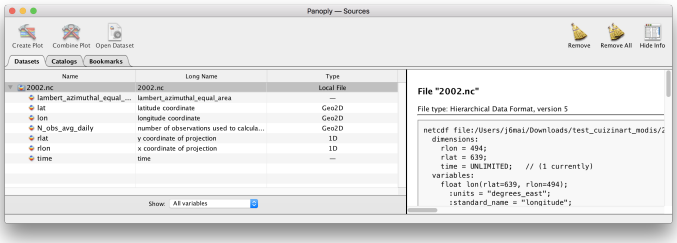
Department Civil & Environmental Engineering
Juliane Mai
Phone: +1 (519) 888-4567 Ext. 30016
E-Mail: juliane.mai@uwaterloo.ca

1 Checklist for Data Managers of New NetCDF Datasets for GWF Cuizinart

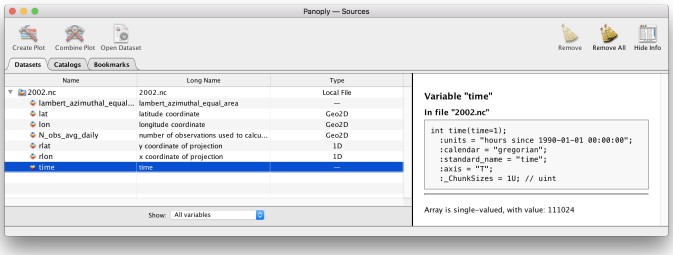
This document describes how Data Managers should check new NetCDF datasets that should be added to the Cuizinart under `http://cuizinart.io`.

1.1 Initial Check

- install Panoply (`https://www.giss.nasa.gov/tools/panoply/`)
- open one of the candidate files
- make sure the dataset has exactly three **dimensions** declared with the exact names `"time"`, `"rlat"`, and `"rlon"`

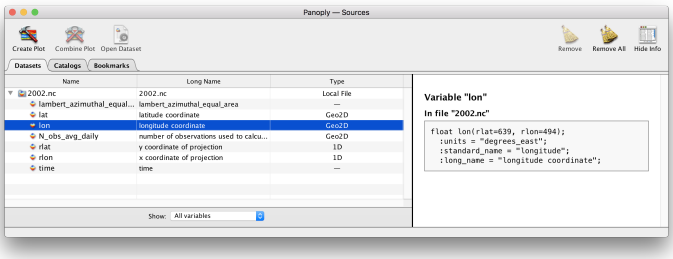


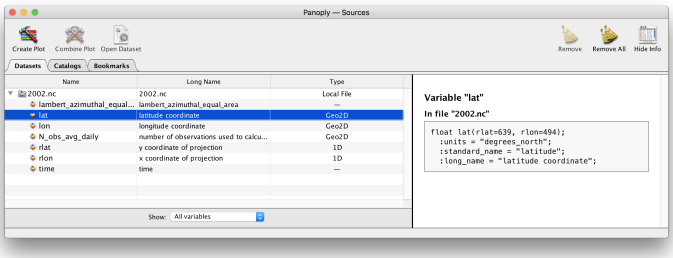
- make sure there is a **variable called** `"time"` (not `"Time"`, not `"t"`, etc) and that its dimension is also `"time"`; the unit of time should look like similar to `"hours since YYYY-MM-DD HH:MM:SS"` (can also be `"minute since"`, `"days since"` etc.)



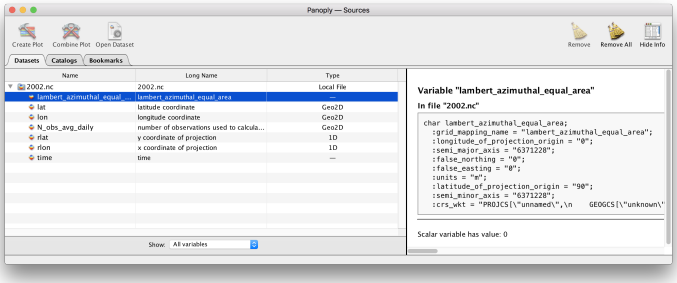
Department Civil & Environmental Engineering
Juliane Mai
Phone: +1 (519) 888-4567 Ext. 30016
E-Mail: juliane.mai@uwaterloo.ca

- make sure there is a **variable called** `"lat"` and a **variable called** `"lon"` (not `"Lat"`/`"Lon"`, not `"lattidude"`/`"longitude"`, etc); both need to be a 2-dimensional variables (not 1-dimensional); the dimensions of both variables need to be `"rlat"` and `"rlon"`; make sure units are in `"degrees"` and that `"long_names"` are set



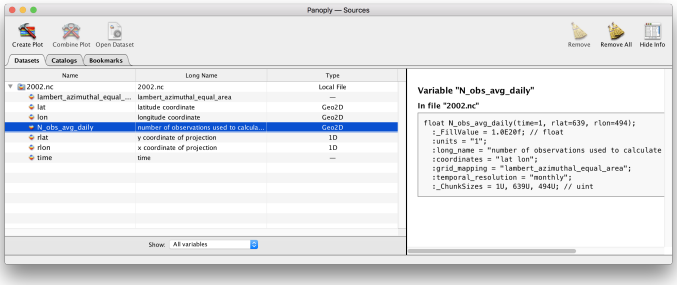


- it is very fancy to have also a **projection attribute** set: the projection information contains all information how one-dimensional latitude and longitude variables `"rlat"` and `"rlon"` (both variables; not dimensions) are transformed such that they result in the 2-dimensional `"lat"` and `"lon"` variables; setting this all correct is pretty sophisticated; I am very happy to help with that; forward data providers to juliane.mai@uwaterloo.ca

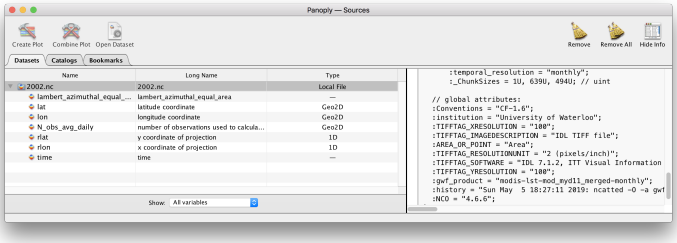


Department Civil & Environmental Engineering
Juliane Mai
Phone: +1 (519) 888-4567 Ext. 30016
E-Mail: juliane.mai@uwaterloo.ca

- **variables of the dataset** need to be exclusively three-dimensional (time, rlat, rlon); make sure they all have a proper description as `"long_name"` (a description a non-modeler understands) and confirm the units with the data provider (this is the most work!)



- make sure that there is one **global attribute** set that is called `"gwf_product"` and contains the exact name of the product



Department Civil & Environmental Engineering
Juliane Mai
Phone: +1 (519) 888-4567 Ext. 30016
E-Mail: juliane.mai@uwaterloo.ca

1.2 CF Compliance Check

- files need to fulfill at least Climate and Forecasts (CF) Metadata Convention CF-1.6
- upload file to webpage `https://pumatest.nerc.ac.uk/cgi-bin/cf-checker.pl` and see that there are NO warnings and NO error messages
- go back and forth with data provider until everything is resolved
- usually files are very large → ask for smaller cut-out, i.e. either fewer time steps or smaller domain

1.3 File Naming Convention

- files are named either `YYYY.nc` (containing data of year `YYYY`) or `YYYYMM.nc` (containing data of month `MM` of year `YYYY`)
- if the product has ensemble members the file naming convention is `YYYY_EEE.nc` or `YYYYMM_EEE.nc` or `YYYY_<ensemble-member-name>.nc`. Ensemble member names are only characters [A-Z], [a-z], and [0-9]. No dashes, no underscore, no blanks etc.
- If files containing full year of month are above 100 GB in size contact me under juliane.mai@uwaterloo.ca and we discuss if we add more possibilities of filenames
- all variables (e.g., temperature, precipitation, etc.) need to in one file
- if the domain of a product is changing over time this will be two products
- product names need to unique → make sure with data provider that if they are revising the product, they will have to give it a new name → make sure that name is not yet taken
- user need to provide fully filled metadata template

History:
May 14, 2019 - J Mai: created