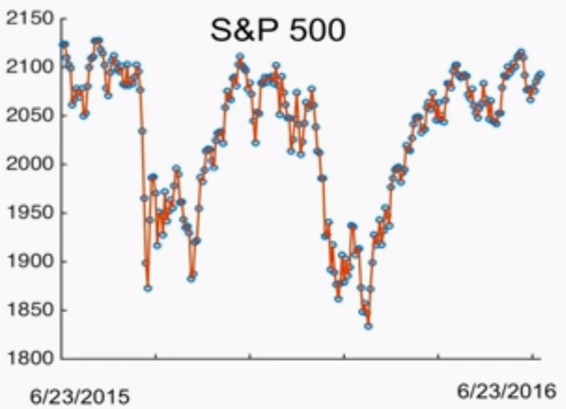
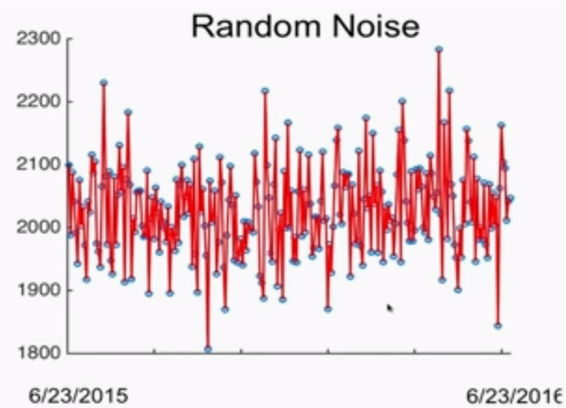
module1 · time series and forecasting

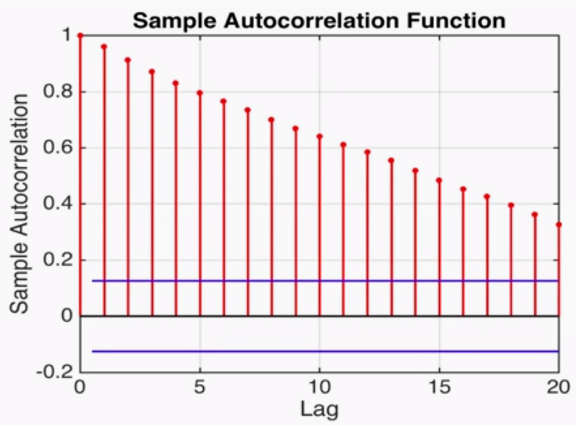
introduction to time series

times series data does not have the same characteristics of random independent numbers

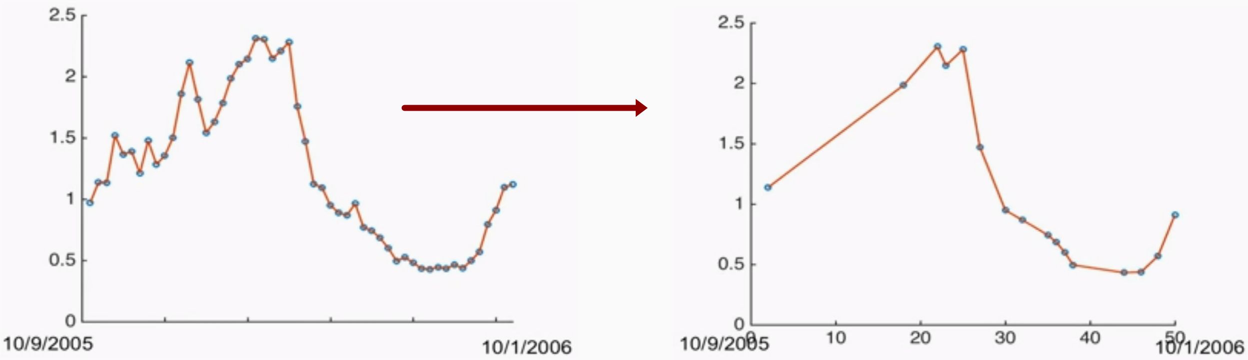




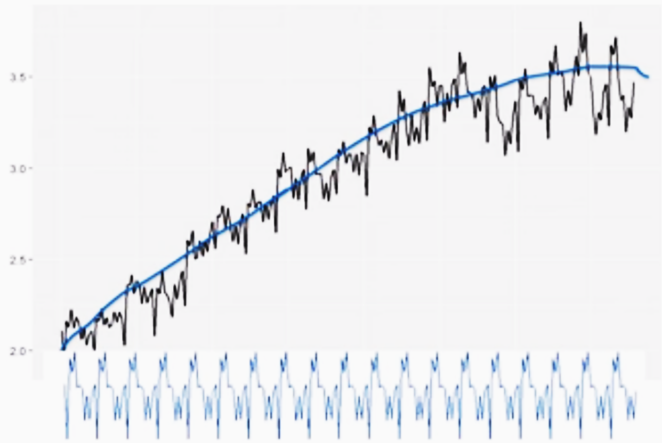
Time series data has periodicity qualities that are capable of being predicted

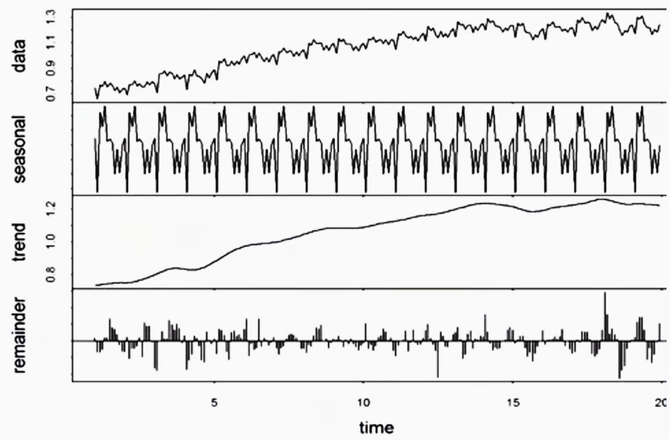


With time series data, irregular reporting times can affect the quality or representation of data



A trendline can be decomposed into various components. For example, the following data consists of a periodic signal and a stationary signal. Periodic referring to a repetitive pattern and stationary referring to data whose mean and standard deviation remain constant over time





Performing trend decomposition automatically is achieved through lowess/loess regression

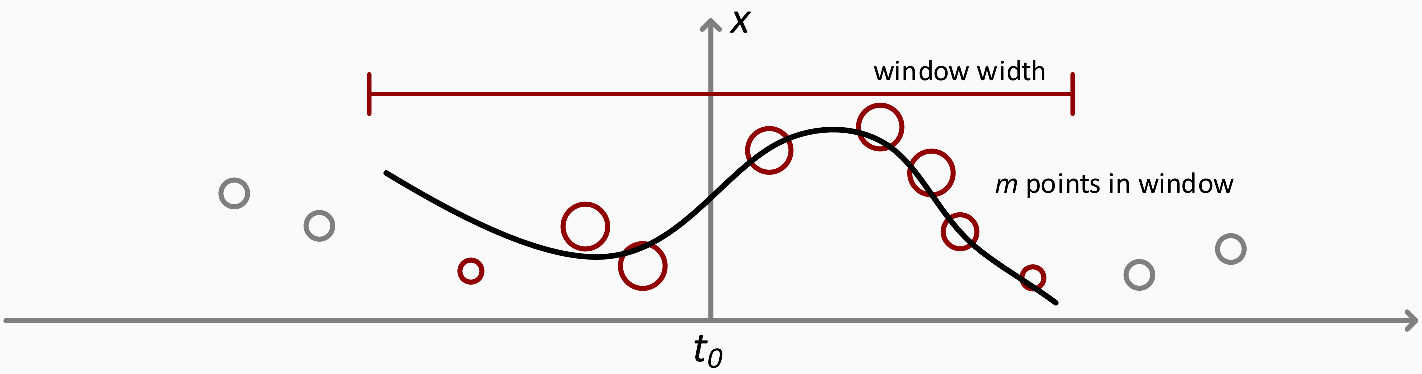
Lowess – locally weighted scatterplot smoothing

Loess – local regression

Each essentially fits local polynomial models and merges accordingly

Focusing on Loess regression:

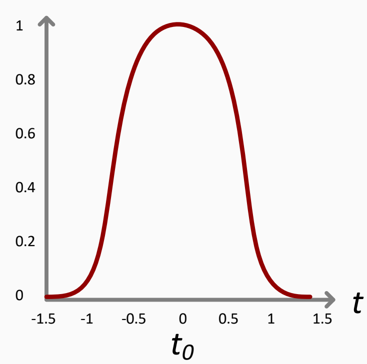
1. the window width defined as m uses m nearest neighbors in the window for local regression
2. a weighting function applies weights to points nearest to the center
3. the data is trended through regression using polynomial terms and the assigned weights
4. repeat the above for each value of t



the weighting function normalizes the data by scaling the data to 1 through the tri-weight function:

$$W(u) = \begin{cases} (1-|u|^3)^3 & \text{if } |u| \leq 1 \\ 0 & \text{if } |u| > 1 \end{cases}$$

$(1 - \text{distance}(t, t_0)^3)^3$



the scaled distance to neighbor i is $u_i = \frac{t_i - t_0}{\text{width}} = \frac{\text{distance from the center}}{\text{scaling factor}}$

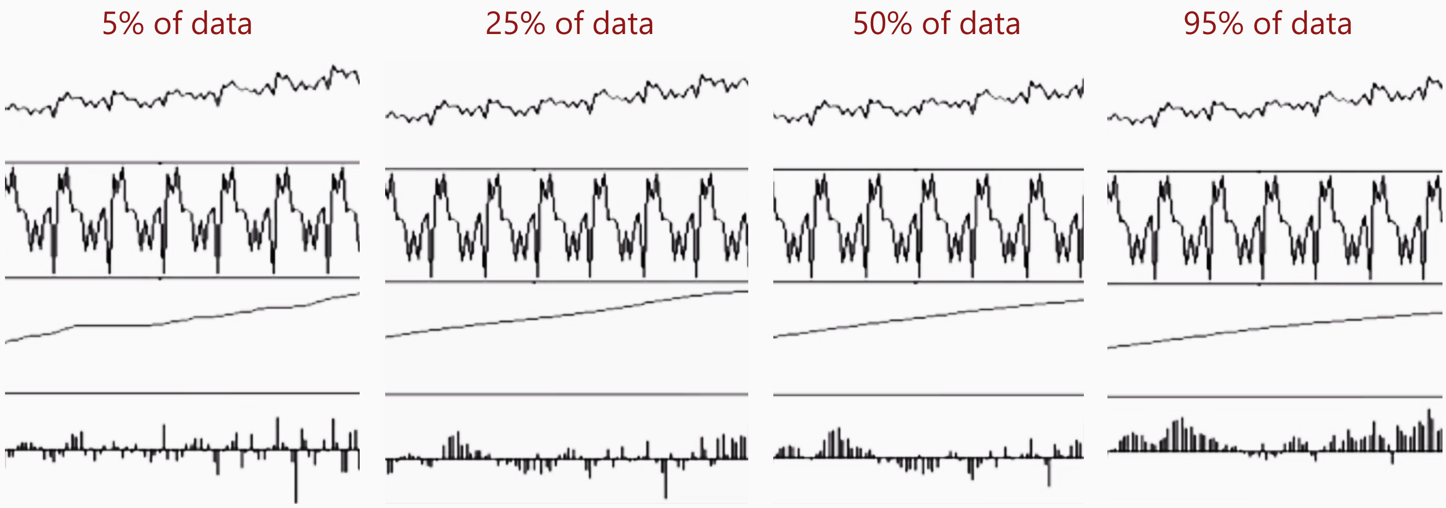
weight for neighbor i is: $w_i = W(u_i)$

the data is fitted through taylor's theorem from calculus:

$$f(t) \approx \beta_0 + \beta_1(t_i - t_0) + \frac{1}{2}\beta_2(t_i - t_0)^2 + \cdots$$

$\hat{\beta} = \arg\min_{\beta_0, \beta_1, \beta_2} \sum_{i=1}^{n} w_i(x_i - \hat{x}_i)^2$ where $\hat{x}_i = \hat{\beta}_0 + \hat{\beta}_1(t_i - t_0) + \frac{1}{2}\hat{\beta}_2(t_i - t_0)^2$ if $t_i = t_0$: $\hat{x}_i = \hat{\beta}_0$

the window with is determined; larger windows will produce flatter trends and larger residuals

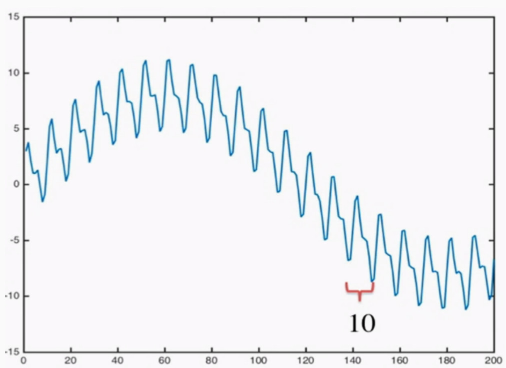


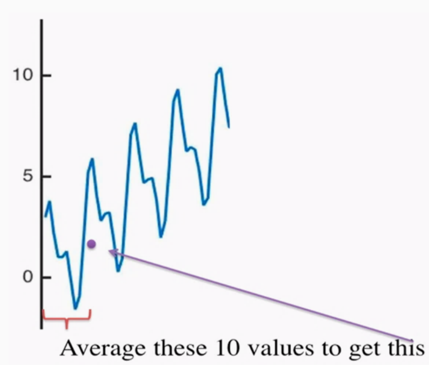
after loess has been applied to understand the trend of the given data, moving average smoothing removes the period signals to obtain a stationary signal

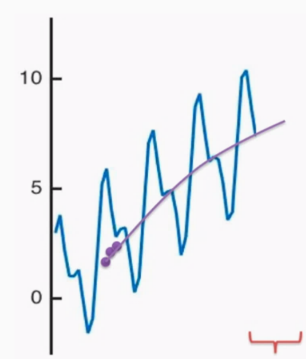
this is another way to remove an obvious periodic signal through the use of a low-pass filter (moving averages)

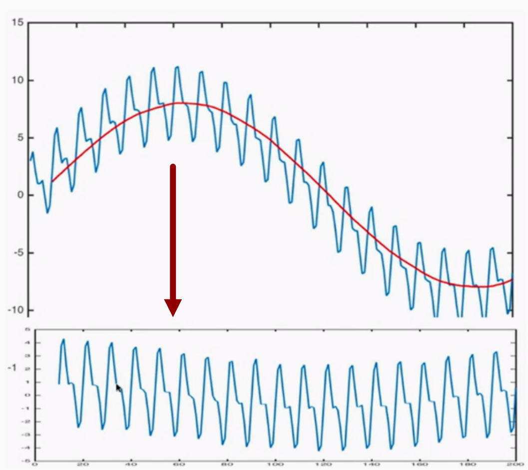
the example (left) illustrates a signal with obvious periodicity. the method is applied simply by averaging the prior 10 values in obtaining each average example (middle figure). the formula is applied forward to compute each successive point and then trended over the plots. the formula for estimate is the average of the previous c points where c is the periodicity of the signal: $\hat{x}_i = \frac{1}{c}(x_i + x_{i-1} + \cdots + x_{i-c+1})$

if c is chosen incorrectly, the formula could fail to compute correctly

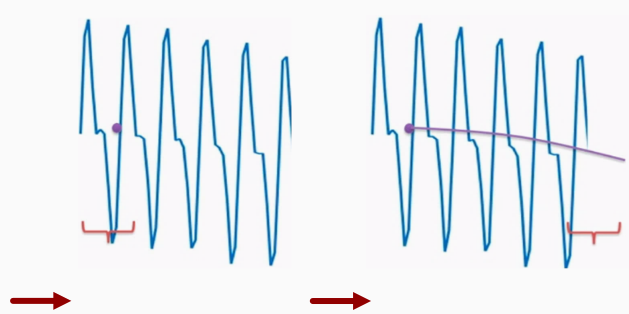


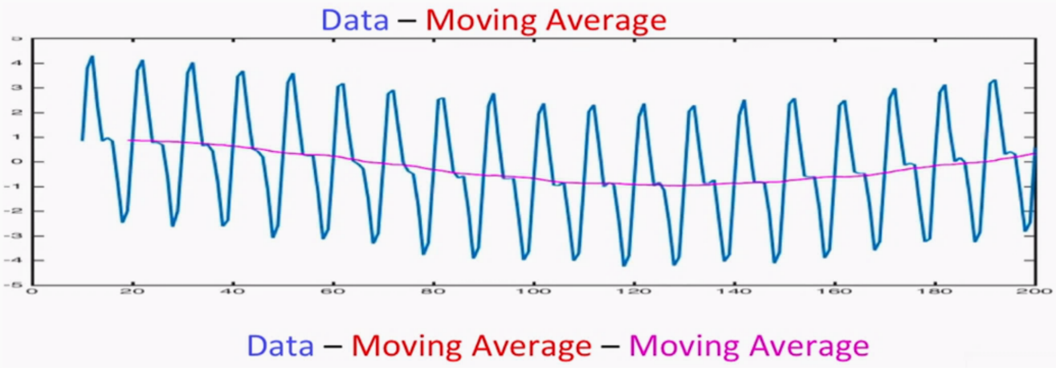




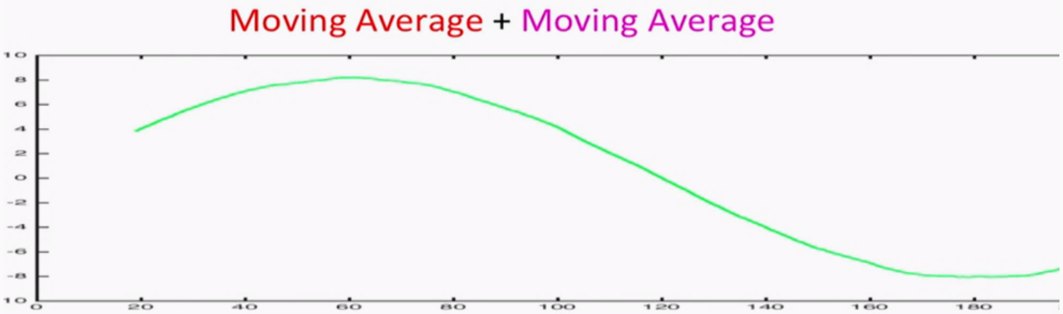


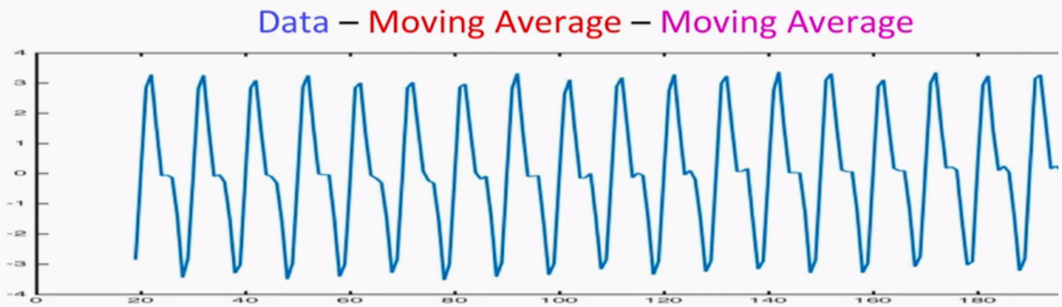
after a first attempt, the moving averaged data still appears to have a trend when the moving average is subtracted from the data (below figure on the left). the approach to refine the model reiterates through the process:



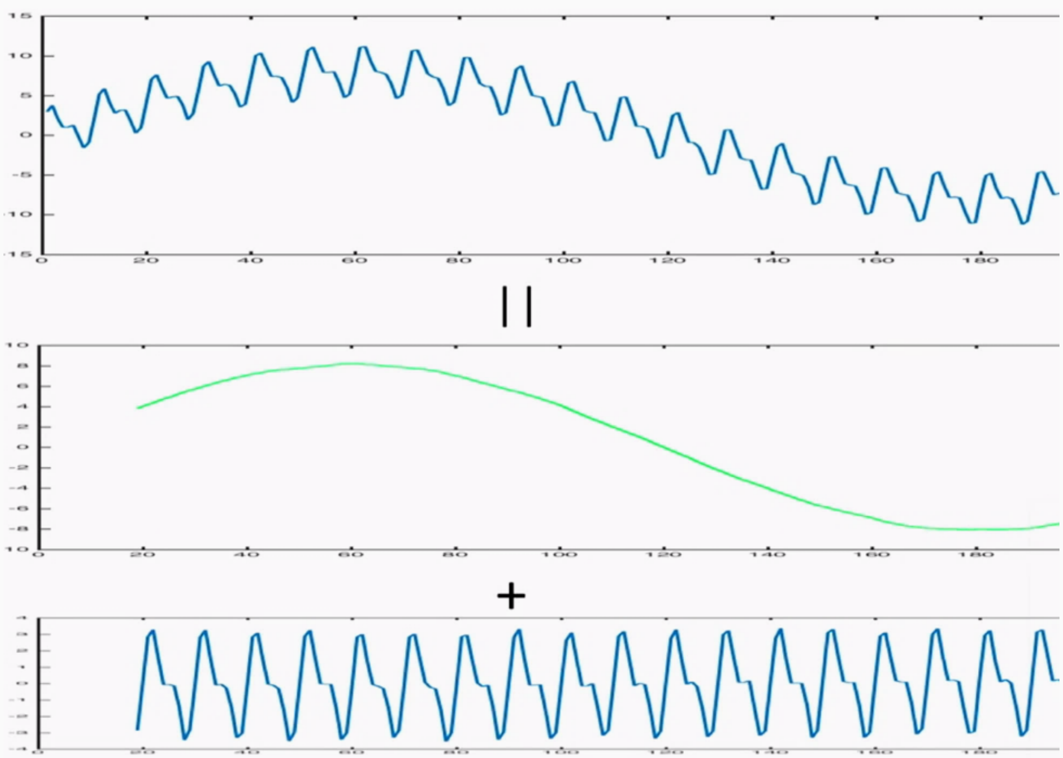


examing the reiterative trend after separating the trend and periodicity contains little to no residual:





the summary decomposition performed through moving averages is as follows:



theorhetical recap:

loess – finds trends using polynomial modeling.

- need to determine bandwidth parameter
- coarse, looks at data over possibly many cycles

moving average smoothing – finds trends by averaging over one cycle length

- will not work well if the trend is not smooth
- the length of each periodic cycle needs to be known
- fine-grained smoothing and looks only at the previous cycle (single cycle length), nothing else

moving average smoothing does not provide long-term smoothing like loess