ROMS Extractor User's Guide

Abstract. In this manuscript, a Python-based Graphical User Interface (GUI) tool for extracting data from ROMS outputs is presented. The tool can be used to produce either 2-D time averages or time series of spatially averaged variables. Data requests are stored in an output NetCDF file and, optionally, a plot can be produced. Available variables include raw variables, such as current velocity, sea surface height, salinity and temperature, but also derived variables, such as density, mixed layer depth, potential energy deficit and a SST-based front index. This tool is aimed at facilitating some common tasks that are usually performed with outputs from numerical ocean models, such as ROMS.

Contact. Diego Pereiro Rodríguez (diego.pereiro@marine.ie)

1. Installation.

In order to use the Marine Institute ROMS Extractor Tool, just place the following five files under the same directory:

- The executable file *romsext.exe*. This file is the result of compiling the code using PyInstaller. You can run the app by just double-clicking this file under Windows OS, even if Python is not installed in your computer. Run the software only after you have properly modified the *config* file (see below).

- You might want to modify the code according to your specific needs. This is why the code file *romsext.py* is also provided in the installation package. You can run this file under any Python IDE.

- The Marine Institute's logo file *icon*.

- The coastline file *mundo_h.dat*. This file is a worldwide coastline database which is used to display the shoreline, both in the area selection pane and in the output plots. This file was produced using the GEODAS Coastline Extractor software. The spatial resolution looks appropriate for regional to global scales. However, it might not be enough for local applications. You can replace this file by another one that satisfies your needs, as long as some format requirements are met. The file is just a plain text with two columns of data: left column is longitude –in -180° to 180° convention–, right column is latitude, separated by one or several whitespaces. Separation between coastal polygons is indicated by a pair of missing values: *nan nan*. If you decide to use a different coastline file, you can either use the same file name –*mundo_h.dat*– or modify the code by introducing the name of your file under the `""" Read coastline """` block.

- The *config* file. The software will read this file to get some essential information about your ROMS setup. Lines starting with '!' will be ignored. Before you run the software for the first time, you should provide the following information in this file:

a. **TRACERS_CONVENTION** (mandatory). Here, you tell the software how your ROMS files are named, including the full path. Usually, output file names from ocean models refer to some point in time. The string after '`TRACERS_CONVENTION =`' will be subject to some substitutions as detailed below.

If the date is, for instance, 15th of February 2019 12:28:03, then:

%Y	will be replaced by	'2019',	i.e.	4-digit year
%y	will be replaced by	'19',	i.e.	2-digit year
%m	will be replaced by	'02',	i.e.	2-digit month
%b	will be replaced by	'Feb',	i.e.	month abbreviation
%d	will be replaced by	'15',	i.e.	2-digit day
%o	will be replaced by	'046',	i.e.	3-digit ordinal date
%H	will be replaced by	'12',	i.e.	2-digit hour
%M	will be replaced by	'28',	i.e.	2-digit minute
%S	will be replaced by	'03',	i.e.	2-digit second
%rt	will be replaced by	a		4-digit ordering number
%rm	will be replaced by	a		4-digit ordering number

So, for example, if your 2016 hindcast is in folder D:\ROMS\2016\ and your 8th May 2016 file name is 'ocean_avg_00128.nc' –where the 3-digit ordinal date naming convention has been used–, then you should use `TRACERS_CONVENTION = D:\ROMS\%Y\ocean_avg_00%o.nc`. Or, if your 15th of January 2020 13:00 file name is D:\ROMS\Connemara\OUTPUT\HISTORY\CONN_2020011513.nc, then you should use `TRACERS_CONVENTION = D:\ROMS\Connemara\OUTPUT\HISTORY\CONN_%Y%m%d%H.nc`.

An additional replacement placeholder is %rt. This is useful if files are named according to a consecutive number scheme. This means that there is an initial file named something like *ocean_avg_0001.nc*. There is one time step per file and the next file is called *ocean_avg_0002.nc*. The appended number will be determined as the number of time steps elapsed since a given initial time, plus one. The initial time is prescribed through the *config* variable 'IDATE' and the time step length is determined through the *config* variable 'STEP' (see below). For example, if time in the initial file *ocean_avg_0001.nc* is the 18th of December 2019 00:00 and the time step is 3 hours, then time in the file *ocean_avg_0200.nc* would be the 11th of January 2020 21:00. Unless you change the behaviour of the `get_running_number` method (see code), the software will expect a file name with a 4-digit number appended, with leading zeros if necessary.

Different replacements other than the ones in the table above have not been implemented yet. In order to enable other file naming conventions, you may either modify the `self.rep` attribute under the `""" Get input file name matching date """` block or contact the author.

b. **MOMENTUM_CONVENTION** (optional). This is the same as TRACERS_CONVENTION but for free-surface and momentum variables. Sometimes, tracer variables (e.g. temperature, salinity) and momentum variables are saved in different files with different time steps. For instance, the hindcasts of the Irish Marine Institute's Northeast Atlantic Model are provided every 3 hours for tracer variables, whereas 1-hour time steps are used for momentum variables. Different file names are used and the software will read tracer variables using the TRACERS_CONVENTION, and momentum variables using the MOMENTUM_CONVENTION.

When using a consecutive-number naming convention, do not use the %rt placeholder in the MOMENTUM_CONVENTION. Use a %rm instead. This will use 'ZSTEP' (see below) instead of 'STEP' to get the number of time steps elapsed since 'IDATE'.

If you make no difference between tracer and momentum variables and both use the same naming convention, just ignore the MOMENTUM_CONVENTION entry by placing a '!' character at the beginning of the line.

c. **IDATE** (mandatory). Initial date of your ROMS application. Use the same format as in the *config* file provided: YYYY-MM-DD HH:MM:SS.

d. **EDATE** (mandatory). End date of your ROMS application. Use the same format as in the *config* file provided: YYYY-MM-DD HH:MM:SS.

e. **STEP** (mandatory). This is the time step (seconds) between successive model outputs, no matter whether they are in different files –single-record files, when there is one file per time step and the length of the time dimension in each file is 1– or in the same file –multiple-record files, when the length of the time dimension in each file is greater than 1–.

The software will create a list of the available time records, from 'IDATE' to 'EDATE' every 'STEP' seconds and the full contents of this list will be displayed in a drop-down list to choose from. Therefore, if you are working with a 45-year hindcast and use `IDATE = 1975-01-01 12:00:00`, `EDATE = 2020-01-01 12:00:00`, and `STEP = 86400`, then you will get a drop-down list with 16436 elements, which might be rather uncomfortable to deal with. Nevertheless, if you plan to get data from the 2016 hindcast only, then it is better to use `IDATE = 2016-01-01 12:00:00` and `EDATE = 2016-12-31 12:00:00` to get a shorter drop-down list of only 366 elements. You can modify the *config* file at any time later and, after restarting the software, changes will be applied.

Furthermore, if your time step between successive model outputs is Δt, you can use a 'STEP' greater than Δt such that 'STEP' = $n \cdot \Delta t$, where n is a positive integer. This is useful if, for instance, your time step between successive model outputs is 1 hour, but you are interested in the average surface temperature at noon only. Then, even if $\Delta t = 3600$ s, use `STEP = 86400`.

Finally, in the aforementioned case of the Irish Marine Institute's Northeast Atlantic Model, where $\Delta t = 1$ h for momentum variables and $\Delta t = 3$ h for tracer variables, 'STEP' should be 10800 seconds, unless there is no interest at all in tracer variables –or variables derived from tracers, such as density–, in which case `STEP = 3600` might be used.

f. **ZSTEP** (optional). When using a different naming convention for tracer and momentum variables, and when using a consecutive-number naming scheme, a time step of 'ZSTEP' seconds will be used to get the number of time steps elapsed since 'IDATE' as a replacement for the %rm placeholder. Place a '!' character at the beginning of the line if you use the same naming convention for tracer and momentum variables, or if you are not using the consecutive-number naming scheme.

g. **OFFSET** (mandatory). ROMS time must be expressed in **seconds** since some time in the past. For instance, if ROMS time is expressed in seconds since the 23rd of May 1968 00:00:00, use `OFFSET = 1968-05-23 00:00:00`. Use the same format as in the *config* file provided: YYYY-MM-DD HH:MM:SS.

h. **VARIABLES** (optional). The tool works with an internal list of 63 variables. These variables are displayed in a drop-down list in the GUI. Since: **a)** you will probably need to search for one or some specific variables in this list each time you run the app and **b)** you might have no interest at all in some of the variables included in the tool's internal list, you can use the 'VARIABLES' entry in the *config* file to save some time. Just select the variables you would like to appear in the drop-down list by indicating its prefixes separated by commas, as shown in the *config* file. For instance, if you are only interested in average, minimum and maximum surface salinities, use `VARIABLES = 16, 17, 18`. Otherwise, if you do not find this feature useful, just place a '!' at the beginning of the line.

After the five files have been placed under the same folder and the *config* file has been modified properly, the tool is ready to run.

2. Usage.

After the software is launched, a new window pops up (see figure below). This window is divided into different frames, as detailed below:

1. The "Choose a file" frame. The tool will always create an output NetCDF file containing the data requested by the user. You should indicate the full path of the output NetCDF file. Just press the "Select output directory" button to select an output folder and enter the desired file name in the "Output file name" entry.

2. The "Select mode" frame. The tool offers two methods, which cannot work simultaneously: the time-averaging method and the time-series method:

- In the time-averaging method, you will get 2-D horizontal slices (surface, bottom or constant-depth) averaged over time. Even though it is called the time-averaging method, you can get not only averages but also minima or maxima over time.

- In the time-series method, you will get time series of 2-D averages, minima or maxima.

3. The "Time period" frame. Here, you can select as many periods of time of any length between 'IDATE' and 'EDATE' as you want. You may choose 'IDATE' and 'EDATE' to be equal to get a single time step output. Under the time-averaging method, you will get separate averages (or maxima or minima) for each period of time selected. Under the time-series method, you will just get one period after another in chronological order, but in different plots (if plotting has been requested).

There are two different procedures to select periods of time: the "Quick selection" method and the "Manual selection" method. The "Quick selection" method automatically selects step-by-step, daily, weekly and/or monthly periods between two months from the same or different years. Weeks are considered to start on Sunday and to finish on Saturday. In the "Manual selection" method, just select a beginning date in the "From" drop-down list, an end date in the "to" drop-down list and press the "Add" button. Repeat this procedure as many times as necessary. You can delete the last entered time period by pressing the "Delete" button. You can also modify the contents of both drop-down lists by properly editing the *config* file (see 1.e above).

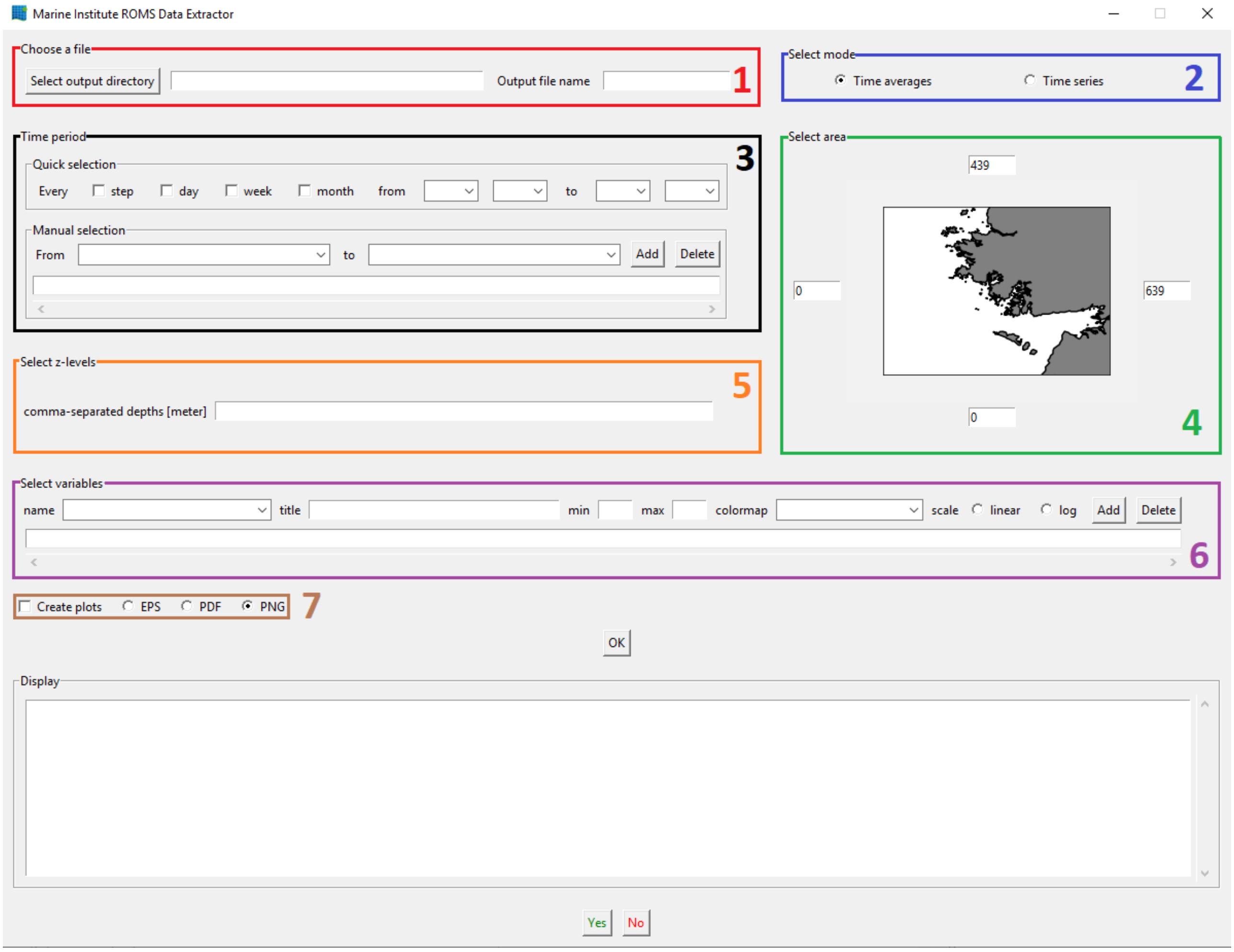
4. The "Select area" frame. Averages, maxima and minima will be taken over the selected area. The entire domain is selected by default. You can select the desired boundaries by entering the i-grid and j-grid coordinates (only non-negative integers are allowed). The selected rectangular area will be drawn in the map.

5. The "Select z-levels" frame. Some variables are explicitly referred to the ocean surface or to the ocean floor. Other variables have no vertical dimension, such as the sea surface height. But there are variables that will require you to introduce at least one depth (otherwise, they will be empty in the output NetCDF file). In this frame, you can enter one or more comma-separated values to get horizontal slices at constant depths from 3-D variables. You can enter either positive or negative values, since the software will convert them to negative anyway. For example, enter '20, 100, 200' to get z-slices at depths 20 m, 100 m and 200 m.

6. The "Select variables" frame. Here, you can select the variables you are interested in. Just click the variable name from the drop-down list "name" and do not forget to click the "Add" button. Repeat this procedure as many times as necessary. You can delete the last entered variable by pressing the "Delete" button. If you do not select any variable, all the variables in the drop-down list will be included by default. You can modify the contents of this drop-down list by properly editing the *config* file (see 1.g above).

Furthermore, you will find some plotting options under this frame that you can just ignore if you are not interested in plotting. Enter the title of the plot in the 'title' entry; enter the colormap limits (or the y-axis limits under the time-series mode) in the 'min' and 'max' entries; choose a colormap from 'colormap' drop-down list, and select the type of the colour scale (or the y-axis scale under the time-series mode): either linear or logarithmic.

7. The plotting frame. Select whether you would like to produce plots or not (a NetCDF file will be created in any case). You can choose from three different file formats: EPS, PDF or PNG.



Finally, click the ‘OK’ button. A summary of actions to be taken will be displayed. You will be asked for permission to proceed. Press the ‘Yes’ button if you agree; otherwise press the ‘No’ button and make the desired changes.

3. Features.

3.1. Variables.

Variables can be classified in either 2-D or 3-D, real or complex, raw or derived:

- **2-D and 3-D variables.** 2-D variables are explicitly referred to the ocean surface or to the ocean floor, or have no vertical dimension, such as the mixed layer depth. 3-D, full-volume variables are temperature, salinity, density, u, v, and velocity. These variables are intended for **z-slicing** only, where constant depth, horizontal slices are obtained through linear interpolation along the water column. You can indicate the depth of the z-slices under the “Select z-levels” frame (see 2.5 above).

- **Real and complex variables.** Real variables are scalars and they are used to produce either colormap plots (under the time-averaging mode) or time-series plots (under the time-series mode). Complex variables are vectors. When plotting a complex variable (e.g. velocity) under the time-averaging mode, you will get a combined quiver-colormap plot, where arrows are used to show direction, and colour is used to show magnitude. When plotting a complex variable under the time-series mode, you will get a feather plot (see 3.2 below).

- **Raw and derived variables.** Raw variables are readily obtained from the ROMS output: sea surface height, temperature, salinity, u and v are raw variables. Derived variables are obtained from raw variables. The following derived variables are available:

  a. **Density.** Density is computed using UNESCO 1983 (EOS-80) equation of state (Fofonoff and Millard Jr 1983; Millero et al. 1980)

  b. **Velocity.** Magnitude and direction are computed from the u, v components of velocity. Average velocity is taken as a vector average, i.e. both magnitude and direction are taken into account. So, if we have two vectors (u_1, v_1) and (u_2, v_2), then the average is ($\frac{1}{2}u_1 + \frac{1}{2}u_2$, $\frac{1}{2}v_1 + \frac{1}{2}v_2$), and average magnitude and direction is calculated from this average.

     However, maximum velocity takes only magnitude into account. Therefore, maximum velocity is taken as simply the vector whose modulus is the largest.

     This software takes grid rotation into account, so that raw u and v are rotated according to the ROMS grid 'angle' value.

  c. **Potential Energy Deficit (PED).** Potential Energy Deficit (Planque, Lazure, and Jegou 2006) is a measure of the stratification of the water column, and it is defined as the amount of energy required to vertically mix the water column, in units of $kg \cdot m^{-1} \cdot s^{-2}$. PED is calculated from density according to:

$$\boldsymbol{PED = \frac{1}{H}\left(\sum_{0}^{-H}(\rho - \rho_z)\, gz\right)}$$

     where z is depth, ρ is the vertically-averaged water column density, ρ_z is density at depth z, g is acceleration due to gravity and H is the bottom depth or 200 meters where the water column is deeper (Bonanno et al. 2014).

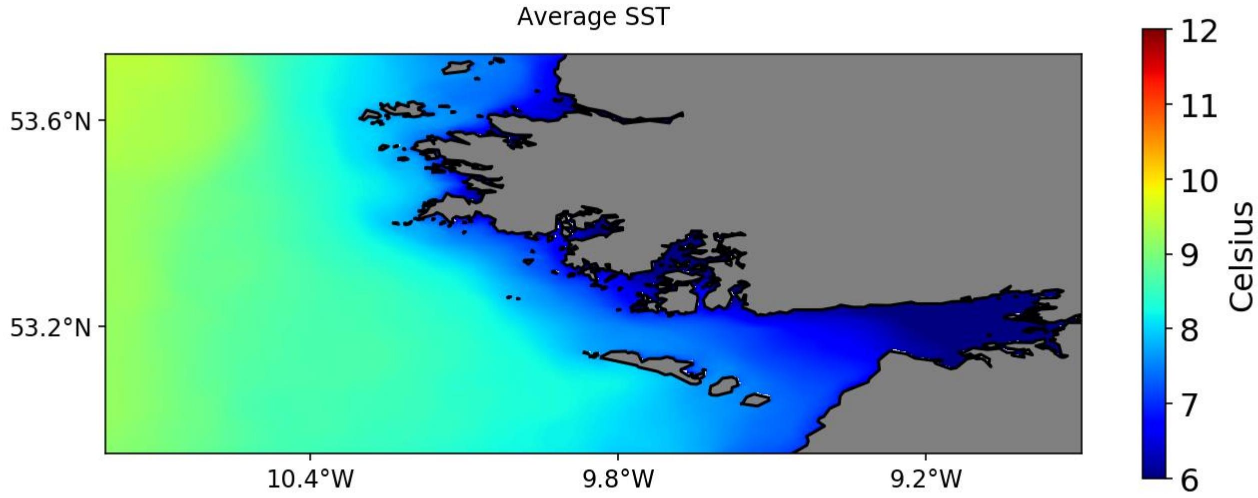
  d. **Mixed Layer Depth.** Mixed Layer Depth is defined as the depth at which the temperature change from the surface temperature is 0.5 °C (Levitus 1982).

  e. **SST-based front Index.** The front index has been defined in terms of the magnitude of the surface temperature gradient computed using the Sobel operator (Nieto, Demarcq, and McClatchie 2012).

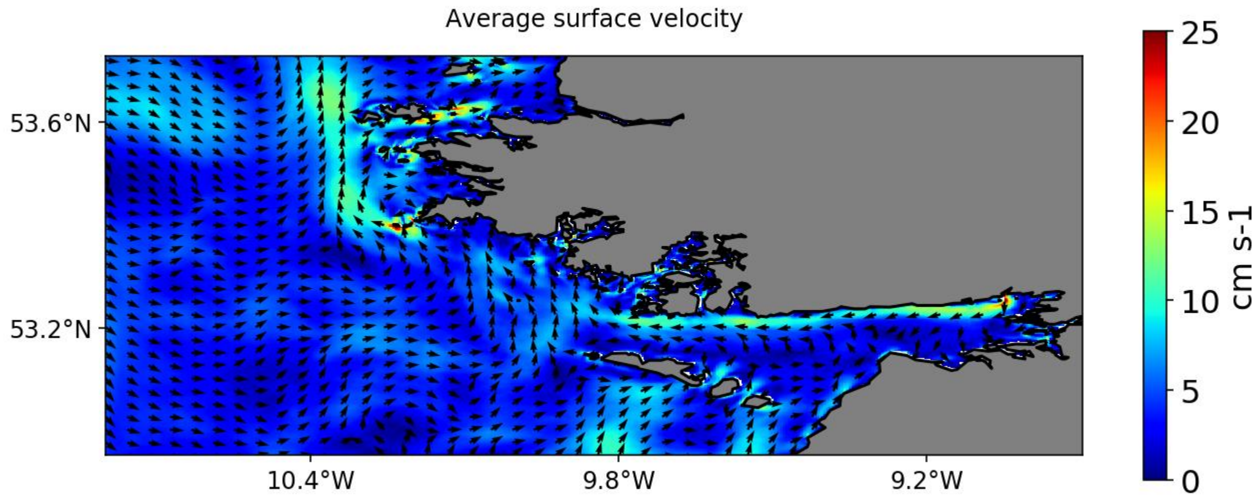
3.2. Plots.

Four different types of plots can be produced:

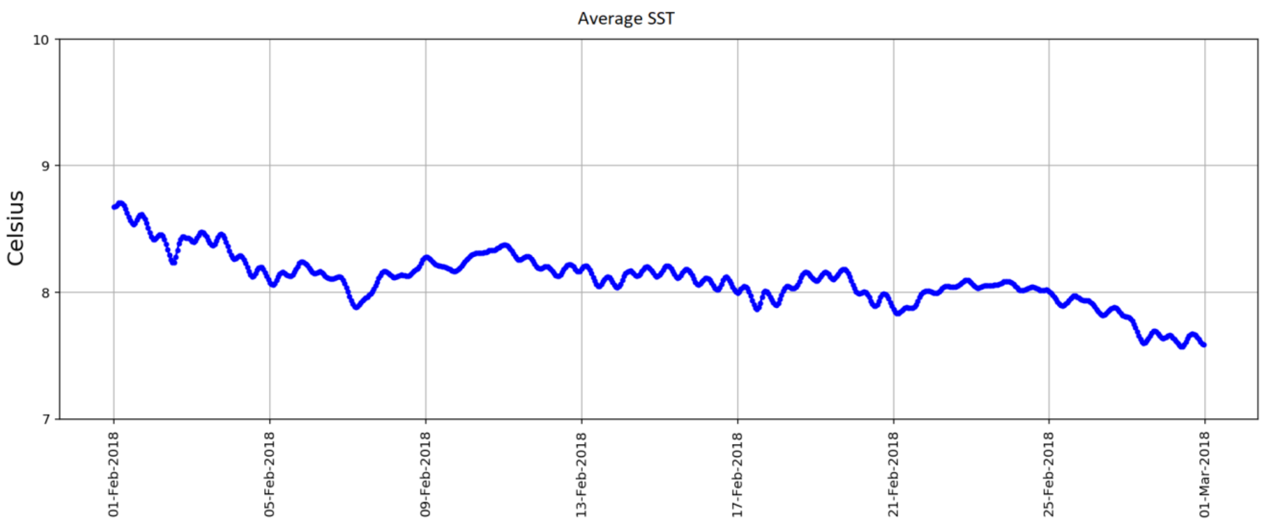
a. **Colormap plot.** Plot a scalar under the time-averaging mode to get a colormap plot.



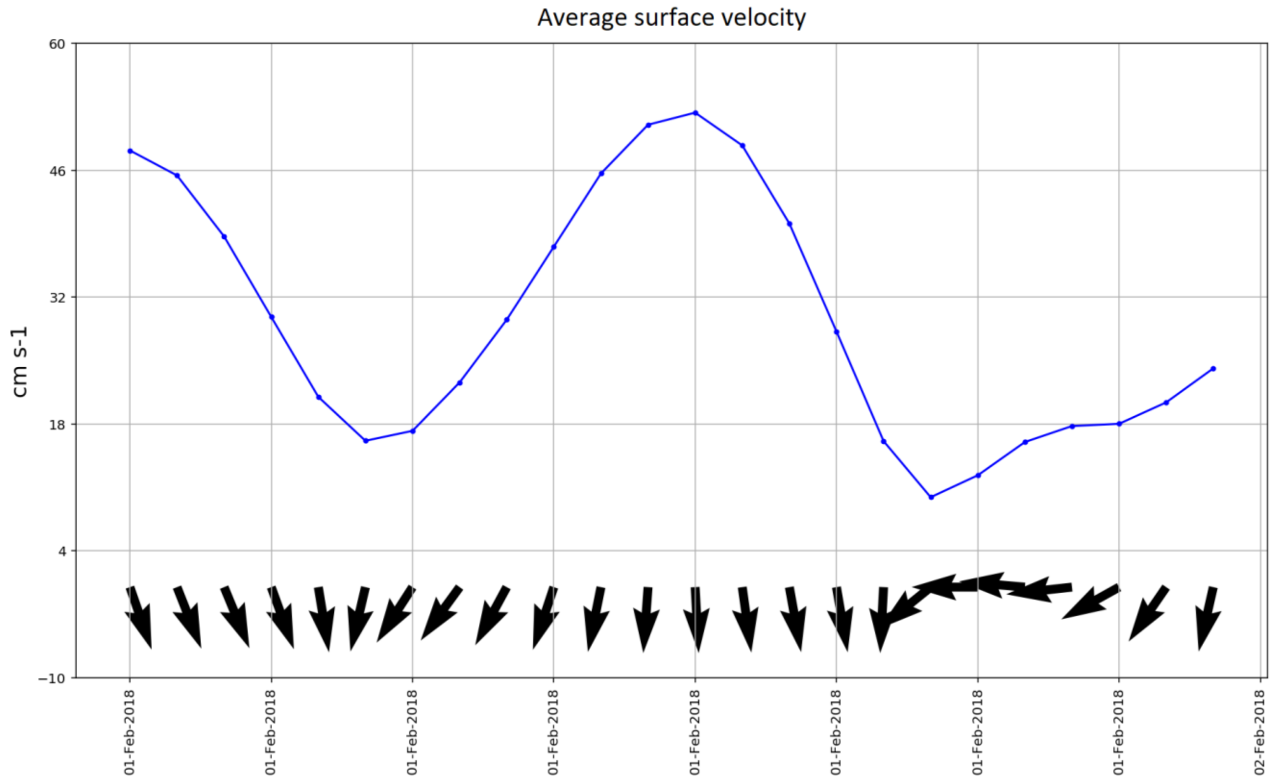
b. Combined quiver-colormap plot. Plot velocity under the time-averaging mode to get a combined quiver-colormap plot. Colour shows magnitude, while arrows show direction.



c. Time-series plot. Plot a scalar under the time-series mode to get a time-series plot.



d. Feather plot. Plot velocity under the time-series mode to get a feather plot, as below. Magnitude is represented as a time series. Direction is represented as arrows at y = 0. So, in order to visualize the arrows, you should choose a negative bottom limit for the y-axis.



Acknowledgements

I would like to thank Dr. Antonio Fernández Álvarez (University of Vigo) for his effort, enthusiasm and well-doing while teaching his Introductory Course to Python in May-June 2019.

References

Bonanno, Angelo et al. 2014. "Habitat Selection Response of Small Pelagic Fish in Different Environments. Two Examples from the Oligotrophic Mediterranean Sea." *PLoS One* 9(7): e101498.

Fofonoff, Nicholas Paul, and R. C. Millard Jr. 1983. "Algorithms for the Computation of Fundamental Properties of Seawater."

Levitus, Sydney. 1982. 13 *Climatological Atlas of the World Ocean*. US Department of Commerce, National Oceanic and Atmospheric Administration.

Millero, Frank J., Chen-Tung Chen, Alvin Bradshaw, and Karl Schleicher. 1980. "A New High Pressure Equation of State for Seawater." *Deep Sea Research Part A. Oceanographic Research Papers* 27(3–4): 255–264.

Nieto, K., Hervé Demarcq, and S. McClatchie. 2012. "Mesoscale Frontal Structures in the Canary Upwelling System: New Front and Filament Detection Algorithms Applied to Spatial and Temporal Patterns." *Remote Sensing of Environment* 123: 339–346.

Planque, Benjamin, Pascal Lazure, and Anne-Marie Jegou. 2006. "Typology of Hydrological Structures Modelled and Observed over the Bay of Biscay Shelf." *Scientia Marina* 70(S1): 43–50.