Delft University of Technology

Data assimilation
WI4475

Assignment: Tidal wave

Authors:
Martin Verlaan and Arnold Heemink

March 26, 2023

Delft University of Technology

Abstract

The Western Scheldt estuary in the south-western part of the Netherlands contains an important shipping channel to Vlissingen and Antwerp. At the same time, the estuary is surrounded by low polder areas, that potentially are at risk of flooding during severe storms in combination with spring tides. Here we will model the hydrodynamics of these tides and storm surges with the linearized shallow water equations. Along the coast of the Western Scheldt the waterlevel is measured in a number of tide gages. In this project, we only consider a limited number of them. The data provided with the project files are the actual observed values as made available by Rijkswaterstaat (`https://waterinfo.rws.nl/#!/thema/Waterveiligheid/`) The period studied in this project is the storm Xaver of December 6 2013, also known as the Sinterklaasstorm in the Netherlands (`https://en.wikipedia.org/wiki/Cyclone_Xaver`). The peak waterlevel in Vlissingen was second highest one over the last century, only second to the storm of 1953.

This description of the wave1d project for WI4475 should be accompanied by a zip file with measurements and program files in Python and Julia. The choice of programming language is up to you. We advice to work in pairs to share the workload, exchange ideas and complement each-other. To finish the project you should together write a report describing what you did, the results and conclusions. The report will be discussed individually in an oral exam.

1 Introduction

Figure 1: Western Scheldt estuary

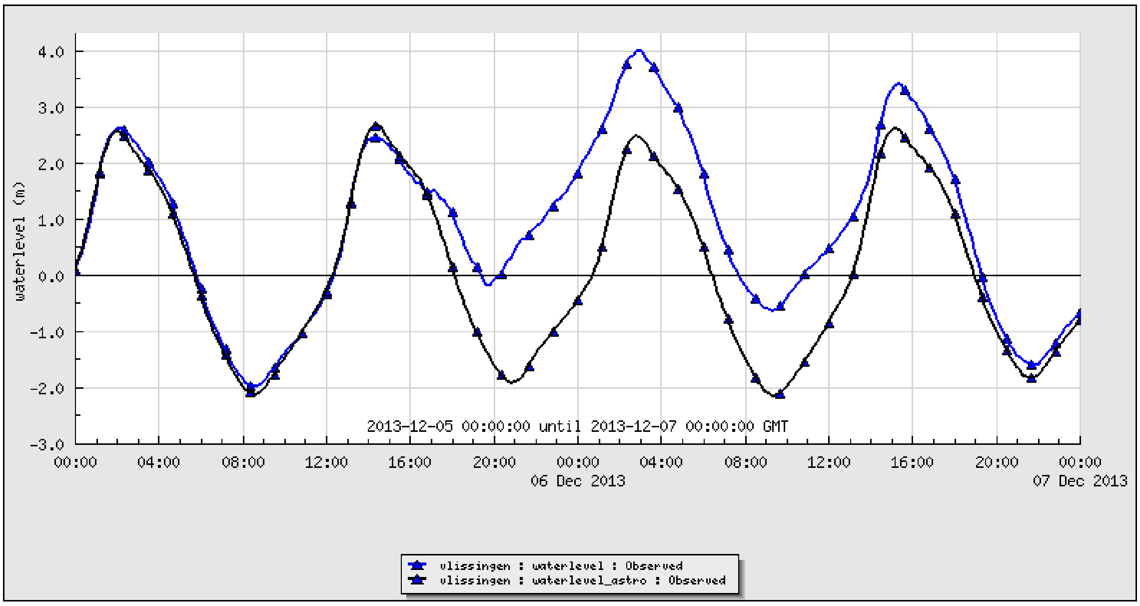

Figure 2: Time-series of measured waterlevel during the storm of Dec 5 2013.

The linearized one-dimensional shallow water equations, also known as De St Venant equations, are given by:

$$\frac{\partial h}{\partial t} + D\frac{\partial u}{\partial x} = 0 \tag{1}$$

$$\frac{\partial u}{\partial t} + g\frac{\partial h}{\partial x} + fu = 0 \tag{2}$$

More about the background can be found at https://en.wikipedia.org/wiki/Shallow_water_equations.

Question 1

Show that the linearized shallow water equations are equivalent to a wave equation, when the friction is neglected. Then compute (analytically) the propagation speed of the tidal wave. And finally substitute the values used in the model code and compare this to the propagation speed in the numerical model. How can you 'measure' the propagation speed in the numerical model?

Question 2

Let's consider the accuracy of the model or the model error. This very simple model is definitely not perfect, so there should be room for improvement. A common approach to improve the output of a model in comparison to observations is to improve the model itself (physics, numerics, input data). Write down three aspects how you think this specific model could be improved.

Identification of the uncertain parts of a model and modelling these as system error is an important part of any realistic data-assimilation application. One can also think of the system noise as the control variables of the data-assimilation problem, the knobs that a data-assimilation algorithm can use to create a better fit to the observations. Our main tool to achieve this is to check for the consistency of the error statistics. Let's look at this for this application.

Question 3

As a reference, we want to see how well the model is doing without data-assimilation. We will first look at calm weather conditions, for this purpose we can apply 'tidal-analysis' as a filter to keep only the tidal frequencies. The model files provided contain a filtered boundary condition and filtered observations (with 'tide' in filename). Run the model and quantify the accuracy of the model for this case in terms of the bias and RMSE. This gives us a first quantitative idea about the accuracy of the model. Name a few alternative statistics that you could you have used? Do these have benefits over bias and RMSE?

Question 4

Before applying an Ensemble Kalman Filter, we first need to model the uncertainty with a stochastic extension of the model. In this project, you'll add an AR(1) type stochastic forcing to the western boundary.

$$N(k+1) = \alpha N(k) + W(k) \tag{3}$$

Use $\alpha = exp(-dt/T)$ with dt the timestep of the model and $T = 6[hours]$. How can one select the standard deviation σ_W for $W(K)$ $N(0, \sigma_W^2)$ so that $\sigma_N = lim_{k\to\infty}\sqrt{E(N^2(k))}$ equals $0.2[m]$? Implement this additional forcing term in the model, run for an ensemble of size 50 and plot some relevant results. Is the uncertainty as modelled with the ensemble comparable to your results of question 3? Compute some statistics (for the ensemble spread) and explain the results.

Question 5

One key requirement for the use of the Kalman filter is the Markov property. Show how you fit the model in the standard system notation and explain that your implementation satisfies the Markov property.

Question 6

Implement an Ensemble Kalman filter and set up an identical twin experiment to test the implementation. Check some relevant statistics to verify your implementation. Why should the EnKF work perfectly (for a large ensemble) in this experiment? How can you verify this for your output?

Question 7

The Ensemble Kalman filter is supposed to get more accurate with a larger number of ensemble members. To what solution should the EnKF converge for a large ensemble in this case? Why? What is the expected convergence rate? How can we verify this? Perform an experiment to check this.

Question 8

For your answer to question 6, you selected an initial condition for the Kalman filter. What choice did you make there? Describe an alternative choice for the initial condition. Also perform an experiment with this alternative choice and study the impact on the results. Can you explain what happens, using the theory?

Question 9

Now, we move on to the real observation with the storm effects included (files with 'waterlevel' in the name). Note that the meassurements at the boundary (Cadzand) are missing. Modify your implementation to assimilate these observations. Make some plots to study the results. How accurate is the result in comparison to running the model without a Kalman filter. Quantify this with some statistics. Comment on your results, also in relation to the theory.

Question 10

In real life, the most important aspect of an application of data-assimilation is probably the accuracy of the forecasts (predicting into the future). Now try to mimic this situation in an experiment. Make some forecasts starting some hours before the peak of the storm and check how the accuracy of forecasts for the peak waterlevel depend on lead-time (=time of peak minus start of forecast). Explain the results and provide advice on how to further improve them.