NCAR UCAR

Diagnostics from the CESM Postprocessing Tool

Keith Oleson

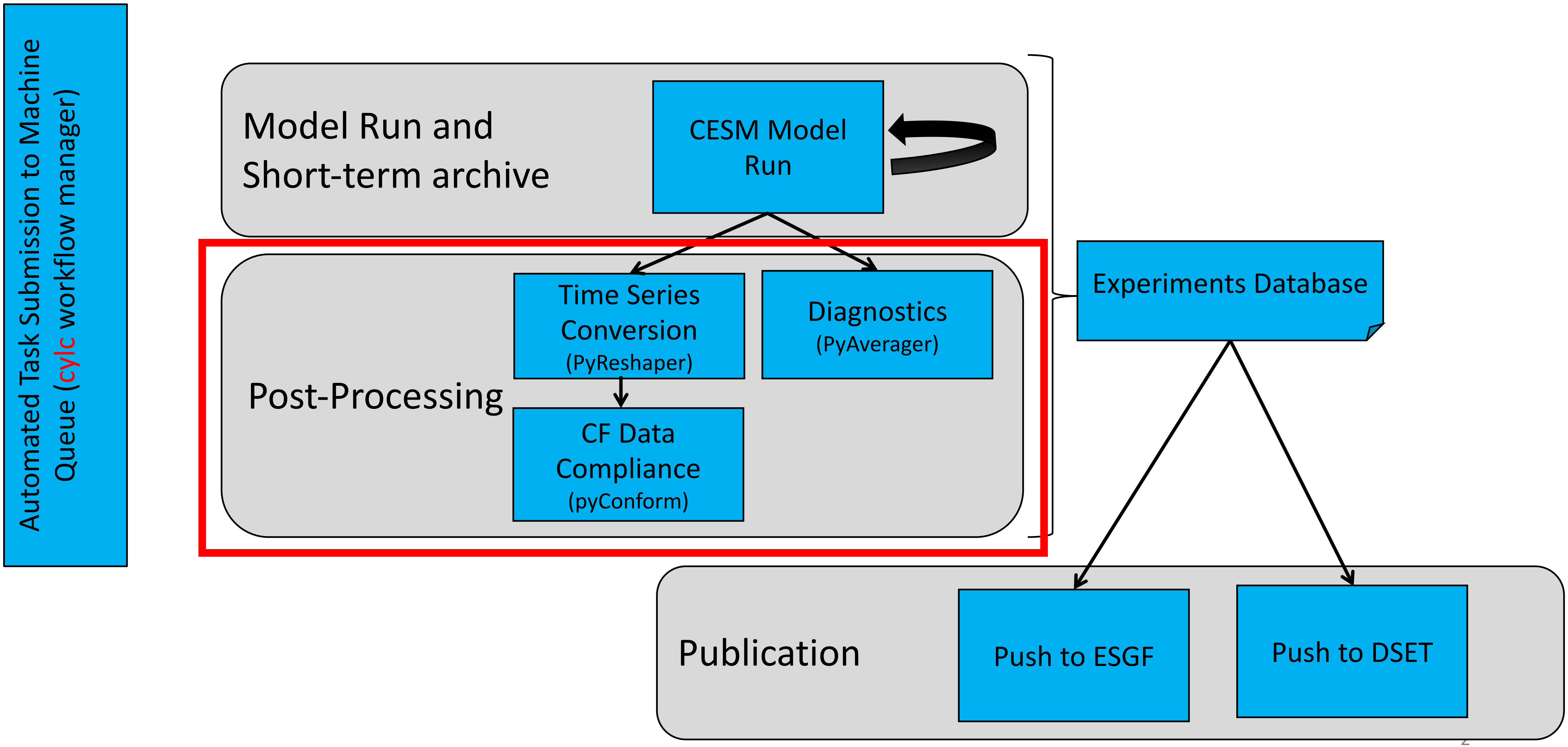
Terrestrial Sciences Section (TSS)

Land Model Working Group Liaison

Thanks to Alice Bertini and Adam Phillips

CGD Climate & Global Dynamics

CESM2 and CMIP6 Experiments Workflow



CESM python based parallel post-processing tools

http://github.com/NCAR/CESM_postprocessing

- Python virtualenv framework (a tool to create isolated Python environments) currently supported on Cheyenne and DAV (Casper) that incorporates **community developed tools** including:
  - CISL ASAP parallel python tools to create
    - single variable timeseries (pyReshaper)
    - averages and climatologies (pyAverager)
    - CF (climate and forecast) compliant output variables (pyConformer)
  - NCL based diagnostics plotting packages from AMWG, LIWG, **LMWG**, OMWG, PCWG and WAWG.
  - **ILAMB (International Land Model Benchmarking)**/IOMB python based diagnostics benchmark comparison plotting package for Land and Ocean
  - Ocean high-resolution diagnostics plots (0.1 degree)

- **Cheyenne and DAV quickstart guide:** https://github.com/NCAR/CESM_postprocessing/wiki/cheyenne-and-DAV-quick-start-guide
- **Post-processing user's guide:** https://github.com/NCAR/CESM_postprocessing/wiki/CESM-Python-Post-Processing-Users-Guide

CESM History Files vs. Timeseries Files

History files contain all variables for a component for a particular frequency, and are output directly from the model.

Timeseries files are created offline from the model, either by the official CESM post-processing tool (run on Cheyenne/DAV machines), or by individual user-generated scripts. Timeseries files span a number of timesteps, and contain only one (major) variable.

Timeseries files are compressed thus saving significant disc space. Each file includes some additional key variables (e.g., landfrac, area, which are needed to calculate global/regional averages/sums).

The diagnostics package will operate on either monthly history files or timeseries files.

Example history file: clm5.clm2.h0.1993-11.nc

- 1 monthly timestep (Nov. 1993)
- 480+ CLM variables (e.g. GPP, TSA, RAIN, etc.)

Example timeseries file: clm5.clm2.h0.GPP.185001-201412.nc

- 165X12 monthly timesteps (Jan 1850 – Dec 2014)
- 1 CLM variable (GPP), along with auxiliary variables (time,lat, lon, landfrac, area, etc.)

Diagnostics Packages

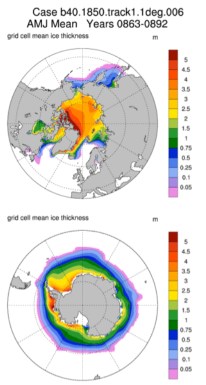
What are they?

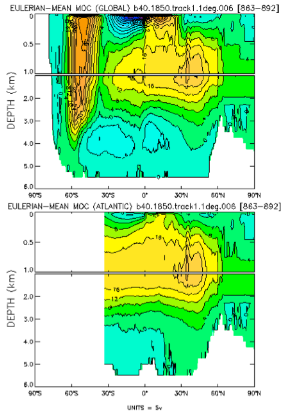
A set of NCL/python scripts that automatically generate a variety of different plots from model output files that are used to evaluate a simulation.

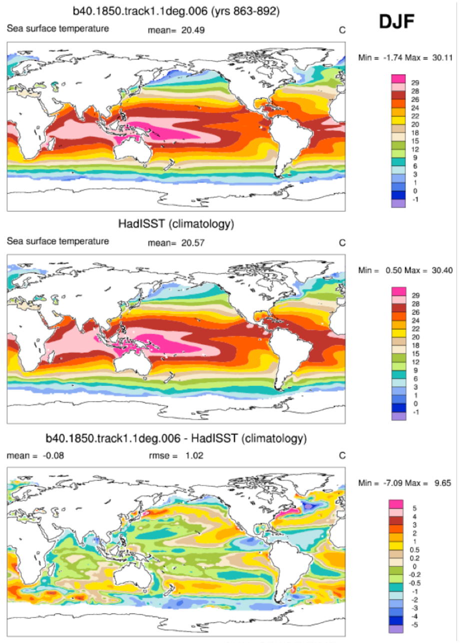
Why are they used?

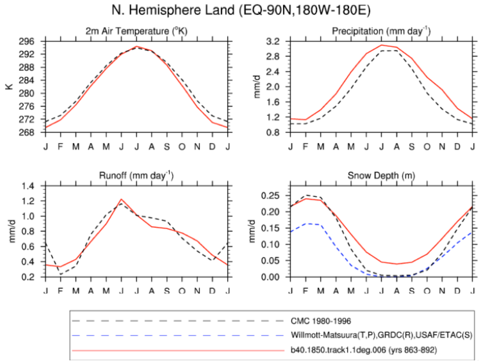
The diagnostics are the easiest and fastest way to get a picture of the mean climate of your simulation. They can also show if something is wrong.

Note: The component diagnostics packages can be used as the first step in the research process, but the general nature of the calculations does not lend itself to in-depth investigation.







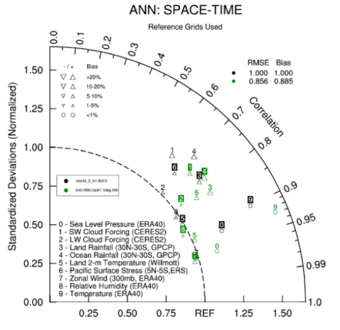


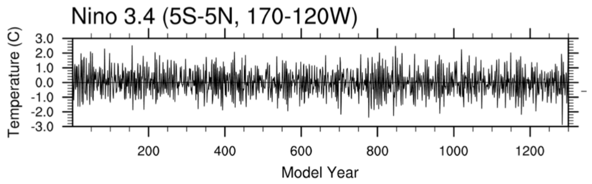

```
++++++++++++++++++++++++++++++
CPL7 Surface Energy Budget for Years:  1 1300
  (+ ==> model gains heat, - ==> model loses heat)
           FREEZE     MELT    RUNOFF        SW        LW       LAT       SEN       NET  (W/m^2)
         -------- -------- -------- -------- -------- -------- -------- --------
   ATM:     0.000    0.000    0.000 -167.337   64.023   85.696   17.776    0.157
   OCN:     0.128   -0.992   -0.062  124.207  -41.138  -71.886  -10.366   -0.109
 ICE_N:    -0.062    0.365    0.000    0.495   -0.862   -0.101    0.159   -0.005
 ICE_S:    -0.066    0.627    0.000    0.609   -1.313   -0.265    0.399   -0.009
   LND:     0.000    0.000    0.062   42.027  -20.710  -13.444   -7.969   -0.034

 O+I+L:    -0.000    0.000    0.000  167.338  -64.023  -85.696  -17.777   -0.158

  OCN NET renormalized:   -0.15341

    Mean ATM frac:    1.00000
    Mean OCN frac:    0.65350
  Mean ICE_N frac:    0.02278
  Mean ICE_S frac:    0.03198
    Mean LND frac:    0.29174
++++++++++++++++++++++++++++++
```

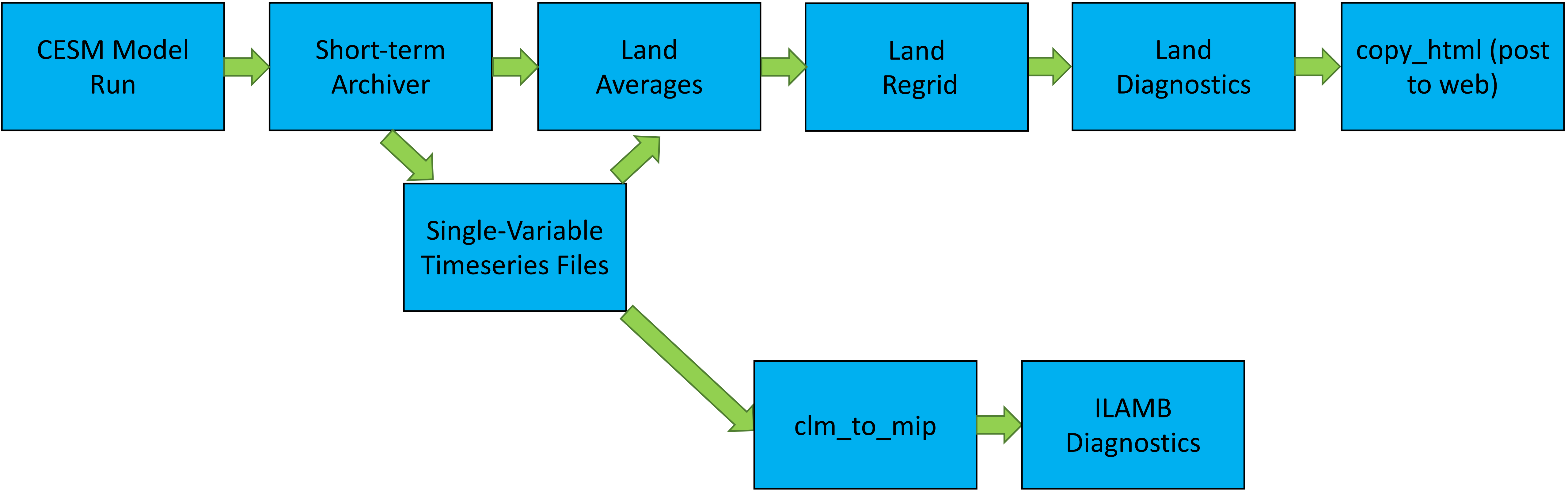




CESM python based parallel post-processing tools

http://github.com/NCAR/CESM_postprocessing

Land



LMWG Diagnostics Package Output

NCAR
National Center for Atmospheric Research
Boulder, Colorado

b.e20.BHIST.f09_g17.20thC.297_01 and b.e20.BHIST.f09_g17.20thC.289_02

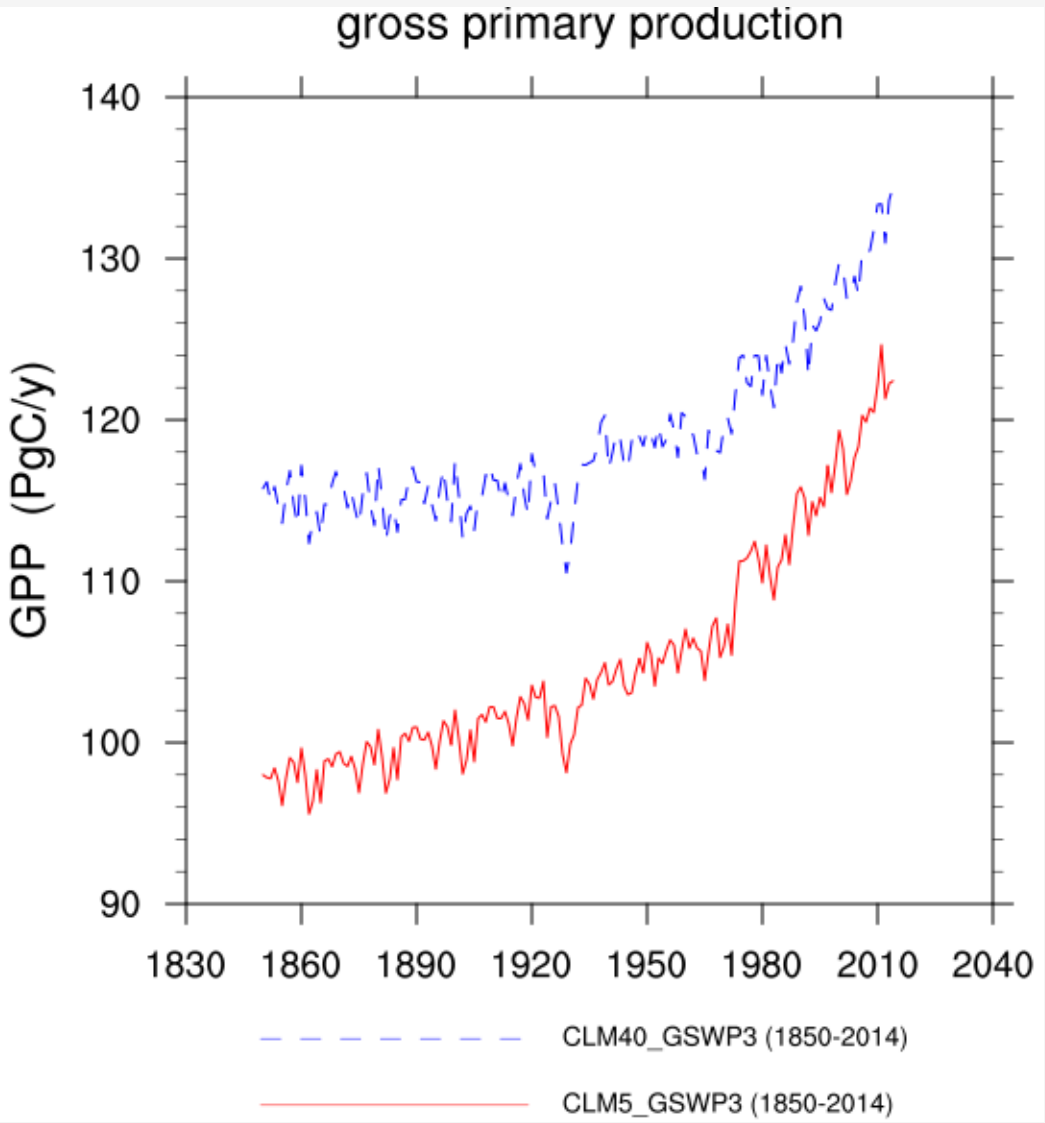
LND_DIAG Diagnostics Plots Source: /glade/p/cesm/postprocessing_ch/lnd_diag/

Set Description

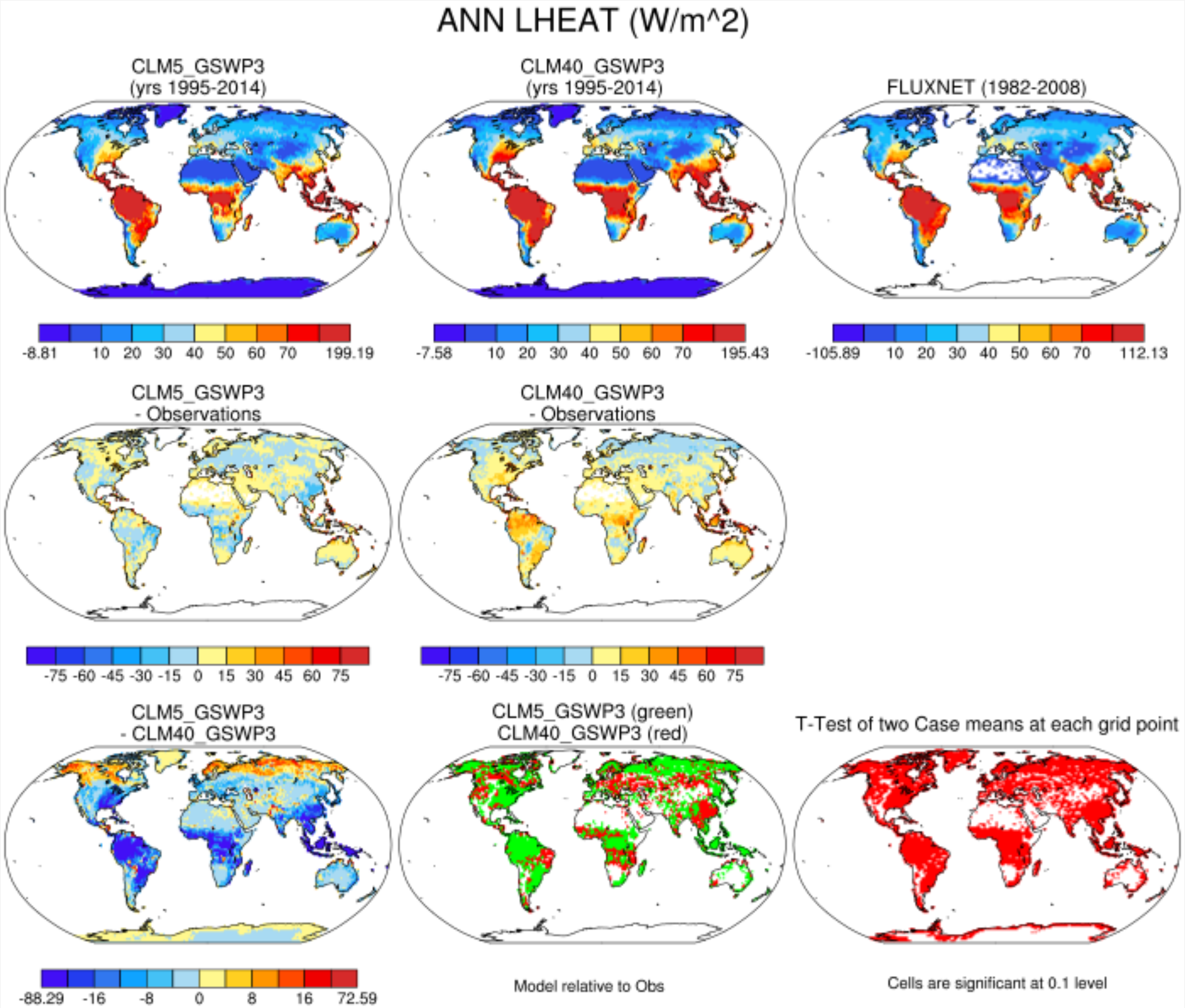
1 Line plots of annual trends in energy balance, soil water/ice and temperature, runoff, snow water/ice, photosynthesis
2 Horizontal contour plots of DJF, MAM, JJA, SON, and ANN means
3 Line plots of monthly climatology: regional air temperature, precipitation, runoff, snow depth, radiative fluxes, and turbulent fluxes
4 *(Inactive)* Vertical profiles at selected land raobs stations
5 Tables of annual means
6 Line plots of annual trends in regional soil water/ice and temperature, runoff, snow water/ice, photosynthesis
7 Line plots, tables, and maps of RTM river flow and discharge to oceans
8 *(Inactive)* Line and contour plots of Ocean/Land/Atmosphere CO_2 exchange
9 Contour plots and statistics for precipitation and temperature. Statistics include DJF, JJA, and ANN biases, and RMSE, correlation and standard deviation of the annual cycle relative to observations
10 Horizontal contour plots of DJF, MAM, JJA, SON, and ANN means, zoomed in on the Greenland ice sheet
11 Horizontal contour plots of DJF, MAM, JJA, SON, and ANN means, zoomed in on the Antarctic ice sheet

LMWG Diagnostics Package Output

Set 1: Global Average/Sum Trends

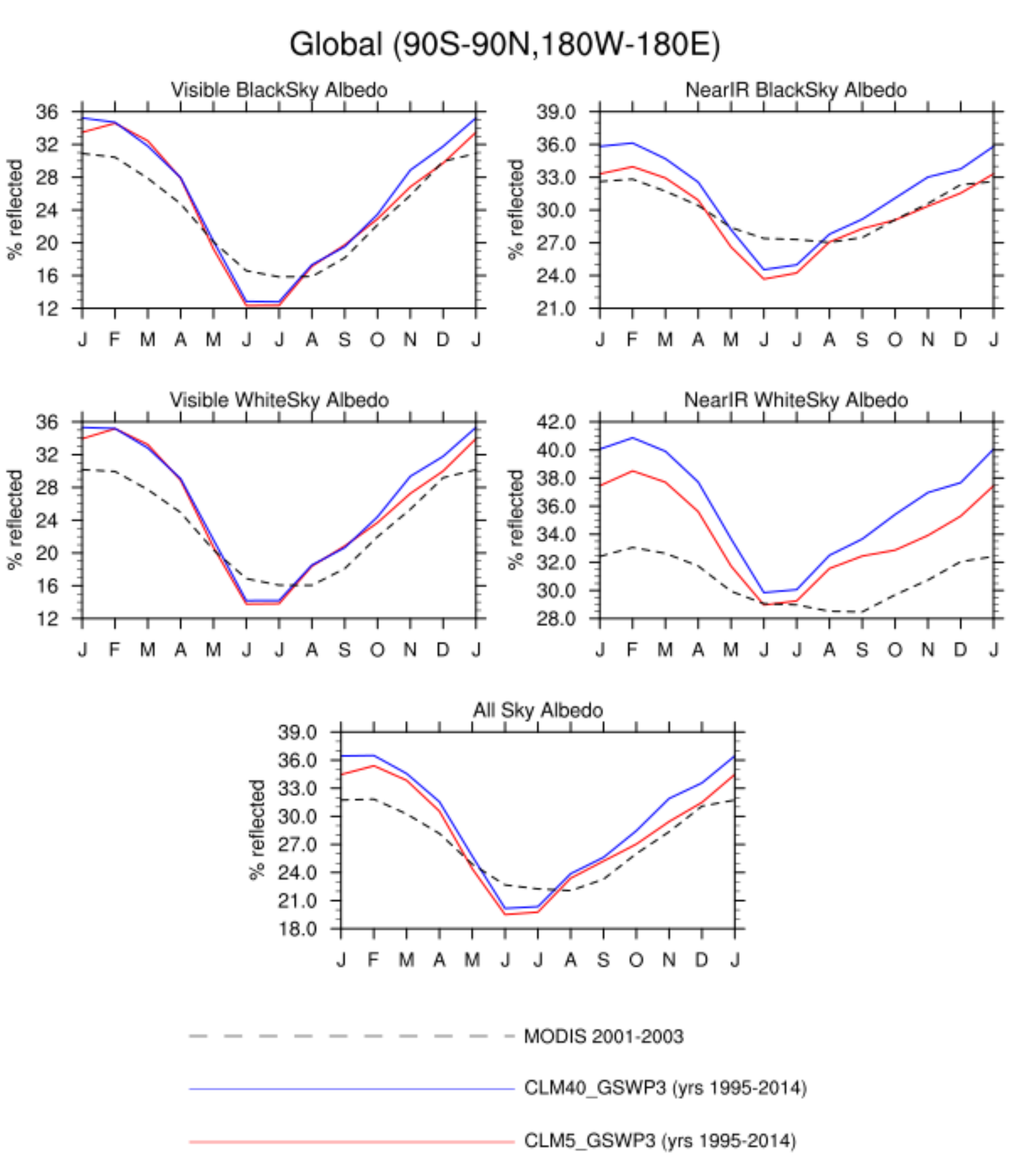


Set 2: Seasonal Mean Contours

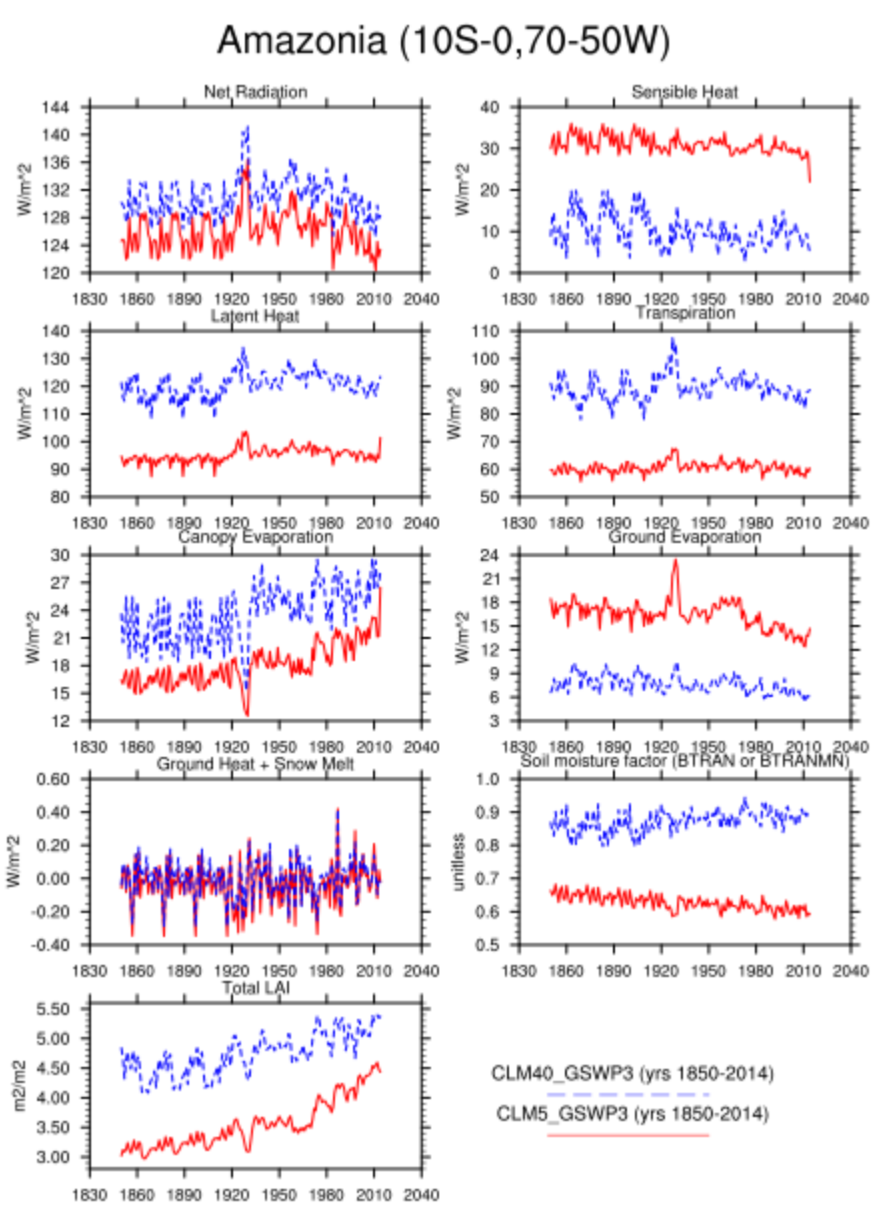


LMWG Diagnostics Package Output

Set 3: Regional Monthly Climatology

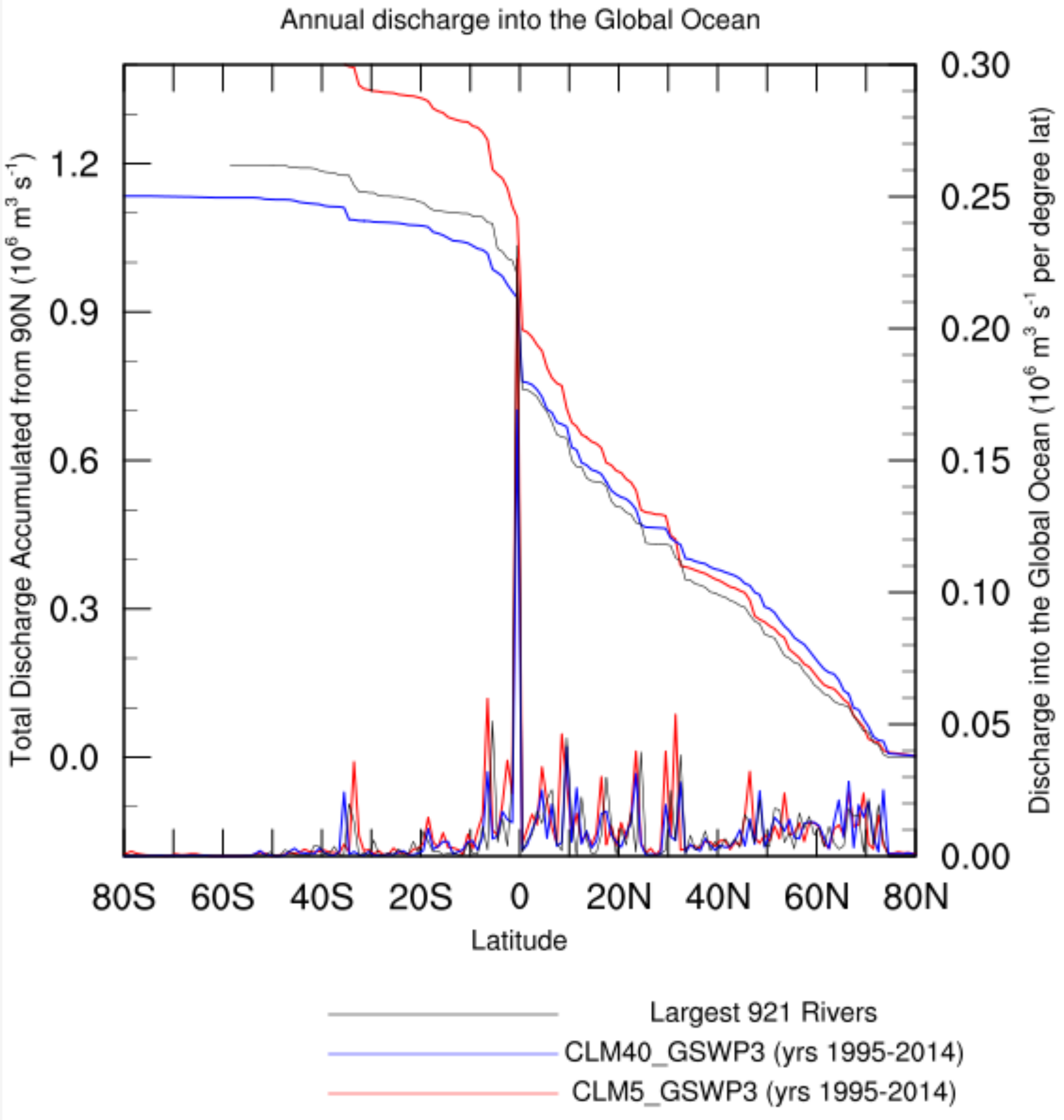


Set 6: Regional Annual Trends

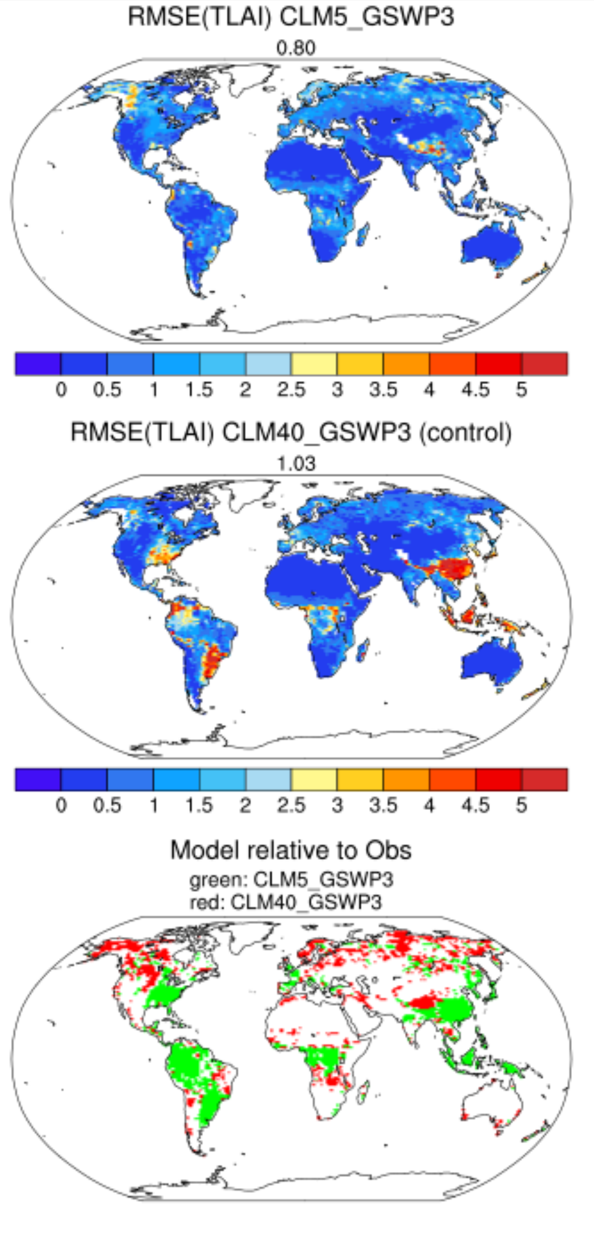


LMWG Diagnostics Package Output

Set 7: RTM/MOSART River Flow & Ocean Discharge

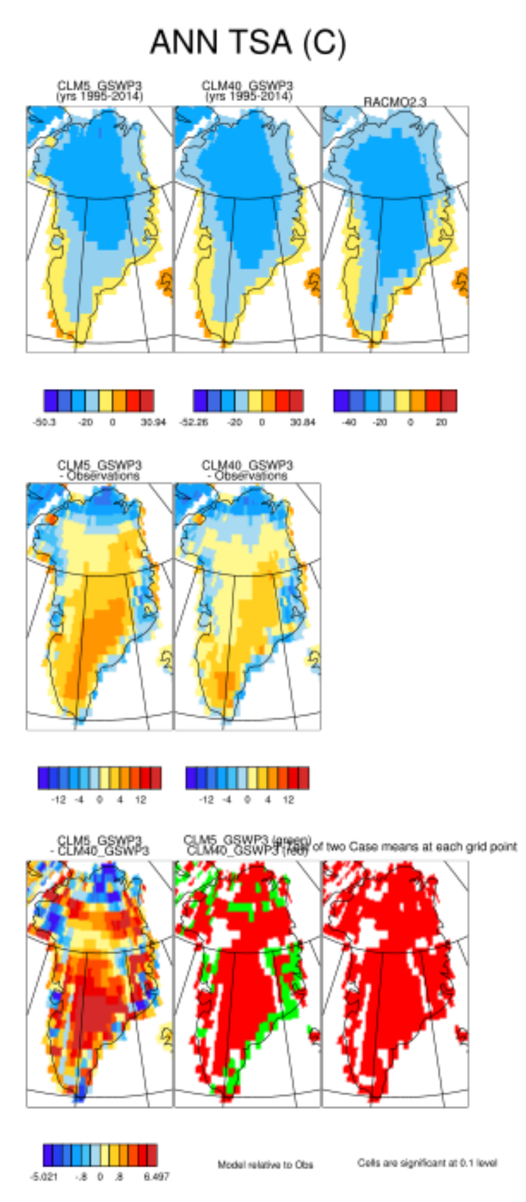


Set 9: Evaluation

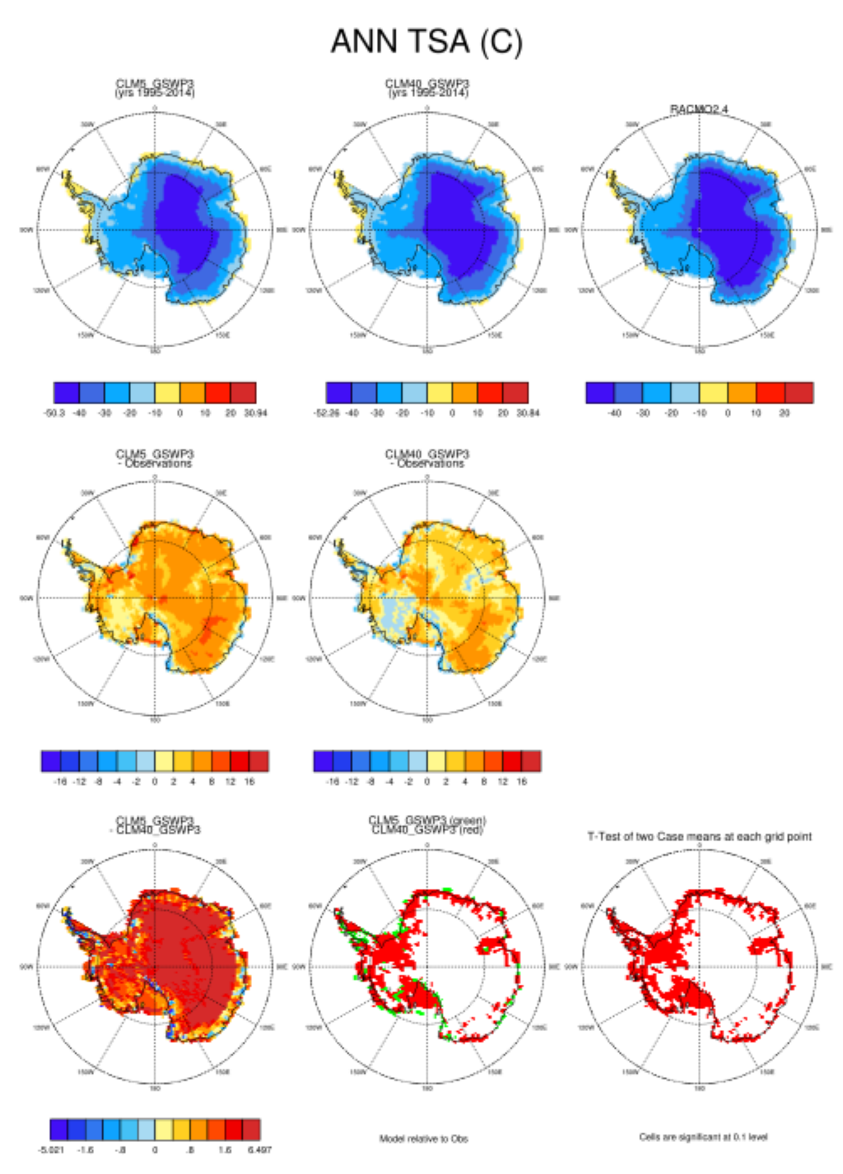


LMWG Diagnostics Package Output

Set 10: Seasonal Mean Contours - Greenland

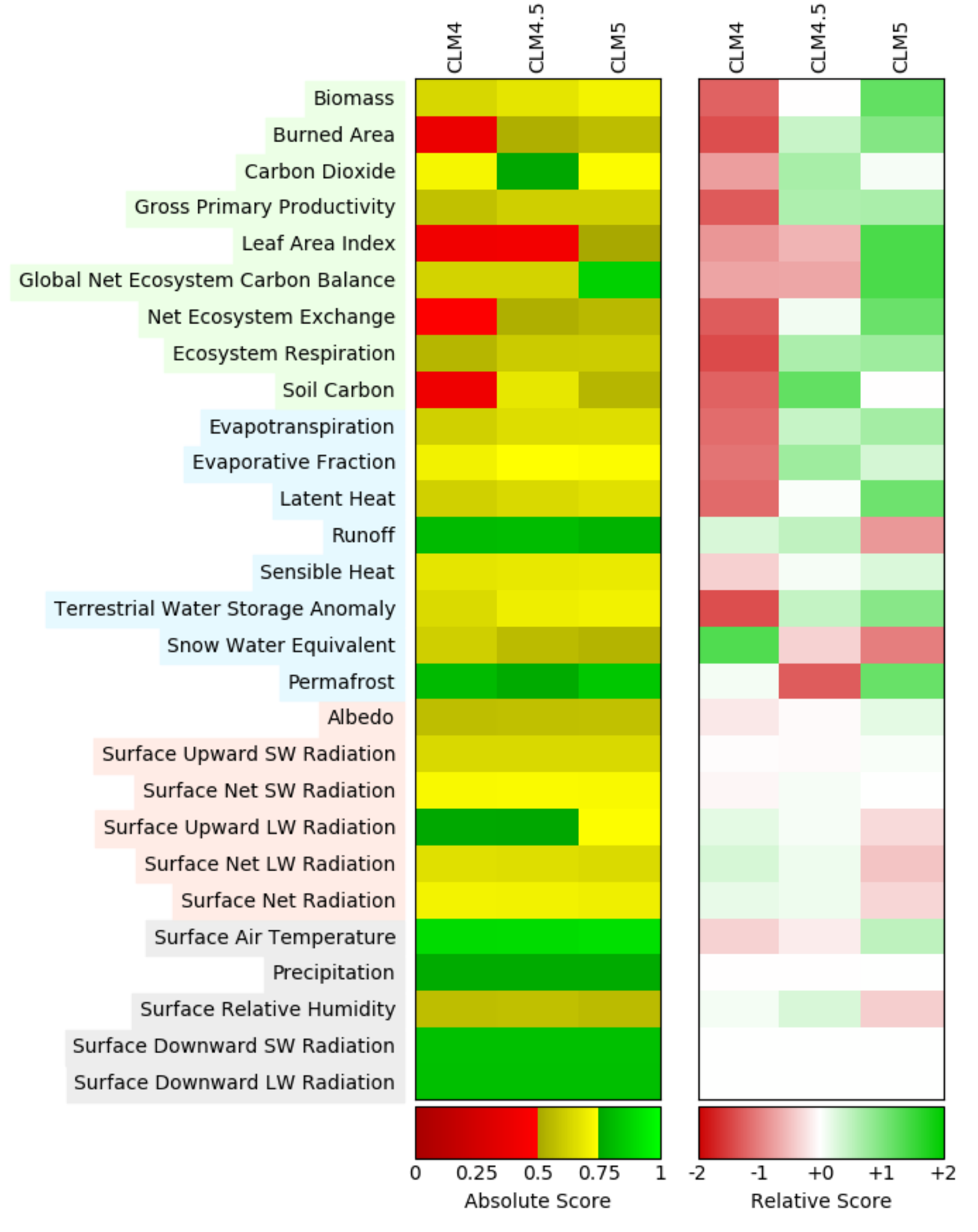


Set 11: Seasonal Mean Contours - Antarctica



courtesy: Jan Lenaerts

ILAMB Output



Albedo / CERES / 2000-2012 / global

	Download Data	Period Mean (original grids) [1]	Model Period Mean (intersection) [1]	Model Period Mean (complement) [1]	Benchmark Period Mean (intersection) [1]	Benchmark Period Mean (complement) [1]	Bias [1]	RMSE [1]	Phase Shift [months]	Bias Score [1]	RMSE Score [1]	Seasonal Cycle Score [1]	Spatial Distribution Score [1]	Interannual Variability Score [1]	Overall Score [1]
Benchmark	[-]	0.147													
CLM4	[-]	0.276	0.276	0.00	0.262	0.0914	0.0138	0.0637	2.24	0.49	0.39	0.65	0.99	0.63	0.59
CLM4.5	[-]	0.271	0.271	0.00	0.262	0.0914	0.00878	0.0611	2.30	0.51	0.38	0.64	0.99	0.65	0.59
CLM5	[-]	0.268	0.268	0.00	0.262	0.0914	0.00606	0.0584	2.21	0.51	0.40	0.65	0.99	0.64	0.60

ILAMB code: https://bitbucket.org/ncollier/ilamb
ILAMB docs/tutorial: https://www.ilamb.org/doc
Collier et al. 2018, JAMES.

Thank You!

Questions?

Getting Started With Land Diagnostics

To set up your environment for today's lab:

1) Log in to cheyenne using either your yubikey or Duo two-factor authentication

2) **For tcsh users:** You may have a .tcshrc file already present in your home directory. If you do not, please copy over the following file:

   cp /glade/p/cgd/tss/CTSM_tutorial2019/CESM_Postprocessing_Tool/tcshrc ~/.tcshrc

   change to your home directory and source the file:

   cd; source .tcshrc

   If you have an existing .tcshrc file and do not wish to overwrite it please copy the **contents** of the
   /glade/p/cgd/tss/CTSM_tutorial2019/CESM_Postprocessing_Tool/tcshrc
   file to your .tcshrc file.

   For bash users: Add the following to your .profile:

   export POSTPROCESS_PATH=/glade/p/cesm/postprocessing
   alias cesm_pp_activate='. $POSTPROCESS_PATH/cesm-env2/bin/activate‘
   PROJECT=UCGD0004;export PROJECT

Getting Started with Land Diagnostics

3) Copy over and rename the following file (hluresfile - sets NCL defaults):

```
cp /glade/p/cgd/tss/CTSM_tutorial2019/CESM_Postprocessing_Tool/hluresfile ~/.hluresfile
```

Running the Land Diagnostics

Note that the general CESM (component) diagnostics instructions are located here:
https://github.com/NCAR/CESM_postprocessing/wiki/cheyenne-and-DAV-quick-start-guide
Customized instructions for the tutorial are given over the next few slides. You will need to change all settings that are encased in < >.

1) Set up your python environment:
   cesm_pp_activate

2) Create a directory to house the CESM postprocessing code:
   mkdir /glade/scratch/<logname>/cesm-postprocess

Running the Land Diagnostics

3) Decide which simulation you will run the diagnostics on, either your own run or the run at:
/glade/p/cgd/tss/CTSM_tutorial2019/CESM_Postprocessing_Tool/I1850CLM50_001
Then run create_postprocess to set up your post-processing directory, and cd to that directory as follows:

```
create_postprocess --caseroot /glade/scratch/<logname>/cesm-postprocess/<model-run>
cd /glade/scratch/<logname>/cesm-postprocess /<model-run>
```

For instance, if you are running running diagnostics on your I1850CLM50_001 simulation:

```
create_postprocess --caseroot /glade/scratch/<logname>/cesm-postprocess/I1850CLM50
cd /glade/scratch/<logname>/cesm-postprocess /I1850CLM50_001
```

Reminder: Your model data location: **/glade/scratch/<logname>/archive/<model-run>**
Note the "--" syntax (two dashes not separated by a space)
Note that the simulation example used here is an "SP' simulation

Running the Land Diagnostics

4) You will now set options in various .xml files in preparation for running. You can do the modifications by hand, or you can do them by using the pp_config command. It is *highly recommended* that you use the pp_config command as that will check that your changed settings are valid.

   The first file that needs modification is env_postprocess.xml. (Note that if you alternatively set up your cesm-processing directory (step 3) within the archive directory of your model run, you can skip this step as everything should be set automatically.)

Set the location of the model data:

./pp_config --set DOUT_S_ROOT=<full path of model run archive path to be analyzed>
(Example: ./pp_config --set DOUT_S_ROOT=/glade/scratch/<logname>/archive/I1850CLM50)

Tell the diagnostics what kind of grid to expect. Our tutorial simulations use 1.9x2.5:

./pp_config --set LND_GRID=1.9x2.5

Running the Land Diagnostics

The land diagnostics need at least 14 months to run and you can only specify complete years. The steps to run the land diagnostics are as follows:

1) The following commands edit settings in env_diags_lnd.xml.

```
./pp_config --set LNDDIAG_OUTPUT_ROOT_PATH=/glade/scratch/<logname>/diagnostics-output/lnd
./pp_config --set LNDDIAG_clim_first_yr_1=<set to first year to be analyzed>
./pp_config --set LNDDIAG_clim_num_yrs_1=<set to # of years to be analyzed>
./pp_config --set LNDDIAG_trends_first_yr_1=<set to first year to be analyzed>
./pp_config --set LNDDIAG_trends_num_yrs_1=<set to # of years to be analyzed>
```

For this particular run:

```
./pp_config --set LNDDIAG_OUTPUT_ROOT_PATH=/glade/scratch/<logname>/diagnostics-output/lnd
./pp_config --set LNDDIAG_clim_first_yr_1=2
./pp_config --set LNDDIAG_clim_num_yrs_1=4
./pp_config --set LNDDIAG_trends_first_yr_1=1
./pp_config --set LNDDIAG_trends_num_yrs_1=5
```

Since this particular run is an “SP” simulation, we need to turn off the “CN” flag

```
./pp_config --set LNDDIAG_CN=0
```

Running the Land Diagnostics

2) You will need to change the queue and account number in lnd_averages and lnd_diagnostics

   #PBS -q R4231039
   #PBS -A UCGD0004

3) Before the land diagnostics can be run, annual, seasonal, and monthly climatologies must be calculated and written to netCDF files. To run the land averages script:

   qsub lnd_averages

   To monitor the status of your submission you can type qstat. You can check progress by checking the newest log file in logs/. If in a log file you notice that things have gone wrong, you can stop your submission by typing qdel <Job ID retrieved from qstat>

Running the Land Diagnostics

4) Once the averages have successfully completed (check the end of the newest log file), you can submit the diagnostics script:

```
qsub lnd_diagnostics
```

5) Again monitor the status of your submission by checking the newest log file in the logs/ directory. Do not be concerned by various error messages (like convert error messages) from individual scripts in the log files. If the submission completed successfully the log file will end with "Successfully completed generating land diagnostics".

6) Once the diagnostics are complete, cd to the location of the diagnostics:

```
cd /glade/scratch/<logname>/diagnostics-output/lnd/diag/<model-run>-obs.<yr1>_<yr2>
```

and open the setsIndex.html in firefox to examine the output:

```
firefox setsIndex.html &
```

Porting

The CESM post-processing suite is currently *only supported* on NCAR machines due to limited resources. The python software-stack uses vanilla python 2.7 (not anaconda) and the virtualenv package:

https://help.dreamhost.com/hc/en-us/articles/215489338-Installing-and-using-virtualenv-with-Python-2

Once you have a virtualenv, then you can install additional python package dependencies. Some of the packages require underlying C library builds, such as netCDF4-python and PyNIO, so you will need to work with your systems administrator to make sure these libraries are built correctly.

NCAR has chosen to move away from the module system for managing and loading python packages in favor of virtualenvs that can be cloned into any user's environment and once activated, contain all the necessary python packages required by the CESM post-processing suite of tools. This change is included in the most recent Github tagged versions of the CESM post-processing suite, v1.0.z.