Exercise 3

Download the file *dt_global_allsat_msla_h_y1993_2018_05deg.nc* (netcdf format, ~52 MB) or *dt_global_allsat_msla_h_y1993_2018_05deg.txt* (ascii, ~93 MB) from Blackboard, or directly via these links:

https://surfdrive.surf.nl/files/index.php/s/i5PAFE6pJaH1DWz (ascii)
https://surfdrive.surf.nl/files/index.php/s/IVXcyZpqanju5e5 (netcdf)

The data in this file represent monthly sea surface height anomalies (1993-2018) with respect to a 20-year mean reference period (1993-2012), observed by a series of altimetry satellites. In the ascii-file, the date of the observations is given in a decimal representation, in the netcdf as days since 01 Jan 1950. The first date corresponds 15 January 1993, the last one to 15 Dec 2018. The original data come at a 0.25° resolution, and cover the global oceans, and were downloaded from:

https://cds.climate.copernicus.eu/cdsapp#!/dataset/sea-level-daily-gridded-data-for-the-global-ocean-from-1993-to-present?tab=overview

More background on the data can be found here:

http://datastore.copernicus-climate.eu/c3s/published-forms/c3sprod/satellite-sea-level-global/product-user-guide.pdf

In this exercises, we'll focus on the tropical Pacific Ocean (here defined as 100° to 280°E; -29.75° to 29.75°N). The data has been downsampled to 0.5° resolution to reduce the file size and computational load.

1. Inspect the data by plotting a number of months. Provide a plot for Dec 2018, using a symmetric color scale. Describe the features in the equatorial band between -15° to 15° N and those outside this band.

In order to reduce the data set and extract its main geophysical signals, we will apply a Empirical Orthogonal Function (EOF) analysis, sometimes also referred to as Principal Component Analysis (PCA).

As a first step, create a data matrix where each row contains the grid data of a time step. Next, form the covariance matrix and derive its eigenvalues and -vectors. The eigenvectors represent the *empirical orthogonal functions* (i.e., spatial modes of variability), the eigenvalues give a measure of the total variance in the data explained. In other words, eigenvectors with a large eigenvalue explain more variability of the sea surface height anomalies than those with a small eigenvalue. How each EOF field evolves in time, can be found by projecting the eigenvector on the data matrix. This yields the *principal component (PC) timeseries*. A short discussion of the method can be found here (sections 1-2.3):

https://surfdrive.surf.nl/files/index.php/s/gTWD7lheLGagw9R or in the seminal book of Rudolf Preisendorfer, *Principal Component Analysis in Meteorology and Oceanography* (1988).

Don't attempt to do this in Excel (let alone by hand!), it can easily be done in mathematical software such as *Matlab* or *R* using the *princomp* command, or in Python by importing *PCA* from *sklearn* library, for example. Alternatively , it can be done by applying a Singular Value Decomposition to the data matrix.

Once you have derived the EOFs and PCs, order them according to the variance explained and apply a normalization to the EOFs and PCs so that the absolute maximum value of each PC equals 1 (keep in mind that if you scale the PC by a factor x, the EOF needs to be scaled by 1/x). Then, do the following:

2. Make a plot of the first EOFs (i.e., the spatial maps) and its PCs (time series). Discuss the main patterns that stand out in the EOF and provide an interpretation of what this mode represents (hint: look at the timing of the peaks in the PC)
3. Plot the EOFs and PCs for mode 2 and 3. Perform a spectral analysis of the PCs and determine the dominant frequency of the modes (include the plots of the spectral analysis in your answer). What do these modes represent?
4. Make plot of the cumulative fraction of explained variance of all modes.
5. Rebuild the data set using modes 1-6. How much of the total variance is contained in this reconstructed data? Provide a plot of the reconstructed data for Dec 2018, and of the difference between the original and reconstructed data. Discuss these differences.

When returning your answers, please DO include the code you used to generate your results.