

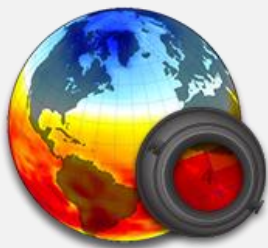
Dataset of *E. huxleyi* blooms: spatio-temporal distribution and their impact on high-latitude marine environments (1998-2016)

Data Access Tips with free software



Nansen International Environmental and Remote Sensing Center
Saint Petersburg, Russian Federation
Contact us: adm@niersc.spb.ru

Our data are distributed in NetCDF file format, which is a standard well-known scientific data representation format. For more usability we prepare data with full GDAL-support, so you can access the spatial representation of it in any popular GIS and mapping software package. This document describes three most relevant, in our opinion, options of accessing such data. All of them are freely available and cross-platform.



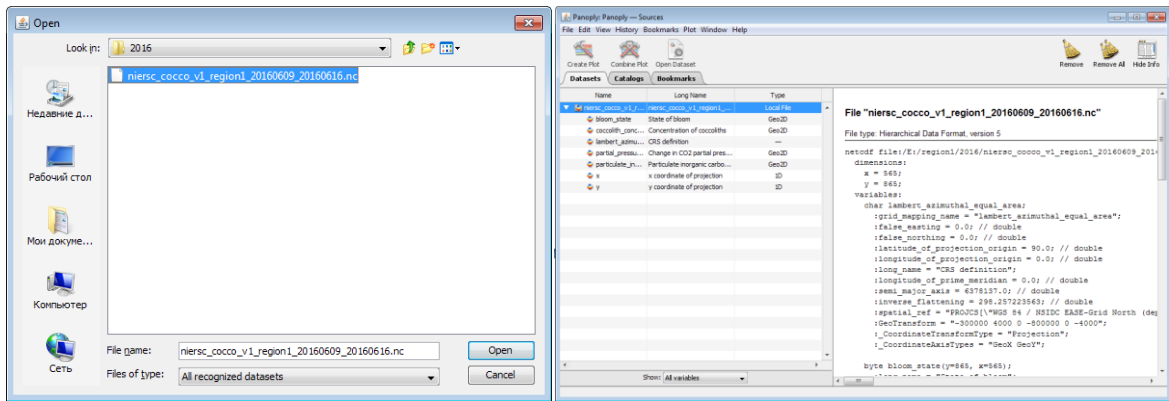
Panoply

Universal tool for viewing and plotting NetCDF, HDF, GRIB and other scientific datasets developed by Robert B. Schmunk (NASA Goddard Institute for Space Studies).

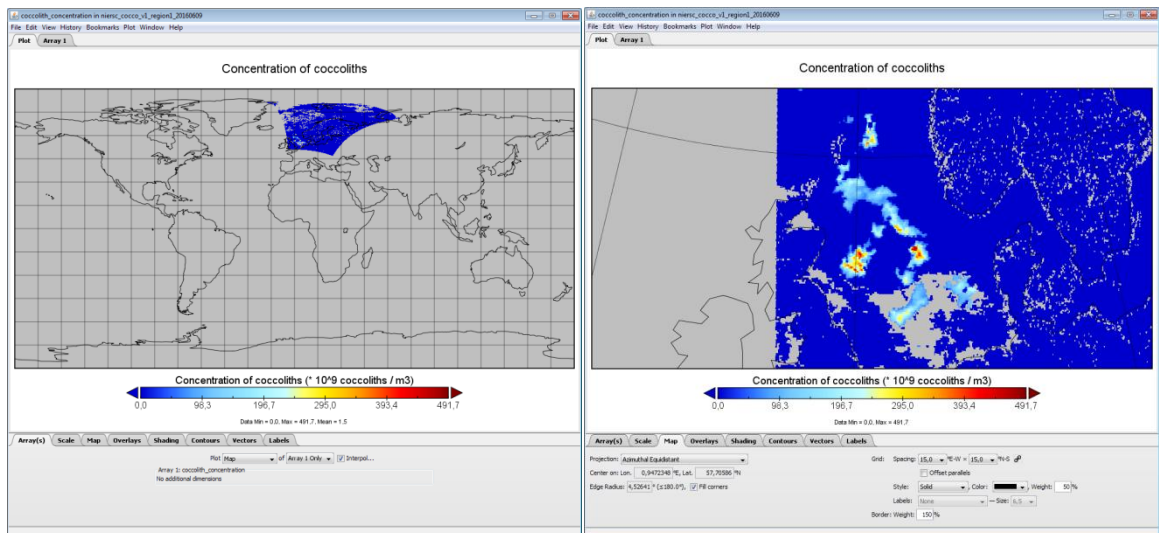
Web: <https://www.giss.nasa.gov/tools/panoply/>
Platforms: Windows, Linux, MacOS

Panoply is lightweight and easy-to-use software for quick access to NetCDF datasets. It allows to take a look at metadata and variables values, and to visualize data as plots or maps. The mapping system allows to choose the map projection, locations of interest zooming and many other things.

After installation, just run Panoply. First thing you'll see is file selecting dialog. Choose the file with the region and date you are interested in. After file selection, you'll see all data it contains. Click on the root item in the list to access the main metadata records of dataset.



Then choose the variable, for example "coccolith_concentration". On the right panel you'll see all its metadata (like units, references, etc.). Double click on the item (or push the button "Create plot") to open the dialog, where you can create a plot or a map. At the window's bottom you'll find many useful map instruments. With Ctrl+Left Mouse Button you can draw a rectangle on the map to zoom in.



You could also export any variable as CSV/CDL/Labeled text (in context menu of variable).

Panoply is great for fast data overview and understanding its relevance to you.



QGIS

The most powerful and popular open-source desktop GIS platform for today, with big community and great experience around.

Web: <https://www.qgis.org>

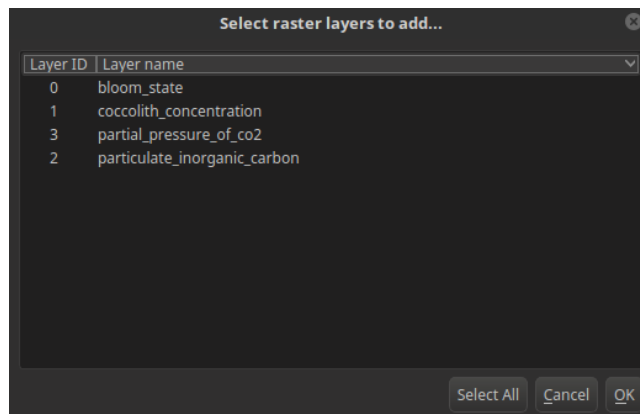
Platforms: Windows, Linux, MacOS

QGIS is a classical desktop GIS platform, which allows to operate with spatial data: from storing and visualization to processing and publishing. It's community-based and open-source, so its functionality is

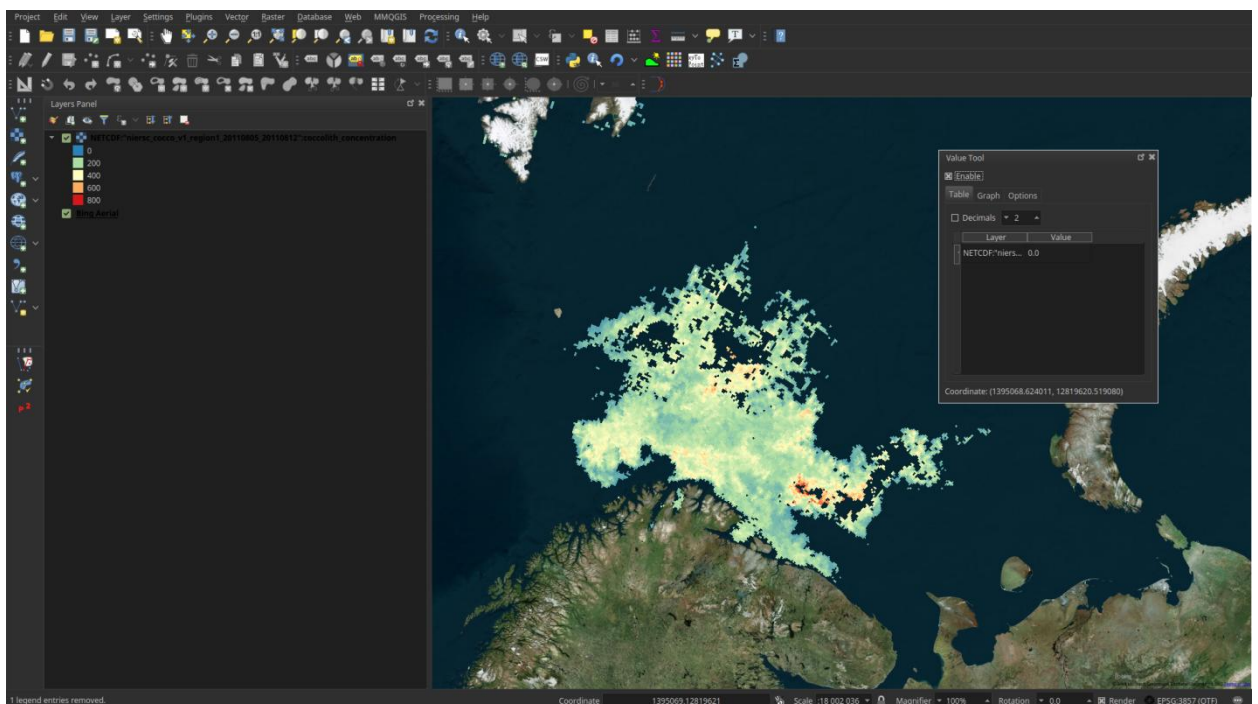
always growing up with plugins and extensions. More than that, nowadays powerful systems GRASS GIS and SAGA GIS are integrated in the QGIS interface.

With QGIS you could quickly visualize our data products with any additional spatial data (basemaps, vector and raster layers, etc.), process it with raster tools (raster arithmetics, reprojecting, zonal statistics, reclassifying and many more), prepare maps, publish them on the web, export to any raster format and so on.

To start working with our data in QGIS, just add niersc_cocco NetCDF file in QGIS with drag-n-drop or menu item "Layer - Add Layer - Add Raster Layer". QGIS will invite you to select the variable of your interest. You could choose one or several.



After that, selected datasets will be added to your QGIS project as raster layers. You can access all metadata in the context menu item "properties". You could also apply any style to your data, export them, process them with available tools and much more.





Python

Widely used in science a universal programming language with a most comprehensive ecosystem (libraries, frameworks, community) in industry.

Web: <https://www.python.org>

Platforms: Windows, Linux, MacOS and many more

Python, in fact, is a new standard in scientific programming languages, mostly because of wide and mature community and low "entry threshold". Python offers hundreds of special libraries and tools for geoscientists, including libraries to work with different dataset formats. With Python you can do everything about data: visualize it, refactor, fuse with other datasets, include in machine learning models, process with any algorithms... You could find much information on the Internet about how to apply Python for Environmental research.

Reading NetCDF files is possible with the library named NetCDF4. You can install it with a standard package utility named pip. In command line:

```
>>pip install netcdf4
```

And, then, work with NetCDF datasets directly in your code. Here is an example of how to open and visualize our dataset with Python and netcdf4 library:

```
# importing library
from netCDF4 import Dataset

# reading netCDF dataset into variable
cocco_sample = Dataset('niersc_cocco_v1_region1_20110805_20110812.nc')

# get variable metadata
cocco_sample.variables['coccolith_concentration'].units

'* 10^9 coccoliths / m3'

# get variable data
coccolith_concentration=cocco_sample.variables['coccolith_concentration'][:]
# what is inside? Some data array
print coccolith_concentration
```

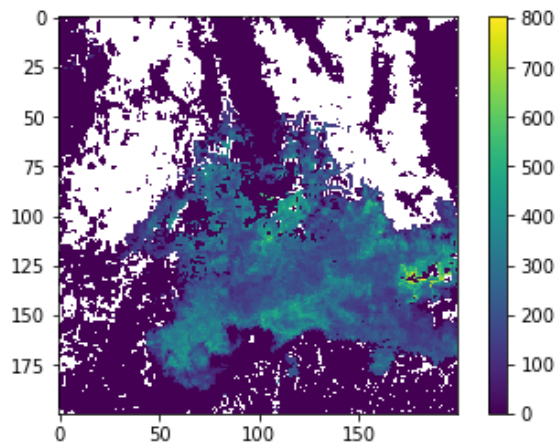
Out:

```
[[ nan  nan  nan  ...,  0.  0.  0.]
 [ nan  nan  nan  ...,  0.  0.  0.]
 [ nan  nan  nan  ...,  nan  nan  0.]
 ...,
 [ 0.  0.  0.  ...,  0.  0.  0.]
 [ 0.  0.  0.  ...,  0.  0.  0.]
```

Out:

```
[ 0.  0.  0. ...,  0.  0.  0.]  
# let's draw it. Import plotting library  
import matplotlib.pyplot as plt  
  
# draw subset of array with coccolith concentration  
plt.imshow(coccolith_concentration[100:300,200:400])  
plt.colorbar()  
plt.show()
```

Out:



```
# some basic statistics. Import library  
import numpy as np  
# Calculate statistics  
print 'Mean: %s' % np.nanmean(coccolith_concentration[100:300,200:400])  
print 'Max: %s' % np.nanmax(coccolith_concentration[100:300,200:400])  
print 'Stdev: %s' % np.nanstd(coccolith_concentration[100:300,200:400])  
print 'etc.'
```

Out:

```
Mean: 99.1991  
Max: 802.784  
Stdev: 133.578  
etc.
```