

Exercises with the CGMS Statistical Toolbox

Allard de Wit & Raymond van der Wijngaart



Installing the CGMS Statistical Toolbox

- WinXP users: Run the CST_setup_1_5WinXP.exe
- Windows 7/Vista users:
 - Unzip CST_Win7.zip to a tmp folder (C:\TEMP)
 - Copy Folder CGMSStatTool\ to C:\Program Files (x86)\Alterra
 - Copy Folder Alterra\ to C:\ProgramData\
 - Create ODBC link “CGMS_Local_Database” pointing to C:\ProgramData\Alterra\CgmsStatTool\cgms.mdb

Exercises with the CGMS Statistical Toolbox

Part 1 Trend model selection

E1: Wheat – Germany (Deutschland)

- Analyze the trend model 1991-2004
- What are the P-values for the linear and quadratic trends
- Which trend is selected and why
- What happens when the threshold for significance is lowered to 0.01
- Is the trend model significant over a shorter time-space as well?

Grain maize - France

- Select period 1990-2004
- What are the p-values of the linear and quadratic trends
- What happens when you exclude the year 2003 (use “view data/exclude more years”)
- What happens when you select the period 1991-2004
- Would you select the no trend, linear trend or the quadratic trend model?

Wheat – Poland (Polska)

- Select the period 1990-2004
- What trend is selected by default
- Find out how stable the selected trend is, by changing the start or end years
- What would you choose: none, linear or quadratic.

Exercises with the CGMS Statistical Toolbox

Part 2

Yield forecasting using regression models

Wheat in Germany (1)

- CSI trend settings:
 - Period 1991-2004
 - Select the default trend model (linear)
 - Select decade III in September (end of season in Europe)
- Go regression tab and add all indicators to the “Free indicators”
- Look at the correlation matrix.
 - Which indicators have the highest correlation, is this logical?
 - Which indicators have the highest correlation with yield?

Wheat in Germany (2)

- CST regression analysis settings:
 - Select “Single free indicator”
 - default settings for ordering and summary statistics.
 - Click “Next”
- What is the best model and how does the RMSE for prediction compare to the “None” model
- What is the second best model, can you explain this from the correlation matrix?
- Why are the lower two models not good.

Winter-barley in Danmark (1)

- CST trend settings:
 - Period 1990-2004
 - Select the default trend model (no trend)
 - Select decade III in September (end of season in Europe)
- Go regression tab and add all indicators to the “Free indicators”
- Look at the correlation matrix. Which indicators have the highest correlation, is this logical?
- Click “Next”

Winter-barley in Danmark (2)

- CST regression analysis settings:
 - Select “Single free indicator”
 - default settings for ordering and summary statistics.
 - Mark all indicators as “positive sign” and click “Next”
- What is the best model, is it significant
- What is the difference with the None model according to the “RMSE for prediction”
- Would you choose this model or the None model?

Grain maize in Italy (1)

- CST trend settings:
 - Period 1994-2004
 - Select the default trend model (quadratic)
 - Select decade III in October (end of season in Europe)
- Go regression tab and add all indicators to the “Free indicators”
- Click “Next”

Grain maize in Italy (1)

- CST regression analysis settings:
 - Select “Single free indicator”
 - default settings for ordering and summary statistics.
 - Click “Next”
- What is the best model and how does the RMSE for prediction compare to the “None” model
- Although all T-values for the model are significant, is this model a realistic one?

Grain maize in France (1)

- CST trend settings:
 - Period 1991-2004
 - Select the default trend model (No trend)
 - Select decade III in October (end of season in Europe)
- Go regression tab and add all indicators to the “Free indicators”
- Click “Next”

Grain maize in Italy (1)

- CST regression analysis settings:
 - Select “Single free indicator”
 - default settings for ordering and summary statistics.
 - Click “Next”
- Write down the R^2 and T-values of the best model?
- Now go back to time trend analysis, force a linear trend and redo the regression analysis
- Compare the R^2 and T-values with the ones of the previous analysis. Can you explain this?