New Methods to Combine, Regrid, and Split Climos

Charlie Zender and Henry Butowsky

Departments of Earth System Science and Computer Science, UC Irvine

ACME All-Hands Meeting
Potomac, MD
June 5-7, 2017

On Web
New Methods

1. Regrid and "renormalize" partial gridcells
2. Conservatively regrid sub-gridscale distributions
3. Combine existing climos without using raw data
4. Reshape/split multi-field output into timeseries
5. High-frequency climos
Naïve Regrid

MPAS default

Renormalize: None, 0%, 50%, 99%

Valid = 50%

Valid = 99%
Regrid and "renormalize" partial grid cells

Problem: Regridded fields appear non-physical in destination cells with incomplete source coverage (e.g., 50% coverage causes SST ~ 140 K).

Solution: Renormalize values with tunable threshold

> ncremap --rnr_thr=0.5 # Source coverage > 50%
> ncremap --rnr_thr=0.0 # Any coverage OK (MPAS)
New Methods

1. Regrid and "renormalize" partial gridcells
2. Conservatively regrid sub-gridscale distributions
3. Combine existing climos without using raw data
4. Reshape/split multi-field output into timeseries
5. High-frequency climos
Sub-Gridscale Remapping

Land Fraction

Sea-ice Fraction

Land Mask

Sea-ice Mask
Conservatively regrid sub-gridscale distributions

Problem: ALM/CICE archive all/some fields valid for spatiotemporally varying gridcell fractions. Naïve regrids are non-conservative, e.g., along coastlines.

Solution: Sub-gridscale (SGS) regridding conserves (area x value) integrals, re-computes binary masks

```bash
> ncremap --sgs_frc=landfrac --sgs_msk=landmask
> ncremap -P alm  # ALM convenience option
> ncremap --sgs_frc=aicen001 --sgs_msk=tmask \ --src_nrm=100  # CICE
```
New Methods

1. Regrid and "renormalize" partial grid cells
2. Conservatively regrid sub-gridscale distributions
3. Combine existing climos without using raw data
4. Reshape/split multi-field output into timeseries
5. High-frequency climos
Combine existing climos without accessing raw data

**Problem:** Long simulations take weeks and consume extensive disk space. Method needed to compute and re-combine climos incrementally.

**Solution:** Binary combinations of pre-existing climos

```
> ncclimo -S 41 -E 50 -s 51 -e 60 -i ${pp}
> ncclimo -S 41 -E 50 -x ${pp}/0041-0050 \
- s 51 -e 60 -i ${pp}/0051-0060
> ncclimo --yr_srt_prv=41 --yr_end_prv=50 \
- --drc_prv=${pp}/0041-0050 --yr_srt=51 \
- --yr_end=60 --drc_in=${pp}/0051-0060 ...
```
New Methods

1. Regrid and "renormalize" partial gridcells
2. Conservatively regrid sub-gridscale distributions
3. Combine existing climos without using raw data
4. Reshape/split multi-field output into timeseries
5. High-frequency climos
Reshape/split multi-field output into timeseries

**Problem:** MIPS want per-variable timeseries in fixed-length (e.g., 50 yr) segments. More useful with ancillary variables (e.g., area, PS).

**Solution:** Parallel split input. Exclusion options:

```
--no_cell_measures, --no_formula_terms, --no_native_grid, --no_staggered_grid
```

```
> ncclimo -s 1 -e 250 --ypf_max=50 *.nc
> ncclimo --yr_srt=1 --yr_end=250 --ypf_max=50 *.nc
```
New Methods

1. Regrid and "renormalize" partial gridcells
2. Conservatively regrid sub-gridscale distributions
3. Combine existing climos without using raw data
4. Reshape/split multi-field output into timeseries
5. High-frequency climos
High-frequency climos

Problem: Researchers need statistics from high frequency timeseries in multiple temporal resolutions (daily to sub-daily) and lengths (timesteps-per-file).

Solution: Average input to 365 day-of-year outputs

> ncclimo -j 8 -c name -C dly -s 2001 -e 2009 *.nc
> ncclimo --job_nbr=8 --caseid=name --clm_md=daily \ --yr_srt=2001 --yr_end=2009 *.nc
4.4 `ncclimo netCDF Climatology Generator`

**SYNTAX**

```bash
ncclimo [-a dec_md] [-C clm_md] [-c caseid] [-d dbg_lvl]
        [-E yr_prv] [-e yr_end] [-f fml_nm] [-h hat_nm] [-i drc_in]
        [-j job_nbr] [-l link_flg] [-m mdl_nm] [-n nco_opt]
        --no_coll_mar ||--no_frm_term ||--no_ntv_tms ||--no_stg_grd
        [-0 drc_rgr] [-o drc_out] [-p par_typ] [-r rgr_opt] [-S rgr_map]
        [-s yr_prv] [-s yr_srt] --stdin ||-t thr_nbr ||-t dfr
[-v var_lst] ||--version ||-x cf_flg ||--x drc_xtn ||--x drc_prv
[-Y rgr_xtn] [-y rgr_prv] --ypf=ypf_max
```

**DESCRIPTION**

In climatology generation mode, `ncclimo` ingests “raw” data consisting of a monthly or annual timeseries of files and from these produces climatological monthly means, seasonal means, and/or annual means. Alternatively, in timeseries reshaping mode, `ncclimo` will subset and temporally split the input raw data timeseries into per-variable files spanning the entire period. `ncclimo` can optionally regrid all output files in either mode. The primary documentation is here.

The `ncclimo` usage for climatology generation looks like

```bash
ncclimo -c caseid -s srt_yr -e end_yr -i drc_in -o drc_out
ncclimo -m mdl_nm -c caseid -s srt_yr -e end_yr -i drc_in -o drc_out
ncclimo -v var_lst -c caseid -s srt_yr -e end_yr -i drc_in -o drc_out
ncclimo --case=cseid --start=srt_yr --end=end_yr --input=drc_in --output=drc_out
```

In climatology generation mode, `ncclimo` constructs the list of input filenames from the argument to the date and model-type options. `ncclimo` automatically switches to timeseries reshaping mode if it receives a list of files from stdin, or, alternatively, placed as positional arguments (after the last command-line option), or if neither of these is done and no `caseid` is specified, in which case it assumes all *nc files in drc_in constitute the input file list.

Options come in both short (single-letter) and long forms. The handful of long-option synonyms for each option allows the user to imbue the commands with a level of verbosity and precision that suits her taste. A
Supplementary Slides