02417: Time Series Analysis
Rune Grønborg Junker — DTU Compute
Spring 2023 — Technical University of Denmark

Assignment 4: Kalman Filter

Sensor platforms that can be deployed for extended periods of time and preferably sending real time measurements is a great tool. Some sensors requires regular maintenance which may be rather costly due to the time spent. This assignment will focus on a bouy that is located in the water near Kulhuse - in opening of Roskilde Fjord.
Every half an hour the sensor platform records the following:

- `Temp` is the water temperature [degC]
- `Sal` is the water salinity (Amount of salt) [PSU] = [g/kg]
- `Depth` is the depth of the sensor platform [m]
- `pH` is the water pH
- `Chl` is the water cholorphyll concentration [mg/L]
- `ODOsat` is the percent dissolved oxygen saturation [%]
- `ODO` is the dissolved oxygen [mg/L]
- `Battery` is the voltage of the battery in the sensor platform [V]
- `DateTime` is the time stamp

Measurements from a period of time is made available in the file `A4_Kulhuse.csv`.
This assignment will only look into the salinity and the dissolved oxygen. The focus of the first 5 questions is on filtering the salinity measurements, while in the last question we will explore how a more advanced state space model for the dissolved oxygen can be deduced.

Question 4.1: Presenting data Load the data and present it by making some plots for salinity and dissolved oxygen.

Comment on what you see.

Question 4.2: Random walk state-space model of salinity At first it is assumed that the salinity can be described by a random walk process that is observed at equidistant time points.

Write the system as a state-space model according to section 10.1. It is sufficient if you define the relevant matrices.

Question 4.3: Pure Kalman filter Filter the salinity with a Kalman filter. You are not required to write your own implementation, but as you will see in Question 4.4, you will have to do some manual adjustments to the Kalman filter and so it might be easier to simply make your own implementation.

Use the first observation as initial value and use the system variance as initial variance. Fix the system variance to 0.01 and the observation variance to 0.005.

1. Plot the one step predictions along the data. Do include 95% PI.

2. Plot the standardized one step prediction errors. (Prediction errors normalized with the standard error of the prediction.)
3. Repeat the two plots with the same content but zooming into observations 800 to 950.
4. Report the values that defines the final state of the filter (at observation 5000).

Comment on what you see.

Question 4.4: Skipping outliers when filtering Some observations are very unlikely. Make your own implementation of the Kalman filter. The implementation should treat observations that are more than six standard deviations away from the prediction as missing (as in don't perform a reconstruction step at these points in time).

Filter the data with your version of the filter.

1. Present the 1-step predictions for index 800 to 950 as in question 4.3.4.
2. Report the indexes for the first five detected outliers (observations that are skipped because they are more than six standard deviations away).
3. Do also report the number of observations that are skipped.
4. Report the values that defines the final state of the filter (at observation 5000).

Comment on your results

Question 4.5: Optimizing the variances Maximum likelihood estimation of the two variance parameters.

1. What is a sensible lower bound for the observation variance?
2. Find the ML estimates of the two parameters using the first 800 observations.
3. Filter the data with the optimal parameters.
4. Plot as in Q4.3.4 for observations 800 through 950.
5. Report the values that defines the final state of the filter (at observation 5000).

Comment on your results

Question 4.6: Model for dissolved oxygen The dissolved oxygen concentration varies during the day due to oxygen production (Called primary production) from chlorophyll and sunlight and oxygen consumption from all biological matter (Called respiration). Furthermore, there is exchange of oxygen with the atmosphere so that the dissolved oxygen concentration approaches the saturation concentration.

The saturation concentration for dissolved oxygen depends on the temperature and the salinity. Here you can treat it as a known input.

Specify a state-space model for dissolved oxygen that includes primary production as a linear function of the sun intensity and exchange of oxygen with the atmosphere. The model should include respiration as a random walk which is to be used to identify the amount of biomass.

Use the following naming of variables in your model formulation:

DO_t for dissolved oxygen at time t
I_t for intensity of sunlight at time t
$DOsat_t$ for saturation concentration of dissolved oxygen at time t
R_t for respiration at time t