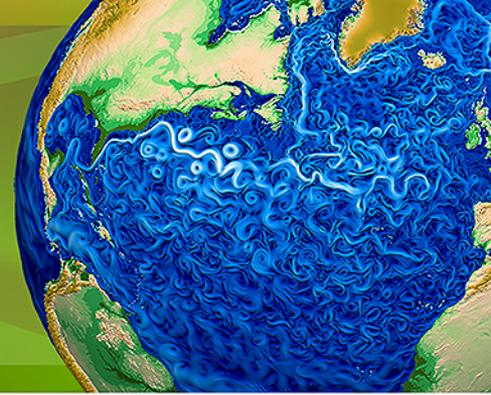




Accelerated Climate Modeling
for Energy



V0.3 AMIP simulations

Peter Caldwell, Kate Evans, Marcia Branstetter, Qi Tang, Pat Worley, Mark Taylor, Charles Doutriaux, Jeff Painter, Sasha Ames, John Harney, Jayesh Krishna, ...



Overview

Lead = Peter Caldwell, Emeritus Lead = Kate Evans

- Motivation:
 - to create data for building/testing tier 1b diagnostics
 - to serve as the baseline against which to test v1 simulations
 - to build experience with production runs
 - because we have 97 M core hours on Mira to use up by June 31st
- Setup:
 - AMIP runs from 1979-2005 (with possible extension through 2013)
 - 4 ensemble members each @ ne30 (1^o) and ne120 (1/4^o) resolution
 - Lots of monthly + a little daily output... no sub-daily (additional runs needed)
 - tuned for energy balance in pre-industrial atmos-only configuration
 - *Note: since tunings are different for ne30 vs ne120, use for resolution-sensitivity studies is debatable...*
 - We hope to do another set of ne30 runs with ne120 tunings
- Why v0.3?
 - SE decided to tag the model v0.3 for these runs. We haven't made any science changes to the model yet, so this is still a reasonable baseline

Publishing (1)

Lead = Sasha Ames and John Harney, Advisor = Peter Caldwell

Before we can publish...

1. Facets need to be defined (Peter has drafted, discussion needed)
2. Estimate of data size needed (=1.4 TB for ne30, 22 TB for ne120)
3. Discussion needed about future considerations
 1. Data should also be archived to HPSS where produced
 2. We may continue the runs for another 8 yrs

Comments:

- AMIP task isn't done until ALL raw output is published
- We also intend to publish climos (on native grid)
- Do we need to publish regridded climos(?)



Publishing (2)

Lead = Sasha Ames and John Harney, Advisor = Peter Caldwell

Plan:

- People running model (Marcia + Qi) will:
 1. move raw data to CADES using GLOBUS
 2. check that the files look ok
 3. notify John and Sasha that data is ready to be published
- Workflow people (John and Sasha) will then:
 1. Organize files into the correct directory structure on CADES
 2. Alter ESGF publication config file per facet metadata
 3. Run publication scripts
 4. Ensure datasets (and files contained within those datasets) are visible to everyone in ACME in both the index and ESGF data store.
 5. Make sure that the content of datasets/files is correct

Making Climo Files

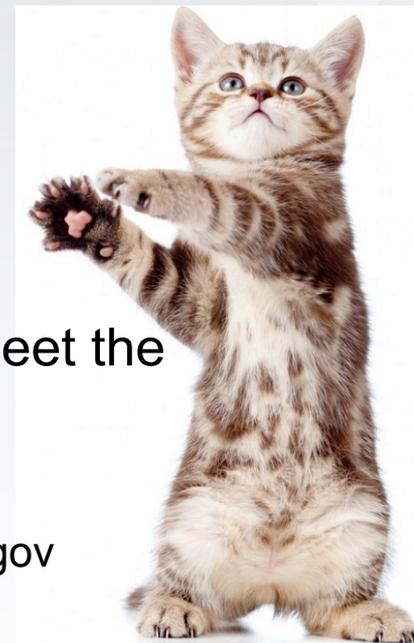
Lead = Jeff Painter, Advisor = Peter Caldwell

Diagnostics operate on climatology files (climos) rather than raw output. We need a utility to (in essence):

- create annual, monthly, and seasonal averages on the native grid
- include all variables in the original file
- be extremely fast (probably through parallelization)
- include all metadata from original files

Progress:

- Workflow has 2 routines for this, but they need work to meet the above requirements
- We are using an NCO-based solution as a stopgap
 - it takes 1.5 min to create a 5 yr ne30 Jan climo file on crunchy.llnl.gov
 - it takes 5 min to create a 2 yr ne120 Jan climo file on tukey.anl.gov
 - parallelizing over month and season => creating all files takes 3x as long



Regridding

Lead = Mark Taylor, Charles Doutriaux, Advisor = Peter Caldwell, Qi Tang

Native-grid SE data is unstructured → very hard to plot. We will:

- use ESMF (or eventually Tempest) to generate weights
- use UV-CDAT to regrid by multiplying original data by weights
- use 3 equal-angle target grids: 1.4° , 0.7° , and 0.225°
 - Bilinear interpolation will be used to grid to higher resolution. Conservative mapping will be used to coarsen data.

Progress:

- ESMF weight generator works. Mapping files are done.
- UV-CDAT regridding works(?)
 - It takes 37 min per ne120 climo file on sierra.llnl.gov (>2x faster than the NCL version we have been using!)
 - The NCL version seems to lose all global attributes

ne30 Runs

Lead = Marcia Branstetter, Advisor = Kate Evans

- We planned to use Mira but its queuing/charging is hostile to small jobs, so switched to Edison
 - We have 43M core hours to burn on Edison before Feb 2016
 - We ran all 27 yrs as one 48 hr job for each ensemble member
 - runs sat in queue for ~1 wk, then all ran.
 - Each costs 3,776 pe hrs/yr/ensemble member
 - We get 22 SYPD using ??? cores
- Progress:
 1. All simulations are finished!
 2. All output has been staged to CADES (publication is waiting on facet definition)
 3. Creation/publishing of climo files is still needed
 4. Regridding has not been tested

ne120 Runs

Lead = Qi Tang, Advisor = Peter Caldwell

- We are running on Mira:
 - The ALCC allocation (96 M core-hours) will expire on 7/1/2015
 - We bundle 4 jobs into a single submission for improved queueing
 - Details are on Qi's poster
 - Optimization requires 2048 nodes/job (8192 nodes total, 16 cores/node)
 - We get 1.7 SYPD and each year costs ~2 M core hours
 - Each job produces 1 yr of data and finishes in 15 hours
 - each submission spends 1-2 days in queue
 - making each job depend on one before improves throughput
- The simulations are expected to finish by early June.
- We have 4 years of ensemble data now (4/30)
- We produced 0.7° climo files for the 1st 3 yrs at tukey.anl.gov

ne120 Results

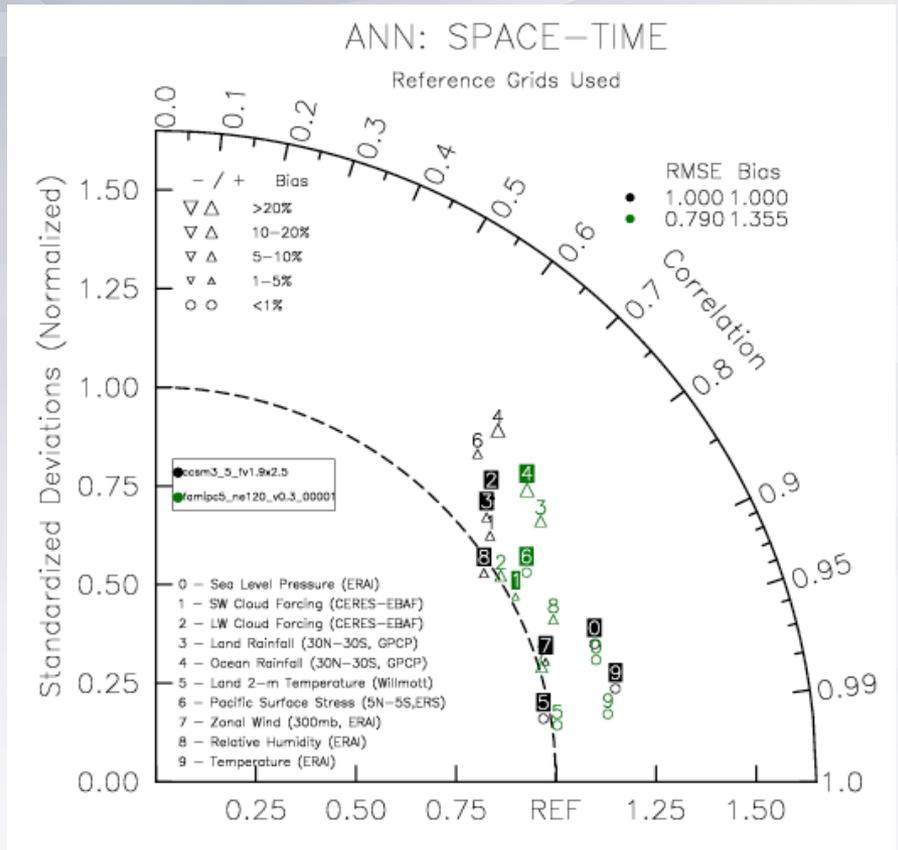
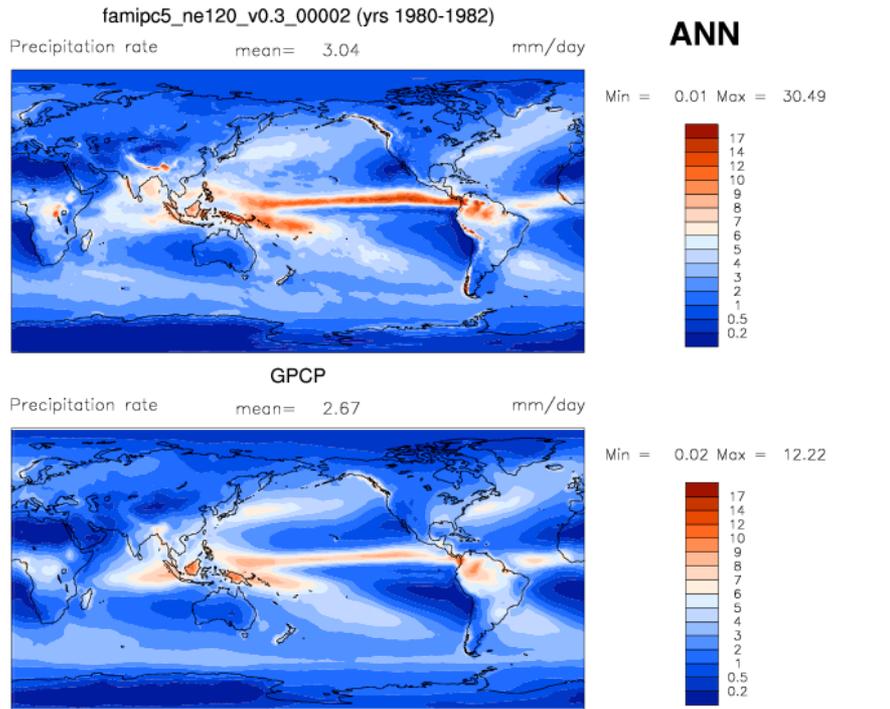


Fig: precip and SW CRE look typical for CAM5 @ 1/4 degree

- A cursory look suggests model output is reasonable. See http://portal.nersc.gov/archive/home/projects/acme/www/famipc5_ne120_v0.3_00001-obs/index.html for more

Verbose-Output Runs

We couldn't add all requested output to these runs because:

1. 27 yrs x 4 ensemble members + lots of hrly and 3 hrly requests = petabytes of data
2. a bug prevents Mira from running threaded with the CFMIP Observation Simulator Package (COSP)
 - This causes a 2-4x slowdown

We plan to do a single (shorter?) simulation at ne30 and ne120 with all desired output

- we need COSP
- we need to bundle ensemble members for throughput on Mira... which precludes a single decade-long run
 - Can we branch an existing run at selected years and re-run years with verbose output in parallel?
 - Can we run on Titan?

Questions to answer!

For Discussion



1. What do we still need to accomplish? What pitfalls should we watch out for?
2. What did we do a good job on?
3. What successes can we export to the rest of the project?
4. What could we do better on next time?

Extra slides

Lessons Learned/To-do:

1. Whole team should have been brought in on experimental design earlier
 - Or would we have stalled out in endless discussion?
 - ACME needs to get better at setting cutoff dates for discussion → action
2. We need to do a better job of setting requirements and interacting with Workflow
 - Can climo task serve as a role model?
 - Would embedding workflow people in atmos tasks work?
3. We need to do better at planning how to use our computing allocations (and following through on that plan)

ACME Publication to ESGF: Requirements

- Information required for publication:

- **Overview:** We have simulations at ne30 and ne120 resolutions each with 4 ensemble members. Raw data will consist of native-grid h0 (monthly) and h1 (daily) files and we will also store h0 climo files and perhaps regridded h0 climo files as well. We will conduct a sister simulation using the same configuration but with more verbose output (6 hrly, 3 hrly, and hrly)
- **Where is data written?** ne30 = Edison (NERSC). ne120=Mira (ANL)
- **How much data (estimate)?** ne30 = X, ne120 = Y
- **What data product?** h0 (monthly) and h1 (daily) files on native grid, h0 climo files on native grid, and perhaps regridded climo files. Sister simulation will have more verbose output
- **Case / experiment name?** amipc5_{ne30,ne120}_v0.3_0000{1-4}
- **Project** ACME
- **Realm** atm
- **Version Number** v0.3
- **How was data regridded / grid type** ne30, ne120. Perhaps also regridded conservatively to 1.4 and 0.7 deg (respectively)
- **What years spanned (if applicable)** Initially 1979-2005 but we may eventually extend to 2013.
- **How to handle directory structure on CADES** I think Workflow needs to suggest an answer to this
- **Future issues** this data will also be stored on HPSS where the simulations ran