Introduction to time series analysis and forecasting

Hugues GARNIER

hugues.garnier@univ-lorraine.fr

UNIVERSITÉ DE LORRAINE

POLYTECH NANCY

Course outline

Introduction to time series analysis and forecasting

I. Main characteristics of time series data

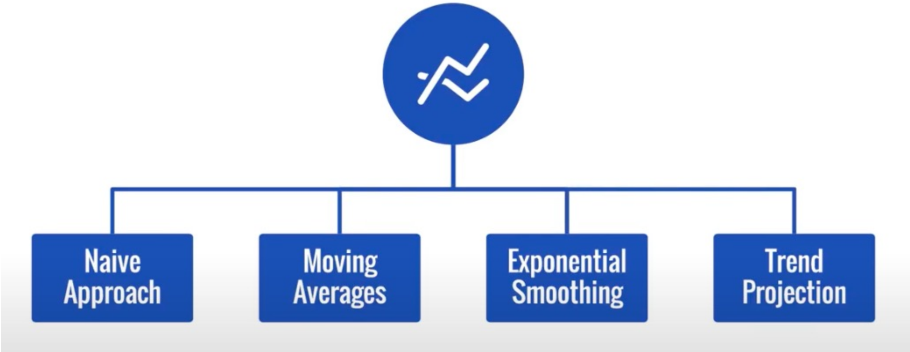
II. Time series decomposition

III. Basic time series modelling and forecasting methods

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

H. Garnier

Basic deterministic model based-forecasting methods



- **Non-parametric model-based approaches**
  - Naïve approach method
  - Moving average smoothing method
  - Exponential smoothing *(Holt-Winter) methods*

- **Parametric model-based approaches**
  - Trend projection or linear regression method

For more information on the exponential smoothing (Holt-Winter's) forecasting methods, see for example

- the free online textbook: otexts.com/fpp2/
- The free online video: bit.ly/2qM9eHL

Forecasting: Principles and Practice
Otexts
Rob J. Hyndman, George Athanasopoulos

UNIVERSITÉ DE LORRAINE

POLYTECH NANCY

Time series modeling: smoothing

From Galit Schmueli, Free online videos supporting her book entitled, Practical time series forecasting with R, 2016

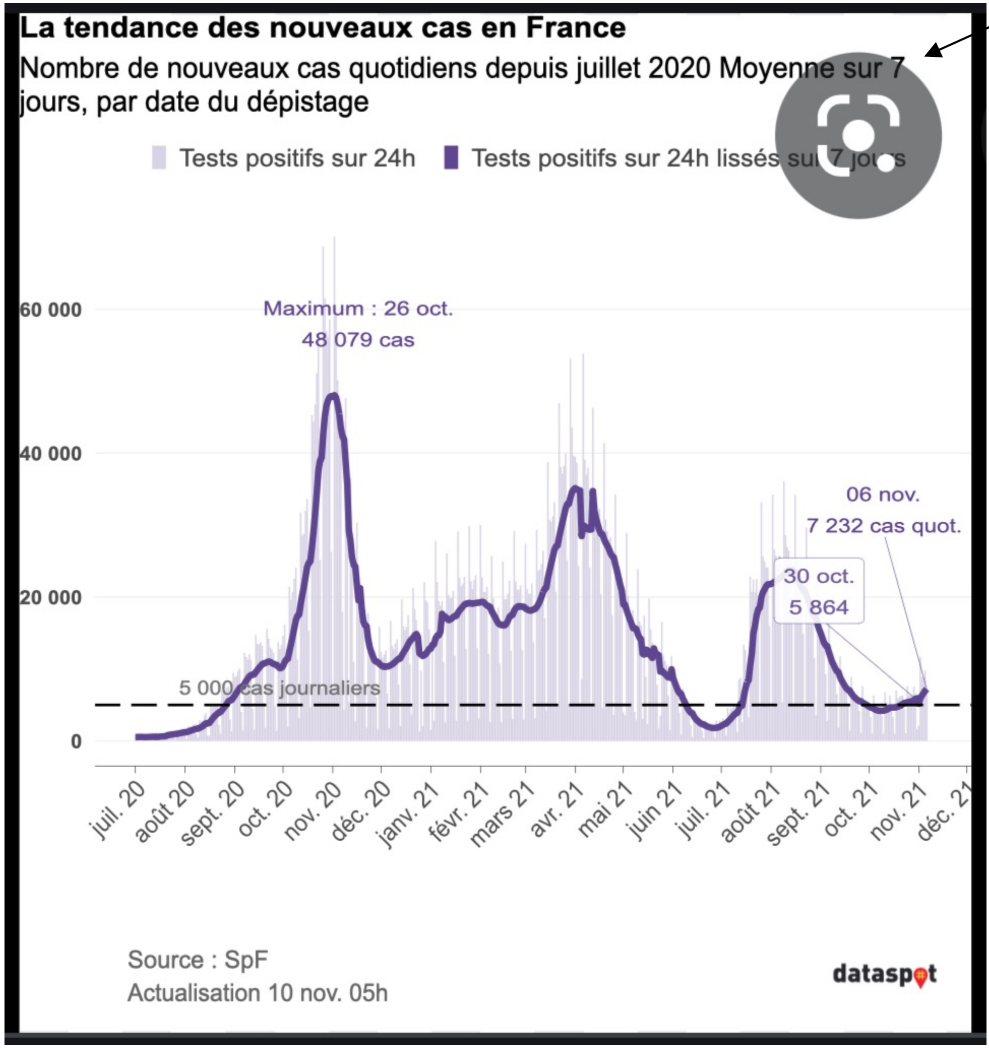
- bit.ly/2qM9eHL

Smoothing for better visualization and forecasting

•y_t

- Smoothing 1: moving average for visualization (4 mn)
  - Centered moving average
  - Trailing moving average
- Smoothing 2: moving average for forecasting (11 mn)

Time series smoothing for better visualization
Example in our daily life



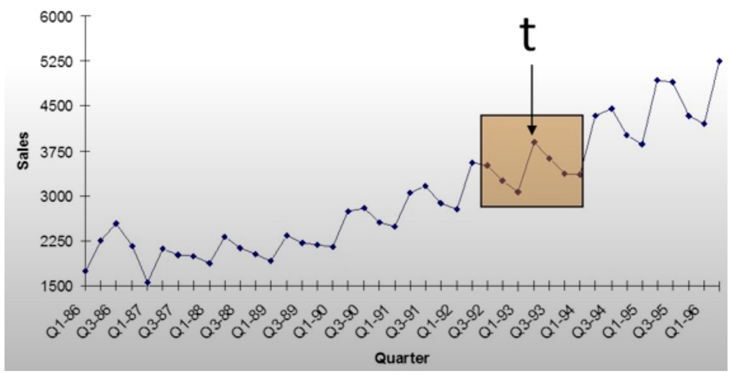
Moving average of width 7 days

Moving average methods

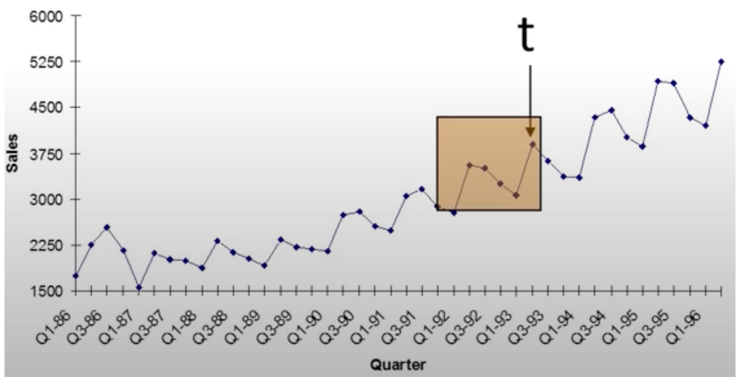
- Idea
  - Smooth or low-pass filter the time series by using an average of several time series data limited to a window of a chosen width
  - This is nothing else than applying a digital low-pass finite impulse response (FIR) filter to the time series
    - *see course on digital signal processing from 4th year IA2R*
- Uses
  - Smoothing for better time series visualization
  - Forecasting
  - Non parametric estimate of the long term trend *(first step of the decomposition procedure)*
- Hyper parameter
  - Width of the window

Moving Average (MA) methods

- Two types of windowing
  - Centered MA: uses a window centered around time t *(the filtering is not causal. This method cannot be used for forecasting)*
  - Trailing MA: based on a window from time t and backwards *(the filtering is causal. This method can easily be used for forecasting)*



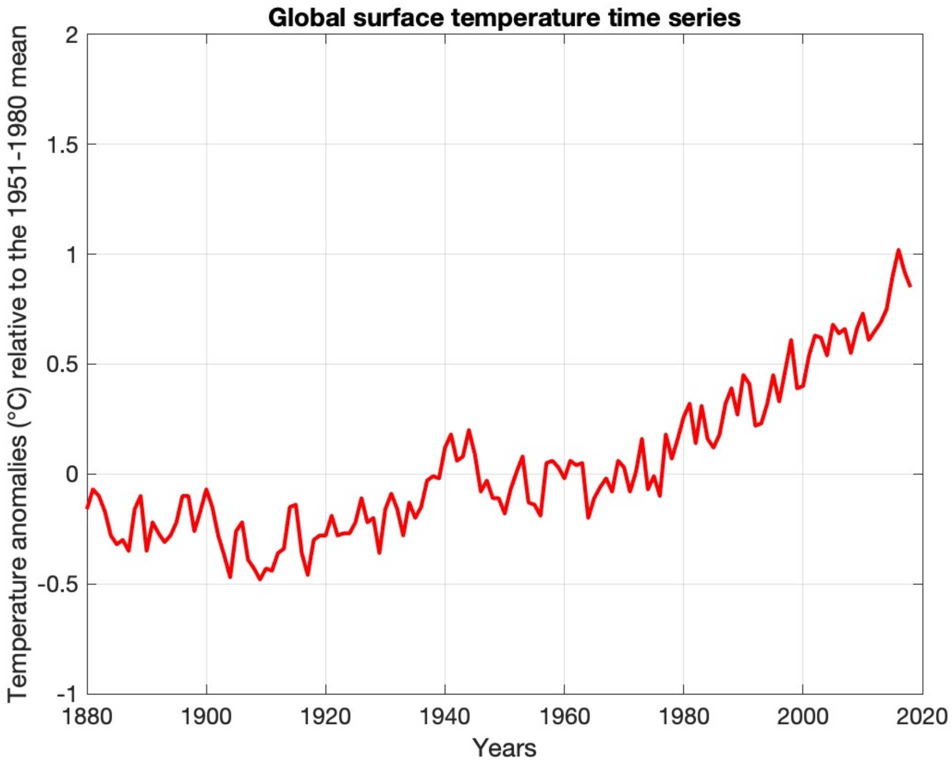
Centered MA



Trailing MA

H. Garnier

Trend model for global surface temperature time series



Global warming seems to be clearly accelerating from 1980 onwards

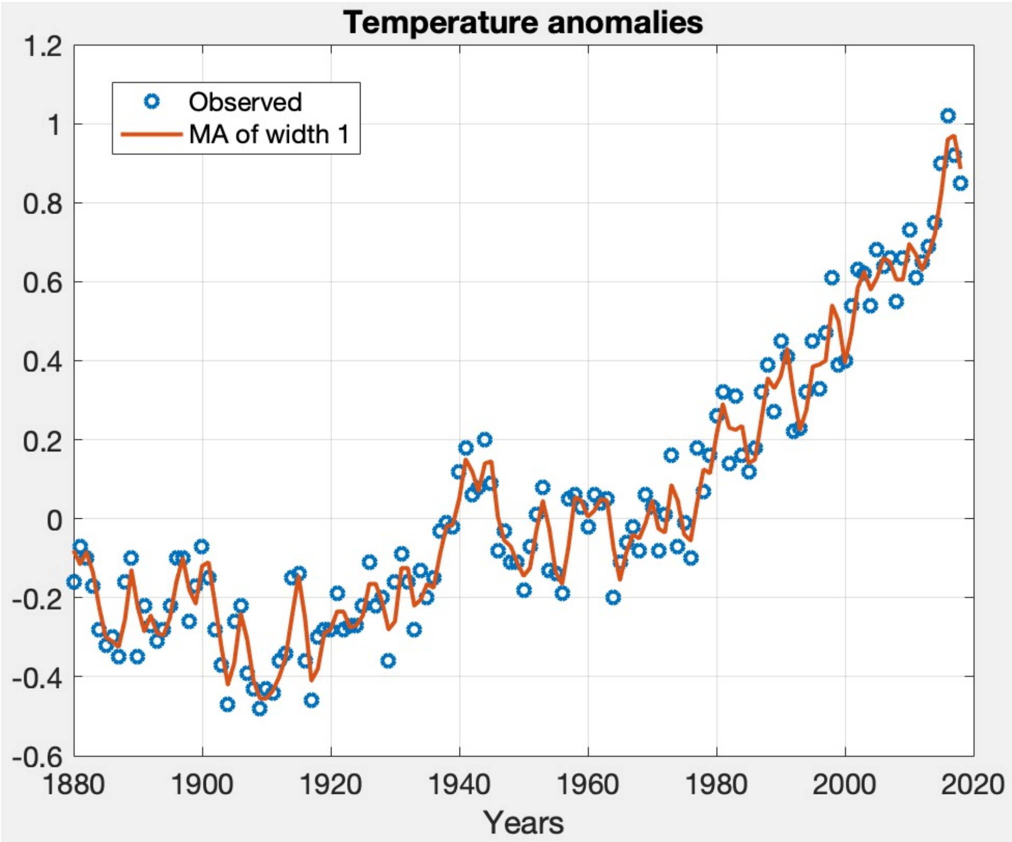
- Global surface temperature time series shown against time is the temperature anomalies (in ° Celsius) relative to the 1951-1980 mean. The series is called GISTEMP after its producer, the NASA, New York, USA
- For more elaborated trend models, see paper by Manfred Mudelsee, *Trend analysis of climate time series: A review of methods*, Earth-Science Reviews 2019

UNIVERSITÉ DE LORRAINE

POLYTECH NANCY

Trailing moving average of width 2 applied to the global surface temperature

The trailing moving average of width 2 is computed by: $w_t = \frac{1}{2} y_t + \frac{1}{2} y_{t-1}$



See routine **movmean** in Matlab

Example

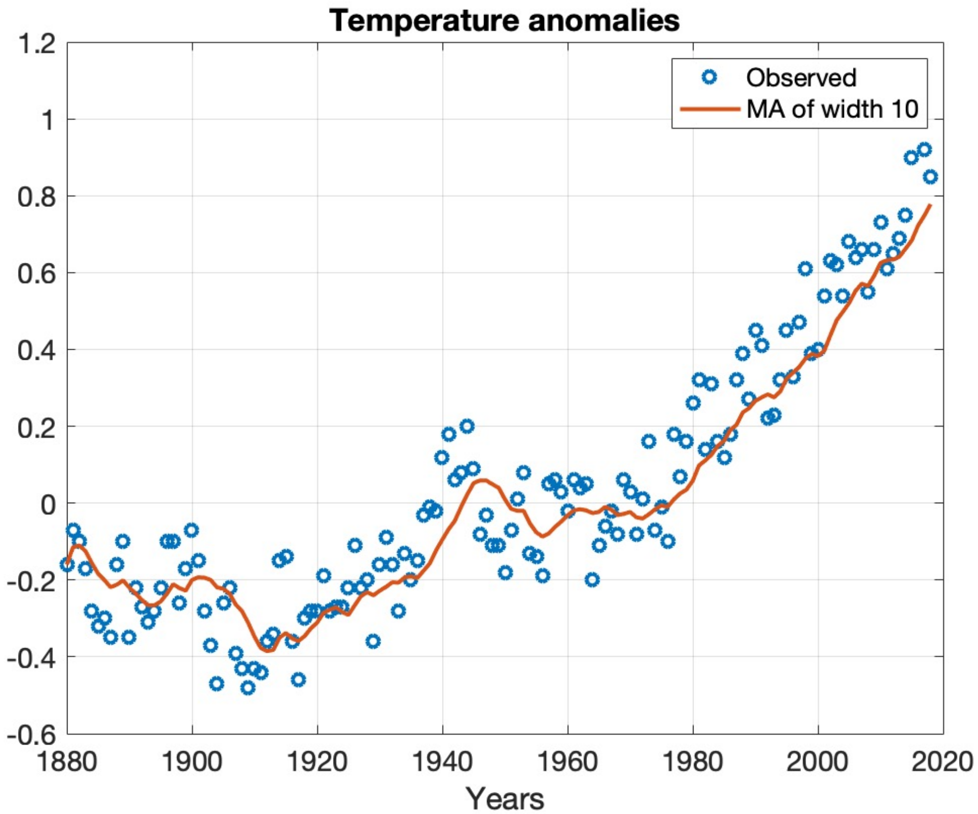
M=2; % *M=width*

w = movmean(y,[M-1 0]);

The long term trend does not appear very clearly

Trailing moving average of width 10 years applied to the global surface temperature

The trailing moving average of width 10 is computed by: $w_t = \frac{1}{10}\sum_{m=0}^{9} y_{t-m}$



The long term trend appears now more clearly

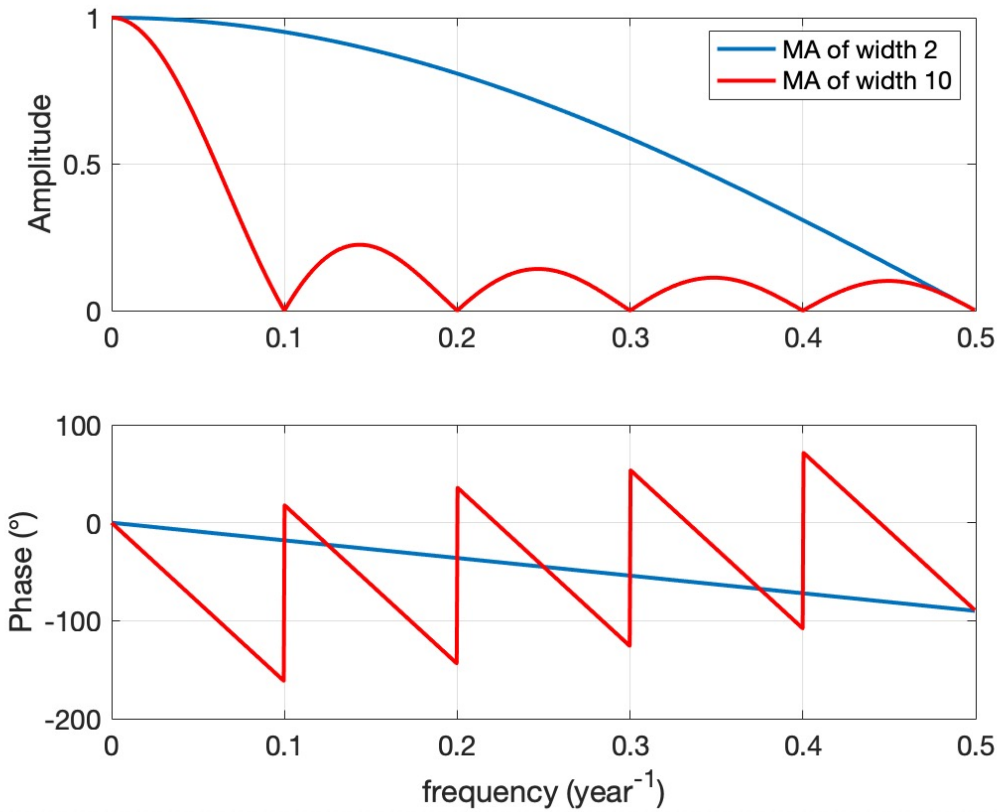
The trailing moving average introduces a time-delay that can be observed in the smoothed time series

Frequency response of the two MA filters

$$H(z) = \frac{1}{10}\sum_{k=0}^{9} z^{-k}$$

Matlab commands:

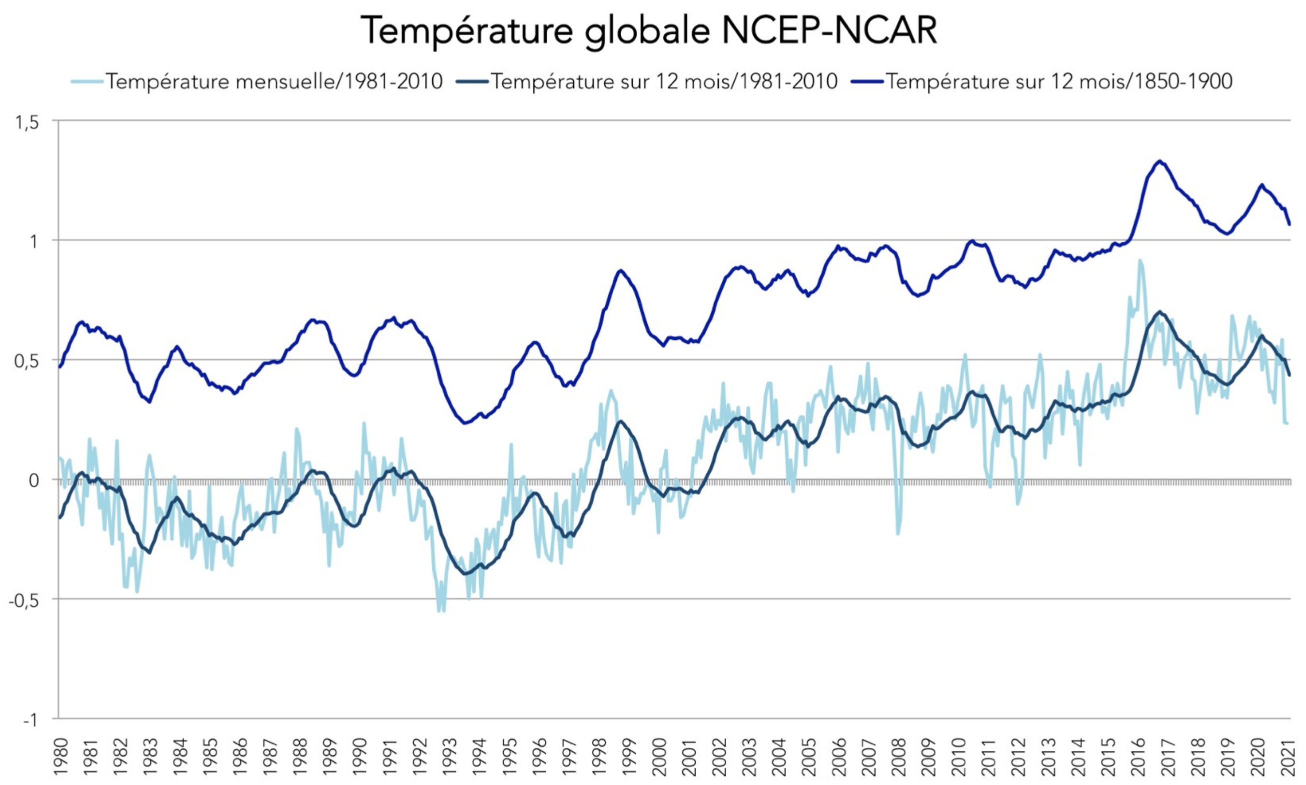
```
N=10;
B=1/N*ones(1,N);
A=1;
fs=1; % Ts=1 year
[H,f]=freqz(B,A,512,fs);
subplot(2,1,1)
plot(f,abs(H)),grid
subplot(2,1,2)
plot(f,angle(H)),grid
xlabel('frequency (year^{-1})')
```



The larger the window width, the lower the low-pass filtering of the time-series but the larger the delay *(proportional to the phase)* introduced

In Matlab, an alternative is to use the routine **filtfilt** so that no delay is introduced

Trailing moving average are commonly used to smooth time series data – An example found on Internet



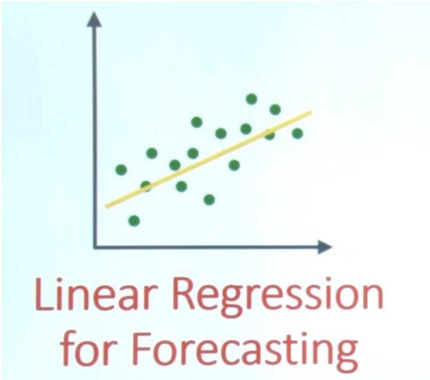
Source: https://global-climat.com/temperature-mondiale-actuelle/

The trailing moving average has been computed with a width of 12 months. The introduced time-delay can be clearly observed in the smoothed time series

Time series modeling: regression models

From Galit Schmueli, Free online videos supporting her book entitled, Practical time series forecasting with R, 2016

– bit.ly/2qM9eHL



Linear models for forecasting

- Regression 1: regression for forecasting (5 mn)
  - Predictors are temporal patterns: t^0, t^1,…, t^k
  - Model estimated from training data
  - Estimated models used to forecast
- Regression 2: linear trend models (7 mn)
- Regression 3: other trend models (8 mn)
- Regression 4: models for capturing seasonality (8 mn)

Linear regression
A brief review

- Prediction of variable y on the basis of information provided by other measured variables $\varphi_1, ..., \varphi_d$.

- Collect $\varphi = \begin{bmatrix} \varphi_1 \\ \vdots \\ \varphi_d \end{bmatrix}$.

- Problem: find function of the regressors $g(\varphi)$ that minimises the difference $y - g(\varphi)$ in some sense.
  So $\hat{y} = g(\varphi)$ should be a good prediction of y.

- Example in a stochastic framework: minimise $E[y - g(\varphi)]^2$.

UNIVERSITÉ DE LORRAINE

POLYTECH NANCY

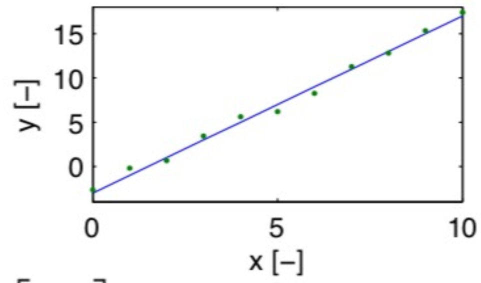
Linear regression
A brief review

- Regression function $g(\varphi)$ is parameterised. It depends on a set of parameters
$$\theta = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_d \end{bmatrix}.$$
- Special case: regression function $g(\varphi)$ is *linear in the parameters* θ. Note that this does **not** imply any linearity with respect to the variables from φ.
- Special case: $g(\varphi) = \theta_1\varphi_1 + \theta_2\varphi_2 + \ldots + \theta_d\varphi_d$
So $g(\varphi) = \varphi^T\theta$.

H. Garnier

Linear regression - Example

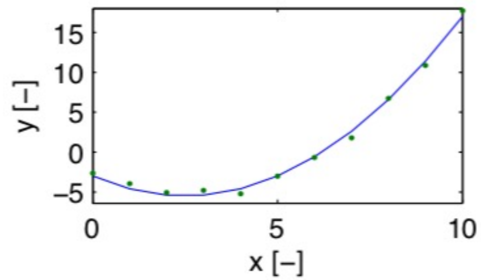
Linear regression — Examples:



- Linear fit $y = ax + b$.

  Then $g(\varphi) = \varphi^T \theta$ with input vector $\varphi = \begin{bmatrix} x \\ 1 \end{bmatrix}$

  and parameter vector $\theta = \begin{bmatrix} a \\ b \end{bmatrix}$. So: $g(\varphi) = \begin{bmatrix} x & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix}$.



- Quadratic function $y = c_2 x^2 + c_1 x + c_0$.

  Then $g(\varphi) = \varphi^T \theta$ with input vector $\varphi = \begin{bmatrix} x^2 \\ x \\ 1 \end{bmatrix}$

  and parameter vector $\theta = \begin{bmatrix} c_2 \\ c_1 \\ c_0 \end{bmatrix}$. So: $g(\varphi) = \begin{bmatrix} x^2 & x & 1 \end{bmatrix} \begin{bmatrix} c_2 \\ c_1 \\ c_0 \end{bmatrix}$.

Least squares estimate (0)

- N measurements $y(t), \varphi(t), \quad t = 1, ..., N.$
- Minimise $V_N(\theta) = \frac{1}{N} \sum_{t=1}^{N} [y(t) - g(\varphi(t))]^2.$
- So a suitable θ is $\widehat{\theta}_N = \arg\min V_N(\theta).$
- Linear case $V_N(\theta) = \frac{1}{N} \sum_{t=1}^{N} [y(t) - \varphi^T(t)\theta]^2.$

Least squares estimate (1)

- In the linear case the "cost" function $V_N(\theta) = \frac{1}{N}\sum_{t=1}^{N}[y(t) - \varphi^T(t)\theta]^2$ is a quadratic function of θ.
- It can be minimised analytically: All partial derivatives $\frac{\partial V_N(\theta)}{\partial \theta}$ have to be zero in the minimum:
  $$\frac{1}{N}\sum_{t=1}^{N} 2\varphi(t)[y(t) - \varphi^T(t)\theta] = 0$$
  The solution of this set of equations is the parameter estimate $\widehat{\theta}_N$.

Least squares estimate (2)

- A *global* minimum is found for $\widehat{\theta}_N$ that satisfies a set of linear equations, the *normal equations*

$$\left[\frac{1}{N}\sum_{t=1}^{N}\varphi(t)\varphi^T(t)\right]\widehat{\theta}_N = \frac{1}{N}\sum_{t=1}^{N}\varphi(t)y(t).$$

- If the matrix on the left is invertible, the LSE is

$$\widehat{\theta}_N = \left[\frac{1}{N}\sum_{t=1}^{N}\varphi(t)\varphi^T(t)\right]^{-1}\frac{1}{N}\sum_{t=1}^{N}\varphi(t)y(t).$$

Least squares estimate – Matrix formulation

- Collect the output measurements in the vector $Y_N = \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix}$,

  and the inputs in the $N \times d$ regression matrix $\Phi_N = \begin{bmatrix} \varphi^T(1) \\ \vdots \\ \varphi^T(N) \end{bmatrix}$.

- Normal equations: $\left[\Phi_N^T \Phi_N\right] \hat{\theta}_N = \Phi_N^T Y_N$.

- Estimate $\boxed{\hat{\theta}_N = \Phi_N^\dagger Y_N}$

  (Moore-Penrose) *pseudoinverse* of Φ_N: $\Phi_N^\dagger = \left[\Phi_N^T \Phi_N\right]^{-1} \Phi_N^T$.

  Note: $\Phi_N^\dagger \Phi_N = I$.

Least squares estimate in Matlab

Solution x of overdetermined $Ax = b$ with rectangular matrix A,
so more equations than unknowns, or
more rows than columns, or
A is m-by-n with $m > n$ and full rank n

Then least squares solution $\hat{x} = A^{\dagger}b$

In Matlab:

```
x = A\b;                % Preferred
x = pinv(A)*b;
x = inv(A'*A)*A'*b;
```

Example 1: the "well-known" linear fit $y = ax + b$

Measurements x_i and y_i for $i = 1, ..., N$.

Cost function $V_N = \frac{1}{N} \sum (y_i - ax_i - b)^2$.

1) "Manual" solution: $\frac{\partial V_N}{\partial a} = 0$ and $\frac{\partial V_N}{\partial b} = 0$, so

$$\begin{cases} \sum -x_i(y_i - ax_i - b) & = & 0 \\ \sum -(y_i - ax_i - b) & = & 0 \end{cases} \Leftrightarrow \begin{bmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & \sum 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} \sum x_i y_i \\ \sum y_i \end{bmatrix}$$

Parameter estimate: $\begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} \sum x_i^2 & \sum x_i \\ \sum x_i & \sum 1 \end{bmatrix}^{-1} \begin{bmatrix} \sum x_i y_i \\ \sum y_i \end{bmatrix}$

Example 1: The "well-known" linear fit $y = ax + b$
Matrix solution *to be preferred*

Measurements x_i and y_i for $i = 1, ..., N$.

Cost function $V_N = \frac{1}{N}\sum(y_i - ax_i - b)^2$.

2) Matrix solution: $Y_N = \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix}, \Phi_N = \begin{bmatrix} x(1) & 1 \\ \vdots & \vdots \\ x(N) & 1 \end{bmatrix}$ and $\theta = \begin{bmatrix} a \\ b \end{bmatrix}$.

Cost function (in vector form) $V_N = \frac{1}{N}||Y_N - \Phi_N\theta||_2^2$.

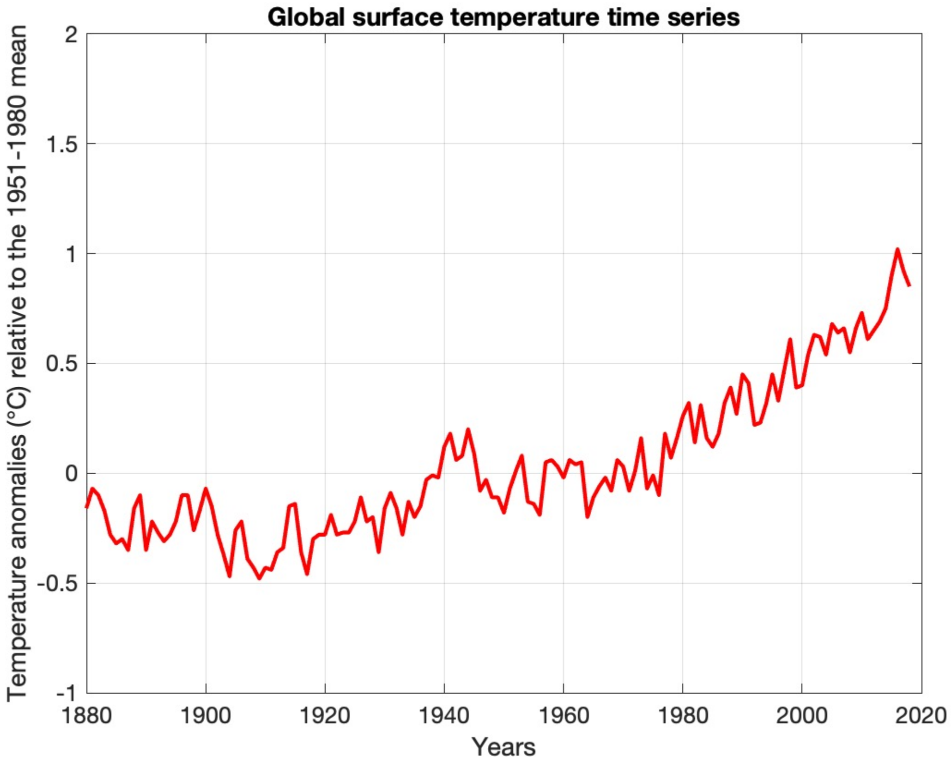
Estimate $\hat{\theta}_N = \Phi_N^{\dagger} Y_N = \left[\Phi_N^T \Phi_N\right]^{-1} \Phi_N^T Y_N$.

In Matlab: `theta = Phi\Y;`

UNIVERSITÉ DE LORRAINE

POLYTECH NANCY

Trend model for global surface temperature time series



- Global surface temperature time series shown against time is the temperature anomalies (in ° Celsius) relative to the 1951-1980 mean. The series is called GISTEMP after its producer, the NASA, New York, USA
- For more elaborated trend models, see paper by Manfred Mudelsee, *Trend analysis of climate time series: A review of methods*, Earth-Science Reviews 2019

Modelling trend components

Review from the previous lecture

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

- 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

Estimation of a linear trend model by LS applied to the global surface temperature

$$V\left(\theta,Z^N\right)=\frac{1}{N}\sum_{k=1}^{N}\left(y(t_k)-(\alpha\times t_k+\beta)\right)^2 \qquad \theta=\begin{bmatrix}\alpha\\ \beta\end{bmatrix}$$

- At the minimum of the criterion, all partial derivatives have to be zero:

$$\frac{\partial V(\theta,Z^N)}{\partial\alpha}=\frac{2}{N}\sum_{k=1}^{N}-t_k\left(y(t_k)-(\alpha\times t_k+\beta)\right)=0$$

$$\frac{\partial V(\theta,Z^N)}{\partial\beta}=\frac{2}{N}\sum_{k=1}^{N}-\left(y(t_k)-(\alpha\times t_k+\beta)\right)=0$$

$$\begin{bmatrix}\sum_{k=1}^{N}t_k^2 & \sum_{k=1}^{N}t_k\\ \sum_{k=1}^{N}t_k & \sum_{k=1}^{N}N\end{bmatrix}\begin{bmatrix}\alpha\\ \beta\end{bmatrix}=\begin{bmatrix}\sum_{k=1}^{N}t_k y(t_k)\\ \sum_{k=1}^{N}y(t_k)\end{bmatrix}$$

- The least squares estimates are given by :

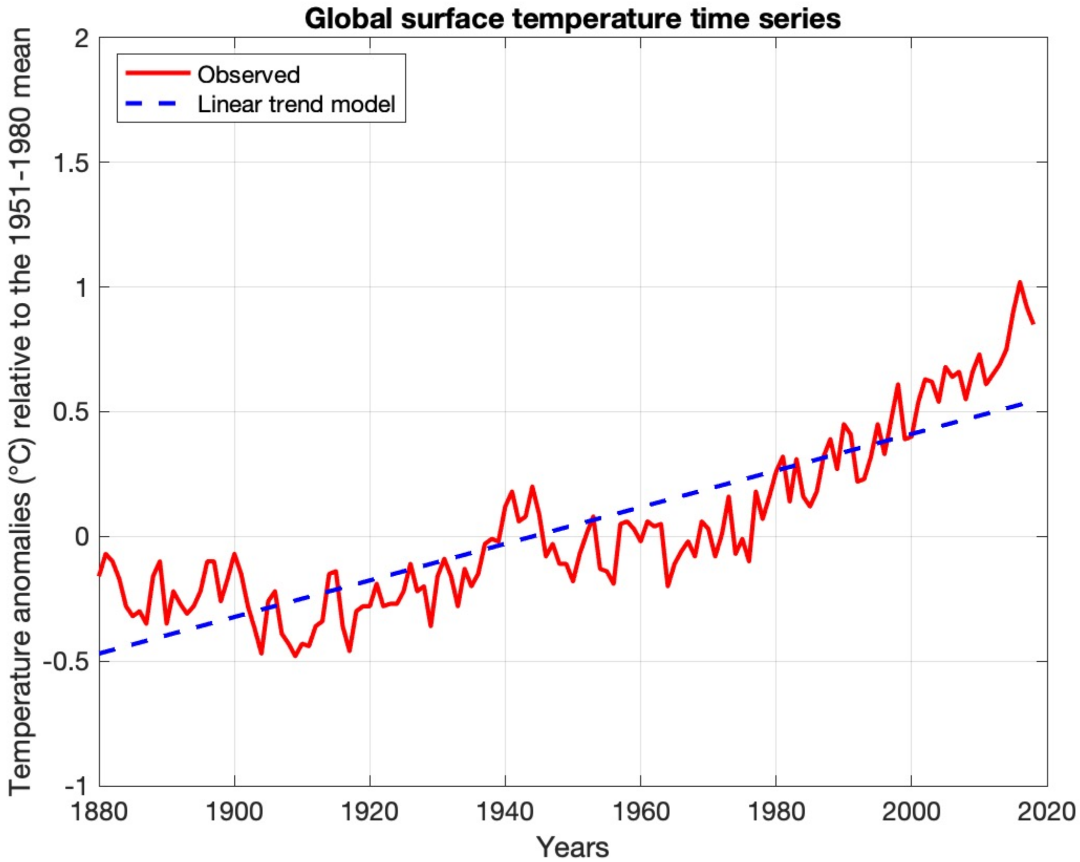
$$\begin{bmatrix}\hat{\alpha}\\ \hat{\beta}\end{bmatrix}=\begin{bmatrix}\sum_{k=1}^{N}t_k^2 & \sum_{k=1}^{N}t_k\\ \sum_{k=1}^{N}t_k & N\end{bmatrix}^{-1}\begin{bmatrix}\sum_{k=1}^{N}t_k y(t_k)\\ \sum_{k=1}^{N}y(t_k)\end{bmatrix}$$

Sous Matlab

```
theta_hat = inv([sum(years.^2) sum(years);
            sum(years) N])*...
            [sum(years.*T);sum(T)]
T_hat = theta(1)* years + theta(2);
plot(years,T,'o',years,T_hat)
```

Estimation of a linear trend model by LS applied to the global surface temperature

$$V\left(\begin{bmatrix} \alpha \\ \beta \end{bmatrix}, Z^N\right) = \frac{1}{N}\sum_{k=1}^{N}\left(y(t_k) - (\alpha \times t_k + \beta)\right)^2$$

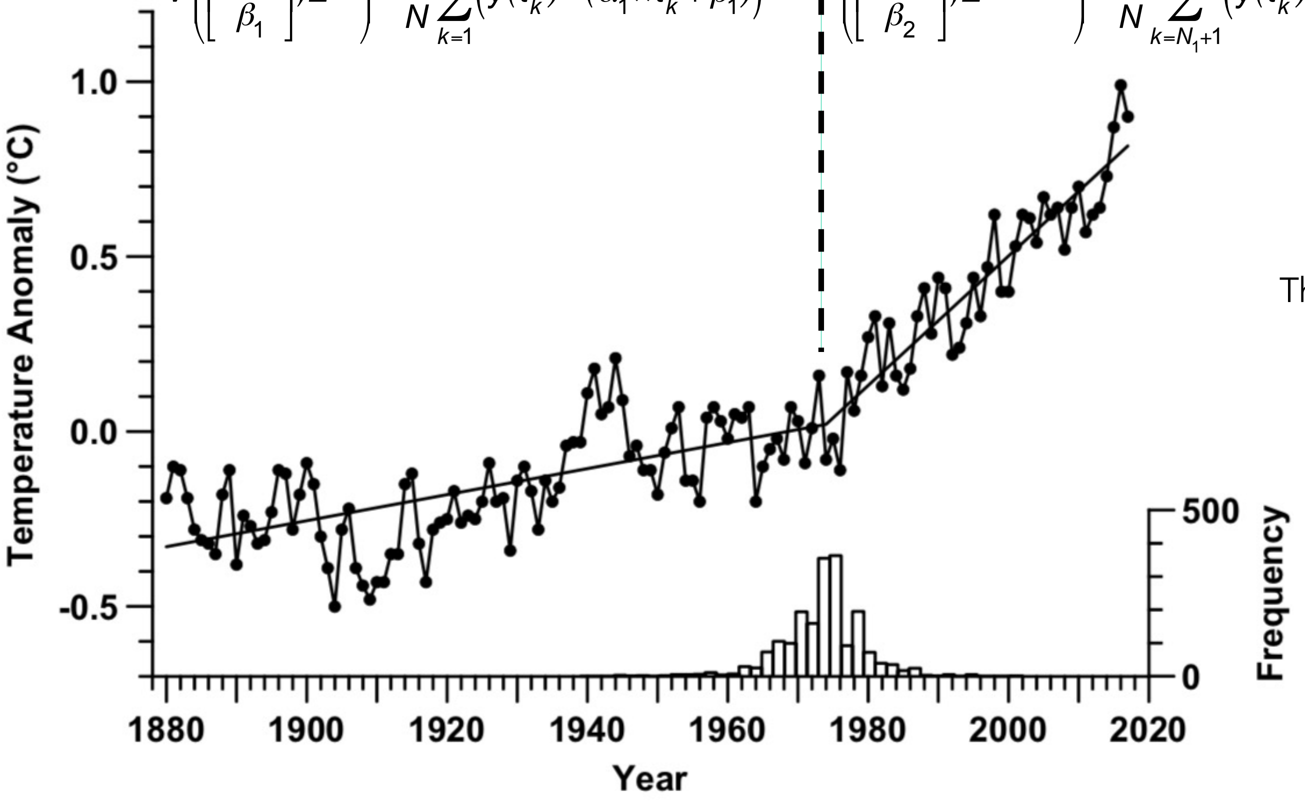


We can do better !

Estimation of piecewise linear trend models applied to the global surface temperature

$$V\left(\begin{bmatrix}\alpha_1\\ \beta_1\end{bmatrix}, Z^{N_1}\right)=\frac{1}{N}\sum_{k=1}^{N_1}\left(y(t_k)-(\alpha_1\times t_k+\beta_1)\right)^2$$

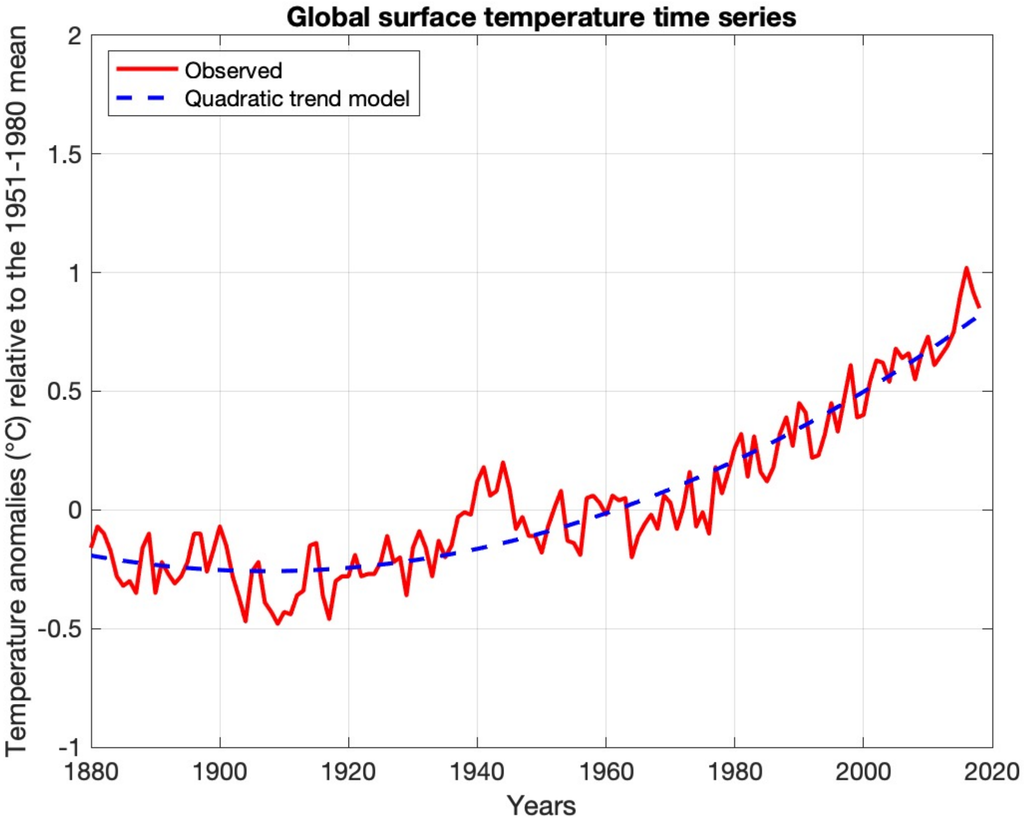
$$\left(\begin{bmatrix}\alpha_2\\ \beta_2\end{bmatrix}, Z^{N_2-N_1-1}\right)=\frac{1}{N}\sum_{k=N_1+1}^{N_2}\left(y(t_k)-(\alpha_2\times t_k+\beta_2)\right)^2$$



This is much better !

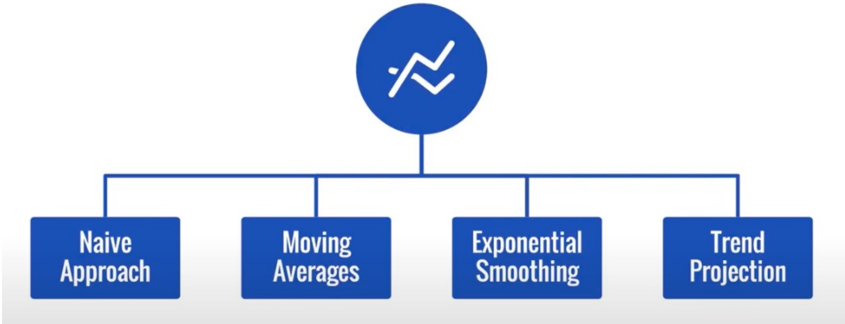
Estimation of a quadratic trend model applied to the global surface temperature

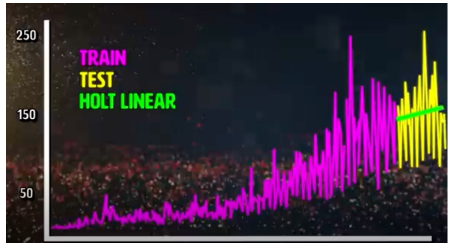
$$V\left(\begin{bmatrix}\alpha\\ \beta\\ \gamma\end{bmatrix}, Z^N\right) = \frac{1}{N}\sum_{k=1}^{N}\left(y(t_k) - (\alpha t_k^2 + \beta t_k + \gamma)\right)^2$$



This is also much better than the basic linear trend model !

Basic deterministic model based-forecasting methods for long term forecast





- **Naïve approach forecast**

Forecasts equal to last observed value : $\hat{y}_{t+h} = y_t \qquad \forall\ h > t$

- **Naïve seasonal approach forecast**

Forecasts equal to last value from same season : $\hat{y}_{t+h} = y_{t+h-km} \quad \forall\ h > t$

m = seasonal period and $k = \left\lfloor \frac{h-1}{m} \right\rfloor + 1$ where $\lfloor x \rfloor$ is the integer part of x

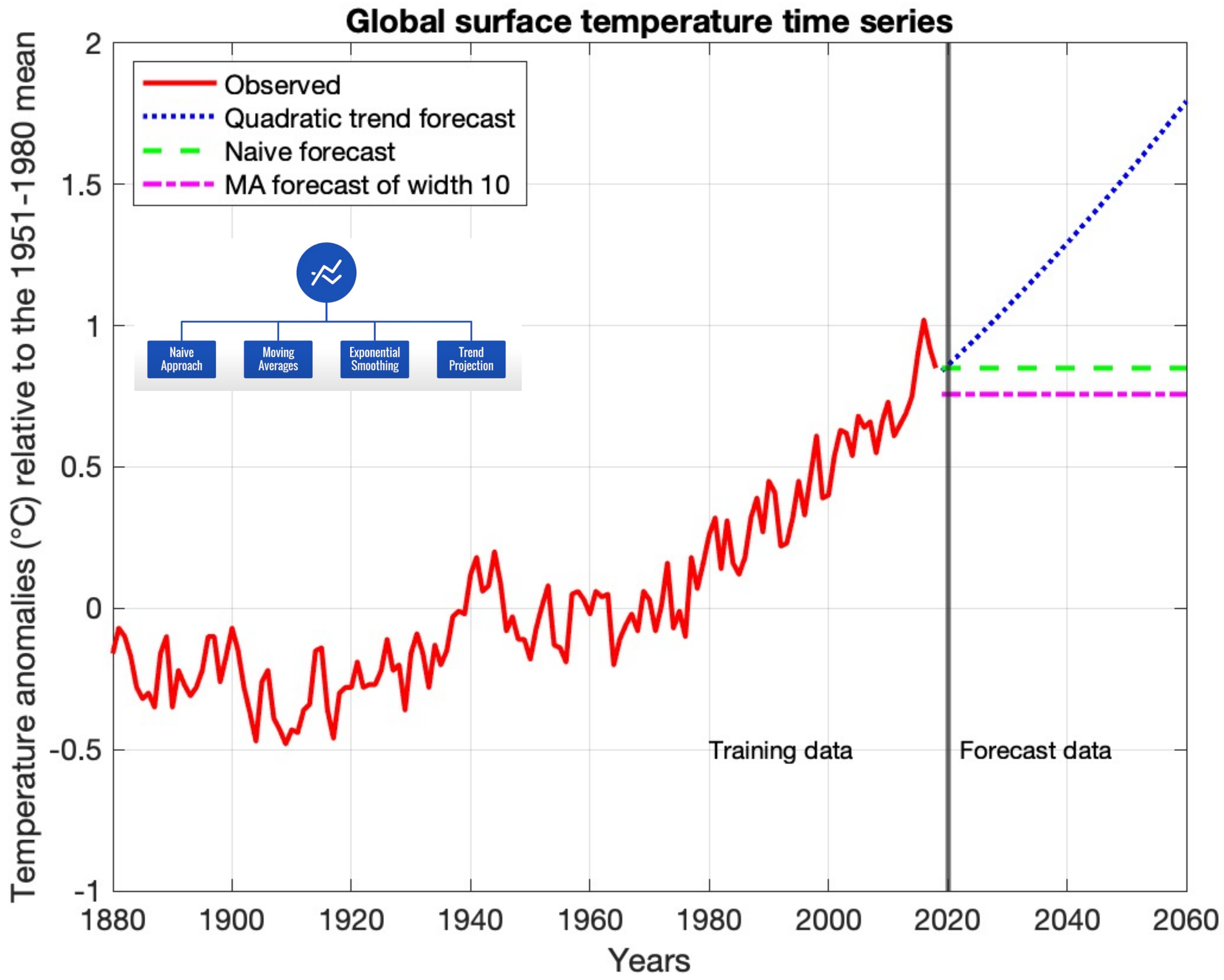
- **Moving average forecast**

$$\hat{y}_{t+h} = \frac{y_t + y_{t-1} + \cdots + y_{t-M+1}}{M} \qquad \forall\ h > t$$

- **Trend projection forecast**

$$\hat{y}_{t+h} = \hat{\beta}_0 + \hat{\beta}_1(t+h) + \hat{\beta}_2(t+h)^2 + \cdots + \hat{\beta}_p(t+h)^p \qquad \forall\ h > t$$

Basic deterministic model based-forecasting for global surface temperature time series

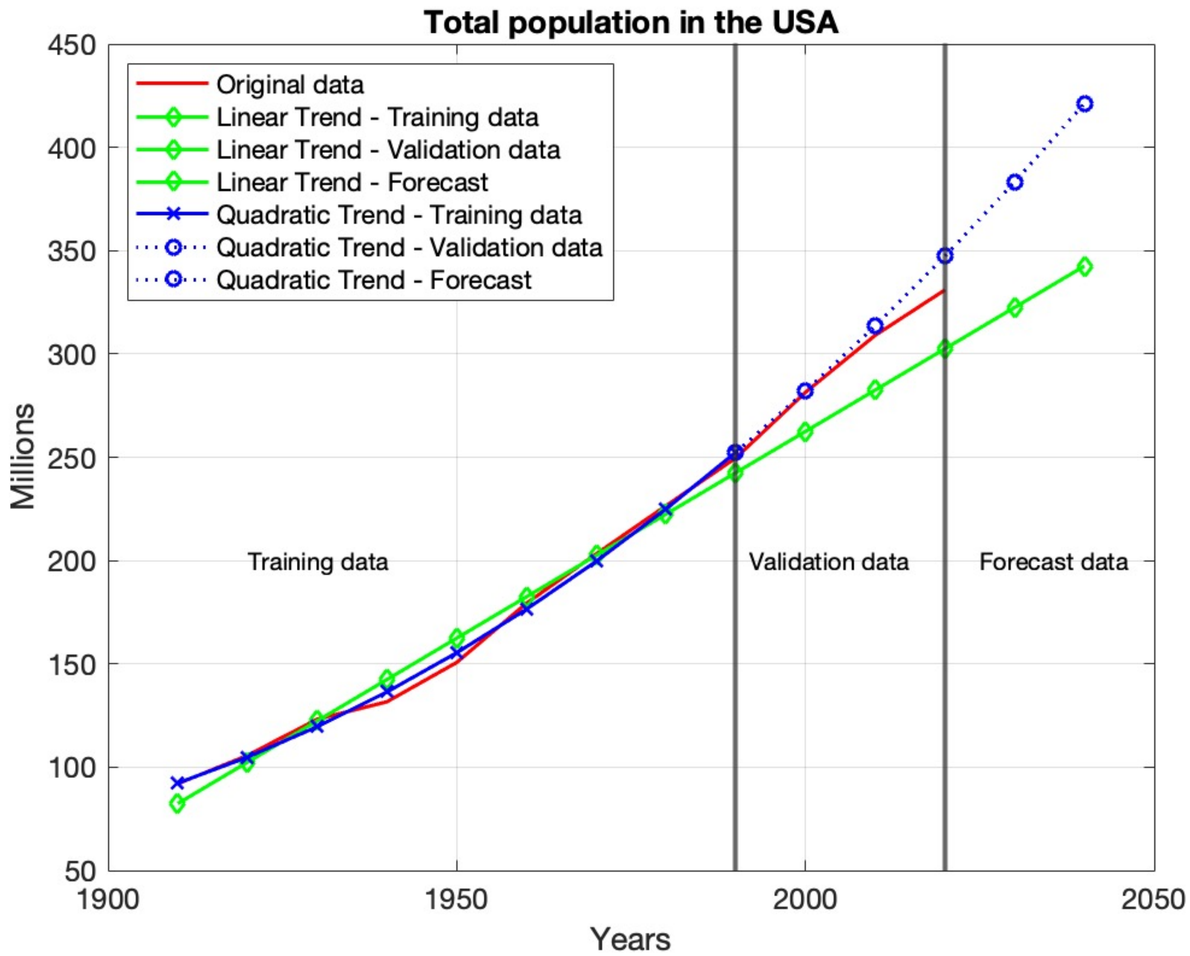


They are quite easy to implement, but the forecast cannot be trusted in long term and these basic methods do not provide any uncertainty intervals

UNIVERSITÉ DE LORRAINE

POLYTECH NANCY

Basic polynomial trend model-based forecasts for the US population



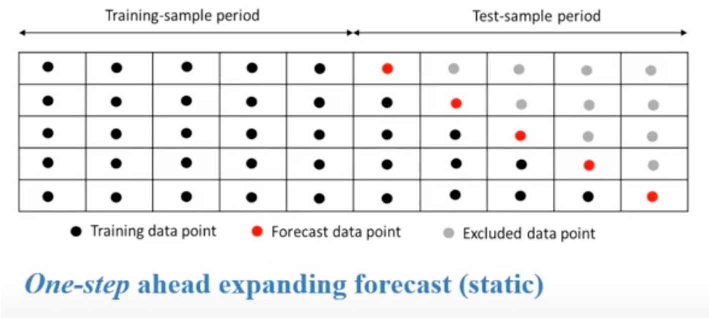
Long term trend models are quite easy to estimate, but the forecast cannot be trusted in long term and these basic methods do not provide any uncertainty intervals

Types of forecasts

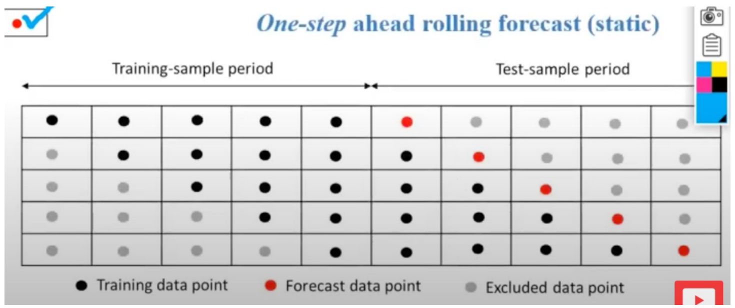
- One-step-ahead forecast
  - A one-step-ahead is a forecast for the next observation only
- Expanding one step-ahead forecast
  - An expanding one step-ahead (or recursive window) forecast means that the initial estimation date is fixed but the additional observations are added one by one to the estimation time span
- Rolling one step-ahead
  - A rolling window is where the estimation time period is fixed but the start and end dates successively increase by 1
- Multi-step-ahead forecast
  - A multi-step-ahead forecast is for 1,2,3,…h steps ahead

Expanding versus rolling forecast for one step ahead prediction

- One-step ahead expanding forecast:
  - train set expands each time step. The data used for computing the one time step ahead forecast is expanded accordindly



- One-step ahead rolling forecast:
  - train set expands each time step but the data used for computing the one time step ahead forecast is fixed



Naive and moving average methods for one step ahead forecast

- **Naïve approach forecast**

Forecasts equal to last observed value : $\hat{y}_{t+1} = y_t \qquad \forall\ t$

- **Naïve seasonal approach forecast**

Forecasts equal to last value from the last season : $\hat{y}_{t+m} = y_{t+h-m} \quad \forall\ t$

m = seasonal period

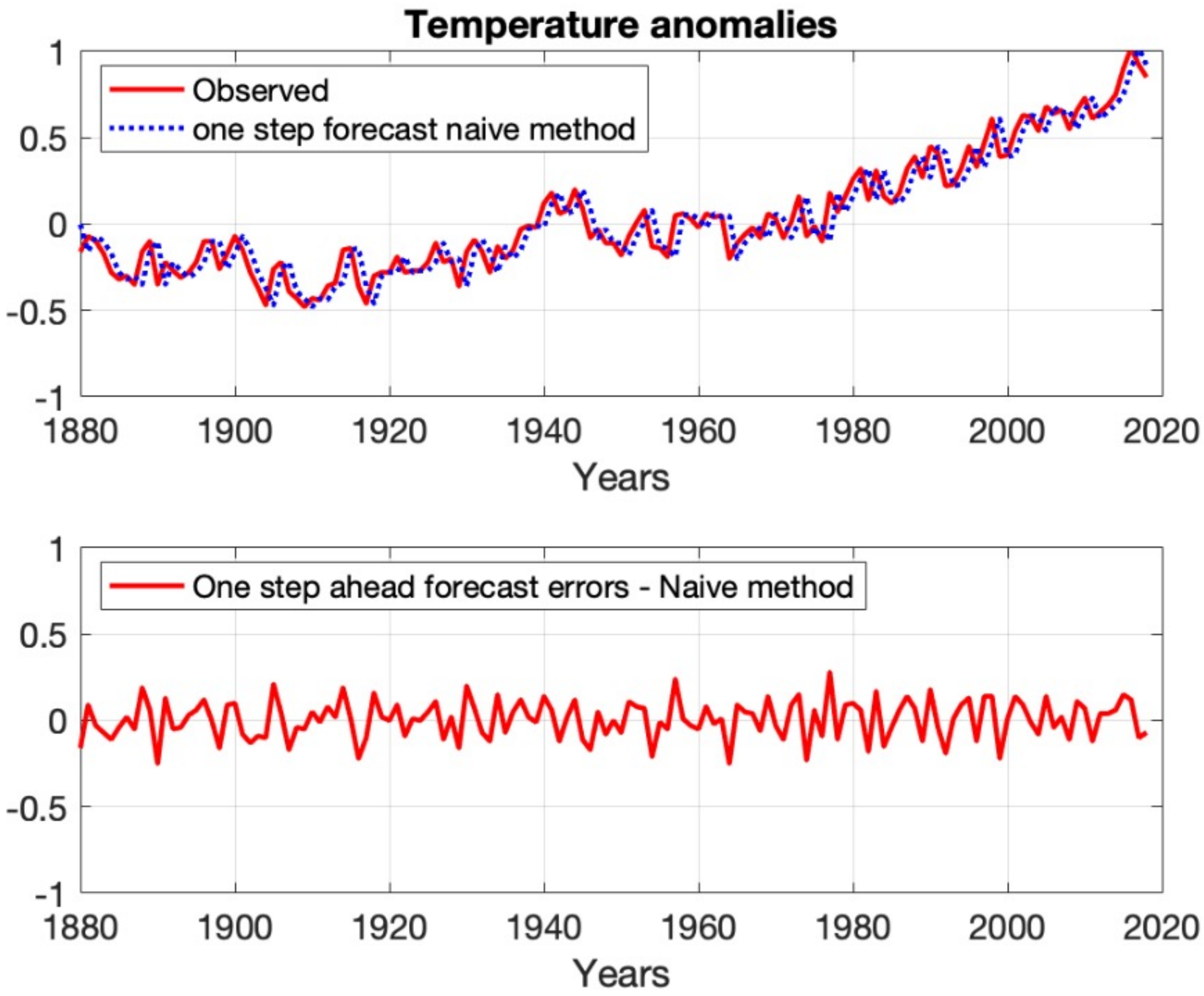
- **Moving average forecast**

$$\hat{y}_{t+1} = \frac{y_t + y_{t-1} + \cdots + y_{t-M+1}}{M} \qquad \forall\ t$$

- **Trend projection forecast**
  - The model is first fitted on the training data set. It can be retrained as new data becomes available

$$\hat{y}_{t+1} = \hat{\beta}_0 + \hat{\beta}_1(t+1) + \hat{\beta}_2(t+1)^2 + \cdots + \hat{\beta}_p(t+1)^p \qquad \forall\ t$$

One step ahead forecast based on the naïve method for global surface temperature time series



One step ahead forecast based on based on the MA method for global surface temperature time series