Analysis Methods in Atmospheric and Oceanic Science

AOSC 652

HDF & NetCDF files
Week 11, Day 1

- **Today: HDF & NetCDF file description**
- **Wed: Advanced Plotting, Python and IDL**
- **Fri: HW #10 review**

7 Nov 2016

AOSC 652: Analysis Methods in AOSC

File Formats

- So far we've only used ASCII files.
- For many years, this is how data was archived.
- With the advent of larger, complex data sets, binary file formats have come into favor
- Two most commonly used binary file formats are **HDF** and **netCDF**
- An ability to read (or extract information) from binary files is ESSENTIAL for a graduate student in Atmospheric and Oceanic Science

AOSC 652: Analysis Methods in AOSC

HDF: *Hierarchical Data Format*

- Developed by the National Center for Supercomputing Applications
- More than a “file format”; HDF includes a software library and various applications
- Several versions of HDF (hdf,hdf4, and hdf5) but most often hdf5 is used in AOSC
- HDF5 files have extension *.h5 or *.he5
- HDF homepage: http://www.hdfgroup.org
  - Can access support, file descriptions, and software to extract information from hdf files at the linux command line
  - FORTRAN, MATLAB, IDL, or Python

AOSC 652: Analysis Methods in AOSC

Please copy

~rjs/aosc652/week_11/omi/ *.*

to your working directory, and at the command line type:

% **hdfview**

A gui utility should pop up. Under the file menu, navigate to your current working directory and double-click the OMI*he5 file you just copied

Click on HDFEOS ... GRIDS ... OMI Column Amount ... Data Fields ...
double click on ColumnAmountO3

Under the “Table” menu on the new spread sheet, select **Export Data to File**

You've now created an ascii file that contains the total column ozone for this date

AOSC 652: Analysis Methods in AOSC

To access he5 file content from the command line, we can use software provided by the HDF Group. More information can be found here

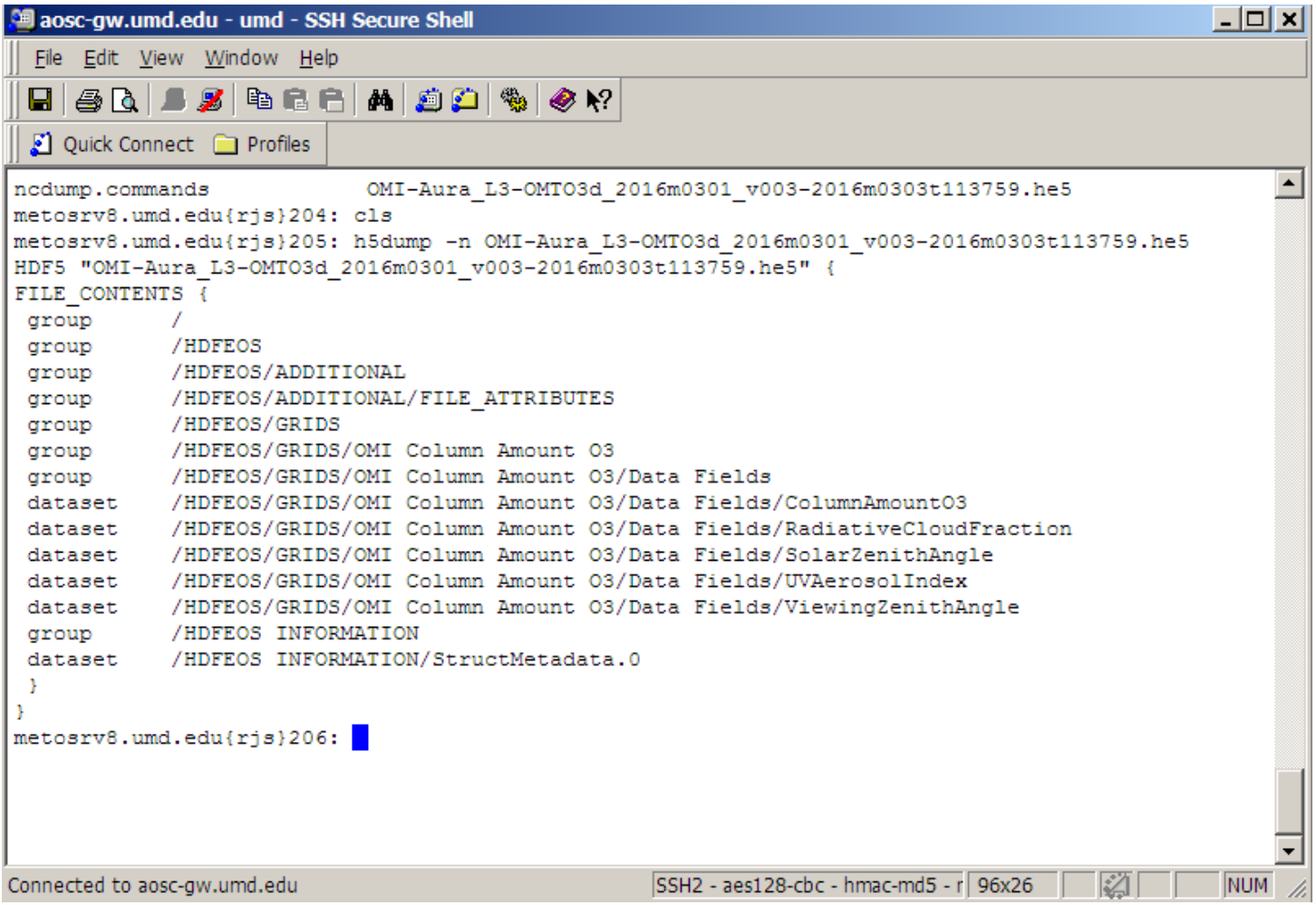
http://www.hdfgroup.org/HDF5/doc/RM/Tools.html

To determine the structure of the file, type we'll use the h5dump command with the -n option

AOSC 652: Analysis Methods in AOSC

The HDF Group

% h5dump -n OMI-Aura_L3-OMTO3d_2016m0301*.he5



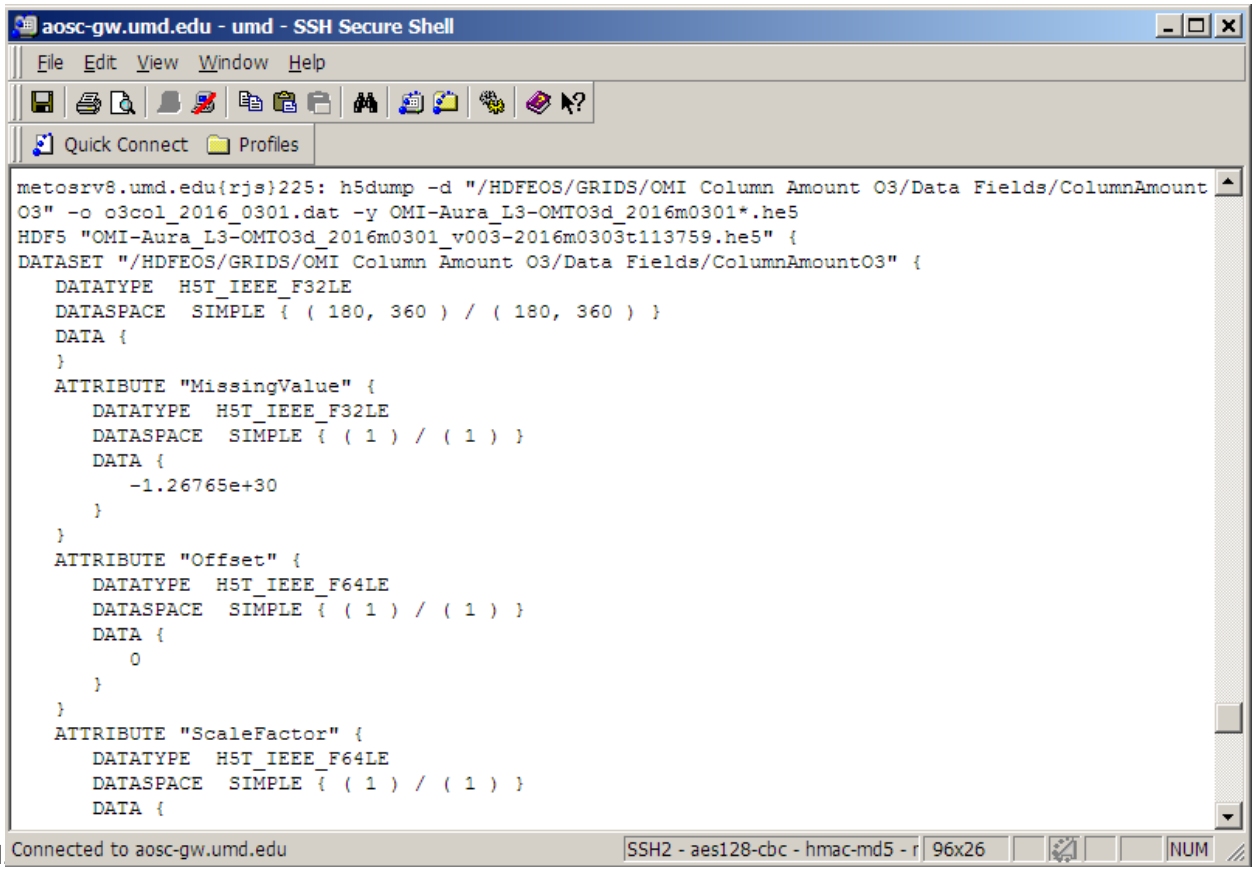
AOSC 652: Analysis Methods in AOSC

The HDF Group

To extract data from this file:

```
h5dump –d  "/HDFEOS/GRIDS/OMI Column Amount O3/Data Fields/ColumnAmountO3"
       –o  o3col_2016_0301.dat  –y  OMI–Aura_L3–OMTO3d_2016m0301*.he5
```

All one line in linux



AOSC 652: Analysis Methods in AOSC

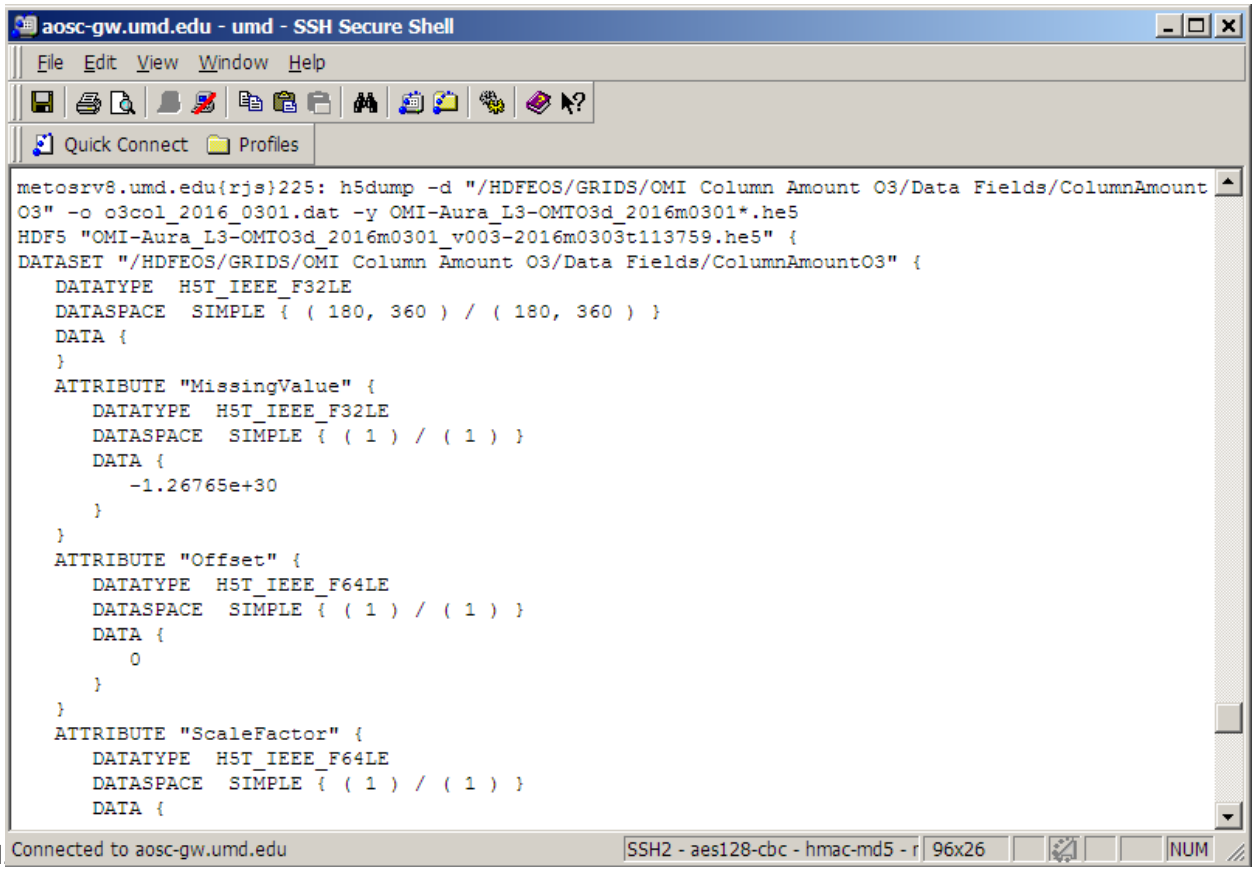
The HDF Group

To extract data from this file:

```
h5dump –d  "/HDFEOS/GRIDS/OMI Column Amount O3/Data Fields/ColumnAmountO3"
       –o  o3col_2016_0301.dat  –y  OMI–Aura_L3–OMTO3d_2016m0301*.he5
```

This command places ozone column data into an ascii file called o3col_2016_0301.dat

All one line in linux



AOSC 652: Analysis Methods in AOSC

FORTRAN:

The HDF group has support for interfacing hdf files with FORTRAN 90. Support for FORTRAN 77 is limited.

Can read hdf files directly in FORTRAN 90: various software libraries provided at http://www.hdfgroup.org/ftp/HDF5/examples/src-html/f90.html

AOSC 652: Analysis Methods in AOSC

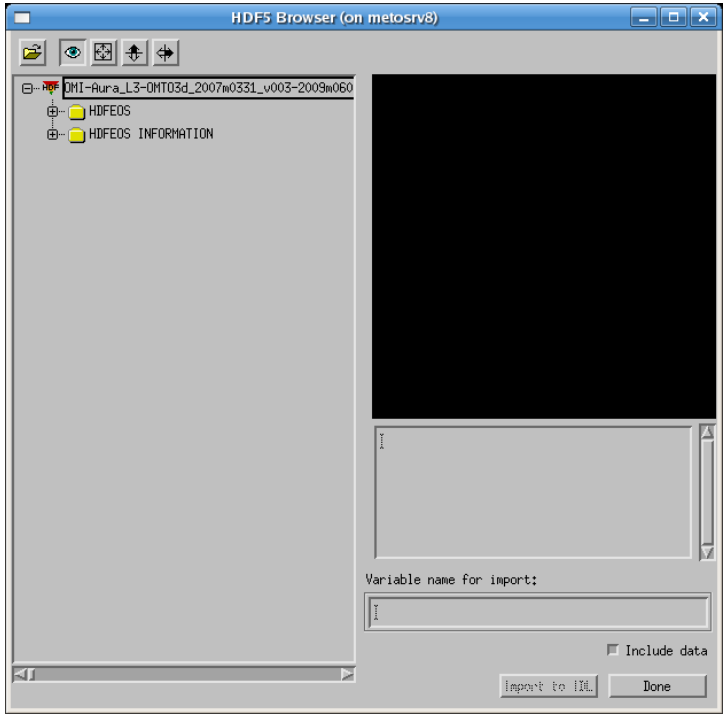
The HDF Group

IDL has some nice functions to read hdf files

Start IDL

To view information about the hdf file, type

```
browse=h5_browser('OMI-Aura_L3-OMTO3d_2016m0301*.he5')
```



AOSC 652: Analysis Methods in AOSC

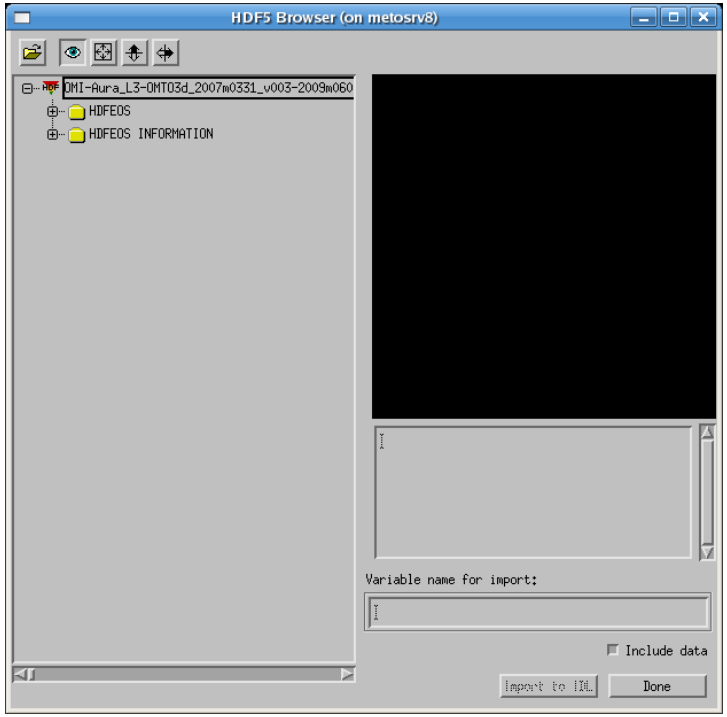
The HDF Group

IDL has some nice functions to read hdf files

Start IDL

To view information about the hdf file, type

result=h5_browser('filename')

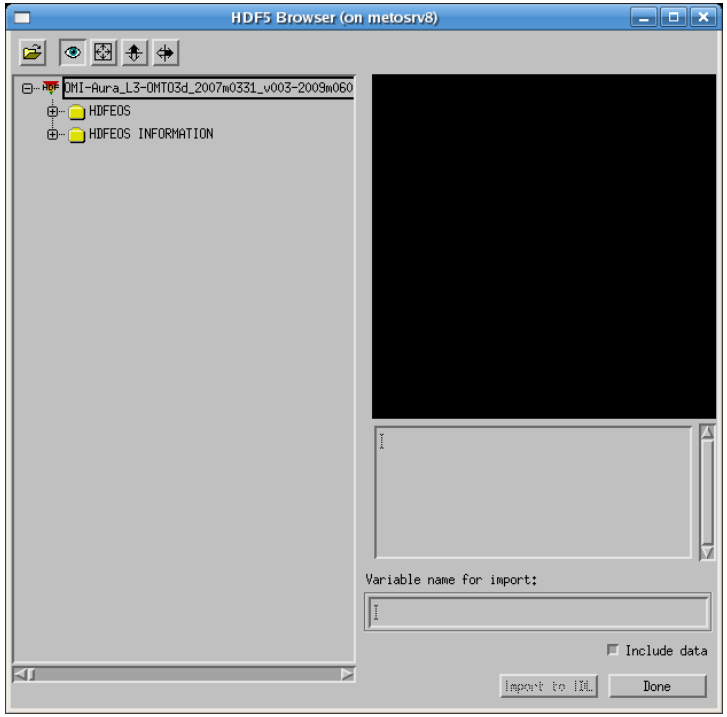


AOSC 652: Analysis Methods in AOSC

The HDF Group

IDL has some nice functions to read hdf files.

You can click through the menus to see what fields are available as well as the metadata (information about the data sets, such as variables names, missing data values, etc). This information is useful when reading in data into an IDL program.



AOSC 652: Analysis Methods in AOSC

IDL

Please have a look at **read_omi_hdf5.pro** (copied from my omi directory)

To read in the data, we need to extract the data from the .he5 file and put into a variable

```
result=h5_parse('OMI-Aura_L3-OMTO3d_2016m0301_v003-2016m0304t143959.he5',/read_data)
```

Contains all of the data in the HDF5 file

AOSC 652: Analysis Methods in AOSC

IDL

Please have a look at **read_omi_hdf5.pro** (copied from my omi directory)

To read in the data, we need to extract the data from the .he5 file and put into a variable

```
result=h5_parse('OMI-Aura_L3-OMTO3d_2016m0301_v003-2016m0304t143959.he5',/read_data)

o3_column=result.hdfeos.grids.OMI_Column_Amount_O3.Data_Fields.ColumnAmountO3._data
```

Variable o3_column contains only ozone column data from HDF5 file

AOSC 652: Analysis Methods in AOSC

IDL

Please have a look at **read_omi_hdf5.pro** (copied from my omi directory)

To read in the data, we need to extract the data from the .he5 file and put into a variable

```
result=h5_parse('OMI-Aura_L3-OMTO3d_2016m0301_v003-2016m0304t143959.he5',/read_data)

o3_column=result.hdfeos.grids.OMI_Column_Amount_O3.Data_Fields.ColumnAmountO3._data
```

Variable o3_column contains only ozone column data from HDF5 file

This syntax allows for selection of fields of interest of data file (subsetting)

AOSC 652: Analysis Methods in AOSC

IDL

Please have a look at **read_omi_hdf5.pro** (copied from my omi directory)

To read in the data, we need to extract the data from the .he5 file and put into a variable

```
result=h5_parse('OMI-Aura_L3-OMTO3d_2016m0301_v003-2016m0304t143959.he5',/read_data)

o3_column=result.hdfeos.grids.OMI_Column_Amount_O3.Data_Fields.ColumnAmountO3._data
```

Variable o3_column contains only ozone column data from HDF5 file

This syntax allows for selection of fields of interest of data file (subsetting)

Can then analyse and display the ozone column data

AOSC 652: Analysis Methods in AOSC

IDL

Please have a look at **read_omi_hdf5.pro** (copied from my omi directory)

To read in the data, we need to extract the data from the .he5 file and put into a variable

```
result=h5_parse('OMI-Aura_L3-OMTO3d_2016m0301_v003-2016m0304t143959.he5',/read_data)

o3_column=result.hdfeos.grids.OMI_Column_Amount_O3.Data_Fields.ColumnAmountO3._data
```

Variable o3_column contains only ozone column data from HDF5 file

This syntax allows for selection of fields of interest of data file (subsetting)

Can then analyse and display the ozone column data

Let's run read_omi_hdf5.pro

AOSC 652: Analysis Methods in AOSC

MATLAB

Please have a look at **read_omi_hdf5.m** (copied from my omi directory)

The procedure to read hdf5 files into MATLAB is a little more complicated

```
% Open the HDF5 file
file_id=H5F.open (FILE_NAME,'H5F_ACC_RDONLY','H5P_DEFAULT')
DATAFIELD_NAME='/HDFEOS/GRIDS/OMI Column Amount O3/Data Fields/ColumnAmountO3'

% Read the dataset
data_id=H5D.open(file_id,DATAFIELD_NAME);
data=H5D.read (data_id,'H5T_NATIVE_DOUBLE','H5S_ALL','H5S_ALL','H5P_DEFAULT');
```

Contains total column ozone

AOSC 652: Analysis Methods in AOSC

MATLAB

Please have a look at **read_omi_hdf5.m** (copied from my omi directory)

The procedure to read hdf5 files into MATLAB is a little more complicated

```
% Open the HDF5 file
file_id=H5F.open (FILE_NAME,'H5F_ACC_RDONLY','H5P_DEFAULT')
DATAFIELD_NAME='/HDFEOS/GRIDS/OMI Column Amount O3/Data Fields/ColumnAmountO3'

% Read the dataset
data_id=H5D.open(file_id,DATAFIELD_NAME);
data=H5D.read (data_id,'H5T_NATIVE_DOUBLE','H5S_ALL','H5S_ALL','H5P_DEFAULT');
```

For more information regarding these commands, please visit:

http://www.hdfgroup.org/HDF5/doc/RM/RM_H5D.html

http://www.mathworks.com/help/matlab/hdf5-files.html

AOSC 652: Analysis Methods in AOSC

MATLAB

Please have a look at **read_omi_hdf5.m** (copied from my omi directory)

The procedure to read hdf5 files into MATLAB is a little more complicated

```
% Open the HDF5 file
file_id=H5F.open (FILE_NAME,'H5F_ACC_RDONLY','H5P_DEFAULT')
DATAFIELD_NAME='/HDFEOS/GRIDS/OMI Column Amount O3/Data Fields/ColumnAmountO3'

% Read the dataset
data_id=H5D.open(file_id,DATAFIELD_NAME);
data=H5D.read (data_id,'H5T_NATIVE_DOUBLE','H5S_ALL','H5S_ALL','H5P_DEFAULT');
```

For more information regarding these commands, please visit:

http://www.hdfgroup.org/HDF5/doc/RM/RM_H5D.html

http://www.mathworks.com/help/matlab/hdf5-files.html

Let's have only the instructor run read_omi_hdf5.m (so we don't crash the system)

AOSC 652: Analysis Methods in AOSC

netCDF: *network common data format*

File format system developed by the University Corporation for Atmospheric Research in Boulder, Co

http://www.unidata.ucar.edu/software/netcdf/

netCDF files have *.nc extension

A variety of netCDF interfaces are available and installed on metosrv

To access, as with HDF, can run either command line scripts or can use readers (programs) in FORTRAN, MATLAB, IDL, or Python

.

AOSC 652: Analysis Methods in AOSC

Please copy:

~rjs/aosc652/week_11/start08/*.*

Contains file: start08_rf01_AWAS_merged_final_v03.nc

To read the header information in a netCDF from the command line:

ncdump -h start08_rf01_AWAS_merged_final_v03.nc

To look at the structure of the file:

ncdump -c start08_rf01_AWAS_merged_final_v03.nc

To extract specific variables and send to an output ascii format file:

ncdump -v LATC start08_rf01_AWAS_merged_final_v03 > latitude_output.dat

AOSC 652: Analysis Methods in AOSC

FORTRAN:

Often in the research world people extract data from netCDF into various ASCII files, then read using codes similar to those used in class

Can read netCDF files directly in FORTRAN: various software libraries provided at http://www.unidata.ucar.edu/software/netcdf/docs/netcdf-f77/

AOSC 652: Analysis Methods in AOSC

MATLAB

Please have a look at **read_start08_netCDF.m** which you have copied

As we in IDL, we must obtain a file ID

```
ncid=netcdf.open('filename','NOWRITE')
```

To inquire as to what variables are in the file:

```
name=netcdf.inqVar(ncid,#)
```

To extract the data:

```
var1=netcdf.getvar(ncid,#)
      or
var1id=netcdf.varinqID(ncid,name)
var1=netcdf.getvar(ncid,var1id)
```

Close the file

```
netcdf.close(ncid)
```

AOSC 652: Analysis Methods in AOSC

MATLAB

Let's run **read_start08_netCDF.m**

This program is written in a fashion **VERY** similar to IDL

Should see a plot similar to the one we created in IDL

AOSC 652: Analysis Methods in AOSC

Different research groups use various file formats.

Many satellite teams use HDF while modelling groups prefer netCDF.

HDF:

- Aura (OMI, MLS, TES, HIRDLS)
- Aqua (AMSR-E, MODIS, AMSU-A, CERES, MODIS)

NetCDF:

- Whole-Atmosphere Community Climate Model (WACCM)
- Regional Air Quality Modeling System (RAQMS)
- Goddard Chemistry Aerosol Radiation and Transport model (GOCART)
- Goddard Earth Observing System Model v5 (GEOS-5)
- Community Multiscale Air Quality Modeling System (CMAQ)

AOSC 652: Analysis Methods in AOSC

Different research groups use various file formats.

Many satellite teams use HDF while modelling groups prefer netCDF.

HDF:

Aura (OMI, MLS, TES, HIRDLS)
Aqua (AMSR-E, MODIS, AMSU-A, CERES, MODIS)

NetCDF:

Whole-Atmosphere Community Climate Model (WACCM)
Regional Air Quality Modeling System (RAQMS)
Goddard Chemistry Aerosol Radiation and Transport model (GOCART)
Goddard Earth Observing System Model v5 (GEOS-5)
Community Multiscale Air Quality Modeling System (CMAQ)

NASA Aircraft Measurement Community uses other formats, the so-called Gaines/Hipskind (or ICARTT) suite of ASCII file formats

http://www-air.larc.nasa.gov/missions/etc/IcarttDataFormat.htm

Break time

Today the Python folks will go to room 3408, because Ross will lead the next segment for IDL and he is all setup on this computer.

IDL

First copy **~rjs/idl/ncdf_browser.zip**
into your **idl** "library" directory.

Unzip the file:

unzip ncdf_browser.zip

Go back to your working directory

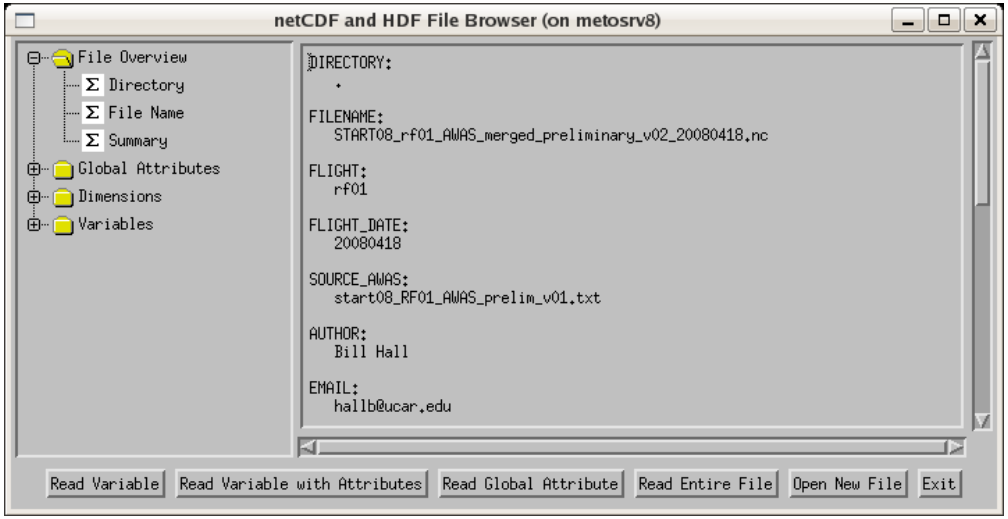
Start IDL and type

ncdf_browser,'start08_rf01_AWAS_merged_final_v03.nc'

52: Analysis Methods in AOSC

IDL

Hopefully, you'll see something like this. You can use this browser to read the metadata (dimensions, fill values, etc.)



52: Analysis Methods in AOSC

IDL

Please have a look at **read_start08_netCDF.pro** which you have copied

To read data in the *.nc file (either in an IDL program or at the command line), need to first get ID number for the *.nc file.

cdfid=ncdf_open('filename') This is similar to assigning a unit number in Fortran (i.e., unit=99)

Use this ID number (variable cdfid) and names of variables to extract data from the *.nc file

ncdf_varget,cdfid,'variable_name', your_new_name_for_this_dataset

example:

ncdf_varget,cdfid,'LATC',lat_airplane

52: Analysis Methods in AOSC

IDL

Please have a look at **read_start08_netCDF.pro** which you have copied

To read data in the *.nc file (either in an IDL program or at the command line), need to first get ID number for the *.nc file.

cdfid=ncdf_open('filename')

Your choice for this variable name

Use this ID number (variable cdfid) and names of variables to extract data from the *.nc file

ncdf_varget,cdfid,'variable_name', your_new_name_for_this_dataset

example:

ncdf_varget,cdfid,'LATC',lat_airplane

52: Analysis Methods in AOSC

IDL

Please have a look at **read_start08_netCDF.pro** which you have copied

This program reads data from *.nc file and assigns variable names

```
filename='start08_rf01_AWAS_merged_final_v03.nc'
cdfid=ncdf_open(filename)
ncdf_varget,cdfid,'GGALT',alt_hia
ncdf_varget,cdfid,'LATC',lat_hia
ncdf_varget,cdfid,'LONC',lon_hia
ncdf_varget,cdfid,'PALT',palt_hia
ncdf_varget,cdfid,'PSXC',press_hia
ncdf_varget,cdfid,'UTMID',time_hia
ncdf_close,cdfid
```

Rest of code (not shown here) plots aircraft flight track and altitude time series.

AOSC 652: Analysis Methods in AOSC

IDL

Please have a look at **read_start08_netCDF.pro** which you have copied

This program reads data from *.nc file and assigns variable names

```
filename='start08_rf01_AWAS_merged_final_v03.nc'
cdfid=ncdf_open(filename)
ncdf_varget,cdfid,'GGALT',alt_hia
ncdf_varget,cdfid,'LATC',lat_hia
ncdf_varget,cdfid,'LONC',lon_hia
ncdf_varget,cdfid,'PALT',palt_hia
ncdf_varget,cdfid,'PSXC',press_hia
ncdf_varget,cdfid,'UTMID',time_hia
ncdf_close,cdfid
```

Rest of code (not shown here) plots aircraft flight track and altitude time series.

Let's run read_start08_netCDF.pro