Handling large netCDF files using climate data operators (cdo) and ncview

Dian Putrasahan

Lunch-bytes seminar

Oct. 21, 2015

Network Common Data Form

Tools that support netCDF files:

- Matlab
- Python
- NCL
- GrADS
- Ferret
- ncview
- Panoply
- ncdump
- NCO
- CDO (climate data operators)

and many more .....

Climate Data Operators (CDO)

CDO is a collection of command-line operators designed to manipulate and analyze climate data.

Why use CDO?

- If you often work on UNIX-type environment, command-line operations are extremely useful.
- CDO has very small memory requirements and can process files larger than the physical memory
- CDO is open source

https://code.zmaw.de/projects/cdo/wiki/cdo

https://code.zmaw.de/projects/cdo/embedded/cdo.pdf

Getting cdo

- Download from https://code.zmaw.de/projects/cdo
  - gunzip cdo-$VERSION.tar.gz
  - tar xvf cdo-$VERSION.tar
  - cd cdo-$VERSION
  - ./configure --with-netcdf=<netcdf_root_dir>
  - make
  - make install
- module load cdo/1.6.5

Syntax:

cdo operator1 [-operatorN] <infile> <outfile>

File information

cdo <operator> <infile>

```
cdo sinfon HRC06.cam2.h0.0109-09.nc
   File format : netCDF
    -1 : Institut     Source   Ttype    Levels Num    Points   Num Dtype : Parameter name
     1 : unknown  CAM    constant    1       1         383     1        F64  : w_stag
     2 : unknown  CAM    constant    1       1         384     2        I32  : nlon
         ......
    13 : unknown  CAM    instant     26      2       221184    4        F32  : CLDLIQ
    14 : unknown  CAM    instant      1      1       221184    4        F32  : CLDLOW
   Grid coordinates :
     4 : lonlat                   : points=221184 (576x384)
                          lon : 0 to 359.375 by 0.625 degrees_east  circular
                          lat : -90 to 90 by 0.469974 degrees_north
   Vertical coordinates :
     1 : surface                  : levels=1
     2 : hybrid                   : levels=26
                          lev : 3.54464 to 992.556 level
   Time coordinate :  1 step
     RefTime =  0001-01-01 00:00:00  Units = days  Calendar = 365_day  Bounds = true
  YYYY-MM-DD hh:mm:ss  0109-09-01 00:00:00
```

cdo griddes HRC06.pop.0109-09.nc

```
#
# gridID 25
#
gridtype  = lonlat
gridsize  = 6480000
xname     = lon
xlongname = longitude
xunits    = degrees_east
yname     = lat
ylongname = latitude
yunits    = degrees_north
xsize     = 3600
ysize     = 1800
xfirst    = 0
xinc      = 0.1
yfirst    = -89.95
yinc      = 0.0999985
cdo griddes: Processed 24 variables ( 0.01s )
```

cdo <operator> <infile>

info
npar
nlevel
ndate
showname
showlevel
showtime
pardes
zaxisdes
.....

cdo showdate TS.full.all.nc

```
0102-01-01  0102-02-01  0102-03-01  0102-04-01  0102-05-01  0102-06-01  0102-07-01
.....
0155-07-01  0155-08-01  0155-09-01  0155-10-01  0155-11-01  0155-12-01
cdo showdate: Processed 1 variable over 648 timesteps ( 0.16s )
```

File operations

cdo <operator> <infiles> <outfile>

cdo mergetime $MLDfilespath/MLD.HRC06.pop.*.nc $mergedfilespath/MLD.HRC06.full.all.nc

MLDfilespath = /bkirtman2/dputrasahan/HRC06/MLD
ls $MLDfilespath
    MLD.HRC06.pop.0102-01.nc
    MLD.HRC06.pop.0102-02.nc
    MLD.HRC06.pop.0102-03.nc
    .....
    MLD.HRC06.pop.0155-10.nc
    MLD.HRC06.pop.0155-11.nc
    MLD.HRC06.pop.0155-12.nc

mergedfilespath = /bkirtman2/dputrasahan/HRC06/merge
ls -lh $mergedfilespath
    -rw-r--r-- 1 dputrasahan kirtman 16G Oct 14 15:53 MLD.HRC06.full.all.nc

copy
cat
replace
merge
.....

Selection

cdo <operator>,specs <infile> <outfile>

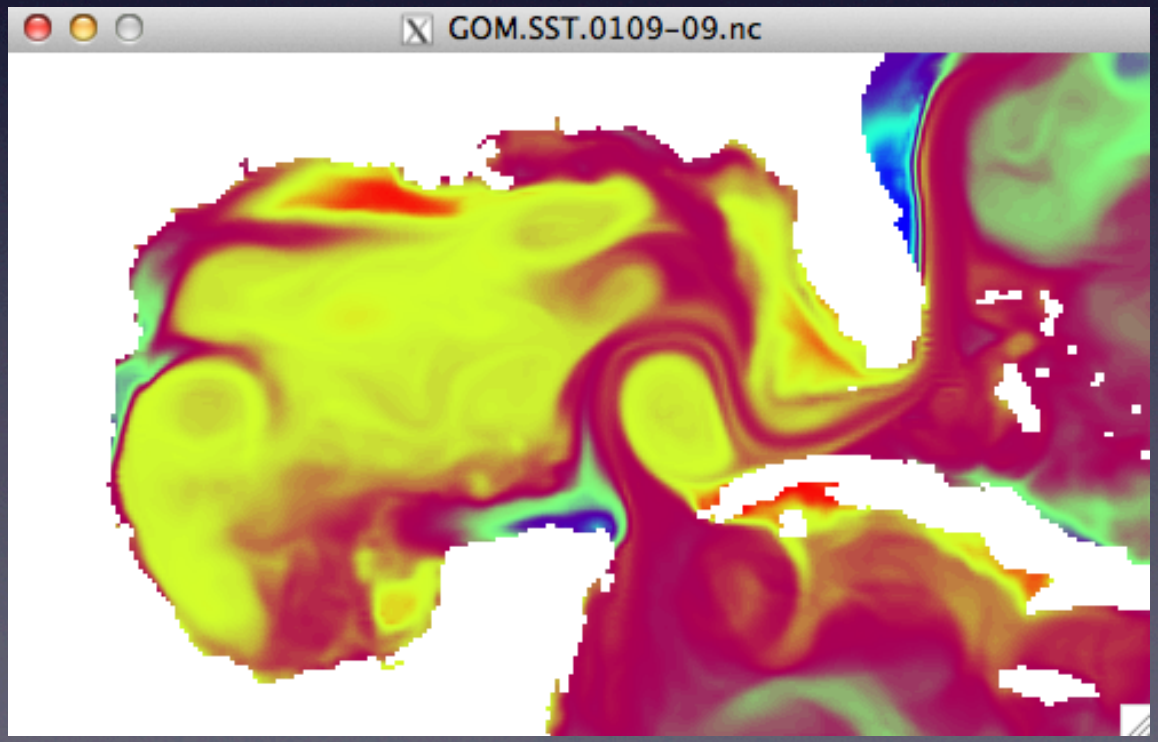
select level

select region

select variable

```
cdo sellevidx,1 -sellonlatbox,-100,-75,17,32 -selname,TEMP $origpath/
HRC06.pop.0109-09.nc $destpath/GOM.SST.0109-09.nc
```

ncview GOM.SST.0109-09.nc



selmon
selyear
seltime
selseas
selparam
sellevel
.....

Arithmetic and Statistics

cdo <operator> <infile1> <infile2> <outfile>
cdo <operator-stat> <infile> <outfile>

Vertical integral of ocean temperature:

vertical sum

multiply two fields

multiply by a constant (convert cm to m)

```
cdo vertsum -mul $origpath/TEMP.HRC06.0109-09.nc -mulc,0.01
dZcm.nc $destpath/OHC.HRC06.0109-09.nc
```

addc (add a constant)
subc (subtract a constant)
mulc (multiply by a constant)
divc (divide by a constant)
cdo <operator>,c <infile> <outfile>

add (add two fields)
sub (subtract two fields)
mul (multiply two fields)
div (divide two fields)
cdo <operator> <infile1> <infile2> <outfile>

Arithmetic and Statistics

cdo <operator> <infile1> <infile2> <outfile>
cdo <operator-stat> <infile> <outfile>

Remove seasonal cycle:

subtract multi-year monthly time series

multi-year monthly average

```
cdo ymonsub $origpath/SST.HRC06.full.all.nc -ymonavg $origpath/
SST.HRC06.full.all.nc $destpath/SST.HRC06.rmseas.all.nc
```

Statistical functions <stat>:
min, max, sum, mean, avg, var, std

<operator-stat> = <operate on><stat>
vert<stat> (vertical statistical values)
vertmean (vertical mean)

<operate on>:
tim (time)
fld (field)
zon (zonal)
mer (meridional)
run (running statistics)
ymon (multi-year monthly)

Correlation and Covariance

cdo <operator> <infile1> <infile2> <outfile>

Correlation between surface temperature and latent heat flux:

```
cdo timcor $origpath/TS.atmHRC06.rmseas.all.nc  $origpath/
LHF.atmHRC06.rmseas.all.nc $destpath/corr_TS_LHF.atmHRC06.all.nc
```

```
ls -lah TS.atmHRC06.rmseas.all.nc
      547M Oct 21 09:45 TS.atmHRC06.rmseas.all.nc

ls -lah LHF.atmHRC06.rmseas.all.nc
      547M Oct 21 09:47 LHF.atmHRC06.rmseas.all.nc

ls -lah corr_TS_LHF.atmHRC06.all.nc
      869K Oct 21 09:36 corr_TS_LHF.atmHRC06.all.nc
```

Correlation and Covariance

cdo <operator> <infile1> <infile2> <outfile>

Correlation between surface temperature and latent heat flux:

```
cdo timcor $origpath/TS.atmHRC06.rmseas.all.nc  $origpath/
LHF.atmHRC06.rmseas.all.nc $destpath/corr_TS_LHF.atmHRC06.all.nc
```

Lag-correlation between TAUX and Nino3.4:

```
cdo timcor -seldate,0102-01-01,0154-12-01 $origpath/nino3.4.all.nc
-seldate,0103-01-01,0155-12-01 $origpath/TAUX.HRC06.all.13pts.nc
$destpath/lagcorr12mths_nino3.4_TAUX.HRC06.all.nc
```

<operator>:
timcor (correlation over time)
fldcor (correlation in grid space)
timcovar (covariance over time)
fldcovar (covariance in grid space)

Regression

regress Y on X = covar(X,Y)/var(X)

regress X on Y = covar(Y,X)/var(Y)

Regression of latent heat flux on TS
(how LHF changes with TS):

covariance of LHF and SST

```
cdo div -timcovar $origpath/TS.atmHRC06.rmseas.all.nc $origpath/
LHF.atmHRC06.rmseas.all.nc -timcovar $origpath/
TS.atmHRC06.rmseas.all.nc $origpath/TS.atmHRC06.rmseas.all.nc
$destpath/regress_LHFonTS.atmHRC06.all.nc
```

variance of SST

Regression over time/ trends

$y = a + bt$

cdo <operator> <infile1> <outfile>

Regression of SLA with time:

cdo mulc,12000 -regres $origpath/reconSLA_1950to2009.nc $destpath/SLAtrend_b2.1950to2009.nc

Detrend SLA:

cdo detrend $origpath/reconSLA_1950to2009.nc $destpath/SLAdetrend.1950to2009. nc

Trend in SLA:

cdo trend $origpath/reconSLA_1950to2009.nc $destpath/SLAtrend_a.1950to2009.nc $destpath/SLAtrend_b.1950to2009.nc

Interpolation/Remapping

cdo <genweights>,gridfile <infile> <remapwgtfile>
cdo remap,gridfile,remapwgtfile <infile> <outfile>

Remap SST onto atmospheric grid:

- Generate weights for atmospheric grid

cdo genbil,atmHRC06.grid.all.nc $origpath/PRECC.atmHRC06.rmseas.all.nc $destpath/atmHRC06.grid.weights.all.nc

- Remap from ocean to atmospheric grid

cdo remap,atmHRC06.grid.all.nc,atmHRC06.grid.weights.all.nc $origpath/SST.HRC06.rmseas.all.nc $destpath/SST.atmHRC06.rmseas.all.nc

<genweights>:
genbil (bilinear interpolation)
genbic (bicubic interpolation)
gendis (distance-weighted average)
gennn (nearest neighbour)
.....

cdo in shell-script

Ocean heat content in upper 400m (HC400.cdo)

```
#!/bin/sh
# Vertical sum of integrated temperature
# integrate to ~400m => z-level 18
HRCPATH=/bkirtman2/kirtman/HRC06
MYHRCPATH=/projects/rsmas/kirtman/dputrasahan/HRC06/HC400
GRIDPATH=/bkirtman2/dputrasahan/grids
export HRCPATH MYHRCPATH GRIDPATH
YYYs=102
MMs=1
YYYe=155
YYYint=$YYYs

while [ $YYYint -le $YYYe ] ; do
  MMint=$MMs
  while [ $MMint -le 12 ]; do
     if [ $MMint -le 9 ]; then
        MMint=`eval echo "0$MMint"`
     fi
cdo vertsum -sellevidx,1/18 -mul -selname,TEMP $HRCPATH/
HRC06.pop.0$YYYint-$MMint.nc  $GRIDPATH/dzd.HRC06.nc
$MYHRCPATH/HC400.0$YYYint-$MMint.nc
     MMint=`expr $MMint + 01`
  done
  YYYint=`expr $YYYint + 01`
done
```

Running shell-script

"Background mode":

```
screen -S ohc400      (start new screen session)
chmod 754 HC400.cdo   (make script executable)
./HC400.cdo
   ctl + a + d        (detaches screen session)
screen -r ohc400      (reattaches screen session)
exit
```

Happy computing!