M1 - Marine Physics 2019–2020
Data Analysis

#3 Statistical analysis of Argo data

This activity will be evaluated and marked. Send by email (gula@univ-brest.fr) your notebook or python script(s) and the figures you plotted.

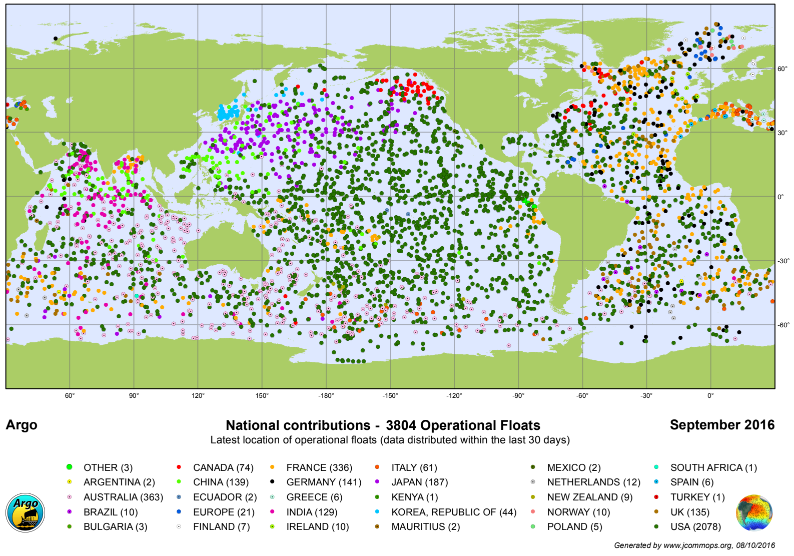
Due date : Wednesday November 20th 2019.

1 The data

The data you will analyze are stored here

`http://stockage.univ-brest.fr/~gula/TS1/argo_subdataset.mat`

The file contains $n = 33\,028$ Argo vertical profiles from the North Atlantic Ocean. Each profile k has been acquired by a certain profiler (information not given in this datafile) at a certain time t_k, a certain location (ϕ_k, λ_k). Each profile has been interpolated on $m = 62$ levels z_l, the same for all profiles. The data of each profile consists in : absolute salinity (SA), the conservative temperature (CT) and a virtual density (ρ) (see IAPSO 2010 for more informations on the definition of these variables). The file also contains the corresponding bathymetry on a $1/6°$ resolution grid from ETOPO2v2.



name	size	meaning
`time`	n x 1	time of profile acquisition (in days)
`lat`	n x 1	latitude of the profile
`lon`	n x 1	longitude of the profile
`z0`	62 x 1	depth of each data in a profile
`ct`	n x 62	conservative temperature
`sa`	n x 62	absolute salinity
`rho`	n x 62	virtual density
`htopo`	241 x 301	elevation on a 1/6° grid
`xtopo`	301 x 1	longitude of the elevation
`ytopo`	241 x 1	latitude of the elevation

Table 1 – List of variables in the data file

2 Descriptive information

Each plot should have proper **labels, title, units, colorbar if any. When using colors, control the color axis** (`caxis`). Save each figure in a PNG file as 'figX.Y.png', where X is the section, Y the question.

1. do a horizontal plot showing the location of each profile (use a 'o' for each profile)
2. do a histogram showing the number of samplings over the whole Argo period of this dataset with a binsize of 1 year
3. Most of the vertical profiles start at -2000m but not all of them. Some start less deep. Do the histogram of the maximum depth of each profile (with a binsize of 50 m). You can use the vector of indices `ind` corresponding to the maximum density sampled for each profile from the command matlab : `[~,ind]=max(rho,[],2);`, python : `ind = np.nanargmax(rho,1)`. You should find that about 50% of them reach 2000 m.
4. do a scatter plot showing each profile location as a colored dot where the color indicates the maximum depth of each profile matlab : `scatter();` , python : `plt.scatter()`

3 Climatology

1. do a monthly climatological histogram : number of profiles for each month of the year (twelve bins). You can use module datetime in python to convert the time into the month value.
2. do a seasonal climatological histogram : number of profiles for each season of the year (winter, spring, summer, autumn).
3. do a scatter plot where the color indicates the temperature at the surface
4. do scatter plots where the color indicates the temperature at the surface for the different seasons (winter, spring, summer, autumn)

4 Mode Water

A mode water is defined as an anomalous volume of water with homogeneous characteristics, usually temperature and salinity. Subtropical mode waters are located in the thermocline and play a key role in biology. The mode water formation process usually occurs at the mid-end of the winter through buoyancy loss at the sea surface (temperature loss and/or salinity gain) which triggers mixing and homogenize the water deep in the water column (then a mode water). Among them is the North Atlantic Subtropical Mode Water, often called 18°C water, which is characterized by a temperature of 17.8°C, a salinity of 36.5PSU and a potential density of 26.45 kg/m^3.

1. do a similar scatter plot where the color indicates the temperature at 200 m depth. Can you see the mode water (around 18°) ?
2. plot the probability density function (PDF) of the temperature measured between 100 m and 500 m. Can you identify the dominant class of temperature (= mode of the distribution) ?
3. do a scatter plot for the salinity at 200 m depth.
4. plot the PDF of salinity measured between 100 m and 500 m. Can you identify the dominant class of salinity ?
5. do a scatter plot for the density (ρ) at 200 m depth.
6. plot the PDF of density measured between 100 m and 500 m. Can you identify the dominant class of density ?

5 Averaging and gridding

The protocol We now want to average and bin the argo data on a regular grid. Let (x_i, y_i) be the location of the grid points. The goal is to estimate $T_i(z_l)$ the temperature at (x_i, y_i, z_l). For that we introduce a weigthing matrix

w_{ik} whose coefficients are

$$w_{ik} = e^{-d_{ik}^2/(2\sigma^2)} \tag{1}$$

with d_{ik} the distance between (x_i, y_i) and (ϕ_k, λ_k) and σ a smoothing length. Note that all the distances will be measured in degrees. To be exact we should consider the spherical distance but we will assume that the euclidean distance is a good estimate of it

$$d_{ik}^2 = (x_i - \phi_k)^2 + (y_i - \lambda_k)^2 \tag{2}$$

Once this weight matrix is set, the temperature T_i (for a given depth z_l) is

$$T_i = \frac{\sum_k w_{ik} T_k}{N_i}, \qquad \text{with} \qquad N_i = \sum_k w_{ik} \tag{3}$$

In practice the summation over k should be done only on points whose value is not NaN and the T_i should be computed only on ocean points (although the algorithm is able to give a value on land).

Implementation

1. First of all you will create arrays for longitude and latitude `xtopo1`, `ytopo1` on a 1° grid. This is the grid on which you will compute the averaged and binned horizontal fields.
2. Coarsen the topography elevation `htopo` to get topography `htopo1` on the 1° grid.
3. Implement the following algorithm to average and bin the data :
   (a) for a given z_l, find the k indices whose T_k is not NaN
   (b) find the i indices whose T_i are ocean points
   (c) construct the d_{ik} matrix and the w_{ik}
   (d) compute T_i and N_i
   (e) reshape T_i and N_i on the grid

   Do a map of $T_i = T(x_i, y_i)$ at the surface (set the land points to NaN) with $\sigma = 1/2°$.
4. Redo the map with $\sigma = 1°$.
5. Repeat the operation for the surface salinity.
6. Repeat the operation for the temperature at 200 m during winter.
7. Do a map of N_i. Interpret this result : what is the meaning of N? [check what happens when you change σ].
8. Pick an ocean point of the grid i_0 and do a scatter plot of the profile locations using $w_{i_0 k}$ as a color.
9. Do the reverse operation : pick a profile k_0 and do a map (gridded) of w_{ik_0}.
10. Equation (3) can be thought as the mean operation where P_{ik}

$$P_{ik} = w_{ik}/N \tag{4}$$

   is a probability density function (PDF). Indeed we have $P_{ik} \geq 0$ and $\sum_k P_{ik} = 1$.
   Adapt the protocol to do a map of the surface temperature standard deviation

$$\sigma_i^T = \sqrt{\sum_k P_{ik}(T_k - \bar{T}_i)^2} \tag{5}$$

   with $\bar{T}_i$ the mean temperature computed with (3).