EarthWorks Progress

David Randall[1], James Hurrell[1], Donald Dazlich[1], Lantao Sun[1], Andrew Feder[1], William Skamarock[2], Sheri Mickelson[2], Chris Fisher[2], Mariana Vertenstein[2], Thomas Hauser[2], Richard Loft[3], Andrew Gettelman[5], Phil Jones[4], Luke van Roekel[4]

[1]Colorado State University
[2]National Center for Atmospheric Research
[3]AreandDee, LLC
[4]Los Alamos National Laboratory
[5]Pacific Northwest National Laboratory

COLORADO STATE UNIVERSITY

http://hogback.atmos.colostate.edu/earthworks/

NCAR

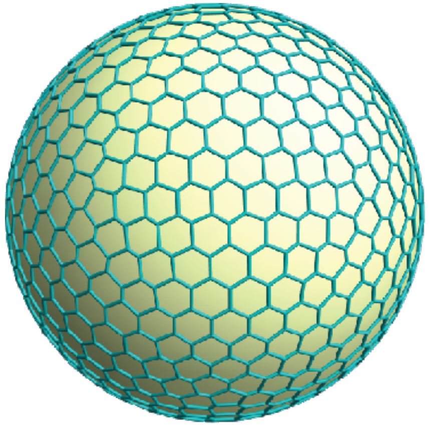
See related talks by
Bill Skamarock,
Brian Medeiros, and
Andrew Gettelman

EarthWorks

EarthWorks is a five-year university-based project, supported by NSF CISE, to develop a global storm-resolving coupled model that uses *a single nearly uniform global grid* for the atmosphere, ocean, sea ice and land surface.

Earthworks consists of:

- CAM6 with the MPAS non-hydrostatic dynamical core
- The MPAS ocean model developed at LANL
- The MPAS sea ice model, based on CICE
- The Community Land Model (CLM)
- The Community Mediator for Earth Prediction Systems (CMEPS)



Earthworks is preserving compatibility with the evolving CESM code base.

EarthWorks

GOAL:

- Develop the capability to perform 3.75 km fully coupled simulations utilizing GPU-enabled components with end-to-end workflow portability across US leadership computing systems

SOME INITIAL OBJECTIVES:

- Port MPAS ocean and sea ice models into the CESM framework ✓
- Assemble a working CPU version of the EarthWorks configuration ✓
- Complete fully coupled simulations at relatively coarse grid resolutions ✓
- Test MPAS-CAM6 physics at convection-permitting spatial scales (Very good progress)
- Port critical-path portions of EarthWorks to GPUs using OpenACC (Very good progress)

Summary of simulations to date

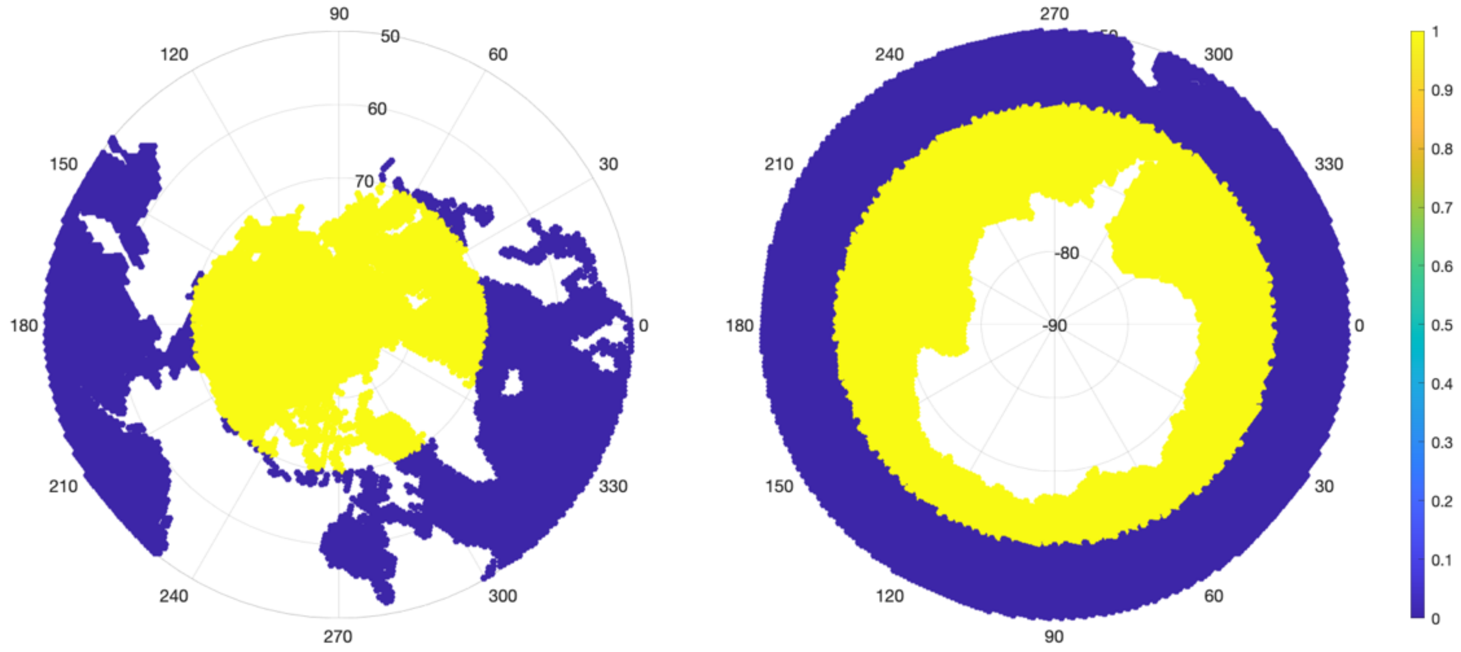
We have done both AMIP simulations and fully coupled simulations with 120 km, 60 km, and 30 km grids.

We have done Aquaplanet simulations with all grids from 120 km down to 3.75 km.

We have done three simulations with regional refinement.

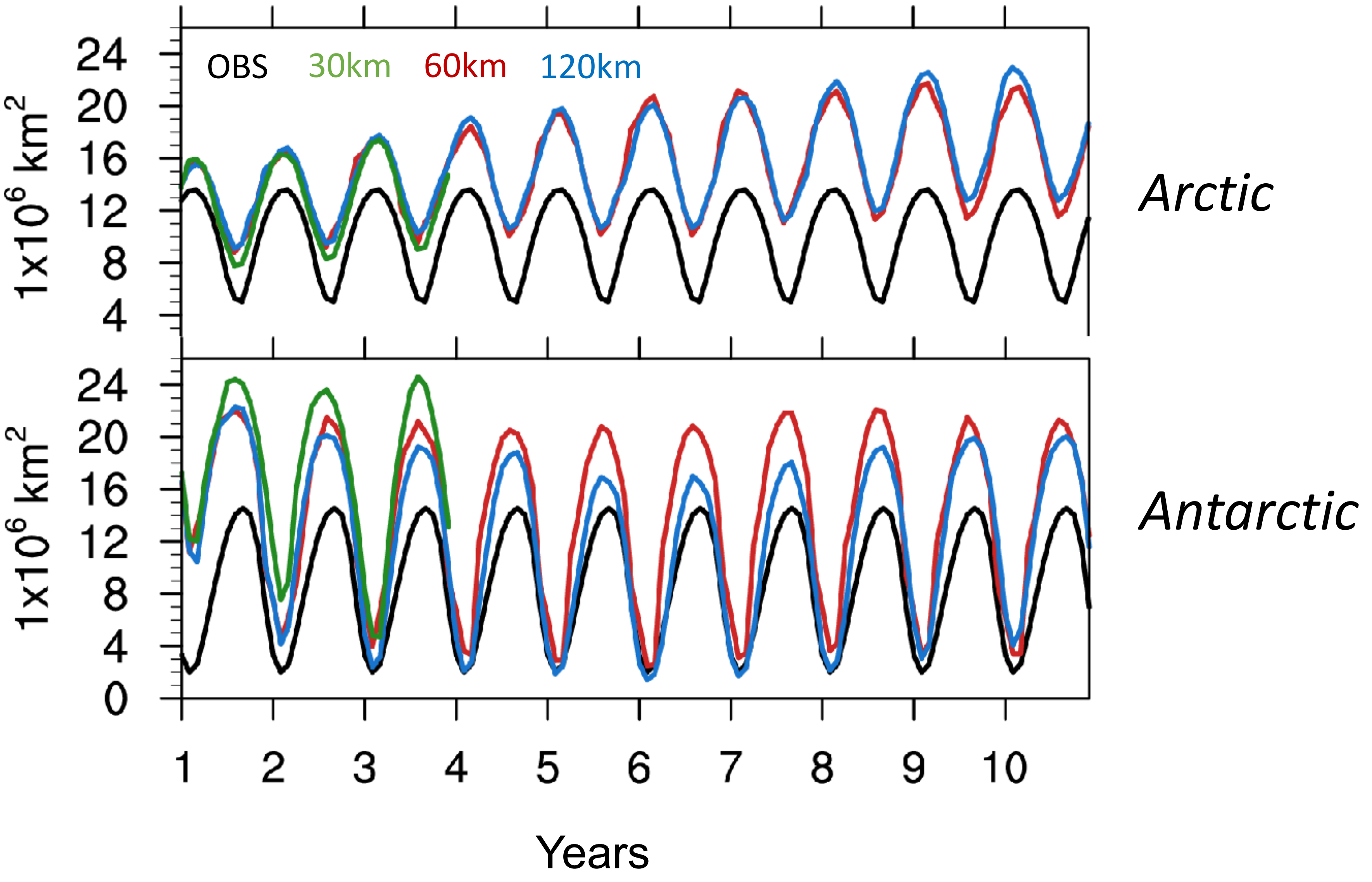
Initial sea ice distribution on January 1

60 km grid



The initial ice thickness was specified as 2.5 m everywhere that ice was present.

The sea-ice area starts too high and then expands.



Ocean & sea ice initialization, take 2

Following advice from Gokhan, we are now implementing the OMIP initialization protocol described by Griffies et al. (2016):

- Ocean, World Ocean Atlas 2013 version 2
- Seaice - previous distribution January simulation, (if available, else disk IC).
- No motion

Restoring:

- Surface salinity restoring damped to a monthly climatology. Examples are given in Table 2 of Danabasoglu et al. (2014).

Forcing (CORE II):

- Interannually varying atmospheric state of Large and Yeager (2009), 62 years long (1948-2009). This is available as a CESM data-atmosphere.
- River runoff - Dai and Trenberth (2002), also available in CESM
- Biogeochemical forcing - not applicable

Simulation length:

- Five cycles of the 62-year forcing

Start with the 120-km grid. Interpolate to the finer grids.

Danabasoglu, G., et al., 2014: North Atlantic simulations in Coordinated Ocean-ice Reference Experiments phase II (CORE-II). Part I: Mean states. *Ocean Modelling*, **73**, 76-107.

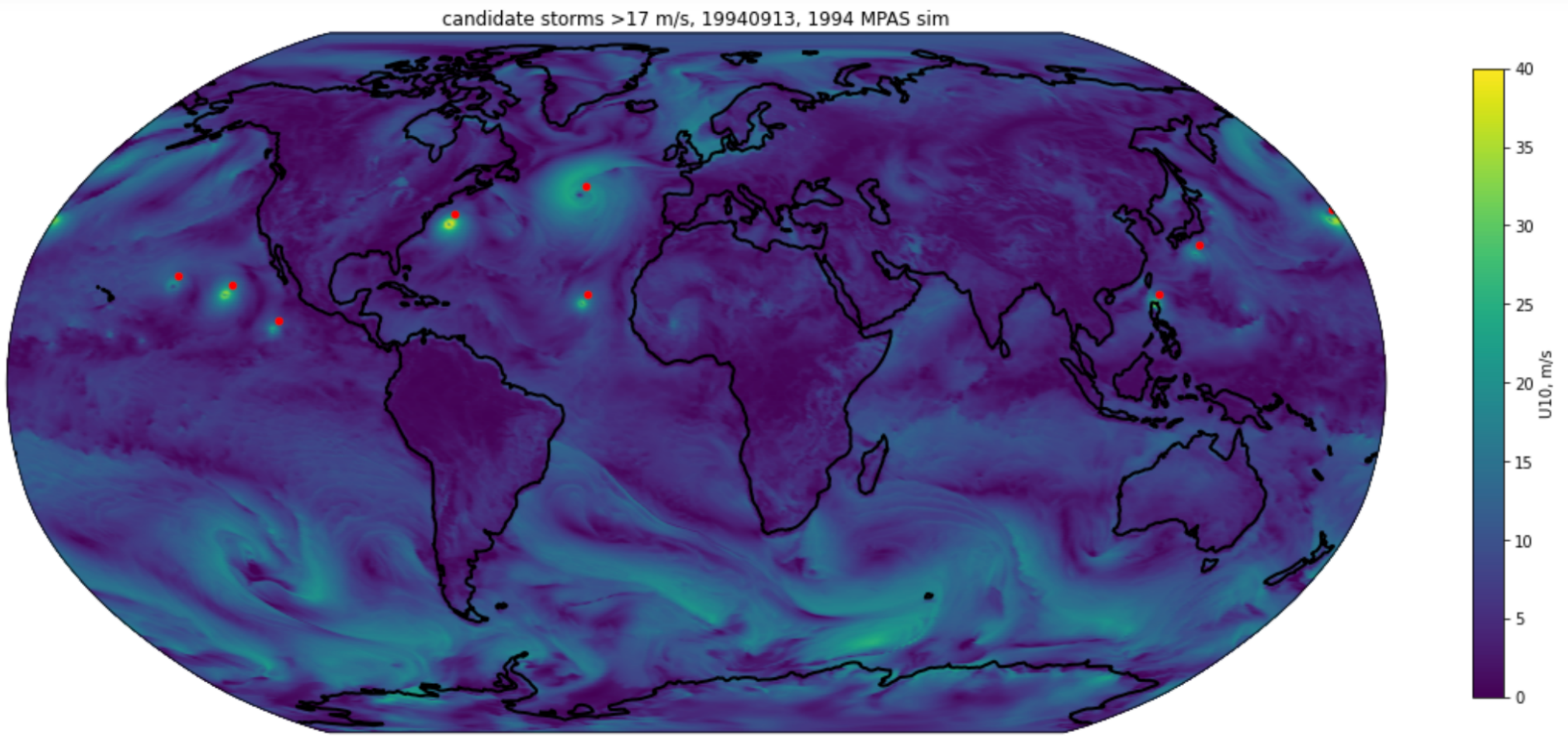
Dai, A., & Trenberth, K.E., 2002:. Estimates of Freshwater Discharge from Continents: Latitudinal and Seasonal Variations. *Journal of Hydrometeorology*, **3**, 660-687.

Griffies, S., et al., 2016: OMIP contribution to CMIP6: experimental and diagnostic protocol for the physical component of the Ocean Model Intercomparison Project. *Geoscientific Model Development*, **9**, 3231-3296.

Large, W.G., & Yeager, S.G., 2009: The global climatology of an interannually varying air–sea flux data set. *Climate Dynamics*, **33**, 341-364.

Tropical cyclones in EarthWorks with a 30 km grid

Work by graduate student Andrew Feder



10-m wind speed on an especially stormy day in an AMIP simulation of 1994.

The TC-detection algorithm of Zhao et al.[1] has been adjusted and simplified, using Python. Detections are based on vorticity, TC warm cores and low-pressure centers.

With a 30 km grid, EarthWorks produces 80-90 TC's a year.

Cyclones possess eyewalls and tracks are realistic for their basins.

Wind strengths range only up to Cat 2, despite central pressures that suggest stronger storms; this suggests PBL parameterization issues.

[1]Zhao M, Held IM, Lin SJ, Vecchi GA (2009) Simulations of global hurricane climatology, interannual variability, and response to global warming using a 50-km resolution GCM. J Clim 22(24):6653–6678

GPU Porting Targets & Status

- MPAS Dynamics - completed/optimizing
- CAM Physics - in progress
  - PUMAS μphysics - completed
  - RRTMGP radiation - completed/testing
  - CLUBB cloud - porting
- MPAS Ocean - completed/testing

We have recently run a very short aquaplanet simulation (QPC6) with a 3.75 km grid, on 1024 nodes (out of 8008) of Frontera, at TACC.

Frontera runs QPC6 about 1.9x faster than Cheyenne, node for node. A Frontera core is about 22% faster than a Cheyenne core.

We expect Derecho cores to be about 6% faster than Frontera cores.

Memory scaling on Cheyenne

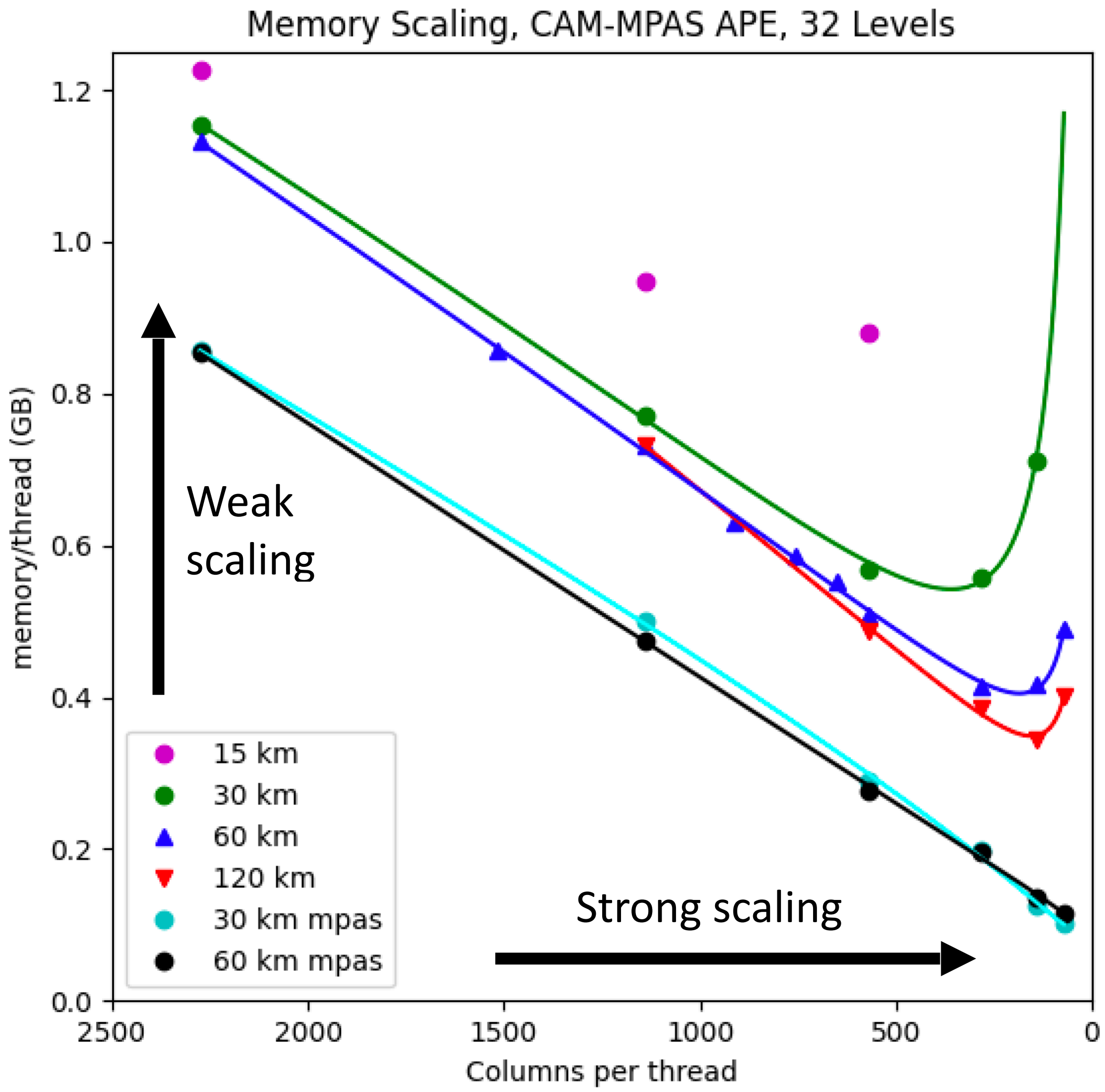
MPAS meshes
120 km, 40962 columns
60 km, 163842 columns
30 km, 655362 columns
15 km, 2621442 columns

Weak scaling: mesh size/#nodes fixed
Strong scaling: mesh size fixed, #nodes vary

CAM-MPAS simulations:
120 km: 1, 2, 4, 8 and 16 nodes
60 km: 2, 3, 4, 5, 6, 7, 8, 16, 64 nodes
30 km: 8, 16, 32, 64, 128 nodes (256 hangs in init)
15 km: 32, 64, 128 nodes (256 hangs in init)

Lines are least-squares fit polynomials:

Memory/thread =
$a_o + a_1$ cols/thrd + a_2 (col+halo)/thrd + a_3 tot_#thrds



Slide from Bill Skamarock

Weak scaling of AP on Frontera

Profiling + timing results

- **Weak Scaling:**
  - increase model resolution holding the columns/node constant.
  - In a **perfectly scalable model,** the throughput, memory use, and GFlops would stay constant.
- **Good News:**
  - Model throughput is weak scaling well, with key performance metrics also holding up: GFlops, Memory and interconnect BW.
- **Bad News:**
  - Initialization takes 95x longer than expected - doesn't seem to be scaling at all;
  - Coupling throughput is 8.2x slower than expected (coupling freq. was held constant).
  - Memory usage (GB/node) has grown 72% during simulation phase. Consistent with Skamarock's observations.

Resolution (km/levels)	Frontera nodes	Columns/ node	Initialization (sec)	TOTAL (myrs/wday)	GB/node (simulation phase)	CPL (myrs/wday)
(30km/32L)	16	40K	33.604	0.51	36	27.49
(15km/32L)	64	40K	TBD	TBD	TBD	TBD
(3.75km/32L)	1024	40K	3193.580	0.095	62	3.35
Expected (Perfect Weak Scaling)			33.6	0.106*	36	27.49

*To compare model throughput we must account for the reduction in timestep (120 sec -> 25 sec) required by the CFL condition.

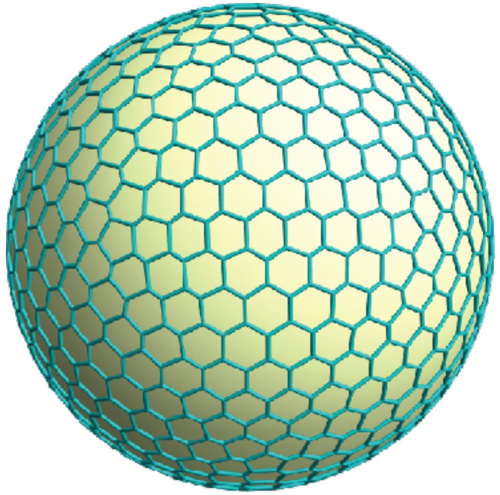
Slide from Rich Loft and Chris Fischer

Projection to EarthWorks goals

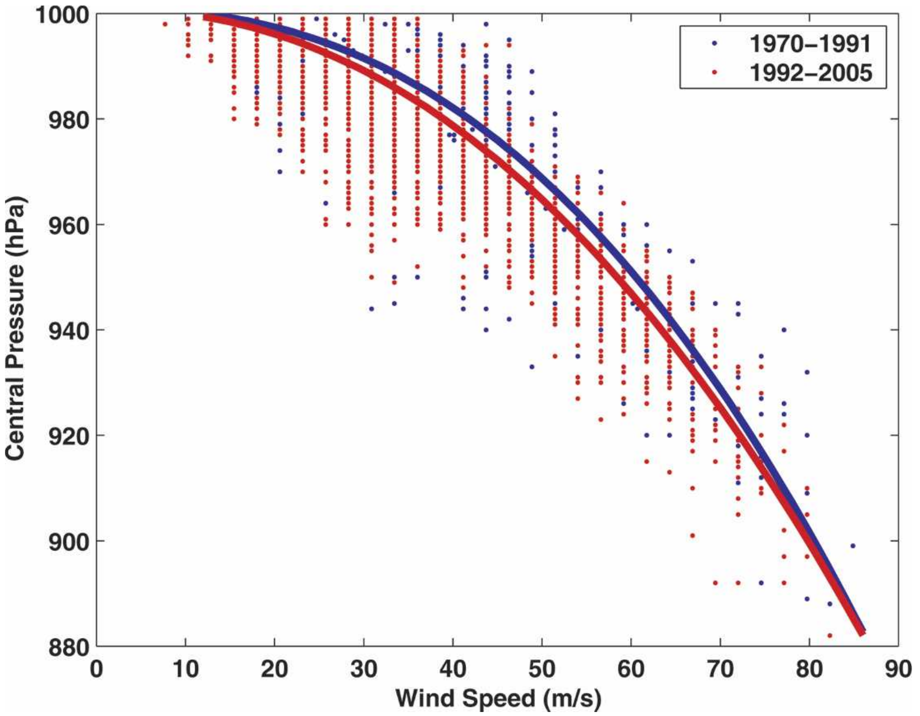
- **EarthWorks goal: demonstrate fully-coupled at 3.75km/L58 @ 0.5 myrs/wday by 2025**
  - atm, land, ocean, sea-ice on same grid
- **L32-> L58 atmospheric levels: 1.75x cost increase**
  - Assumes linear cost function.
- **Estimated overhead AP 58 levels -> fully-coupled: 17%**
  - Based on observed fully-coupled overheads with 30km/32L runs on Cheyenne.
- **Resulting Estimated Frontera Throughput:**
  - 0.046 myrs/wday on 1024 nodes: 10.7x too slow
  - *This is within the plausible range of GPU acceleration on 1024 nodes each with 4 GPUs.*

Summary

- EarthWorks is up and running in the CESM framework.
- We are revising our approach to initializing the ocean and sea ice.
- We are analyzing many aspects of the results, including tropical cyclones (shown here).
- Porting to GPUs is essentially complete except for CLUBB.
- Initialization is unreasonably slow on the 3.75-km grid.
- EarthWorks is on track to produce at least one simulated year per day on a 3.75 km grid by the completion of our first five years.



Extra slides



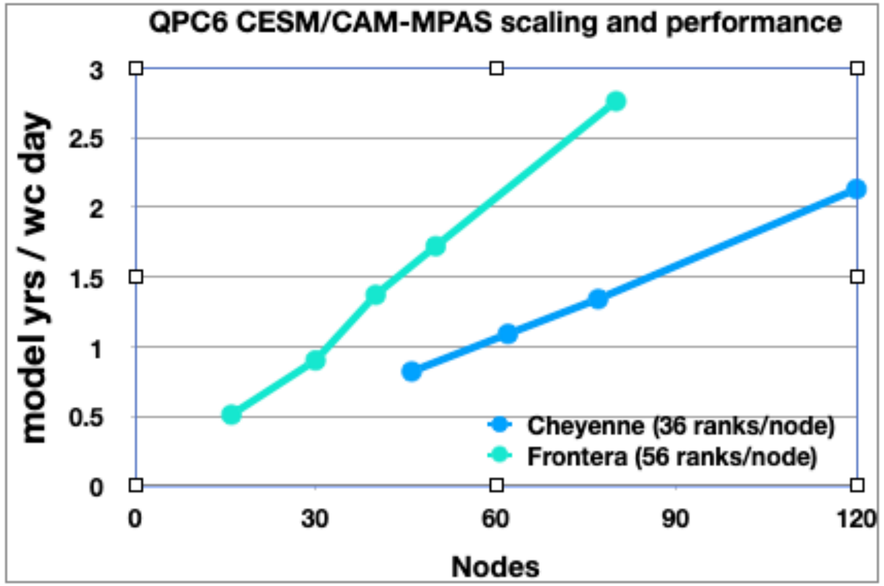
Plug in our model's minimum PSL's, 890-900 mb in several months of 1994, and that should be yielding Cat 4s, with wind speeds of up to 65 m/s.

$$V = 5.95\sqrt{(1010 - p_c)}$$

Kerry Emanuel, 2007: Environmental Factors Affecting Tropical Cyclone Power Dissipation. *Journal of Climate*

Performance of Aqua Planet at 30 km/32 levels

Cheyenne-Frontera inter-comparison



What AP benchmarks about Cheyenne vs Frontera:

- Frontera has more cores/node (56) Compared to Cheyenne (36)
- Both have Intel Xeon processors, but Frontera's Cascade Lake are newer and clocked faster (3.1 GHz vs 2.3 GHz) than Cheyenne's Broadwell processor.
- **Node-for-node** on AP: Frontera is **1.72x faster**
- **Core-for-core** on AP: Frontera is **11% faster**

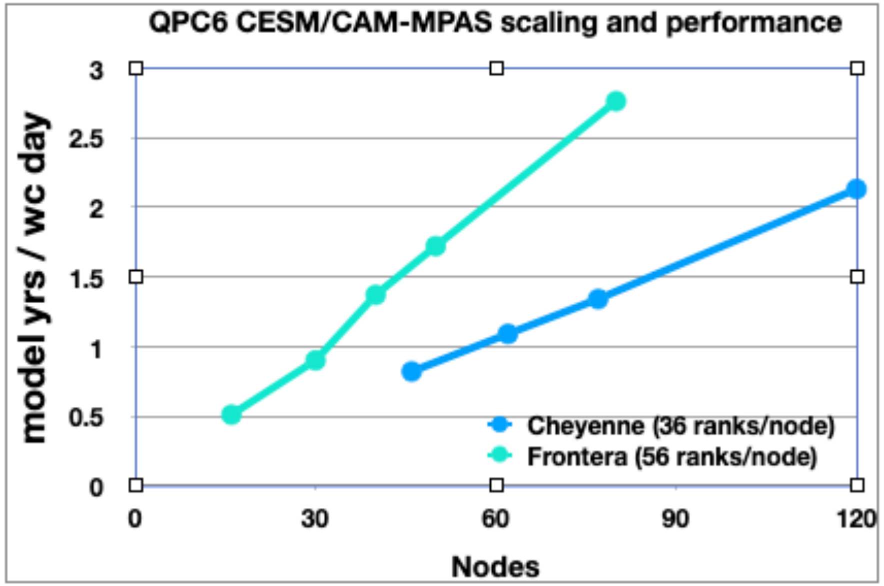
Acknowledgement: Chris Fischer, NCAR

Time steps & coupling frequency

Grid spacing	ATM dynamics time step, seconds	Coupling frequency & ATM physics time step, seconds	OCN barotropic solver time step, seconds	OCN physics/advection time step, seconds
120 km	450	1800	90	1800
60 km	225	900	60	900
30 km	120	360	60	360
15 km (planned)	90	180	30	180
7.5 km (planned)	45	90	15	90
3.75 km	22.5	45	5	45

Performance of Aqua Planet at 30 km/32 levels

Cheyenne-Frontera inter-comparison



What AP benchmarks about Cheyenne vs Frontera:

- Frontera has more cores/node (56) Compared to Cheyenne (36)
- Both have Intel Xeon processors, but Frontera's Cascade Lake are newer and clocked faster (3.1 GHz vs 2.3 GHz) than Cheyenne's Broadwell processor.
- **Node-for-node** on AP: Frontera is **1.72x faster**
- **Core-for-core** on AP: Frontera is **11% faster**

Acknowledgement: Chris Fischer, NCAR

Diagnostic packages

ADF (atmosphere)

CAM Diagnostics

Test Case: mpas_aos60

Control Case: obs

Plot Types: Tables **LatLon** LatLon_Vector Zonal NHPolar

Clouds

CLDHGH
ANN DJF MAM JJA SON
CLDLOW
ANN DJF MAM JJA SON
CLDMED
ANN DJF MAM JJA SON
CLDTOT
ANN DJF MAM JJA SON

No category yet

FLDS
ANN DJF MAM JJA SON
FSUTOA
ANN DJF MAM JJA SON
TMQ
ANN DJF MAM JJA SON

- B2000-120km
- B2000-60km

MPAS-Analysis (ocean/ice)

MPAS-Analysis Diagnostics

Run: MPAS_aos60_analysis

Components

- B2000-120km
- B2000-60km
- B2000-30km
- DATM-120km
- DATM-60km

CVDP package (climate variability)

NCAR UCAR | CGD's Climate Analysis Section
Climate Variability Diagnostics Package

Methodology and Definitions
Metrics Tables: Pattern Correlations | RMS
Namelists: Input | Derived
Climatological Period Used: Full
Created: Thu Jun 2 12:55:02 MDT 2022
CVDP Version 5.2.1_beta

mpas_aos60

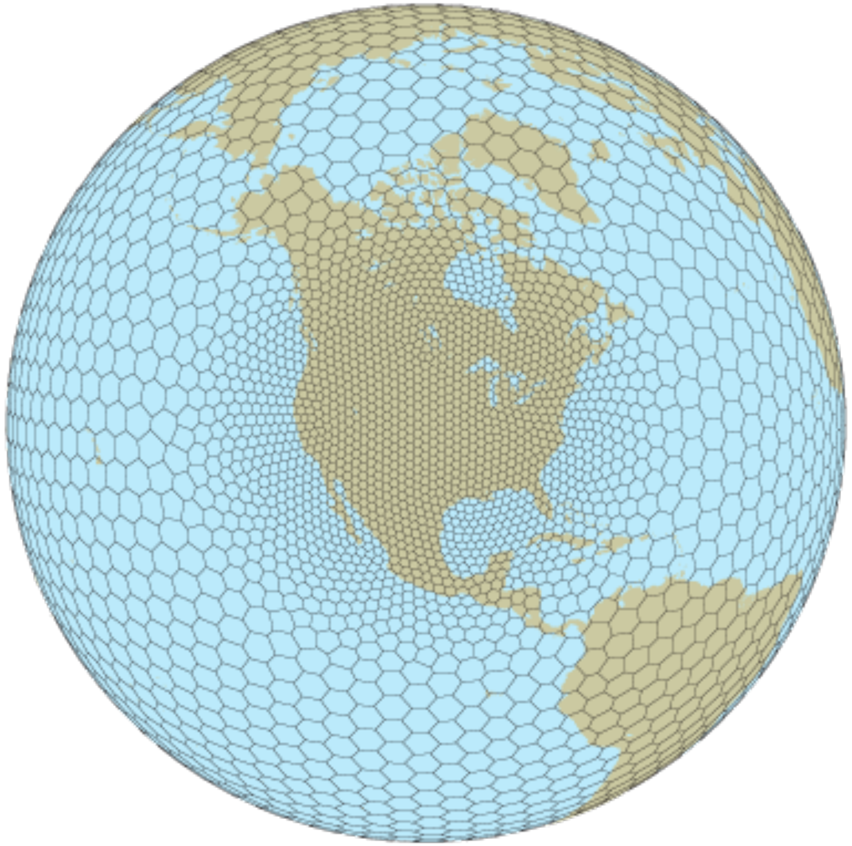
Means

SST	DJF	MAM	JJA	SON	Annual
TAS	DJF	MAM	JJA	SON	Annual
PSL	DJF	MAM	JJA	SON	Annual
PR	DJF	MAM	JJA	SON	Annual

- FHIST-240km, 120km, 60km
- B2000-120km, 60km, 30km, and FHIST-30km

Data Analysis - Project Raijin

Community Geoscience Analysis Tools for Unstructured Mesh Data



Scalable Python tools for analyzing and plotting geoscience data on unstructured grids

★ Climate and global weather modeling communities focus (also works with regional data)

★ Generalizes NCAR's GeoCAT analysis package to support unstructured mesh data

★ Builds on the ubiquitous Xarray package

For more information, see Clyne et al. poster at 2022 EarthCube Annual Meeting

This work supported through the National Science Foundation Award 2126458.