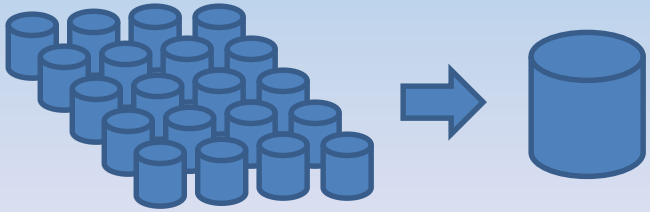
Exploratory analysis of ensembles using pyFerret or Ferret

Steve Hankin
UW/JISAO
NOAA/PMEL/Sci.Dat.Integ.Grp.

April 2016

Ensembles can be explored with
little more effort than an individual
Climate/forecast model run.

Strategy: treat a collection of model runs
as a single dataset,
aggregated along a new ‘E’ (ensemble) axis



Topic 1: creating a single model ensemble

yes? ENSEMBLE my_ensemble = my_run1.nc, my_run2.nc, my_run3 …

```
yes? SHOW DATA
currently SET data sets:
    6> my_ensemble  (default)  Ensemble aggregation
 name   title                I        J     K    L        M     N
 HEIGHT height              ...      ...   ...  ...      1:5   ...
 TAS    Air Temperature   1:256    1:128   ...  1:1140   1:5   ...
```

Lon axis

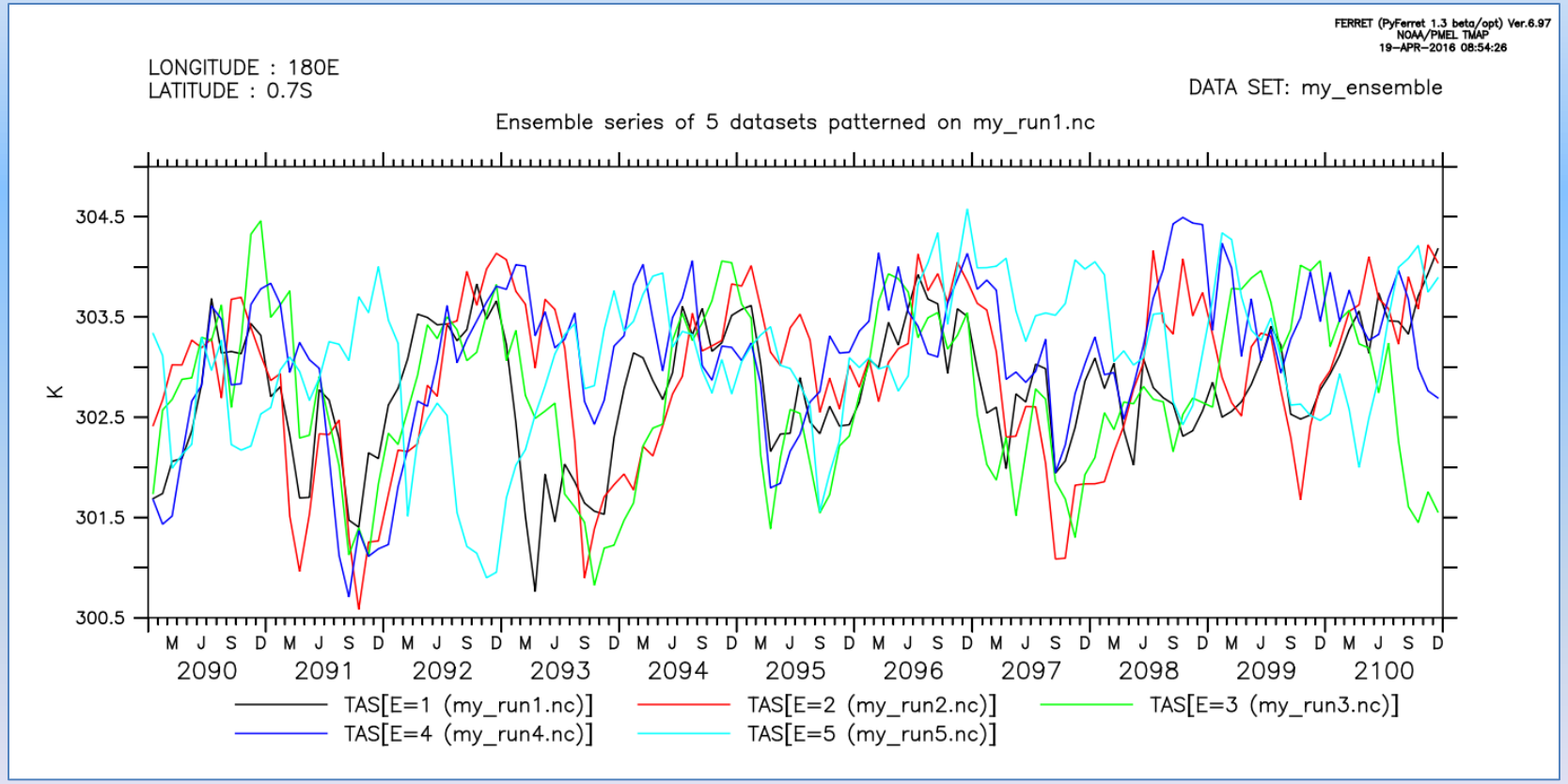
Lat axis

Time axis

Ensemble member axis

Ready to begin exploring the ensemble …

yes? PLOT/ALONG=t/X=180/Y=0 tas



Topic 2: Creating a multi-model ensemble

```
yes? ! Open a set of model output datasets
yes? ! They might be OPeNDAP url's

yes? USE GFDL_CM3,      \
       GFDL_ESM2G,     \
       GFDL_ESM2M,     \
       ACCESS10,      \
       BCC_CSM1m,      \
       CanESM2,        \
       CCSM4,          \
       CESM1_CAM5,     \
       CMCESM1_WACCM,\
       CNRM_CM5,       \
       …
```

Grids differ between models

GFDL:

```
yes? show grid

 name          axis          # pts   start          end
 LON           LONGITUDE     144mi   1.25E          1.25W
 LAT           LATITUDE       90 i   89S            89N
 TIME          TIME         1140 i   16-JAN-2006    16-DEC-2100
```

other models:

```
               LONGITUDE     192mr   0E             1.875W
               LATITUDE      145 i   90S            90N
               LONGITUDE     320mr   0E             1.125W
               LATITUDE      160 i   89.142S       89.142N
               TIME         1128 i   16-JAN-2006    16-DEC-2099

                  etc …
```

CF calendars include ‘NOLEAP’, proleptic Gregorian, and ‘360_day’

Ensemble member variables of the same name must share the same coordinate grid.
We choose to regrid to the GFDL grid (*)

e.g. to change the grid of variable ‘tas’ in dataset 5 to GFDL’s (‘d=1’)

```
> pyferret
   yes? SET VARIABLE/NAME=tas_orig   tas[d=5]
   yes? LET/d=5 tas = tas_orig[g=tas[d=1], gt=@asn]
```

Regrid lat and lon

“associate” calendars(*)

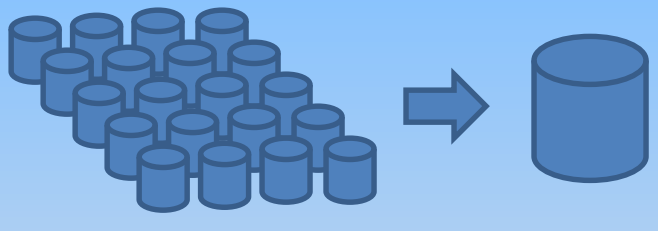
(*) The default regridding is multi-linear interpolation, but TIME, where the calendars types differ, requires special treatment. For monthly data we choose to ignore the differences, and we regrid time axes by simple association (‘@asn’) of points.

Now, regrid all the models with these two commands(*) …

```
yes? REPEAT/NAME=dset/RANGE=2:12    \
    ( SET VAR/NAME=tas_orig tas[d=`dset`]; \
      LET/d=`dset` tas = tas_orig[g=tas[d=1],gt=@asn]  )
```

(*) pyFerret uses ‘delayed mode analysis’, so this regridding loop is making definitions only. The regridding operation will be done when data is loaded, e.g. on a PLOT call.

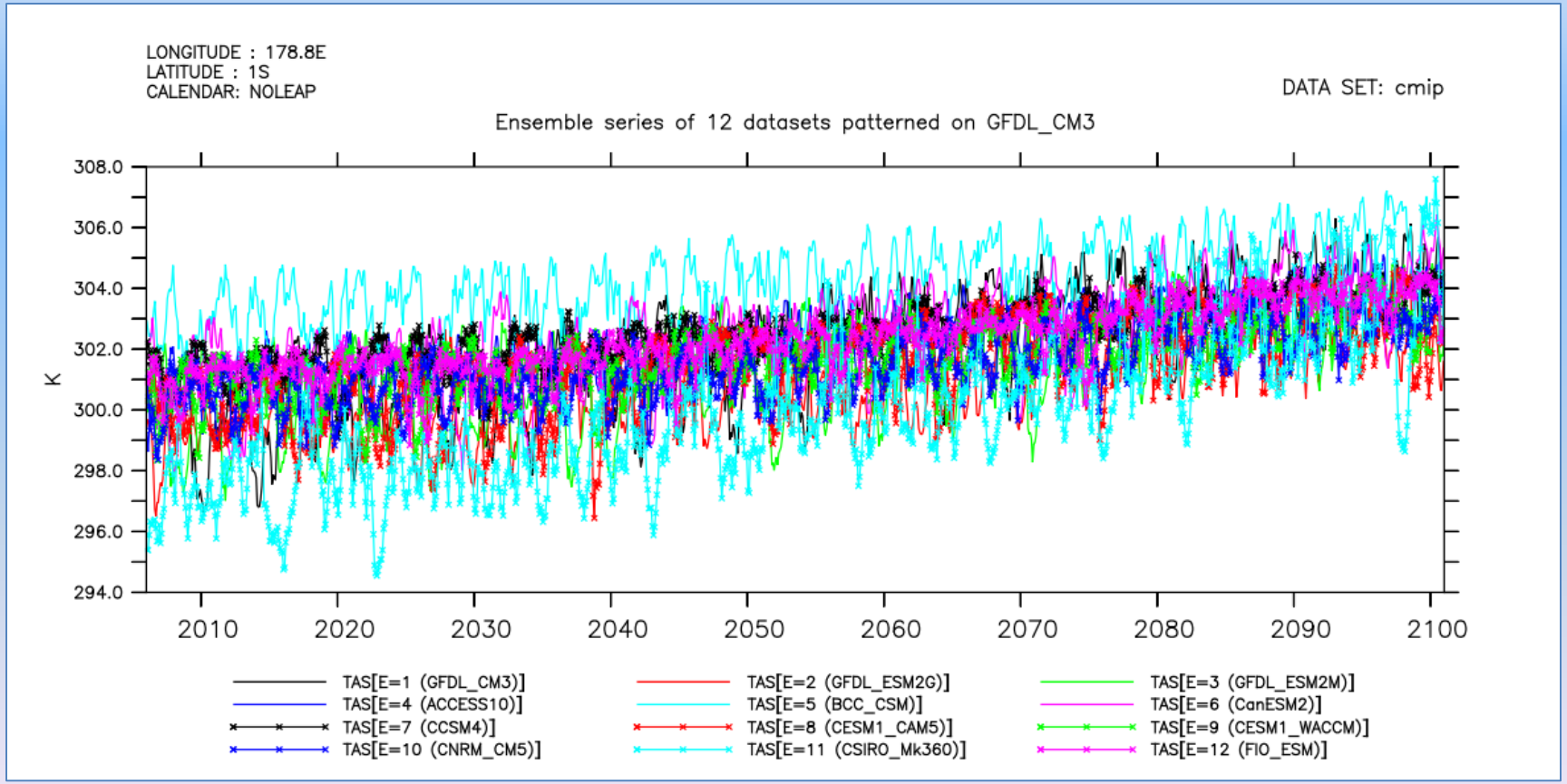
yes? ENSEMBLE cmip = 1, 2, 3, … , 12



We have created a 12 member multi-model ensemble, “cmip”

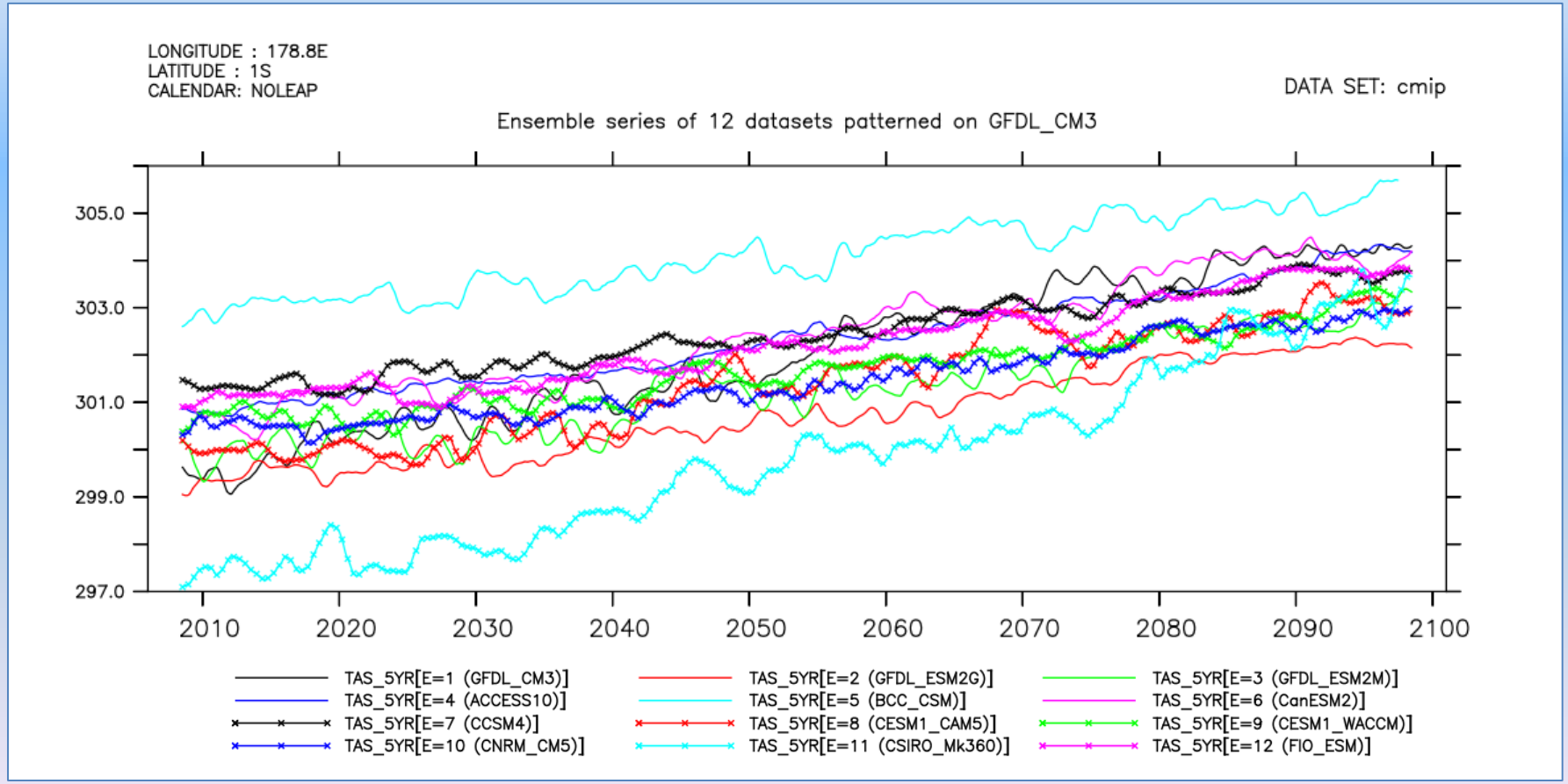
Begin exploring the multi-model ensemble

yes? SET REGION/X=180/Y=0
yes? PLOT/ALONG=t tas



Apply a 60 month (5 yr) running average

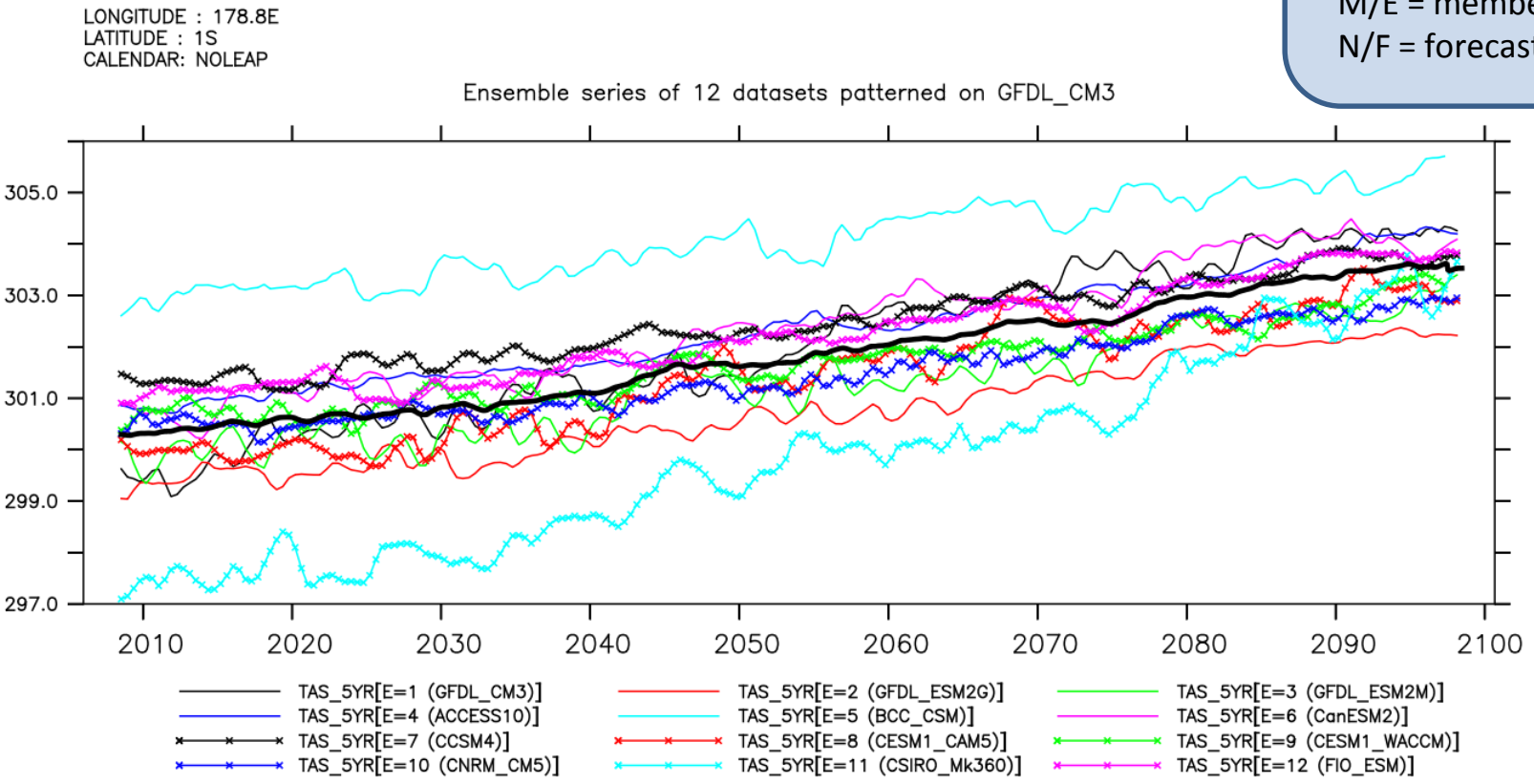
yes? LET tas_5yr = tas[t=@SBX:60]
yes? PLOT/ALONG=T tas_5yr



Overlay the ensemble mean

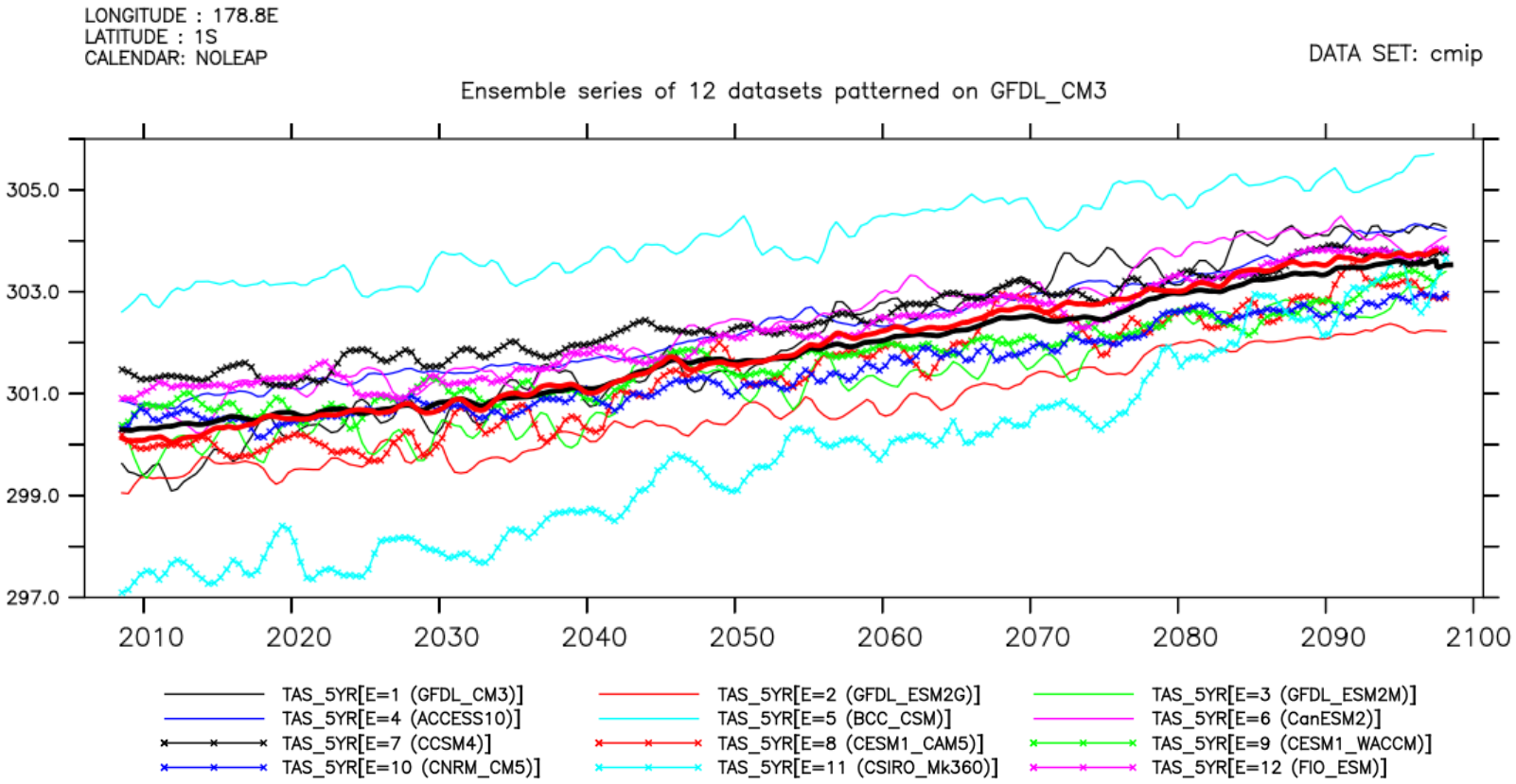
yes? PLOT/OVER/COLOR=black tas_5yr[m=@ave]

I/X = longitude
J/Y = latitude
K/Z = depth
L/T = time
M/E = member no.
N/F = forecast run



Create and overlay a weighted ensemble average

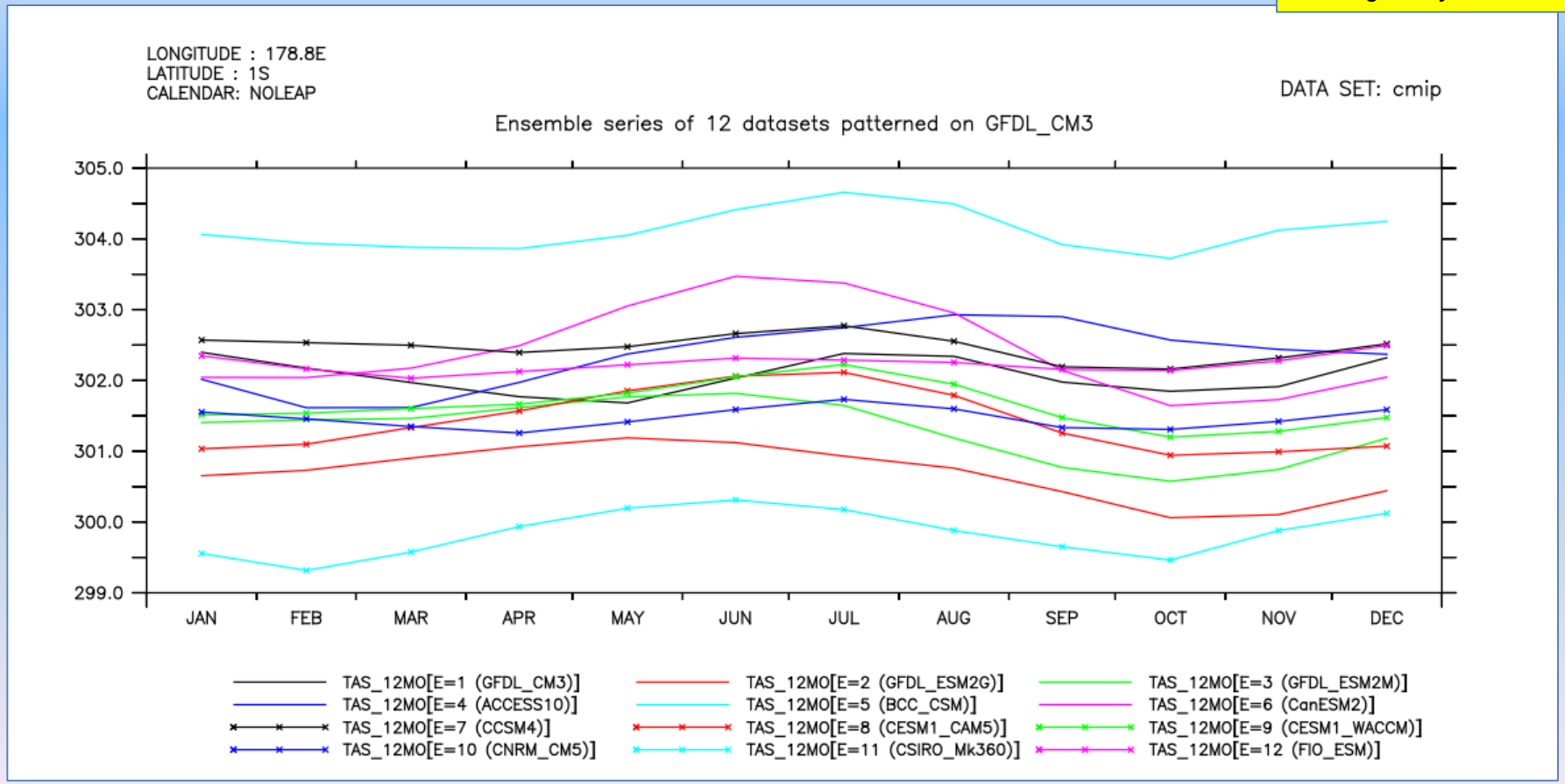
yes? LET weights = ESEQUENCE({5, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1})
yes? LET wgt_runs = tas_5yr * weights / weights[m=@sum]
yes? PLOT/OVER/COLOR=RED wgt_runs[m=@sum]



Plot the 12 month climatological average at a point

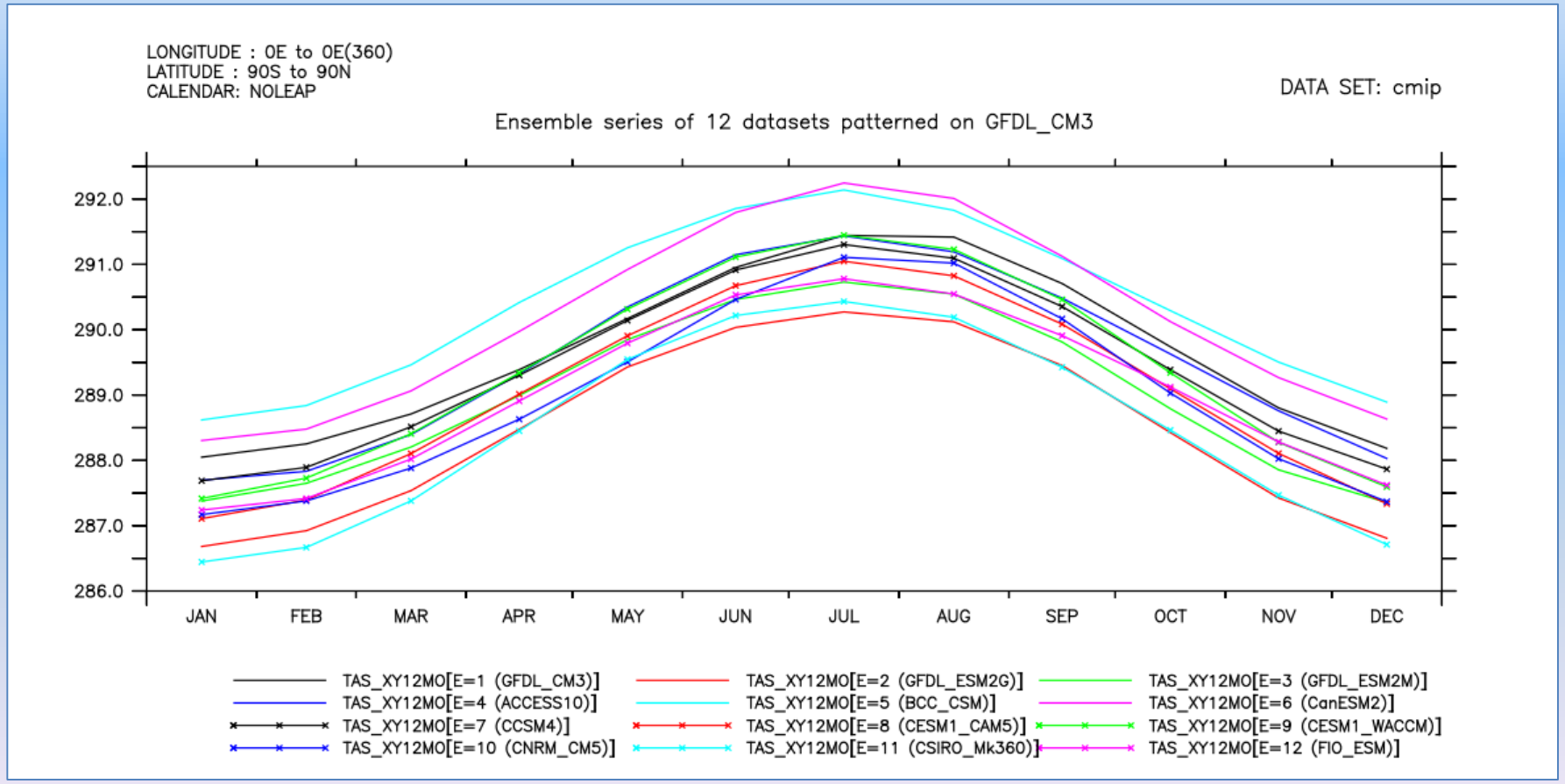
yes? LET tas_12mo = tas[gt=MONTH_NOLEAP@mod]
yes? PLOT/ALONG=T tas_12mo

Weak annual cycle on the equator. What about globally?



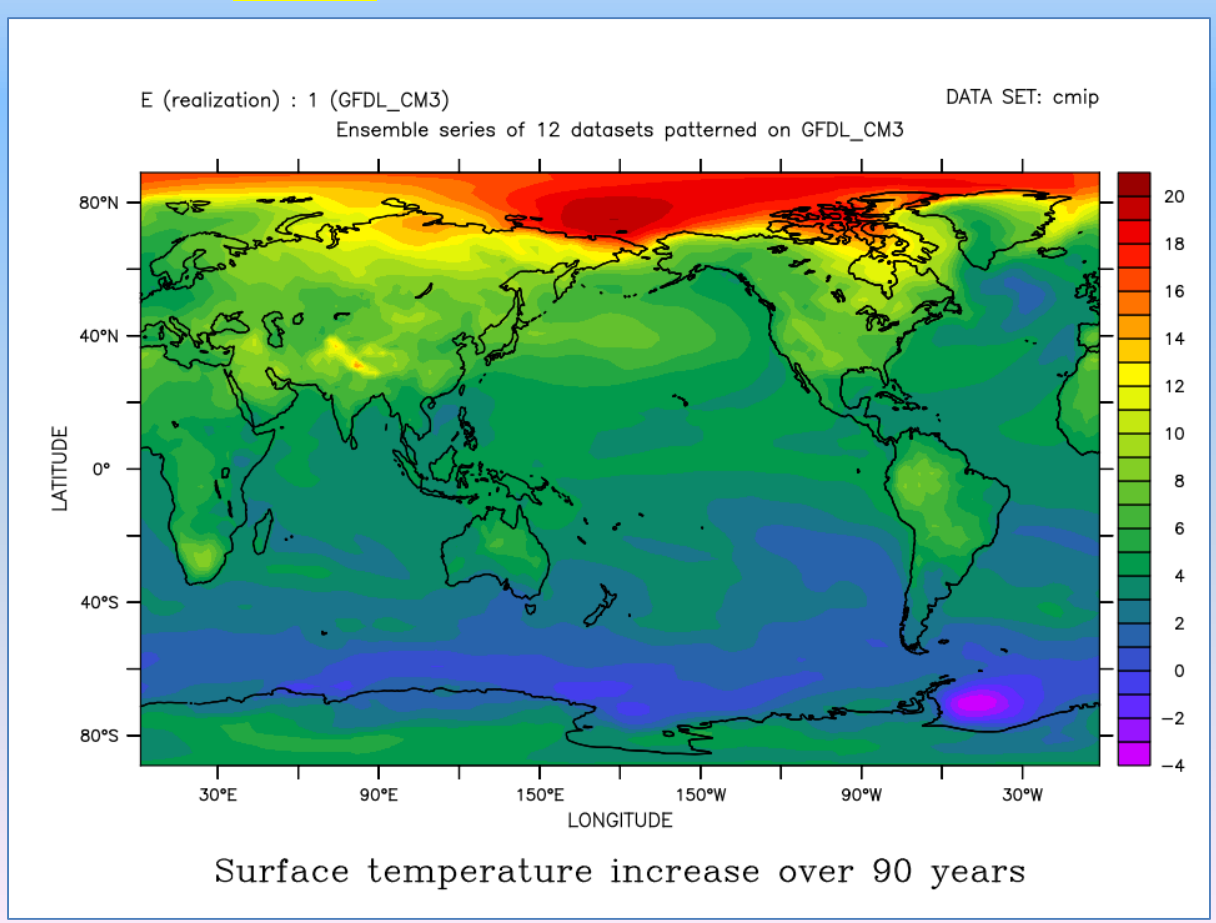
Global average climatological cycle

yes? LET tas_xyav = tas_12mo[x=@ave,y=@ave]
yes? PLOT/ALONG=T tas_xyav



GFDL model ("m=1") temperature change over 90 years …

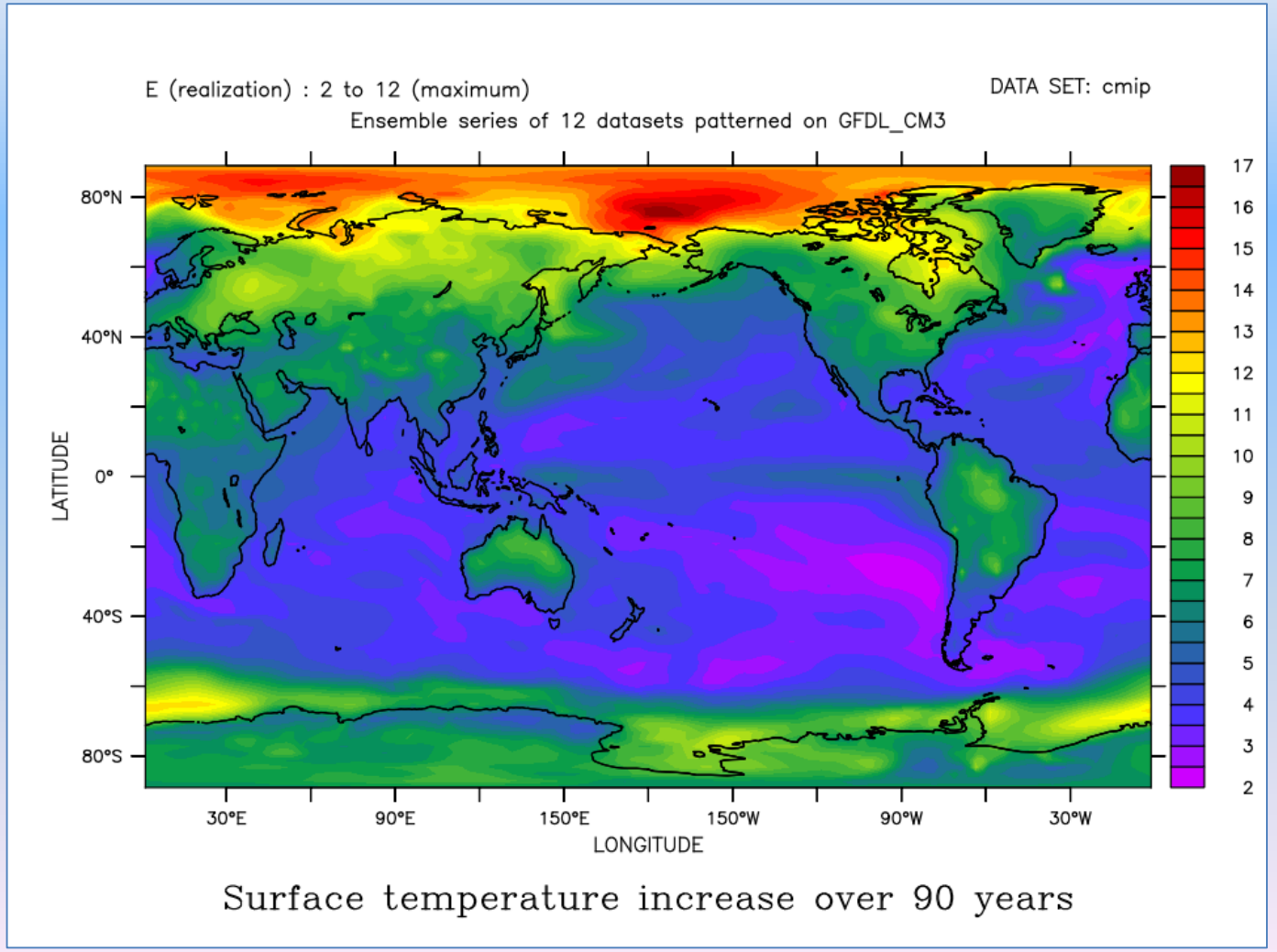
```
yes? LET tas_2007 = tas[t=1-jan-2007:31-dec-2007@ave]
yes? LET tas_2097 = tas[t=1-jan-2097:31-dec-2097@ave]
yes? LET/TITLE="…" tas_90yr = tas_2097 - tas_2007
yes? FILL tas_90yr[m=1]
```



Surface temperature increase over 90 years

Ensemble-maximum of the other models …

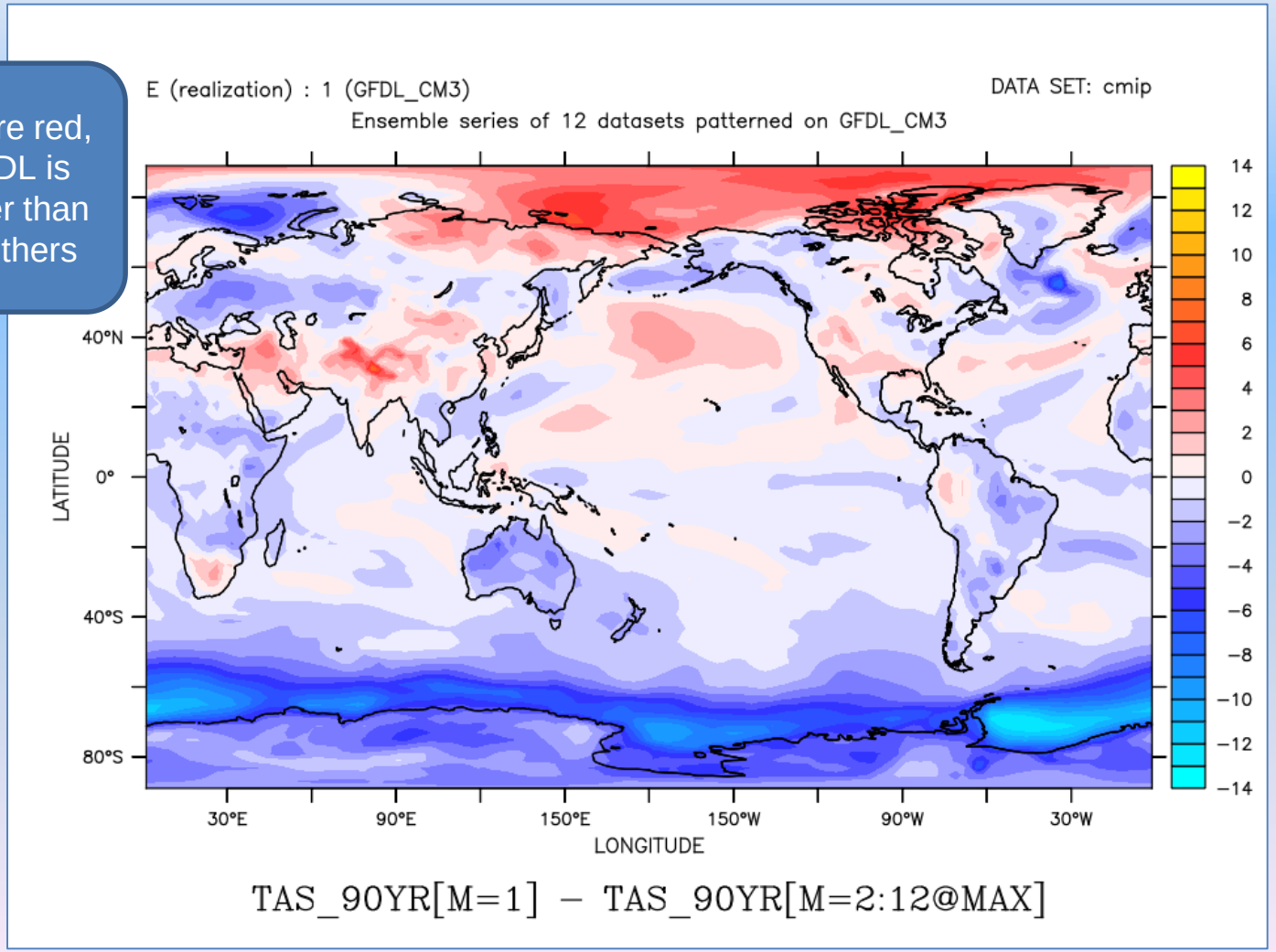
yes? FILL tas_90yr[m=2:12@max]



Anomaly of GFDL relative to ensemble max

yes? FILL tas_90yr[m=1] - tas_90yr[m=2:12@max]

Where red, GFDL is hotter than all others



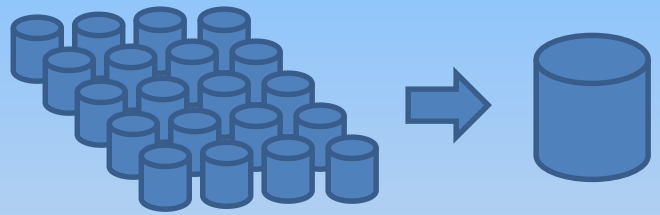
Ensembles of ensembles…

Institutions may contribute ensemble runs (typ. 3-5 members) to the multi-model modeling effort, so the multi-model ensemble is actually an ensemble of ensembles

Topic 3: creating an ensemble of ensembles

Define a run-ensemble for the model FIO_ESM (*)

yes? ENSEMBLE FIO_ESM_ens = SPAWN("ls FIO_ESM_run?.nc")



… and similarly for

- CanESM2
- CNRM_CM5

(*) SPAWN executes the indicated Linux shell command

Redefine ‘tas’ as the ensemble run average from FIO_ESM

Step 1. Rename the native "tas" variables

```
yes? set var/name=tas_orig    tas[d=FIO_ESM_ens]
```

Step 2. Compute the ensemble means

```
yes? LET/D=FIO_ESM_ens      tas_Mave = tas_orig[m=@ave]
```

Step 3. Regrid to the GFDL grid

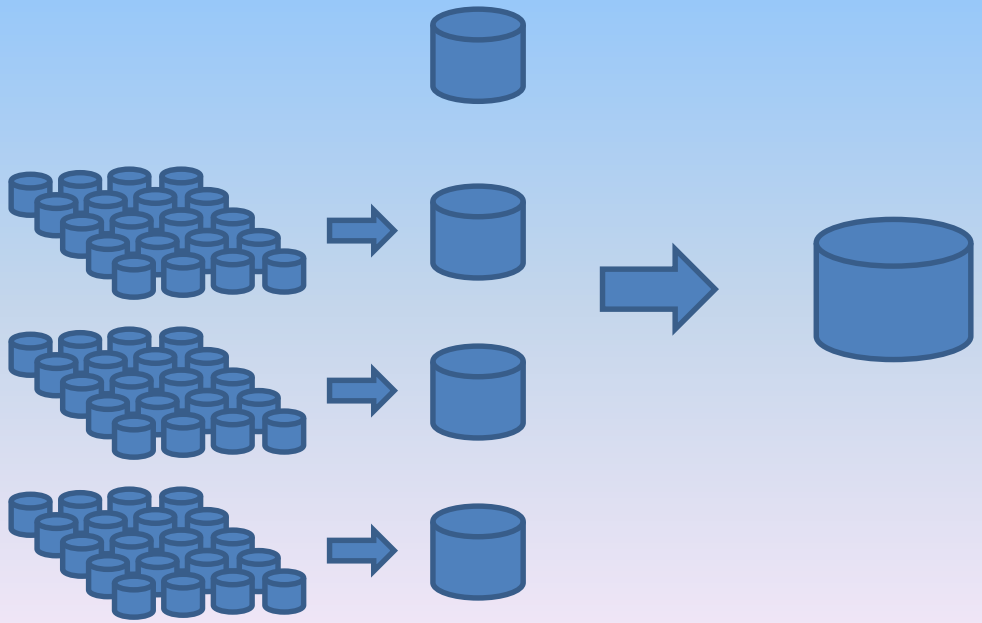
```
yes? LET/D=FIO_ESM_ens      tas = tas_Mave[g=tas[d=1], gt=@asn]
```

Now use the identical 3 steps to create ensemble run means of the other models

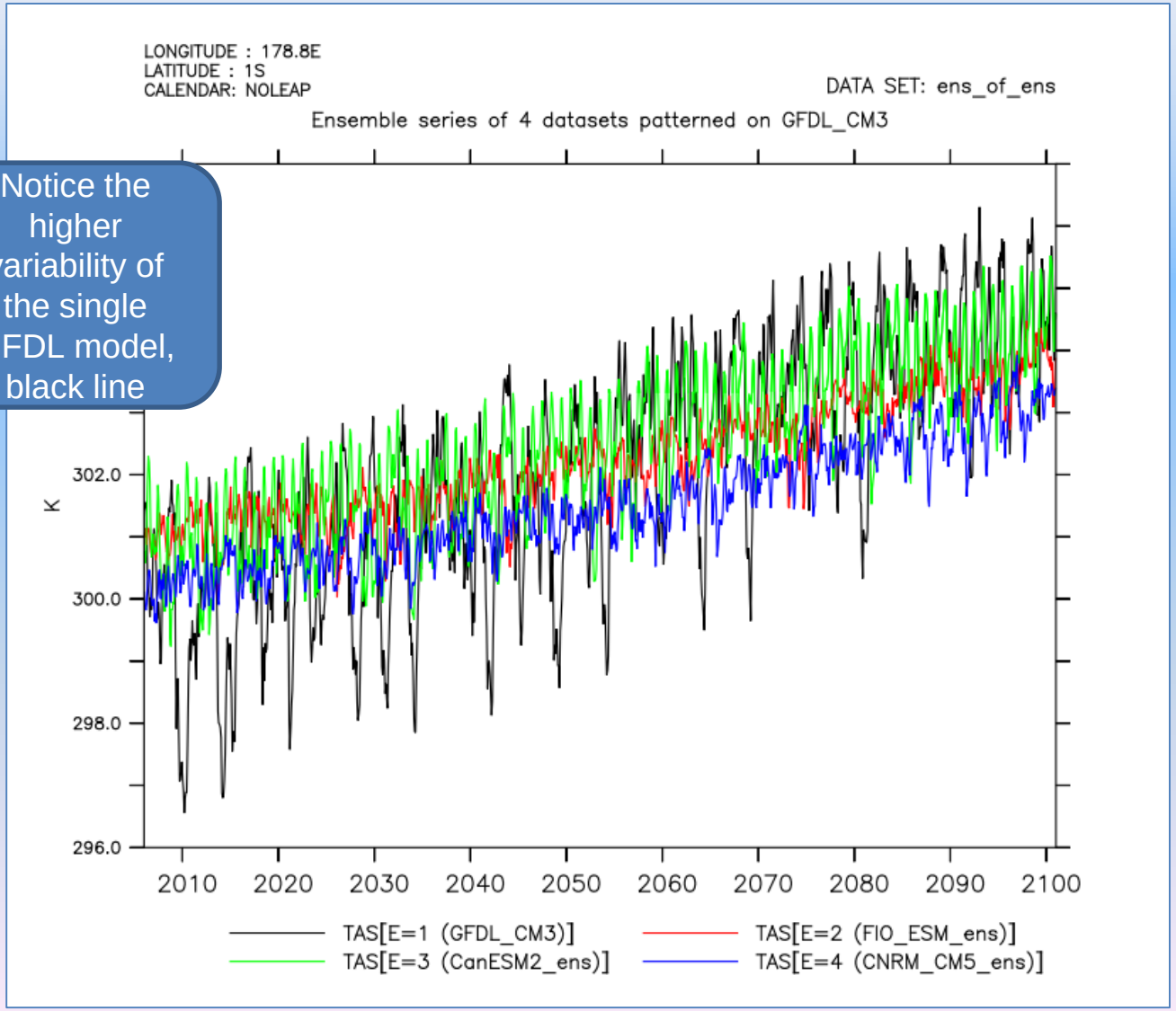
- CanESM2
- CNRM_CM5

Build the ensemble of ensembles

```
yes? ENSEMBLE ens_of_ens =  GFDL_CM3,     \
                            FIO_ESM_ens, \
                            CanESM2_ens, \
                            CNRM_CM5_ens
```



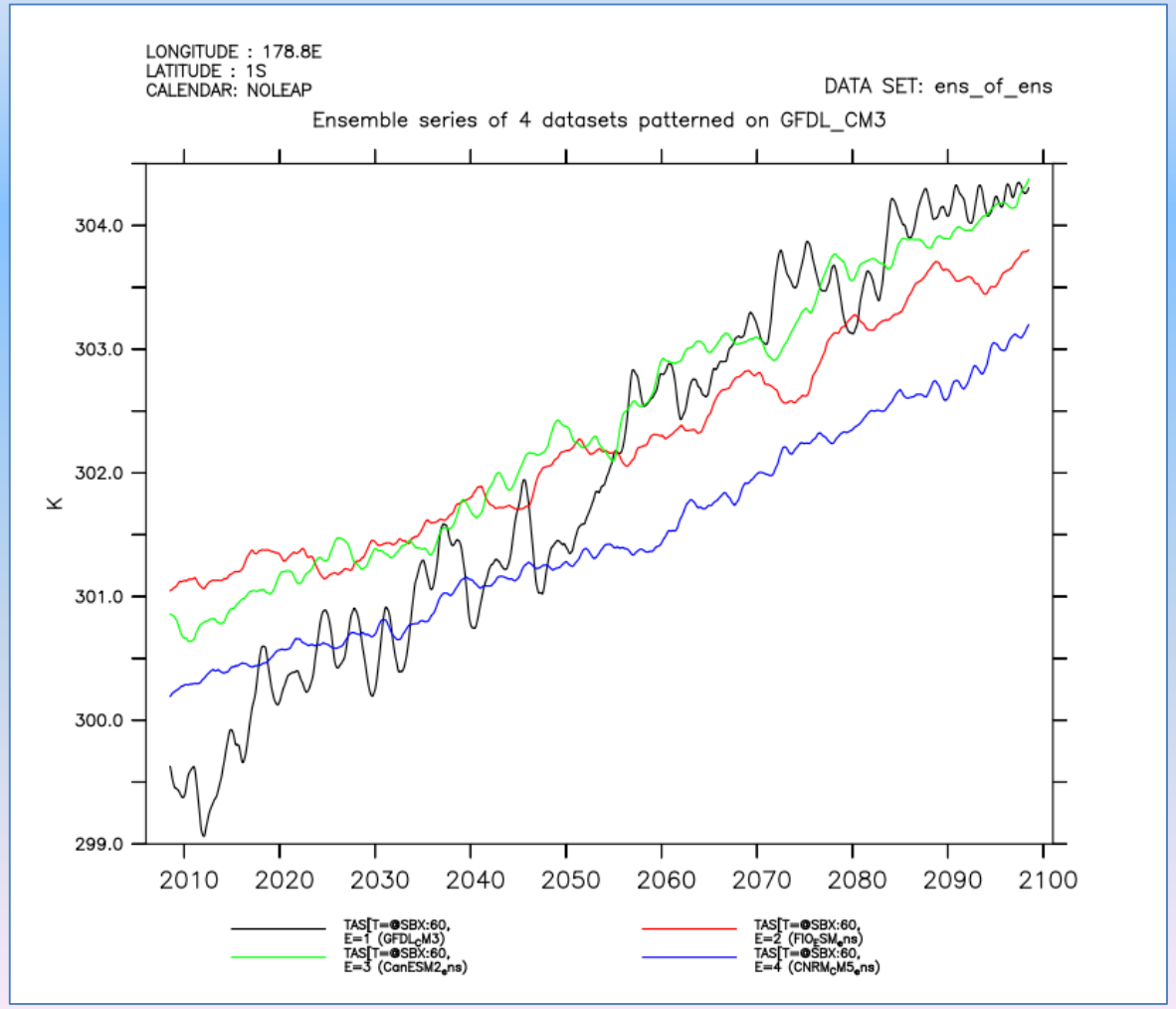
yes? SET REGION/X=180/Y=0
yes? PLOT/ALONG=t tas



Notice the higher variability of the single GFDL model, black line

Again, a 60 month (5 yr) running average smoother

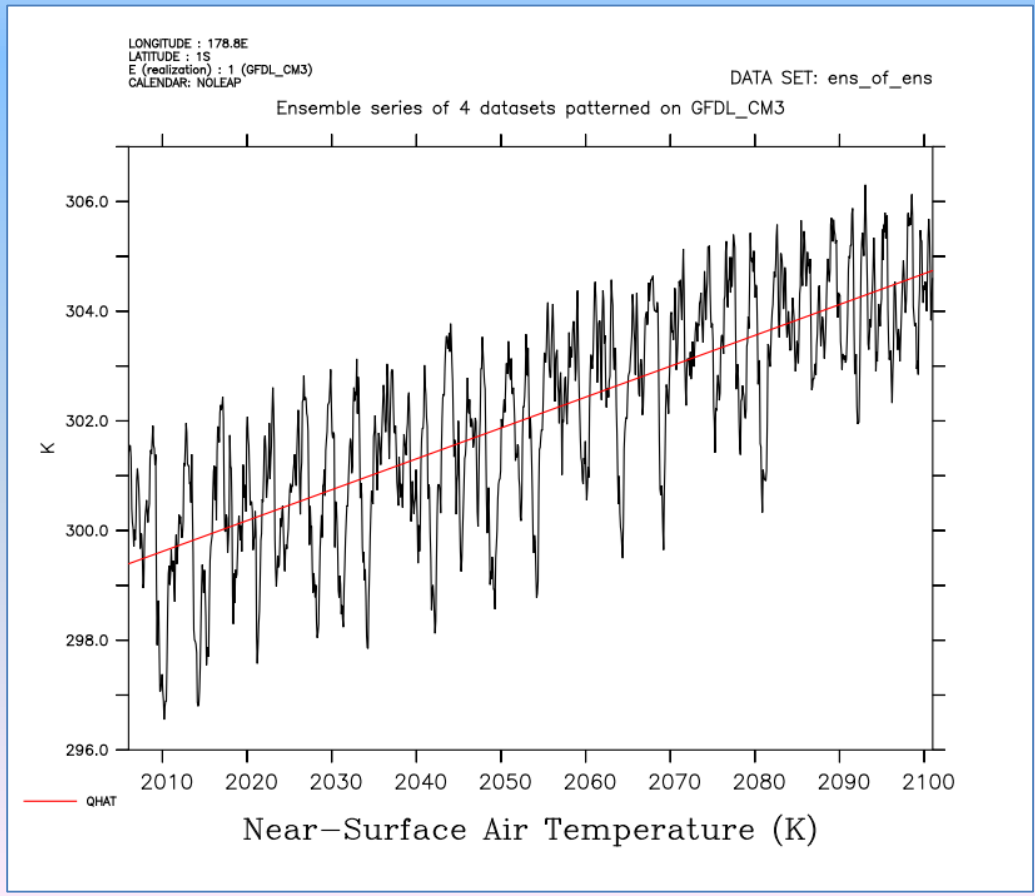
yes? PLOT/ALONG=T tas[t=@SBX:60]



Linear-regression trends
Start with the GFDL model, [m=1] …

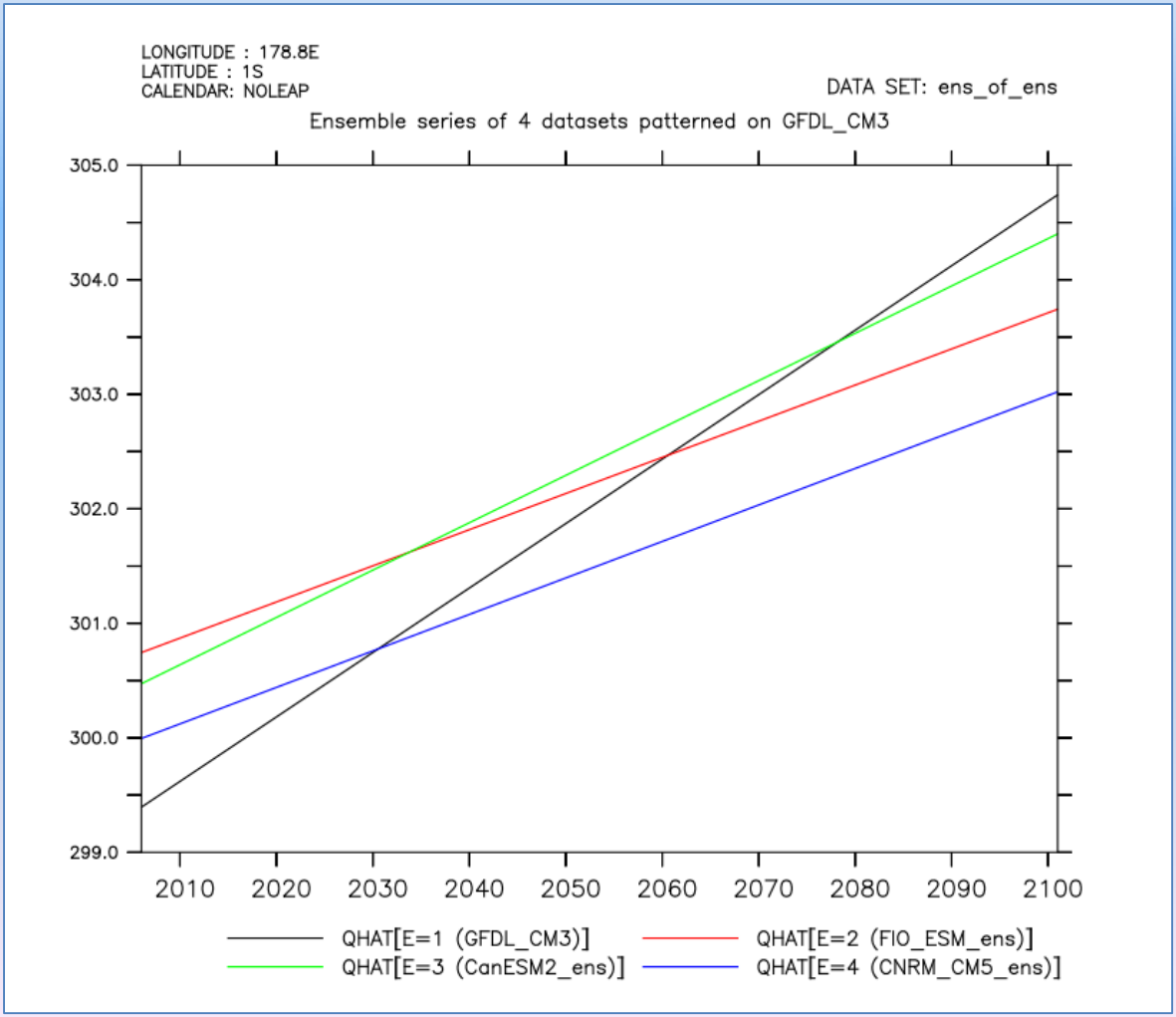
yes? GO regresst
yes? LET q = tas; LET p = T[g=tas]
yes? PLOT tas[m=1]
yes? PLOT/OVER qhat[m=1]

The script ‘regresst’ defines variables ‘slope’ and ‘qhat’ in terms of variables ‘p’ and ‘q’



The trend lines of all ensemble members at this point…

yes? PLOT/ALONG=T qhat

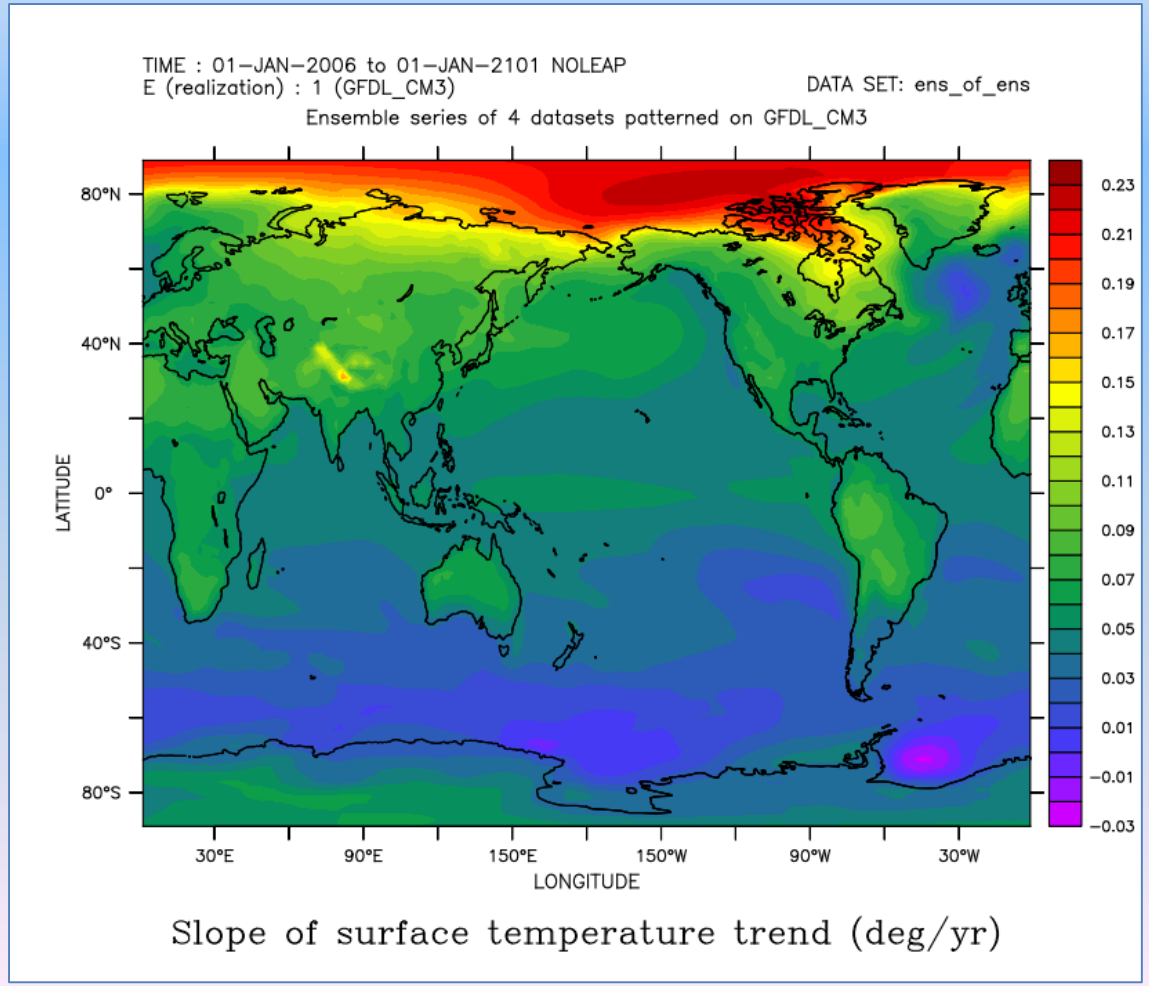


Slope of the 100 yr temperature trend in the GFDL model

yes? LET/title="…" slope_yr = 365*slope

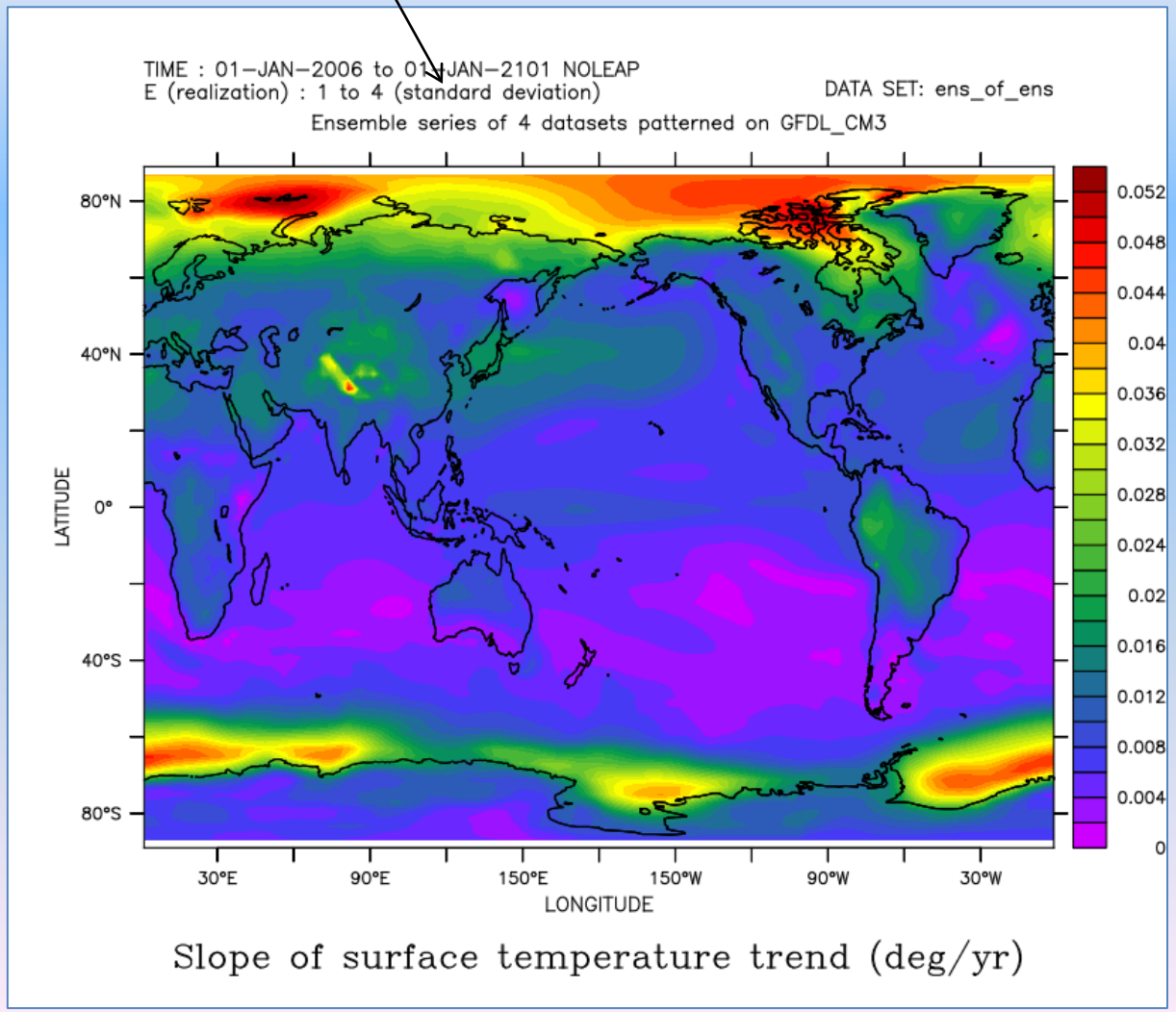
yes? FILL slope_yr[m=1]

Change slope units from 'per day' to 'per year'



Ensemble variability of the slope

yes? FILL slope_yr[m=@std]



Topic 4: ensembles of native-grid analyses

When analyzing a non-linear function

$$\underline{f(x)} \neq f(\underline{x})$$

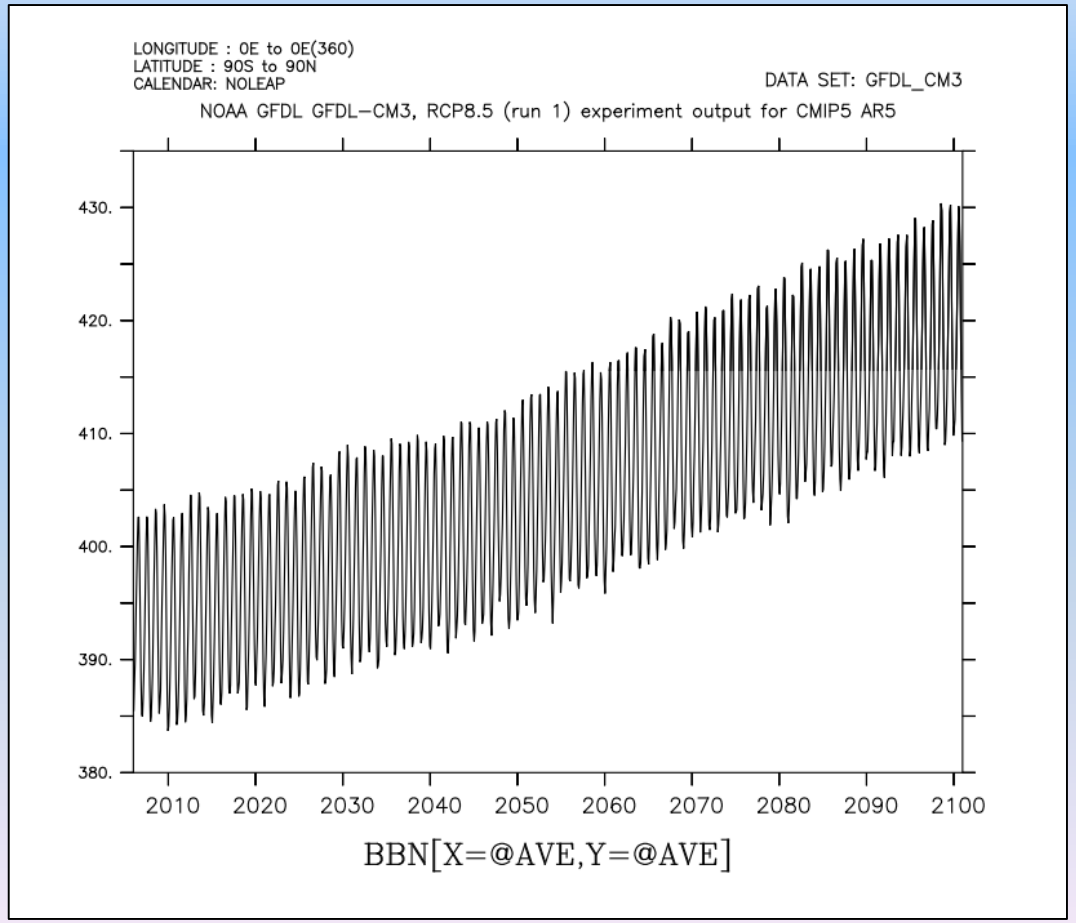
Sometimes we must compute ensemble-member variables on their native grids before defining the ensemble

Our test data includes only surface temperature, “tas”.
so we will calculate black body radiation.

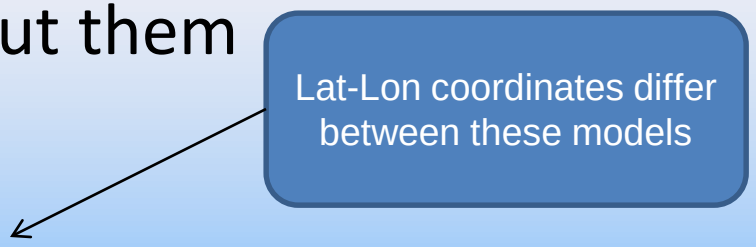
```
yes? LET/TITLE="Black Body Nonsense" bbn = sigma * tas^4
yes? LET sigma = 5.67E-8
```

Globally-averaged Black Body Nonsense from the GFDL model:

yes? USE GFDL_CM3.nc
yes? LET/D=GFDL_CM3 bbn_ave = bbn[x=@ave,y=@ave]
yes? PLOT bbn_ave



Globally-average bbn in three models on their native coordinates. Put them onto a common time axis.




```
yes? USE ACCESS10.nc,   BCC_CSM.nc,  CanESM2.nc

yes? LET/D=ACCESS10 bbn_ave = bbn[x=@ave,y=@ave,gt=time@asn]
yes? LET/D=BCC_CSM  bbn_ave = bbn[x=@ave,y=@ave,gt=time@asn]
yes? LET/D=CanESM2  bbn_ave = bbn[x=@ave,y=@ave,gt=time@asn]
```

Create an ensemble of the results:

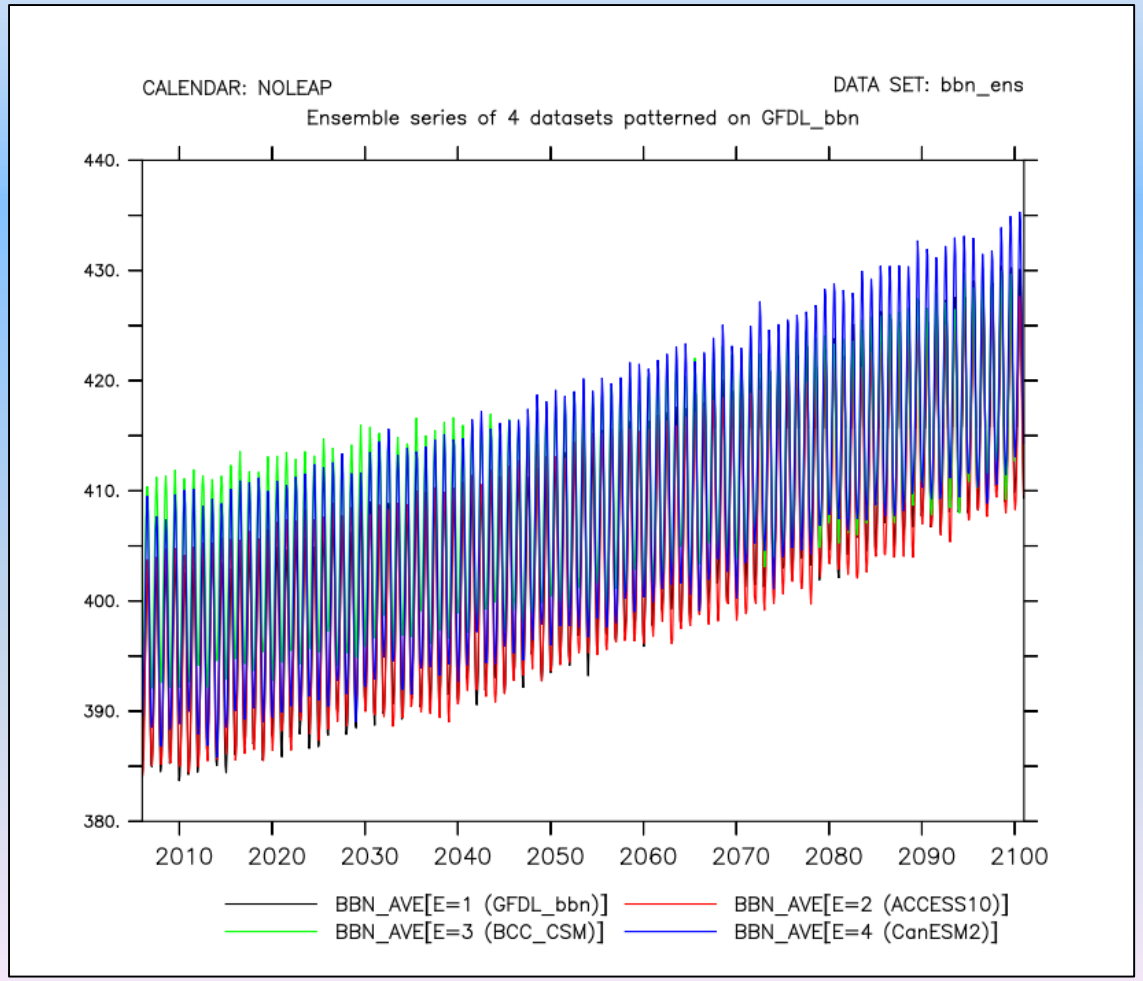
```
yes? SAVE/file=GFDL_bbn.nc    bbn_ave[d=GFDL_CM3]     !  (*)

yes? ENSEMBLE   bbn_ens = GFDL_bbn, ACCESS10, BCC_CSM, CanESM2
```

(*) An ensemble variable must exist in a netCDF file of at least one ensemble member.

Plot the native-averaged ensemble variable:

yes? PLOT/ALONG=t bbn_ave



… again apply a 5-year smoother:

yes? LET bbn_yr = bbn_ave [l=@sbx:60]
yes? PLOT/ALONG=t bbn_y5r