Regrid Curvilinear, Rectangular, and Unstructured Data (CRUD) with **ncremap**, a new netCDF Operator

Charlie Zender <zender@uci.edu>, Departments of Earth System Science & Computer Science, UC Irvine

Once upon a time Earth Science Data spatial grids were rectangular in latitude-longitude. Now researchers increasingly analyze Curvilinear, Rectangular, and Unstructured Data (CRUD) on different model and measurement grids. Tools to generate remapping weights (e.g., **ESMF_RegridWeightGen** and **TempestRemap**) are accurate, yet researchers still struggle to remap data because datasets are not fully annotated, the tools are intricate and inflexible, and this makes the regridding process time-consuming and error-prone. We introduce a netCDF Operator, **ncremap**, that automatically remaps CRUD and requires minimal, if any, user-intervention or regridding expertise.

Invocation Modes:

1. Free-will: Infer source and destination grids to generate mapfile
```
ncremap -s src.nc -d dst.nc
```

2. Old grid: Uses known-good grid(s) to generate mapfile then regrid)
```
ncremap -g grd.nc -s src.nc -d dst.nc
```

3. New grid: Generate source-grid from ncks parameter string
```
ncremap -G '--rgr latlon=40,40 --rgr snwe=30.0,70.0,-130.0,-90.0'
```

4. Pre-Destination: Apply supplied mapfile to all input files (fastest)
```
ncremap -s src.nc -m map.nc
```

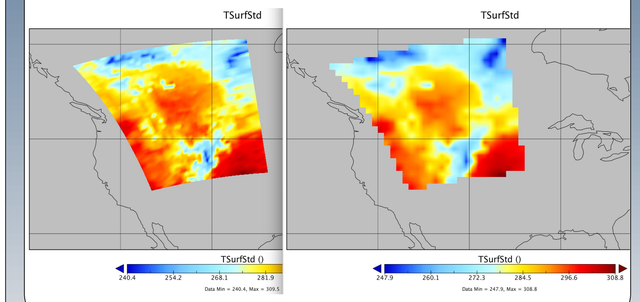
Procedure to Infer Grid Properties:

ncremap first looks in any (optionally) supplied gridfile or mapfile, and supplements this information (if any) with grid details sometimes provided in CF metadata (e.g., "bounds" variables) from the datafiles themselves. **ncremap** next tests for known rectangular grid types (equiangular, FV, offset, Gaussian) and supplements metadata with exact information derived from inferred gridtype, if any. This fails for curvilinear and unstructured grids, for which **ncremap** must extrapolate grid properties from cell-center locations.

Regional Swath Data

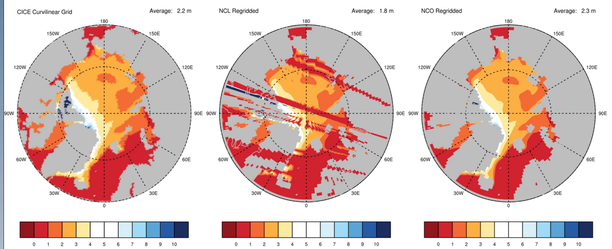
Regrid NASA AIRS Level 2 Swath Data in raw HDF4 format from regional curvilinear 45x30 source grid to equiangular 1x1 degree:

```
% ncremap -s AIRS.2014.10.01.202.L2.RetStd.v6.0.11.0.G14275134307.hdf -d dst_1x1.nc
```



Global Sea-Ice (CICE) Data

CESM simulated sea-ice thickness in June 2005



Native CICE grid has _FillValue in mask variable (tmask), which NCL function curvilinear_to_SCRIP() and ESMF do not understand. Unless user first manually sets _FillValue to 0, NCL generates incorrect grid, and regridding produces subtly biased results. **ncremap** handles missing values in masks without user intervention.

Examples

```
ncremap -s swath1.nc -d swath2.nc # Curvilinear → Curvilinear
ncremap -v temperature -s mpas_ocean.nc -d cam_se.nc # Unstructured → Unstructured
ncremap -s AIRS.2014.10.01.202.L2.RetStd.v6.0.11.0.G14275134307.hdf -d 1x1.nc # Curvilinear → Rectangular
ncremap -v StepTwoO3 -s OMI-Aura_L2-OMTO3_2015m0731t0034-o58727_v003-2015m0731t080836.he5 -d cam_fv.nc # Subsetting
ncremap -x TSurfStd_ct -s AIRS.2014.10.01.202.L2.RetStd.v6.0.11.0.G14275134307.hdf -d cam_se.nc # Extensive variables (beta)
```

Regridder Bake-Off

Three regridders (UV-CDAT, NCL, NCO) on CAM-SE unstructured grid datasets from ~1-13 GB on multicore nodes:

Table 2: Latest Regridders	UV-CDAT	NCL	NCO
map_ne30np4_fv129x256_aave.150418.nc	0m19s-0m27s	0m50s-1m30s	0m05s-0m06s
map_ne30np4_fv257x512_bilin.150418.nc	0m25s-0m33s	1m55s-2m00s	0m10s-0m11s
map_ne120np4_fv257x512_aave.150418.nc	1m10s-1m35s	4m50s-7m20s	0m45s-0m50s
map_ne120np4_fv801x1600_bilin.150418.nc	2m40s-4m00s	15m40s-41m00s	1m30s-1m40s

Parallelism

ncremap is threaded over variables with OpenMP and scales well up to 8-16 threads:

Table 3: Parallel Configurations	NCO (serial, nc3)	NCO (threads=8, nc3)	NCO (threads=8, nc4)	NCO (threads=16,nc4)
map_ne30np4_fv129x256_aave.150418.nc	0m20s-0m35s	0m09s-0m12s	0m06s-0m07s	0m05s-0m06s
map_ne30np4_fv257x512_bilin.150418.nc	1m30s-1m40s	0m32s-0m33s	0m12s-0m13s	0m10s-0m11s
map_ne120np4_fv257x512_aave.150418.nc	5m00s-7m30s	0m50s-1m20s	0m48s-0m50s	0m45s-0m50s
map_ne120np4_fv801x1600_bilin.150418.nc	14m50s-24m50s	4m10s-7m15s	2m00s-2m05s	1m30s-1m40s

Summary

ncremap improves previous regridding solutions in multiple ways:

1. Single command regrids Curvilinear, Rectangular, and Unstructured Data (CRUD)
2. Infers accurate grids from CRUD without user intervention
3. Threading across variables makes it fastest regridder tested

Features

1. Curvilinear, Rectilinear, and Unstructured grids
2. Global and regional source and destination grids
3. Generates Accurate Gaussian grid boundaries (only known regridder that does, uses Newton-Raphson technique)
4. Diagnoses convex polygon area for any grid boundaries (necessary, e.g., with ESMF bilinear interpolation maps)
5. Built-in database of dimension/coordinate names
6. Adds latitude-weights (rectangular grids only)
7. Adds bounds variables when possible (even if missing from mapfile)
8. Propagates input metadata to output (adds NCO provenance info)
9. Subsets variables using regular expressions
10. CF metadata annotation (bounds, axis, cell_area, cell_methods)
11. OpenMP threading (OMP_NUM_THREADS=8 works well)
12. Compatibility: all tested versions of ESMF, TempestRemap
13. Extensive variables (nascent support)

Smoke-tested

ncremap tested on notable CRUD including: AIRS, CAM-FV, CAM-SE, CERES, CICE, CMIP5, MPAS-O/I, OMI, POP, and WRF.

Vaporware

More extensive variable support
Multi-grid support (e.g., simultaneous cell center + edge regridding)
Automatic dimension permutation for non-trailing horizontal dimensions
Radius-of-influence interpolation algorithms
MPI-I/O via PIO2

Support

Contributions from P. Caldwell and W. Wang, Funding: DOE ACME DE-SC0012998, NASA ACCESS NNX14AH55A, NSF IIS-0431203.