Xraster PLUGIN FOR QGIS

Multidimensional Raster GIS

USERS MANUAL

This plugin requires additional Python libraries.

version: 0.2
January 2025

Table of Contents

INTRODUCTION

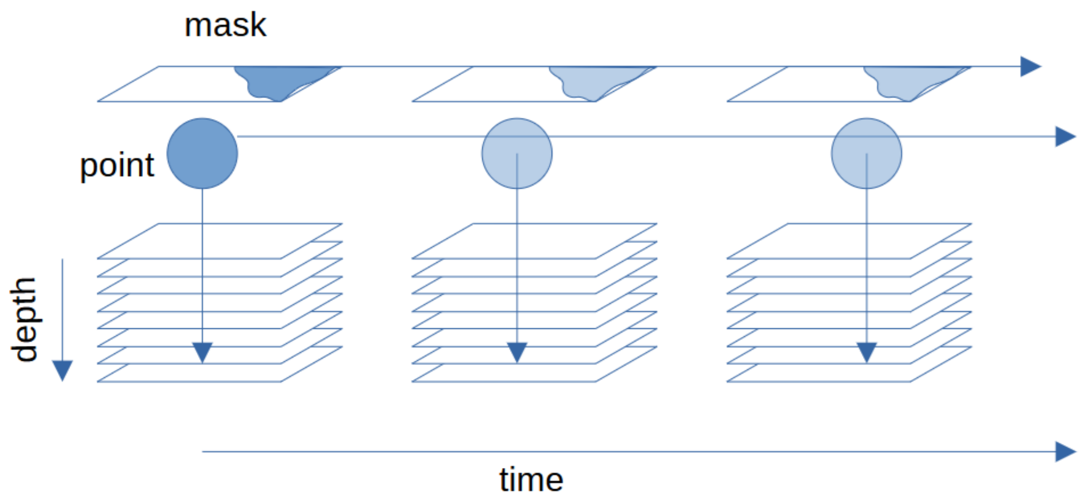
Xraster is designed for the analysis and processing of multidimensional raster data, where the primary dimension is time. This characteristic distinguishes such data from multispectral satellite images, where the raster dimension corresponds to the position along the wavelength axis (channels). Both types of data are based on a basic raster (with x and y axes) defined in the same way as a raster layer is understood in GIS. The concept of a basic raster is used in this manual to refer to this element of the data structure.

Multidimensional data can appear as 3-dimensional data, representing a time series of rasters, or as 4-dimensional data, where for each moment in time there is an additional depth dimension (a separate raster for each depth).

Such data constitute the most important source of information about the state of the atmosphere and oceans, both past and present. These data are provided by various organizations and scientific institutions (e.g., the Copernicus Programme of the European Union) in the NetCDF format, which is one of the primary formats for such data and is supported by **Xraster.** The primary tools for spatial data analysis are Geographic Information Systems (GIS). **Xraster** is embedded as a **plugin** in the **QGIS** software, enabling bidirectional integration in data processing. The plugin uses raster and point layers created in QGIS and generates data (raster layers and text data) that can be utilized within the program, creating a simple raster-based multidimensional GIS.

Xraster is built upon the functionality of the Python library (module) **xarray**. It serves as an interface to this module, allowing the entire process of data processing and analysis to occur without the need for programming.

The analysis process in this module, which serves as the engine for **Xraster**, is based on creating data subsets.



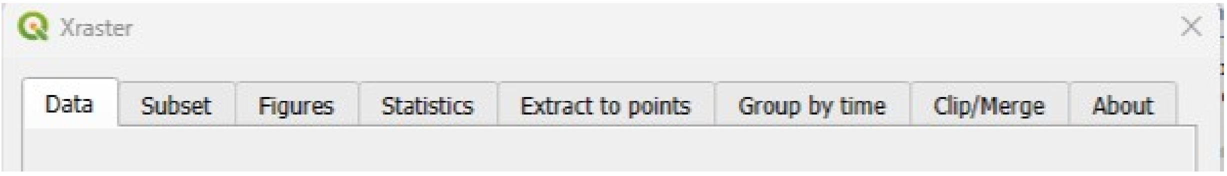
The diagram above illustrates the structure of multidimensional raster data with two accompanying elements: a mask raster and a point layer. If the raster is defined only on the surface, we will have 3-dimensional rasters (x, y of the primary raster and time). If the rasters are also defined for different depths (or heights), we will have 4-dimensional rasters (x, y of the basic raster, time, and depth or height).

Simplifying the entire process, geoprocessing involves two stages: creating a data subset and analyzing it. The type of subset determines the possibilities for data visualization and analysis.

Subset creation is based on 3 or 4 dimensions. In each dimension, a slice (a portion of the original space) or a time period can be defined. A subset corresponding to a specific point in space or a moment on the time axis can also be designated. Different types of data subsets allow for various visualization methods (e.g., histogram, map, time series, vertical profile, depth-time cross-section) and statistical analysis.

Working with the program involves creating subsets, visualizing them, and conducting statistical analysis. The results of these analyses can be visualized as described above or saved as raster or text layers.

The plugin consists of several tabs, each dedicated to a specific stage of working with the program. The program includes the following tabs,



which perform the following tasks.

Data

- Import of 3-dimensional (x, y, time) and 4-dimensional (x, y, z, time) data in NetCDF (.nc) format and obtaining their description.
- Changing the spatial resolution of the primary raster.
- Import of a raster mask (zone map) used for data analysis.
- Import of an additional variable representing the surface area of cells in the primary raster. This is used with geographic coordinates to calculate areas (e.g., ice cover) or as a weight for calculating weighted averages.

Subset

- Selection of the variable to be analyzed.
- Recalculation of the variable using an algebraic expression.
- Creation of a subset of the imported 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 (visualizations can be saved as high-resolution .jpg files).
- Maps can be exported as GeoTIFF rasters. Histogram, profile, time series, and cross-section data can be exported as CSV text files.
- For data with geographic coordinates, visualization is possible using the Polar Stereographic projection, including the addition of coastline lines on maps.

Statistics

- For a subset, spatial statistics (Spatial Summary) can be calculated for each day across the entire space (x, y, z) or for areas defined by mask zones (Zonal Statistics). Results are presented as a line chart over time and saved as text files.
- Temporal statistics for a given surface (x, y) produce a map, while at a point, they create an average vertical profile. Maps can be saved as GeoTIFF files.
- For data with a depth dimension, statistical maps of the seabed can be created.

Extract to points

- Extraction of data from a subset into a dataset of points, saved as a text file.

Group by time

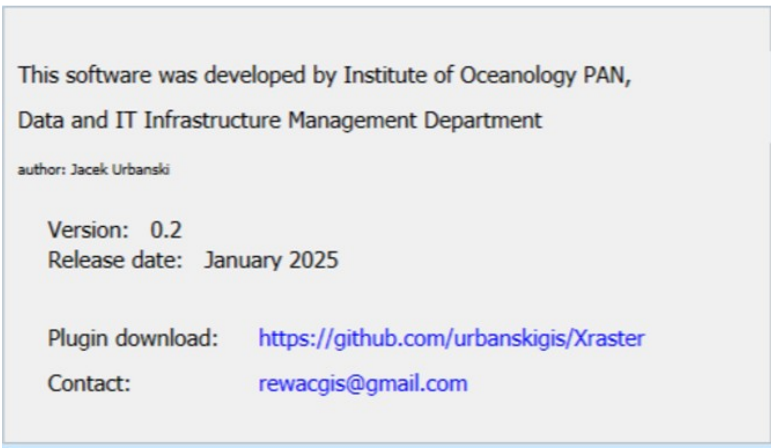
- Aggregation of data into groups (e.g., by months, years, climate seasons, or days) and calculation of statistics for each group. For each group, a map (at a specified depth) or a histogram of any statistic can be created.

- 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 as text files.

Clip/Merge

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

About



[TOC]

INSTALLATION

---

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:

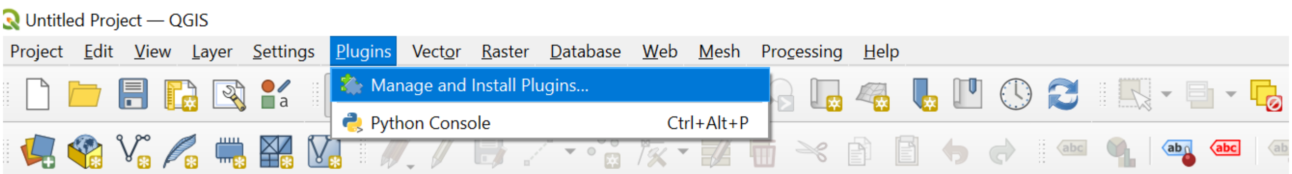
- **Run** osgeo4w-setup
- **Choose** Express install
- **Select Packages** : QGIS and GDAL (at least)
- Continue installation

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

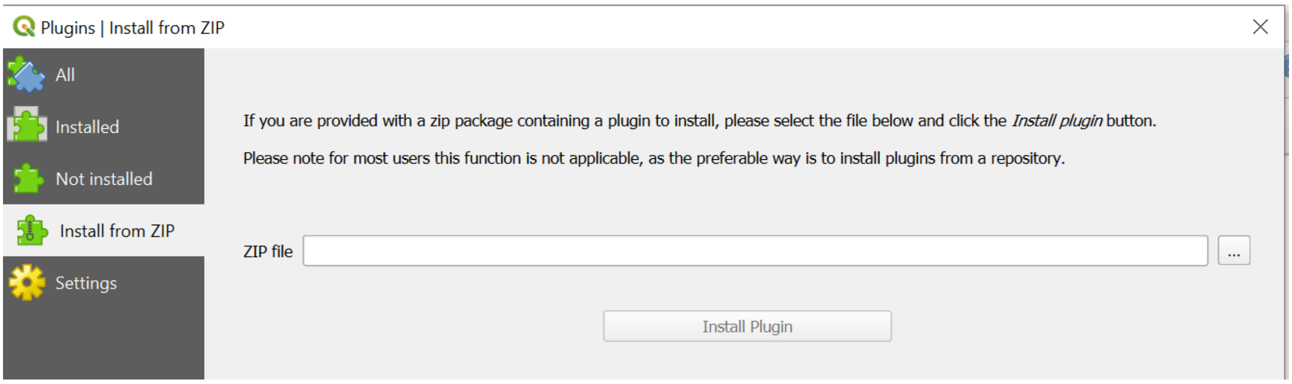
2. Xraster INSTALLATION PROCEDURE

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

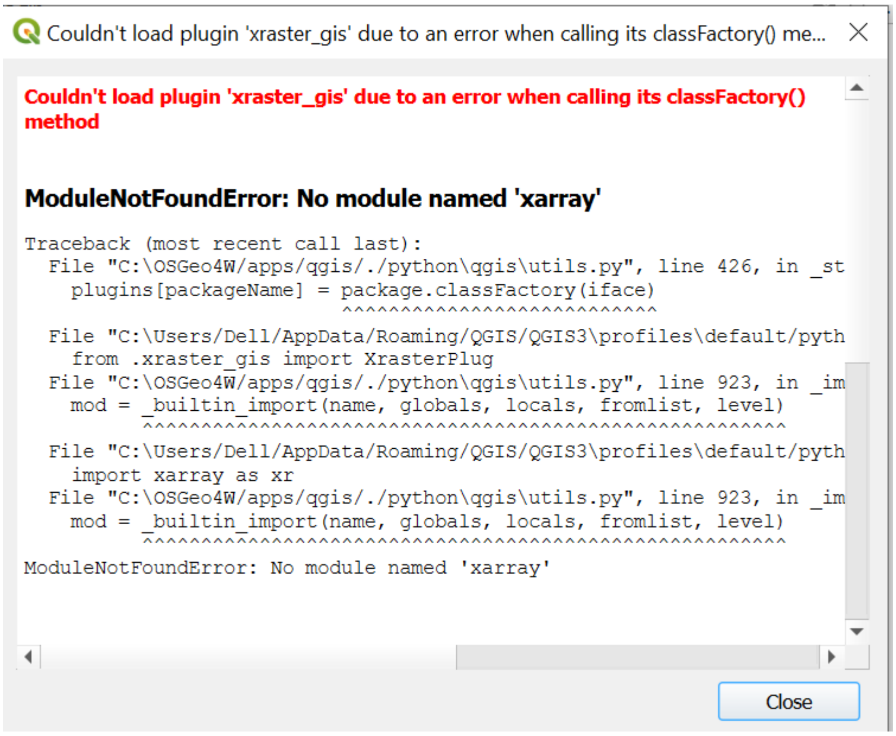
- You need to download the packaged plugin ***xraster_gis_?.zip*** from the website: https://github.com/urbanskigis/Xraster to any directory.
- Open QGIS 3.xx (minimum version 3.20)
- Open "Plugin" menu from the top bar



- 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**, **h5netcdf** and **Cartopy** using the following three commands executed in sequence:

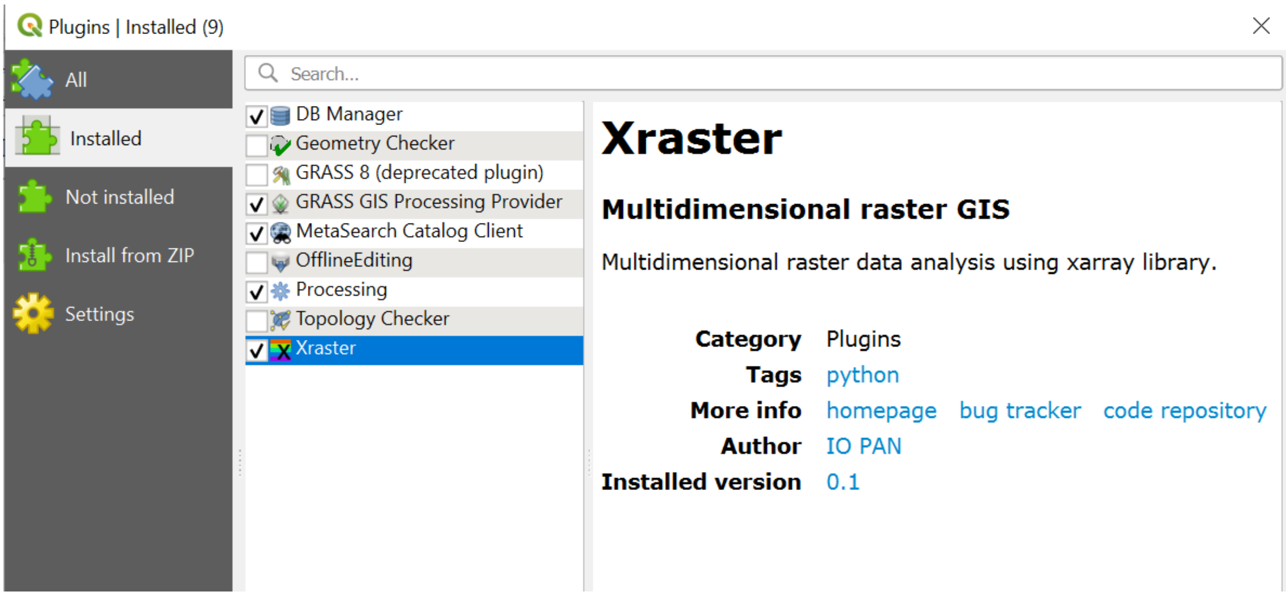
```
C:\OSGeo4W>pip install xarray
```

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

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

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

Lanch again QGIS and the plugin installation should run fine. When the installation is complete, the ***Xraster*** tool appears in the installed plugin list (The current version of the program is best checked on the About tab, page 5) ,



[TOC]

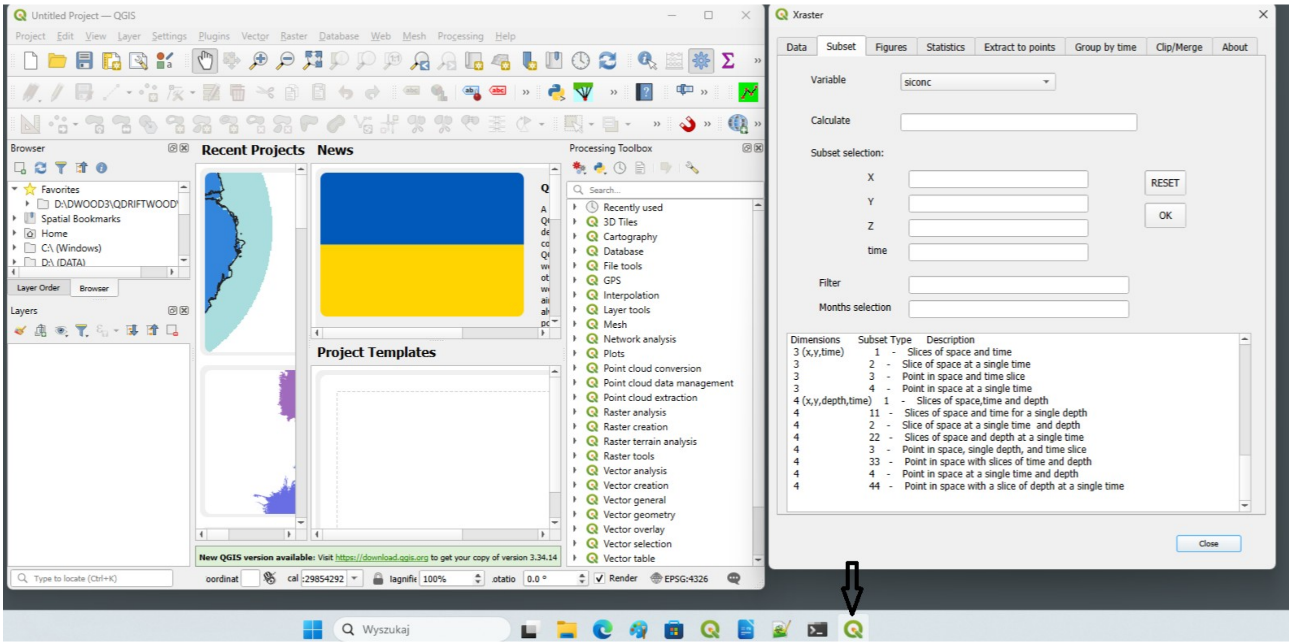
Xraster GENERAL PRESENTATION

---

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.**

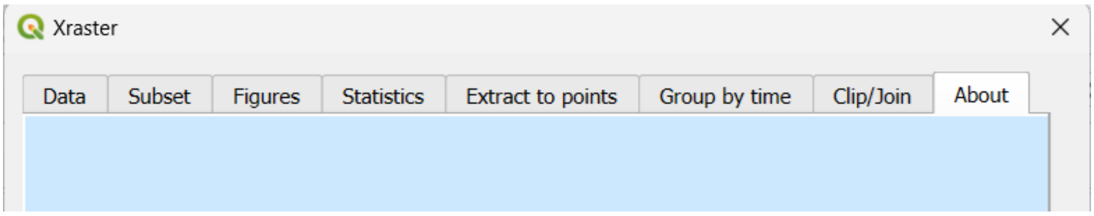


,

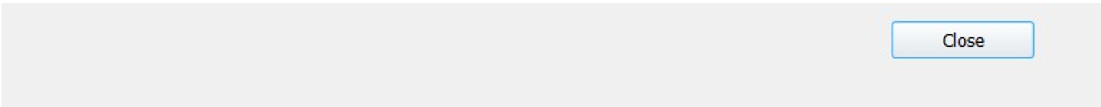
[TOC]

2. TOOL STRUCTURE AND FUNCTIONS

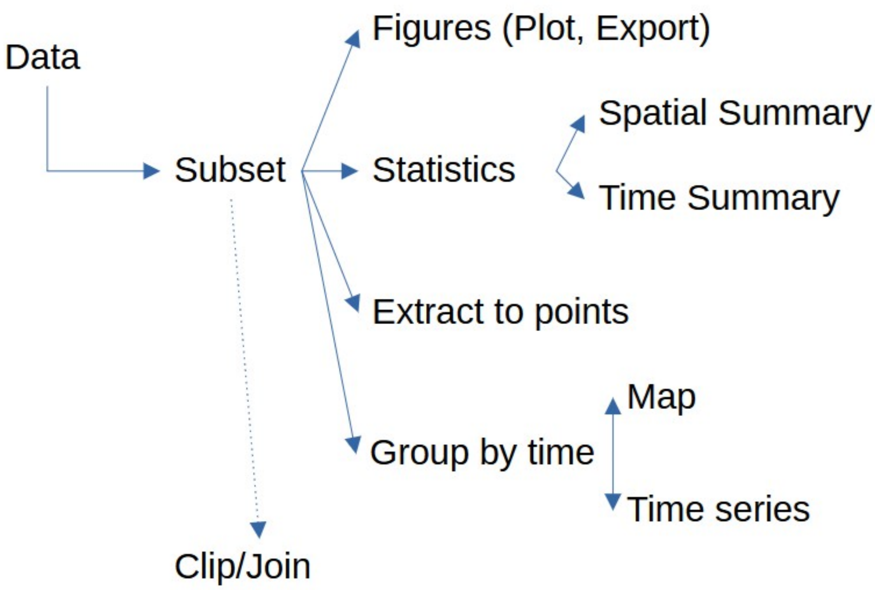
The **Xraster** plugin consists of seven tabs that implement different groups of functions of the software. (Currently, Clip/Join has been replaced by Clip/Merge.)



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/Merge** tab is a separate subprogram designed to assist in creating .nc files for analysis in Xraster. Downloaded NetCDF (.nc) files may cover large areas, contain numerous variables, and be downloaded for separate time periods or individual days. As a result, they are often impractical for use in the program. Effective analysis requires combining the files into one while simultaneously minimizing its size. This can be achieved by limiting the area and the number of variables. The tab performs four functions:

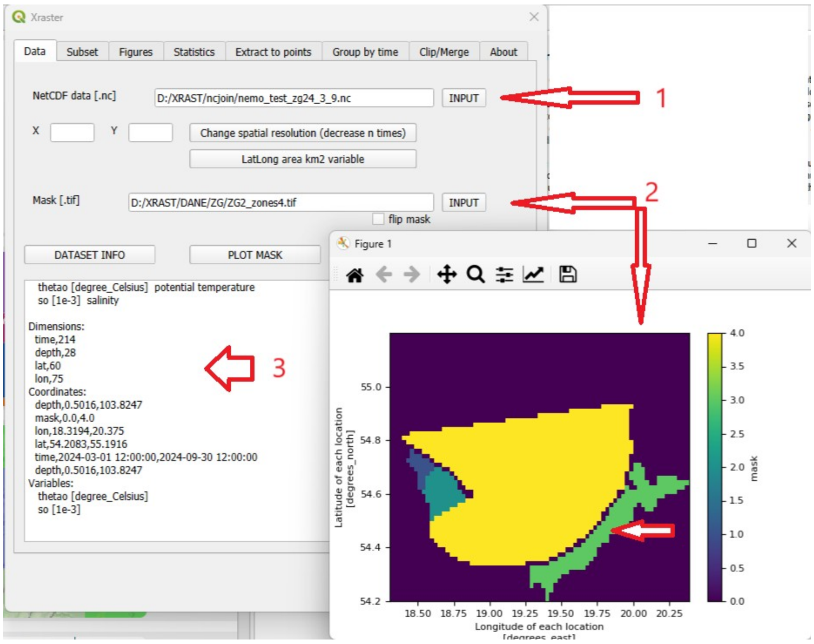
1. **Organizing File Paths**: It allows you to organize paths to downloaded files into a single text file, assuming the files may be stored across multiple directories.
2. **Area Clipping and Variable Selection**: It enables you to clip the area of interest, select the required variables, and merge them into a single file. The spatial extent (clipping) of the data is taken from the current extent defined in the Subset tab.
3. **Merging by Variables**: It facilitates merging files with different variables but a common time dimension into a single .nc file.
4. **Saving Subsets**: It allows saving the current subset created in the Subset or Group by Time 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**.

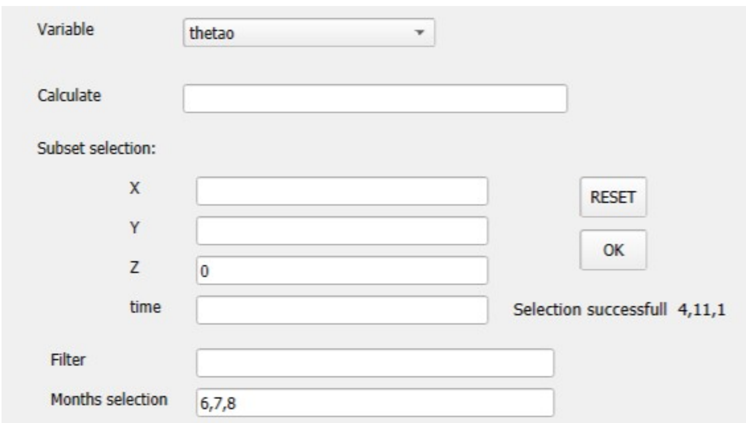
Data analysis begins with data input, examining the data structure, and optionally obtaining a mask.



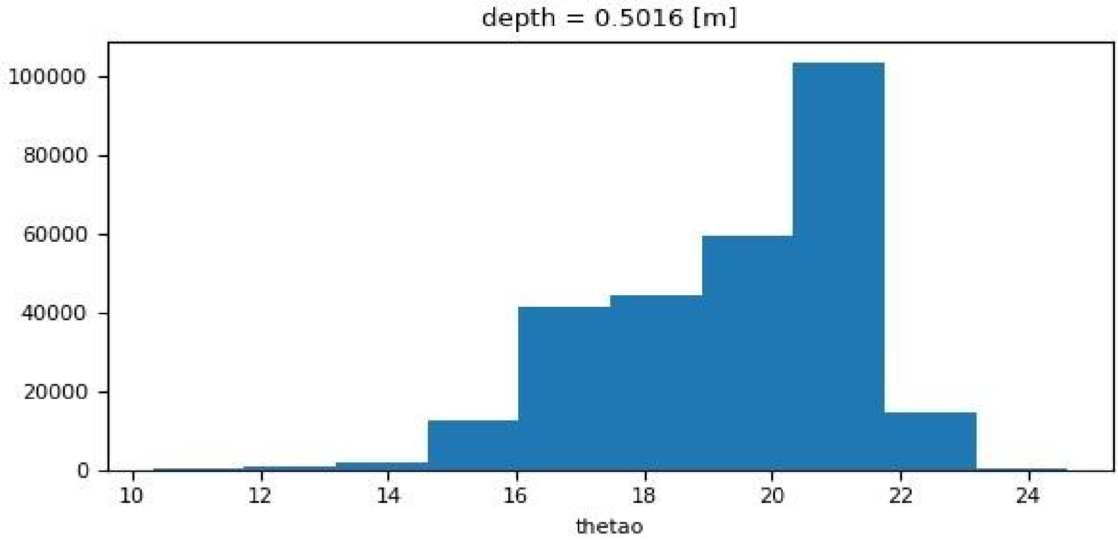
The first step is **data input** (**1**) and, if needed, **inputting a mask** to designate analysis zones (**2**). A description of the data will then be displayed (**3**). Our data dimensions are area (60 x 75 cells), depth (28 levels), and time (214 days). Each data cell contains two variables: temperature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

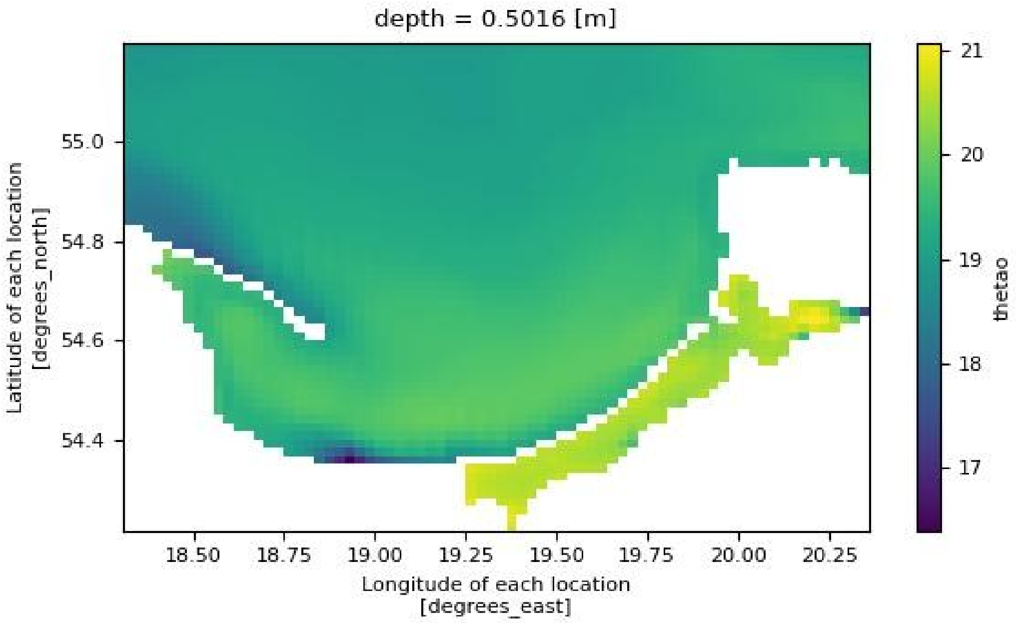
a. We select surface temperature for the months of June, July, and August.



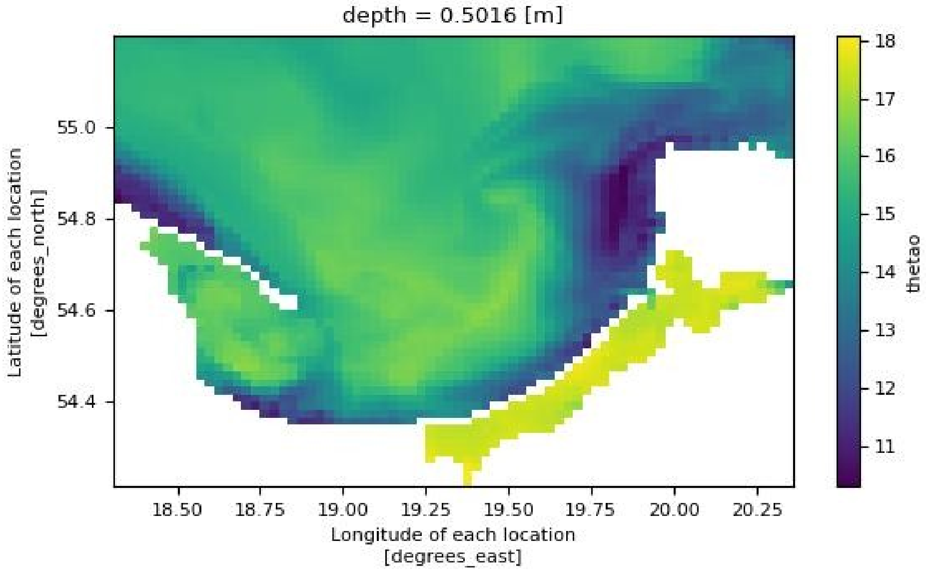
b. Display a histogram (Plot from the **Figures** tab) of surface temperature values.



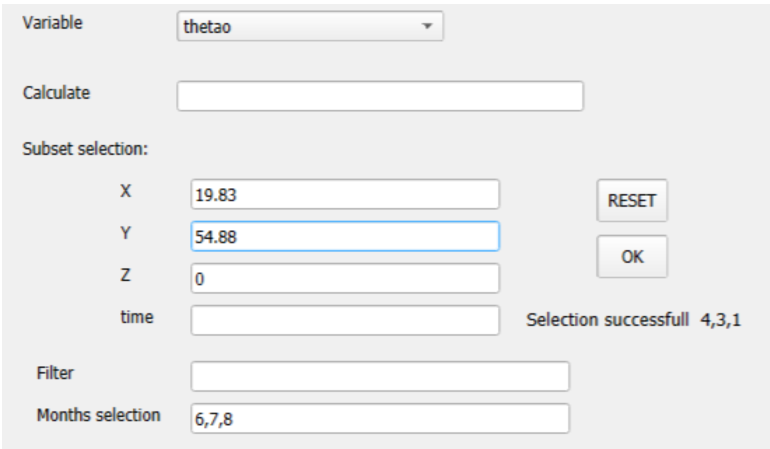
c. Display the average temperature distribution map for the entire time period (Plot - **mean** from the **Statistics** tab).



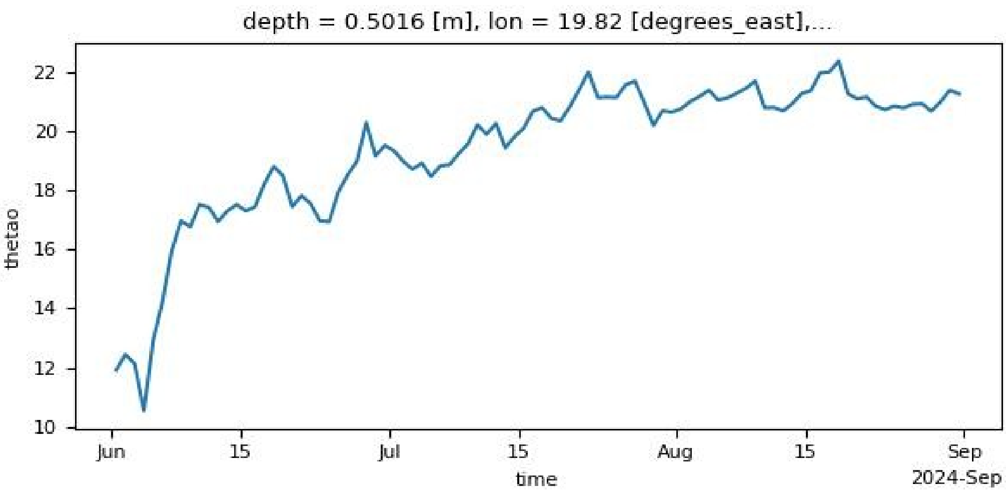
d. Display the map of minimum surface temperature (Plot - **min** from the **Statistics** tab).



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

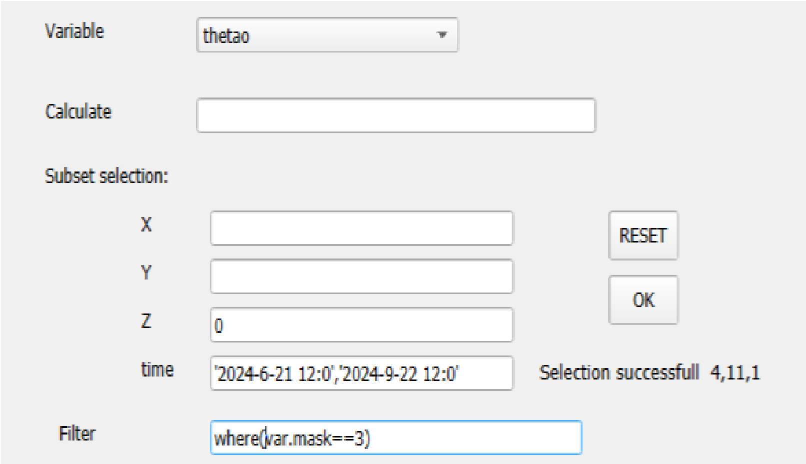


and plotting a time series (Plot from the **Figures** tab).



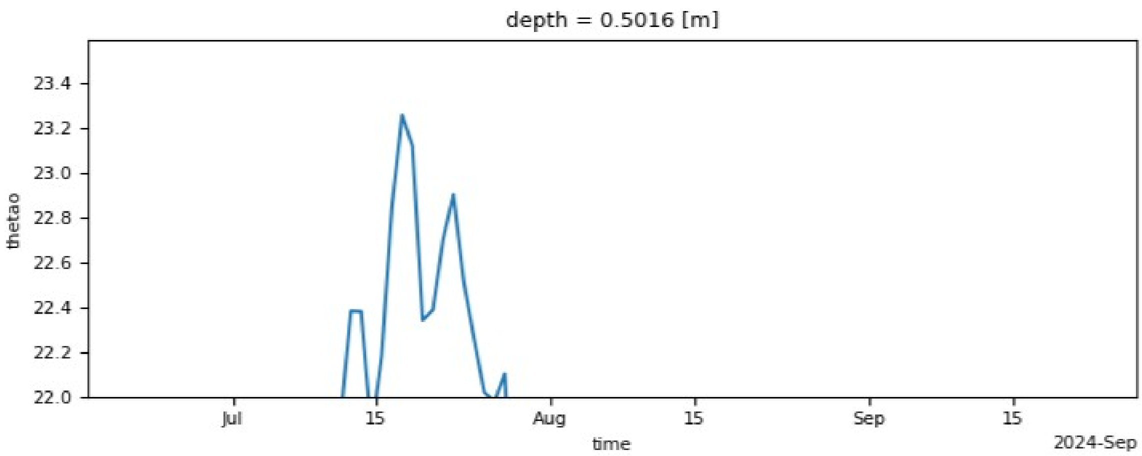
In a similar way, you can continue exploring the data.
Data analysis is based on defined queries. For example, I would like to know how many days the average surface water temperature of the Vistula Lagoon (mask ID = 3) was above 22 degrees during the period from June 21 to September 22 (summer). I am 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).

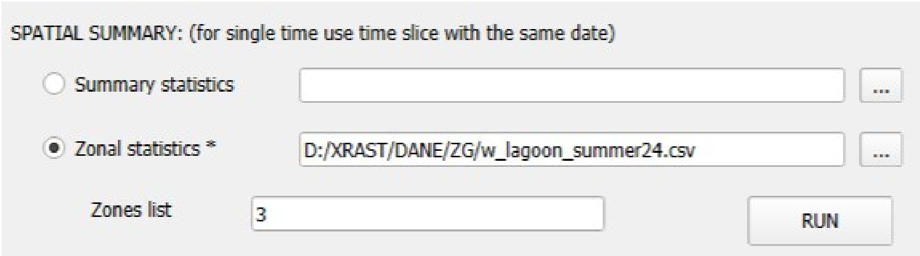


A required time range is specified. The filter retains only the data that meet the condition: Zone 3 (Vistula Lagoon).In the **Statistics** tab, there are two options for calculating and presenting spatial statistics, separately for each moment (day) within the specified time range. The first method involves calculating a specific statistic and generating a chart of its variation over time.

On the resulting chart, the temperature axis can be set to start at 22 degrees. As a result, we obtain:



The figure shows that such a situation occurred in the second half of July. More detailed results can be obtained using the **Zonal Statistics** function (for Zone 3 only), which generates a text file with calculated statistics for each day of the analyzed period.



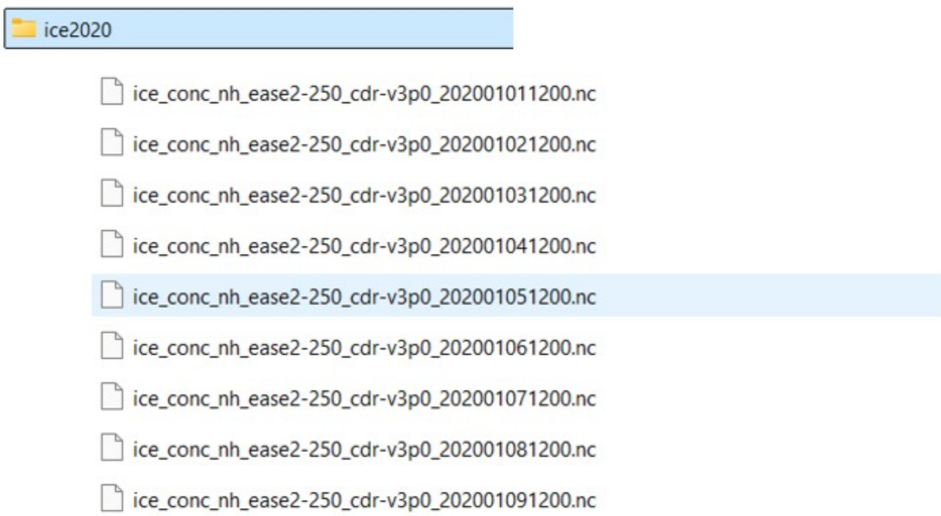
The result is a text file that can be analyzed in other programs.

	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

Data engineering in the **Xraster** program is used to prepare data for analysis and is performed in the **Clip/Merge** 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.

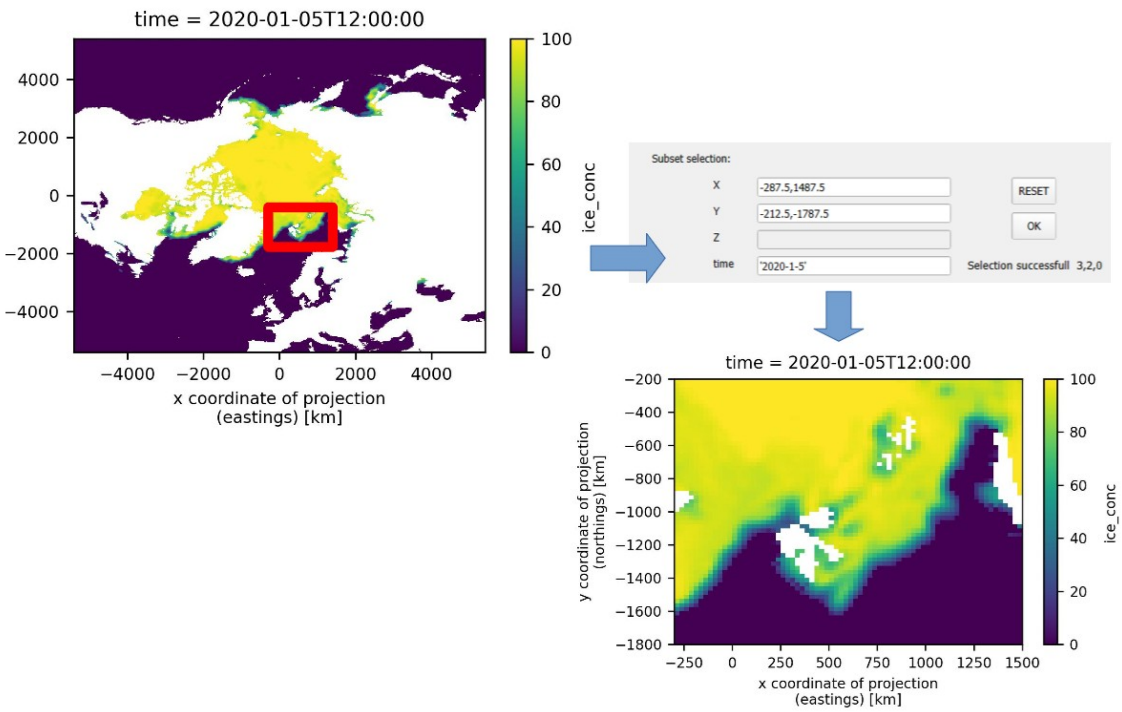
In other cases, it's possible to select a spatial and temporal window and choose specific variables as needed. Some data, when using the simplest methods, may only be downloadable for the entire area with all variables, as separate files for each day. This requires subsequent extraction of the desired 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.

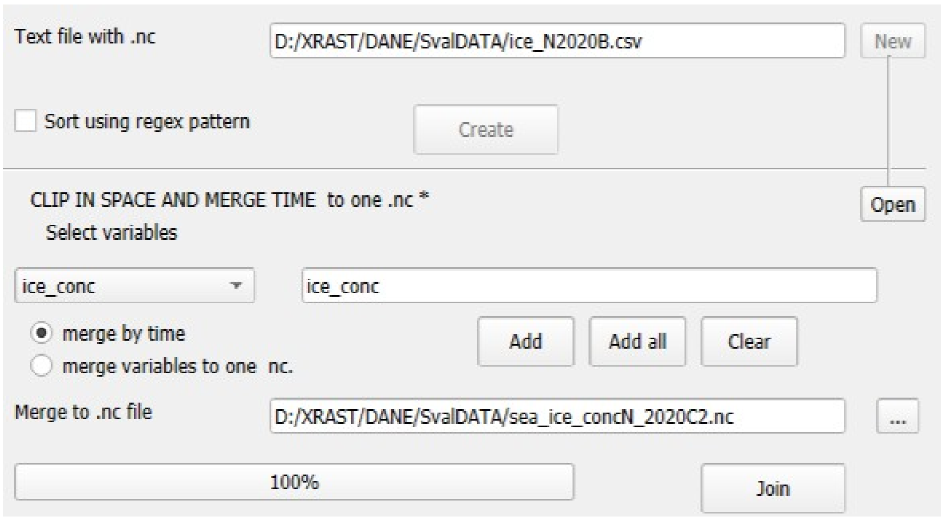


The first step is to create a text file listing all .nc files (with full paths) located in the specified directory. All subdirectories within this directory structure are also searched.

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.

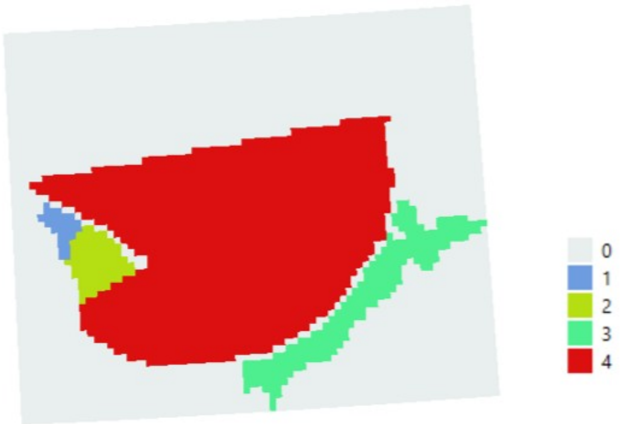


[TOC]

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. <u>It is best not to use NoData values on the mask and replace them with zeros.</u> 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.

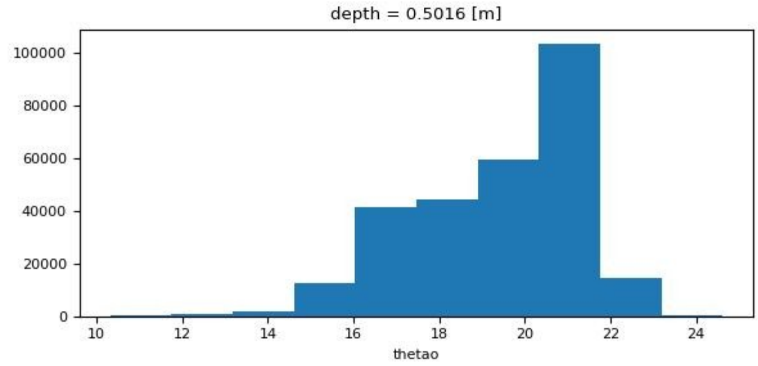
[TOC]

5. PLOTS, SAVE AND EXPORT PLOT

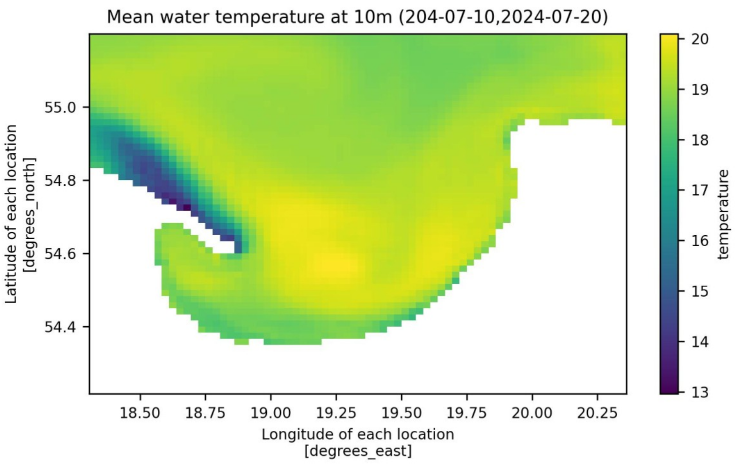
Images generated by Xraster

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;

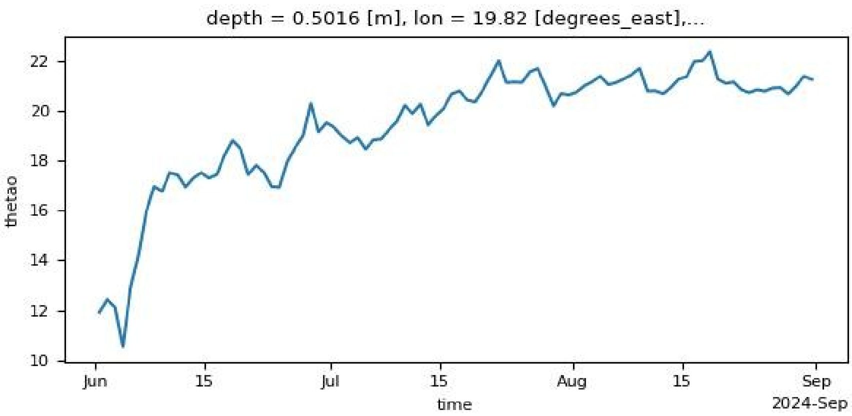
Histogram,



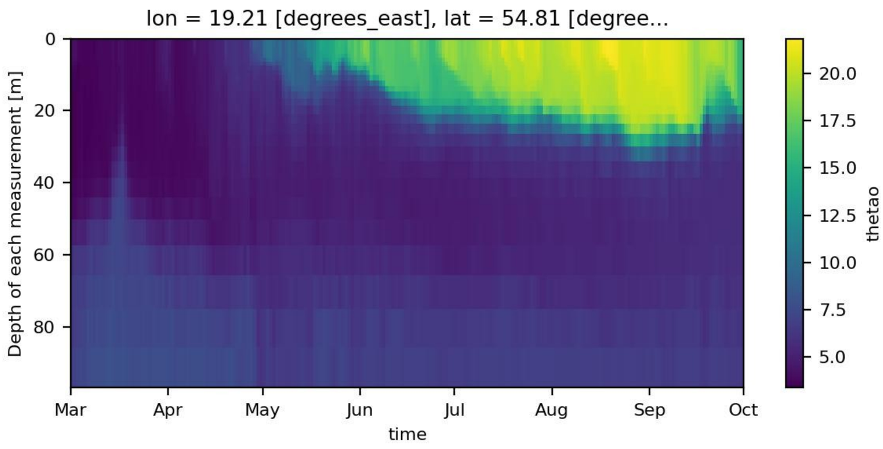
Map,



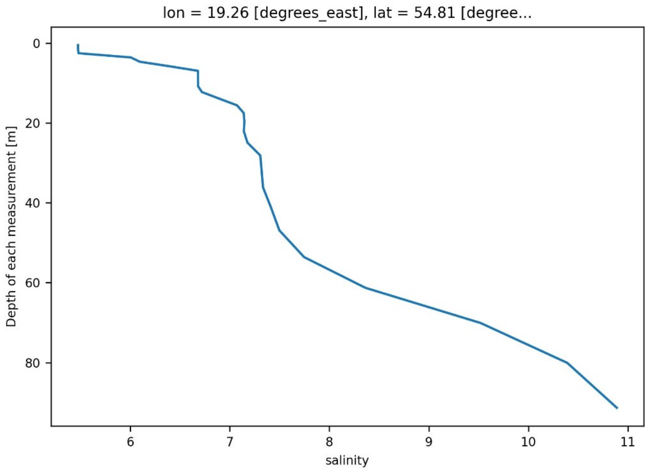
Time series,



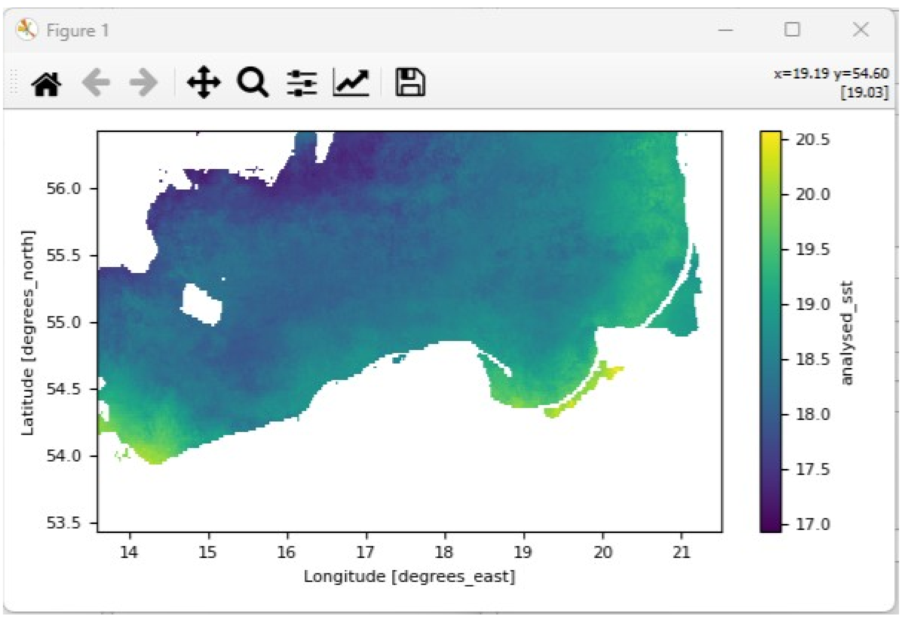
Depth-time-cross-section,



Vertical profile,



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

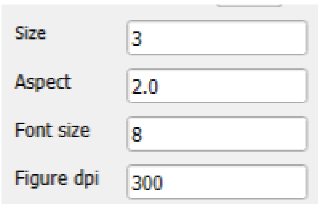


This window displays the image along with axis labels and a color scale with descriptions. The top toolbar includes icons for interacting with the image, as well as coordinates for the cursor position on the image and the corresponding variable value.

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.

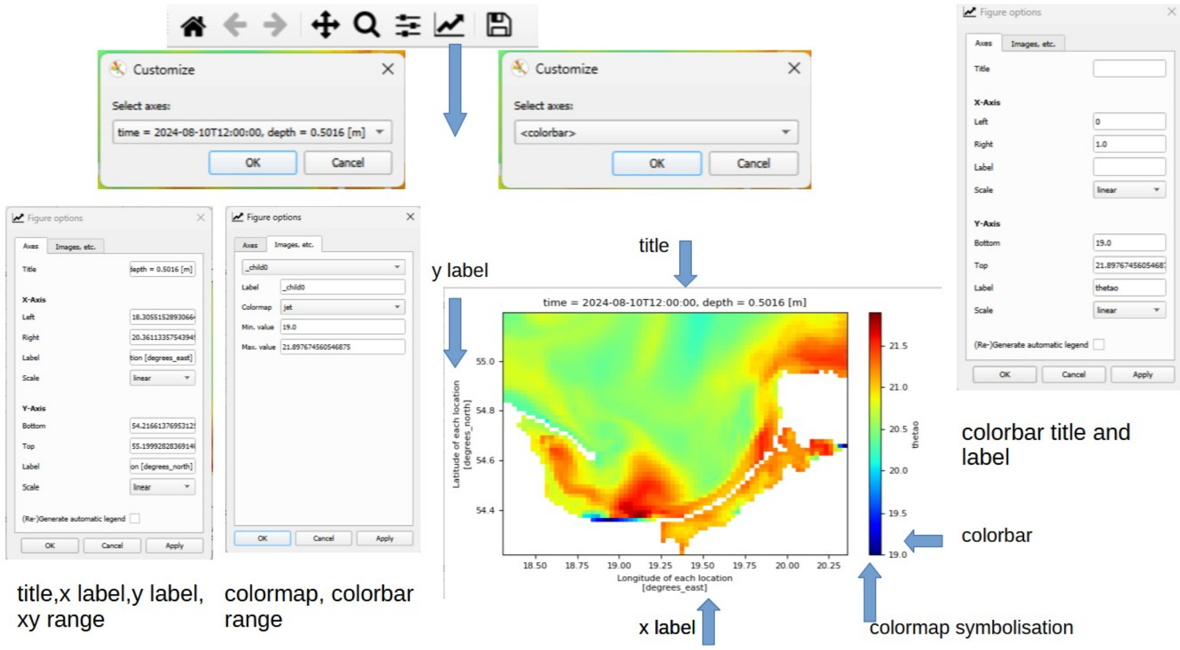


This is done using a separate 'Save plot (file)' option in the program.

[[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



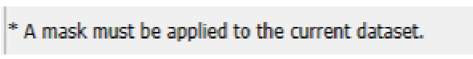
We begin the process by selecting the first option (the window on the left). Here, it's possible to change the map title and labels for both axes. For maps, you can also define the color scale (colormap) and its range. The second option adjusts based on the first selection and allows you to add a description 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/Merge** tab creates a text file containing all NetCDF (.nc) files in a given directory, including subdirectories. Additionally, the data can be sorted by time.

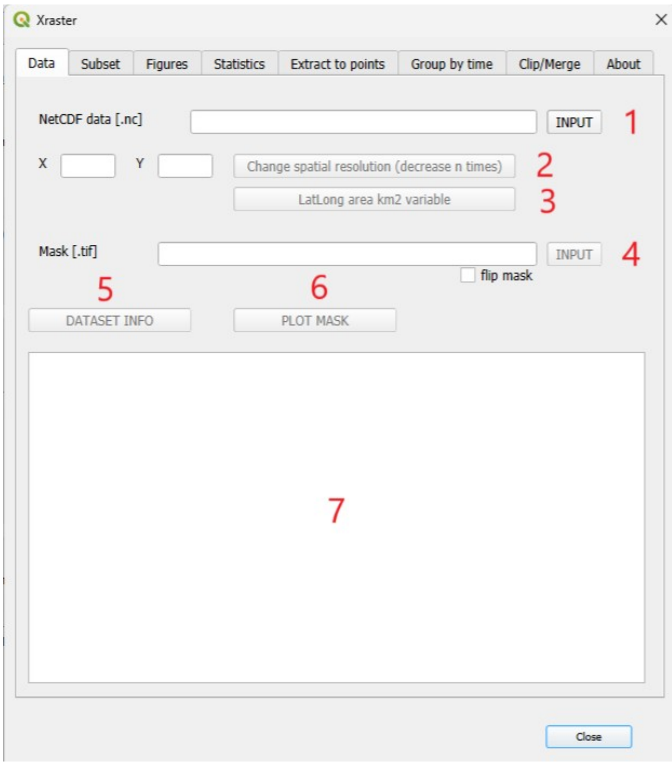
[TOC]

USING Xraster (Tabs)

---

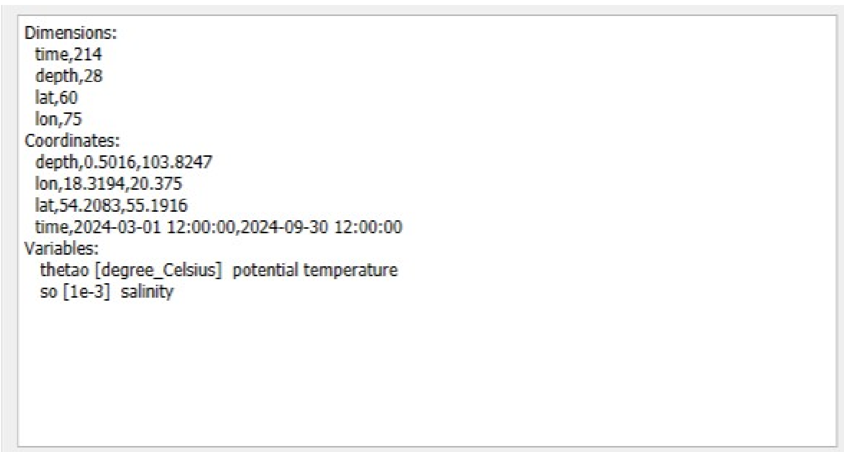
1. Data

This tab has been modified compared to version 0.1. There is no longer a need to confirm the order of dimensions, and two new features have been added. Examples from older videos are still valid but without the need to confirm the order of dimensions.



Basic operation.

(1) The process starts by importing a NetCDF (.nc) file into the program. If an invalid file is imported, the text field **(7)** will display the message: **!!!! Not valid NetCDF [.nc] file or no time dimension !!!!**. After importing the file, the text field will show information about the file. Once the data is loaded, you can proceed to the next step by navigating to the **Subset** tab.



Dimensions refer to the names of the coordinate axes of the input data. In the example, we have a basic raster with dimensions of 60x70 cells (axes: lat, lon). There is a time axis with 214 time points (time axis). At each time point, there are 28 depth levels (depth axis). In total, our data structure consists of 214 x 28 = 5,992 basic rasters (similar to those in GIS), oriented by time and depth.

Coordinates refer to the assignment of labels. The displayed information provides the minimum and maximum values for each coordinate. In the example, the coordinate names correspond to the axis names. However, this is not always the case. Axes without assigned labels can exist, and coordinates may not correspond to axes. Nonetheless, coordinate ranges often provide information about the value ranges along the axes.

Variables represent the variables within the spatial structure defined by the axes. These are the values found in each cell of the basic rasters. In the example, there are two variables: temperature and salinity. Variable names include metadata, usually describing the variable and its unit of measurement.

All additional options in the tab, except the PLOT MASK button, will become active. Once the data is loaded, you can proceed to the next step by navigating to the **Subset** tab.

Additional options.

The additional options include:

- **Changing the resolution of the basic raster (2):** Allows for adjusting the size or scale of the raster grid.
- **Creating a km2 variable (3):** For geographic coordinates, this generates a variable containing the surface area (in square kilometers) of each cell.
- **Introducing a mask (4):** Enables the application of a mask to the data for filtering or limiting specific regions.

The resolution of the basic raster (defined by the x and y axes) **is changed** by entering values into the X and Y fields. These values specify the divisor by which the current number of cells along each axis will be divided.

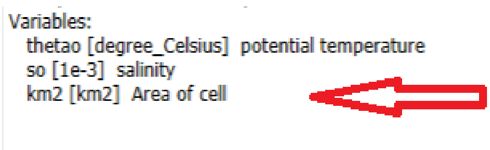
- Entering 1 in both fields will result in no change to the resolution.
- After specifying the desired values, press the **Change spatial resolution (2)** button to apply the changes.

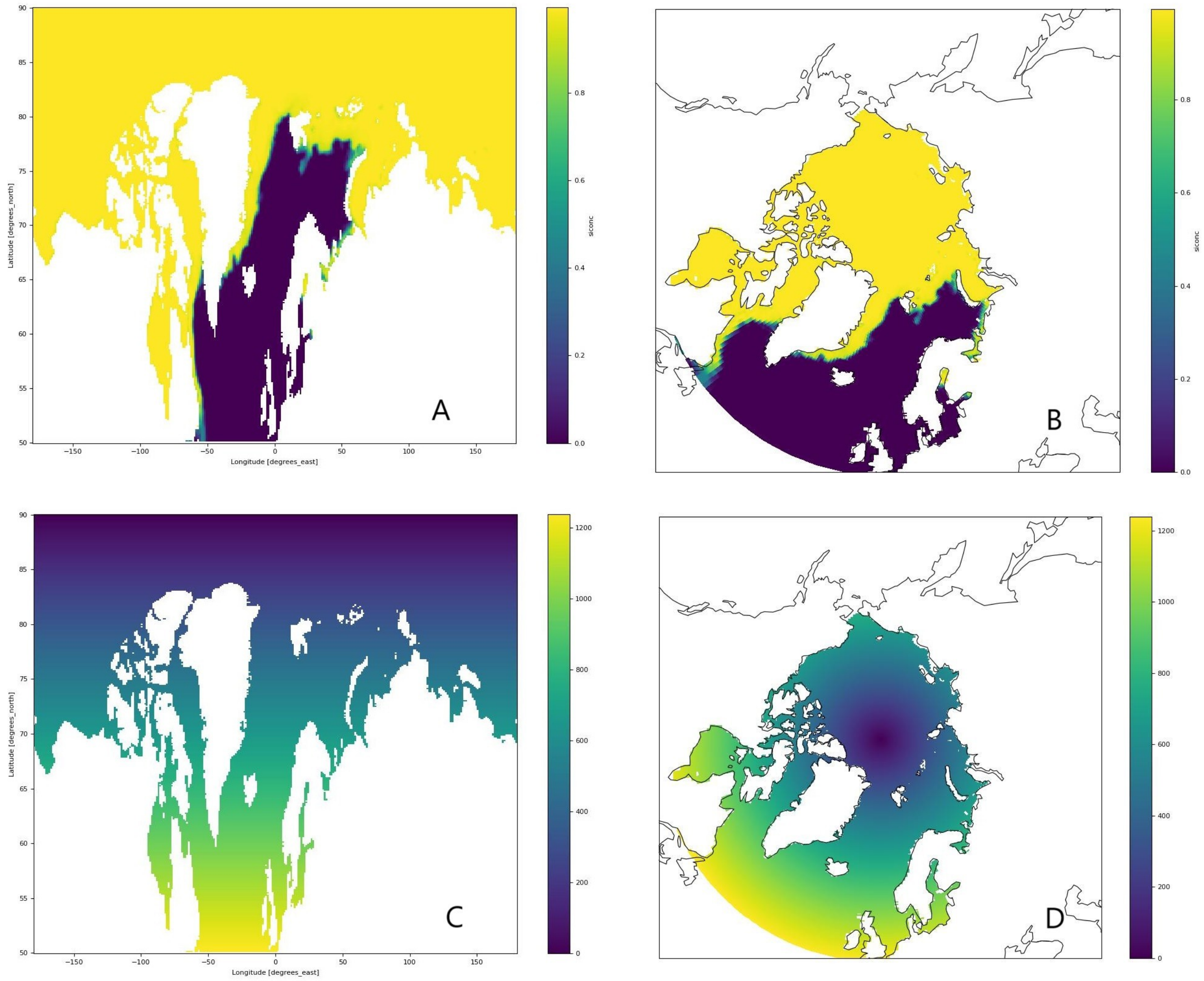
Pressing the DATASET INFO button again will redisplay the information about the dataset. The size of the basic raster will reflect the updated resolution.

To revert to the previous resolution, the data must be reloaded **(1)** .

Reducing the resolution is necessary when using the Polar Stereographic projection. Due to the interactivity of the Matplotlib GUI window, the program's performance significantly slows down after creating a plot with this projection. In practice, the data should be reduced to resolutions smaller than 300 x 300 cells in the basic raster. The reduction values for the X and Y axes do not need to be equal.

The km2 variable is a new variable that can be useful for geographic coordinates in the basic raster. This variable represents the area of a cell in square kilometers. Since the cell area varies depending on the location in rasters with geographic coordinates, this variable can be used for surface area calculations and for determining weighted averages (e.g., the weighted average temperature in geographic coordinates). To create this variable press the LatLong km2 VARIABLE **(3)** button. After pressing the DATASET INFO button, the new **km2** variable will appear in the list of variables.



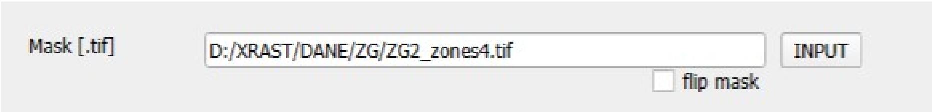


The figure above illustrates how the **km2** variable can be utilized:

- **Figure A:** Displays ice concentration data (values ranging from 0 to 1) in geographic coordinates.
- **Figure B:** Shows a map of the same ice concentration data projected using the Polar Stereographic projection.
- **Figure C:** Depicts the **km2** variable in geographic coordinates, representing the surface area of raster cells (ranging from 0 to 1200 km^2).
- **Figure D:** Shows a map of the **km2** variable using the same Polar Stereographic projection.

To calculate the **total ice-covered area (in km^2)** from the data in **raster A**, multiply **raster A** by **raster C** and sum the values across the entire resulting raster.

Introducing the mask is optional (4).



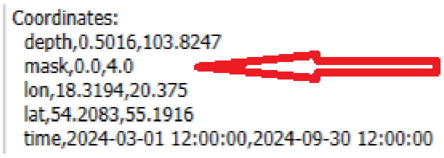
Its introduction enables:

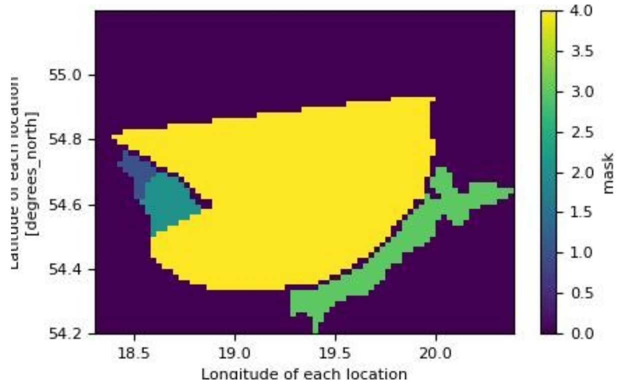
- Using it when creating subsets (**Subset**).
- Exporting surface maps to raster layers in GIS as GeoTIFF (**Figures, Statistics, Group by time**).
- Applying zonal statistics operations (**Statistics**).

The mask must have the same number of rows and columns, as well as the same georeferencing, as the surface data (x, y) in the input data (.nc). If the number of rows and columns do not match, an error message will appear:

- !!!! Not valid mask [.tif] file. !!!!
- !!!! Mask has different dimension x=75, y=60 !!!!

After entering a valid mask, it can be displayed using the PLOT MASK (6) button. Pressing DATASET INFO (5) again will show that the mask has been added to the coordinates. The mask contains values ranging from 0 to 4. Zero represents the background, while the other values define the zones (areas of interest).



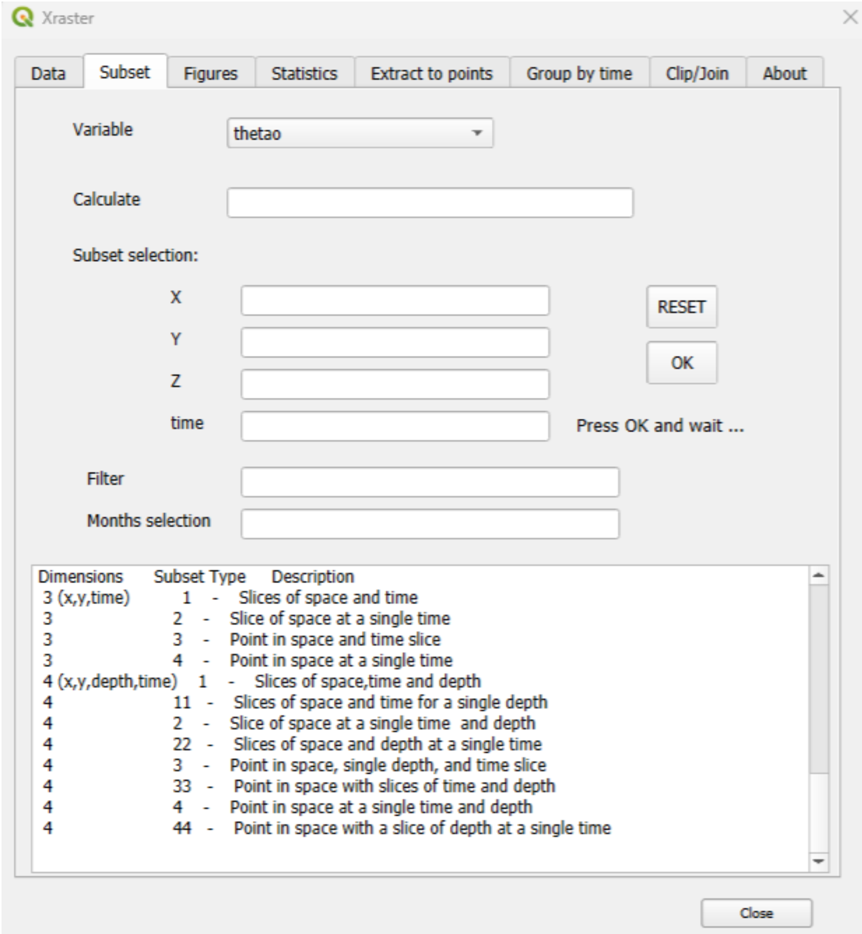


The introduction of the mask includes an option to **flip** it. To do so, you need to check the **flip mask** box located below the mask name field. This modification is sometimes necessary due to the different orientations of the Y axis.

[TOCTabs]

2. Subset

Creating a data subset is a key stage of analysis. The data subset is created for a selected variable. This variable, even in a four-dimensional dataset, can have a three-dimensional nature (e.g., referring exclusively to the sea surface).



The basic principles of creating subsets are presented in the text window. The program works with three-dimensional data (surface defined by x and y dimensions and time) and four-dimensional data (surface defined by x and y dimensions, depth, and time).

The primary method of defining a subset is by specifying intervals (slices) in space. This applies to the surface (restriction defined by coordinates), depth range, and time interval. It is also possible to define a subset at a surface point, at a given depth, and at a specific moment in time. If data is unavailable at a given surface point, depth, or time, the closest available data will be used.

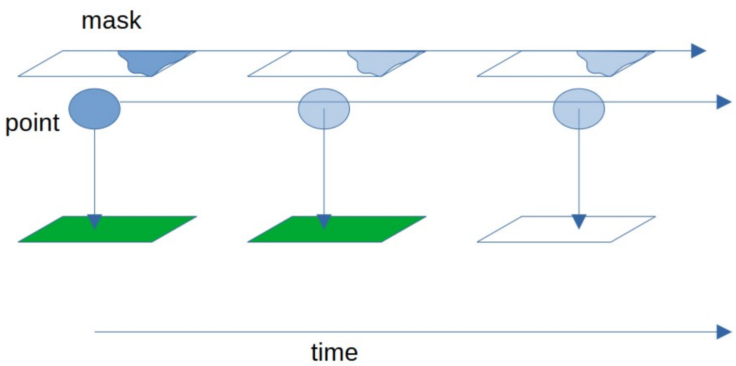
In addition to spatial operations, data filtering is possible. Filtering can apply to a variable or use a mask. It can also utilize all variables. This allows the creation of a subset of data located only within a specific geographic area. Additionally, filtering by months is available, enabling the selection of data from specific months only.

<ins>In the program, the type of subset is defined by a code in which the first digit represents the data dimension, and the second digit specifies the method of defining the dimension.</ins> For three-dimensional data, four subset types are possible, while for four-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

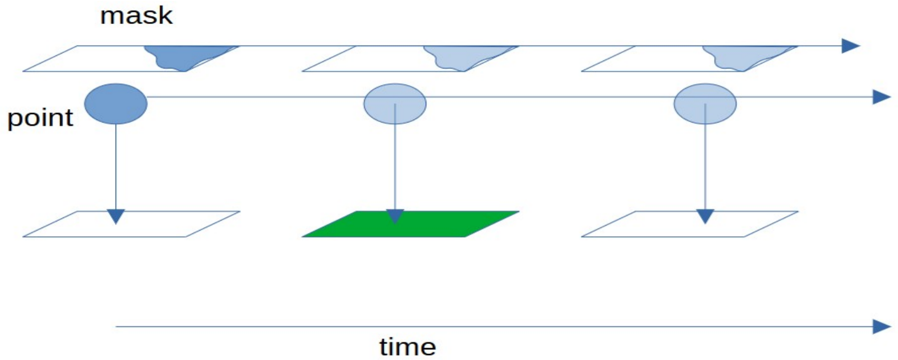
Three dimensions (x, y, time).

For three-dimensional data, the most general subset is a time series of basic rasters (**3.1 Slices of space and time**). It is possible to select a portion of the basic raster (x and y axes) and choose a specific time range.

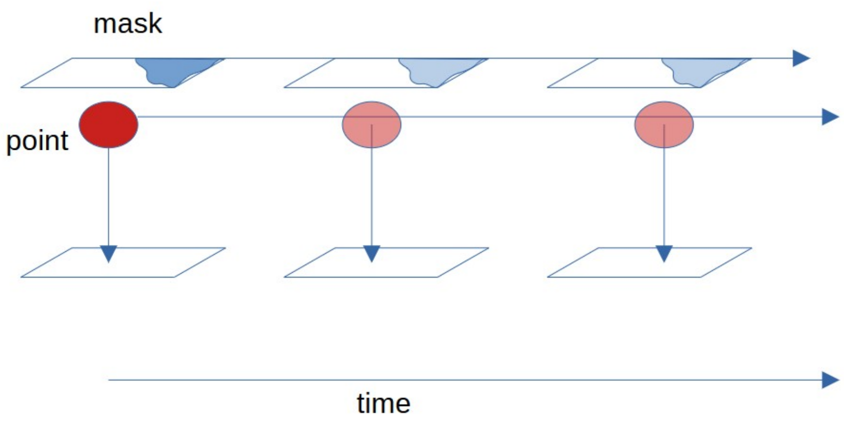


An example of such data is a time series of rainfall over a specific area during a defined period. Using filtering, the data can be further restricted to specific months, such as the summer months.

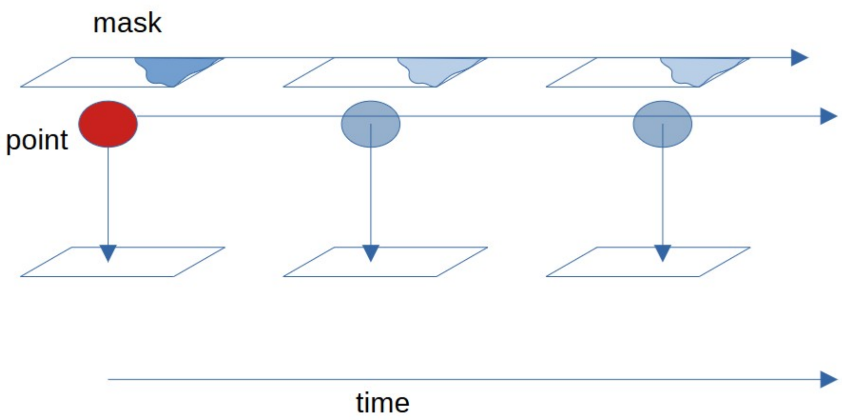
A three-dimensional subset can be defined for a specific moment in time (**3.2 Slice of space at a single time**). In this case, there is only a single basic raster, corresponding to a raster in GIS. Of course, the analysis can focus on a portion of the raster, defined by the x and y range or a mask.



The next two subsets are defined at a specific point (x, y). This means that from each basic raster, the variable at this point is extracted. The extracted values can form a time series of the variable, such as daily air temperature in a given city (defined as a point). This corresponds to **3.3 Point in space and time slice**.

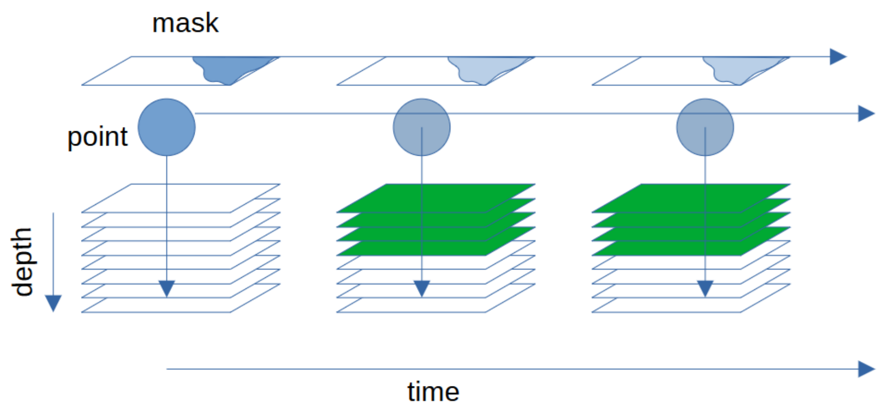


The extreme case is when a variable is extracted at a specific point for a given moment in time (**3.4 Point in space at a single time**). In this case, we obtain a single value of the variable.



Four dimensions (x, y, time, depth).

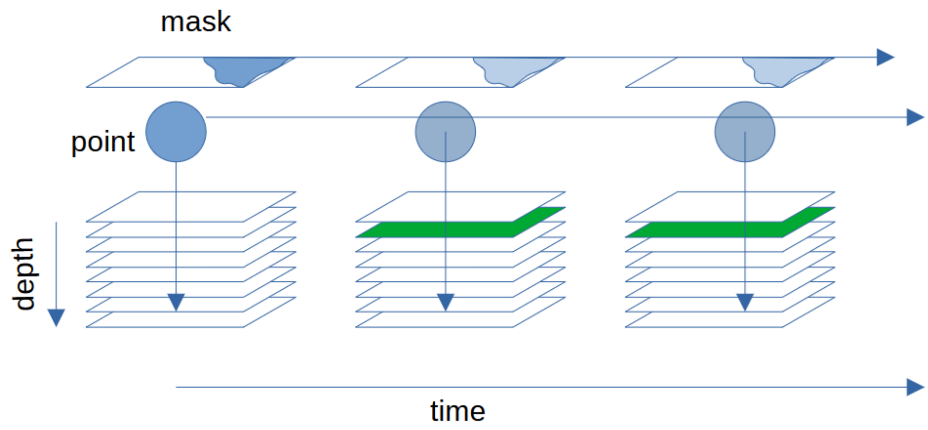
Four-dimensional data provide more possibilities for defining subset sizes.



The most general subset is defined across all dimensions (**4.1 Slices of space, time, and depth**). For different moments in time, there are corresponding basic rasters (x, y) at each depth within the depth range, or portions of these rasters.

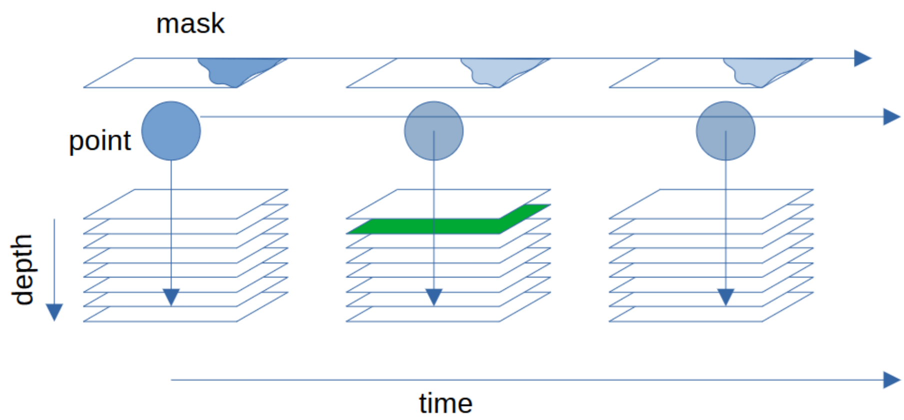
An example is an oceanographic variable changing with depth, such as a monthly (t) temperature field (x, y) over a certain area for a range of depths.

The primary way to restrict a four-dimensional subset is to limit it to a single depth (**4.11 Slices of space and time for a single depth**). This creates a subset analogous to **3.1**.

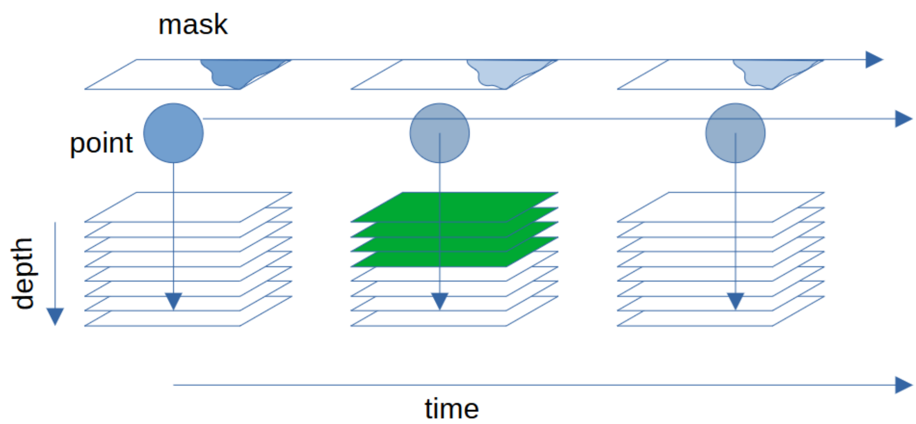


An example could be a subset from the previous example limited to a depth of 10 meters.

Restricting a four-dimensional subset to a specific moment in time and depth results in a single basic raster (x, y) (**4.2 Slices of space at a single time and depth**). As a result, we obtain a single basic raster, similar to **3.2**.

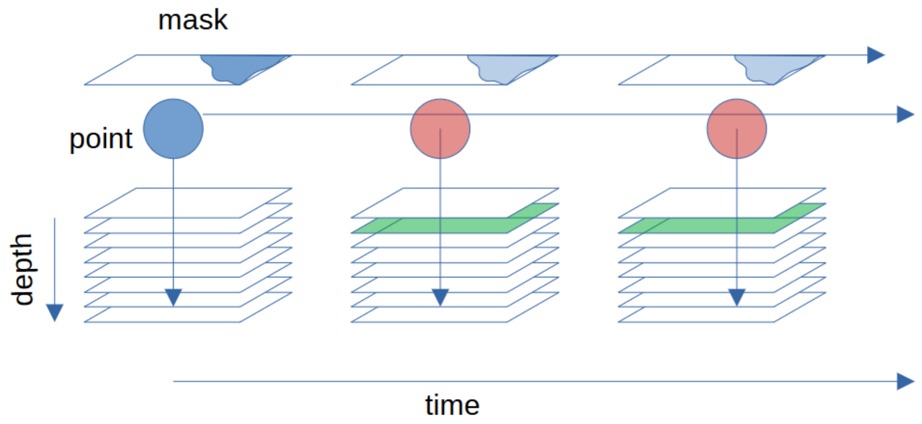


The last variant of the four-dimensional subset, which includes basic rasters (x, y), is to restrict it to a single moment in time (**4.22 Slices of space and depth at a single time**).



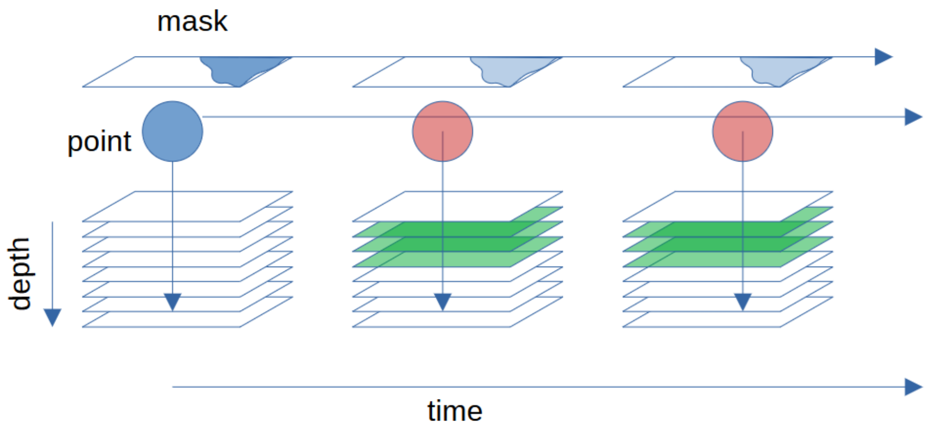
As a result, we obtain a series of basic rasters, each corresponding to a different depth within the depth range. An example of such data could be the average salinity field in a given area for a range of depths on a specific day.

The remaining subsets will be obtained similarly to those in three-dimensional data, for specific points. The analogue of subset **3.3** would be the subset **(4.3 Point in space, single depth, and time slice)**.



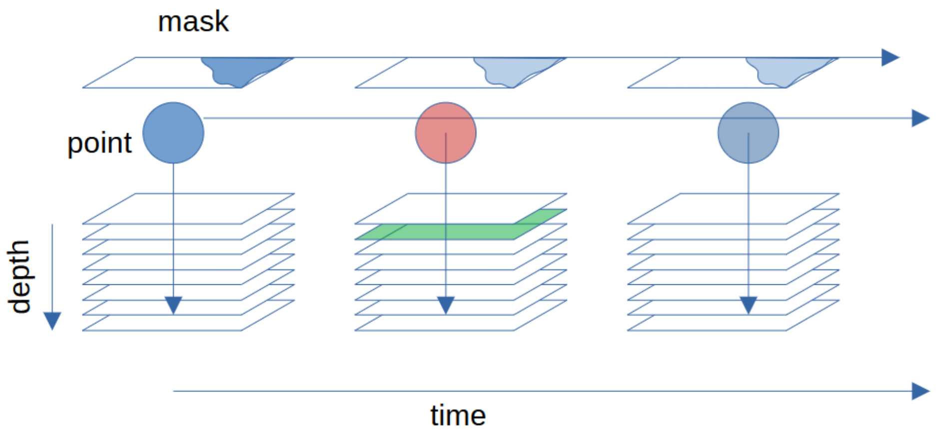
From each basic raster, the variable at a specific point will be extracted, creating a time series of data, such as salinity for each moment in time.

If we include data at different depths for each point, we obtain a subset that, for each time, will contain data for the point at each depth (**4.33 Point in space with slices of time and depth**).

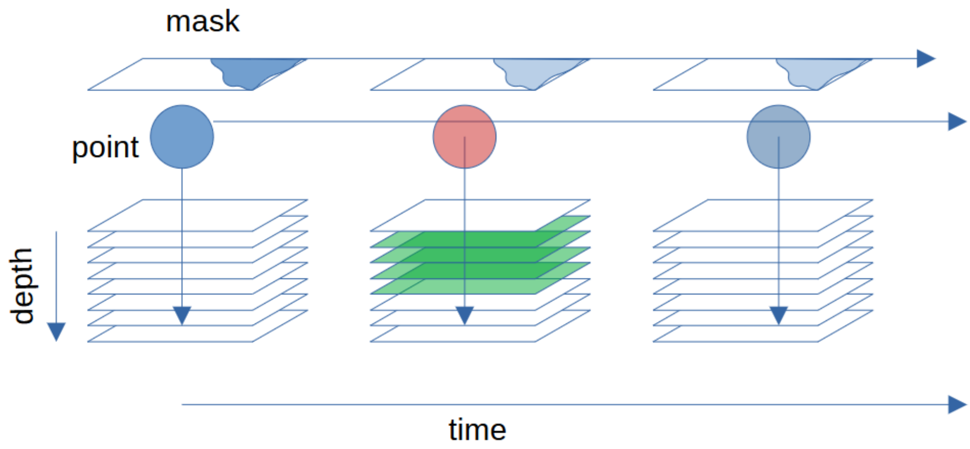


The resulting subset of data can be represented as a **depth-time cross-section**.

Restricting the data extraction at a point to a single moment in time and depth (**4.4 Point in space at a single time and depth**) will, similarly to **3.4**, yield a single value of the variable.



The final option is to extract data at a point at a specific moment in time from different depths (**4.44 Point in space with a slice of depth at a single time**). As a result, data will be retrieved that forms a **vertical profile**.



Additionally, when creating a subset, a third digit (1/0) is used to indicate the presence or absence of a mask. This notation is then used in analytical functions to inform about their availability for a specific type of data.

The fields defining the subset are described below, with the first two used to specify the variable and its modification:

- **Variable** – The variable is selected from a dropdown list. The variable ***km2*** will not appear on the list if it has been created, but access to it in the **Calculate** and **Filter** fields, as well as other variables, is provided by the syntax **ds2.variable_name**, e.g., **ds2.km2**. This variable is referred to below as **var**.
- **Calculate** – This field recalculates the current variable using an expression in the form of **variable rest_of_expression** (where only **rest_of_expression** is entered). **Calculate** also allows the inclusion of another variable from the retrieved data set, referred to as **ds2**, with the variable as **ds2.variable_name**. This variable syntax can also be used in filtering. The expression (**rest_of_expression**) takes the form of a Python expression using algebraic operators, comparison, and logical operators. If the field is left empty, the expression **variable * 1** is executed. Below are several examples of expressions entered into the **Calculate** field and their applications. The examples are taken from videos demonstrating the use of the **Xraster** plugin.

Expression	Application
- 273.5	conversion of Kelvin degrees to Celsius degrees
* 32/100	the conversion of oxygen concentration in water from mmol/m^3 (or µmol/L) to mg/L
* 80e3	total rainfall in m^3 per pixel with an area of 80 km^2 over an 24-hour period (using daily rainfall variable in mm), 1[mm]*10^{-3} m * 80[km^2]* 10^6 m^2 = 80000 m^3

(Starting an expression by multiplying by zero assigns the variable **var** a value of zero, which allows for the creation of more complex expressions. If there is a need to use a selected variable in these expressions, we proceed as with other variables by prefixing the variable name with **ds2**.)

* 0 + 1	changing all variable values to 1 (NoData remains NoData)
* 0 + (ds2.tx >= 35)	ds2.tx represents the variable tx, if the condition is met, the result will be 1, otherwise 0
* 0 + (ds2.o2 * 32 / 100 < 2)	ds2.o2 represents the concentration of oxygen in water in mmol/m^3 (or µmol/L), the expression converts the unit to mg/L and assigns a value of 1 if the concentration is less than 2, or 0 if it is greater than or equal to 2, this is useful for identifying hypoxia zones
* 0 + (ds2.o2 * 32 / 100 < 2) * 4	similar to the previous case, but it assigns a value of 4 (representing an area of 4 km^2) , to the cell if the condition is met; otherwise, the cell value is 0, summing the values provides the total area of hypoxia zones
* 0 + ds2.km2	ds2.km2 represents the area of a cell in km^2, a reference map of cell areas can be created to assess how much their sizes vary.
* 0 + (ds2.t2m - 273.15) * ds2.km2/1.33e7	ds2.t2m represents the temperature in Kelvin, the expression is used to calculate the weighted average temperature, where the weight is the cell area (ds2.km2), the weighted average value is obtained by summing the values across all cells (optionally filtered using a mask), the cell area values are divided by the total area (1.33e7) and then summed.

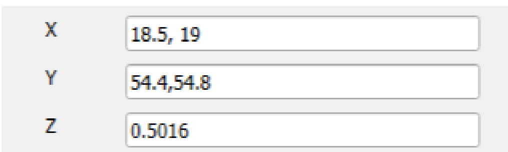
In raster GIS terminology, Calculate performs the functionality of both Raster Calculator and Reclassify.

- **X, Y, Z**

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

Y, Z – Follow the same rules as **X**.

Input format,



There are situations where the X, Y range must be entered from a higher value to a lower value (e.g., for negative coordinates). You can check the input rule using RESET. If, after entering the values and resetting, you receive "Selection failed," you should reverse the signs, for example:

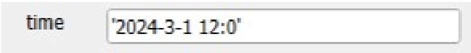
Change -89.9375, 89.9375 to 89.9375, -89.9375.

A similar issue may occur when entering Z as 0, particularly when depths start from a value like 0.5. In this case, you should provide any valid value within the relevant range.

- **time**

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

To enter the time correctly, ensure that the year, month, and day are provided in this format, without leading zeros for the month and day.



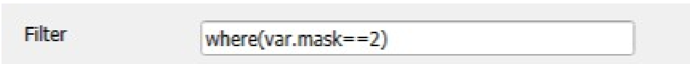
To use certain program functionalities, for a subset at a single point in time, it may be necessary to enter the time as, for example: '2004-7-1', '2004-7-1'. This format represents the same date as both the start and end time.

- **Filter**

The filter function uses the **where()** function along with comparison and logical operators (e.g., **|** for OR, **&** for AND). It can refer to the variable **var**, its mask **var.mask**, or any variable **ds2.variable**. Here are some example applications:

where((var.mask == 1) | (var.mask == 2))	This filters the area where the mask value is either 1 or 2.
where(var <= 5)	This filters the variable for values less than or equal to 5.
where(ds2.variable_name > 3)	This filters the ds2.variable_name where the value is greater than 3.

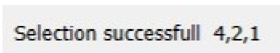
Input format,



- **Month selection**

Month selection limits data to only the months specified in the list. The input format for selecting data from December, January, February, and March would be:

Selection is made by pressing the **OK** button. A message will appear with a code corresponding to the subset type, for example:



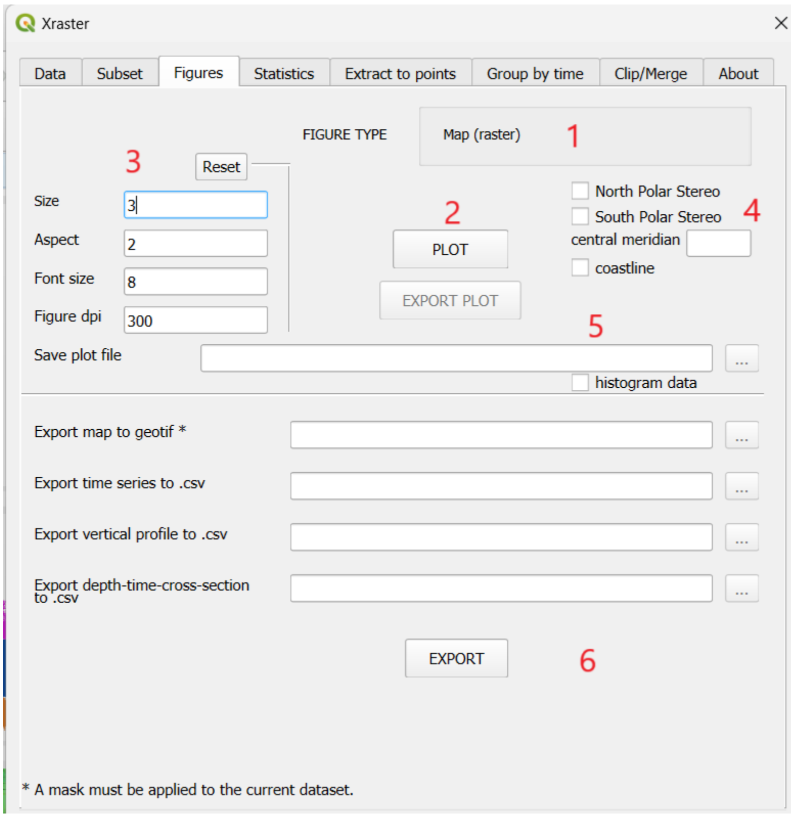
For **four dimensions, type 2** – this represents the area at a specific depth at a given time (there is a mask for the data).

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

[TOCTabs]

3. Figures

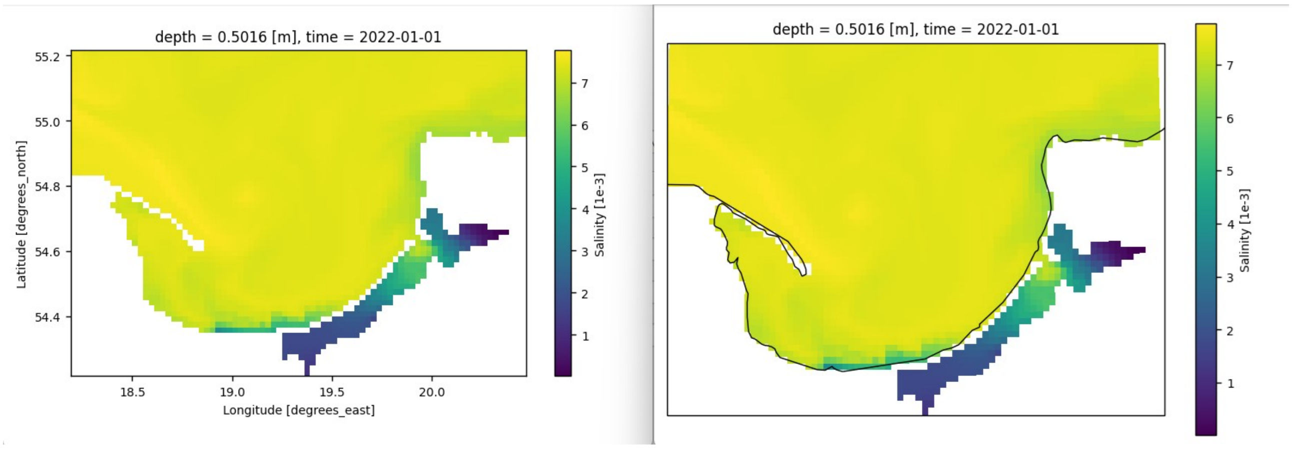
This tab is used for visualizing a data subset and conducting exploratory data analysis. It also allows for the export of any subset area (x, y) to a GIS raster layer (a mask is required) and the saving of data from the drawing as text files (for histograms, time series, vertical profiles, and depth-time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



After selecting the data and creating its subset, the FIGURE TYPE panel **(1)** will determine the type of plot based on the created data subset. The types of subsets, their codes, and the corresponding plot types are summarized in a table at the end of the subsection. To generate the plot, click PLOT **(2)**. The plot will be created in the Matplotlib GUI window with default parameters **(3)** described in the fields: **Size**, **Aspect**, and **Font size**. Size and Aspect can be changed by resizing the window or modifying values in the fields. Font size can only be adjusted by changing the value in the font size field. All values can be reset to their default settings using the RESET button. The plot can be saved either in the GUI window (diskette icon) at the resolution displayed on the screen, or by entering a name in the **Save plot file** field and clicking **EXPORT PLOT** **(5)** to save it with a higher resolution defined in the **Figure dpi** field. Additional customization of the plot is described in section 3.6.

When saving the histogram plot, there is an option to save the data from which the histogram was created (by checking the **histogram data** checkbox). The text file will contain 10,000 randomly selected data points (or fewer) and will be named after the entered histogram filename with the suffix _data (e.g., if the histogram is saved as *histzg_ice.jpg*, the data will be saved as *histzg_ice_data.csv*). The data will be saved with 100 numbers per row, separated by commas (up to 100 rows).

If the data has geographical coordinates, map visualizations can be performed using the **Polar Stereographic projection** **(4)** (for both the northern and southern hemispheres - select the corresponding checkbox). You must specify the central meridian (if not provided, it will default to 0). Additionally, you can check the **coastline** box to add the coastline. The map below shows the geographical coordinates and the Polar Stereographic projection with the coastline added.



The above option has been introduced for the visualization of polar regions. Some of the data from these areas is provided in geographical coordinates and cannot be effectively visualized without applying the appropriate projection.

There is a significant limitation with this solution. Interaction with the plot is effective for raster sizes up to 400 x 400 pixels. For larger rasters, the Matplotlib GUI requires a long waiting time. It is best not to work with rasters larger than 200 x 200 pixels. The raster size can be adjusted after input in the Data tab.

At the bottom of the tab, there are four export fields along with buttons for entering the file name. Only the fields that are applicable at the given moment are active. You need to enter the file name and click the EXPORT **(6)** button.

- **Export map to geotif** – If the plot is a map (xy surface) and a mask has been applied, it can be exported as a raster layer in .tif format (in the original coordinate system).
- **Export time series to .csv** – If the plot is a Time Series, the data can be exported to a .csv file in the following format:

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 plot is a Vertical Profile, the 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 plot is a Depth-time-cross-section, the 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

………………………………………………

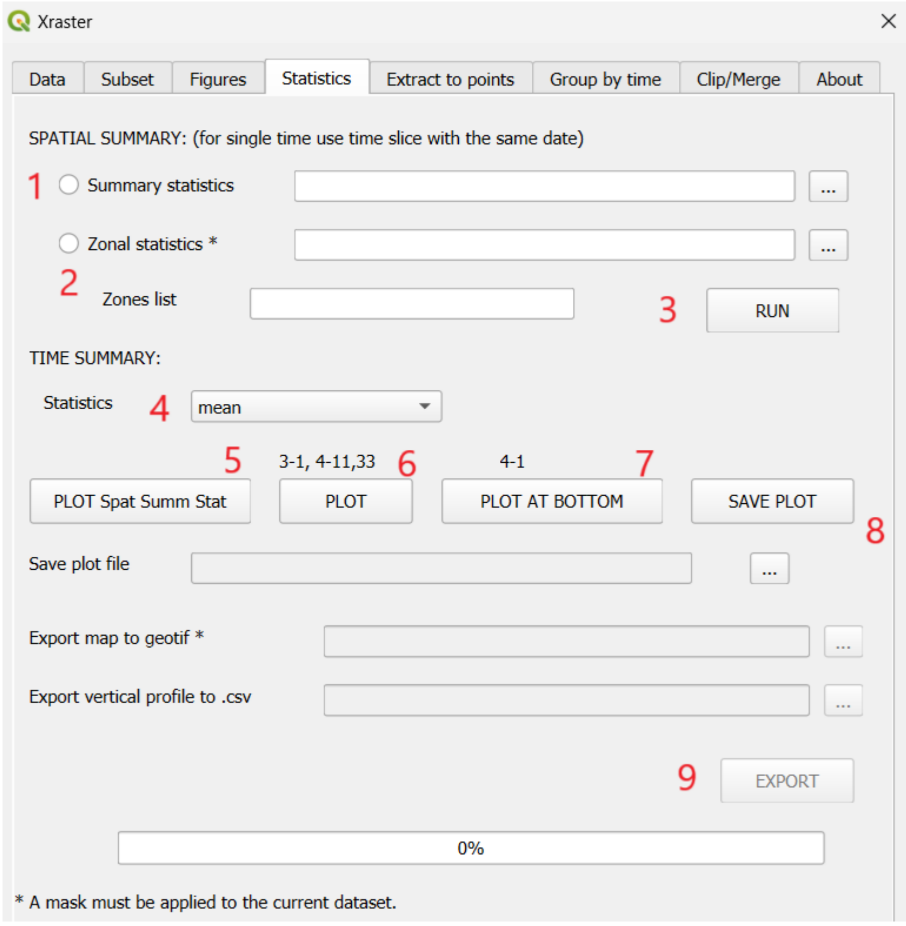
Types of subsets, codes, and their corresponding plot types:

Description of subset	Subset Type	Figure
Slices of space and time	3-1	Histogram
Slice of space at a single time	3-2	Map (basic 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 (basic 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

In raster GIS terminology, this tab is used, among other things, to implement the functionality of **Zonal Statistics** and **Cell Statistics**. It calculates statistics in space (**SPATIAL SUMMARY**) and time (**TIME SUMMARY**). Spatial statistics are calculated from the data of the entire space (x, y, z), possibly limited through filtering, separately for each time point. Time statistics calculate statistics separately for each location in space from time-varying data (similar to **Cell Statistics** in ArcGIS and **r.series** in QGIS).



Summary statistics (1) analyzes all the data from the subset for a given time point and calculates statistics for them. 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

So, if we have data x, y, z (e.g., temperature in a reservoir for a given day, at different depth levels), the average temperature will be calculated from all the data (x, t, z). You need to specify the file name to be created in field **(1)** and click **RUN (3)**.

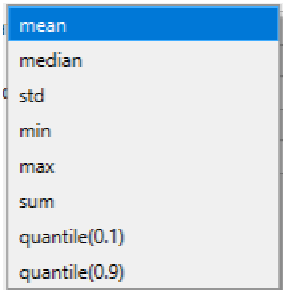
The **SUMMARY STATISTICS** results can be presented as a time series. To do this, select the desired statistic from the dropdown menu **(4)** and press the PLOT SPAT SUMM STAT button **(5)** (note that this option does not work for quantiles). This will generate a plot showing the selected summary statistic over time.

Similarly, **Zonal Statistics (2)** works by dividing the data into zones based on a mask. In the **Zones list** field,

you need to enter the zone numbers for which calculations will 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 for the calculation of statistics over time from the surface data (x, y, time) and from the points (z, time). The following statistics **(4)** can be selected:



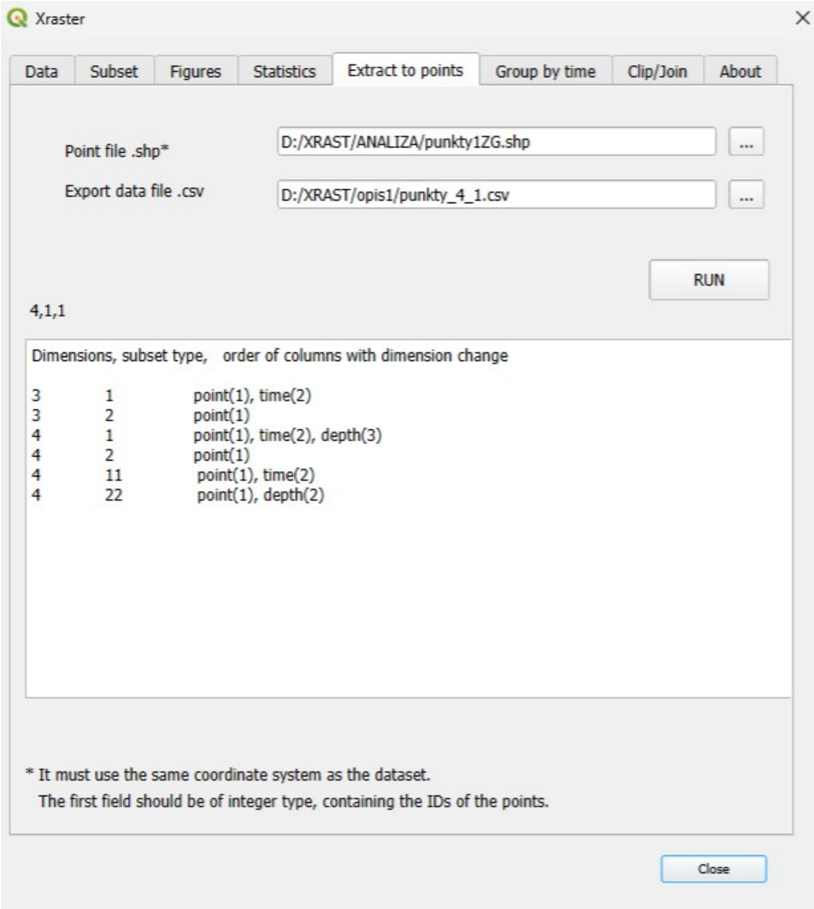
The first option (from the surface) generates a map, while the second one creates a vertical profile. Both products can be created using the PLOT **(6)** button. The map and vertical profile can be saved as an image with a specified resolution (on the **Figures** tab) using the SAVE PLOT **(8)** button. Additionally, the map can be exported to a GIS raster layer in tif 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 **(9)** button.

There is also an additional option to create maps of the variable statistics at the bottom from data of type (4-1). This option works exclusively for the following statistics: mean, median, std, min, max. Calculations are triggered by the PLOT AT BOTTOM **(7)** button. Similar to the map, the result can be saved as an image (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.

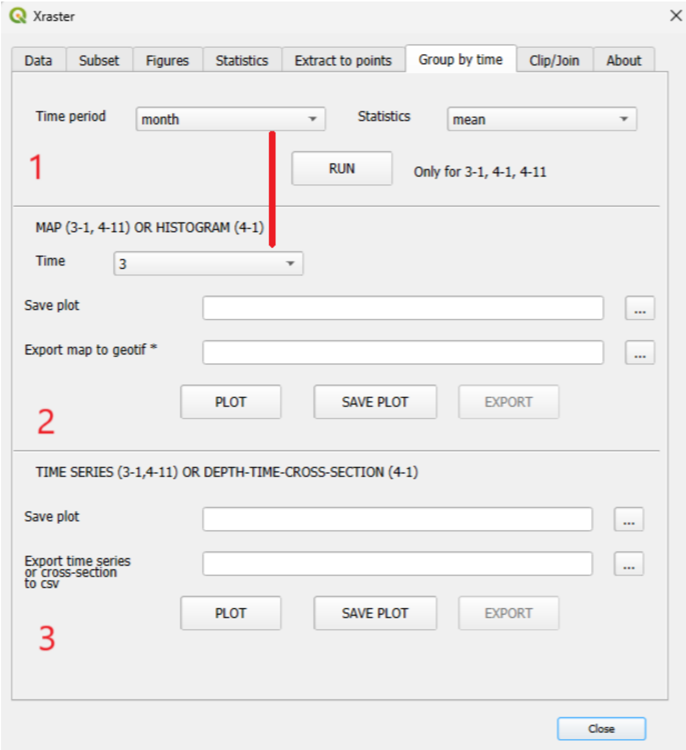
Subset description	Subset Type	Order of columns with dimension change
Slices of space and time	3-1	point(1), time(2)
Slice of space at a single time	3-2	point(1)
Slices of space,time and depth	4-1	point(1), time(2), depth(3)
Slices of space and time for a single depth	4-11	point(1), time(2)
Slice of 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 time dimension in GIS allows for the grouping of data. The key for this grouping can be based on the year, month, or climate season. Each group is then assigned a selected statistical measure, resulting in its statistical representation. It is possible to analyze the statistics of each group separately or combine them into a new data structure. This type of analysis is suitable for time series of basic rasters or full four-dimensional data.

The card consists of three parts and performs statistical analysis for groups defined by years, months, days, or climate seasons (JJA – June-August, SON – September-November, DJF – December-February, MAM – March-May). Separate groups are created from a data subset based on their keys (e.g., individual months), and then statistics (mean, median, standard deviation, minimum, maximum, sum) are calculated for each group separately. After computing the statistics, the keys with the calculated values are combined into a single subset, which can then be visualized and exported.



Grouping can be performed for data subsets such as slices of space and time for three-dimensional data and slices of space, time, and depth or slices of space and time for a single depth for four-dimensional data. The analysis always begins with the top section of the card **(1)**, where the grouping key is defined using the **Time period** option, and the desired statistic is selected under **Statistics**. After pressing the RUN button, a subset consisting of grouped data is created. In the middle section of the card **(2)**, under the **Time** field, all generated groups (e.g., statistics calculated for each month) will be displayed.

In the second part of the **middle section**, a **map or histogram** can be created using the **PLOT** button for the selected group in the **Time** field. The generated map or histogram can be saved using SAVE PLOT. If a mask is applied, the map can also be exported to a GIS raster layer via the EXPORT button.

In the bottom section of the card **(3)**, visualization of the analysis results is created for all groups together. For three-dimensional data and data at a specified depth (types **3-1, 4-11**), this visualization takes the form of time series. For four-dimensional data (4-1), it generates a depth-time cross-section (this is not created for climate seasons, and in the case of a single group, it will be a vertical profile). The visualizations can be saved using SAVE PLOT, and the data can be exported to text files with the EXPORT button.

[TOCTabs]

7. Clip/Merge

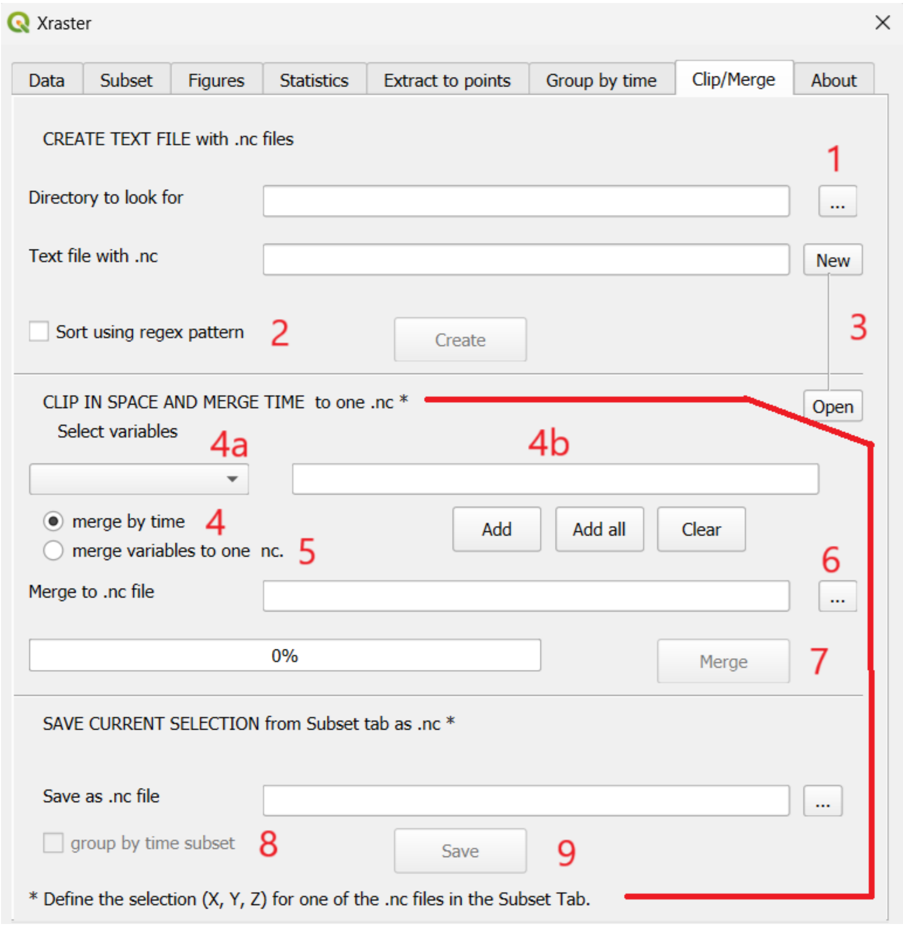
This module carries out a range of tasks related to the preparation and storage of NetCDF (.nc) data. In general, it performs data engineering on NetCDF data. The tasks include:

1. Preparing a NetCDF file tailored to a specific project from the collected data

2. Saving the current data subset (in Xraster) in NetCDF format (always with a single variable)

3. Merging saved subsets (for different variables) from the previous task into a single NetCDF file

4. Saving a "group by time" subset as a NetCDF file (for opening and working in QGIS)



1. Preparing a NetCDF file tailored to a specific project from the collected data

Depending on the method used, the data retrieval process can result in different outcomes. When using the *Copernicus Marine Toolbox*, we obtain a ready-to-analyze dataset in the form of a NetCDF (.nc) file. However, other methods may produce files that cover large areas, contain numerous variables, or be retrieved for separate time periods (days, hours). As a result, these files are not practical for direct use in a program.

Effective analysis requires merging these files into a single one while minimizing its size. This can be achieved by limiting the spatial extent and the number of variables. The module carries out these tasks in three steps:

1. The first step compiles the file paths of the retrieved data into a single text file, assuming they may be stored in a directory structure.

2. The second step extracts the area of interest and selects the necessary variables.
3. The third step merges them into a single file.

The spatial extent (clipping) of the data is taken from the current extent defined in the **Subset** module.

CREATE TEXT FILE with .nc file (first step)

The purpose of this section is to create a text file with paths to the subsequently downloaded files (.nc). Additionally, they should be sorted by time (starting from the oldest). Sorting can be done manually or automatically using the **Sort using regex pattern (2)** option.

A **regex pattern** is an abbreviation for **regular expression**, which is a sequence of characters that defines a match pattern in text. It is used to detect the date, after which the files are sorted by date. This option is experimental—if it does not work, an empty file is created.

We start by entering the directory under which (including all subdirectories) the required .nc files are located. This is done in the **Directory to look for (1)** field. Then, we enter the name of the text file in the **Text file with .nc** field using the NEW **(3)** button, where these files will be stored along with their full paths. They may require manual sorting.

If we select the **Sort using regex pattern (2)** option, a file with the same name plus the suffix **"*_sort*"** will be created next to the previously generated text file. This file will contain the time-sorted files (or be empty if sorting fails).

The text file is created using the CREATE button.

CLIP IN SPACE AND MERGE IN TIME to one .nc (second and third steps)

The purpose of this section is to download the prepared text file with time-sorted .nc files, select the required variables, trim the data to the needed dimension range, and create a single output file.

We start by selecting the option **merge by time (4)**. The file is loaded in the **Text file with .nc** field using the OPEN **(3)** button. The next step is to select variables in the **Select variables (4a)** field. We select the required variables and add them to the right-side field **(4b)**, under which there are buttons **Add, Add all, Clear**. These buttons are used to create a list of required variables (names separated by commas). The list can also be edited manually.

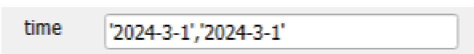
It is also necessary to define the area and possibly the depth range in the output data. These ranges are defined in the **Subset** panel, using one of the downloaded files for merging. Selection must be confirmed by clicking **OK** in the **Subset** panel.

We enter the name of the new file (.nc) (6) and click the MERGE **(7)** button.

2. Saving the current data subset (in Xraster) in NetCDF format (always with a single variable)

Saves the current selection as a new NetCDF (.nc) file. Enter the file name and press the SAVE **(9)** button.

If you want to save a subset for a single point in time, create a time interval with the same value at the start and end, e.g.,



3. Merging saved subsets (for different variables) from the previous task into a single NetCDF file

Xraster saves subsets with a single variable. If there is a need to merge them into one file, proceed as follows:

Prepare a text file containing the files to be merged (with full paths).

e.g.

D:\XRAST\temp\iconc_zg.nc

D:\XRAST\temp\sal_zg.nc

D:\XRAST\temp\temp_zg.nc

The file is loaded in the **Text file with .nc** field using the OPEN **(3)** button. Select the **merge variables to one .nc (5)** option. Enter the name of the new .nc file **(6)** and click the MERGE **(7)** button.

The merged files must have the same size (in space and time) and coordinate system.

4. Saving a "group by time" subset as a NetCDF file (for opening and working in QGIS)

The main issue with using NetCDF data in QGIS is that, although they can be opened, they often contain hundreds or even thousands of rasters, making processing extremely difficult. The solution is to reduce their number through grouping.

Grouping by time (years, months, climate seasons) using the required statistics in Xraster allows for preparing a NetCDF file optimized for easy analysis in QGIS. For example, 365 daily air temperature rasters can be aggregated into 12 monthly average rasters, which can be processed in QGIS without difficulty.

In the **Group By time** panel, create a group by clicking the **RUN** button. Then, check the **group by time subset (8)** option. Enter the file name and press the SAVE **(9)** button.

[TOC]