

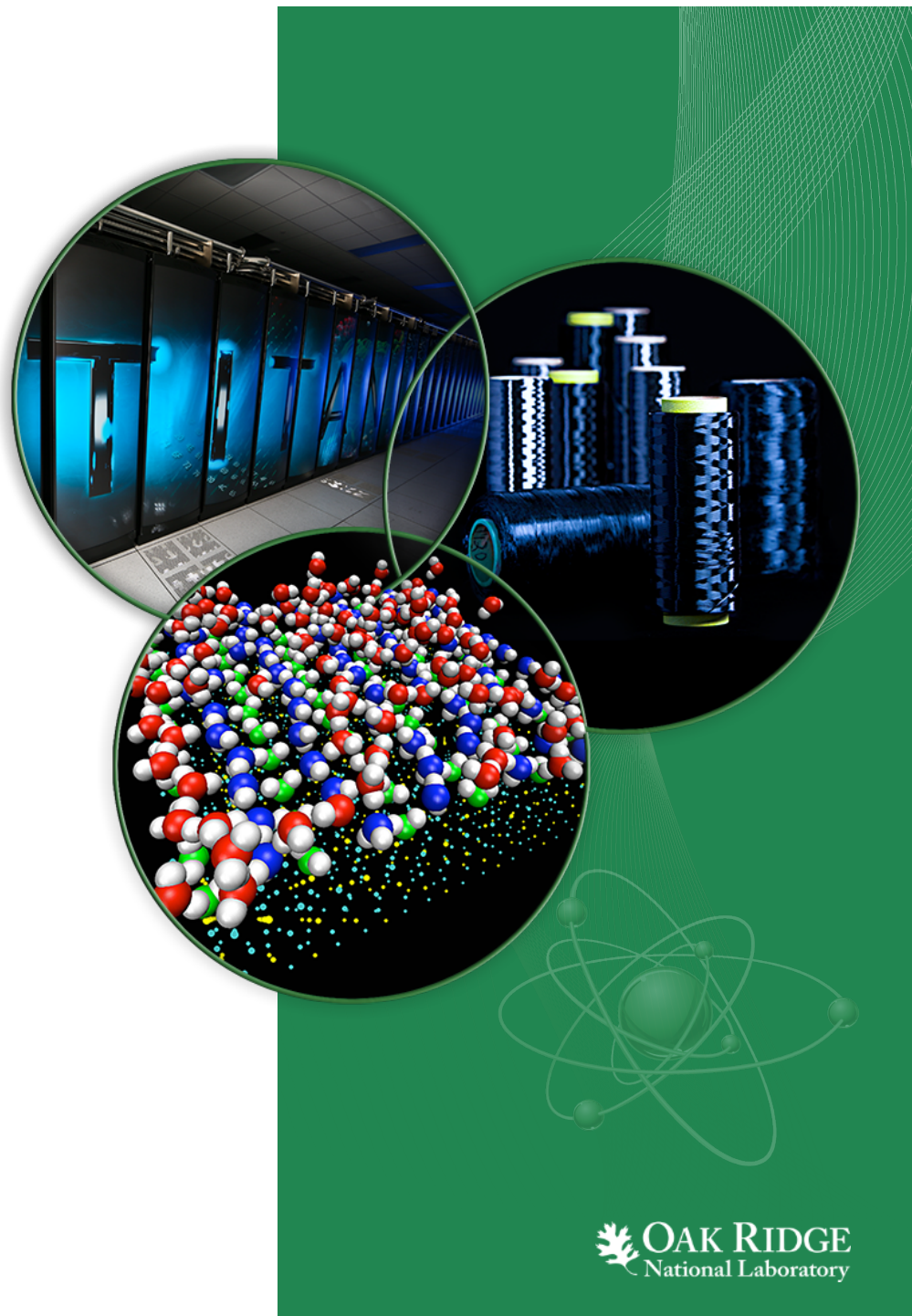
Diagnostics and Exploratory Analysis Infrastructure for ACME Workflow

ORNL: **Brian Smith,**
John Harney, Brian
Jewell

LLNL: Jeffrey Painter,
James McEnerney,

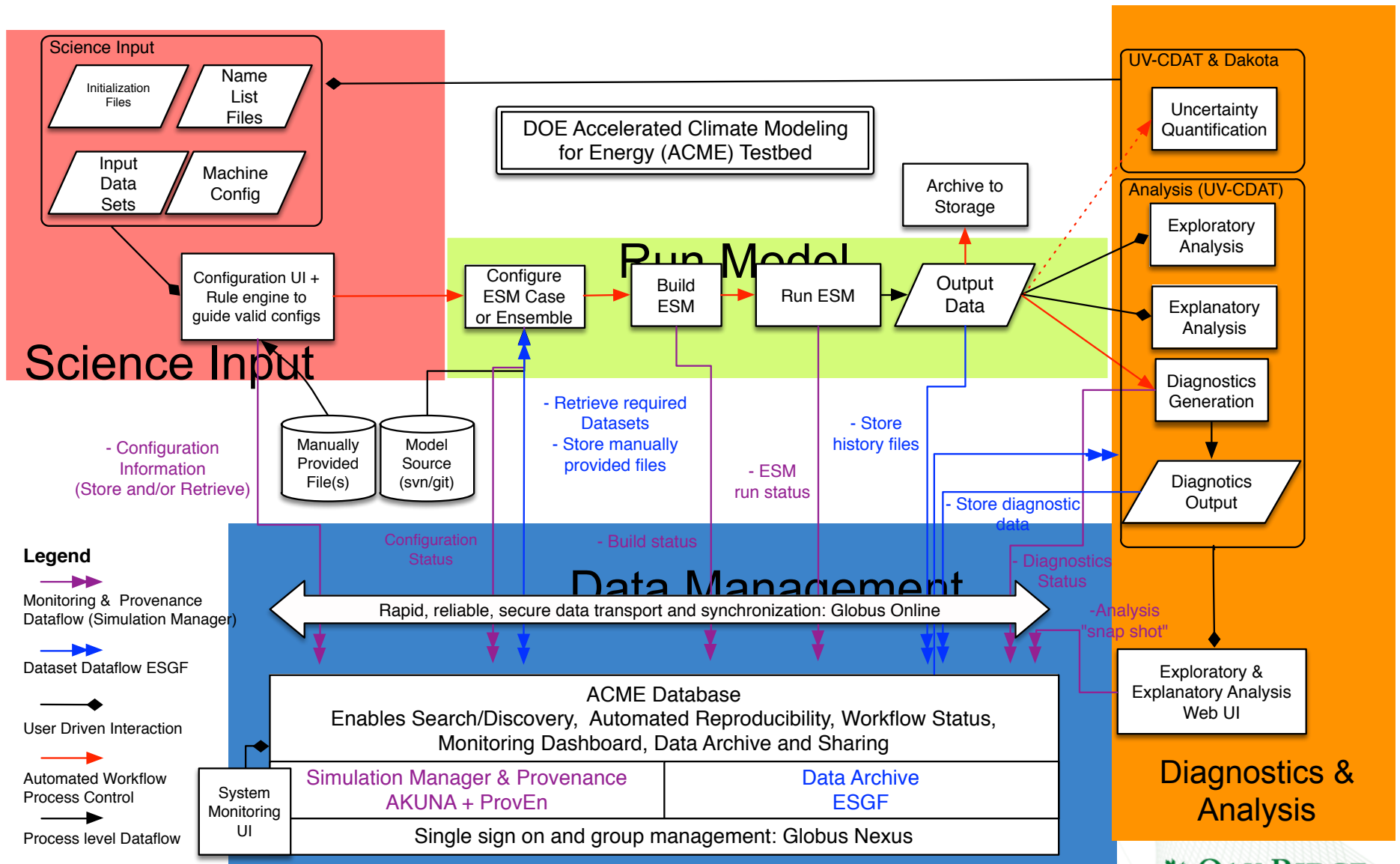


ORNL is managed by UT-Battelle
for the US Department of Energy

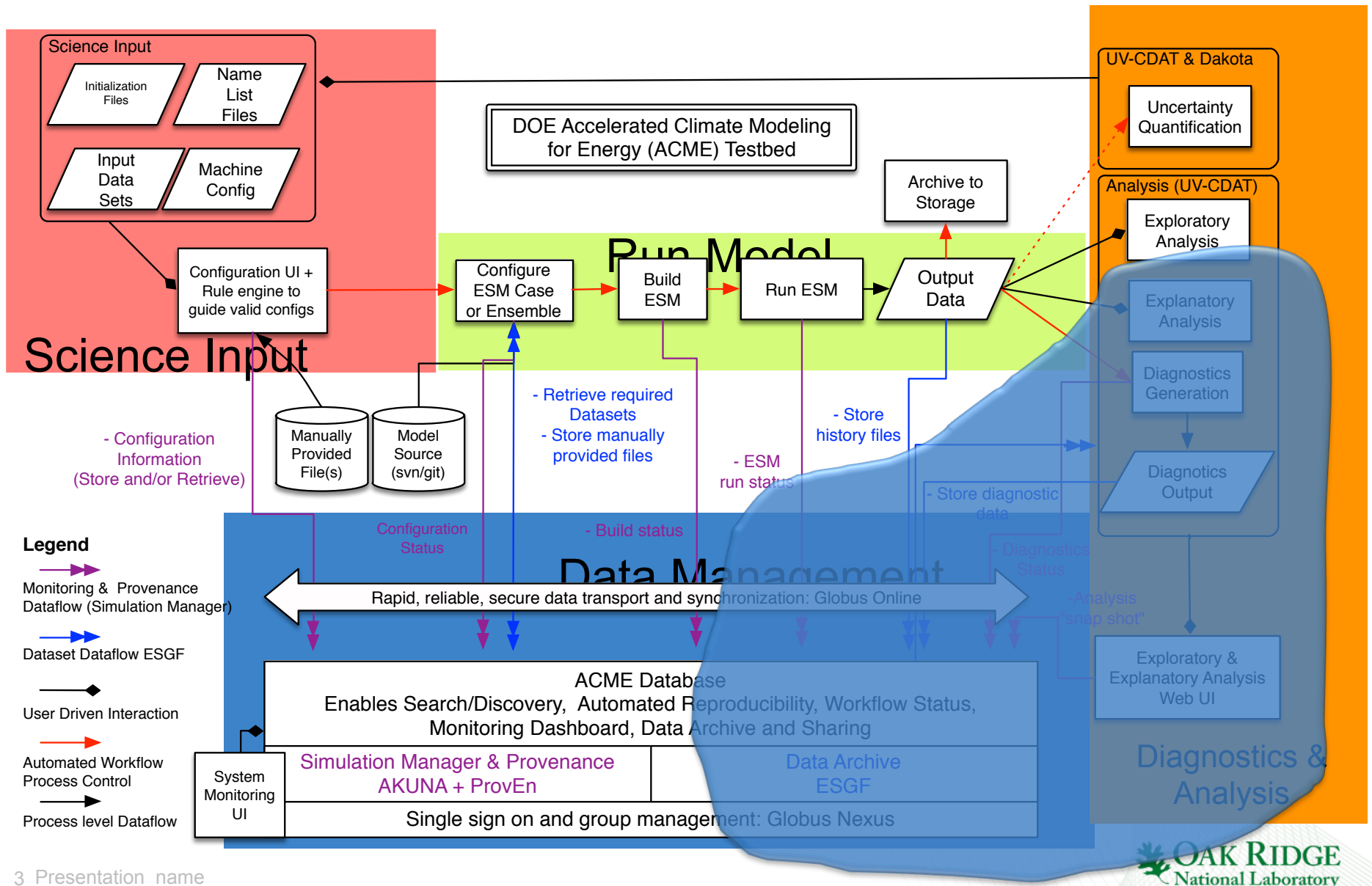


 **OAK RIDGE**
National Laboratory

ACME End-to-end workflow



ACME End-to-end workflow



Workflow Infrastructure Support

- Several command line tools have been developed to support workflow infrastructure tasks for diagnostics and EA
 - Climatology generation
 - Data movement and ESGF publishing
 - Individual diagnostics generation
 - Diagnostics “collections” generation (e.g. AMWG, “Tier 1A”)
 - Visualization
 - Exploratory analysis support
- Primarily Python scripts utilizing CDAT

Climatology.py

- Used to generate climatology files
 - Seasonal, monthly, and annual temporal averaging of model data
- Uses CDAT to do the averaging and NetCDF data management
- Produces CF-compliant output files
- Generates subset of climatology files with subset of variables per user specifications

Diags.py

- Generates individual diagnostics plots.
- Command line tool
 - Takes path(s), package type, plot types, variables, seasons, regions, etc
 - Produces PNGs, NetCDF summary files, JSON summary files
- Typically run against climatology files

Metadiags.py

- Wrapper script around the diags.py script
- Generates a series of plots in “collections”
- Controlled by single input dictionary file
- Tries to group plots to help with IO caching effects
- Creates opportunities for exploiting trivial parallelism based on “collection” and variable within collection.

Collections Dictionary

- Used by metadiags to generate diagnostics plots
- Also used by “classic viewer” to organize diagnostics plots
- Users define “collections” which are lists of variables, the plot types associated with them, which observation sets to use, and any other parameters (e.g. seasons, regions, etc) needed
- Initial work was in recreating NCAR AMWG and LMWG diagnostics collections
- Expanding to support new collections in a “scientist-friendly” way

File format (proposed)

```
# *** Collection so (southern ocean) ***
```

```
collection['so'] = {}
```

```
collection['so']['mixed_plots'] = True
```

```
collection['so']['mixed_regions'] = True
```

```
collection['so']['mixed_seasons'] = True
```

```
collection['so']['desc'] = 'Tier 1B Diagnostics (Southern Ocean)'
```

```
collection['so']['preamble'] = " # Web page text
```

```
collection['so']['seasons'] = ['ANN']
```

```
collection['so']['regions'] = ['Global']
```

```
collection['so']['packages'] = ['AMWG']
```

```
collection['so']['SHFLX'] = {'plottype': '3', 'obs': ['LARYEA_1'], 'regions': ['S._Hemisphere_Land']}
```

```
collection['so']['QFLX'] = {'plottype': '5', 'obs': ['LARYEA_1'], 'regions': ['S._Hemisphere_Land']}
```

```
collection['so']['FSNS'] = {'plottype': '7', 'obs': ['LARYEA_1']}
```

```
collection['so']['T'] = {'plottype': '4', 'obs': ['ERA1_1'], 'seasons': ['DJF']}
```

```
collection['so']['SURF_WIND'] = {'plottype': '6', 'obs': ['NCEP_1'], 'regions': ['S._Hemisphere_Land']}
```

```
collection['so']['CLDTOT'] = {'plottype': '9', 'obs': ['CLOUDSAT_1'], 'regions': ['S._Hemisphere_Land']}
```

Classic View

Dataset:

cam5se

Package:

amwg

Variables:

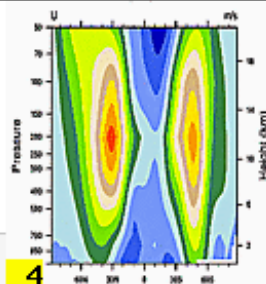
17 selected

Times:

17 selected

Plot Dataset

[Back to Atm Home](#)



cam5se and OBS data

DIAG Set4 Vertical contour plots of DJF, JJA and ANN zonal means

ECMWF Reanalysis 1979-93

DJF JJA ANN

U	Zonal Wind	plot	plot	plot
SHUM	Specific humidity	plot	plot	plot
RELHUM	Relative humidity	plot	plot	plot
OMEGA	Pressure vertical velocity	plot	plot	plot
T	Temperature	plot	plot	plot

JRA25 Reanalysis 1979-04

DJF JJA ANN

U	Zonal Wind	plot	plot	plot
SHUM	Specific humidity	plot	plot	plot
OMEGA	Pressure vertical velocity	plot	plot	plot
T	Temperature	plot	plot	plot

AIRS IR Sounder 2002-06

DJF JJA ANN

SHUM	Specific humidity	plot	plot	plot
RELHUM	Relative humidity	plot	plot	plot
T	Temperature	plot	plot	plot

NCEP Reanalysis 1979-98

DJF JJA ANN

U	Zonal Wind	plot	plot	plot
SHUM	Specific humidity	plot	plot	plot
RELHUM	Relative humidity	plot	plot	plot

Exploratory Analytics Infrastructure Support

- Goal: Provide mechanisms for the EA tools to process and display climate data quickly and efficiently
 - Side benefit: “API” usable by other standalone scripts
- Makes calls to diagnostics framework as needed or calls to CDAT directly
- Integrated with django EA applications
- Designed to make use of cached and/or pre-generated data whenever possible

Exploratory Analytics Infrastructure Support (cont)

- Interfaces to generate diagnostics plots and trees of available plots (tree viewer), temporal/spatial averaging for animations/region selection (geospatial view), and organization/creation of web pages (classic view)
- Interfaces to get information about a dataset (variable list, other metadata) and return JSON objects related to dataset
- More details on EA in John Harney's talk which follows this one

Diagnostics Tree Viewer

UV-CDAT:EA

Home

Geo

Diagnostics Tree

Heatmap

Logout

Create Tree

Bookmarks ▾

Saved Images ▾

Diagnostics Tree

Dataset:

tropics_warming_th_q_co2

Package:

lmwg

Variables:

6 selected

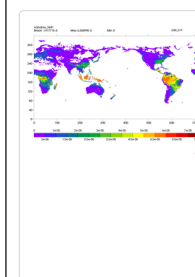
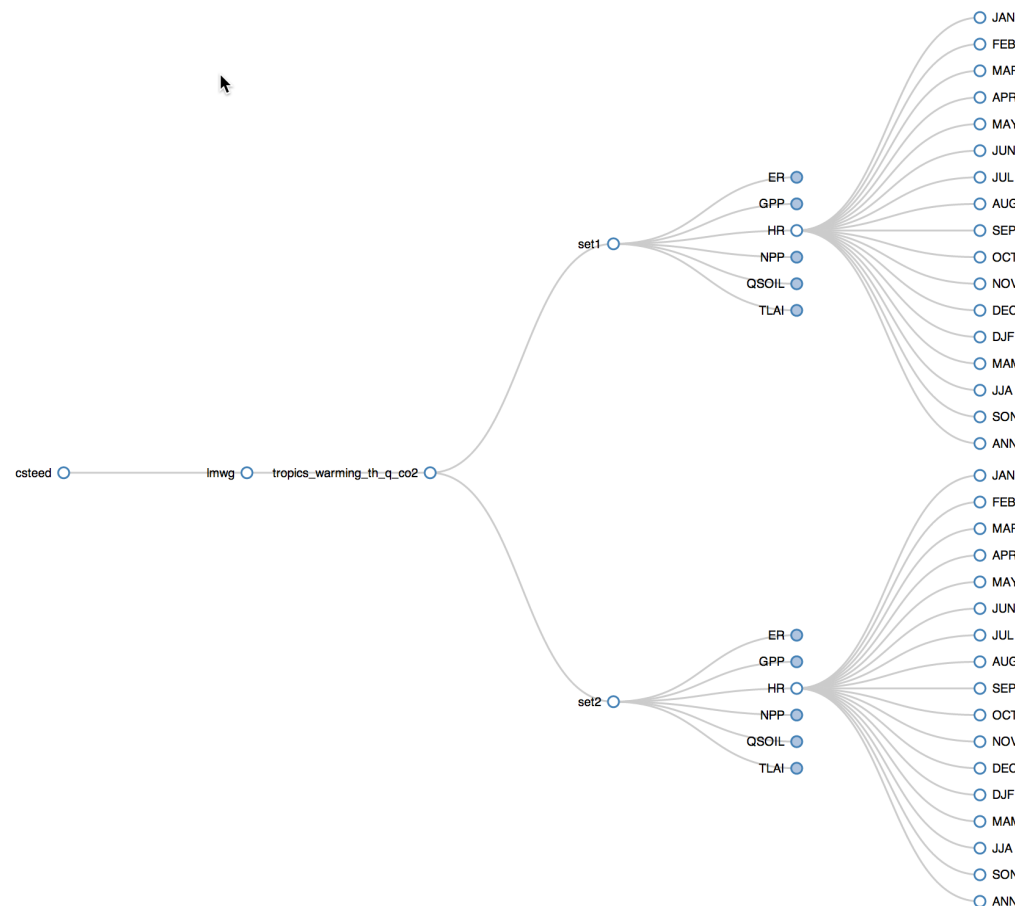
Times:

17 selected

tropics

Save

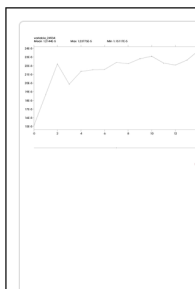
Tree Name: tropics



land_lmwg_set2_JAN_HR.png

Remove

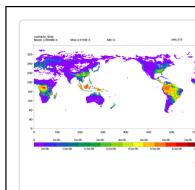
View



land_lmwg_set1_AUG_HR.png

Remove

View



Future Work

- Finish implementing new “collections” format
- Parallelization of climatology.py and metadiags.py
- Other speedup in climatology.py and (meta)diags.py
- More functions needed by EA
 - Correlation calculations (heatmap)
 - Document and create an actual API for the interfaces?

Acknowledgements

This research is sponsored by the U.S. Department of Energy, Office of Science, Biological and Environmental Research (BER) program and performed at Oak Ridge National Laboratory (ORNL). The work was performed at the Oak Ridge National Laboratory, which is managed by UTBattelle, LLC, for the Department of Energy, under Contract No. DEAC0500OR22725. This research used resources of the Center for Computational Sciences at Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DEAC0500OR22725.

- Questions?