Tools for analyzing climate model output I: accessing netCDF with MATLAB

2013 ASP Summer Colloquium

K. Lindsay (NCAR)

Tutorial Outline

- Why MATLAB
- netCDF (a common file format)
- Reading netCDF files w/ MATLAB
- Examples

MATLAB is an Integrated System for

- Computation, analysis
  - Scalars, vectors, matrices
  - Many functions are built in (Std. Dev., Corr, EOF, etc.)
- Visualization
  - 1 independent variable (line, scatter, histogram, etc.)
  - 2 independent variables (contour, mesh, surface, etc.)
- Programming
  - Modular programming, scripts & functions, via m-files
  - Variety of control constructs (if, else, for, while, switch)

Full Disclosure

- matlab is discussed because of previous participant exposure to it

- It has its own strengths (array handling, many builtins, quick graphics, etc.) & weaknesses (minimal netCDF support)

- Other packages commonly used in the field
  - NCL, ferret, GrADS
  - IDL

Selected MATLAB Materials

- /project/colloquium/hands-on/lindsay
- matlab-cheatsheet.pdf (1 page)
- Krasny_matlab.pdf (3 pages)
- driscoll_matlab_notes.pdf (66 pages)
- getstart.pdf (214 pages)-
  http://www.mathworks.com/access/
  helpdesk/help/pdf_doc/matlab/getstart.pdf

Modular programming:
m-files for scripts and functions

- A collection of matlab commands in a file whose filename has an .m extensions constitutes an m-file.

- Useful for sequence of commands that will be run repeatedly.

- m-files can accept arguments, making them functions, type 'help function' for more info

netCDF (network Common Data Form)

- Set of software libraries and self-describing, machine-independent data formats that support the creation, access, and sharing of array-oriented scientific data.

- /project/colloquium/hands-on/lindsay/netcdf_factsheet.pdf

- http://www.unidata.ucar.edu/software/netcdf/index.html

netCDF features

- *Self-Describing*. A netCDF file includes information about the data it contains.

- *Portable*. A netCDF file can be accessed by computers with different ways of storing integers, characters, and floating-point numbers.

- *Scalable*. A small subset of a large dataset may be accessed efficiently.

- *Appendable*. Data may be appended to a properly structured netCDF file without copying the dataset or redefining its structure.

- *Sharable*. One writer and multiple readers may simultaneously access the same netCDF file.

- *Archivable*. Access to all earlier forms of netCDF data will be supported by current and future versions of the software.

Basic Properties of a Dataset

- A dataset consists of dimensions and variables defined on those dimensions.

- Variables have metadata associated with them, which consist of named attributes that describe the variable, e.g. long_name, units, scale_factor, coordinates, …

- Various metadata conventions exist that state what different attributes mean, e.g. COARDS (1995), Climate and Forecast (CF) (2003)

ncdump −h file.nc

- netcdf speciescont.cam2.h0.TS.1698-01_cat_1897-12 {
- dimensions:
- time = UNLIMITED ; // (2400 currently)
- lat = 48 ;
- lon = 96 ;
- variables:
- double time(time) ;
- time:bounds = "time_bnds" ;
- time:calendar = "noleap" ;
- time:units = "days since 0001-01-01 00:00:00" ;
- time:long_name = "time" ;

cont.

- double lat(lat) ;
-         lat:units = "degrees_north" ;
-         lat:long_name = "latitude" ;
-     double lon(lon) ;
-         lon:units = "degrees_east" ;
-         lon:long_name = "longitude" ;
- float TS(time, lat, lon) ;
-         TS:cell_method = "time: mean" ;
-         TS:long_name = "Surface temperature (radiative)" ;
-         TS:units = "K" ;
-         TS:_FillValue = 1.e+20f ;
-         TS:missing_value = 1.e+20f ;

NCO suite, http://nco.sf.net/

- ncks in.nc out.nc   # extract/append vars
- ncrcat in1.nc … out.nc   # concatenate files
- ncra in1.nc … out.nc   # averaging
- ncwa in.nc out.nc   # weighted averaging
- Many available command line options
  - Operate on subset of variables, -v
  - Operate on subset of dimensions, -d
- Online manual, -h gives brief help

Accessing netCDF with MATLAB

- help netcdf
- help netcdf.getVar

- interface to netCDF is rudimentary
- to read a variable
  - open file
  - look up variable
  - read variable
  - close file
- This provides no metadata!

Helper m-files

- Single precision values (float) are always promoted to double precision

- nc_read_var(fname, varname)
  - read an entire variable

- nc_read_var_slice(fname, varname, start, count)
  - read a variable slice
  - start is starting indices
    - 0-based, order opposite from ncdump output
    - -1 is replaced with the end of the dimension
  - count is length of index span
    - -1 is replaced to read to the end of the dimension

Helper m-files

- nc_read_var_att(fname, varname, attname)
  - read a variable's attribute

- Read variable's associated dimensions:
  - nc_read_var_lon(fname, varname)
  - nc_read_var_lat(fname, varname)
  - nc_read_var_time(fname, varname)

Handling time

- Example time metadata
  - units = "days since 1850-01-01 00:00:00"
  - calendar = "noleap"

- Read attributes from a variable's time coordinate
  - nc_read_var_time_calendar(filename, varname)
  - nc_read_var_time_units(filename, varname)

- time_to_year_offset(in_time, units, calendar)
  - convert time value to years of offset
  - handles calendar types 'noleap', '365_day', 'julian'

Miscellaneous

- comp_masked_weighted_mean
  - Area weighted regional averaging
- lonlat_to_logical
  - convert lon-lat coordinates to logical indices
- overlay_coastlines
  - overlay contintental coastlines on a figure