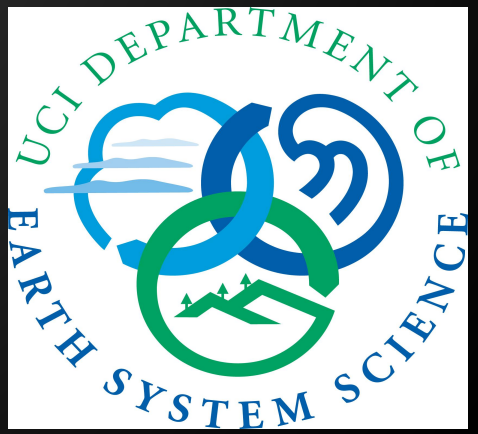
New Methods to Combine, Regrid, and Split Climos

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

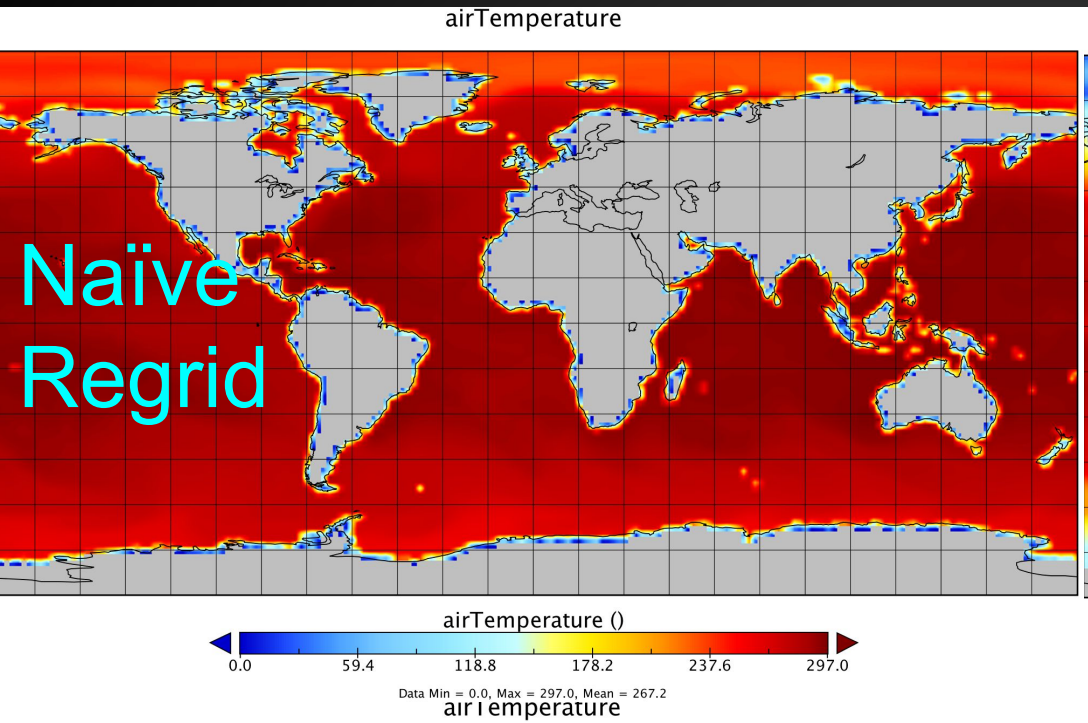
On Web

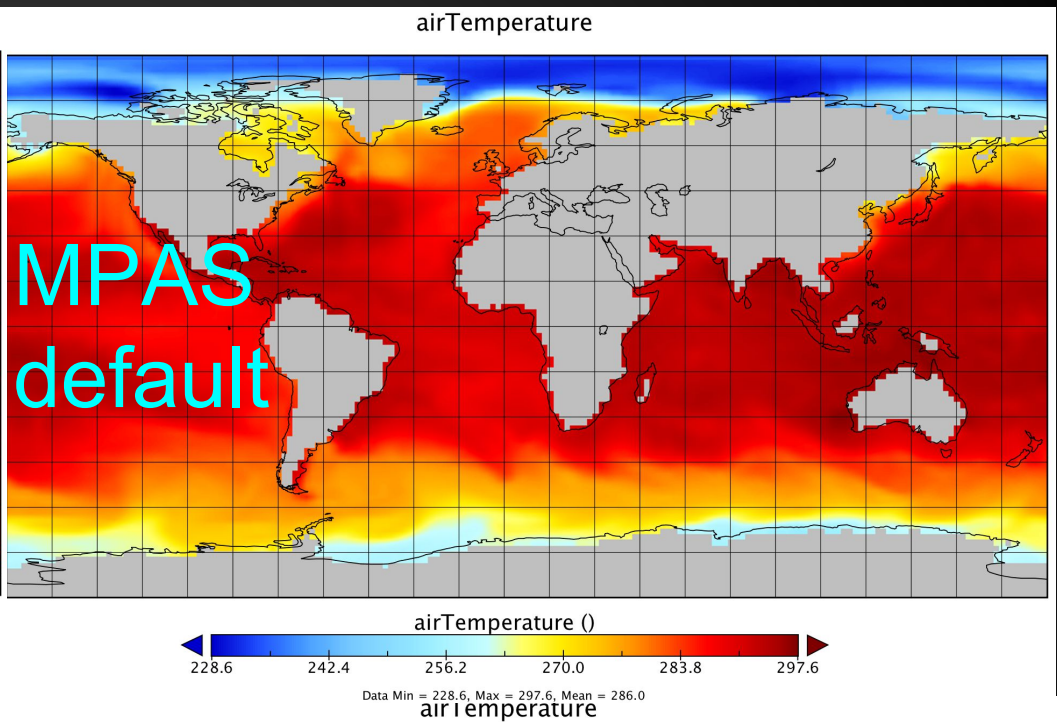


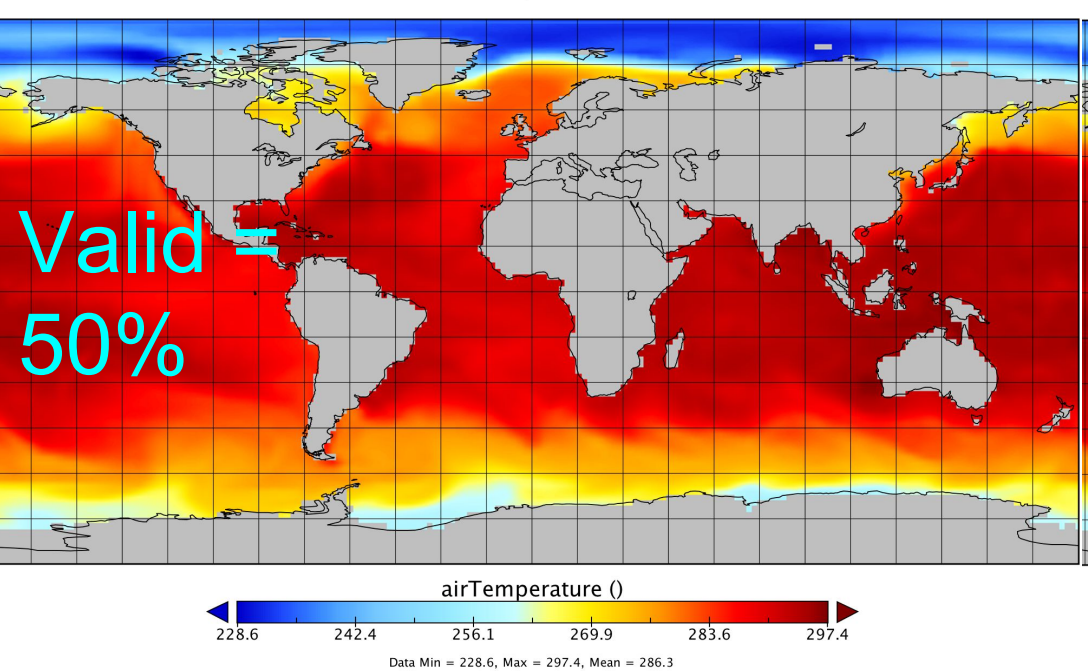
New Methods

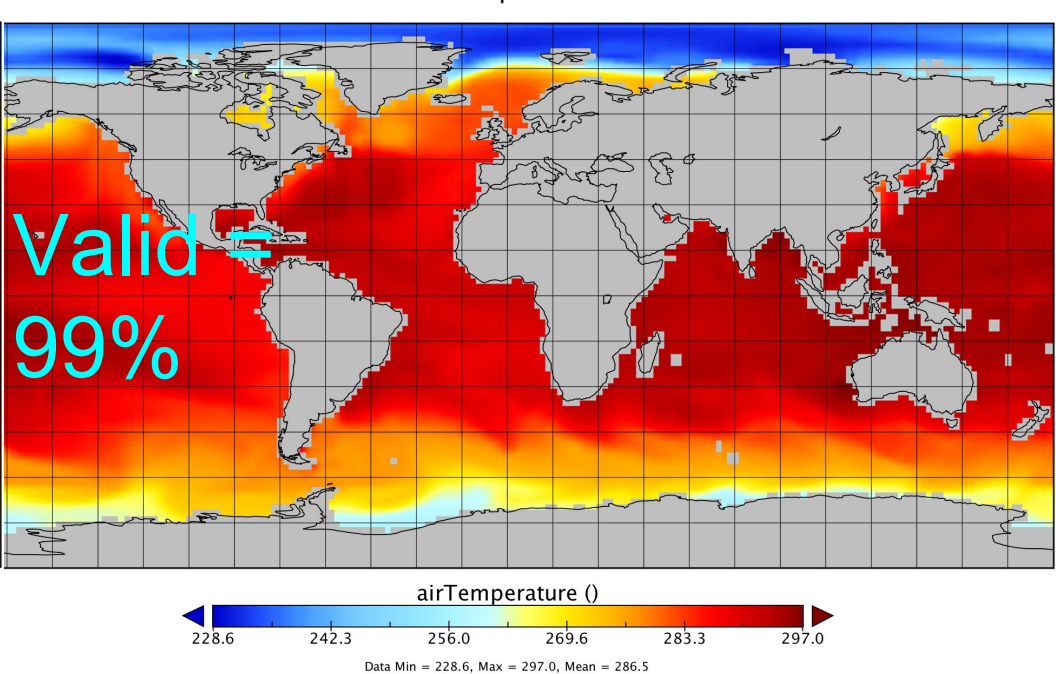
1. Regrid and "renormalize" partial gridcells
2. Conservatively regrid sub-gridscale distributions
3. Combine existing climos without using raw data
4. Reshape/split multi-field output into timeseries
5. High-frequency climos

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









Regrid and "renormalize" partial gridcells

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

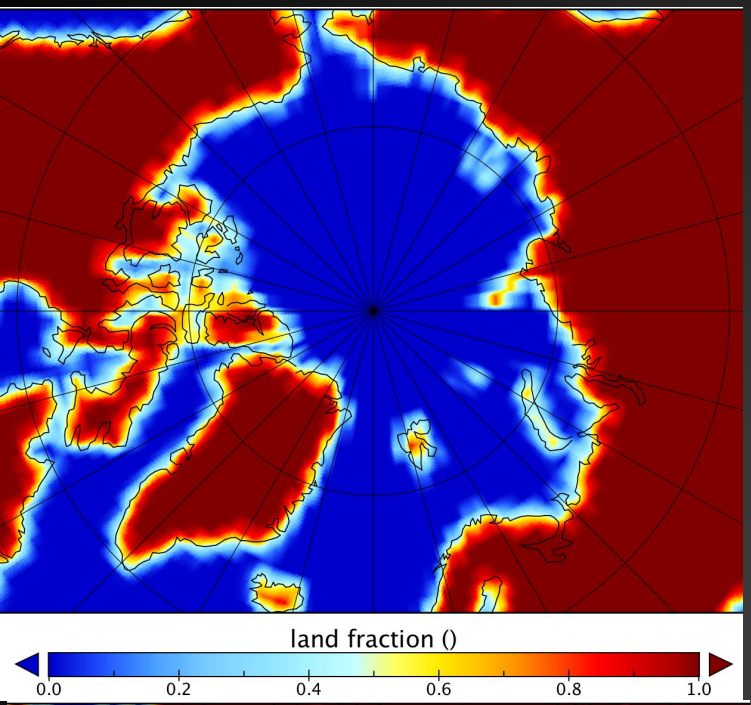
Solution: Renormalize values with tunable threshold

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

New Methods

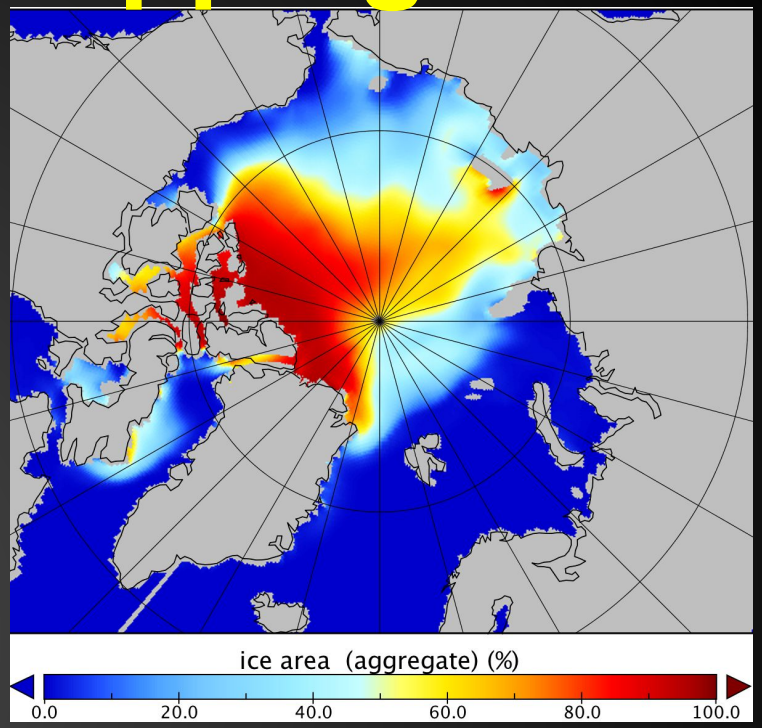
1. Regrid and "renormalize" partial gridcells
2. Conservatively regrid sub-gridscale distributions
3. Combine existing climos without using raw data
4. Reshape/split multi-field output into timeseries
5. High-frequency climos

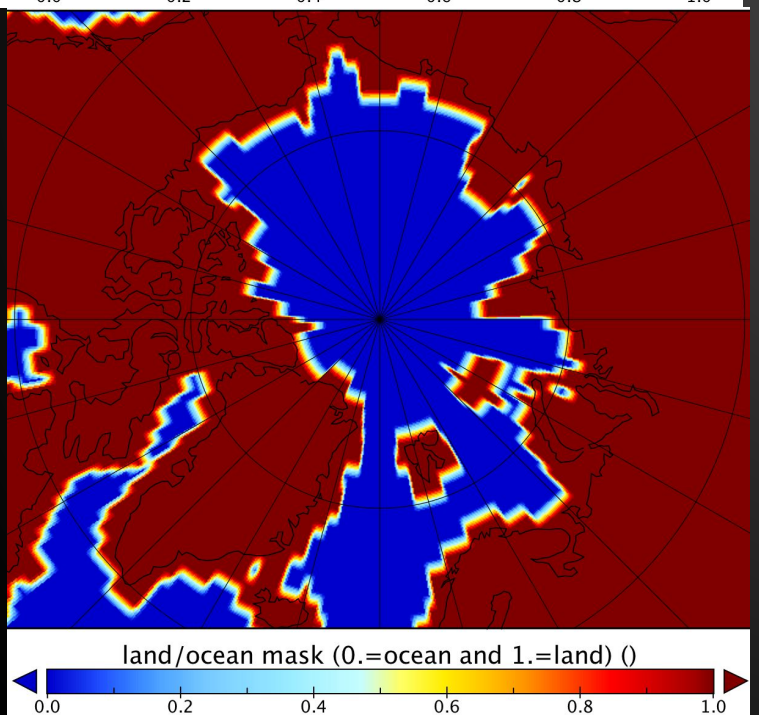
Sub-Gridscale Remapping



Land Fraction

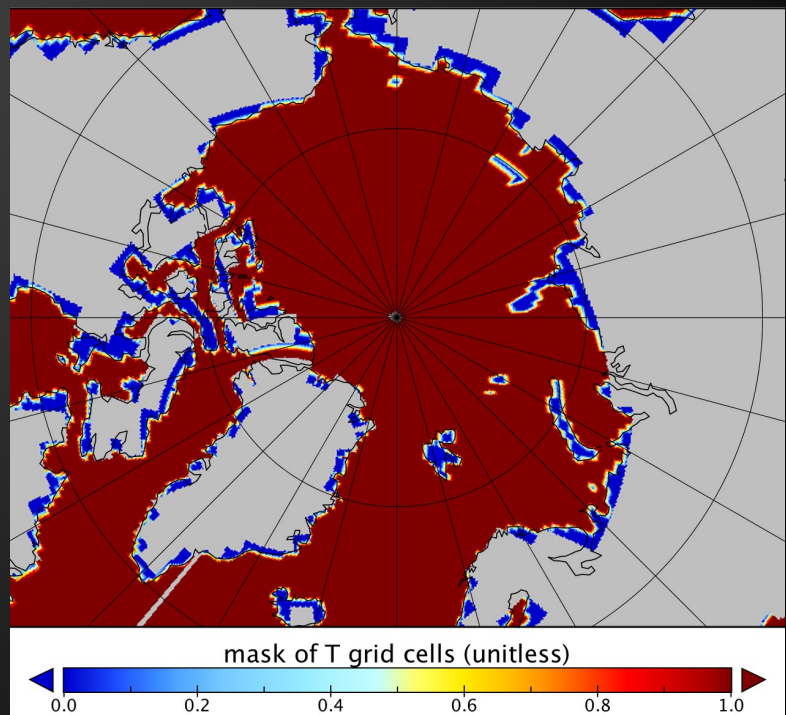
Sea-ice Fraction





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

```
> 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 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

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
```

ACME Docs

Complete Docs

Secure https://acme-climate.atlassian.net/wiki/di...

ACME Confluence Spaces More Create

Pages Edit Save for later Watching Share

/ Coupled Simulation Group / How-to articles

Generate, Regrid, and Split Climatologies (climo files) with ncclimo and ncremap

Created by Peter Caldwell, last modified by Charlie Zender on Jun 01, 2017

Overview:

Based on extensive evaluation of AMWG, UV-CDAT, and NCO codes for generating climatology files (see here), we have determined that NCO provides the most correct answers, has the best metadata, and is fastest. Until UV-CDAT bests NCO in these measures we advocate using NCO for creating climatologies.

In climatology generation mode, the NCO operator **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. This presentation, given at the Albuquerque workshop on 20151104, conveys much of the information presented below, and some newer information, in a more graphical format.

Prerequisites:

Use **ncclimo** if possible. It requires and comes with NCO version 4.6.0 and later. Its predecessor **climo_nco.sh** (which is deprecated) requires NCO version 4.5.2 or later. The newest versions of NCO are installed on rhea/titan.ccs.ornl.gov at ORNL, pileus.ornl.gov (CADES at ORNL), cooley/mira.alcf.anl.gov at ANL, cori/edison.nersc.gov (NERSC), aims4.llnl.gov (LLNL), roger.ncsa.illinois.edu (NCSA), and yellowstone.ucar.edu (NCAR). The **ncclimo** and **ncremap** scripts are hard-coded to find the latest versions automatically, and do not require any module or path changes. To use other (besides the **ncclimo** and **ncremap** scripts) NCO executables from the command-line or from your own scripts may require loading modules. This is site-specific and not under my (CZ's) control. At OLCF, for example, "module load gcc" helps to run NCO from the command-line or scripts. For other machines check that the default NCO is recent enough (try "module load nco", then "ncks --version") or use developers' executables/libraries (in ~zender/[bin,lib] on all machines). Follow these directions on the NCO homepage to install on your own machines/directories. It can be as easy as "apt-get install nco", "dnf install nco", or "conda install -c conda-forge nco", or you can build/install from scratch with "configure;make install".

Climatology generation mode (produce monthly + seasonal + annual

Posters_Layout_and_A....pdf pst_nco_acme_201706....pdf pst_nco_acme_201706.pdf Show All

nco.sourceforge.net/nco.html#ncclimo

Next: ncecat netCDF Ensemble Concatenator, Previous: ncbo netCDF Binary Operator, Up: Reference Manual [Contents][Index]

4.4 `ncclimo` netCDF Climatology Generator

SYNTAX

```
ncclimo [-a dec_md] [-C clm_md] [-c caseid] [-d dbg_lvl]
[-E yr_prv] [-e yr_end] [-f fml_nm] [-h hst_nm] [-i drc_in]
[-j job_nbr] [-l lnk_flg] [-m mdl_nm] [-n nco_opt]
[--no_cll_msr ] [--no_frm_trm] [--no_ntv_tms] [--no_stg_grd]
[-O drc_rgr] [-o drc_out] [-p par_typ] [-R rgr_opt] [-r rgr_map]
[-S yr_prv] [-s yr_srt] [--stdin] [-t thr_nbr] [--tpd=tpd_dly]
[-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.

There are five required options (‘`-c`’, ‘`-s`’, ‘`-e`’, ‘`-i`’, and ‘`-o`’)) to generate climatologies, and many more options are available to customize the processing. Options are similar to `ncremap` options. Standard `ncclimo` usage for climatology generation looks like

```
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=caseid --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

Posters_Layout_and_A....pdf pst_nco_acme_201706....pdf pst_nco_acme_201706.pdf Show All

Supplementary Slides

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