Time Series Analysis

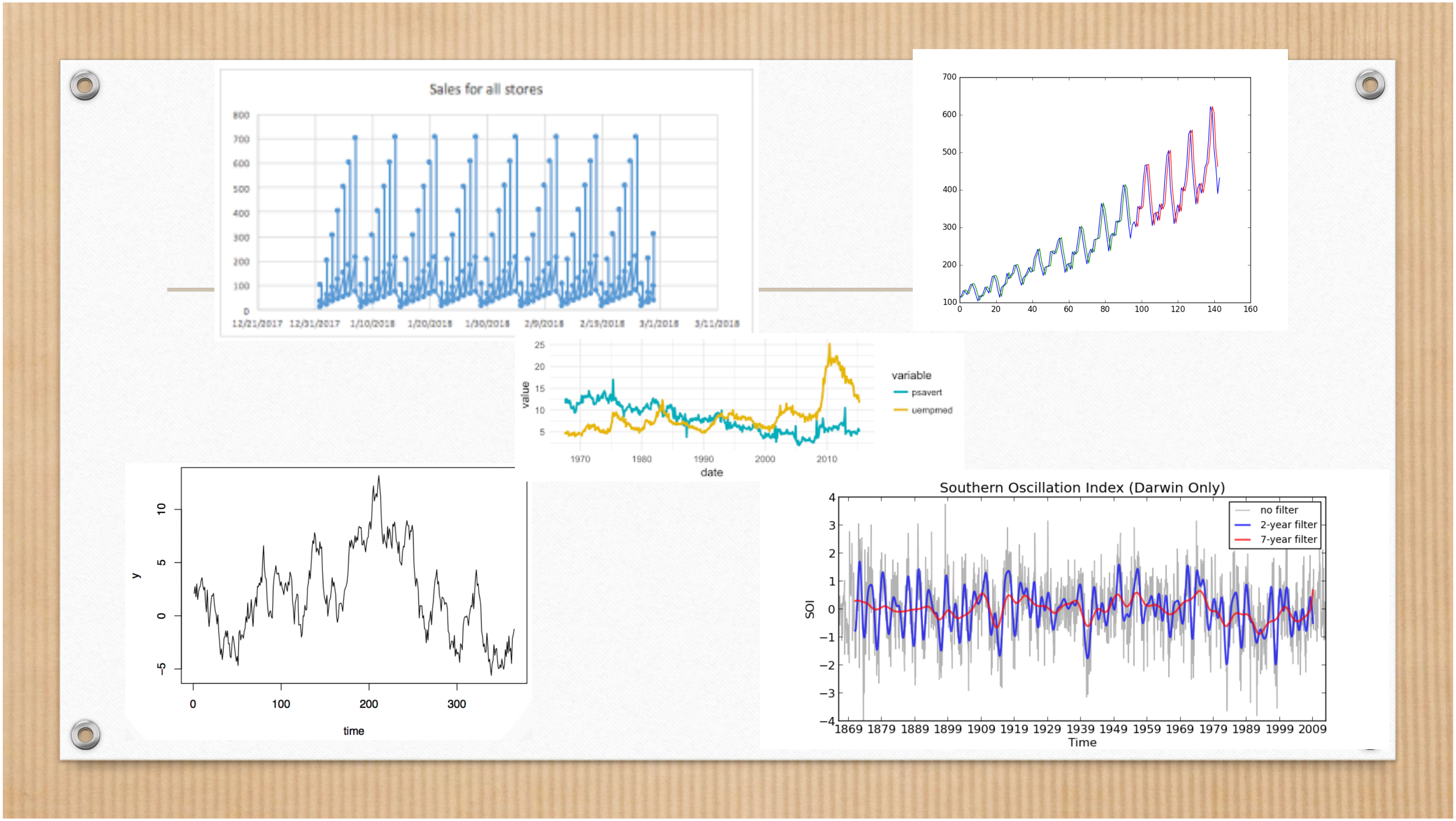
~Abhishek Kumar

Scope

- What is Time series?
- Why it is different than regression?
- What is Signal & Noise?
- Autocorrelation, Seasonality, Trending
- Preprocessing & Filtering

Time Series

- An order sequence of values of a variable at equally spaced time interval


Sales for all stores
0
100
200
300
400
500
600
700
800
12/21/2017
12/31/2017
1/10/2018
1/20/2018
1/30/2018
2/9/2018
2/19/2018
3/1/2018
3/11/2018
100
200
300
400
500
600
700
0
20
40
60
80
100
120
140
160
value
date
variable
psavert
uempmed
1970
1980
1990
2000
2010
y
time
0
100
200
300
Southern Oscillation Index (Darwin Only)
no filter
2-year filter
7-year filter
SOI
Time
1869
1879
1889
1899
1909
1919
1929
1939
1949
1959
1969
1979
1989
1999
2009

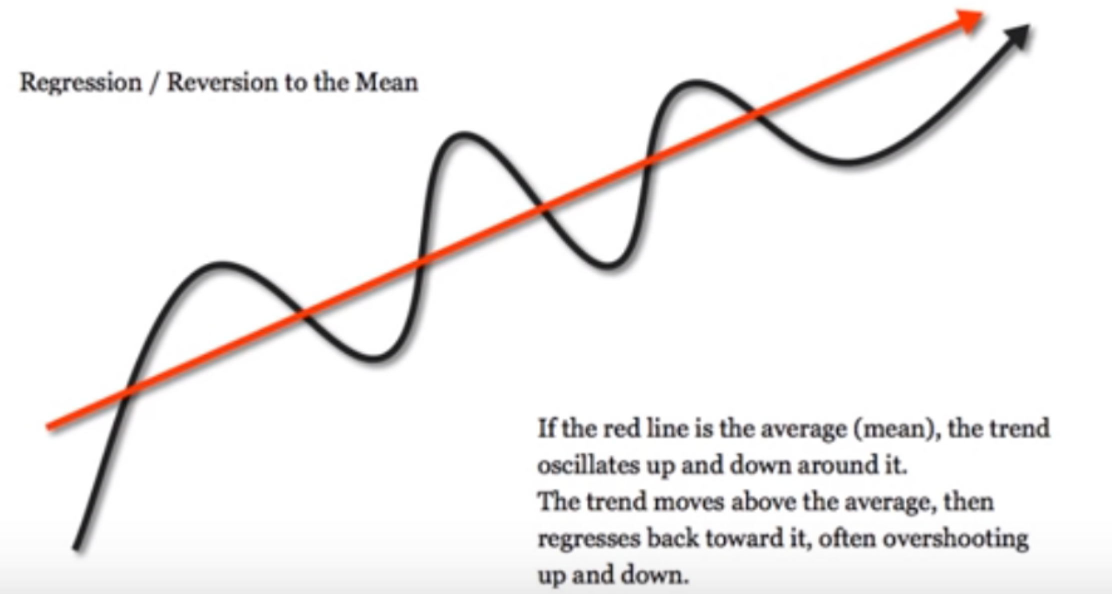
Like Regression?

- It is like regression, time series is often focused on identifying underline trends and patterns, describing them mathematically and then make a prediction/forecast about future events.

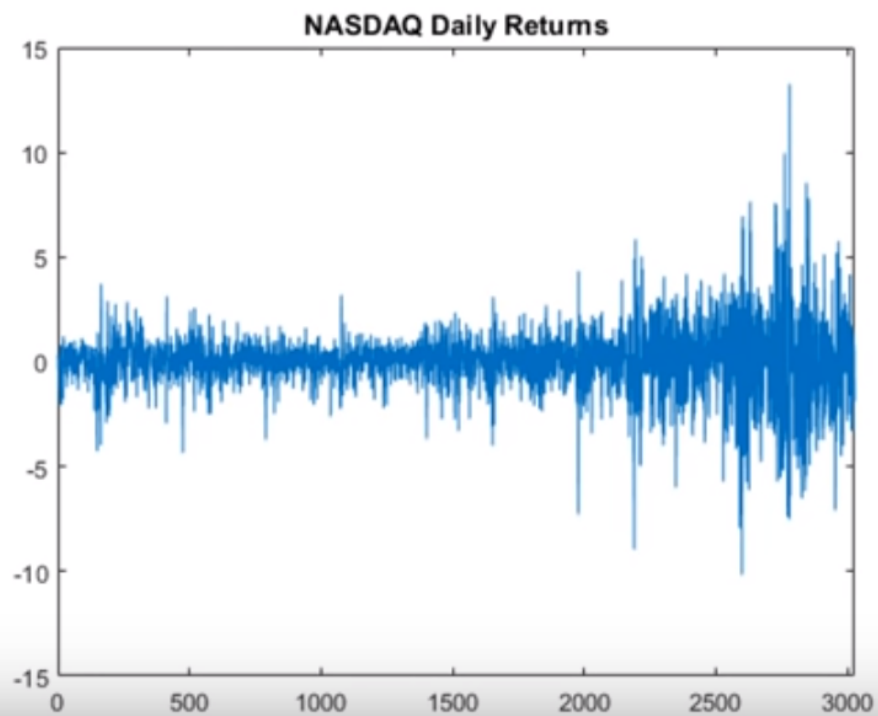
Whats different?

- In many cases, we can describe parts of time series processes in terms of randomized variable with statistical moments.
- An important features of many time series processes is that their mean and variance change through time

Mean is varying, but oscillation around the mean (the standard deviation) looks constant.

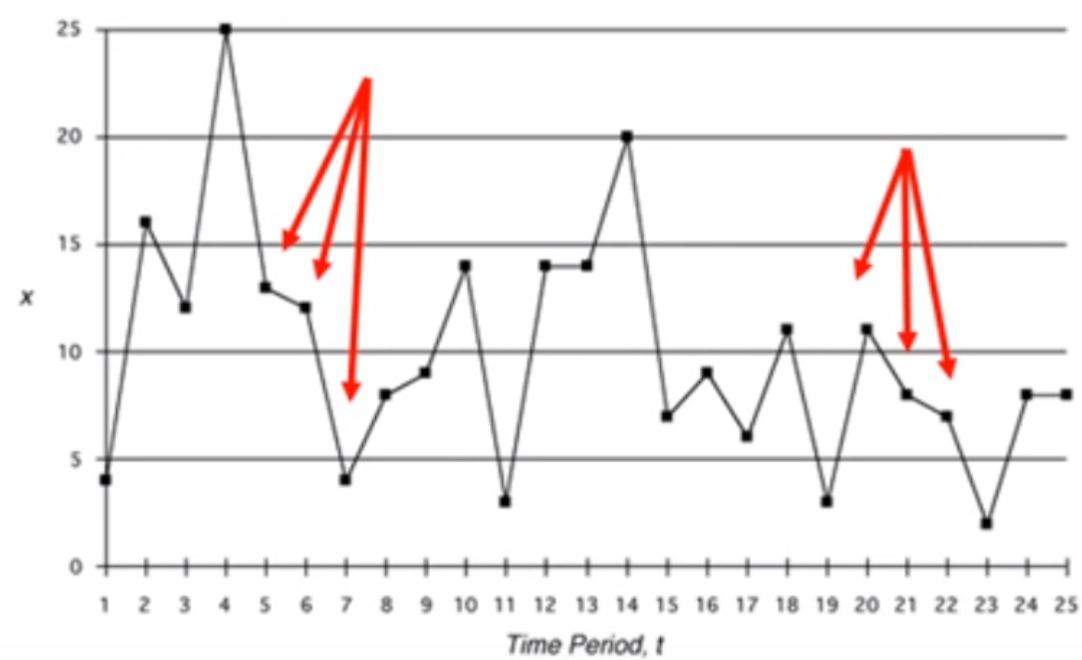


Constant mean, changing standard deviation (“volatility”)

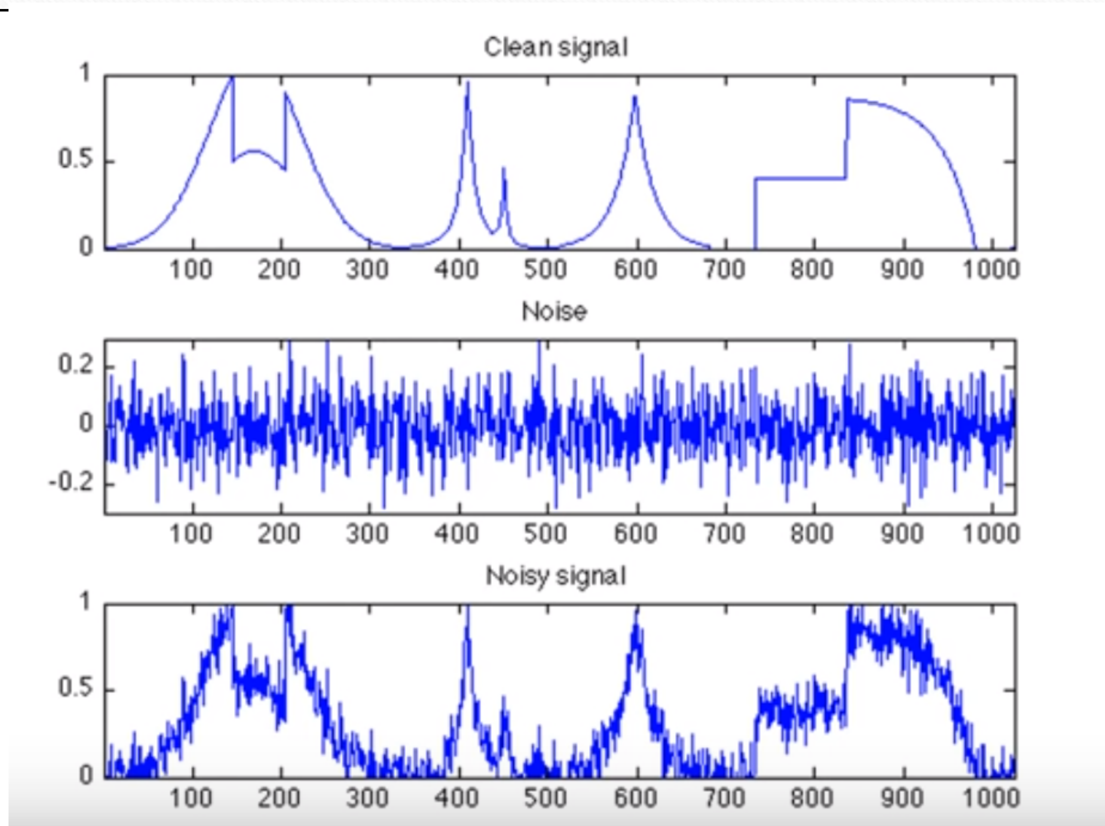


Not if the value of x(t) depends in any significant way on the value of x(t-1).

Are these related?



Time series allow us to replicate every element of the process by decomposing the mathematical process into combination of signals and noise without necessarily knowing the underline causes.


Clean signal
Noise
Noisy signal
1
0.5
0
0.2
-0.2
100
200
300
400
500
600
700
800
900
1000

Signal & Noise

- Signal: Most of we interested in signal
- Noise: random process, with mean 0 and constant variance
- A person should be able to tell the difference between signal and noise.
- Model can never be perfect, because there is always something left and that we call random process or noise/ white noise

Autocorrelation

- If value of x is dependent of any given point of time depends in anyway to the value of x to the time period before.
- If trends tend to persists within time series but overall there is no trend this is called memory in time series or autocorrelation.
- To detect randomness in the data
- Both Noise and signal can have autocorrelation.

Non-stationary

- Change over time, may be your signal remain constant but your noise can vary with time (NASDAQ)

Seasonality

- Periodic patterns

Preprocessing & Filtering

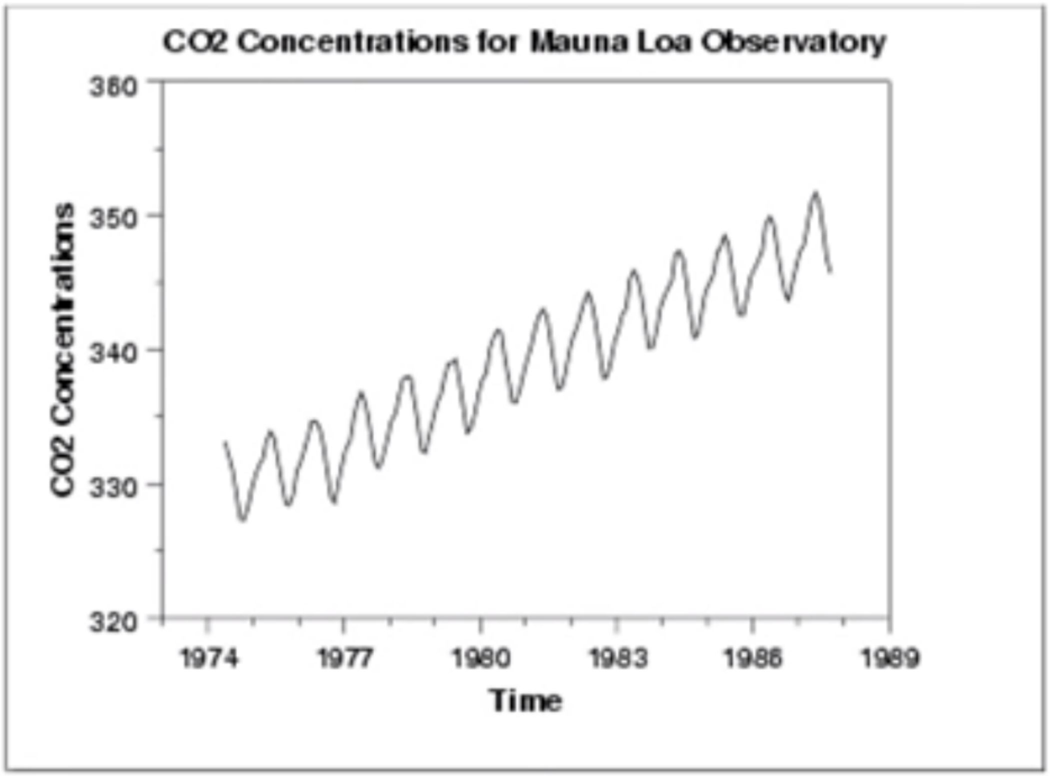
- Take most of the time.
- Filtering: Changing the attribute of a time series or deconstructing it into its component part

Preprocessing & Filtering

- Detrending
- Autocorrelation: take all the signal out.
- Outliers
- “Low pass” filters/ Smoothing

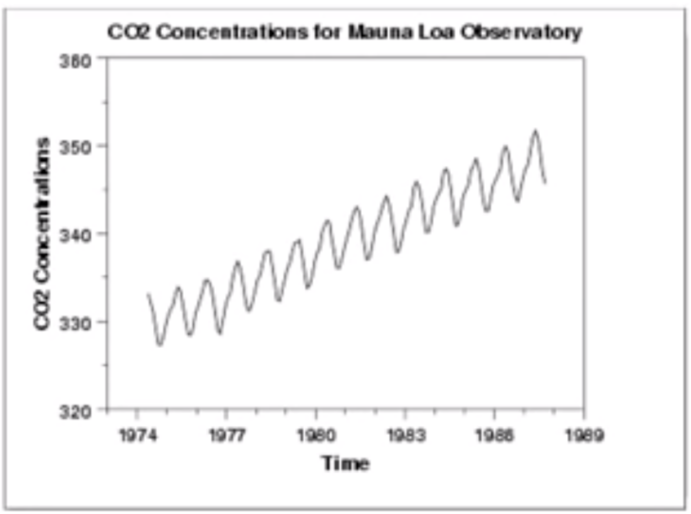
Detrending

- It is used to make series stationary


CO2 Concentrations for Mauna Loa Observatory
CO2 Concentrations
360
350
340
330
320
1974
1977
1980
1983
1986
1989
Time

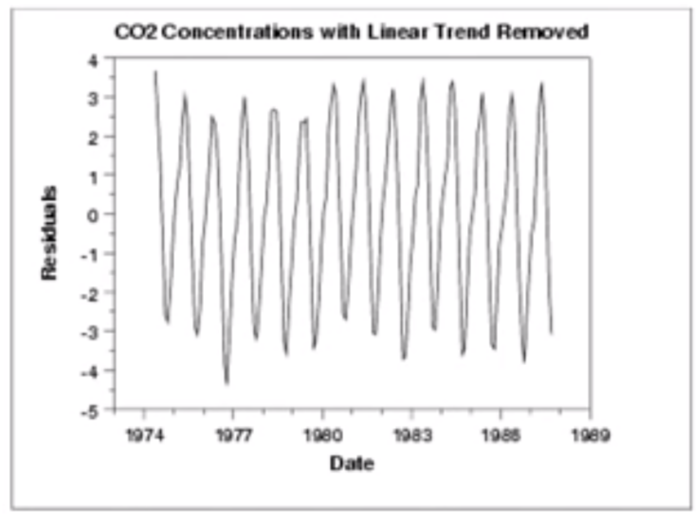
We can fit a linear model to the data using least squares regression:

$$CO_2 = b(Time) + C$$



We can subtract this linear trend from our original data to get "de-trended" CO_2 data.

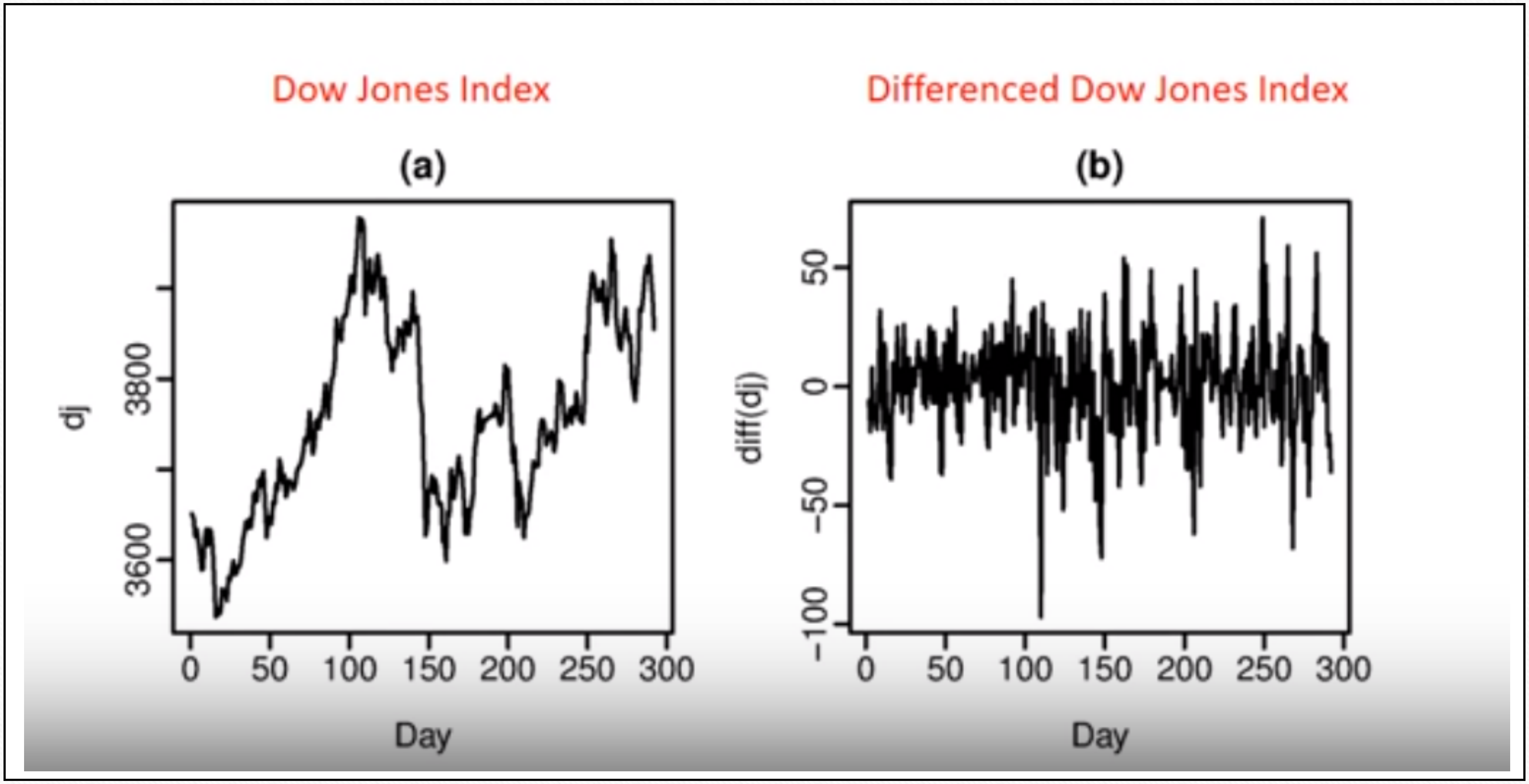
Trend Stationary



Now we're left with seasonality

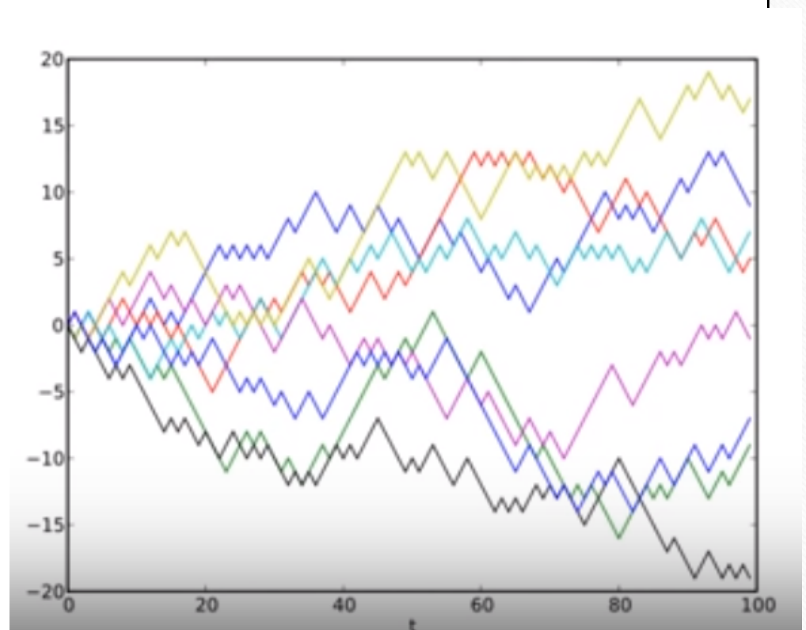
Differencing

- If the mean, variance, autocorrelation of the original time series is not constant in time, even after detrending, in that case we have to transform it into a series of period to period and season to season differences.
- Such a series is called difference stationary.

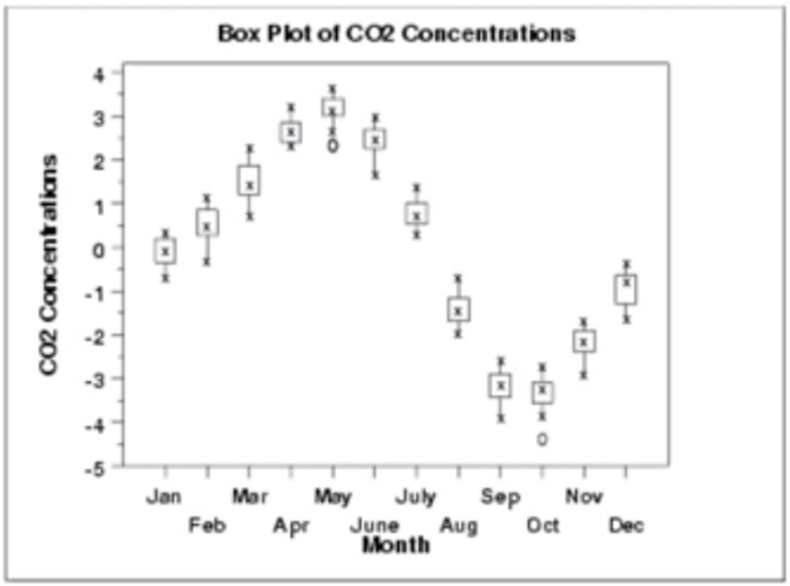

Dow Jones Index
(a)
dj
3600
3800
0
50
100
150
200
250
300
Day
Differenced Dow Jones Index
(b)
diff(dj)
50
0
−50
−100
0
50
100
150
200
250
300
Day

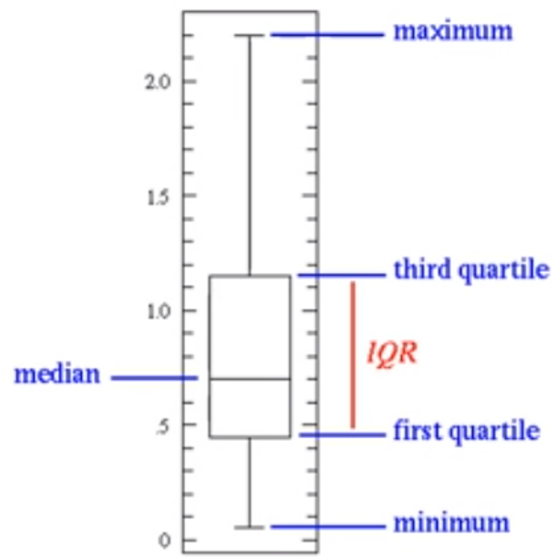
- If the differences of Y is stationary and also completely random(not autocorrelated) then Y is described as Random walk model, which means each value is a random step away from previous value.

$$y_t = y_{t-1} + w_t$$

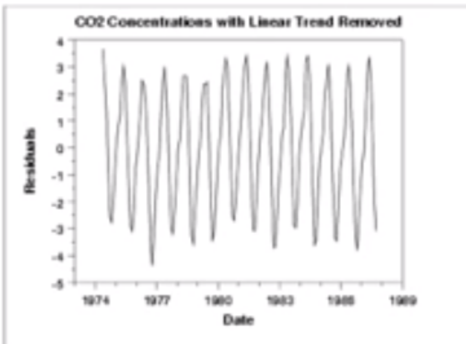


Many time series display **seasonality** (periodic fluctuations). If seasonality is present, it must be incorporated into a time series model. **How do we detect it?**

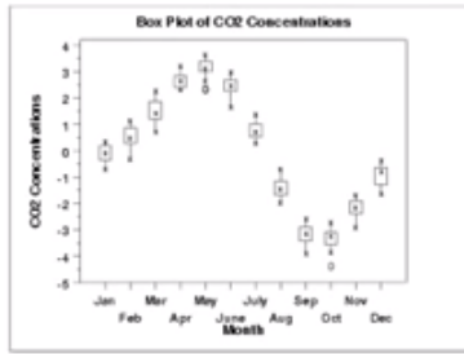




We can likewise subtract the seasonal or periodic trend from the data, leaving a de-trended process.



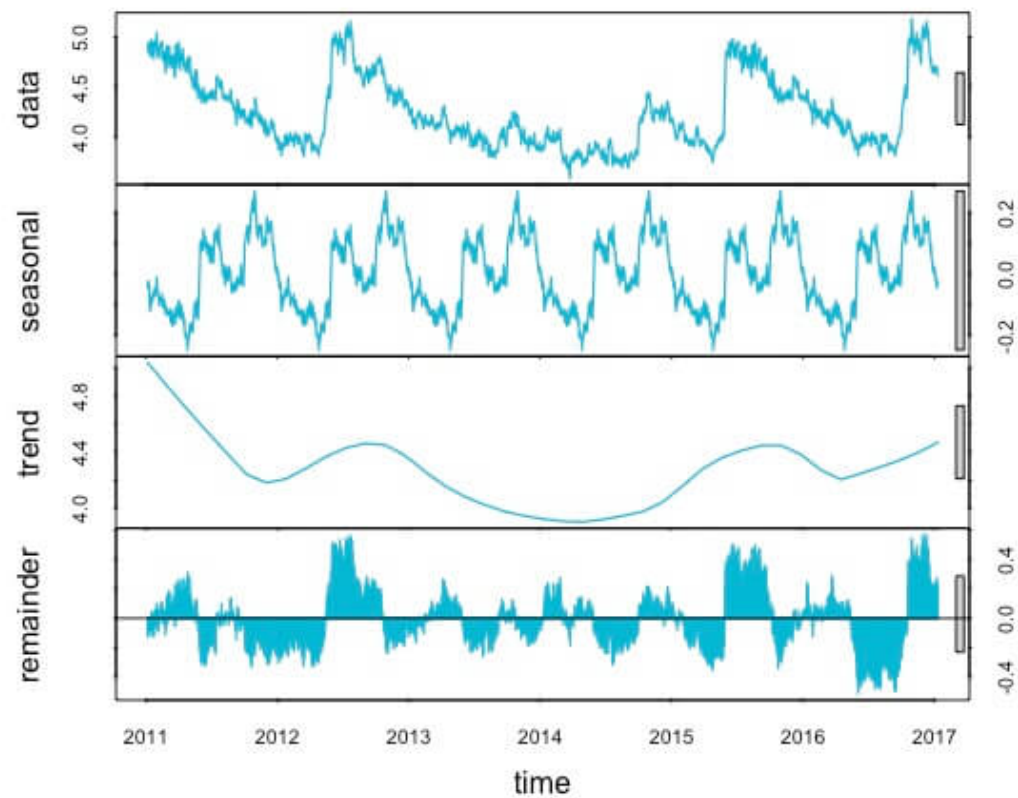
–



=

We might get something that looks like this:

Is this a random process? (can we replicate it by just sampling from a known distribution?)


data
seasonal
trend
remainder
4.0
4.5
5.0
-0.2
0.0
0.2
4.0
4.4
4.8
-0.4
0.0
0.4
2011
2012
2013
2014
2015
2016
2017
time

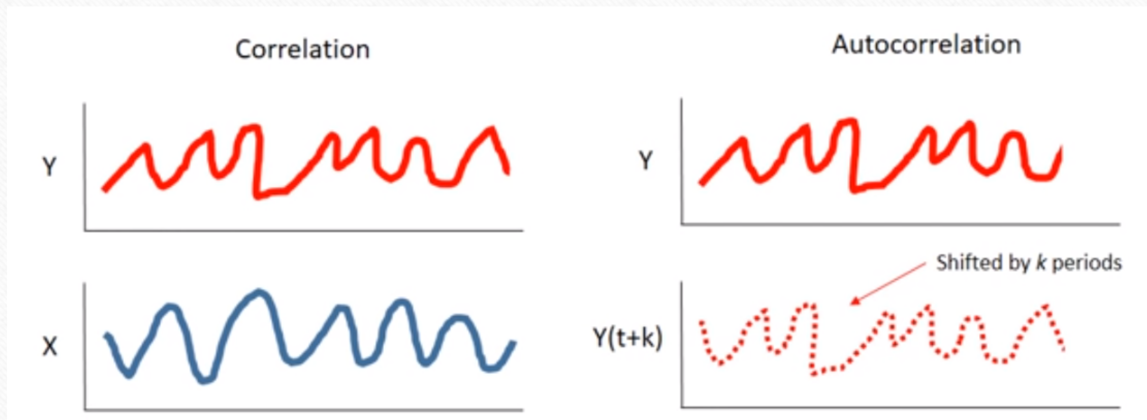
Autocorrelation

$$r_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

$$r_{X,Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \; \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

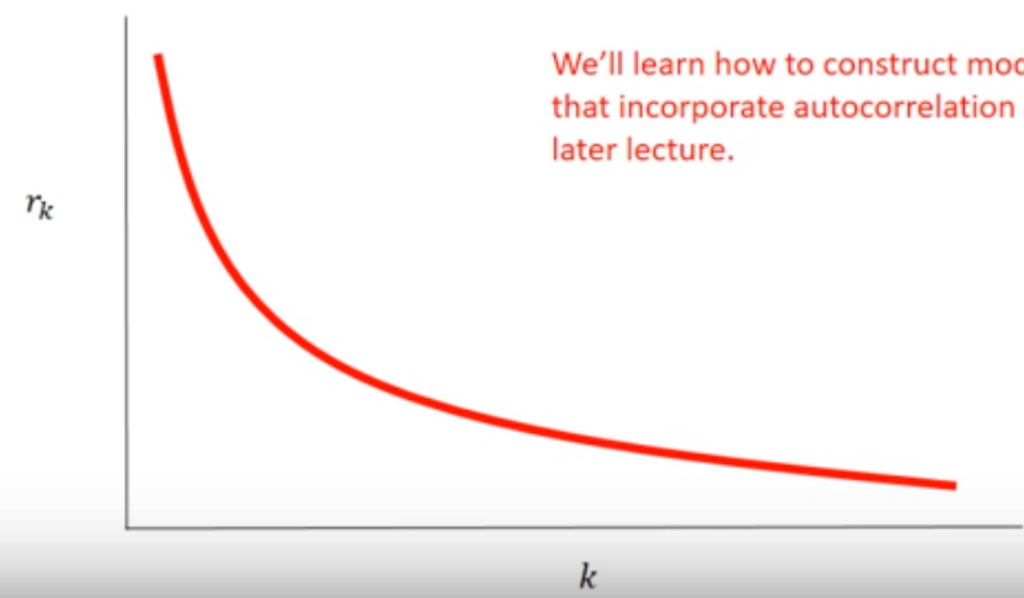
Correlation vs Autocorrelation

- We are just measuring the correlation between time series and itself, at different time lags.



We can test the autocorrelation at as many lags as we want, depending on the length of the time series. Then we can plot autocorrelation as a function of lag.

What does this tell us?



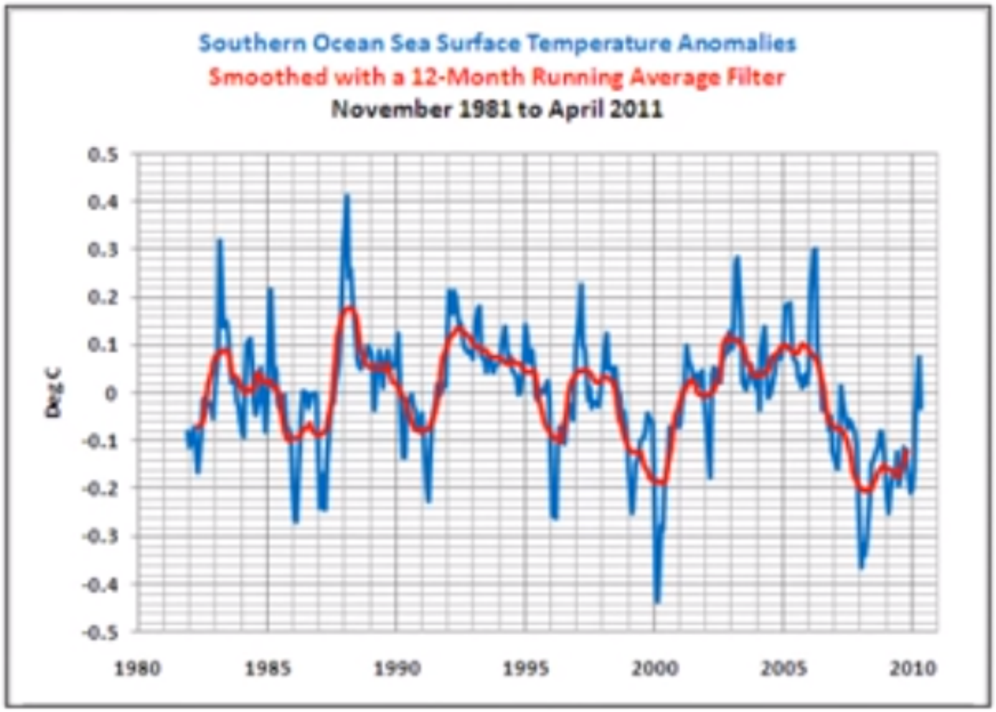
We'll learn how to construct models that incorporate autocorrelation in a later lecture.

White Noise

- If we eliminate the signal(trends, periodicity, autocorrelation) then we are left with white noise.
- White Noise is a completely random process, whose sample are regarded as a sequence of uncorrelated random variables with zero mean and constant variance.

Smoothing

- A form of filtering which produces a time series in which the importance of the spectral components at high frequency is reduced. Also called low pass filer because low frequency are allowed to pass through


Southern Ocean Sea Surface Temperature Anomalies
Smoothed with a 12-Month Running Average Filter
November 1981 to April 2011
Deg C
0.5
0.4
0.3
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
-0.5
1980
1985
1990
1995
2000
2005
2010

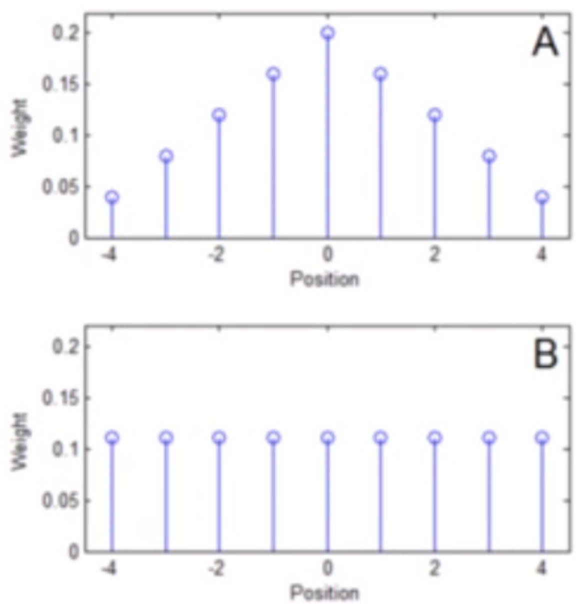
Digital Filter

Table 1. Filtering

Year	Filter	Time Series	Filtered Values
1		12	
2	.25 x	17	14.00
3	.50 x	10	14.75
4	.25 x	22	17.25
5		15	15.75
6		11	13.75
7		18	18.50
8		27	21.50
9		14	

Filter proceeds by sliding alongside the time series one value at a time, each time computing a cumulative product.

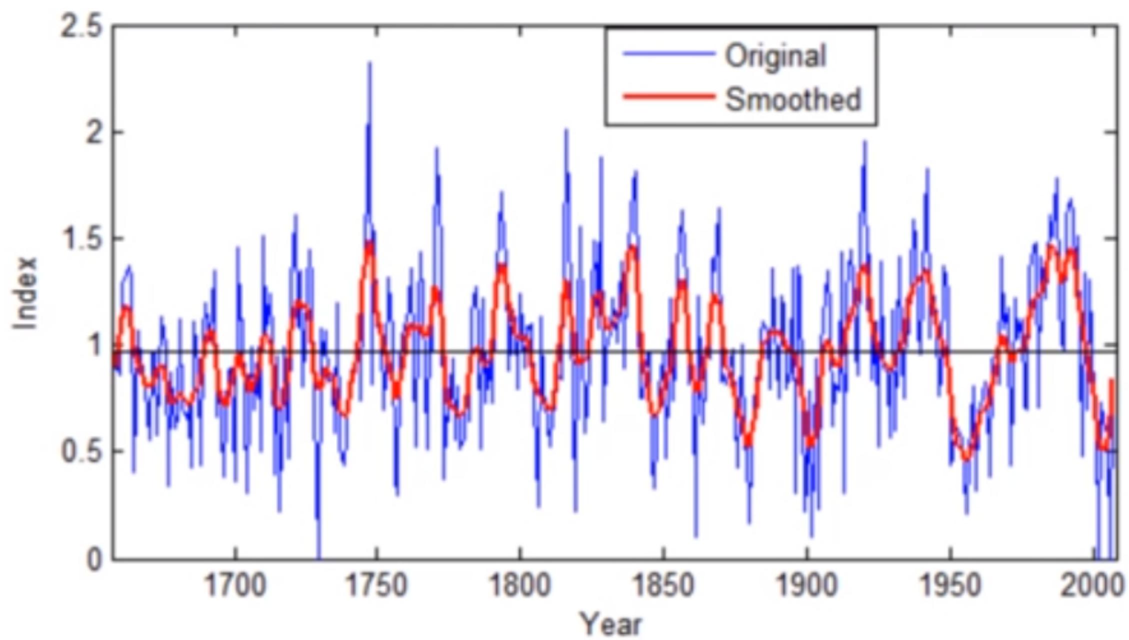
Smoothing filters can be designed lots of different ways.

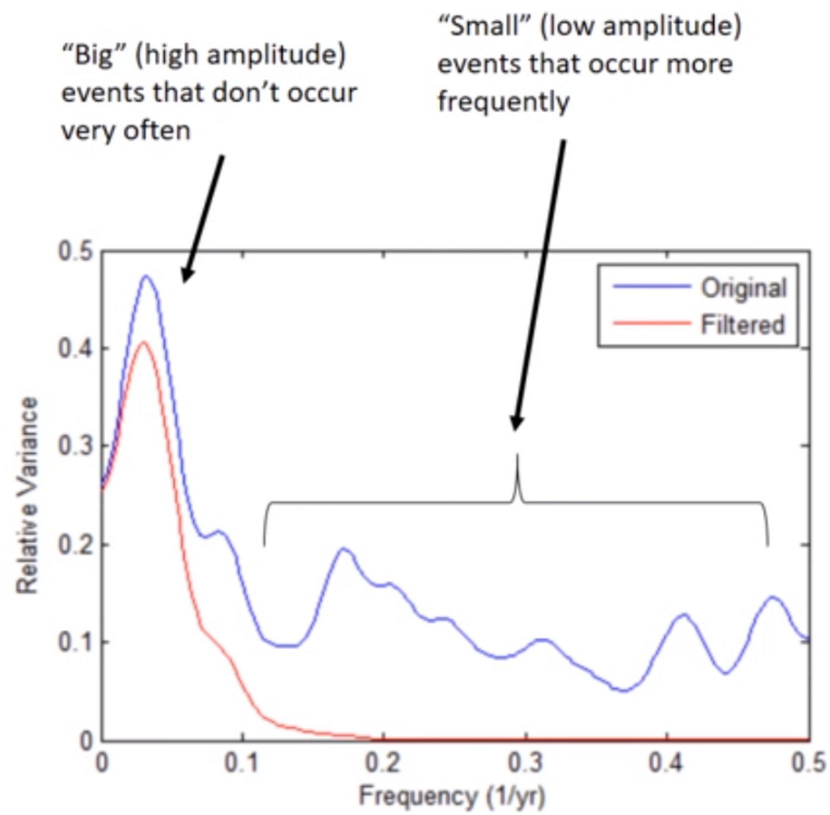


Linear decay.

Moving average.

End result. Why do this in the first place?





"Big" (high amplitude) events that don't occur very often
"Small" (low amplitude) events that occur more frequently
Original
Filtered
Relative Variance
Frequency (1/yr)
0
0.1
0.2
0.3
0.4
0.5

Issue with smoothing

- You loose some data/infromation

Discussion

Thank you!!!!!!