Xraster PLUGIN FOR QGIS

USERS MANUAL

This plugin requires additional Python libraries.

version: 0.1
November 2024

Table of Contents

1. INTRODUCTION 3
2. INSTALLATION 5
    1. QGIS VERSION
    2. **Xraster** INSTALLATION PROCEDURE
3. **Xraster** GENERAL PRESENTATION
    1. START AND QUIT THE TOOL IN QGIS 8
    2. TOOL STRUCTURE AND FUNCTIONS 8
    3. TYPICAL WORKFLOWS IN **Xraster** 10
    4. QGIS LAYERS USED IN **Xraster** 16
    5. PLOTS, SAVE AND EXPORT PLOT 17
    6. PLOT CUSTOMISATION 20
    7. EXPORT GEOTIFF RASTERS 20
    8. EXPORT TEXT FILES 21
4. USING **Xraster (Tabs)**
    1. **Data** 22
    2. **Subset** 24
    3. **Figures** 27
    4. **Statistics** 30
    5. **Extract to points** 32
    6. **Group by time** 34
    7. **Clip/Join** 35

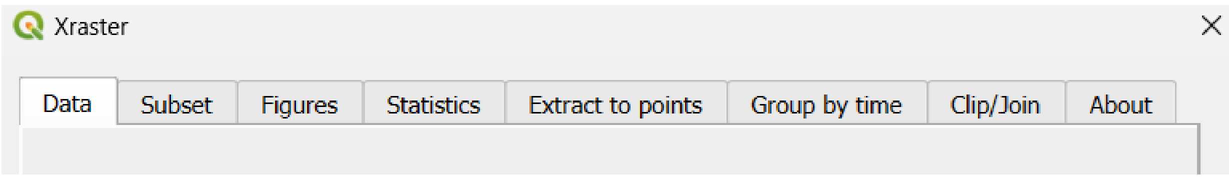
INTRODUCTION

Xraster is a tool designed for scientists and specialists to analyze multidimensional data describing the marine environment and climate. This data is provided by various organizations and scientific institutions (e.g., the European Union's Copernicus Programme) in the NetCDF format, which allows for the creation, access, and dissemination of array-oriented data. The array represents a grid on the Earth's surface, with variables assigned to each cell. Depending on spatial factors, the grid can be 2-dimensional (e.g., sea surface temperature) or 3-dimensional (e.g., water temperature including depth). An additional dimension characterizing this data is time, resulting in data sets that are typically 3-dimensional (x, y, and time) or 4-dimensional (x, y, z, and time). Geographic Information Systems (GIS) are the primary tools for spatial data analysis.

Xraster is embedded as a plugin in **QGIS**, enabling two-way integration for data processing. The plugin uses raster and point layers created in QGIS and generates data (raster layers and text data) that can be used within the program, creating a simple Multidimensional GIS. Xraster is based on the functionalities of the Python module **xarray**, providing an interface to this module, allowing the entire data processing and analysis workflow to be carried out without the need for programming.

The analysis process in this module, reflected in **Xraster**, is based on creating data subsets. Subsetting is carried out across 3 or 4 dimensions. Each dimension can be defined as a slice (a portion of the original space) or a period of time. It can also be defined as a point in space or a specific moment on the time axis. Different types of data subsets allow for various visualization methods (such as histograms, maps, time series, vertical profiles, and depth-time cross-sections) and statistical analysis.

Working with the program involves creating subsets, visualizing them, and conducting statistical analysis, the results of which can be visualized in the ways described above, saved as raster layers or text files. The plugin consists of several tabs, each dedicated to different stages of working with the program. The program includes the following tabs,



which perform the following tasks.

Data
- Input of 3-dimensional (x, y, time) and 4-dimensional (x, y, z, time) data in NetCDF (.nc) format and obtaining their description.
- Input of a raster mask (zone map) used for data analysis.

Subset
- Creation of a subset from the input data using: the selected variable, spatial range (x, y), time, depth (z), mask zones, and specified months.

Figures
- Depending on the selected subset, the following data visualizations are possible: histogram, map (x, y), vertical profile, depth-time cross-section, and time series (figures can be saved as high-resolution .jpg files).

- The map can be exported as a geoTIFF raster. Profile, time series, and cross-section data can be saved as CSV text files.

Statistics

- For the subset, statistics (Summary statistics) can be calculated for each day across the entire space (x, y) or within areas defined by mask zones (Zonal statistics). The results are saved as text files.
- Calculating statistics over time for a given area (x, y) generates a map, while at a specific point, it produces an average vertical profile. Maps can be saved as geoTIFF files.

Extract to points

- Extraction of data from the subset to points as a text file.

Group by time

- Aggregation of data into groups (for months, years, climatic seasons, or days) and calculation of statistics for each group. For each group, a map (at a specified depth) or a histogram of any statistic can be created.
- Groups can also be visualized as time series of a given statistic or as depth-time cross-sections. Time series and depth-time cross-sections can be saved to text files.

Clip/Join

- The program offers a range of tools for transforming .nc files. They can be efficiently cropped and merged. The created data subset can be saved as a new .nc file.

[TOC]

INSTALLSTION

---

1. QGIS VERSION

The **Xraster** tool is a plugin within QGIS. QGIS software is available from this site: https://qgis.org/ . The minimum version of QGIS required to run **Xraster** is 3.20 (the tool was designed and tested in version 3.34). The efficiency of working with the plugin depends on the computer's capabilities, particularly the available RAM (ideally 16-30 GB). However, the program can also run with 8 GB of RAM. The plugin operates on the Windows operating system.

QGIS may be installed using either the standalone installer or the OSGeo4W installer. **To use Xraster, you must use the OSGeo4W installer, which allows the addition of Python libraries used by the tool.** If you have QGIS installed via the standalone installer, you should first remove the software and download OSGeo4W. If you have QGIS installed via OSGeo4W, you do not need to remove QGIS. Next:

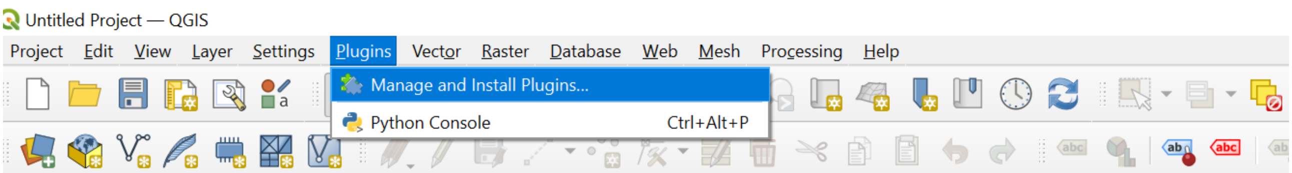
- **Run** osgeo4w-setup
- **Choose** Express install
- **Select Packages** : QGIS and GDAL (at least)
- Continue installation

  After installing the program, you should switch to the English language version (recommended).

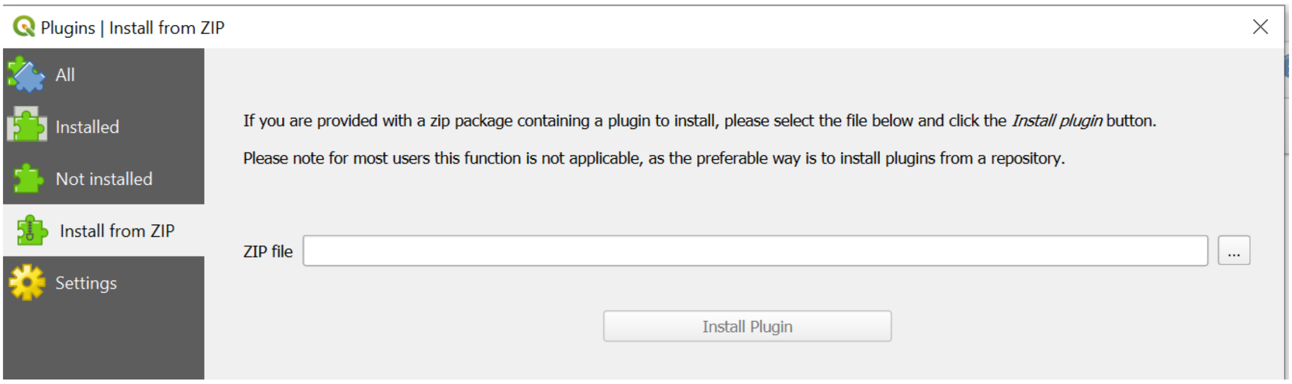
2. Xraster INSTALLATION PROCEDURE

The installation process of **Xraster** in QGIS follows these steps:

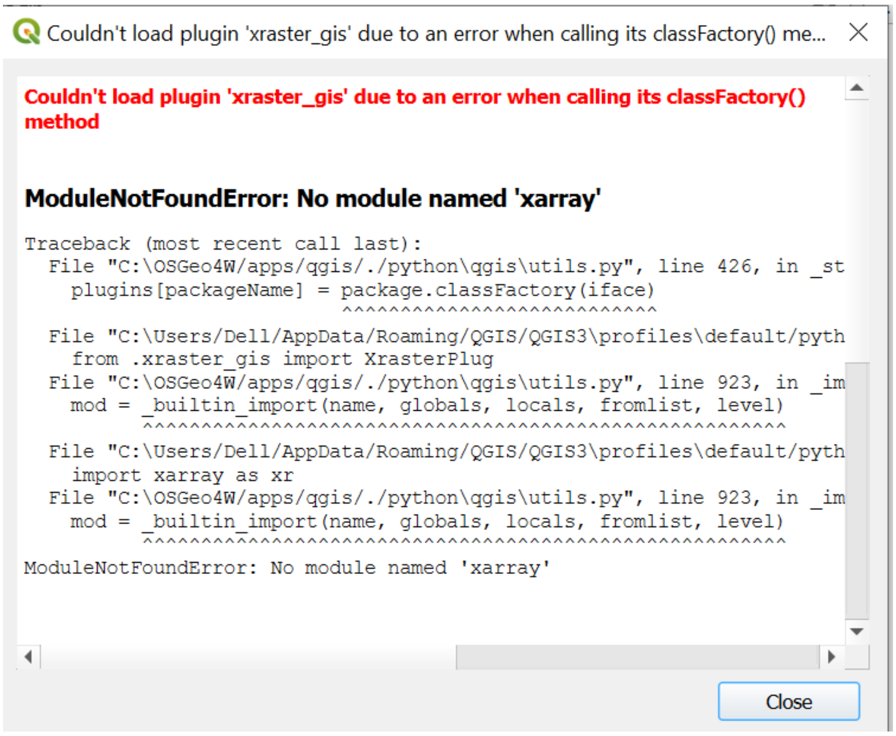
- You need to download the packaged plugin ***xraster_gis_?.zip*** from the website: https://github.com/urbanskigis/Xraster to any directory.
- Open QGIS 3.xx (minimum version 3.20)
- Open "Plugin" menu from the top bar



- Select "Manage and Install Plugins...
- Go to "Install from ZIP"
- Select ***xraster_gis_?.zip***
- Click install



Since the plugin requires additional Python libraries that are not installed by default, if they are not found, a message will be displayed,



In such a case, you should:

- Go to the **C: /OSGeo4W** directory where the OSGeo installer has installed QGIS
- Run **OSGeo4W.bat**
- Check with **pip list** packages installed

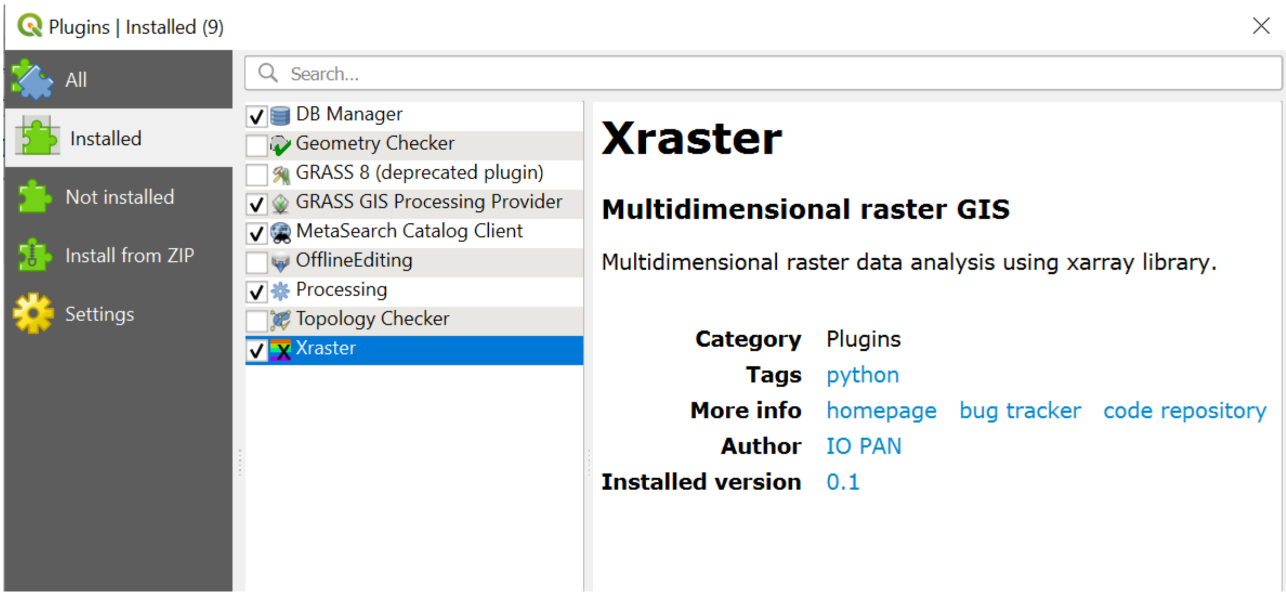
You need to install the packages: **xarray**, **netCDF4**, and **h5netcdf** using the following three commands executed in sequence:

```
C:\OSGeo4W>pip install xarray
```

```
C:\OSGeo4W>pip install netCDF4
```

```
C:\OSGeo4W>pip install h5netcdf
```

Lanch again QGIS and the plugin installation should run fine. When the installation is complete, the ***Xraster*** tool appears in the installed plugin list,



[TOC]

Xraster GENERAL PRESENTATION

---

1. START AND QUIT THE TOOL IN QGIS

After the installation, ***Xraster*** is made available in the tool bar by showing this button with the icon:

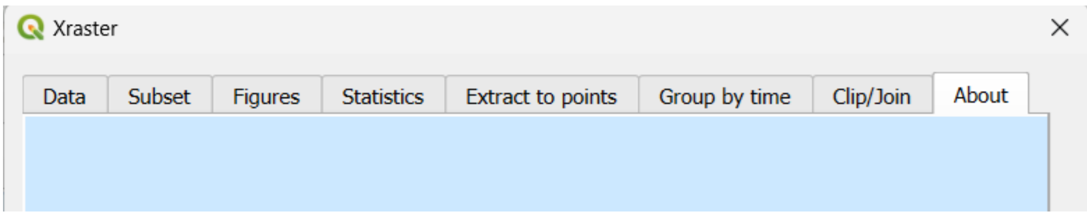
To start ***Xraster*** click on this button. If the button is not displayed, the tool can be selected in the plugin top bar menu. To quit the tool click **Close** click the button at the bottom of the window. **In QGIS, plugin windows may be hidden beneath the main program window, but they are visible on the bar at the bottom of the screen, from where they can be brought to the front.**

,

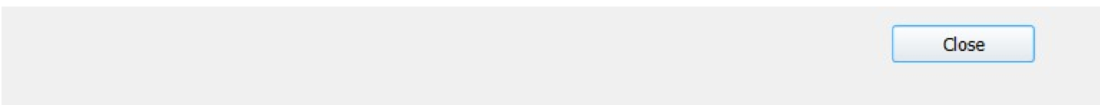
[TOC]

2. TOOL STRUCTURE AND FUNCTIONS

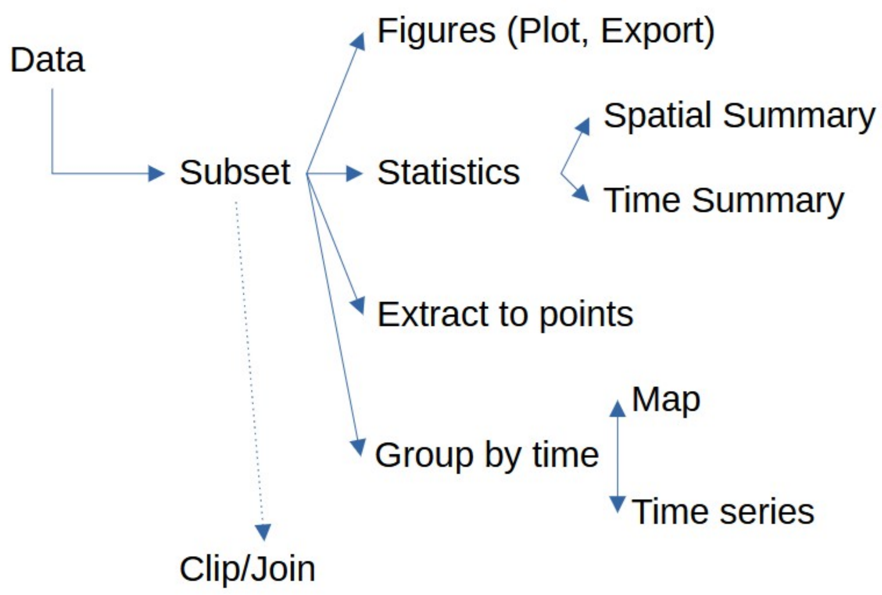
The **Xraster** plugin consists of seven tabs that implement different groups of functions of the software.



and a common **Close** button on the right side at the bottom for exiting the program.



The structure of the program is presented below,



The first tab **Data** is used for inputting data into the program. The program uses data in NetCDF (.nc) format and also uses raster masks in .tif format with georeferencing. The next tab **Subset** plays a fundamental role in the analysis process by creating a data subset. The primary dimensions of the data are the geographic coordinates defining a surface on the Earth (x, y), time (t), and depth (z). The input data can contain 3 (x, y, t) or 4 (x, y, z, t) dimensions. The defined data subset can refer to a surface or a point. It can pertain to a specific period of time or a particular moment in time. It may encompass a certain range of depths or a selected level. Additionally, when defining the data subset, it is possible to limit the data to a specific area and specific months using a mask. As a result, the data subset can take very different forms, which will determine the methods for its visualization and analysis. In searching for answers to a specific question, it is essential to first create the appropriate data subset.

Exploratory data analysis takes place in the **Figures** tab. Depending on the defined data subset, a specific graphical representation is created. The map, which represents the spatial distribution of the data, can be exported as a raster layer. The data used to create figures (e.g., vertical profile or time series) can be saved as text files.

There are two tabs dedicated to statistical analysis of the data subset: **Statistics** and **Group by Time**. The **Statistics** tab is used to calculate both spatial and temporal statistics. In the spatial analysis, a typical GIS function called Summary Statistics is performed, which computes statistics for a variable from a given area at a specified moment in time. The area can be defined using a mask. It is also possible to calculate statistics simultaneously for multiple mask zones (Zonal Statistics). Statistics are computed separately for each time point in the data subset and are saved as a text file. Temporal statistics are calculated separately for corresponding pixels in the subset, resulting in a map or vertical profile of the statistics.

The **Group by Time** tab groups data into categories based on a relevant key defined as a period on the timeline. This could be, for example, years or months. For each group, a statistic is calculated. It is possible to create maps or histograms of statistics for each group separately, or to create charts for all time periods (e.g., as time series).

The **Extract to points** tab extracts variables from a data subset to points defined in a QGIS vector point layer. It corresponds to the typical GIS function of extracting data from raster to point. The result is a text file containing variable values with a complete set of coordinates assigned to each point.

The **Clip/Join** tab is a separate subprogram that assists in creating .nc files for analysis in Xraster. Downloaded NetCDF (.nc) files can cover large areas, contain many variables, and may also be downloaded for separate time periods or specific days. As a result, they are often impractical for use directly in the program. Effective analysis requires combining these files into a single file while minimizing its size. This can be achieved by limiting the area and number of variables. The tab performs three functions:

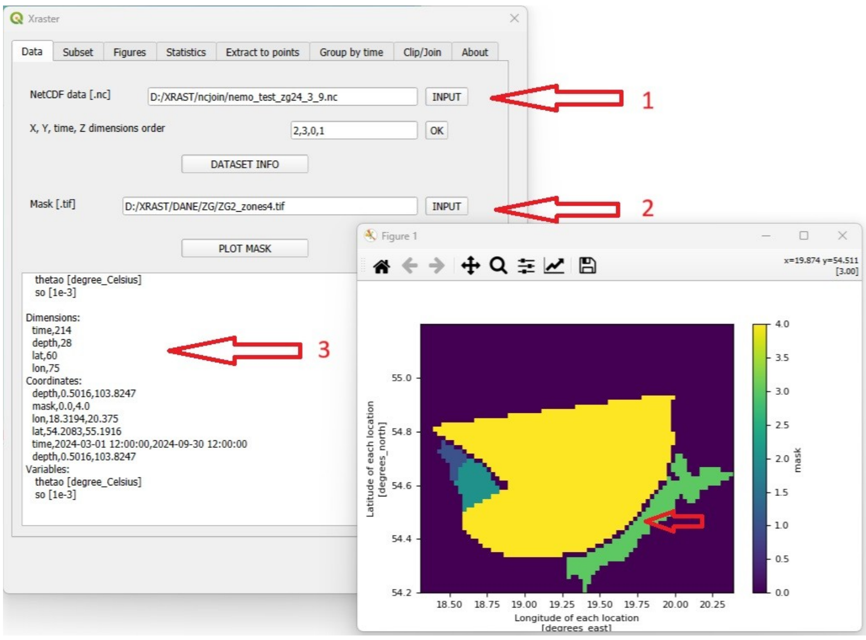
1. Organize file paths – Allows organizing paths to downloaded files into a single text file, assuming they may be saved across multiple directories.
2. Clip and Combine – Allows clipping an area of interest, selecting necessary variables, and merging them into one file. The spatial extent (clipping) of the data is taken from the current extent defined in the **Subset** tab.
3. Save Subset as New File – Allows saving the current subset created in the **Subset** tab as a new NetCDF (.nc) file.

[TOC]

3. TYPICAL WORKFLOWS IN Xraster

Working with the **Xraster** program can proceed in two primary ways: **data analysis** and **data engineering**.

Data analysis begins with data input, examining the data structure, and optionally obtaining a mask.

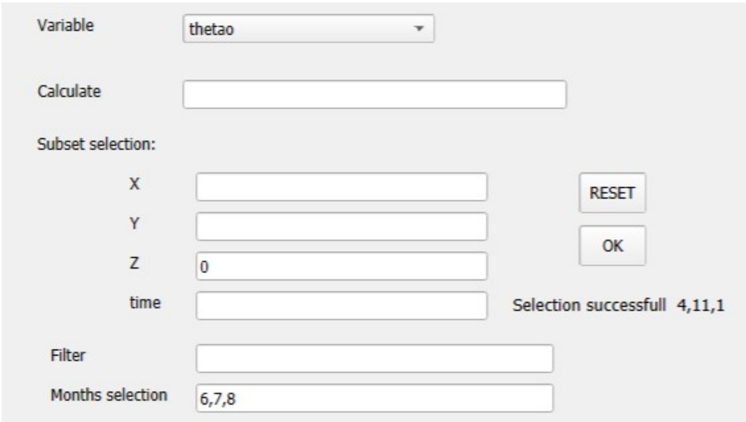


The first step is **data input** (**1**) and, if needed, **inputting a mask** to designate analysis zones (**2**). A description of the data will then be displayed (**3**). Our data dimensions are area (60 x 75 cells), depth (28 levels), and time (214 days). Each data cell contains two variables: temperature and salinity. The depth range is from 0.5 to 103 meters. The data represent conditions from March 1 to September 30, 2024.

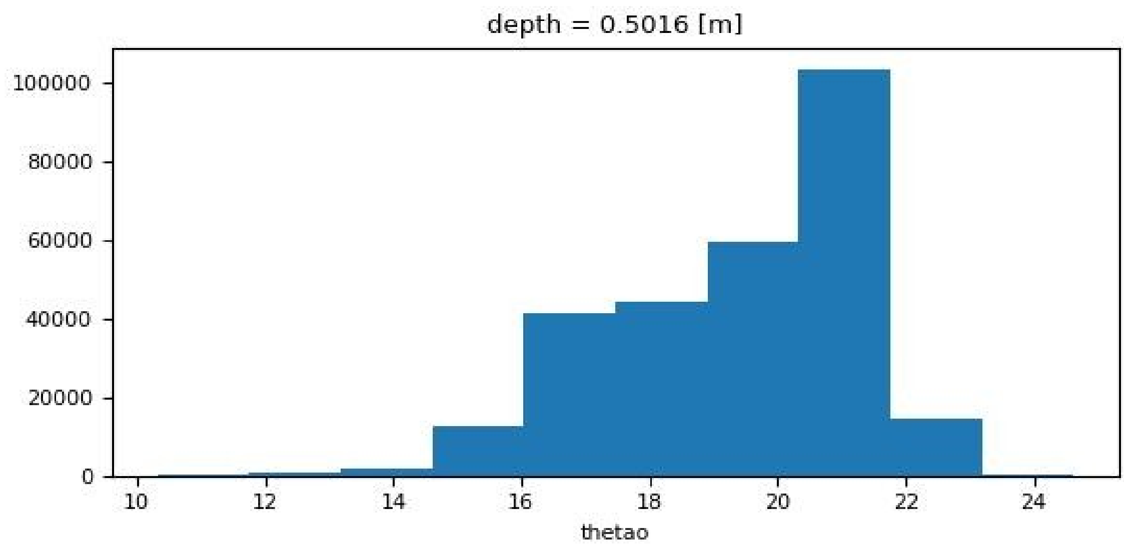
The primary data structure is a two-dimensional array (represented in GIS as a raster with dimensions 60 x 75). In total, the dataset contains 2 variables x 28 depth levels x 214 days = 11,984 rasters organized by depth and time. The input mask (if not applied, the analysis will cover the entire area or a selected point) enables the definition of geographical analysis areas. The displayed interactive mask map shows that the Vistula Lagoon has an identifier of 3.

After data input, the second step is typically **exploratory data analysis**. In this step, we aim to familiarize ourselves with our data by performing various visualizations. Each analysis process will be based on the defined data subset in the **Subset tab**.

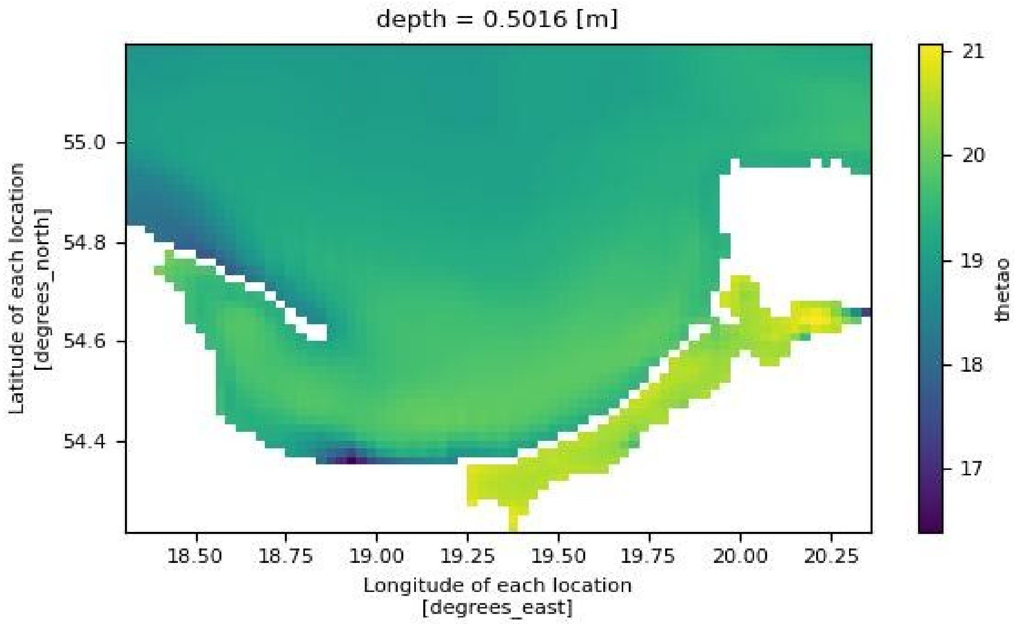
a. We select surface temperature for the months of June, July, and August.



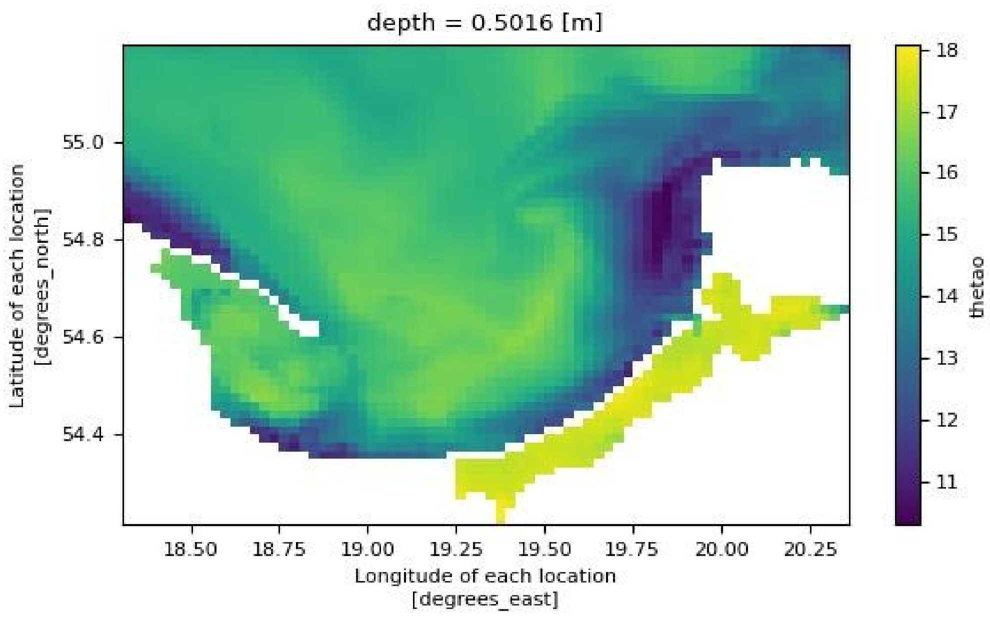
b. Display a histogram (Plot from the **Figures** tab) of surface temperature values.



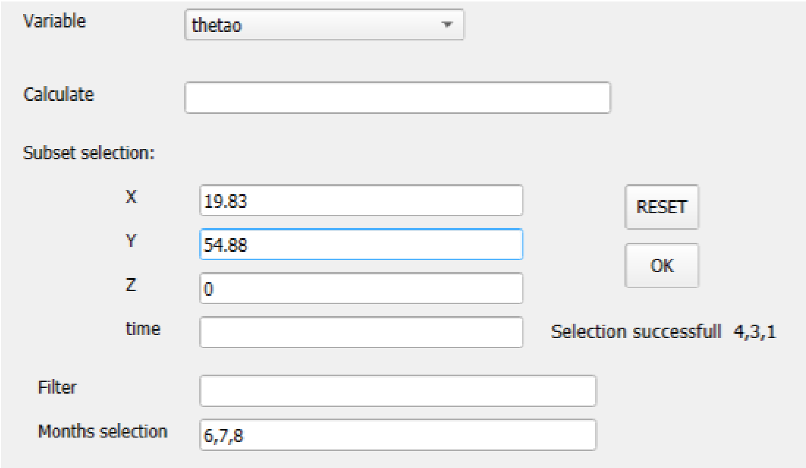
c. Display the average temperature distribution map for the entire time period (Plot - **mean** from the **Statistics** tab).



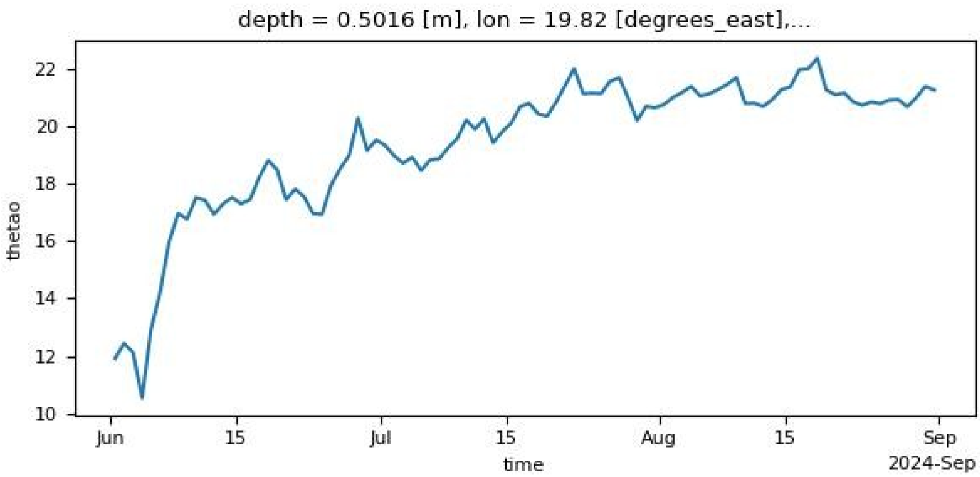
d. Display the map of minimum surface temperature (Plot - **min** from the **Statistics** tab).



Checking how the temperature has changed over time at coordinates 54.88N, 19.83E. I am modifying the data subset (**Subset** tab).



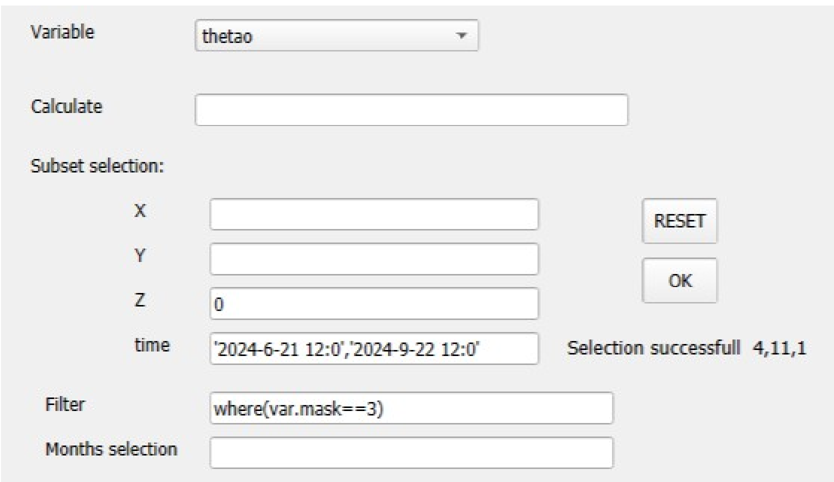
and plotting a time series (Plot from the **Figures** tab).



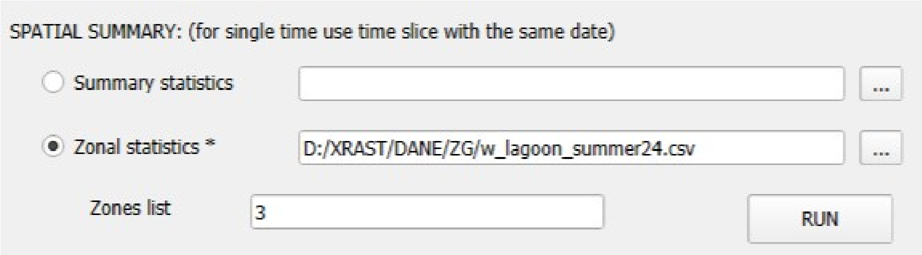
In a similar way, you can continue exploring the data.

Data analysis is based on defined queries. For example, I would like to know how many days the average surface water temperature of the Vistula Lagoon (mask ID = 3) was above 22 degrees during the period from June 21 to September 22 (summer). I am interested in which specific days these were and what the longest continuous period of this phenomenon was.

We define a new data subset (**Subset** tab).



We use the Zonal Statistics function (for zone 3 only) to create a text file with calculated statistics for each day of the analyzed period.



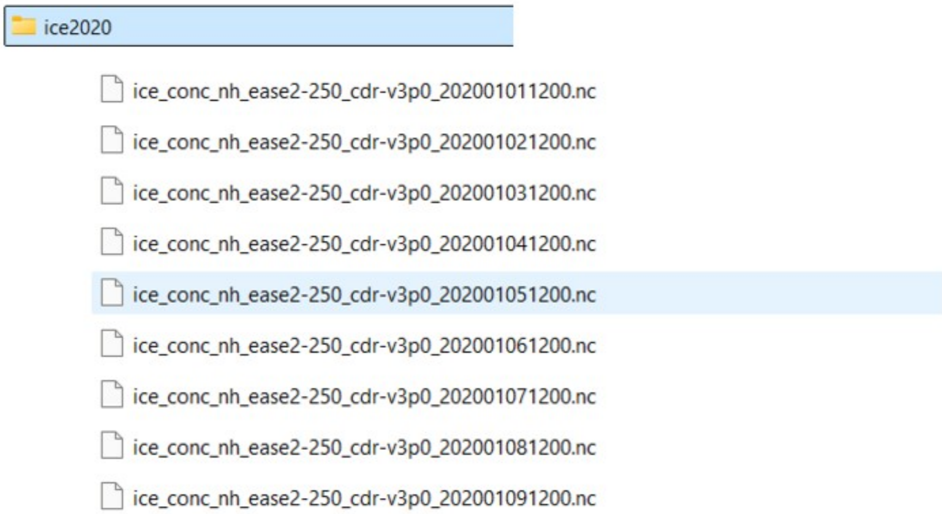
The result is a text file that can be analyzed in other programs.

1	zone	time	mean	min	max	median	std
2	3	2024-06-21 12:00:00	19.1301	17.9176	19.8132	19.1887	0.2763
3	3	2024-06-22 12:00:00	18.9035	17.6225	19.7632	18.9386	0.3971
4	3	2024-06-23 12:00:00	18.0806	16.9108	19.0385	18.1087	0.3491
5	3	2024-06-24 12:00:00	17.6978	16.3191	18.7195	17.6035	0.3674
6	3	2024-06-25 12:00:00	18.381	16.9448	19.5094	18.3451	0.3212
7	3	2024-06-26 12:00:00	19.4606	17.9824	20.6486	19.4588	0.3156
8	3	2024-06-27 12:00:00	20.635	19.0013	22.2339	20.6055	0.3791
9	3	2024-06-28 12:00:00	21.3244	19.6767	23.6299	21.2773	0.4491
10	3	2024-06-29 12:00:00	21.5457	20.0523	22.6589	21.5124	0.3412
11	3	2024-06-30 12:00:00	21.7714	20.2585	22.7662	21.7176	0.322
12	3	2024-07-01 12:00:00	21.7912	19.8334	22.8084	21.7887	0.3923
13	3	2024-07-02 12:00:00	20.754	19.1984	21.4251	20.7988	0.3048
14	3	2024-07-03 12:00:00	19.5977	18.5387	20.3986	19.6071	0.2696
15	3	2024-07-04 12:00:00	19.3826	17.9182	20.0618	19.4113	0.2577
[illegible]	3	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

Data engineering in the **Xraster** program is used to prepare data for analysis and is performed in the Clip/Join tab. The data available for download can vary in nature. Some data may be processed almost to a ready-to-use form before download. If tools like the **Copernicus Marine Toolbox** are used to generate the data, further transformation may not be necessary.

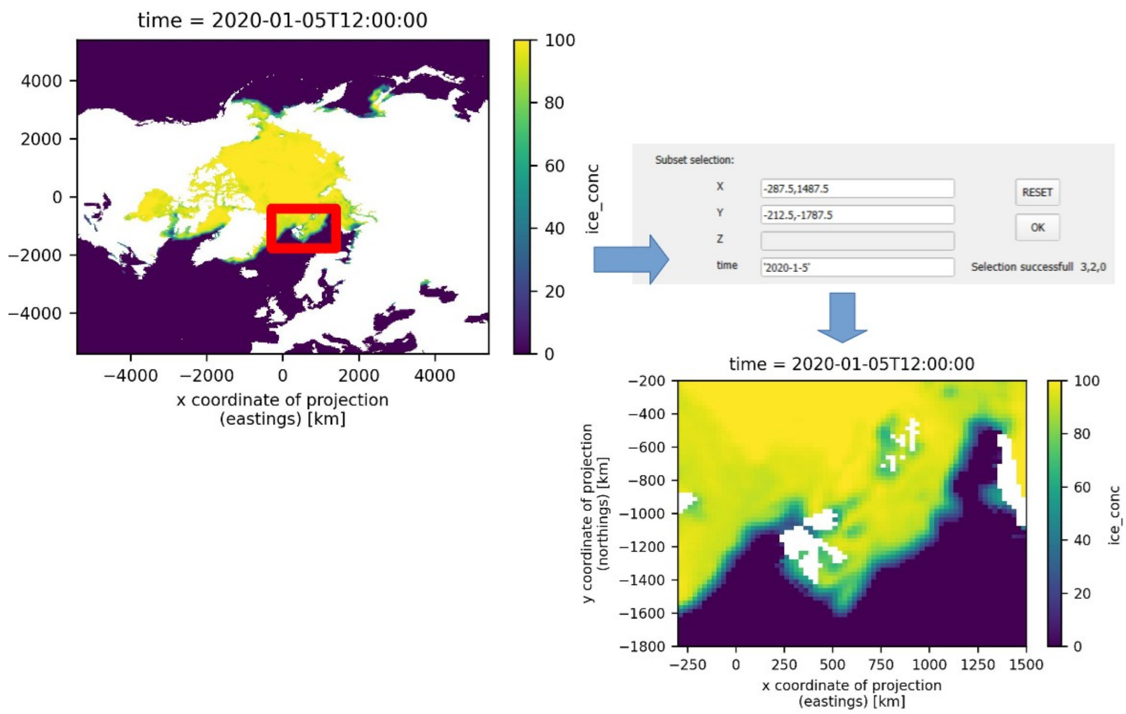
In other cases, it's possible to select a spatial and temporal window and choose specific variables as needed. Some data, when using the simplest methods, may only be downloadable for the entire area with all variables, as separate files for each day. This requires subsequent extraction of the desired area and selection of relevant variables. A crucial aspect of data preparation is optimizing its size, as data handling efficiency is closely linked to data volume.

Let's assume that the available data is downloaded in full for each day.

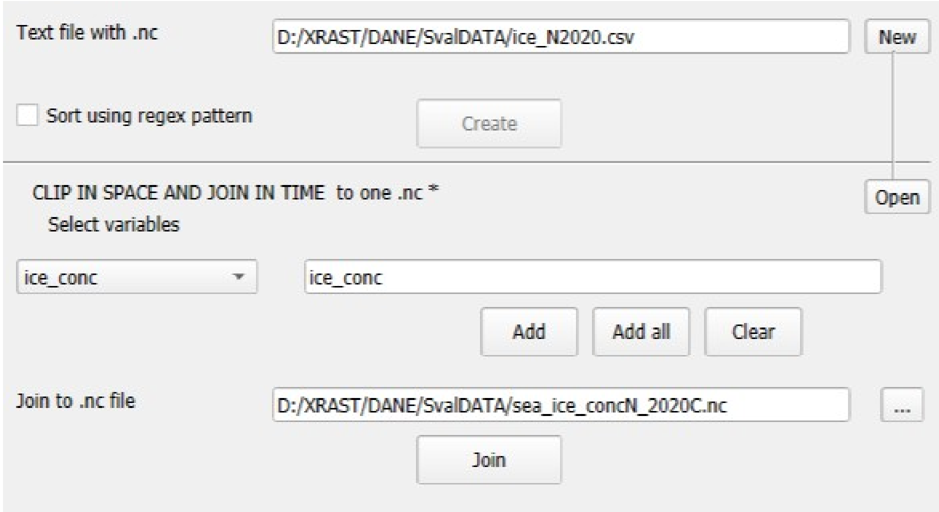


The first step is to create a text file listing all .nc files (with full paths) located in the specified directory. All subdirectories within this directory structure are also searched.

The next step is to load a single data file and display it to select the area to be clipped. You can define both the area of interest (indicated by a red rectangle) and, if necessary, the required depth range. Then, a map of the clipped area is displayed. The process will proceed as follows:



The selection parameters will be used to create subsets from individual files before they are merged. In the final step, we choose the variables we need from the available ones and perform the clipping and merging operation.



[TOC]

4. QGIS LAYERS USED IN Xraster

A raster mask representing the analysis zones

Xraster uses two types of GIS layers. An important role is played by the raster mask layer, which defines spatial zones for analysis. The mask is created as a raster layer (.tif), with the same coordinate system and the same number of rows and columns as the surface (x, y) layer in the NetCDF (.nc) data. The analysis zones are marked with identifiers having fixed values (though the mask itself can be of either integer or float type). An example of a mask for the Gdańsk Bay is shown below.

The mask is originally in the GCS coordinate system, but it has been displayed in the UTM 33N coordinate system. It contains four zones corresponding to specific geographic areas. Areas outside the zones have an identifier of zero. It is best not to use NoData values on the mask and replace them with zeros. The mask is used in the program as,

- in the **Subset** tab to define a data subset, taking the zones into account,

- for performing the Zonal Statistics function in the **Statistics** tab**,**
- in all map exports to raster as .tif files (the mask is used as the source of georeferencing).

Vector point layer

This layer is used for data extraction at specific points.

The figure also shows it with a polygon layer outlining the Gdańsk Bay. The first field in the attribute table must be of integer type and contain a unique point identifier (the field name is arbitrary).

	ID
1	10
2	11
3	12
4	13
5	14
6	15
7	16

Similarly to the mask, the point layer must have the same coordinate system as the NetCDF (.nc) data.

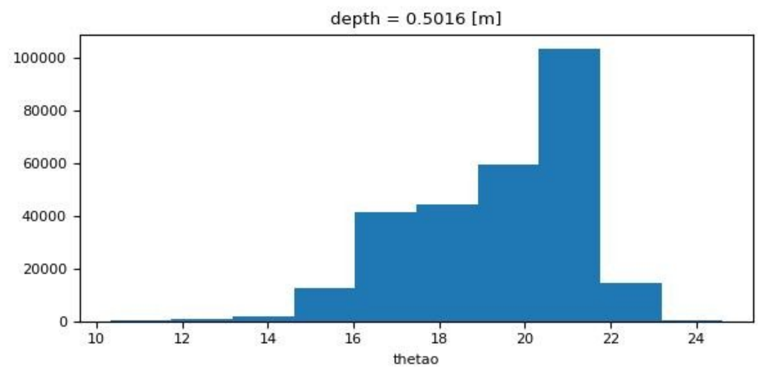
[TOC]

5. PLOTS, SAVE AND EXPORT PLOT

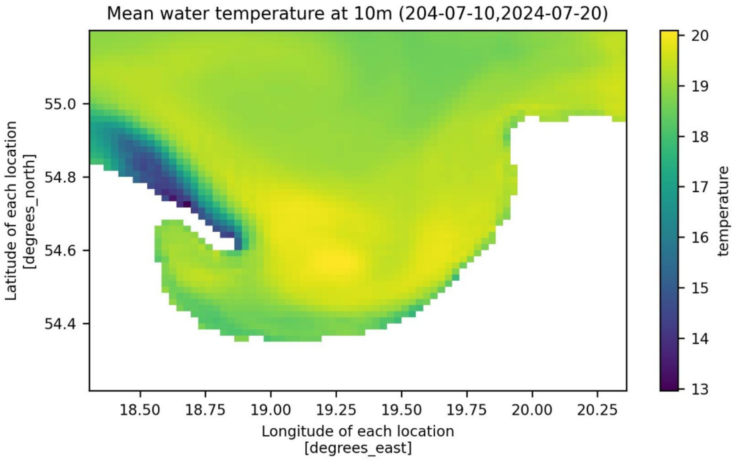
Images generated by Xraster

Xraster generates images using the methods and capabilities of the xarray package. The xarray package selects the type of image based on a subset defined by selection or analysis results. The following types of images are created;

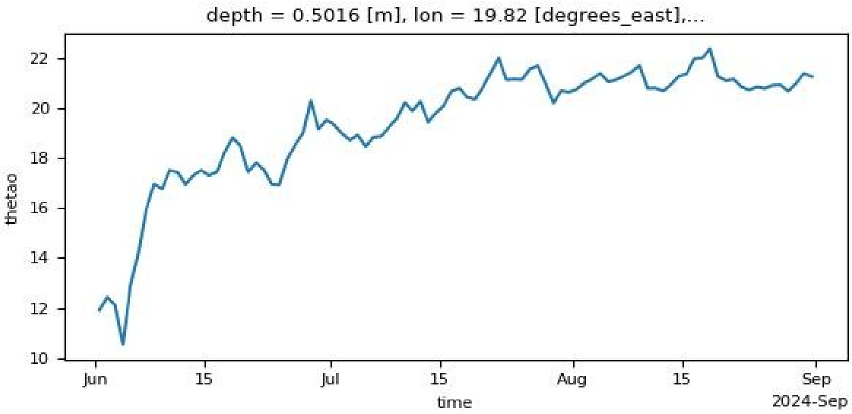
Histogram,



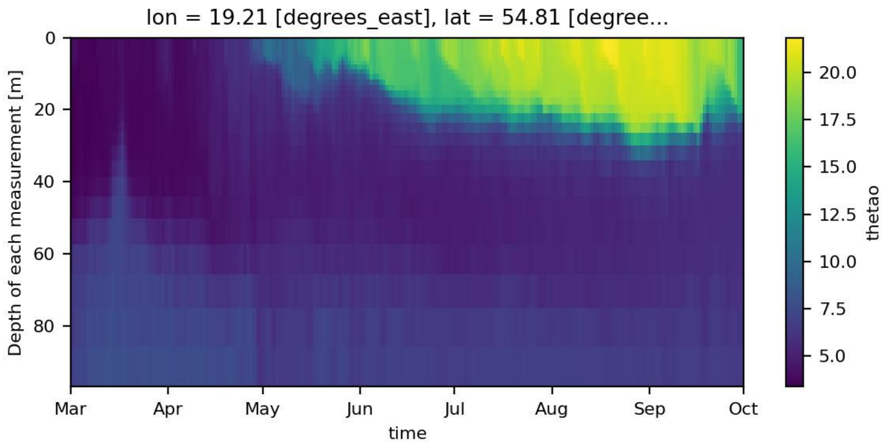
Map,



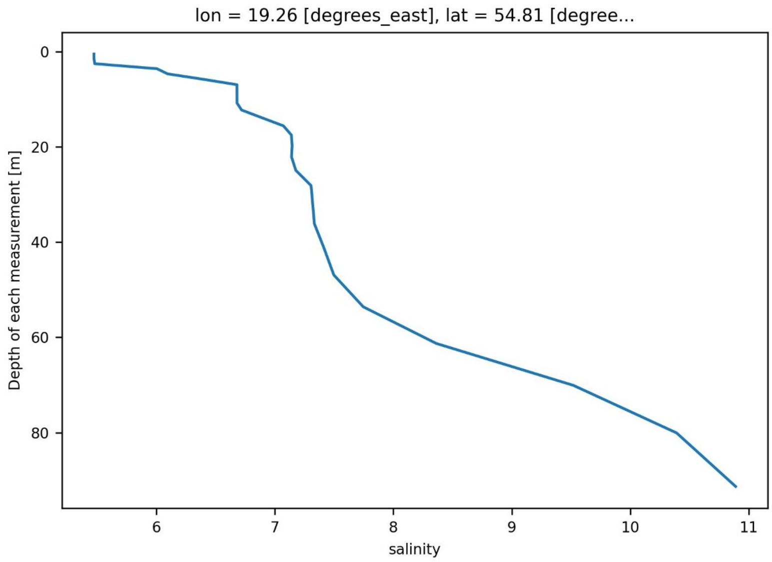
Time series,



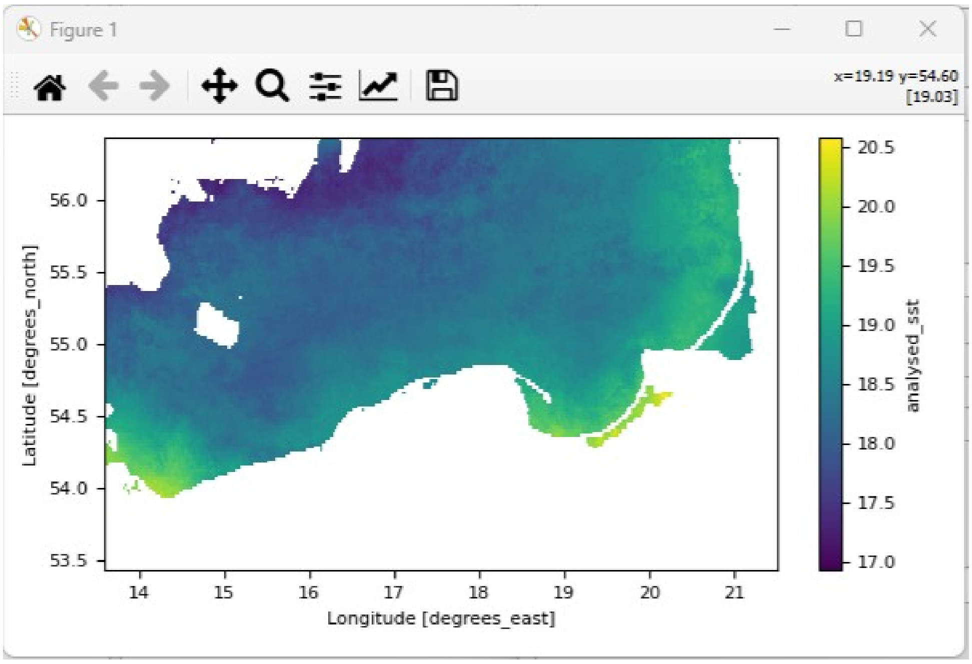
Depth-time-cross-section,



Vertical profile,



To display images in QGIS, the Matplotlib GUI window is used.

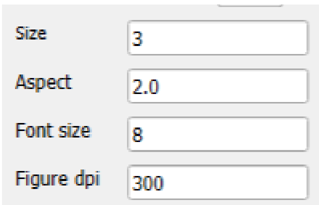


This window displays the image along with axis labels and a color scale with descriptions. The top toolbar includes icons for interacting with the image, as well as coordinates for the cursor position on the image and the corresponding variable value.

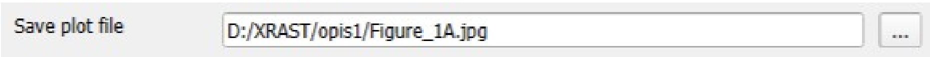
The first five icons are used for navigation: return to the original view (1), previous or next image (2, 3), panning (4), and zooming with a rectangular selection (5).

Save and Export plot

The last icon, number 8 (a floppy disk), allows you to save the image in several formats at screen resolution. Images can also be exported as .jpg files at higher resolutions and with specific parameters, such as size, aspect ratio, font size, and dpi, which are defined on the **Figures** tab.



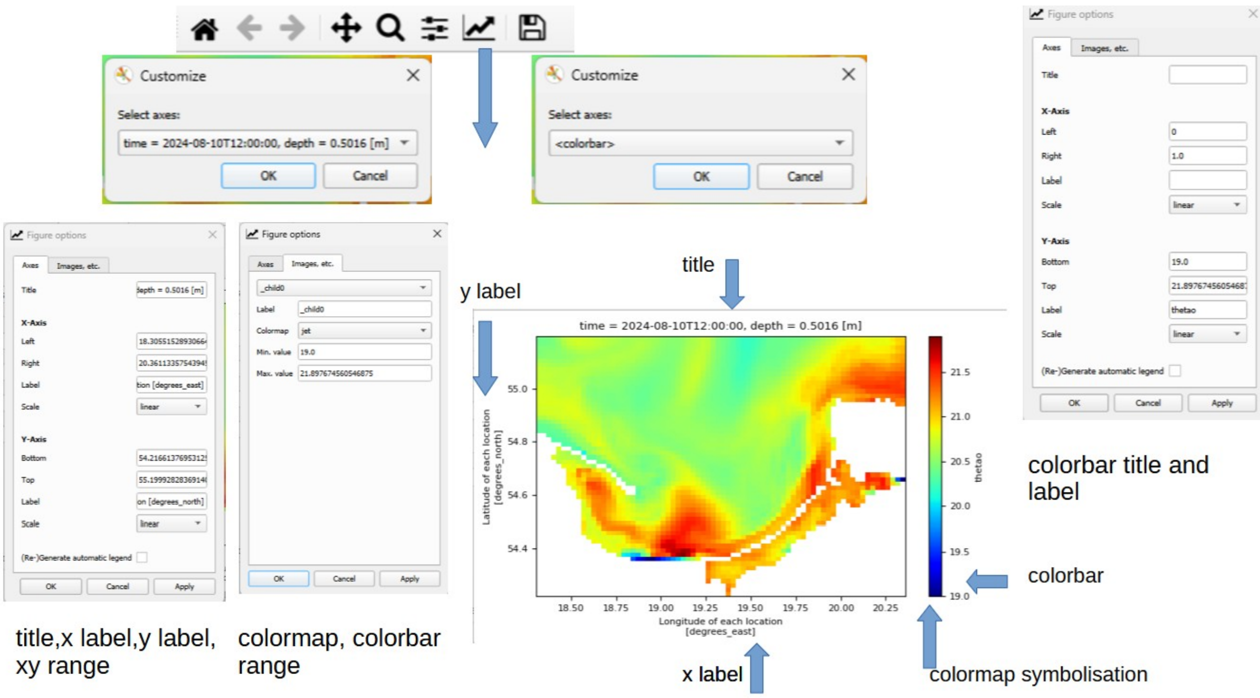
This is done using a separate 'Save plot (file)' option in the program.



[TOC]

6. PLOT CUSTOMISATION

The Matplotlib GUI window allows customization of created images. The penultimate icon enables editing of axis labels, titles, and the styling of lines and rasters. Below are methods for customizing maps, though this process applies to all types of images

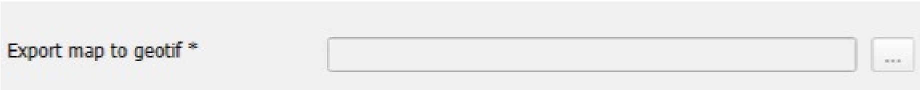


We begin the process by selecting the first option (the window on the left). Here, it's possible to change the map title and labels for both axes. For maps, you can also define the color scale (colormap) and its range. The second option adjusts based on the first selection and allows you to add a description to the color scale. Customization applies to both image-saving options.

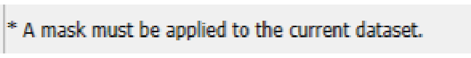
[TOC]

7. EXPORT GEOTIFF RASTERS

Exporting a two-dimensional map to a raster layer is possible after creating the map in the **Figures**, **Statistics**, and **Group by Time** tabs.



This option will only be active if the map was created using PLOT or PLOT AT BOTTOM, and if a mask was previously applied.



The size of the created raster layer and its georeferencing are derived from the mask.

[TOC]

8. EXPORT TEXT FILES

Exporting text files (.csv) occurs in the following situations:

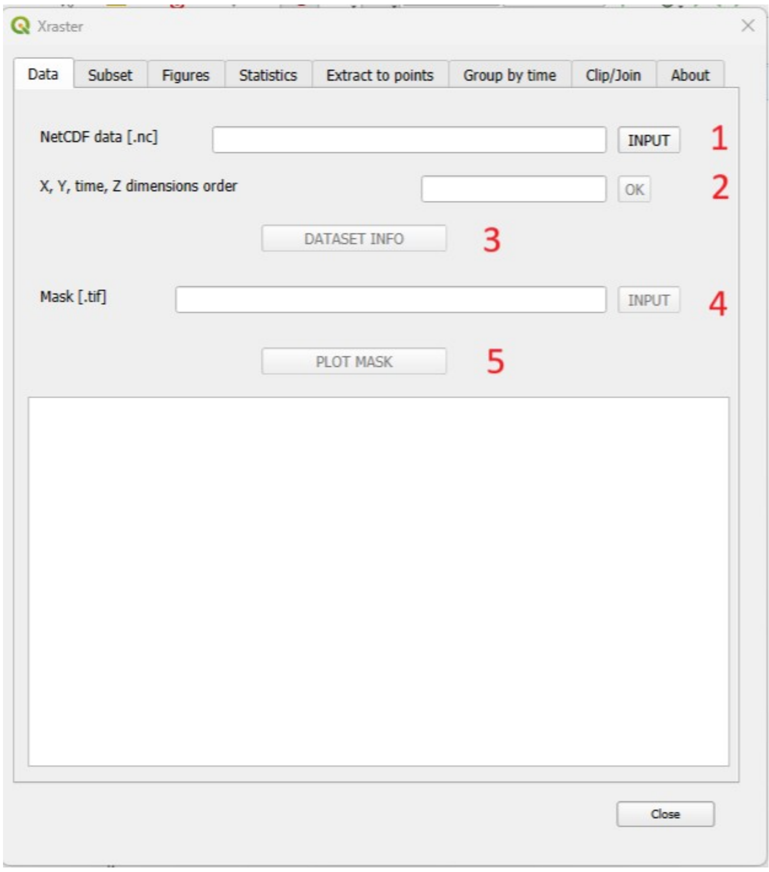
- Exporting data from images to a text file occurs for the following types: time series, vertical profile, depth-time cross-section (**Figures**, **Statistics**, **Group by Time**).
- Eksport rezultatów spatial summary: summary statistics i zonal statistics. Dla każdego czasu obliczone są statystyki i zapisywane w pliku tekstowym. W przypadku zonal statistics dodatkowo zapisywany jest numer strefy.
- Exporting data in the **Extract to Points** tab exports point data along with all its dimensions.
- The Clip/Join tab creates a text file containing all NetCDF (.nc) files in a given directory, including subdirectories. Additionally, the data can be sorted by time.

[TOC]

USING Xraster (Tabs)

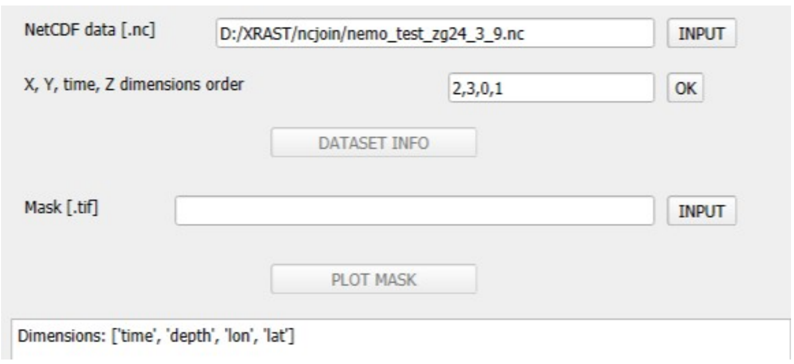
---

1. Data



(1) The work begins by importing a NetCDF (.nc) file into the program, which contains a time series (t) of surface layers (x, y). If a valid file is not provided, a message will appear in the text field. , !!!! Not valid NetCDF [.nc] file or no time dimension !!!!.

After importing a valid file, the order of the equivalents for the dimensions X, Y, time, and Z appears in window **(2)** in the text field below, listed as Dimensions: [.....].



The dimensions are displayed in the text field. It is necessary to check if the order suggested by the program is correct; if not, it should be adjusted. In the example provided, X (lon) is in the second position, Y (lat) in the third, time in the first position, and Z (depth) in the second. If the order is correct, confirm it by pressing the OK button. This will import the data and display information about the file in the text window. This information can also be accessed via the DATASET INFO button **(3)**. For example, the information appears as follows:

Dimensions:

time,214
depth,28
lat,60
lon,75
Coordinates:
depth,0.5016,103.8247
lon,18.3194,20.375
lat,54.2083,55.1916
time,2024-03-01 12:00:00,2024-09-30 12:00:00
depth,0.5016,103.8247
Variables:
thetao [degree_Celsius]
so [1e-3]
The order of dimensions and coordinates may differ from the order of dimensions defined earlier. This file contains several key pieces of information:

- 214 time moments

- 28 depth levels

- Surface dimensions of 60 rows by 75 columns

Next, the ranges of each coordinate are presented. There may be more coordinates than dimensions, but all dimensions must be represented here. Additionally, all variables and their units are labeled.

NOTE – If values of -1 appear in window (2), it is likely that this file cannot be used.

Mask input is optional (4). Adding a mask enables:

- its use in creating subsets (**Subset**),

- exporting surface maps to raster layers in GIS as GeoTIFFs (**Figures**, **Statistics**, **Group by time**),

- performing zonal statistics operations (**Statistics**).

The mask must have the same number of rows and columns and the same georeference as the surface data (x, y) in the imported (.nc) file. If the number of rows and columns does not match, a message appears:

`!!!! Mask has different dimension x=142, y=62 !!!!`

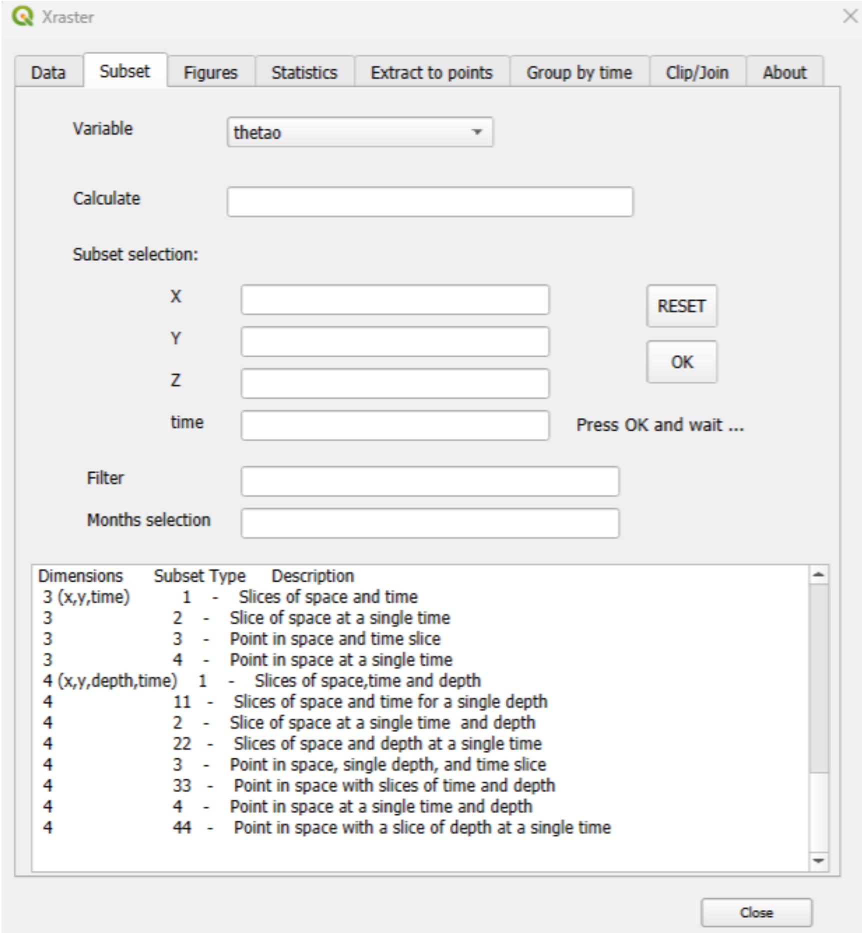
After entering a valid mask, it can be displayed (PLOT MASK **(5)**). Re-displaying the file information (DATASET INFO) confirms that the mask has been added to the coordinates.

Coordinates:

depth,0.5016,103.8247
mask,0.0,4.0
lon,18.3194,20.375
....
[TOCTabs]

2. Subset

Creating a data subset is a crucial step in the analysis. The subset is created for a selected variable. This variable, even within a 4-dimensional dataset, can have a 3-dimensional nature (e.g., it may refer exclusively to the sea surface).



The basic principles of creating subsets are outlined in the text window. The program works with data in three dimensions (surface—defined by dimensions x and y and time) and four dimensions (surface—defined by dimensions x and y, depth, and time). The primary method for defining a subset is to specify ranges (slices) in space. This applies to the surface (restricted by coordinate limits), depth range, and time interval. It is also possible to define a subset at a specific surface point, at a given depth, and at a specific time. If data is unavailable for a particular surface point, depth, or time, the closest available data will be used.

In addition to spatial operations, data filtering is available. Filtering can apply to a specific variable or use a mask, and it can also involve all variables. This allows the creation of a subset containing data only for a specific geographical area. Furthermore, data can be filtered by month, enabling selection of data for specific months only.

In the program, the type of subset is identified by a code, where the first digit represents the data dimensionality, and the second digit specifies the method for defining the dimension. For 3-dimensional data, four types of subsets are possible, while for 4-dimensional data, eight types are available.

Description	Subset Type
Slices of space and time	3-1
Slice of space at a single time	3-2
Point in space and time slice	3-3
Point in space at a single time	3-4
Slices of space,time and depth	4-1
Slices of space and time for a single depth	4-11
Slice of space at a single time and depth	4-2
Slices of space and depth at a single time	4-22
Point in space, single depth, and time slice	4-3
Point in space with slices of time and depth	4-33
Point in space at a single time and depth	4-4
Point in space with a slice of depth at a single time	4-44

Additionally, a third digit (1/0) is used to indicate the presence or absence of a mask. This notation is then utilized in analytical functions, informing whether they are available for a particular data type.

The fields used to define the subset are described below, with the first two fields designated for selecting and modifying the variable.

Variable – The variable is selected from a dropdown list.

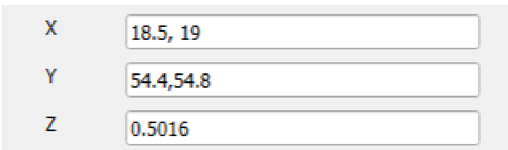
Calculate – This option recalculates the variable using a simple expression in the form of ***variable rest_of_expression*** (only the *rest_of_expression* is entered). This feature was introduced, among other things, for converting Kelvin to Celsius. For example, to convert from Kelvin to Celsius, simply enter `-272.15` (since the full expression would be `temp[K] - 272.15`). The **Calculate** option also allows recalculating a variable using another variable from the selected dataset.

For example, if the dataset contains the variables VHM0 and VTM02, and we need to calculate the expression `0.5 * VHM0$^2$* VTM02`, the modification of the input variable VHMO (variable) would involve entering the expression `**2 * 0.5 * ds2.VTM02` in the Calculate field. The full expression will then be: `VHM0**2 * 0.5 * ds2.VTM02`. Here, `ds2` represents the dataset symbol, and `.VTM02` refers to any variable within the dataset. The notation `ds2.[variable]` can also be used for filtering.

X – The range for the x-coordinate is entered as: Xstart, Xend*(e.g., 18.2345, 19.0), or for a single point, just one value (e.g., 18.2345). If no value is entered, it represents the entire range. The full range can be reset using the RESET button. Note that when entering a range of values, a comma is used.

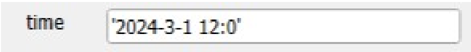
Y, Z – These are entered in the same manner as the X range.

Method of entering,



There are situations when the X, Y range must be entered from a larger value to a smaller one (in the case of negative coordinates). This rule can be verified using the RESET function. If, after entering the range and using RESET, the message *Selection failed* appears, the signs should be swapped. For example, change the input from 89.9375, -89.9375 to -89.9375, 89.9375.

time – The time is entered in the following format: 'YYYY-M-D', where YYYY is the full year, M is the month without a leading zero, and D is the day without a leading zero. For example: a time range: '2004-7-1', '2004-7-31', a single time: '2004-7-1'. Method of entering,



Filter – The filter function uses **where()** along with comparison and logical operators (such as `or` as `|` and `and` as `&`). It can be applied to a variable **var**, the mask **var.mask**, or any variable from the dataset **ds2.variable.** Here are some example uses:

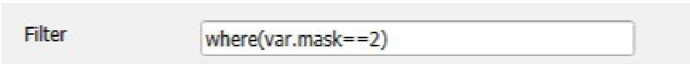
where((var.mask == 1) | (var.mask == 2)) This selects areas where the mask is either 1 or 2.

-where(var <= 5) This filters the variable to include only values less than or equal to 5.

where(ds2.variable_name > 3) This filters the dataset to include values where the variable *variable_name* is greater than 3.

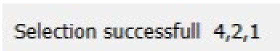
These filtering expressions help narrow down data based on specified conditions.

Method of entering,



Month selection – This feature restricts the data to only the months specified in a list format. To select data from December, January, February, and March, you would enter the months as follows:

Selection is made by pressing the OK button. A message appears with a code corresponding to the subset type, e.g.



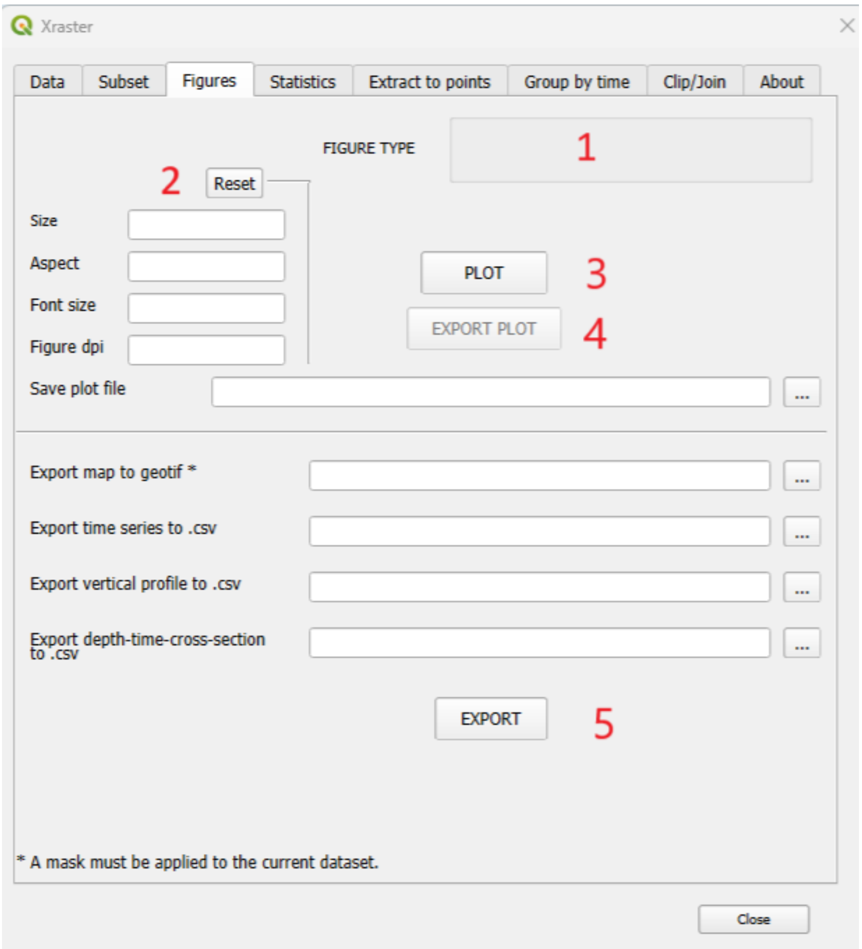
Four dimensions, type 2 – surface at a specific depth at a specific moment in time (a data mask is available).

If the selection process fails, the subset will contain all data from the dataset, and a message will appear,

[TOCTabs]

3. Figures

This tab is used for visualizing the data subset and conducting exploratory data analysis. It also allows exporting any surface (x, y) of the data subset to a GIS raster layer (a mask is required) and saving data from the chart as text files (for time series, vertical profiles, and depth-time cross-sections).



After data selection and subset creation, the figure type is determined in the FIGURE TYPE panel **(1)**, based on the created data subset. The types of subsets, their codes, and corresponding figure types are summarized in a table at the end of the subsection. To generate the plot, press PLOT **(3)**. The figure will appear in the Matplotlib GUI window with default parameters **(2)** specified in the fields: **Size**, **Aspect**, and **Font size**. The Size and Aspect can be adjusted by resizing the window or changing the values in the fields, while Font size can only be adjusted by changing its field value. All values can be restored to default by pressing the RESET button.

The figure can be saved directly from the GUI window (by clicking the diskette icon) in screen resolution or saved with a higher resolution by entering a name in the **Save plot file** field and pressing EXPORT PLOT **(4)**, with the resolution defined by the **Figure dpi** field. Additional customization options for the figure are described in Section 3.6.

At the bottom of the tab, there are four export fields with buttons for entering a file name. Only the usable fields are active at any given time. Enter the file name and click the EXPORT button to proceed.

- **Export map to geotiff** – if the figure is a map (xy surface) and a mask has been applied, it can be exported as a raster .tif layer.

- **Export time series to .csv** – if the figure is a time series, the figure data can be exported to a .csv file as follows:

```
time,thetao
2024-03-01 12:00:00,4.0354
2024-03-02 12:00:00,4.2373
2024-03-03 12:00:00,4.3542
2024-03-04 12:00:00,3.5257
2024-03-05 12:00:00,3.449
2024-03-06 12:00:00,3.5584
..........................................
```

- **Export vertical profile to .csv** – if the figure is a vertical profile, the figure data can be exported to a .csv file in the following format:

```
depth,thetao
0.5,19.9852
1.52,19.9852
2.55,19.9852
3.6,19.9851
4.68,19.985
5.8,19.9848
6.96,19.9847
.................
```

- **Export depth-time-cross-section to .csv** – if the figure is a depth-time cross-section, the figure data can be exported to a .csv file in the following format:

```
time,depth,thetao
2024-08-01 12:00:00,0.5,19.8857
2024-08-01 12:00:00,1.52,19.8849
2024-08-01 12:00:00,2.55,19.8833
2024-08-01 12:00:00,3.6,19.8809
2024-08-01 12:00:00,4.68,19.8701
2024-08-01 12:00:00,5.8,19.8446
2024-08-01 12:00:00,6.96,19.8138
2024-08-01 12:00:00,8.17,19.7778
2024-08-01 12:00:00,9.45,19.7678
2024-08-01 12:00:00,10.81,19.7648
2024-08-01 12:00:00,12.27,19.762
2024-08-01 12:00:00,13.86,19.7502
2024-08-01 12:00:00,15.6,19.6345
.....................................................
```

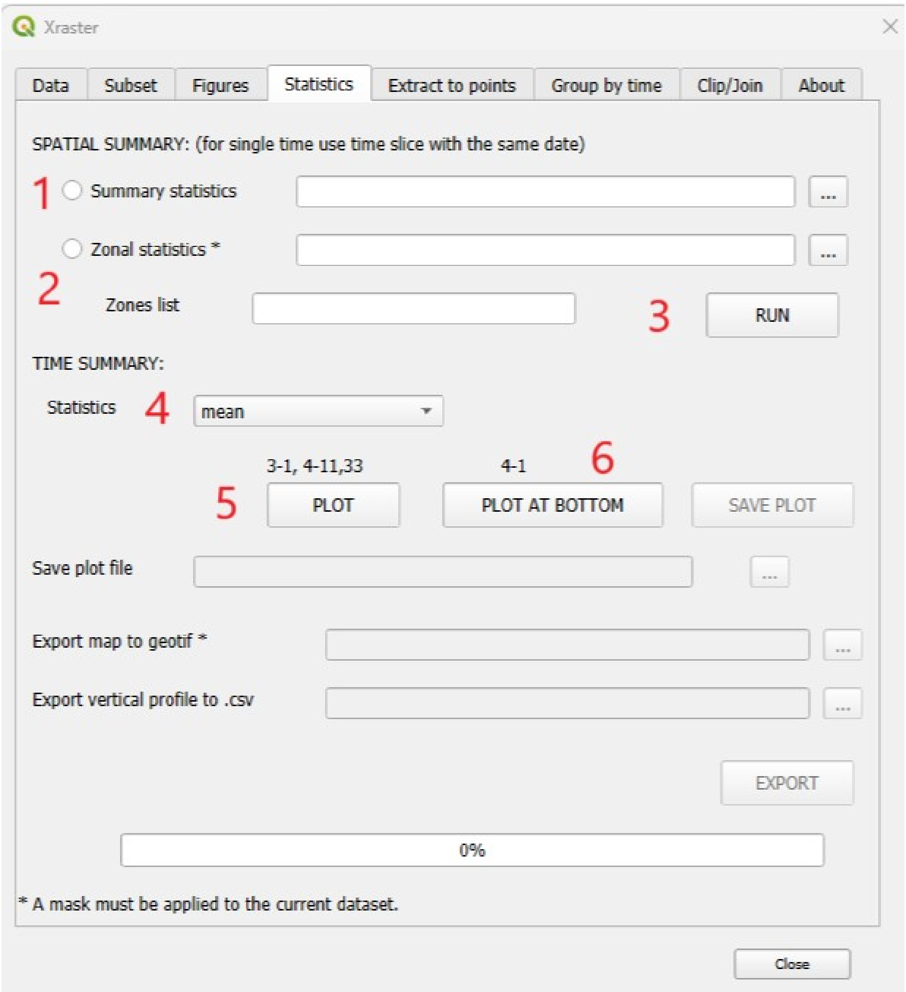
Types of Subsets, Codes, and Corresponding Figure Types

Description of subset	Subset Type	Figure
Slices of space and time	3-1	Histogram
Slice of space at a single time	3-2	Map (raster)
Point in space and time slice	3-3	Time series
Point in space at a single time	3-4	-
Slices of space,time and depth	4-1	Histogram
Slices of space and time for a single depth	4-11	Histogram
Slice of space at a single time and depth	4-2	Map (raster)
Slices of space and depth at a single time	4-22	Histogram
Point in space, single depth, and time slice	4-3	Time series
Point in space with slices of time and depth	4-33	Depth-time-cross-section
Point in space at a single time and depth	4-4	-
Point in space with a slice of depth at a single time	4-44	Vertical profile

[TOCTabs]

4. Statistics

This tab is used for calculating spatial (SPATIAL SUMMARY) and temporal (TIME SUMMARY) statistics. Spatial statistics are calculated from the data across the entire spatial domain (x, y, z) separately for each time point. Temporal statistics are calculated separately for each spatial location from data that varies over time.



In SPATIAL SUMMARY **Summary statistics (1)** analyzes all data from the subset for a given moment in time and calculates statistics for it. The results are saved separately for each time point in a text file in the following format:

time,mean,min,max,median,std,sum

2024-07-01 12:00:00,5.7482,4.781,7.2044,5.6085,0.5863, 43967.62

2024-07-02 12:00:00,5.738,4.7833,7.1345,5.6021,0.574, 43889.88

2024-07-03 12:00:00,5.7299,4.7878,7.1024,5.5954,0.5684, 43828.16

2024-07-04 12:00:00,5.7415,4.7964,9.2777,5.6069,0.5875, 43916.54

2024-07-05 12:00:00,5.7397,4.8054,8.4586,5.6072,0.5876, 43903.03

You need to enter the file name in the field and click RUN **(3)**.

Similarly, **Zonal statistics** **(2)** works in the same way, but it also divides the data into zones based on the mask. In the **Zones list** field,

you need to enter the zone numbers for which the calculations should be performed. The result is in the following format:

zone,time,mean,min,max,median,std,sum

1,2024-07-01 12:00:00,20.2142,18.2626,21.2628,20.2573,0.6859, 2587.42

1,2024-07-02 12:00:00,19.5465,17.6327,20.4764,19.5872,0.6289, 2501.95

1,2024-07-03 12:00:00,18.8658,17.8467,19.6353,18.7756,0.4766, 2414.82

1,2024-07-04 12:00:00,18.5177,17.6609,19.1908,18.5182,0.3849, 2370.27

1,2024-07-05 12:00:00,18.3142,17.9012,18.9311,18.2979,0.2801, 2344.22

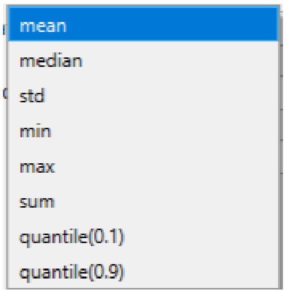
2,2024-07-01 12:00:00,17.3837,8.2579,19.9589,18.1639,2.2035, 16896.95

....

3,2024-07-04 12:00:00,19.3774,17.845,20.0618,19.4105,0.2616, 9184.88

3,2024-07-05 12:00:00,19.0395,17.7076,19.7524,19.0459,0.2543, 9024.71

TIME SUMMARY allows you to calculate statistics over time for surfaces (x, y, time) and points (z, time). The following statistics **(4)** can be selected:



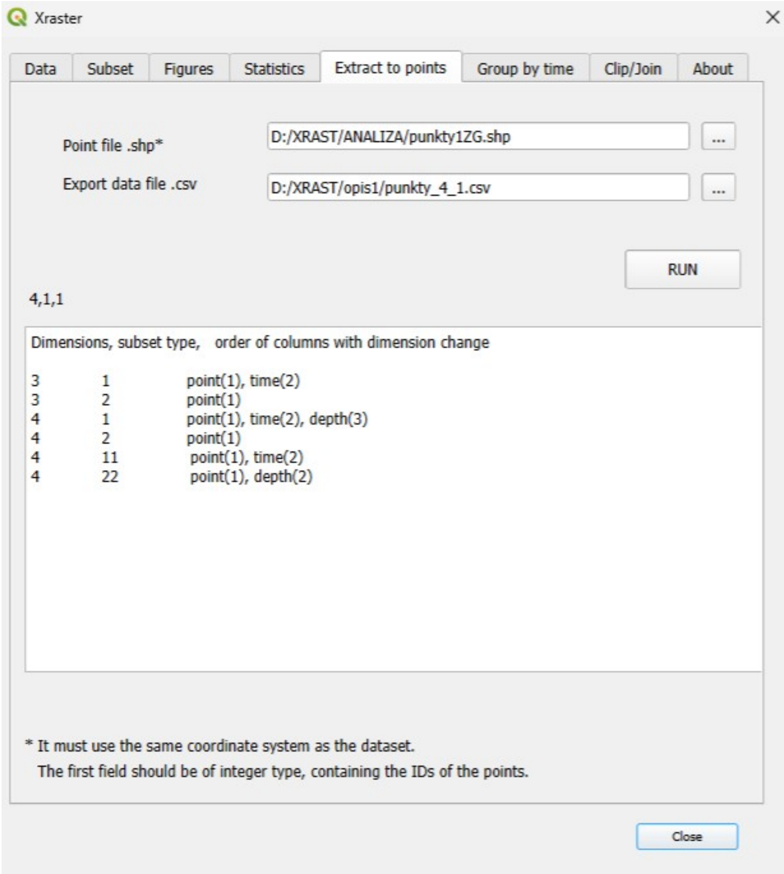
The first option (from the surface) produces a map, and the second option (from points) generates a vertical profile. Both products are created by pressing the PLOT button **(5)** . The map and vertical profile can be saved as a figures with the specified resolution (on the **Figures** tab) by using the SAVE PLOT button. Additionally, the map can be exported to a GIS raster layer in TIFF format (**Export map to geotif**), and the data from the vertical profile can be saved to a text file (**Export vertical profile to .csv**) using the EXPORT button.

There is also an additional option to create maps of the variable statistics on the seabed from data of type (4-1). This option works only for the following statistics: <u>mean, median, std, min, max</u>. Calculations are triggered by the PLOT AT BOTTOM button **(6)**. Similarly to the map, the result can be saved as a drawing (SAVE PLOT) or exported (EXPORT) to a GIS raster layer.

[TOCTabs]

5. Extract to points

The tool performs a common operation in GIS systems: data extraction (retrieval) from rasters at specific points. Due to the multidimensional nature of the data, this operation is more complex.



Two fields are used for data input. The *Point file (.shp)* is for entering point data—locations where data extraction from the dataset subset will occur. The point layer must have the same coordinate system as the dataset. The first field (with a name of your choice) must be an integer type and contain a unique identifier for each point. The second field is for entering the name of the output text file. This text file will have columns with the point number, followed by all dimensions of the given data type. For example, an extraction from data type 4-11, representing *slices of space and time for a single depth*, will yield a file formatted as follows:

IDpoint,X,Y,time,depth,thetao

10,18.9664,54.7731,2024-07-01 12:00:00,0.5,18.7793

10,18.9664,54.7731,2024-07-02 12:00:00,0.5,18.4781

10,18.9664,54.7731,2024-07-03 12:00:00,0.5,18.2965

10,18.9664,54.7731,2024-07-04 12:00:00,0.5,18.7907

....

11,18.5127,54.7078,2024-07-01 12:00:00,0.5,20.6068

11,18.5127,54.7078,2024-07-02 12:00:00,0.5,20.067

11,18.5127,54.7078,2024-07-03 12:00:00,0.5,18.9485

....

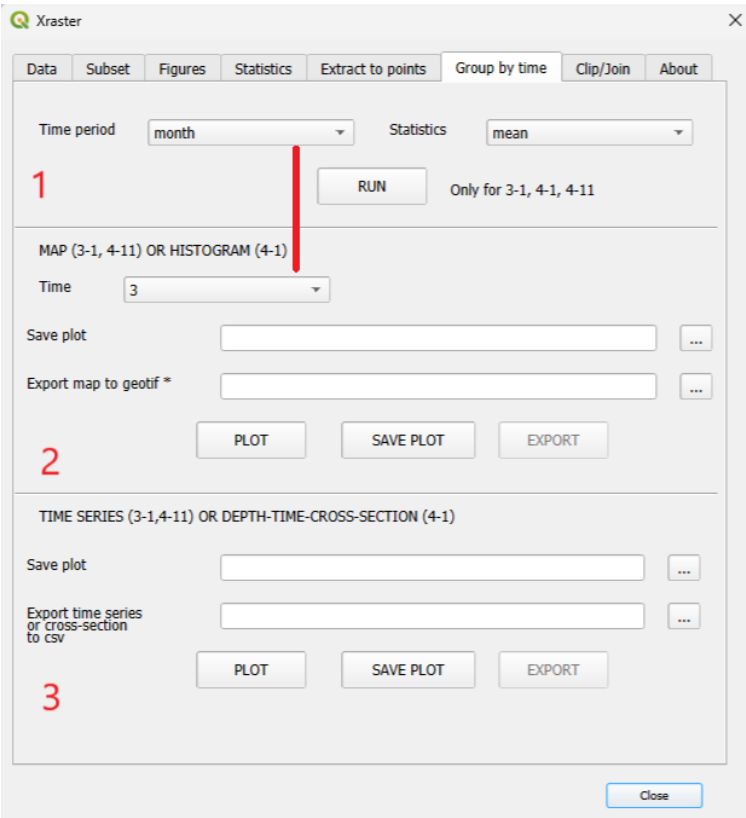
In the first column, the point number is listed, followed by its coordinates. For each point, subsequent columns contain the variable's value for each date. The structure of the output text files is summarized in the table below. This format can be applied to all subsets except those specific to a single point. A schematic of the text file format is also presented on the panel of the interface.

Subset description	Subset Type	Order of columns with dimension change
Slices of space and time	3-1	point(1), time(2)
Slice of space at a single time	3-2	point(1)
Slices of space,time and depth	4-1	point(1), time(2), depth(3)
Slices of space and time for a single depth	4-11	point(1), time(2)
Slice of space at a single time and depth	4-2	point(1)
Slices of space and depth at a single time	4-22	point(1), depth(2)

[TOCTabs]

6. Group by time

The panel consists of three sections and is used to perform statistical analysis for groups defined by year, month, day, or climatic seasons (JJA - June to August, SON - September to November, DJF - December to February, MAM - March to May). Separate groups are created from the data subset based on specific keys (e.g., individual months), and statistics (mean, median, standard deviation, minimum, maximum, sum) are calculated separately for each group. Once the statistics are computed, the keys with calculated statistics are combined into a single subset, which can then be visualized and exported.



Grouping can be applied to subsets of data like *slices of space and time* for three-dimensional data, and to *slices of space, time, and depth* or *slices of space and time for a single depth* for four-dimensional data. The analysis always starts at the top section of the panel **(1)**, where the group key (**Time period**) is defined, and the statistic to be calculated for each group (**Statistics**) is selected. After pressing the RUN button, a subset of groups is created, and in the middle section of the panel **(2)**, the **Time** field will display all groups (e.g., those calculated for each month).

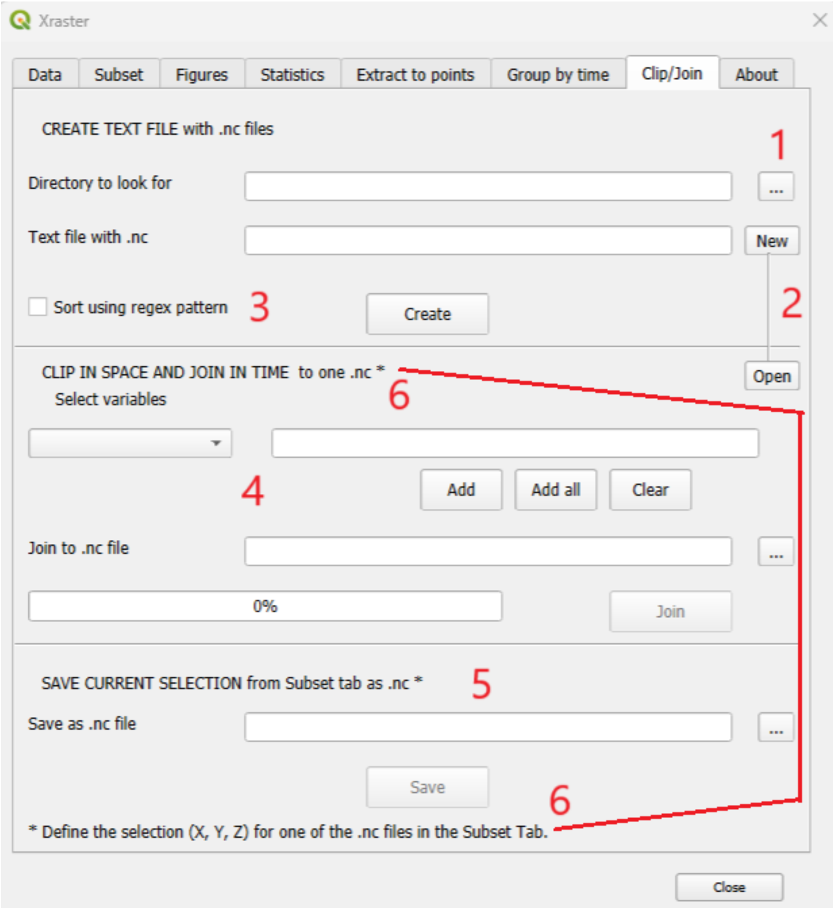
In the second middle section, you can create a map or histogram for the selected group in the **Time** field by using the PLOT button. The map or histogram image can be saved with SAVE PLOT. If there is a mask, the map can also be exported as a GIS raster layer using the EXPORT button.

In the lower section of the panel **(3)**, you can visualize the analysis results for all groups combined. For three-dimensional data and data at a specified depth (types 3-1, 4-11), this will produce time series. For data of type 4-1, a depth-time cross-section is created (not for climatic seasons; if there is only one group, it will be a vertical profile). These visualizations can be saved with SAVE PLOT, and data can be exported to text files using the EXPORT button.

[TOCTabs]

7. Clip/Join

In this panel, two tasks related to preparing and saving NetCDF (.nc) data are performed. The first task involves preparing a NetCDF file tailored to the specific project from the retrieved data, while the second task saves the current data subset in NetCDF format.



Depending on the method used, the data retrieved may vary in format and usability. Using the Copernicus Marine Toolbox, for instance, can provide a ready-to-analyze dataset as a NetCDF (.nc) file. However, other methods may yield files covering large areas, containing many variables, or spanning separate time periods (e.g., days, hours), making them less practical for direct use in the program. Effective analysis requires merging these files into one while minimizing its size, which can be achieved by limiting the area and the number of variables included.

The panel performs three tasks, divided into sections by horizontal lines. The first section allows you to compile the paths to the downloaded files into a single text file, assuming they may be saved within a directory structure. The second section enables you to extract the area of interest, specify necessary variables, and merge them into a single file. The spatial extent (crop) is taken from the current range defined in the **Subset** panel.

The third section, independent of the previous two, allows you to save the current subset created in the **Subset** panel as a new NetCDF (.nc) file.

CREATE TEXT FILE with .nc file

The goal of this section is to create a text file with paths to each of the downloaded (.nc) files, sorted by time (starting from the earliest). Sorting can be done manually or automatically using the **Sort using regex pattern** option. A *regex pattern* (short for regular expression) is a sequence of characters that defines a match pattern in text, used here to detect dates. The files are then sorted by date. This option is experimental; if it fails, an empty file is generated.

The first step is to enter the directory containing the necessary (.nc) files, including all subdirectories, in the **Directory to look for field** **(1)**. Next, enter the name of the text file in the **Text file with .nc** field using the NEW button **(2)**, where the file paths, including their full paths, will be saved. Manual sorting may be required. If you select the **Sort using regex pattern** option **(3)**, an additional file with the same name and "_sort" suffix will be created, containing the time-sorted lines from the file (or empty if sorting is unsuccessful). Use the CREATE button to generate the text file.

CLIP IN SPACE AND JOIN IN TIME to one .nc

The goal of this section is to load a prepared text file with time-sorted .nc files, select the required variables, crop the data to the needed range of dimensions, and create a single output file.

1. Load the Text File : Use the **Text file with .nc** field and press the OPEN button **(2)** to load the prepared file containing paths to the .nc files.

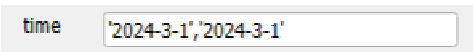
2. Select Variables: In the **Select variables** field **(4)**, choose the needed variables. Add selected variables to the right-hand list using the Add , Add all , and Clear buttons to create the desired variable list (comma-separated). You can also manually edit the list if needed.

3. Define Spatial and Depth Ranges : Specify the area and, if necessary, the depth range in the output data. !!!Define these ranges in the **Subset** tab by selecting one of the files intended for merging.!!! Confirm your selection by clicking OK in the **Subset tab** **(6)**.

4. Name the Output File and Join Files: Enter a name for the new (.nc) output file and click the JOIN button to create the merged file.

SAVE CURRENT SELECTION from Subset tab as .nc

This function saves the current selection as a new NetCDF (.nc) file. To do this, enter the desired file name and press the SAVE button. If you want to save a subset for a single time point, you need to create a time interval with the same value at both the start and end, e.g.,



[TOCTabs]