TRAINING SESSION

L3 CATDS DATA

EXPLORING AND FAST PROCESSING OF NETCDF DATA

Ahmad AL BITAR[1,2],

1. CESBIO Centre d'Etudes Spatiale de la Biosphère

2. CNRS (Centre National de Recherche Scientifique)

Table of content

I- OBJECTIVE

The objective of this exercise is to present tools that enables the user to fully understand the NetCDF format and to make fast processing (subsetting, region extraction, time series, computation, simple math) using the NCO binaries.

II – MATERIALS

-DATA

SMOS CATDS operational level3 Soil moisture:
http://www.catds.fr/
http://www.cesbio.ups-tlse.fr/SMOS_blog/
SMOS CATDS operational salinity data :
http://www.catds.fr/Expertise-Centers-CEC/Ocean-Salinity-Expertise-Center-CEC-O
SMOS CATDS research salinity data :
http://www.salinityremotesensing.ifremer.fr/news/newsmosl3salinityresearchproductsnowavailablefromthecatdscecexpertisecenter

-TOOLS AND SOFTWARES

- **NetCDF (Network Common Data Form)**
  http://www.unidata.ucar.edu/software/netcdf/docs/netcdf-install.html
  Installation:
  Debian Ubuntu: sudo apt-get install netcdf-bin libnetcdf-dev libnetcdf4
  or
  RedHat: yum install netcdf netcdf-devel
- **NCO Binaries**
  netCDF Operator (NCO) version > 4.0.9
  Download from: http://nco.sourceforge.net/
- **GrADS**
  Grid Analysis and Display System is a manipulation and visualization tool for a large array of climate datasets.
  In this session we use it to map 2D variables only .
  http://www.iges.org/grads/gadoc/index.html

- **Visualization tools**
  - **ncview**
  - **IDV :** Integrated Data http://www.unidata.ucar.edu/software/idv/
  - **Panoply:** http://www.giss.nasa.gov/tools/panoply/

EXERCISE 1: OVERVIEW OF THE CATDS NETCDF PRODUCTS

-THE CATDS PRODUCTS

The CATDS products are time synthesis, enhanced and higher-end land and sea products from SMOS. The Products are in NetCDF format. The used gridding system is the EASE 25km grid for L3SM and L3OS operational, for the research product this is a regular grid.

The CATDS products consist of a Header file: *.HDR and a NetCDF file: *.DBL. The Header file contains Metadata information about the product. The NetCDF product contains the data blocks. It is more familiar to have the Netcdf product in nc extension, the extension is expected to be modified in the future.

The content of the HDR file will not be presented here.

-THE NETCDF FORMAT

Here we check the structure and content of the NetCDF file (*.DBL;*.nc).

- Open terminal
- Enter: ncdump and check options
- cd to a L3 data product directory
  (USBStick16GB_Land/Data/SML3/MIR_CLF31A/2011/09/)
- enter: ncdump –h *SELECTED_DBL_PRODUCT_PATH* > product.cdl
- enter: gedit product.cdl

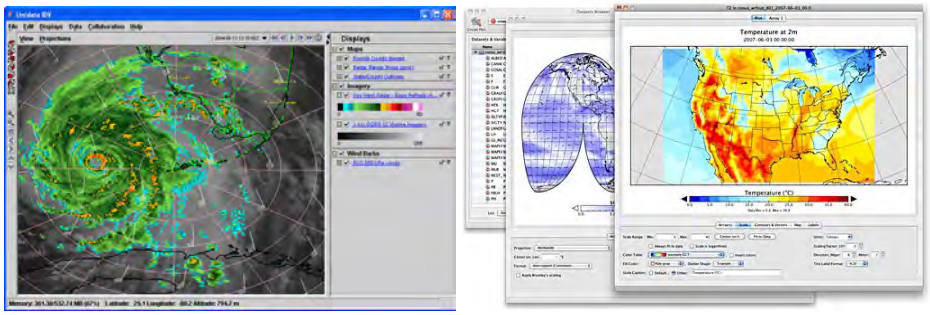
The file is divided into three parts: 1- dimensions, 2- variables, 3- global attributes. Each variable has predefined attributes enabeling the user to decode the file without need for external information.

- From the CDL file we can generate an empty netcdf file:

type: ncgen –o empty_product.nc product.cdl

- get latitudes and longitudes coordinates of the product grid: ncdump –c *SELECTED_DBL_PRODUCT* > coordinates.txt
- extract a variable to text format: ncdump –v Soil_Moisture *SELECTED_DBL_PRODUCT* > Soil_Moisture.txt
- Visualize the NetCDF product content : ncview *SELECTED_DBL_PRODUCT*

In ncview select field and make plot. ncview doesn't offer advance navigation options , use Integrated Data Viewer (IDV) or Panoply for interactive visualization .



IDV and Panoply interfaces

Table giving explanations of the NetCDF format:

```
{
dimensions:
  dim1 =  ;
  dim2 =  ;
  dim3 =  ;
variables:
  short VariableX (dim1, dim2) ;
            VariableX:_FillValue
            VariableX:long_name
            VariableX:units
            VariableX:add_offset
            VariableX:scale_factor
            VariableX :_Unsigned
// global attributes:
            :institution = "SMOS CATDS Processing Chain" ;
            :references = "CAT-DDI-CT-00020-CG" ;
            :source = "CATDS L3SM Optimal Processor" ;
            :title = "UDP L3SM Output Product" ;
            :ease_origin_lat = -85.31227f ;
            :ease_origin_lon = -179.8698f ;
            :ease_resolution = 25.f ;
            :ease_projection = "cylindrical" ;
            :ease_global = "yes" ;
            :conventions = "CF-1.4" ;
            :creation_date = "2012-03-21T08:25:21" ;
            :product_version = "1.0" ;
            :netcdf_version_id = "3.6.2" ;
}
```

Example of a variable definition

```
short Soil_Moisture(lat, lon) ;
        Soil_Moisture:_FillValue = -32768s ;
        Soil_Moisture:long_name = "Retrieved soil moisture value" ;
        Soil_Moisture:units = "m3.m-3" ;
        Soil_Moisture:add_offset = 0. ;
        Soil_Moisture:scale_factor = 3.05185094759972e-05 ;
```

EXERCISE 1: MAPPING USING GRADS

The objective of this exercise is to generate an image from a netCDF file. MIR_CLF3MD

-Go to GrADS scripts directory

Run:

./map_L3SM.sh Soil_Moisture *SELECTED_DBL_PRODUCT OUTPUT_DIRECTORY*

-make images for all products in a directory

Run:

./map_all.sh Soil_Moisture *SELECTED_DBL_PRODUCT OUTPUT_DIRECTORY*

EXERCISE 2: SPATIAL COMPUTATIONS

The objective of this exercise is to extract a region of interest defined by a box of lat and lons

-DEFINE A REGION OF INTEREST INCLUDING YOUR STUDY SITE

- Define a region of interest defined:

minimum_lat
maximum_lat
minimum_lon
maximum_lon

IMPORTANT: PUT a Float (e.g. 40.0) and not an integer (40) ncecat will interprete the integer as a node index
INPUT_PRODUCT: /home/smos/USBStick16Gb_Land/Data/SML3/MIR_CLF3UA/2011/09/
OUTPUT_PRODUCT: /home/smos/output.nc

Run in terminal:

```
ncea -d lat,minimum_lat,maximum_lat -d lon,minimum_lon,maximum_lon   INPUT_PRODUCT OUTPUT_PRODUCT
```

- save the cdl file of the extracted product: ncdump –h OUTPUT_PRODUCT > subset.cdl
- save the cdl file of the original product: ncdump –h INPUT_PRODUCT > original.cdl
- compare the cdl of the files: diff original.cdl subset.cdl
- you can also ncdump –c OUTPUT_PRODUCT to get the new coordinates
- and ncview OUTPUT_PRODUCT to visualize it

type:

```
cd  /home/smos/USBStick16Gb_Land/Data/SML3/MIR_CLF3UA/2011/09/
for f in ./*.DBL
do
        ncea -d lat,minimum_lat,maximum_lat -d lon,minimum_lon,maximum_lon   ${f}  ~/${f}.nc
done
```

EXERCISE 2: TIME AVERAGE AND DIFFERENCES USING OCEAN PRODUCTS

- MANIPULATION OF CATDS - CEC DATA

- Use ncdump -h to extract name of variable : sss

```
ncdump -h SSS_SMOS_L3_Monthly_0.25deg_CATDS_CECOS_2010.07_V02.nc | grep –i salinity
```

- Average over July to September 2010

```
ncea -v sss SSS_SMOS_L3_Monthly_0.25deg_CATDS_CECOS_2010.0[789]_V02.nc SSS_2010.nc
```

- Average over July to September 2011

```
ncea -v sss SSS_SMOS_L3_Monthly_0.25deg_CATDS_CECOS_2011.0[789]_V02.nc SSS_2011.nc
```

- Compute difference between 2011 and 2010

```
ncdiff SSS_2011.nc SSS_2010.nc SSS_2011-2010.nc
```

- Visualize the difference with ncview

```
ncview SSS_2011-2010.nc
```

- MANIPULATION OF - OP DATA - ASCENDING ORBIT

-use of ncdump -h to extract name of variable : Mean_Sea_Surface_Salinity

```
ncdump -h SM_RE01_MIR_CSF3PA_20100701T000000_20100731T235959_240_001_7.DBL
```

- Average over July to September 2010

```
ncea -v  Mean_Sea_Surface_Salinity SM_RE01_MIR_CSF3PA_20100[789]*_001_7.DBL SSS_A_CATDS_2010.nc
```

- Average over July to September 2011

```
ncea -v  Mean_Sea_Surface_Salinity SM_RE01_MIR_CSF3PA_20110[789]*_001_7.DBL SSS_A_CATDS_2011.nc
```

- Difference between 2011 and 2010

```
ncdiff SSS_A_CATDS_2011.nc SSS_A_CATDS_2010.nc SSS_A_CATDS_2011-2010.nc
```

- Visualize the difference with ncview

```
ncview SSS_A_CATDS_2011-2010.nc
```

-ANALYSIS USING CATDS - OP DATA - ASCENDING+DESCENDING ORBIT

- Use of ncdump -h to extract name of variable : Mean_Sea_Surface_Salinity

```
ncdump -h SM_RE01_MIR_CSF3P__20100701T000000_20100731T235959_240_001_7.DBL
```

-Average over July to September 2010

```
ncea -v  Mean_Sea_Surface_Salinity SM_RE01_MIR_CSF3P__20100[789]*_001_7.DBL SSS___CATDS_2010.nc
```

- Average over July to September 2011

```
ncea -v  Mean_Sea_Surface_Salinity SM_RE01_MIR_CSF3P__20110[789]*_001_7.DBL SSS___CATDS_2011.nc
```

- Compute difference between 2011 and 2010

```
ncdiff SSS___CATDS_2011.nc SSS___CATDS_2010.nc SSS___CATDS_2011-2010.nc
```

-Show difference with ncview

```
ncview SSS___CATDS_2011-2010.nc
```

EXERCISE 3: EXTRACT A TIMESERIES OVER ONE NODE

- find index of EASE grid node index:

Go to this training path and use select_latlon:

```
./select_latlon  SELECTED_DBL_PRODUCT    LATITUDE    LONGITUDE
```

you get the index of the EASE Grid

-go to a soil moisture product dir and extract time series over grid index:

```
ncecat -A -u time -d lat,lat_index -d lon,lon_idx -v Soil_Moisture,Mean_Acq_Time_Days -h SM*.DBL -o ~/SM_timeseries.nc
```

IMPORTANT: LAT_INDEX and LON_INDEX, the lat and lon need to be intered as integer

-A mean appending each extraction to the ouput product
-u time means that concatenation will be done ove time dimension
-d is the dimension on which we extract
(type ncecat in terminal to see significance of all options –h –o –d –u -A)
-Go to home directory:

```
ncview the SM_timeseries.nc
```

EXERCISE 4: TIME COMPUTATIONS

-EXTRACT A FIELD OVER TIME

```
ncecat -A -u time -d lat,lat_min, lat_max -d lon,lon_min, lon_max -v Soil_Moisture,Mean_Acq_Time_Days -h SM*.DBL -o ~/subset_soil_moisture.nc
```

-RESAMPLING IN TIME:

```
ncks  -F  -d  time,1,,8   input_file   output_file
```

-AVERAGING IN TIME:

```
ncra  -F  -d  time,1,,1  input_file   output_file
```