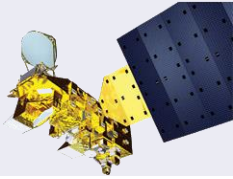
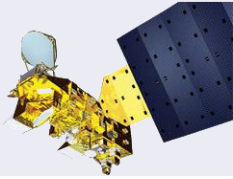


netCDF4 files for AIRS and CrIMSS

Evan M. Manning
Jet Propulsion Laboratory, California Institute of Technology,
4800 Oak Grove Dr., Pasadena, CA 91109

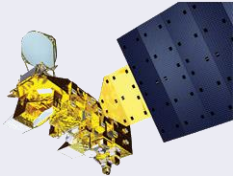
Goal



Our new format for **SounderCDF** is:

- Easy to use
- Interoperable with future products
  - Standard conformant:
    - **CF**
    - **ACDD**
- Flexible – other instruments, products are “the same but different”

Standards



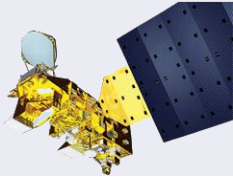
CF:

- **C**limate and **F**orecast Metadata
- COARDS heritage
- Stated goals include:
  - Locate data in space–time and as a function of other independent variables, to facilitate processing and graphics
  - Identify data sufficiently to enable users of data from different sources to decide what is comparable, and to distinguish variables in archives
- Conventions for dimensions, units, and variable and file attributes
- Users will need to consult an interface specification document much less

ACDD:

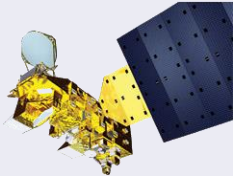
- **A**ttribute **C**onventions for **D**ataset **D**iscovery
- Mostly file-level or collection-level metadata emphasizing letting users find our data set

Basic Format



- netCDF4
  - Built on HDF5 – but users should not use the HDF5 library directly
- Compared to classic netCDF adds:
  - **Strings**
  - **Groups**
  - **Compound types** (structures)
  - etc.
- We minimize use of these new features:
  - **Strings** are used but not arrays of strings
  - **Groups** are used but sparingly – most parameters are in the root group for most granule products
  - **Compound types** are not used.

Granule structure

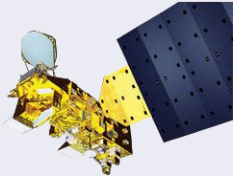


- Like AIRS, SounderCDF Level-1 and Level-2 granule files are six minutes in duration – 45 8-second cycles
- But unlike AIRS these granules are closely aligned with UTC time, the first one starting within 8 seconds of midnight.
- File names have both **start time** to the nearest minute and **granule number**:

SNDR.SNPP.ATMS.20120228T2354.m06.g240.L1B.std.v01_00_00.J.160908005508.nc
SNDR.SNPP.CRIMSS.20130415T1154.m06.g120.L2_RET_CLIMCAPS_NSR.std.v01_23_01.J.180328160143.nc

- The structure is swath-like with primary cross-scan and along-scan dimensions: **atrack**, **xtrack**
  - Plus **fov** for selected variables: clouds, OLR, etc.

Group structure

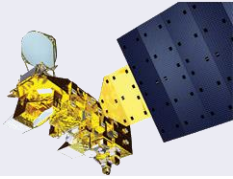


- netCDF4 allows a hierarchical structure – “groups”.
  - This allows useful partitioning of data.
  - But too much use of the group structure imposes a burden on users, so we keep it simple.

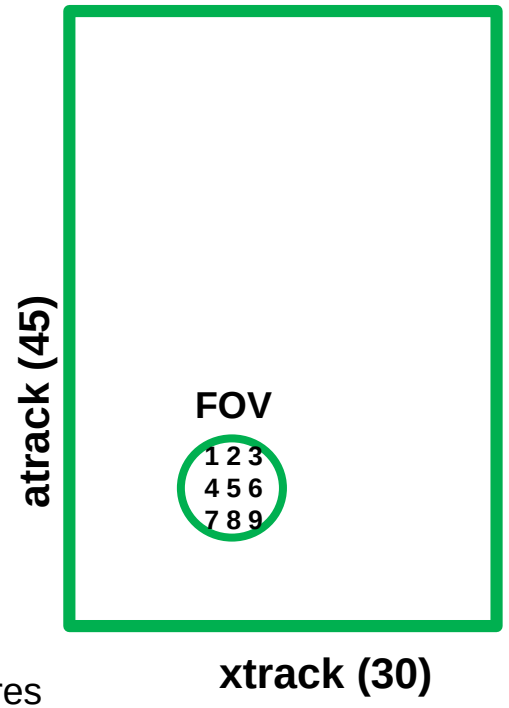
- Granule Level-1 and Level-2 files have all of the most commonly used variables in the **root** group (“**/**”):
    - Data
    - Geolocation
    - QC
  - Casual users can ignore the group feature entirely.
  - It will be easy to order data with any additional groups stripped out.
  - An “**aux**” group holds extra information of interest to the algorithm team but not most users
    - L2 **aux** varies by algorithm
      - L2 **aux** for CHART includes neural net first guess output, QC predictors
      - L2 **aux** for CLIMCAPS includes MERRA2 first guess, Chi-squared
  - L2 also has:
    - A “**mol_lay**” group with column density profiles.
    - A “**mw**” group with outputs from the Microwave-Only retrieval step.
    - An “**ave_kern**” group with averaging kernels. (Not included in initial release.)

- Level-3 also has all of its basic information in the root group: gridded means of all variables.
  - Almost all users will find everything they need in the root group
  - A “**nobs**” group contains information about the **n**umber of **obs**ervations associated with the means in the root group.

Key Level-2 dimensions



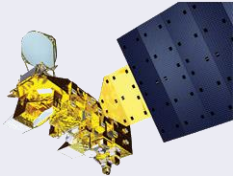
- **atrack**
  - Along-track horizontal dimension: 45
- **xtrack**
  - Cross-track horizontal dimension: 30
- **fov**
  - FOV dimension within FOR: 9
    - Not 3x3 as for AIRS
- **air_pres**
  - 100 pressure levels
  - Coordinate variable **air_pres** has the pressures
- **air_pres_h2o**
  - 66 pressure levels up to 50 hPa
  - Coordinate variable **air_pres_h2o** has the pressures
- **air_pres_lay**
  - 100 pressure layers
  - Coordinate variable **air_pres_lay** has the mid-layer pressures
  - Boundary variable **air_pres_lay_bnds** has the layer boundaries



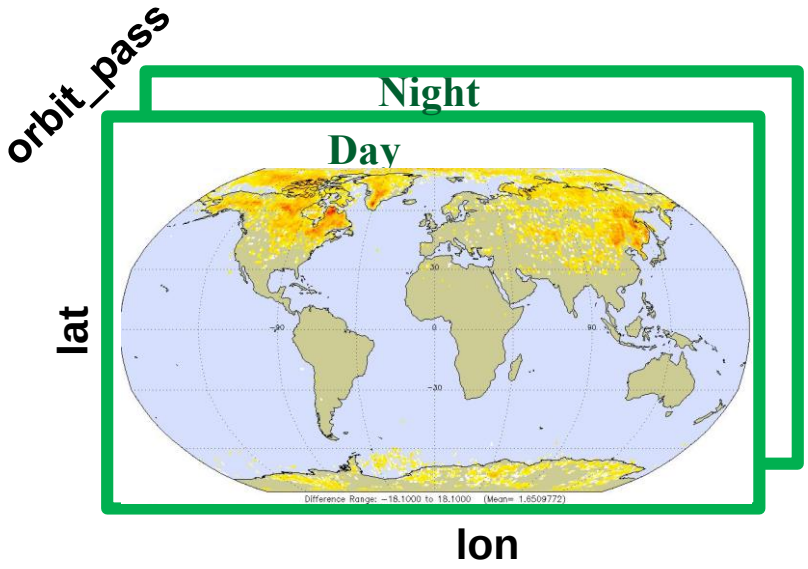
For Cloud-cleared radiances:

- **wnum_lw**, **wnum_mw**, **wnum_sw**
  - CrIS channel dimensions
  - Coordinate variables **wnum_lw** etc. have frequencies in wavenumber

Key Level-3 dimensions

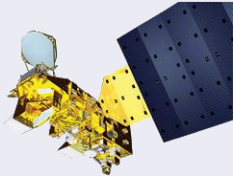


- **orbit_pass**
  - Day/night dimension: 2
  - No longer have 2 grids
- **lon**
  - Logitude dimension: 360
- **lat**
  - Latitude dimension: 180



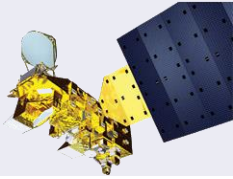
- **air_pres**
  - 100 pressure levels
  - Coordinate variable **air_pres** has the pressures
- **air_pres_h2o**
  - 66 pressure levels up to 50 hPa
  - Coordinate variable **air_pres_h2o** has the pressures

Attributes



- SounderCDF files contain many attributes.
  - Global attributes are like metadata giving info at the file level or higher
    - Many are required or recommended by **CF** or **ACDD**
      - **ACDD** attributes generally describe the data set to a new user: instrument, platform, publisher, …
    - Some are helpful for selecting granules of interest:
      - **geospatial_*** give location as center lat/lon, bounding box, and polygon
  - Variable attributes provide information about each variable
    - **long_name -- for plots**
    - **description -- internal documentation**
    - **standard_name – CF compliance / interoperability**
    - **units – CF uses UDUnits. We use raw SI where possible.**
    - **coverage_content_type -- ISO 19115 {physicalMeasurement, qualityInformation, thematicClassification, referenceInformation, coordinate, modelResult}**
    - **valid_range – min & max**
    - **coordinates – links to coordinate variables, especially lat & lon**
    - **bounds -- optional link to another variable with boundary info for layers, grid cells, FOVs, etc.**
    - **_FillValue – A special value when there is no valid data**
    - **ancillary_variables -- points to error estimate, QC, nobs**
    - **cell_methods – we only use area means & time means for L3**
    - **flag_values, flag_masks, flag_meanings -- for flags**

Variable names



Each variable has several "**names**":

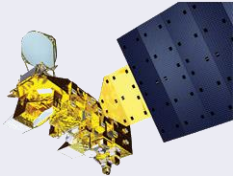
- The actual name of the variable
  - We have complete freedom
  - We go with easy-to-type, short but not cryptic, underscores, no caps:
    - **lat**
    - **air_temp**
    - **h2o_vap_tot**
    - **o3_mmr**

- "**standard_name**"
  - CF maintains a table of all acceptable values. Not all variables have standard names. New names can be added after discussion.
  - Names are precise:
    - **latitude**
    - **air_temperature**
    - **atmosphere_mass_content_of_water_vapor**
    - **mass_fraction_of_ozone_in_air**

- "**long_name**"
  - Freeform, suitable for use as an axis label by automatic plotting tools
  - Long for a name but short for a description:
    - **"latitude"**
    - **"air temperature profile"**
    - **"total water vapor"**
    - **"ozone MMR profile"**

- "**description**"
  - A full description, excluding things like units covered in other attributes
  - Similar to what you might find in a format specification, but right there in the file:
    - **"latitude of FOV center"**
    - **"air temperature profile"**
    - **"total precipitable water vapor"**
    - **"ozone mass mixing ratio to moist air"**

Variable attribute example: o3_mmr



float **o3_mmr**(atrack, xtrack, air_pres);

string o3_mmr : **units** = "1";

string o3_mmr : **ancillary_variables** = "o3_mmr_qc o3_mmr_err";

string o3_mmr : **long_name** = "ozone MMR profile";

string o3_mmr : **standard_name** = "mass_fraction_of_ozone_in_air";

string o3_mmr : **coordinates** = "lon lat";

string o3_mmr : **description** = "ozone mass mixing ratio to moist air";

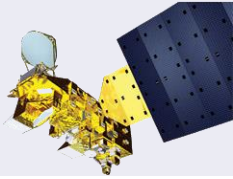
string o3_mmr : **AIRS_name** = "O3mmrLevSup";

float o3_mmr : **_FillValue** = 9.96920996838686905e+36f;

string o3_mmr : **cell_methods** = "area: mean";

string o3_mmr : **coverage_content_type** = "physicalMeasurement";

L2 Ancillary variables

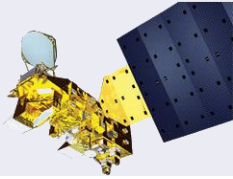


Each L2 science variable typically has two “ancillary variables”:

- *_qc is the equivalent of AIRS *_QC:
  - 0: “Best”
  - 1: “Good”
  - 2: “Do not use”
  - flag_values and flag_meanings attributes are used to communicate this information in a standard way.
- *_err is an error estimate in physical units

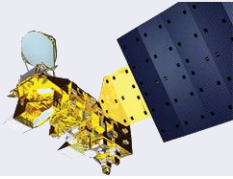
```
byte o3_mmr_qc(atrack, xtrack, air_pres);
      […]
      string o3_mmr_qc : coverage_content_type = "qualityInformation";
      byte   o3_mmr_qc : flag_values = 0,  1,  2;
      string o3_mmr_qc : flag_meanings = "Best Good Do_Not_Use";

float o3_mmr_err(atrack, xtrack, air_pres);
      string o3_mmr_err : coverage_content_type = "qualityInformation";
```

Gran_id, Obs_id

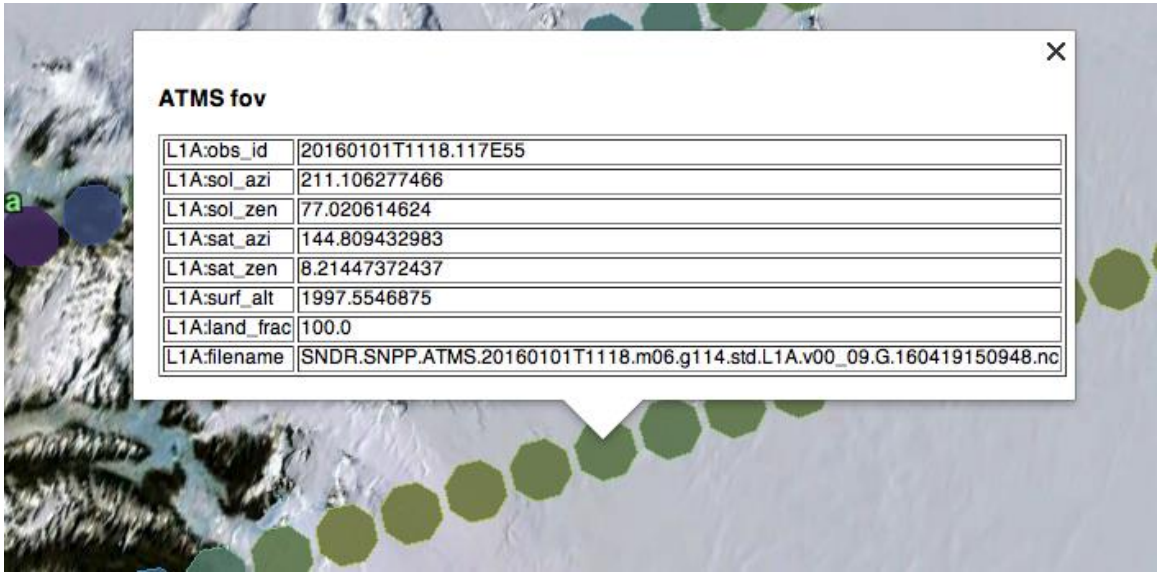


- Each observation (granule, FOR, and FOV) has a unique identifier over the entire mission.
- These will be kept alongside the data in subsets like SNO, match-ups, calibration subset
- They should be used when sample granules, spectra, or profiles are included in papers or presentations
- **gran_id**: yyyymmdd**T**hhmm – **20160914T1124**
- **obs_id**:
  - ATMS: yyyymmdd**T**hhmm.aaa**E**xx -- **20160914T1124.135E96**
  - CrIS FOR, FOR-oriented L2: yyyymmdd**T**hhmm.aa**E**xx -- **20160914T1124.45E30**
  - CrIS FOV, FOV-oriented L2: yyyymmdd**T**hhmm.aa**E**xx.f -- **20160914T1124.45E30.9**

FOV boundaries



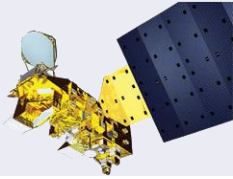
- One key feature of CF is boundaries.
- For each ATMS and CrIS FOV we are able to provide a 8-point bounding polygon.
- This will greatly simplify the task of users who want to plot data on a map.



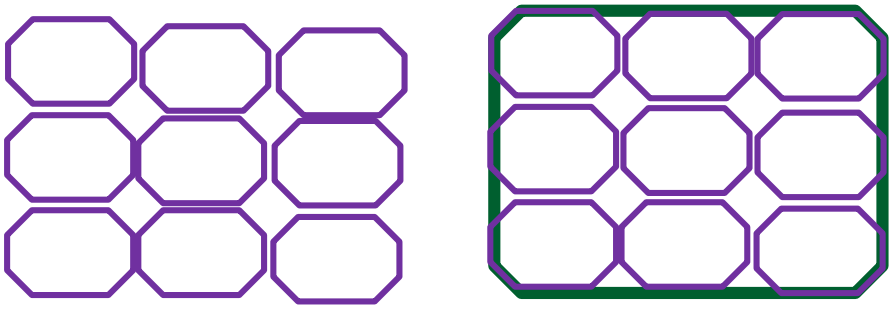
This image shows Elevation data displayed in Google Earth using ATMS FOV polygons

Tool: Jia Zong

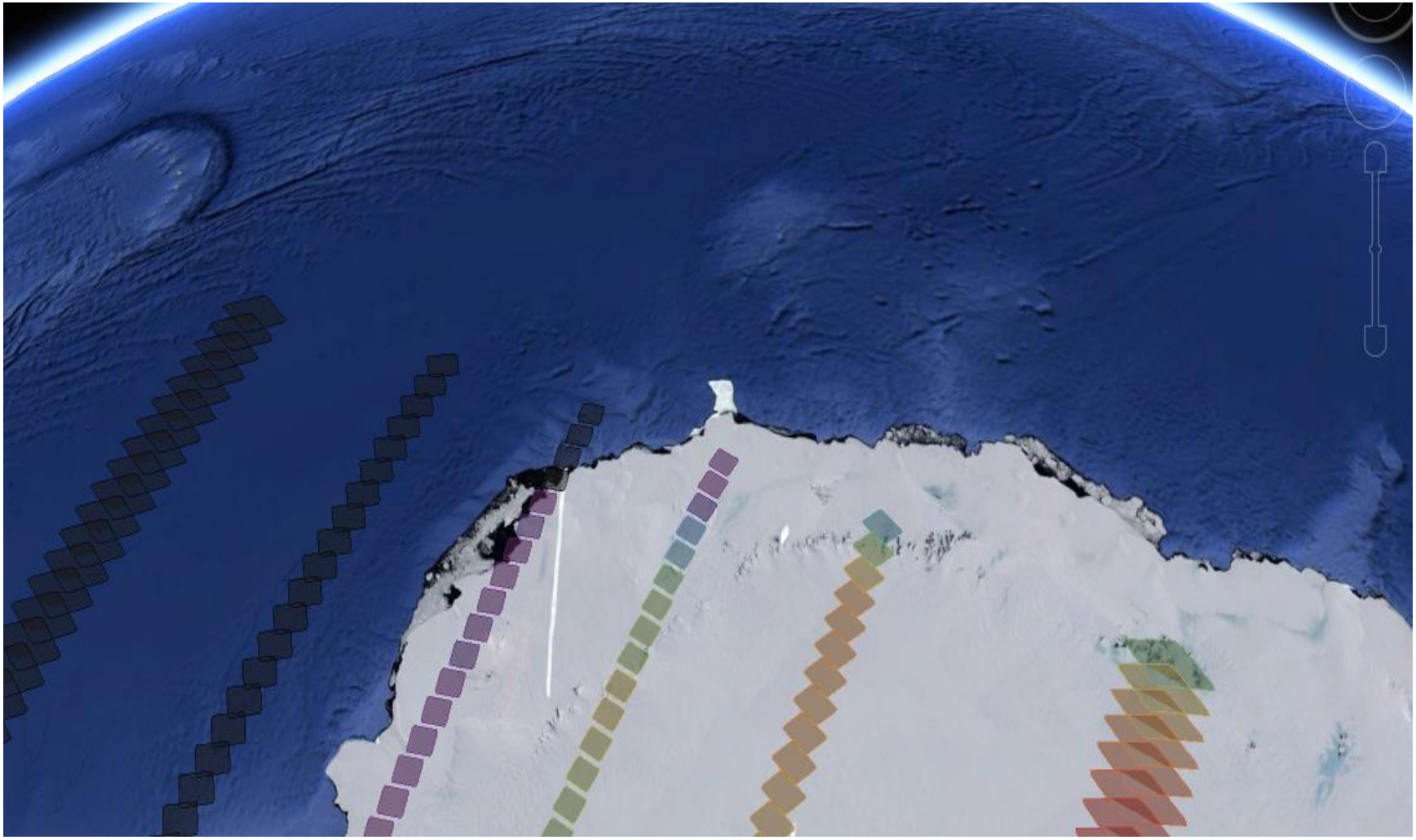
L2 FOR boundaries



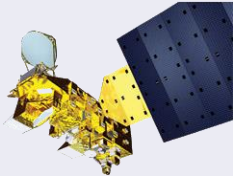
For each L2 FOR we have a 8-side polygon with two points from each of the corner CrIS FOVs.



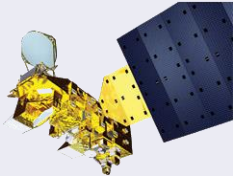
L2 FOR boundaries



SounderCDF in the wild

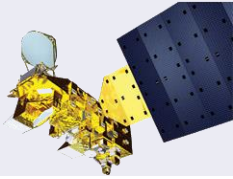


- We have sample data sets of Level-1B for CrIS and ATMS
  - At JPL and public at GES DISC DAAC.
- Level-2 and Level-3 daily CrIMSS data for the 8 comparison months * 2 algorithms is ready at JPL and coming soon to the GES DISC DAAC.
- Level-3 monthly is coming soon.
  - Format is identical to Level-3 daily.
- AIRS will use this format for v7.
  - Hopefully other instruments will also adopt aspects of this format.

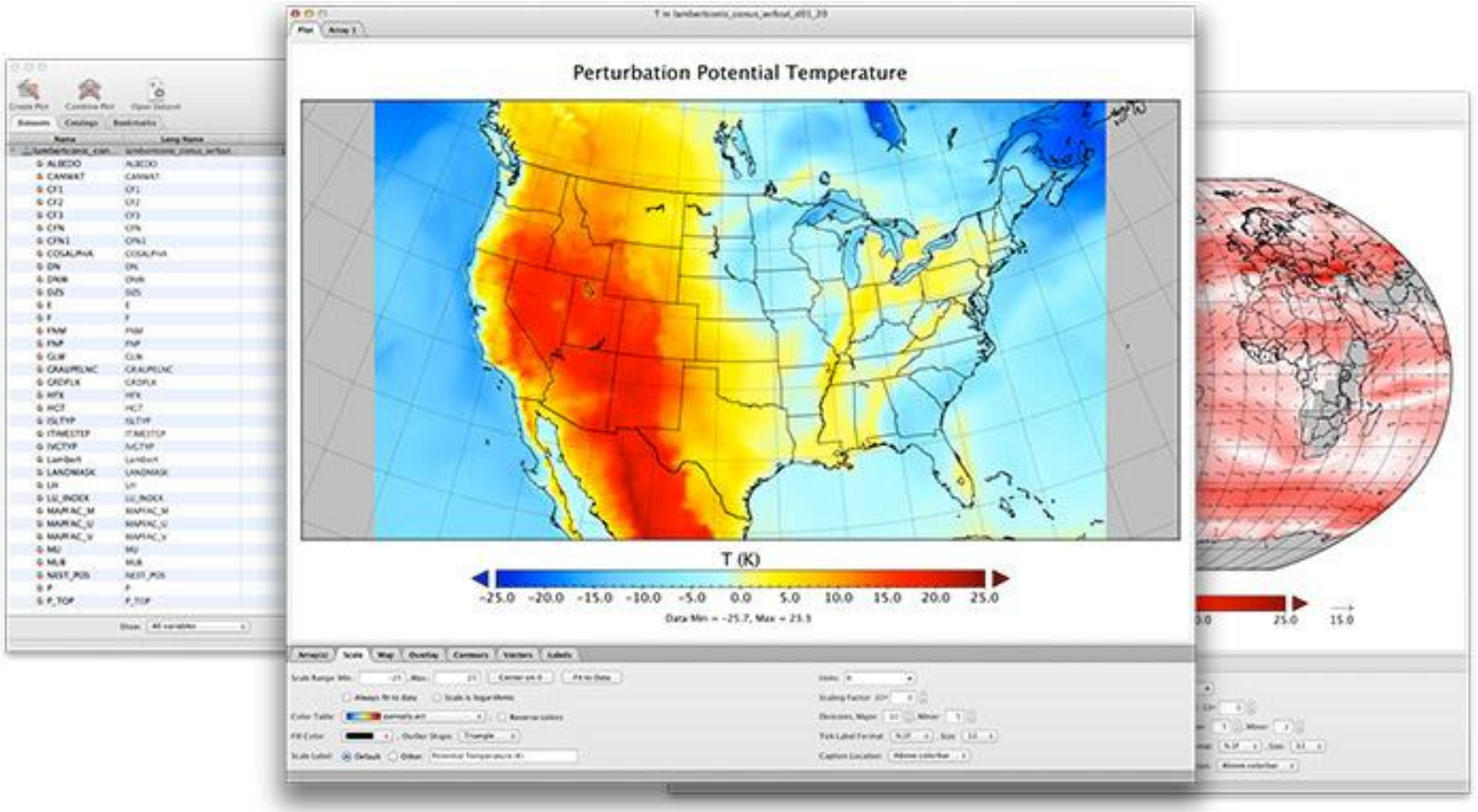


Back Off

Sample tool -- Panoply



- Panoply is a cross-platform geo-aware tool from NASA GISS
- It works well for granule files and even better for Level-3



Sample tool -- Panoply

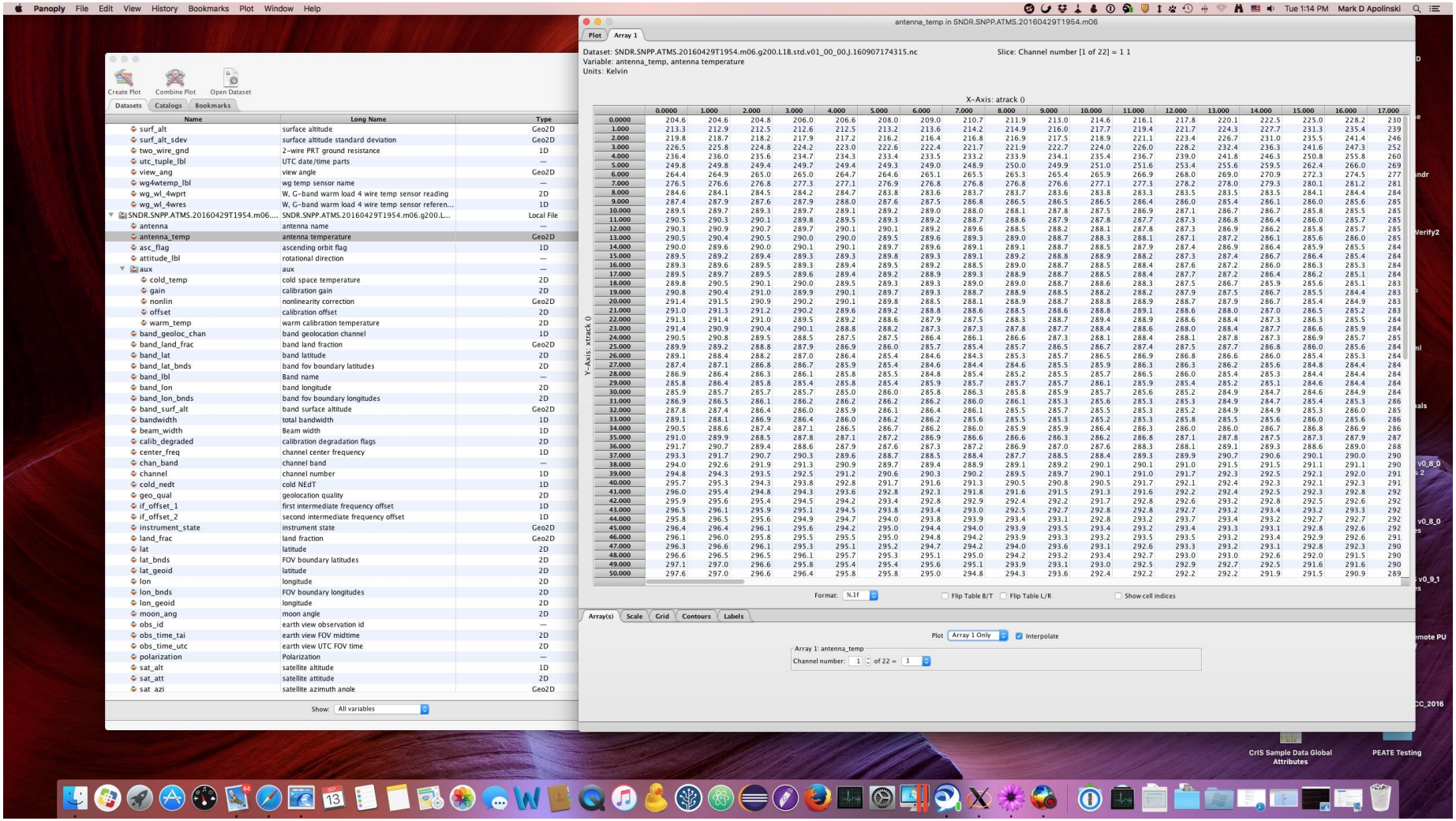


Image: Mark Apolinski

Sample tool -- Panoply

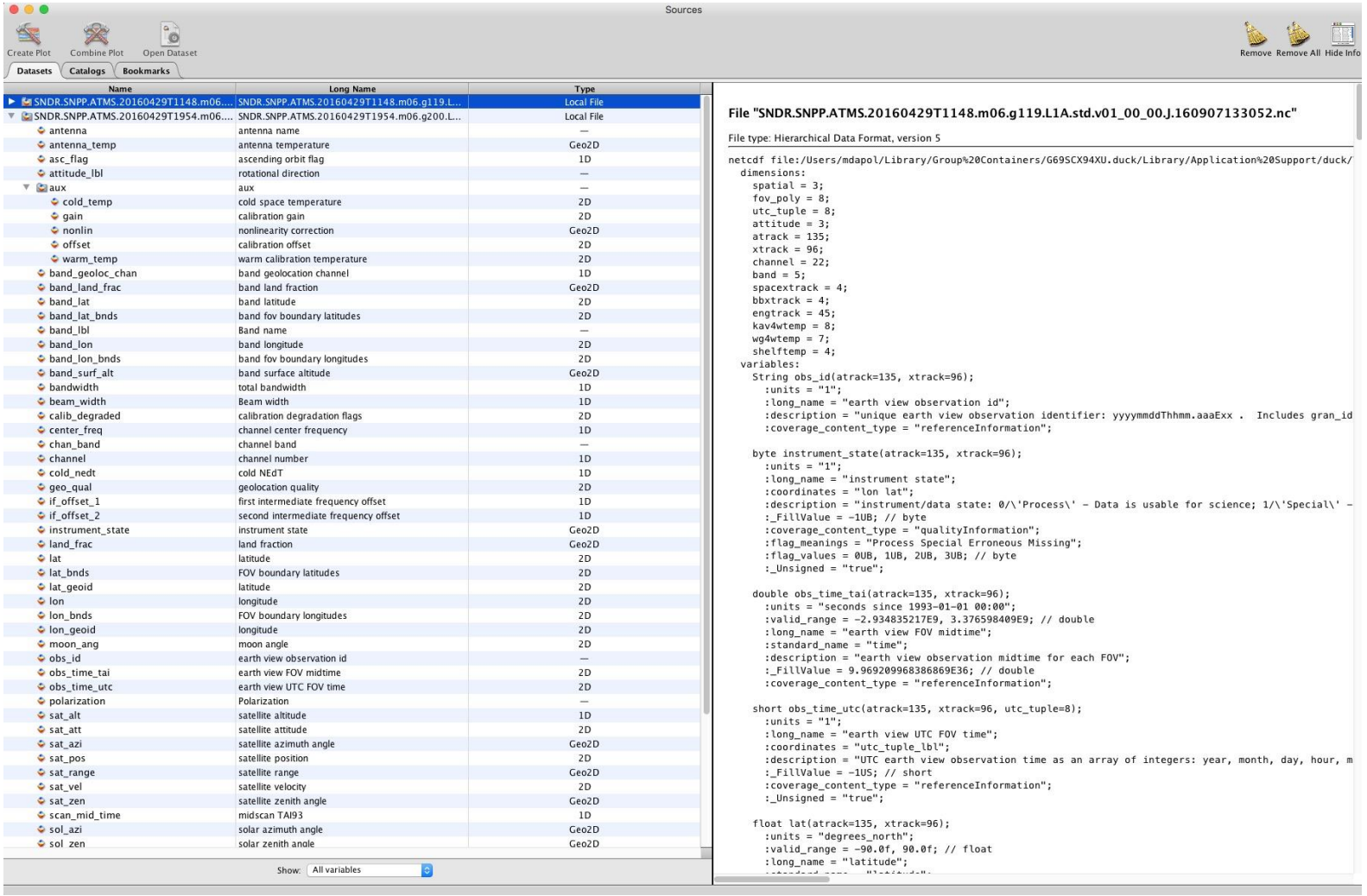


Image: Mark Apolinski

Sample tool -- Panoply

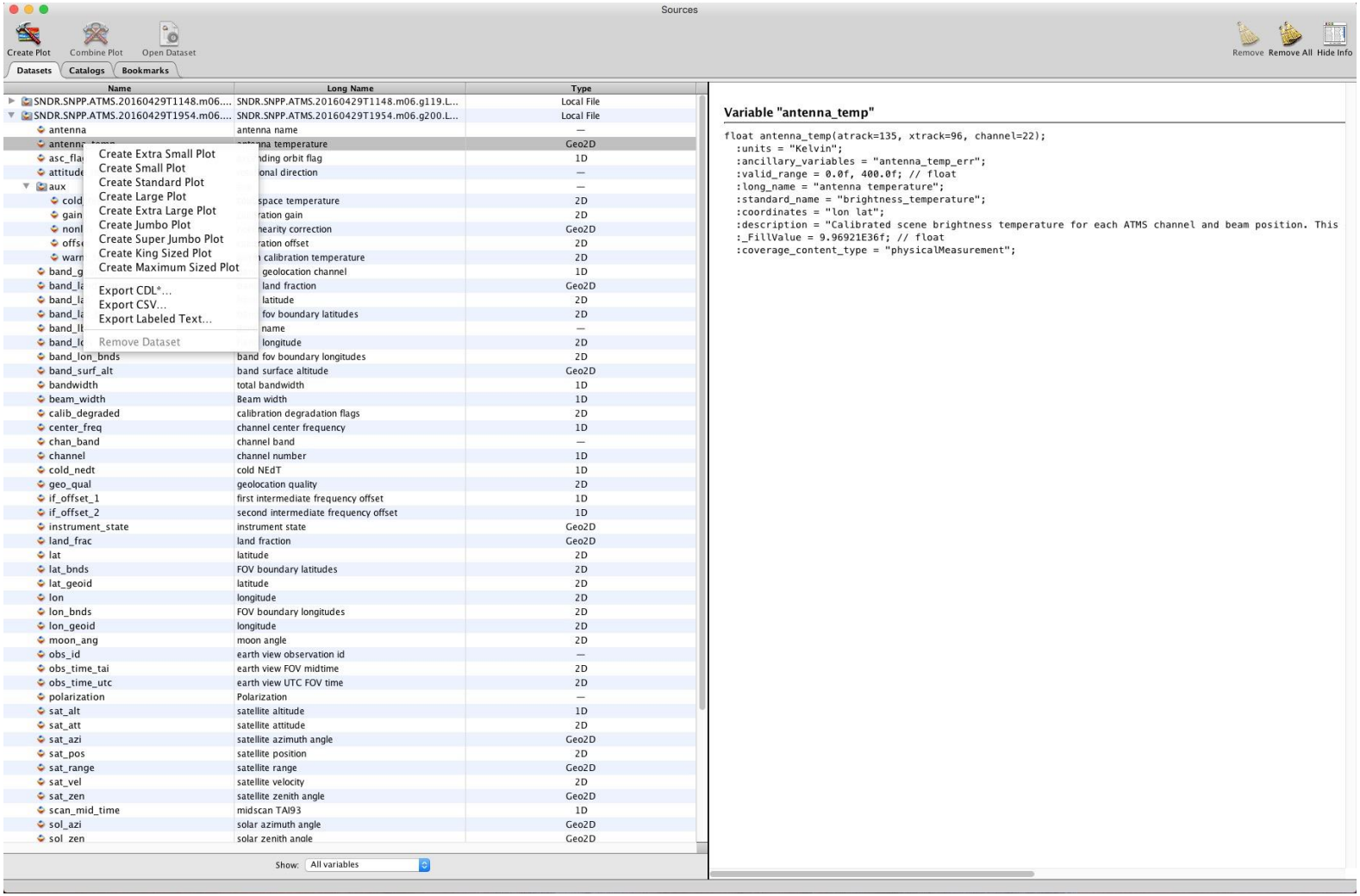


Image: Mark Apolinski

Sample tool -- Panoply

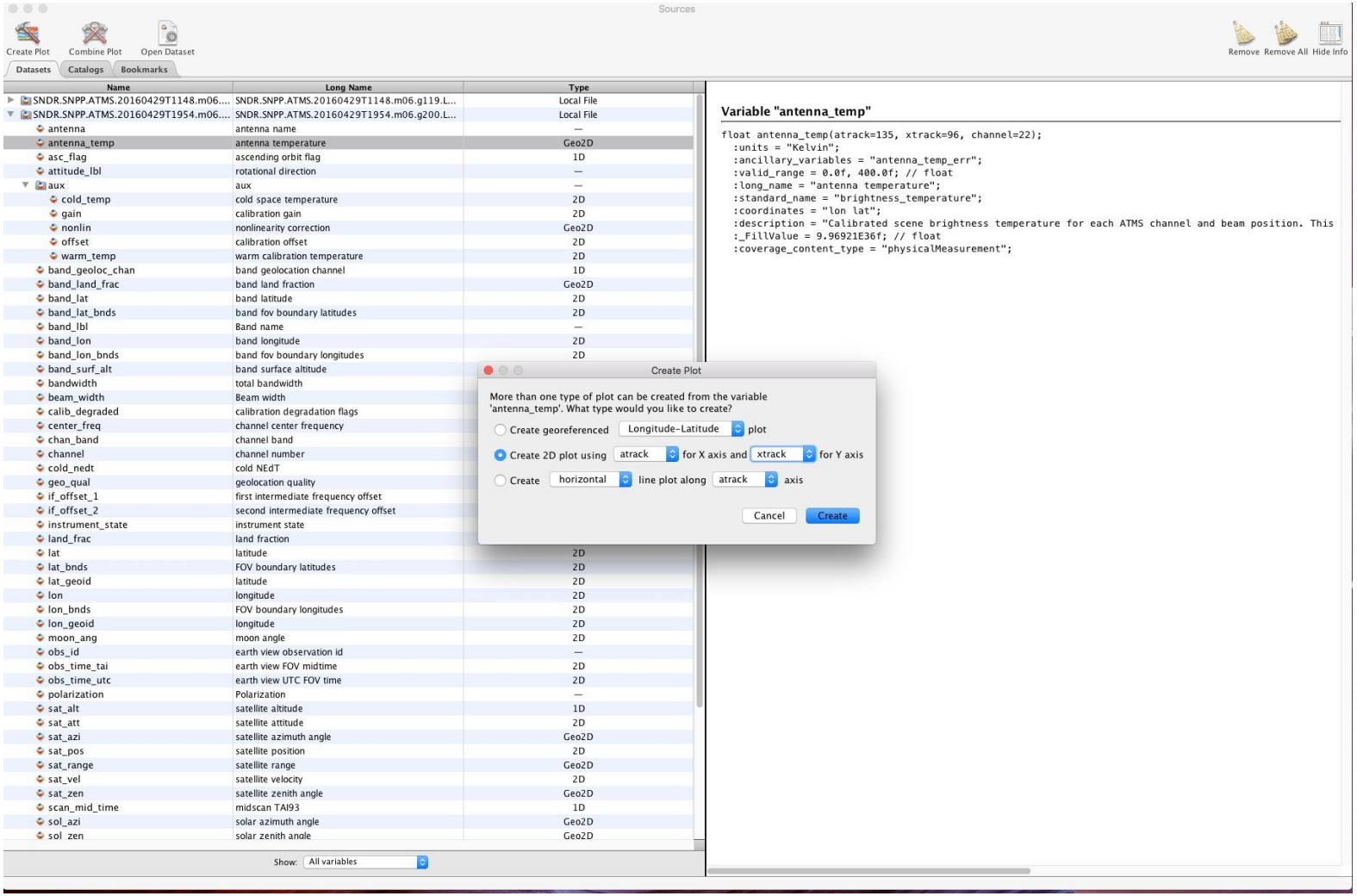


Image: Mark Apolinski

Sample tool -- Panoply

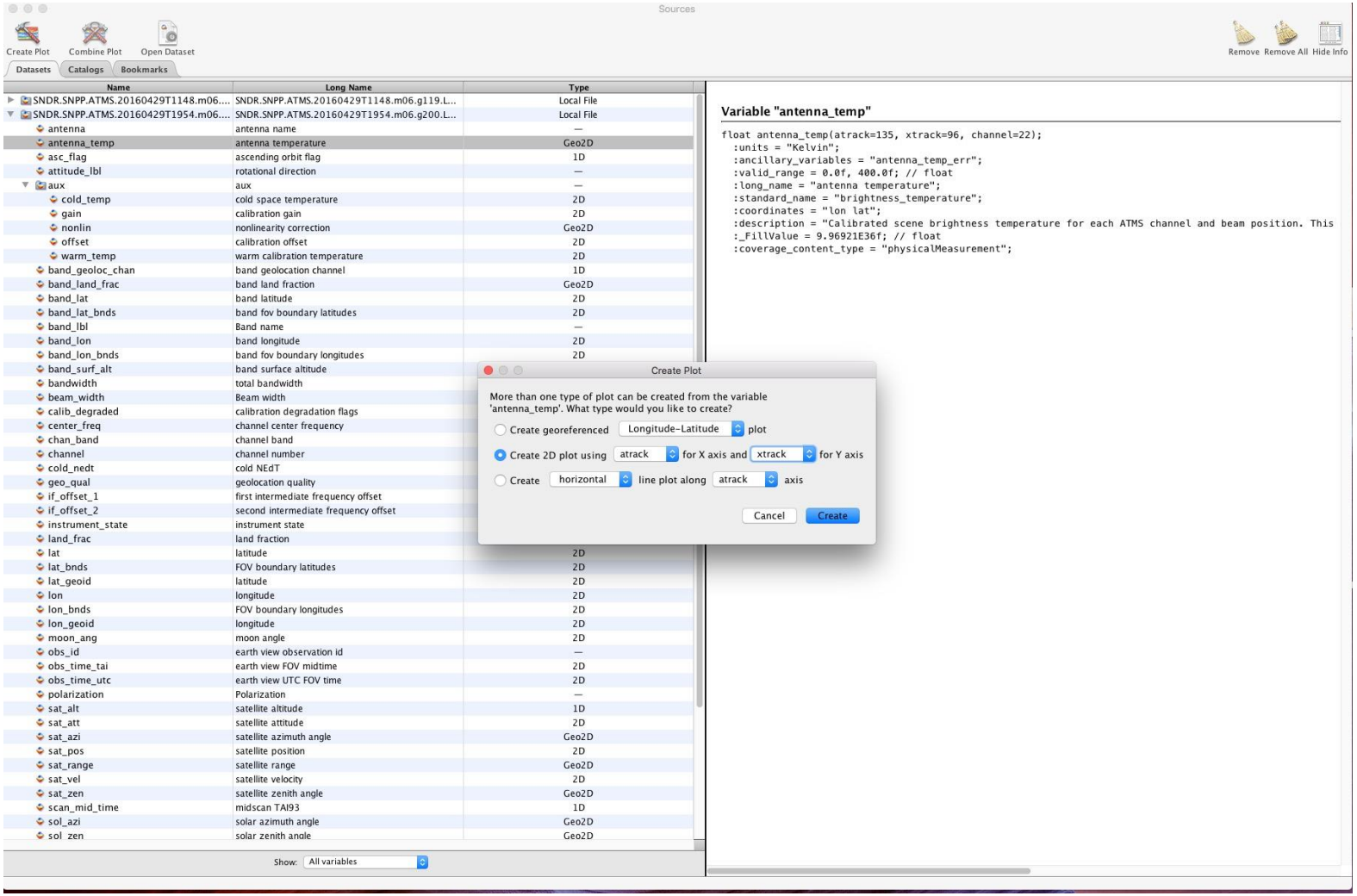


Image: Mark Apolinski

Sample tool -- Panoply

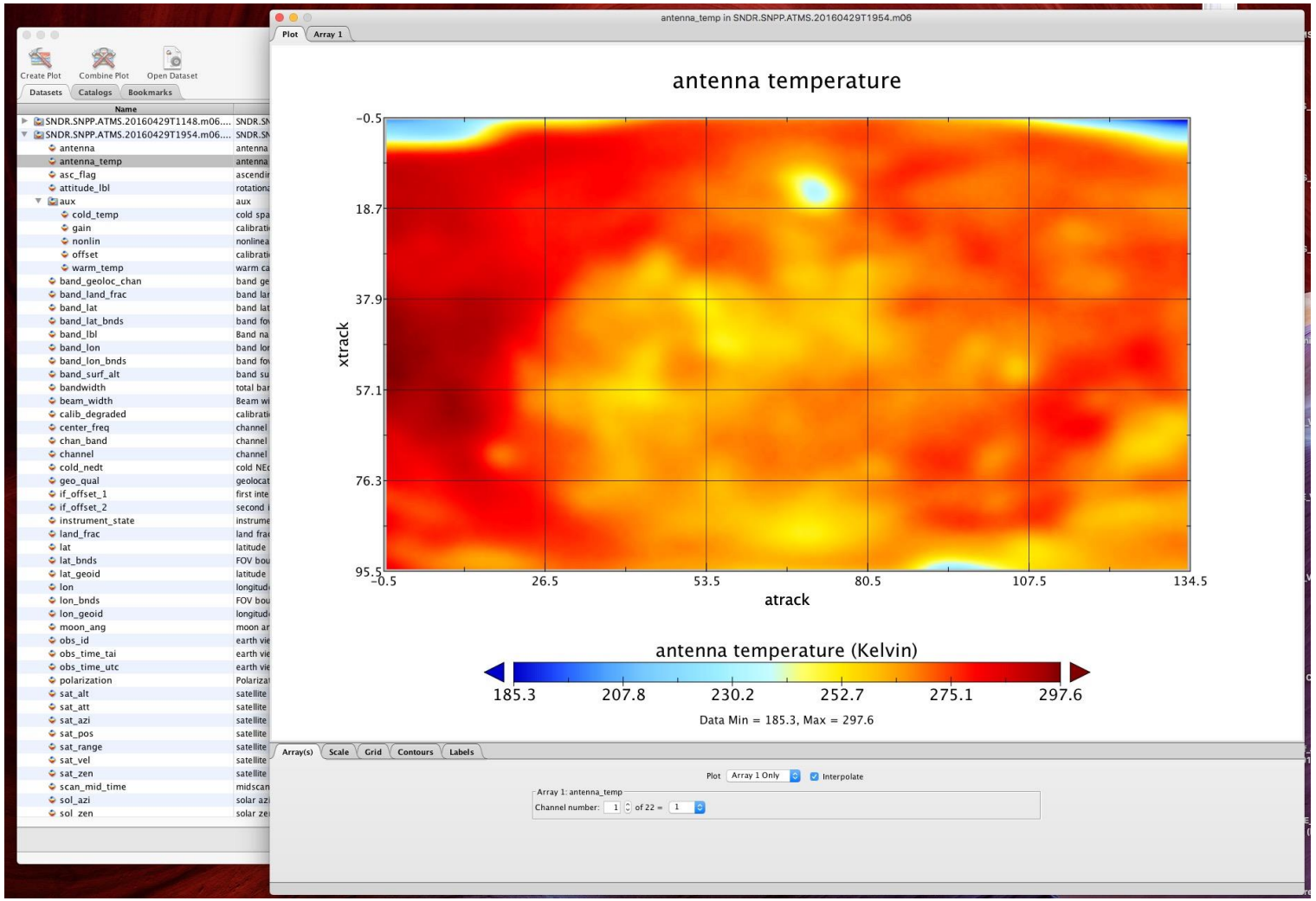
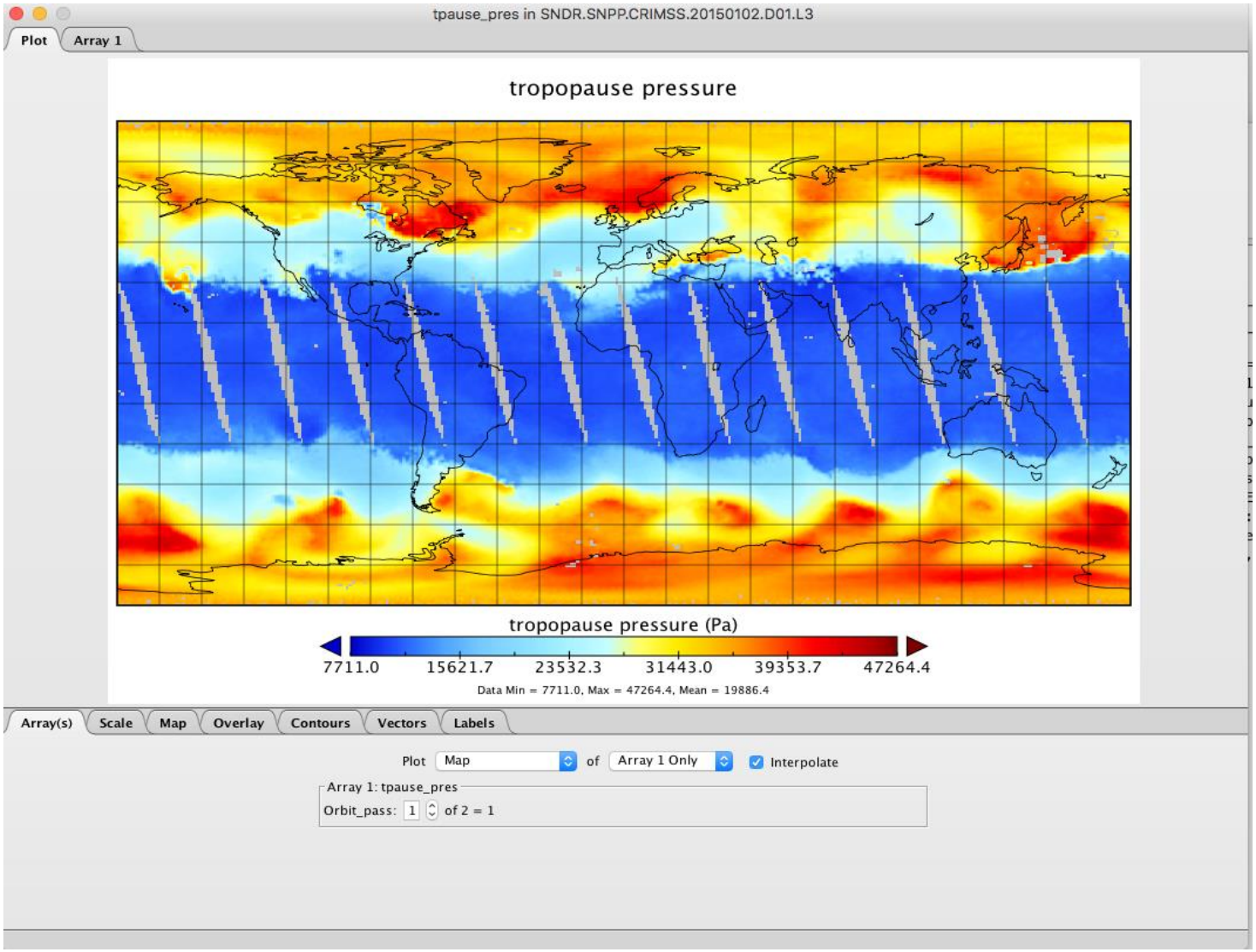
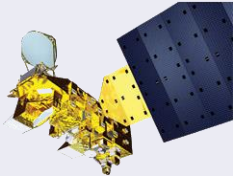


Image: Mark Apolinski

Level-3 sample



Group Structure for subset files



- Subset file types include SNO, Match-Ups, calibration subset
- Subset files add a layer of group structure at the top, so there are groups for L1B ATMS vs L1B CrIS vs L2 CrIMSS and new groups with info on the selection process and per-granule info
- Example: SNO of Aqua AMSU-A with ATMS:
  - **/l1b_amsua** – obs from L1B AMSU-A
  - **/l1b_atms** – obs from L1B ATMS
  - **/select** – info about the match: distance, dtime, …
  - **/l1b_amsua_ingran** – AMSU-A attributes as 240-element arrays: ingran_gran_id, ingran_file_name, ingran_day_night_flag, …
  - **/l1b_atms_ingran**