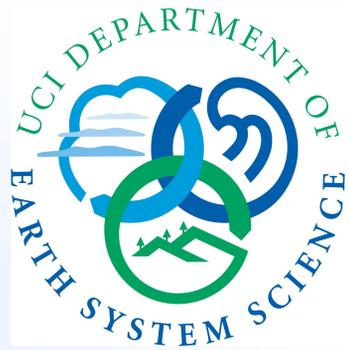# Simplifying and accelerating model evaluation by NASA satellite data

Charlie Zender, Pedro Vicente, and Wenshan Wang

Department of Earth System Science, University of California, Irvine

BACKGROUND

The fidelity of geoscientific model results is increasingly evaluated by comparison to products derived from NASA satellite measurements. The satellite data are archived in HDF-EOS format, which is now a superset of the netCDF format employed by most geoscientific models. Putting NASA-generated (HDF-EOS) data and model-generated (netCDF) data on a common grid, in the same format, for numerical comparison can be arduous because of data format incompatibilities. Many researchers desire a common toolkit for both HDF-EOS and netCDF data that would 1. simplify and accelerate the independent analysis of both data formats (HDF-EOS and netCDF), 2. exploit the strengths of netCDF's underlying HDF data format with easy-to-use netCDF tools, 3. ease evaluations of model predictions (in netCDF format) by NASA-generated data (in HDF-EOS format).

SIMPLIFY

Eliminate unnecessary complexity intercomparing multiple models and measurements. Current techniques require user-specified loops over ensemble members and scenarios to evaluate and attribute observed changes. Such analyses are increasingly common in comparing CMIP5 models to NASA observations, e.g.,

a) NA — Annual Snow Cover (%) vs year, SRESA1B 720ppm Stabilization Scenario

California Mean Surface Air Temperature Change (K) vs year

MATERIALS

ACCESS Project

Our ACCESS project consists of developing software to analyze and manipulate data stored in the EOS-standard HDF format. We are adding HDF5 group support for the netCDF Operators (NCO[12]), and creating NCO wrappers for HDF EOS files. The NCO homepage is http://nco.sf.net.

netCDF Operators (NCO) Software Stack

ACCELERATE

Vectorize Analysis Using HDF/netCDF4 Groups

```
# Old Analysis: Loop over ensemble members
for run in '1 2 3 4 5 6 7 8'; do
  ncdiff CCSM_${run}.nc MODIS.nc CCSM_${run}_MODIS.nc
done

# New Analysis: Put ensemble members in groups
ncecat CCSM_*.nc CCSM_all.nc
ncdiff CCSM_all.nc MODIS.nc CCSM_minus_MODIS.nc

# Old Analysis: Loop over ensemble members & models
for model in 'CESM GISS ECHAM ...'; do
    for run in '1 2 3 4 5 6 7 8'; do
        ncdiff ${model}_${run}.nc MODIS.nc \
                 ${model}_${run}_minus_MODIS.nc
    done
done

# New Analysis: Combine ensembles into groups
ncecat CCSM_all.nc GISS_all.nc ECHAM_all.nc ... CMIP5.nc
ncdiff CMIP5_all.nc MODIS.A2012.nc CMIP5_minus_MODIS.nc
```

File-level Analysis of Native HDF-EOS Formats

```
# NCO wrappers
ncra MOD10CM.A2007*.nc   MOD10CM_2007_avg.nc  # netCDF
hera MOD10CM.A2007*.hdf MOD10CM_2007_avg.hdf # HDF-EOS
```

MILESTONES

M2: Robust installs MacOS, RHEL, Windows 07/12
M4: Single-level hierarchies in `ncks` operator 10/12
M5: Nested hierarchies in `ncks`   01/13
M7: MODIS/CERES snow/albedo scripts   01/13
M8: Nested hierarchies in `ncra` operator   04/13
M10: Verify CMIP5 hierarchical analyses   07/13
M11: Nested hierarchies remaining operators 07/13
M12: Wrappers for MODIS/CERES HDF data   10/13
M15: Present simplification at AGU ESSI   01/14
M16: Wrappers for other (MISR?) HDF data 01/14
M17: Consolidated NCO/HDF release   04/14

STATUS

Accomplished

Group subsetting, hyperslabbing, movement, aggregation, and flattening all work:

```
# Extract variables from groups g1, g2
ncks -g g1,g2 -v v1,v2 in.nc out.nc

# Hyperslab tropical latitudes of g1
ncks -g g1 -d lat,-30.,30. in.nc out.nc

# Move data from g1 to g2
ncks -G g2:-1 in.nc out.nc

# Put data in groups 1985, 1986
ncecat —gag 1985.nc 1986.nc out.nc
```

Next Steps

Address namespace conflicts, renaming, and regular expressions, and then arithmetic.

REFERENCES

1. Zender, C. S. (2008), Analysis of Self-describing Gridded Geoscience Data with netCDF Operators (NCO), Environ. Modell. Softw., 23(10), 1338-1342, doi:10.1016/j.envsoft.2008.03.004.
2. Zender, C. S., and H. J. Mangalam (2007), Scaling Properties of Common Statistical Operators for Gridded Datasets, Int. J. High Perform. Comput. Appl., 21(4), 485-498, doi:10.1177/1094342007083802.