732A62 Time Series Analysis
Division of Statistics and Machine Learning
Department of Computer and Information Science

# Computer lab A

## Instructions

- The lab is assumed to be done in groups.
- Create a report to the lab solutions in PDF.
- Be concise and do not include unnecessary printouts and figures produced by the software and not required in the assignments.
- **Include all your codes as an appendix into your report.**
- A typical lab report should 2-4 pages of text plus some number of figures plus appendix with codes.
- The group lab report should be submitted via LISAM before the deadline specified in LISAM.
- Use 12345 as a random seed everywhere where the result of the simulation differs with the run unless stated otherwise.

## Assignment 1. Computations with simulated data

a) Generate two time series $x_t = -0.8x_{t-2} + w_t$, where $x_0 = x_1 = 0$ and $x_t = \cos\left(\frac{2\pi t}{5}\right)$ with 100 observations each. Apply a smoothing filter $v_t = 0.2(x_t + x_{t-1} + x_{t-2}+x_{t-3} + x_{t-4})$ to these two series and compare how the filter has affected them.

b) Consider time series $x_t - 4x_{t-1} + 2x_{t-2} + x_{t-5} = w_t + 3w_{t-2} + w_{t-4} - 4w_{t-6}$. Write an appropriate R code to investigate whether this time series is casual and invertible.

c) Use built-in R functions to simulate 100 observations from the process $x_t + \frac{3}{4}x_{t-1} = w_t - \frac{1}{9}w_{t-2}$, compute sample ACF and theoretical ACF, use seed 54321. Compare the ACF plots.

## Assignment 2. Visualization, detrending and residual analysis of Rhine data.

The data set **Rhine.csv** contains monthly concentrations of total nitrogen in the Rhine River in the period 1989-2002.

a) Import the data to R, convert it appropriately to *ts* object (use function *ts()*) and explore it by plotting the time series, creating scatter plots of $x_t$ against $x_{t-1}, \dots x_{t-12}$ . Analyze the time series plot and the scatter plots: Are there any trends, linear or seasonal, in the time series? When during the year is the concentration highest? Are there any special patterns in the data or scatterplots? Does the variance seem to change over time? Which variables in the scatterplots seem to have a significant relation to each other?

b) Eliminate the trend by fitting a linear model with respect to $t$ to the time series. Is there a significant time trend? Look at the residual pattern and the sample ACF of the residuals and comment how this pattern might be related to seasonality of the series.

c) Eliminate the trend by fitting a kernel smoother with respect to $t$ to the time series (choose a reasonable bandwidth yourself so the fit looks reasonable). Analyze the residual pattern and the sample ACF of the residuals and compare it to the ACF from step b). Conclusions? Do residuals seem to represent a stationary series?

d) Eliminate the trend by fitting the following so-called seasonal means model:

$$x_t = \alpha_0 + \alpha_1 t + \beta_1 I(month = 2) + \dots + \beta_{12} I(month = 12) + w_t,$$

where $I(x) = 1$ if x is true and 0 otherwise. Fitting of this model will require you to augment data with a categorical variable showing the current month, and then fitting a usual linear regression. Analyze the residual pattern and the ACF of residuals.

e) Perform stepwise variable selection in model from step d). Which model gives you the lowest AIC value? Which variables are left in the model?

# Assignment 3. Analysis of oil and gas time series.

Weekly time series *oil* and *gas* present in the package *astsa* show the oil prices in dollars per barrel and gas prices in cents per dollar.

a) Plot the given time series in the same graph. Do they look like stationary series? Do the processes seem to be related to each other? Motivate your answer.
b) Apply log-transform to the time series and plot the transformed data. In what respect did this transformation made the data easier for the analysis?
c) To eliminate trend, compute the first difference of the transformed data, plot the detrended series, check their ACFs and analyze the obtained plots. Denote the data obtained here as $x_t$ (oil) and $y_t$ (gas).
d) Exhibit scatterplots of $x_t$ and $y_t$ for up to three weeks of lead time of $x_t$; include a nonparametric smoother in each plot and comment the results: are there outliers? Are the relationships linear? Are there changes in the trend?
e) Fit the following model: $y_t = \alpha_0 + \alpha_1 I(x_t > 0) + \beta_1 x_t + \beta_2 x_{t-1} + w_t$ and check which coefficients seem to be significant. How can this be interpreted? Analyze the residual pattern and the ACF of the residuals.