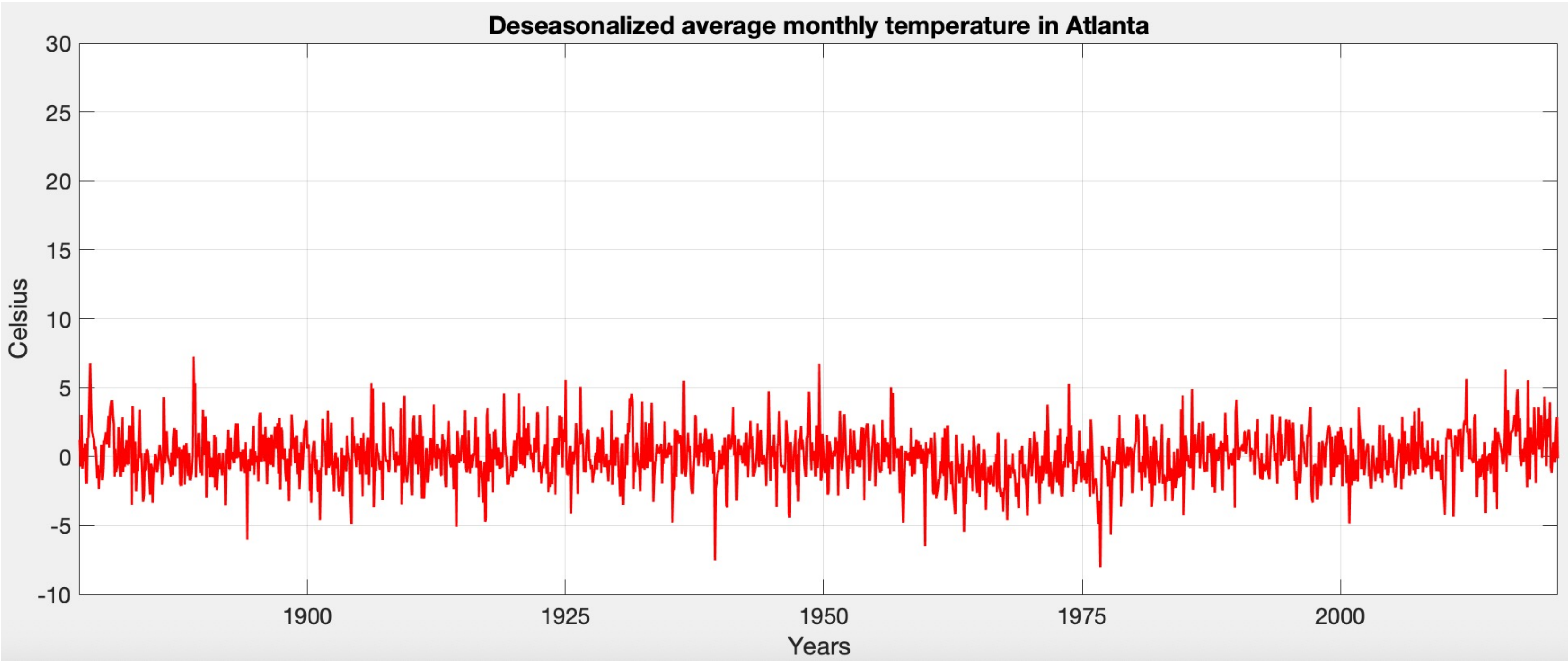
- The slight linear increase trend has been removed from the original time series
- There is still a clear and expected seasonal pattern of 1 year *(12 months) which can be confirmed from the ACF of the detrended time series*

3. Sample autocorrelation function of the detrended time series



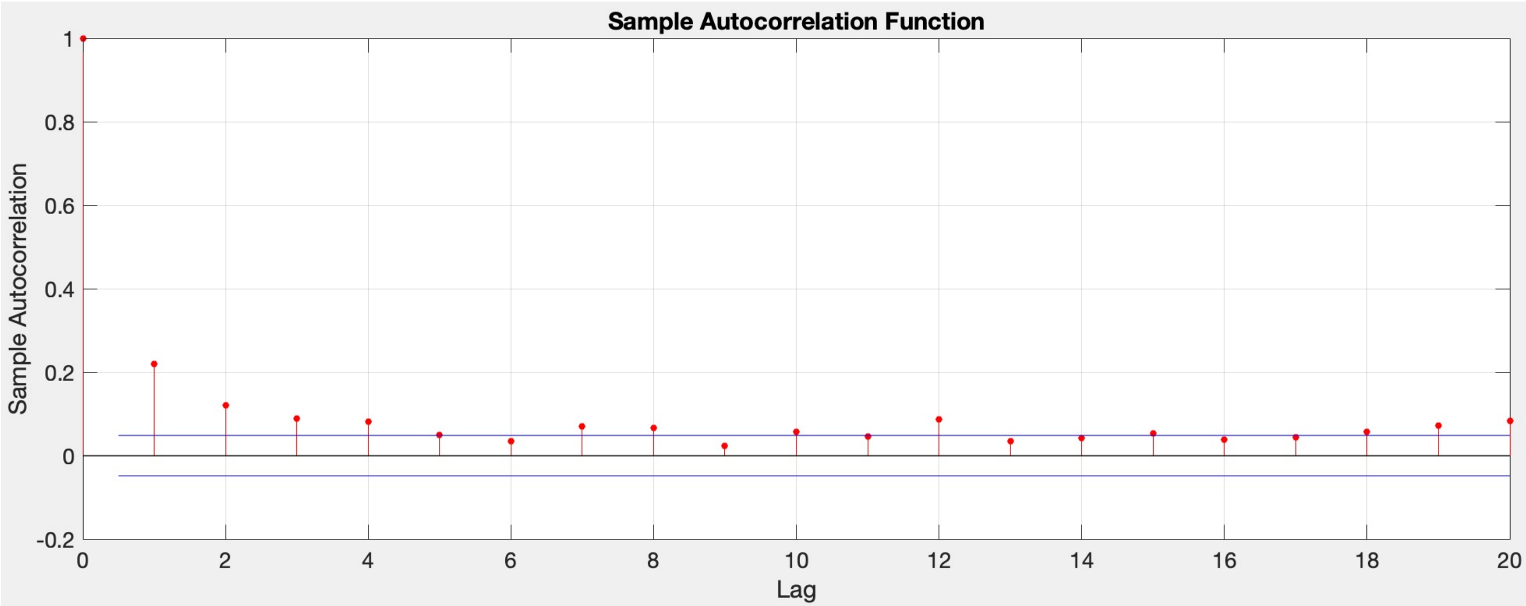
- The seasonal pattern of 12 months is clearly observed from the ACF
  ⇨ Needs to be modelled by a seasonal pattern of 1 year

4. Plot of the detrended and deseasonalized time series



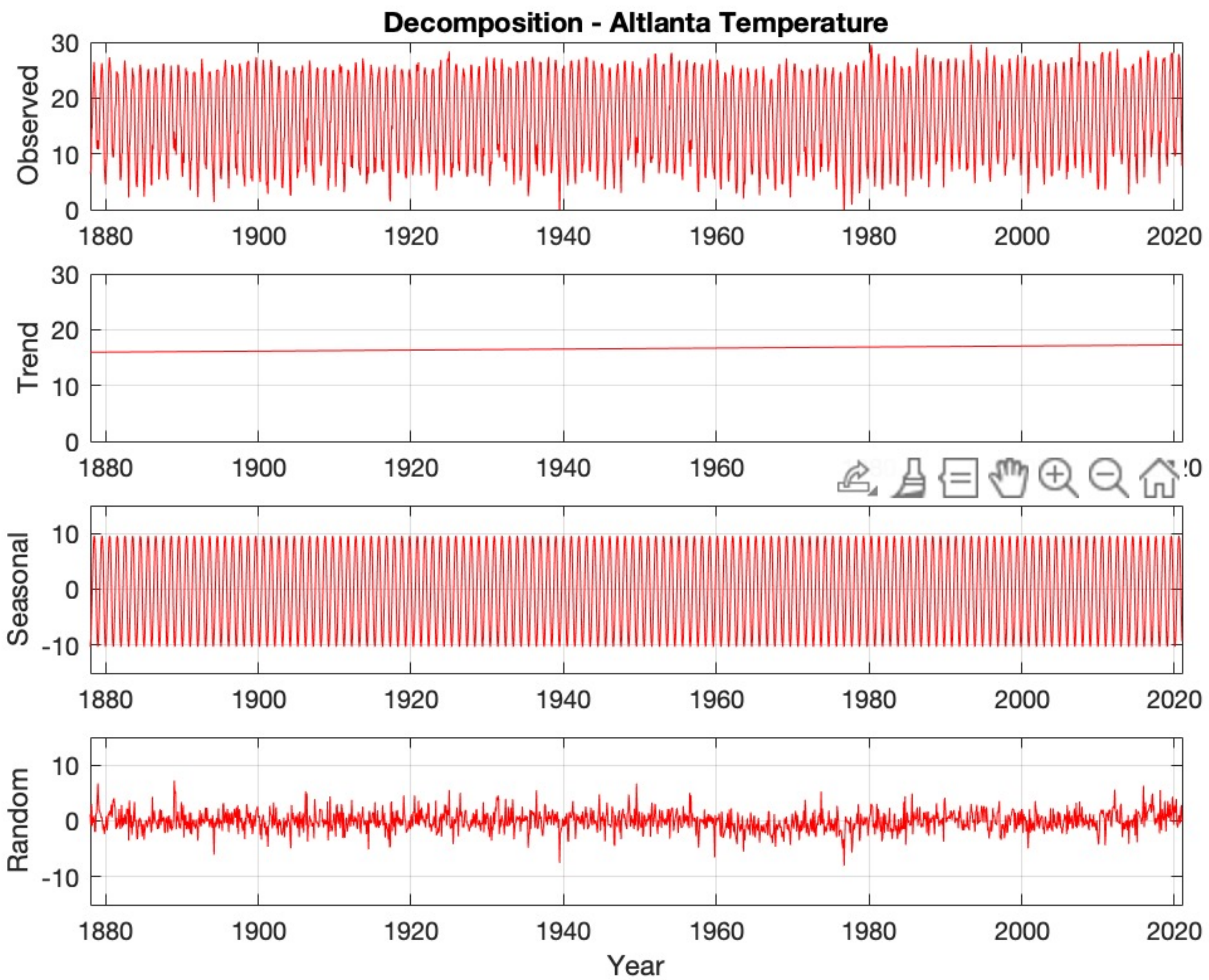
- The estimated trend and seasonal patterns (=estimate of the random component) have now been removed from the original time series
- There is no obvious cyclical pattern

5. Sample autocorrelation functions of the detrended and deseasonalized time series



- There is no more obvious trend, seasonal or cyclical patterns
- The ACF shows that the residuals have some stationarity but they would need further modelling to capture the remaining correlation in the time series
  - ⇨ could be captured by an ARMA model (see next lectures)

Time series decomposition
Monthly average temperature in Atlanta



H. Garnier