Lund University
Centre for Mathematical Sciences
Mathematical Statistics

FMS051/MASM17
2017-11-26/AJ

Project

Prediction of weather may be done globally or locally, based on more or less available information. In this project, you are to model and predict the outdoor temperature in one place in southern Sweden. The global part of the model will be fetched from SMHI (the Swedish weather bureau); together with local temperature measurements, we will then use this to do a local temperature prediction.

The measurements that form the basis for the project are taken during the end of 1993, the whole 1994, and the beginning of 1995 in Sturup and Växjö by SMHI, and in Svedala, Alvesta, and Harplinge by Sydkraft. Furthermore, a couple of times a day, SMHI produces predictions for the outdoor temperature in three hour intervals in Sturup and Växjö. In order to get hourly predictions, production personal at Sydkraft make linear interpolation between the values gotten from SMHI. Hence, the prediction of the outdoor temperature we have at our disposal for a certain hour is the last one that was produced for that particular hour. That means that we cannot give a common prediction horizon for all predictions, but the prediction that is registered is computed with a prediction horizon between 1 and approximatively 10 hours. Your model can be based solely on the measured outdoor temperature, but by using the predictions from SMHI for the same place, or a place close by, as an external signal, you can hope for a better modeling of the temperature, and possible also for a better local temperature prediction.

The task to be fulfilled in the project consists in studying parts of a set of data containing hourly measurements of outdoor air temperature and predictions thereof in Sturup, Svedala, Växjö, Harplinge, or Alvesta. You are supposed to *model* and *predict* the air temperature during some season.

A. Start with the measurements of the temperature. Choose ten weeks for the modeling part, clearly indicating which location and time period you have selected. Beware that there might be missing samples, with some time periods missing more samples than other. How can you deal with such missing samples? Do you need to be concerned with outliers? Construct a model for the temperature data, showing the ACF and PACF for your modeling residual (show lags up to 50). Are the residuals white? Show the one-step and the 8-step predictions as compared to the measured values, and show the ACF for the corresponding prediction errors. Do the ACFs behave as desired? Comment on your results and specify the variance of the three residuals. (*20 marks*)

B. Investigate if your model and the predictions of the temperature are improved if you consider the prognosis made by SMHI as an external *known* signal. Show the ACF and PACF for your model residuals, as well as the ACF for the one-step and 8-step predictions for the resulting model. Also show the one-step and the 8-step predictions as compared to the measured values. Comment on your results and specify the variance of the different residuals. (*20 marks*)

C. Can you improve your model by recursively estimating some, or all, the parameters of your models? Show the results both with and without the use of an external input. Compare your predictions for the models in part A, B, and C during the period 20-30 weeks after the end of your modelling period. Comment on your results and specify the variance of the different residuals. (*20 marks*)

D. Can you build a model that includes the predicted temperatures as an input as well as the measured temperature at another location to make a better model? In which sense is it better? Compare your predictions during the period 20-30 weeks after the end of your modelling period. Comment on your results and specify the variance of the different residuals. (*optional, up to 5 bonus marks*)

The project report should contain a summary of the steps you have taken to get your models. It should be written in a concise way with figures and plots being clearly explained. Resulting models and predictors should be given in detail.

The project can be done in groups of maximally two students. Discussions on the project with anyone other than the teaching staff is prohibited and it is expected that all students refrain from this. During the oral presentation, a random selection of groups will be asked to present their projects. For the selected groups, one randomly chosen group member will be asked to present the entire project.

Data

The data files which you are going to use can be found on the homepage for the course. The following data sets are available:

File	Comments
ptstu93.mat	Prediction of the temperature in Sturup during 1993 (SMHI)
ptstu94.mat	Prediction of the temperature in Sturup during 1994 (SMHI)
ptstu95.mat	Prediction of the temperature in Sturup during 1995 (SMHI)
ptvxo93.mat	Prediction of the temperature in Växjö during 1993 (SMHI)
ptvxo94.mat	Prediction of the temperature in Växjö during 1994 (SMHI)
ptvxo95.mat	Prediction of the temperature in Växjö during 1995 (SMHI)
tstu93.mat	Measured temperature in Sturup during 1993 (SMHI)
tstu94.mat	Measured temperature in Sturup during 1994 (SMHI)
tstu95.mat	Measured temperature in Sturup during 1995 (SMHI)
tvxo93.mat	Measured temperature in Växjö during 1993 (SMHI)
tvxo94.mat	Measured temperature in Växjö during 1994 (SMHI)
tvxo95.mat	Measured temperature in Växjö during 1995 (SMHI)
utempAva_9395.dat	Measured temperature in Alvesta during 1993-95 (Sydkraft)
utempHar_9395.dat	Measured temperature in Harplinge during 1993-95 (Sydkraft)
utempSla_9395.dat	Measured temperature in Svedala during 1993-95 (Sydkraft)
tid93.mat	Time schedule for the measurements for 1993
tid94.mat	Time schedule for the measurements for 1994
tid95.mat	Time schedule for the measurements for 1995

The data files from SMHI contain only one column with temperature data starting on the first hour of the year. As mentioned above, we have 8 values per twenty-four hours, which are interpolated linearly to hourly measurements. The accuracy is 1 ^{o}C.

The data files from Sydkraft are registered every hour apart from the last hour of the day, which is represented by 0. The accuracy is 0.3 ^{o}C. These files have three columns: day, hour, temperature.

The time schedule files have five columns: number of the hour, year, number of the week, day of the week, hour.