Lecture 13: Time series analysis and model exploration

Wengang Mao (Marine Technology)
Department of Mechanics and Maritime Sciences,
Chalmers University of Technology,
Goteborg, Sweden

Outline

- Models for the time series regression/analysis
- Measures for goodness of fit
- Model/data exploration (AR models)
- Smooth data (Moving Average) models

Models for time series regression

- Let assume the output/sample of time series X_t is influenced (dependent) by a series of independent random process (time series) $Z_{t1}, Z_{t2}, \ldots, Z_{tq}$
- The (regression) model to describe their relationship is normally assumed a linear model as
  - $X_t = \beta_0 + \beta_1 Z_{t1} + \beta_2 Z_{t2} +, \ldots, +\beta_q Z_{tq} + W_t$
  - β_0, $\beta_1, \ldots$, β_q are the coefficients to be estimated through the samples of random process
  - $Z_{t1}, Z_{t2}, \ldots, Z_{tq}$ time series of the random process
  - W_t white noise process (often assumed to be Gaussian)
- Normally, three simple models for the description of X_t
  - $X_t = \beta_0 + \beta_1 f(t) + W_t$, *f(t)* is a deterministic value in terms of time *t*
  - $X_t = \beta_0 + \beta_1 f(Z_t) + W_t$, dependent on a function of another RP without time lag/shift
  - $X_t = \beta_0 + \beta_1 f(Z_{t-h}) + W_t$, dependent on a function of another RP with time lag/shift
  - A combination of the above simple linear regression models

Measures for Goodness of the fit

Questions: **for a general model:** $\boldsymbol{X_t = \beta_0 + \beta_1 Z_{t1} + \beta_2 Z_{t2} +, \dots, +\beta_q Z_{tq} + W_t}$.

- ✓Are all the dependent RPs needed?
- ✓If we reduce one or several RPs, can we still make sure that the simplified model good enough for the data?
- ✓If we reduce r RPs from the model with q RPs above, which one is the best?
- ✓A special case, X_t is independent of all the RPs, i.e., it is a white noise, how good is it?

The measures to answer the above questions as follows:

- ✓ANNOVA
- ✓AIC (Akaike’s Information Criterion)
- ✓BIC (Bayesian Information Criterion)

ANOVA: F-test

- **Which model is the best, i.e., balance between the model complexity (more dependent RPs $Z_{t,i}$) and errors from the model regression?**

Table 2.1. *Analysis of Variance for Regression*

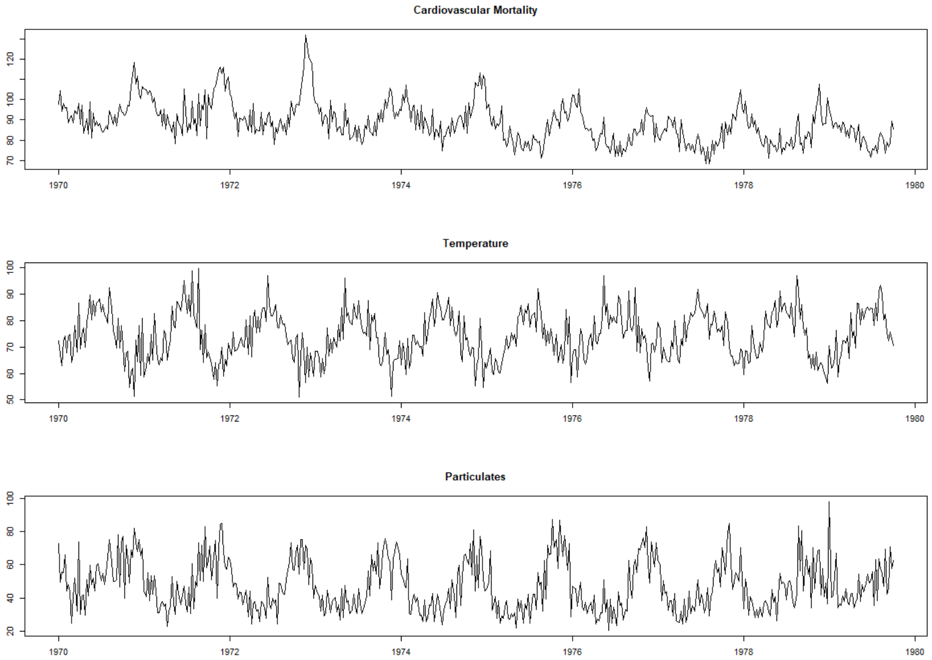
Source	df	Sum of Squares	Mean Square	F
$z_{t,r+1:q}$	$q-r$	$SSR = SSE_r - SSE$	$MSR = SSR/(q-r)$	$F = \frac{MSR}{MSE}$
Error	$n-(q+1)$	SSE	$MSE = SSE/(n-q-1)$	

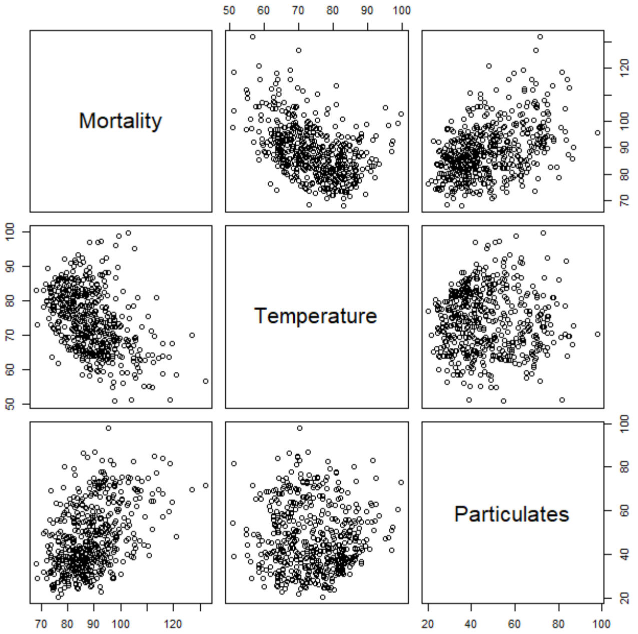
- SSE: Sum of Squared Errors $SSE = \sum_{t=1}^{n}(x_t - \hat{x}_t)^2$, where $\hat{x}_t$ is estimation from the model with *q* dependent random process, i.e., *q+1* degree of freedom including intercept
- SSE_r: sum of squared errors for the simplified model with *r* dependent RP in the model
- $F(q-r, n-(q+1))$ distribution to compare two models with *r* and *q* DOF

Example – goodness of fit (1)

Mortality in terms of pollution and temperature

- Let M_t, T_t and P_t represent the mortality, temperature, and pollution levels in terms of time t





Example – goodness of fit (2)

Mortality in terms of pollution and temperature

- Let M_t, T_t and P_t represent the mortality, temperature, and pollution levels in terms of time t
- According to the scatter plots that present their correlation with the mortality, we propose to use the following models

$$M_t = \beta_0 + \beta_1 t + w_t \tag{2.18}$$
$$M_t = \beta_0 + \beta_1 t + \beta_2(T_t - T_\cdot) + w_t \tag{2.19}$$
$$M_t = \beta_0 + \beta_1 t + \beta_2(T_t - T_\cdot) + \beta_3(T_t - T_\cdot)^2 + w_t \tag{2.20}$$
$$M_t = \beta_0 + \beta_1 t + \beta_2(T_t - T_\cdot) + \beta_3(T_t - T_\cdot)^2 + \beta_4 P_t + w_t \tag{2.21}$$

Model	k	SSE	df	MSE	R^2	AIC	BIC
(2.18)	2	40,020	506	79.0	.21	5.38	5.40
(2.19)	3	31,413	505	62.2	.38	5.14	5.17
(2.20)	4	27,985	504	55.5	.45	5.03	5.07
(2.21)	5	20,508	503	40.8	.60	4.72	4.77

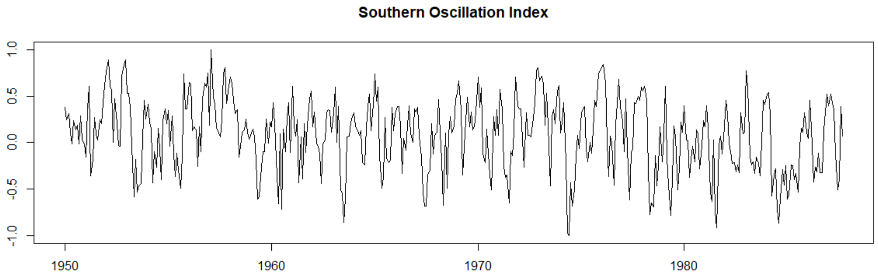
ANOVA: only test the full model VS the time dependent trend model

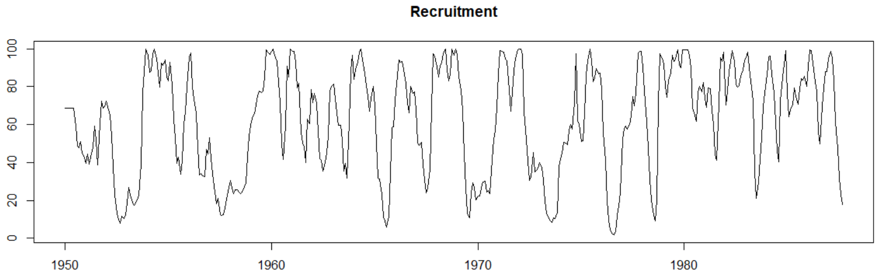
- ✓ H0: $\beta_2 = \beta_3 = \beta_4 = 0$
- ✓ q=4, r=1, n=508
- ✓ $F_{3,503} = \frac{(SSR/(q-r))}{SSE/(508-q-r)}$=160
- ✓ Reject the hypothesis, i.e., the probability to accept H0 is very low.

Example (2): regression with lagged RP

- Relationship: Southern Oscillation Index (SOI) and Recruitment (number of new fishes)
- There is a significant time lag between SOI (S_t) and Recruitment (R_t)
  - $R_t = \beta_0 + \beta_1 S_{t-6} + W_t$
  - Due to the time lag, one needs to align the data

```
> fish = ts.intersect(rec, soiL6=lag(soi,-6), dframe=TRUE)
> summary(fit1 <- lm(rec~soiL6, data=fish, na.action=NULL))

Call:
lm(formula = rec ~ soiL6, data = fish, na.action = NULL)

Residuals:
    Min      1Q  Median      3Q     Max
-65.187 -18.234   0.354  16.580  55.790

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   65.790      1.088   60.47   <2e-16 ***
soiL6        -44.283      2.781  -15.92   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 22.5 on 445 degrees of freedom
Multiple R-squared:  0.3629,    Adjusted R-squared:  0.3615
F-statistic: 253.5 on 1 and 445 DF,  p-value: < 2.2e-16
```





AIC and BIC

- **Akaike's Information Criterion (AIC)**
  - ✓$AIC = log\hat{\sigma}_k^2 + \frac{n+2k}{n}$
  - ✓$\hat{\sigma}_k^2 = \frac{SSE(k)}{n}$, where *k* represents number of parameters(RPs) in the model, and *n* denotes the number of sample data points
  - ✓The value of *k* that gives the minimum value of AIC represent the best model

- **Bayesian Information Criterion (BIC)**
  - ✓$BIC = log\hat{\sigma}_k^2 + \frac{klogn}{n}$
  - ✓The terms in BIC is the same as in AIC. And minimum value of BIC gives the best model

The penalty term in BIC is much larger than in AIC. It means that BIC tends to choose smooth models, i.e., models with less parameters inside.

Outline

- Typical models for the time series regression/analysis
- Measures for goodness of fit
- Model/data exploration (AR models)
- Smooth data (Moving Average) models

Exploratory data to find models

Data exploratory analysis:

- ✓The important measure of a RP (time series) is the autocorrelation with/without lag
- ✓Therefore, time series data should be made **stationary** and not too **fluctuated**
- ✓It means the mean and autocovariance should fulfil stationary conditions
- ✓But actual time series of data are often non-stationary with certain trend

Examples of time series with trend

- ✓Price of certain product or salaries (increase trend due to inflation)
- ✓Global temperature due to climate change
- ✓Stock market index

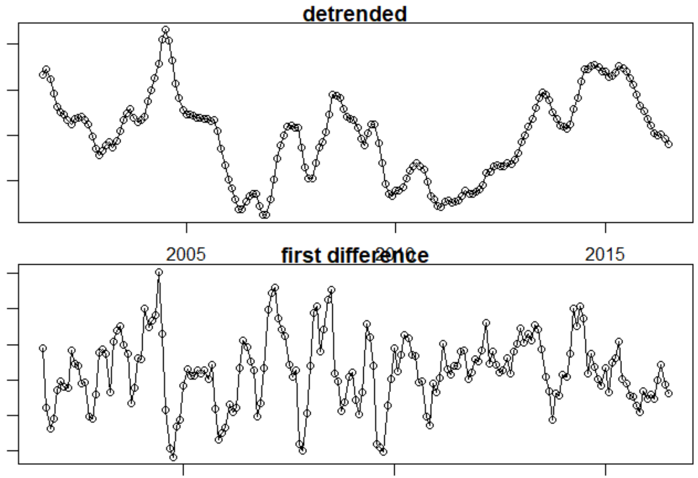
Ways to take away the trend and fluctuation

- ✓Detrend
- ✓Difference of various orders
- ✓Moving average

Nonstationary a time series→ stationary

A non-stationary time series with trend, for example, the chicken price with the model $X_t = \beta_0 + \beta_1 f(t) + W_t$, where f(t)=t, there are two ways to take away the trend as:

- **Detrend the nonstationary process**
  - ✓In the chicken price model, the mean trend is regressed as $\mu_t = \beta_0 + \beta_1 f(t)$
  - ✓After the detrend, new data becomes $Y_t = X_t - \mu_t = W_t$, a stationary white noise
- **Difference the nonstationary process**
  - ✓Let write $X_t = \mu_t + W_t$, where $\mu_t = \beta_0 + \beta_1 f(t)$ with trend
  - ✓Difference (1st order): $X_t - X_{t-1} = \beta_1 + W_t - W_{t-1}$
  - ✓It is easy to prove the stationarity, i.e., $Z_t = X_t - X_{t-1}$ with constant mean and ACF only dependent on time lag
  - ✓If still not stationary, i.e., with trend, higher order difference

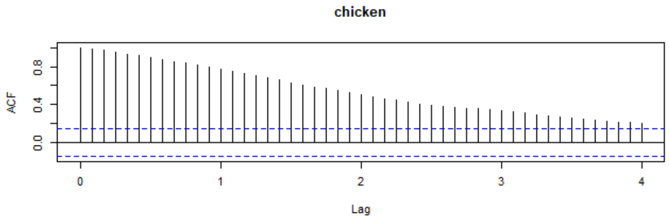


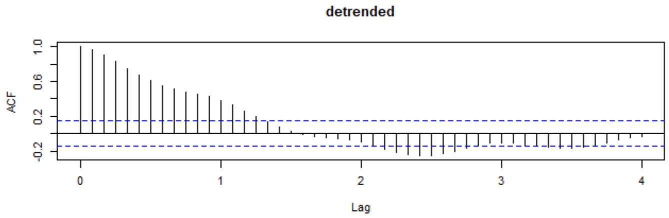
Detrend VS difference

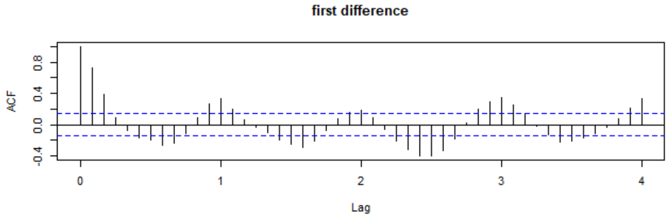
The time series after detrend and difference are difference even though they may be both stationary

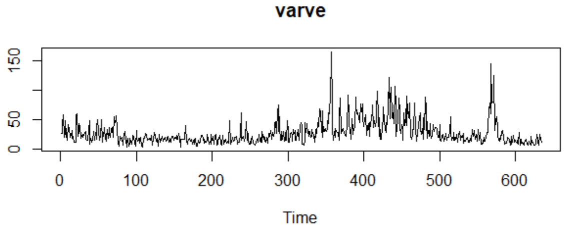
- **Cons and pros of difference**
  - ✓Difference: no parameters to estimate are needed
  - ✓Stationarity of differenced RP depends on the $W_t - W_{t-1}$
- **Alternative to remove the nonstationary: transformation**

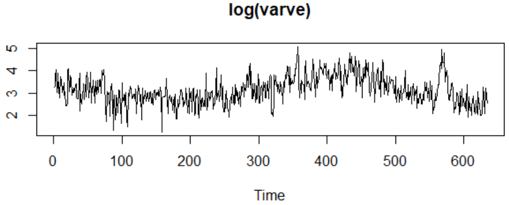
$$y_t = \log x_t, \qquad y_t = \begin{cases} (x_t^\lambda - 1)/\lambda & \lambda \neq 0, \\ \log x_t & \lambda = 0. \end{cases}$$









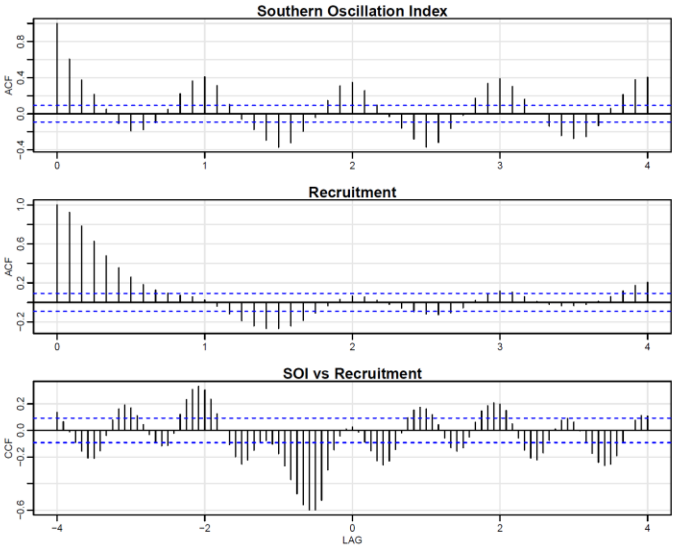


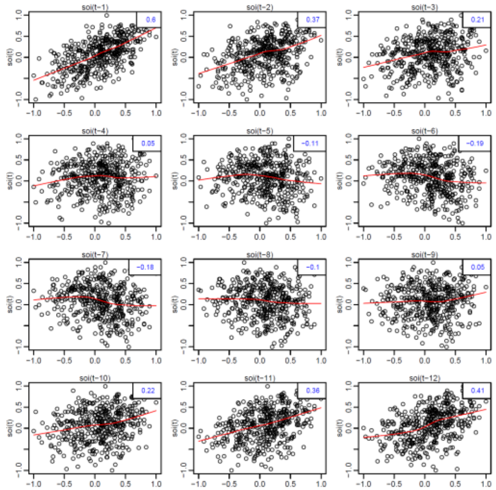
ACF: backshift operator and time lag

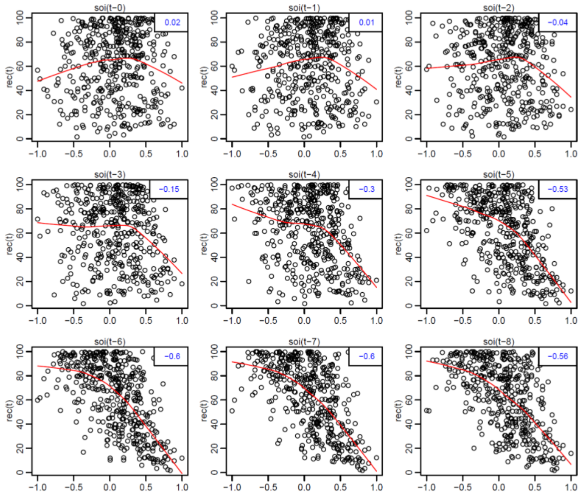
- Let X_t denote a random process
- The backshift operator is defined as:
  - ✓$BX_t = X_{t-1}$, where *B* means backshift
  - ✓$B^2X_t = B(BX_t) = BX_{t-1} = X_{t-2}$, thus
  - ✓$B^kX_t = X_{t-k}$
- The inverse of the backshift operator B^{-1} is known as the forward-shift operator as
  - ✓$X_t = B^{-1}BX_t = B^{-1}X_{t-1}$
  - ✓Then the difference operator: $\nabla X_t = X_t - X_{t-1} = (1-B)X_t$
  - ✓Second order of difference: $\nabla^2 X_t = (1-B)^2 X_t = (1-2B-B^2)X_t = X_t - 2X_{t-1} + X_{t-2}$
  - ✓Similar definition: $\nabla^2 X_t = \nabla(X_t - X_{t-1}) = (X_t - X_{t-1}) - (X_{t-1} - X_{t-2}) = X_t - 2X_{t-1} + X_{t-2}$
- Differences of order *d* can be defined as:
  - ✓$\nabla^d = (1-B)^d$
  - ✓When d=1, it is often dropped from the notation.

Backshift correlation by scatterplot: SOI vs Recruitment

- We know the amount of harvested fish (recruitment) will depends on the ocean index (SOI), but there might be some lag between SOI and recruitment
- First, the SOI itself maybe autocorrelated, $S,\ S_{t-1,} S_{t-2,} S_{t-3}$, ...
- Second the Recruitment may be cross-correlated with lagged SOI, i.e., S(t-h)





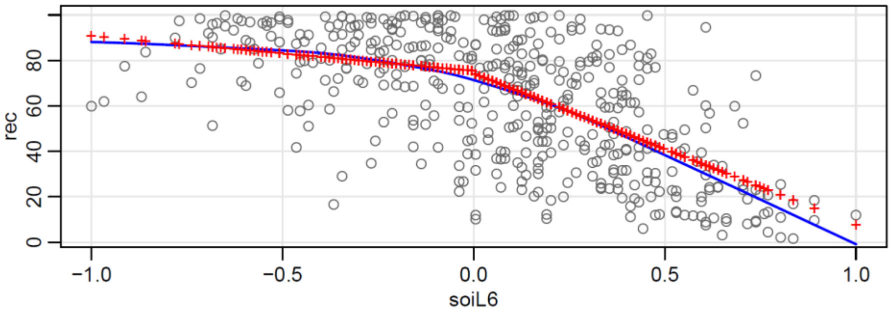


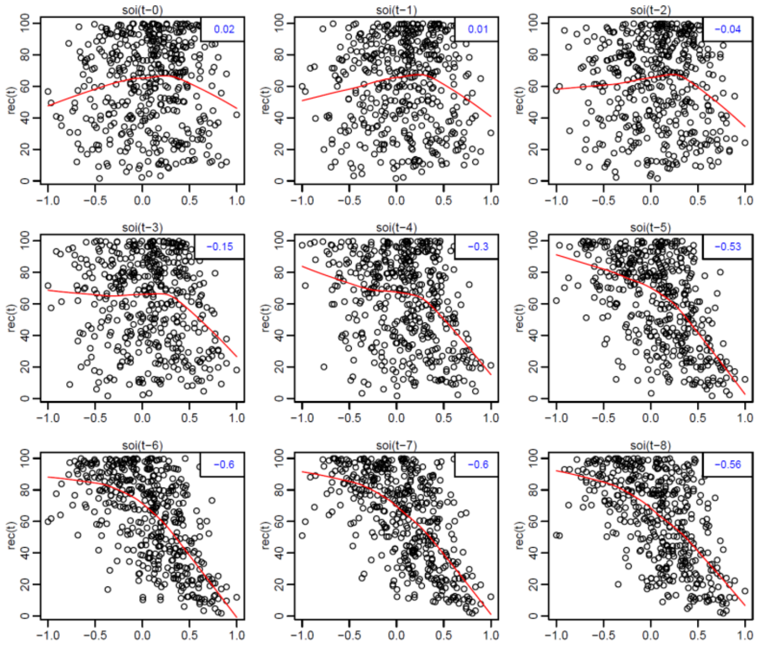
Backshift regression: SOI vs Recruitment

From previous scatterplot, we can see:

- ✓ Recruitment is most correlated to SOI(t-6)
- ✓ There is a separation zone at $S_{t-6} = 0$
- ✓ Divide the model into two piecewise models

$$R_t = \begin{cases} \beta_0 + \beta_1 S_{t-6} + w_t & \text{if } S_{t-6} < 0, \\ (\beta_0 + \beta_2) + (\beta_1 + \beta_3) S_{t-6} + w_t & \text{if } S_{t-6} \geq 0. \end{cases}$$





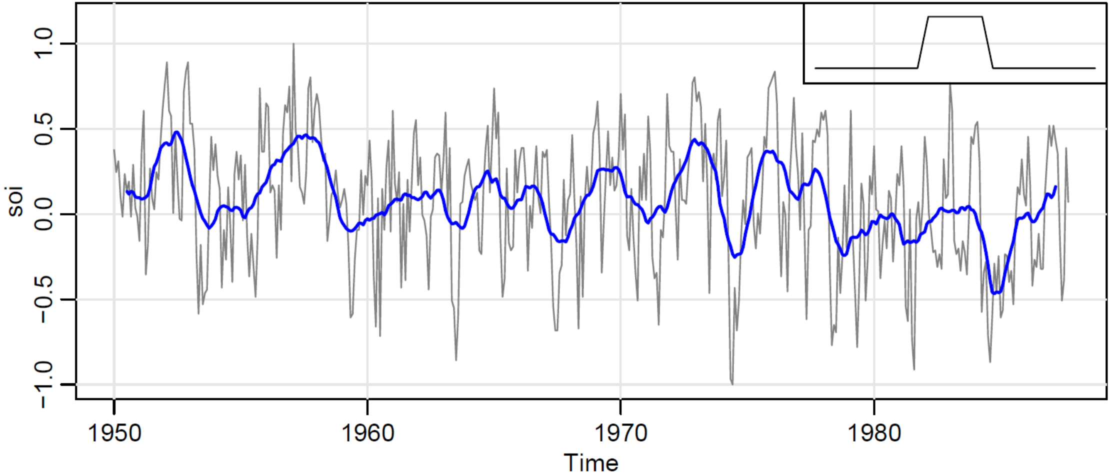
Outline

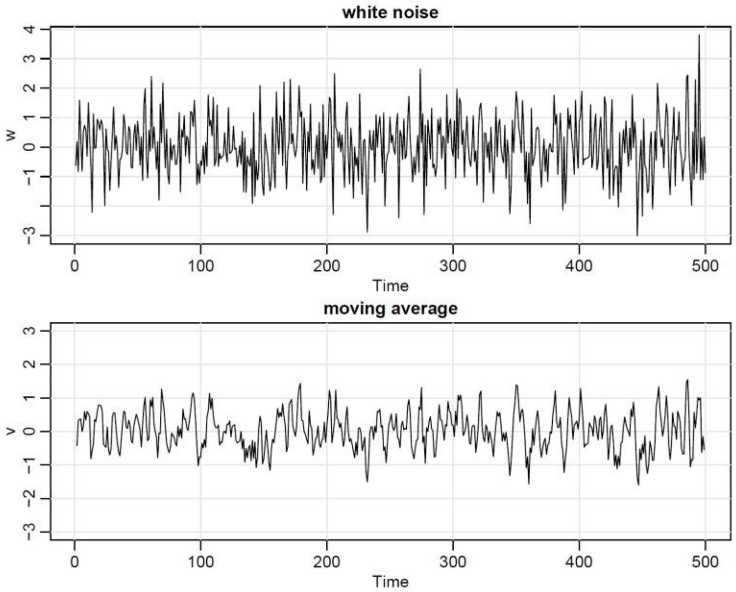
- Typical models for the time series regression/analysis
- Measures for goodness of fit
- Model/data exploration (AR models)
- Smooth data (Moving Average) models

Smoothing the time series (MA)

- For practical problems, often the data will look too noise (fluctuated): the white noise below
- The most simple way to smooth the data is by averaging over an interval

$$m_t = \sum_{j=-k}^{k} a_j x_{t-j},$$





Different ways of smoothing (1)

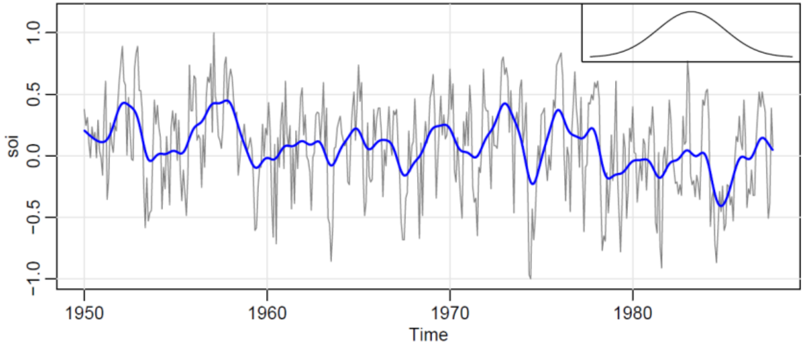
- Moving average smooth

  ✓ $$m_t = \sum_{j=-k}^{k} a_j x_{t-j}, \quad \text{where } a_j = \frac{1}{2k}, j = -k, -k+1, \ldots, k-1, k+1$$

- Kernel smoothing

$$m_t = \sum_{i=1}^{n} w_i(t) x_i,$$

$$w_i(t) = K\left(\frac{t-i}{b}\right) \Big/ \sum_{j=1}^{n} K\left(\frac{t-j}{b}\right)$$

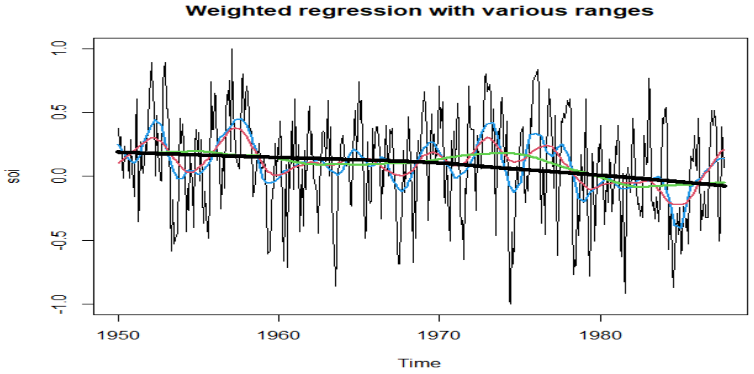


✓Different kernels can be used, such as the normal kernel, uniformly distributed kernel (MA)…

$$K(z) = \frac{1}{\sqrt{2\pi}} \exp(-z^2/2)$$

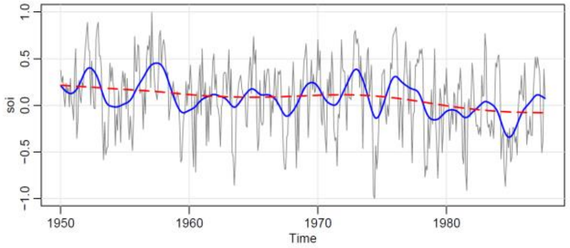
Different ways of smoothing (2)

- Lowess
  - ✓Similar as kernel smoothing
  - ✓But for k-nearest neighbours regression
  - ✓The regression uses a robust weighted reg.
  - ✓The regression to predict X_t,
  - ✓The predicted X_t is used to get the mean m_t



- Smooth splines (similar as the regression to give expected values in terms of RVs, here *t*)
  - ✓We get $X_t = m_t + W_t$, where $m_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3$ from cubic polynomial regression
  - ✓For the smoothness of the fit, we often divide time into various stages (knots), and then fit the cubic
  - ✓For the continuous along the knots, we set up the smoothness function (loss function)

$$\sum_{t=1}^{n} [x_t - m_t]^2 + \lambda \int (m_t'')^2 \, dt,$$



AVANCEZ

1829

CHALMERS

UNIVERSITY OF TECHNOLOGY