Assignment on parameter estimation with sea-level rise as a theme.

Authors: C. Tiberius and S. Verhagen – Dept. of Geoscience and Remote Sensing
Version: 1.0 / Date: January 26th, 2023

Introduction

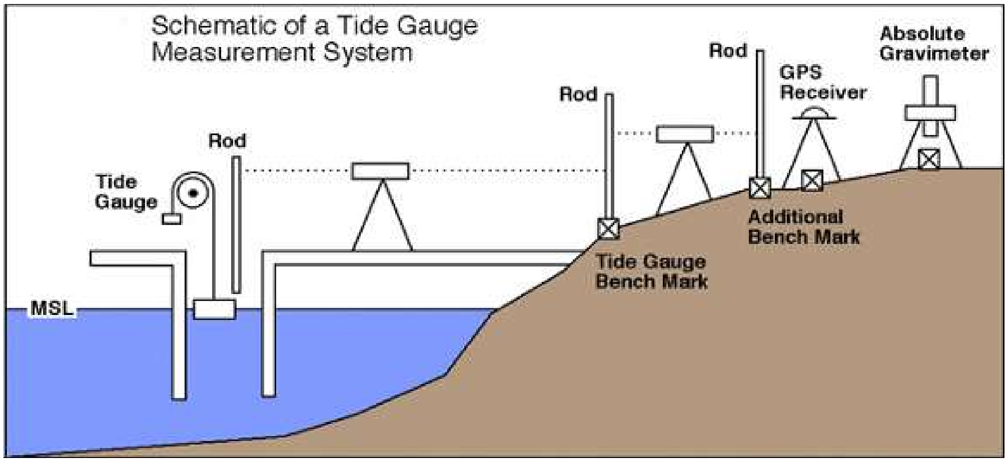
The sea level may vary due to ocean volume change and ocean mass change. Ocean volume change is caused by thermal expansion of the ocean – as ocean water warms up, it expands, and thereby the sea level rises. Change in the mass of the ocean is primarily due to melting of glaciers and ice sheets, with also some contribution from water stored in continental reservoirs and groundwater extraction.


At left: ice sheet in Greenland - photo by Christine Zenino - Wikimedia Commons; at right: Oosterschelde-kering coastal defense in Zeeland.

In this assignment you will work with actual measurements from a tide gauge, to assess whether the sea level is rising indeed, and estimate by how much. This knowledge is important, in particular for the Netherlands, to keep the coastal defense up to date, and future proof. The subject – in the context of climate change – has clearly a societal relevance, and every now and then spectacular sea-level figures are being reported and quoted in the media. Awareness of statistical significance is a great asset in this discussion.

Please prepare a report with your results (description of models used, answers, graphs), reasoning, derivation, and also hand-in your code.

Tide gauge records are the primary source of data to analyse sea-level changes over long periods of time (i.e. over 100-200 years). A tide gauge provides measurements of relative sea-level changes at a particular location on Earth (the word ‘relative’ refers to the fact that the water level can change, but also the land – the Earth’s crust – can move).



A tide gauge, typically installed near a pier, measures sea-level relative to a nearby geodetic benchmark. The height of the tide gauge is related to the benchmark on the coast typically through leveling. The classical tide gauge consists of an object floating in a so-called stilling well. Diagram taken from https://sealevel.colorado.edu/tide-gauge-sea-level.

Data sets

The sea water level is fluctuating – over a time scale of minutes – as a result of wind and waves. The water level inside the tide gauge is damped. Typically a tide gauge yields hourly readings. This is referred here to as Mean Sea Level (MSL) – it is the level of the water, without waves.

The water level is also fluctuating over longer time scales, as a result of tides. In the North Sea, the sea level exhibits a semi-diurnal tidal cycle – a high-low-high-low tide cycle is completed in about a 25 hour period, with a ± 2 meter amplitude. The Moon orbits the Earth in the same direction as the Earth rotates about its axis, and therefore it takes slightly more than one day – about 24 hours and 50 minutes (twice the so-called M_2 period of 12h25m) – to see the Moon at 'the same spot in the sky' again (as seen from Earth).

In this assignment we use monthly data, and these monthly values are obtained by averaging the hourly readings, in order to avoid also tidal effects. Global tide gauge data is archived by the Permanent Service for Mean Sea Level (PMSML): http://www.psmsl.org/ (PSMSL, 2013). The list of tide gauge stations is available at http://www.psmsl.org/data/obtaining/ From the list you can select the station of your choice, and then click on 'Download monthly mean sea level data' (for instance by right-clicking and choosing 'Save target as', you will get e.g. the `22.rlrdata` file in your folder.

For the Netherlands, the following stations are available (with long observation records):

- 9 Maassluis
- 20 Vlissingen
- 22 Hoek van Holland
- 23 Den Helder

- 24 Delfzijl
- 32 IJmuiden

Long record tide gauge stations along the Dutch coast (image from Google Earth).
Most of them have records back to about the year 1865.

A description of the data is given at http://www.psmsl.org/data/obtaining/notes.php The data file, with so-called semi-colon delimiters, typically looks like:

```
1900.0416;  6693; 0;000
1900.1250;  6609; 0;000
1900.2084;  6599; 0;000
1900.2916;  6593; 0;000
1900.3750;  6634; 0;000
1900.4584;  6637; 0;000
```

with the epoch (<u>time</u> label of the measurement) given <u>in decimal years</u> (and each time referring to the middle of the month; e.g. 1900.0416 refers to mid-January), the actual measurement (<u>sea-level height</u>) expressed <u>in millimeters</u> (e.g. 6693), and next two auxiliary values, which are not of interest with this assignment. There are 12 measurements per year (monthly measurements).

The `rlrdata`-file can be inspected (and edited if you wish) for instance in Notepad and Wordpad.

Reading the file in Python can be done using `loadtxt` from the numpy-library, e.g. `timedata, heightdata, missingdata = np.loadtxt(filename, delimiter=';', usecols=(0, 1, 2), unpack=True).`
At the start of the Python script, load the necessary packages, in this case:

```
import numpy as np
from matplotlib import pyplot as plt
```

to be able to do matrix calculations, vector and dot products, and to plot, respectively. A matrix inverse can be obtained via: `np.linalg.inv`.

Objectives

With this assignment you will set up a linear model to estimate both a trend and an annual signal (harmonic function) through a given data set of observations, you will apply linear least-squares parameter estimation, and you will interpret estimation results, apply error propagation, and also apply statistical hypothesis testing to validate the measurements.

Question 1

Estimate a linear trend from the sea-level data, for the station of your choice, based on all measurements recorded in the 20th century. So, the first observation is at 1900.0416 (for Jan 1900), and the last one at 1999.9584 (for Dec 1999).

a) Read-in these data, and make a graph of the recorded sea-level measurements, shown as a function of time.

Next, we will model the sea-level by means of a linear trend, i.e. we will fit a straight line through the recorded sea-level measurements.

b) Write down, symbolically, the model $y = Ax$ for estimating a linear trend (hence with an offset and a slope parameter) from the given sea-level measurements (in vector y). The offset parameter represents the (unknown/true) sea-level at time t_0, in this case 01-JAN-1900 (hence t_0=1900.0), and the slope parameter represents the (unknown/true) rate of change of the sea-level. How many measurements, *m*, are there? Specify them! How many (unknown) parameters, *n*, does the model contain? Specify them! Specify the dimensions m x n of the A-matrix, and specify the coefficients/elements of this matrix (i.e. state how you compute them numerically).

In Python the command `np.ones` may be useful in building the A-matrix. In Python you may, in addition, use `np.vstack` or `np.hstack` and/or `np.transpose`, and `len`.

c) Compute the (unweighted) least-squares solution for the unknown parameters, and report the estimated numerical values (also specify units).
d) Make a graph of the recorded sea-level measurements, as a function of time, and also plot the trend (line) you just estimated.
e) Based on the estimates for the sea-level at time t_0 and the change rate, compute the sea-level for the year 2100 (Jan 1st). By how much will the sea-level rise, compared to the year 1900 (Jan 1st)?

Question 2

A closer look at the graph of the previous question, reveals that the sea-level shows a yearly fluctuation (next to the trend, and noise). In this question, we will model the annual variation of the observed sea-level y by means of a harmonic function: $y(t) = A\cos(2\pi f_o t + \varphi)$ (for a short while, we'll forget about the trend here). The frequency f_o is given to be 1 (1 cycle per year), when time t is expressed in years. There are two unknown parameters, namely the amplitude A, and the phase φ (maximum sea-level does not necessarily occur on Jan 1st, that is why we have to insert also an unknown phase parameter). The phase is included in the argument of the cosine, and this is not convenient (this would lead to a non-linear parameter estimation problem). Therefore we use the goniometric identity $\cos(\alpha+\beta) = \cos\alpha\cos\beta - \sin\alpha\sin\beta$, and turn the harmonic function into $y(t) = A\cos\varphi\cos(2\pi t) - A\sin\varphi\sin(2\pi t)$. Taking again the linear trend into account as well, we now set $x_3 = Acos\varphi$ and $x_4 = -Asin\varphi$ (two unknown parameters are transformed into two other unknown parameters), as well as the 'starting' time t_o, so that $y(t) = x_1 + (t - t_o)x_2 + \cos(2\pi(t - t_o))\, x_3 + \sin(2\pi(t - t_o))\, x_4$. The result is that we have a model which is linear in the unknown parameters.

a) Compute the least-squares solution for the four unknown parameters in the above model, and report the estimated values (also specify units). You can use again the same data for the station of your choice (all measurements from the 20th century).
b) Parameters x_3 and x_4 are not directly meaningful to us. From the obtained estimates, retrieve the values for the amplitude A, and the phase φ. Report them, and specify units.
c) Can you interpret the phase shift? We estimated an annual cycle; in what period of the year is the sea-level at minimum, and when is it at maximum?
d) Can you think of, and report an explanation of the behaviour observed under c)?

For the remainder of this assignment we ignore the yearly fluctuation. We will be working with just the linear trend (offset and slope).

Question 3

The monthly tide gauge sea-level observable has a standard deviation of σ_{y_i}= 60 mm, for $i = 1, \dots, m$. The observables are uncorrelated.

a) Using the Best Linear Unbiased (or Minimum Variance) Estimator, compute the estimates for the offset and slope parameter.
b) Compute the variance matrix of this estimator, and also report the standard deviation of the offset and slope estimator.
c) Compute the correlation-coefficient between the offset and slope estimator, and provide an interpretation and explanation for this result.
d) With Question 1e) we extrapolated the sea-level to the year 2100. Using error propagation (variance propagation law) compute the standard deviation of this estimator for the sea-level in 2100.

Question 4

With the assumption of the functional model $E\left(\underline{y}\right) = Ax$, the variance matrix $D\left(\underline{y}\right) = Q_{yy}$, and that the observables are normally distributed, we can perform statistical hypothesis testing for the purpose of model validation (detection). Can you briefly describe how you would perform a goodness-of-fit test of the assumed model?

Question 5

Next we would like to <u>analyse</u> how the precision of the offset and slope estimator depends on the length of the data-record. To do this we conveniently use a slightly different model. We will use <u>yearly observations</u> (each time taken at Jan 1st), only one observation per year, and we use Jan1st, 1950 at t_0. Derive an analytical expression for the standard deviation of the offset estimator and for the standard deviation of the slope estimator, as a function of N, the number of years we take into account on *either side* of the year 1950. For instance, with N=10, we use data from 1940 to 1960 (both included). And for N=50 we use data from 1900 to 2000 (both included). You may want to use the identity $\sum_{i=1}^{k} i^2 = \frac{1}{6}k(k+1)(2k+1)$. Please explain why the standard deviation of the offset and slope estimator behave differently as a function of *N*.

Question 6

a) In Question 3 you were instructed to use a standard deviation of σ_{y_i}= 60 mm, for $i = 1, \dots, m$ for your measurements. Would your parameter estimates of the offset and slope change if the standard deviation instead would be given as σ_{y_i}= 80 mm, for $i = 1, \dots, m$, and if so, by how much? What would such a change do to the precision of the estimated offset and slope?
b) Now assume that the standard deviation of the measurements is <u>unknown</u>, but still the same for all measurements. Can you describe a method of using your data to determine an estimate of the unknown standard deviation, as well as what then the quality of such estimate is?
c) In Question 4 the data were assumed to be normally distributed. Now assume that the distribution is unknown. Can you describe a method of using your data to determine or verify its probability distribution?

References

Permanent Service for Mean Sea Level (PSMSL), 2013, Tide Gauge Data from http://www.psmsl.org/data/obtaining/.
Simon J. Holgate, Andrew Matthews, Philip L. Woodworth, Lesley J. Rickards, Mark E. Tamisiea, Elizabeth Bradshaw, Peter R. Foden, Kathleen M. Gordon, Svetlana Jevrejeva, and Jeff Pugh (2013). New Data Systems and Products at the Permanent Service for Mean Sea Level. Journal of Coastal Research: Volume 29, Issue 3: pp. 493 – 504. doi:10.2112/JCOASTRES-D-12-00175.1.