Climate modeling and the challenges of exascale

Robert Jacob
Mathematics and Computer Science Division
Argonne National Laboratory

November 19, 2012

An old saying….
"Climate is what you expect, weather is what you get"

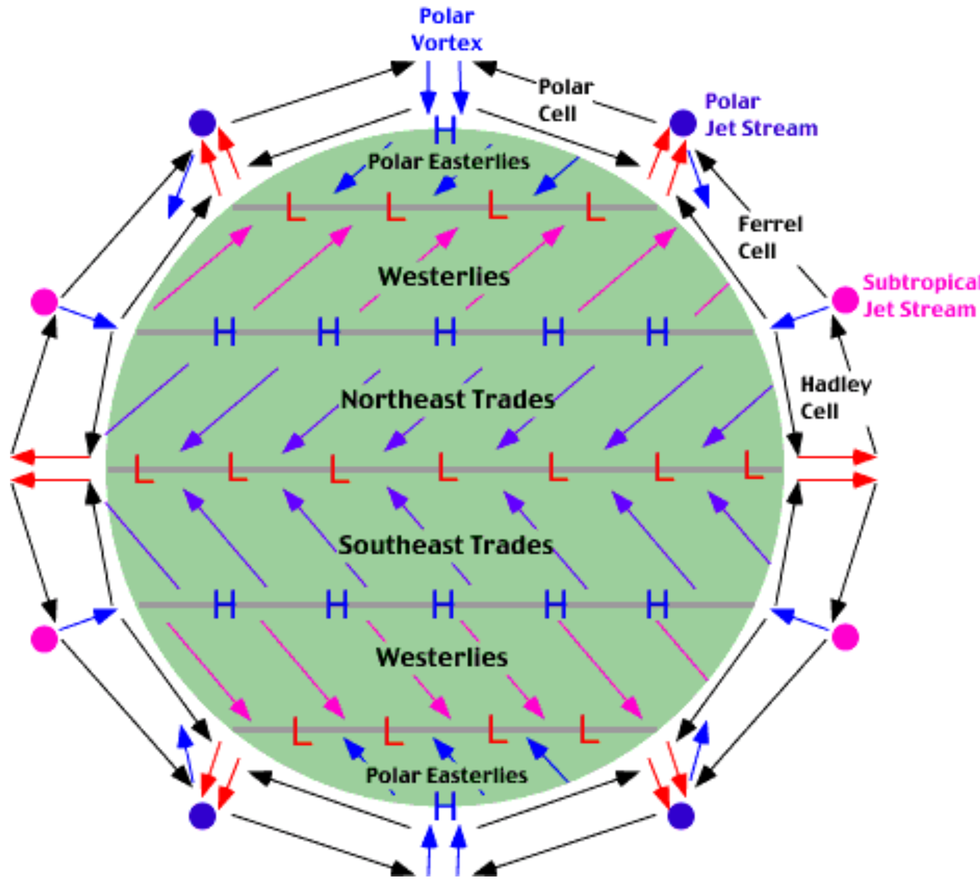
- Climate is the *average* of weather.
- The (predicted) high temperature today, Nov 19th , is 57F
- The average high temperature is 46F.   This is calculated by taking the average of several (usually 30) Nov 19th highs.

$$\frac{(T\ \text{Nov}\ 19^{th}, 1981) + (T\ \text{Nov}\ 19^{th}, 1982) + \ldots + (T\ \text{Nov}\ 19^{th}, 2010)}{30}$$

From NWS site: "Please note, as of forecast May 2011, the climatological reference period has been updated from 1971-2000 to 1981-2010"

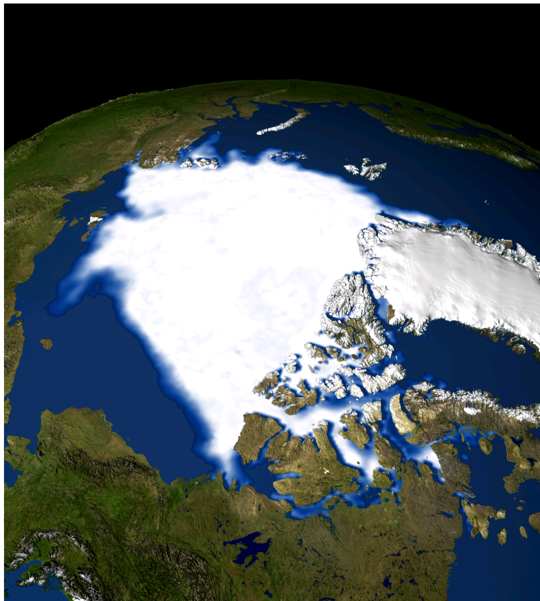
To model the climate system, must model years of global weather

Need to simulate weather-scale phenomena over the entire globe.

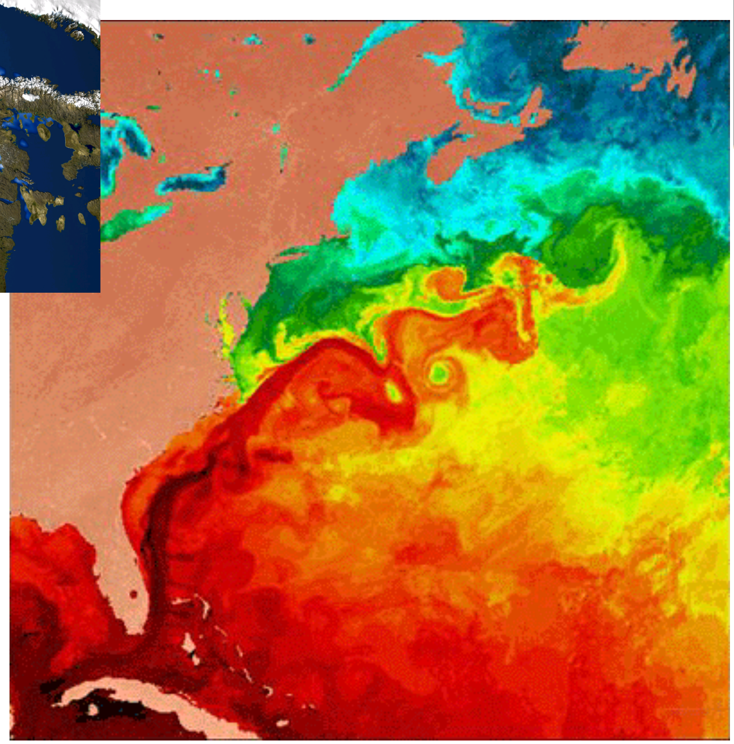


Weather is embedded in the *general circulation* of the atmosphere

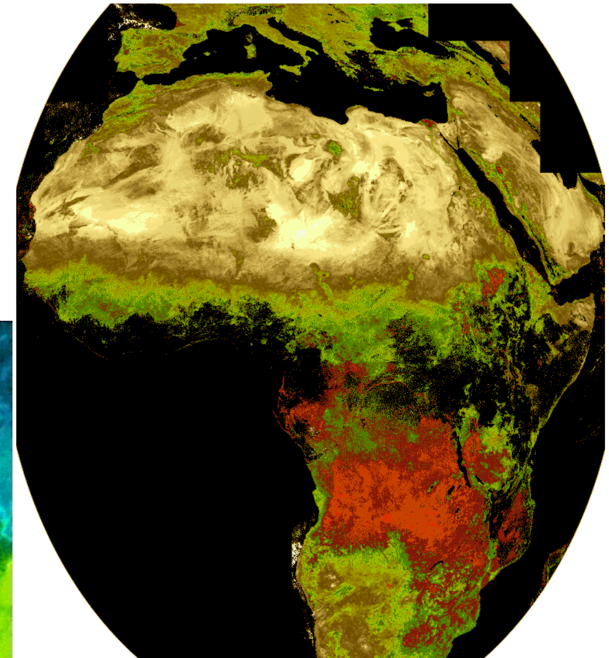
Over may days, months, atmosphere circulation is dominated by interaction with surface.



Sea Ice



Ocean



Land

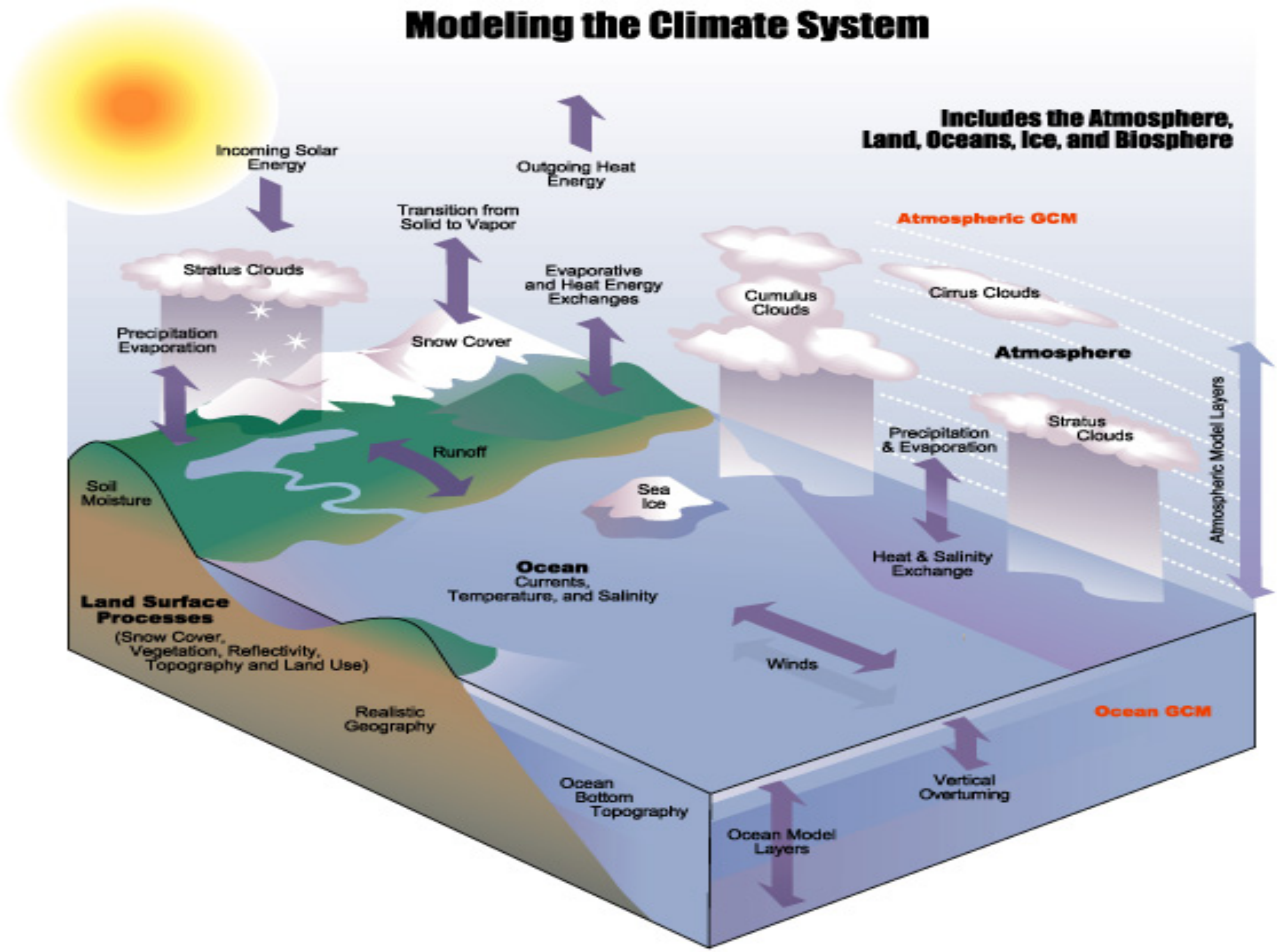


Figure: U.S. Global Change Research Program

GCM: General Circulation Model

- Solves the "primitive equations", a set of non-linear PDEs which ultimately derive from the Navier-Stokes equations.

- Fundamental properties of geophysical fluids:
  - Fluid is rotating
  - Fluid is on a sphere
  - Fluid is acted upon by gravity

- Assumptions:
  - Thin Stratified Fluid
  - Hydrostatic
  - Anelastic and Boussinesq (no sound waves, small aspect ratio, motions are shallow)

- Derived in a non-inertial reference frame rotating with the Earth

The “Primitive equations” in spherical coordinates

momentum equations:

Coriolis force

$$\frac{\partial}{\partial t}u + \mathcal{L}(u) - (uv\tan\phi)/a - fv = -\frac{1}{\rho_0 a\cos\phi}\frac{\partial p}{\partial \lambda} + \mathcal{F}_{Hx}(u,v) + \mathcal{F}_V(u)$$

$$\frac{\partial}{\partial t}v + \mathcal{L}(v) + (u^2\tan\phi)/a + fu = -\frac{1}{\rho_0 a}\frac{\partial p}{\partial \phi} + \mathcal{F}_{Hy}(u,v) + \mathcal{F}_V(v)$$

Advection

$$\mathcal{L}(\alpha) = \frac{1}{a\cos\phi}\left[\frac{\partial}{\partial\lambda}(u\alpha) + \frac{\partial}{\partial\phi}(\cos\phi\, v\alpha)\right] + \frac{\partial}{\partial z}(w\alpha)$$

Horizontal Friction

$$\mathcal{F}_{Hx}(u,v) = A_M\left\{\nabla^2 u + u(1-\tan^2\phi)/a^2 - \frac{2\sin\phi}{a^2\cos^2\phi}\frac{\partial v}{\partial\lambda}\right\}$$

$$\mathcal{F}_{Hy}(u,v) = A_M\left\{\nabla^2 v + v(1-\tan^2\phi)/a^2 + \frac{2\sin\phi}{a^2\cos^2\phi}\frac{\partial u}{\partial\lambda}\right\}$$

$$\nabla^2\alpha = \frac{1}{a^2\cos^2\phi}\frac{\partial^2\alpha}{\partial\lambda^2} + \frac{1}{a^2\cos\phi}\frac{\partial}{\partial\phi}\left(\cos\phi\frac{\partial\alpha}{\partial\phi}\right)$$

Vertical Friction

$$\mathcal{F}_V(\alpha) = \frac{\partial}{\partial z}\mu\frac{\partial}{\partial z}\alpha$$

The "Primitive equations" continued

continuity equation:

$$\mathcal{L}(1) = 0$$

hydrostatic equation:

$$\frac{\partial p}{\partial z} = -\rho g$$

equation of state:

$$\rho = \rho(\Theta, S, p) \;\rightarrow\; \rho(\Theta, S, z)$$ **(Ocean)**

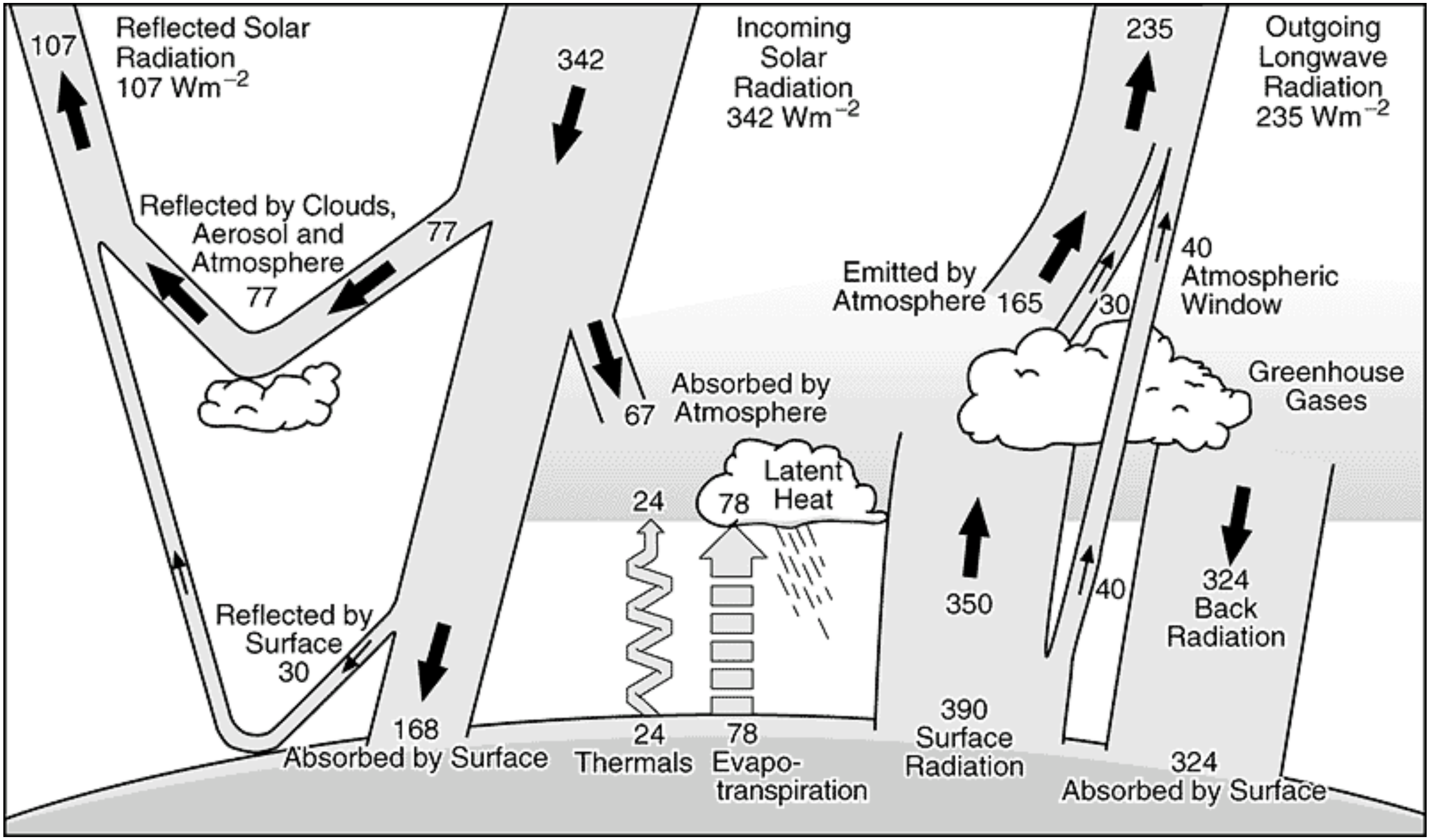
tracer transport:

$$\frac{\partial}{\partial t}\varphi + \mathcal{L}(\varphi) = \mathcal{D}_H(\varphi) + \mathcal{D}_V(\varphi)$$ **+ F(t,u,v,phi)**

$$\mathcal{D}_H(\varphi) = A_H \nabla^2 \varphi$$

$$\mathcal{D}_V(\varphi) = \frac{\partial}{\partial z}\kappa\frac{\partial}{\partial z}\varphi \ ,$$

Heat forcing on the atmosphere: Radiation and other. F(t,u,v,phi)



Longwave radiative flux in the 500-1500 cm^{-1} band.

$$\int_{500}^{1500} (1-\mathcal{T}_\nu)F(B_\nu)d\nu = \int_{500}^{750} (1-\mathcal{T}^1_{CO_2}\mathcal{T}^1_{N_2O}\mathcal{T}_{H_2O}\mathcal{T}^1_{H_2SO_4})F(B_\nu)d\nu$$

$$+\int_{750}^{820} (1-\mathcal{T}^1_{CFC11}\mathcal{T}_{H_2O}\mathcal{T}^*_{H_2SO_4})F(B_\nu)d\nu$$

F(B) : Plank function for emissivity
T: atmospheric transmission

$$+\int_{820}^{880} (1-\mathcal{T}^2_{CFC11}\mathcal{T}_{H_2O}\mathcal{T}^3_{H_2SO_4})F(B_\nu)d\nu$$

$$+\int_{880}^{900} (1-\mathcal{T}^1_{CFC12}\mathcal{T}_{H_2O}\mathcal{T}^3_{H_2SO_4})F(B_\nu)d\nu$$

$$+\int_{900}^{1000} (1-\mathcal{T}^2_{CO_2}\mathcal{T}_{H_2O}\mathcal{T}^3_{H_2SO_4}\mathcal{T}^3_{CFC11}\mathcal{T}^2_{CFC12})F(B_\nu)d\nu$$

$$+\int_{1000}^{1120} (1-\mathcal{T}^3_{CO_2}\mathcal{T}_{O_3}\mathcal{T}_{H_2O}\mathcal{T}^4_{H_2SO_4}\mathcal{T}^4_{CFC11}\mathcal{T}^3_{CFC12})F(B_\nu)d\nu$$

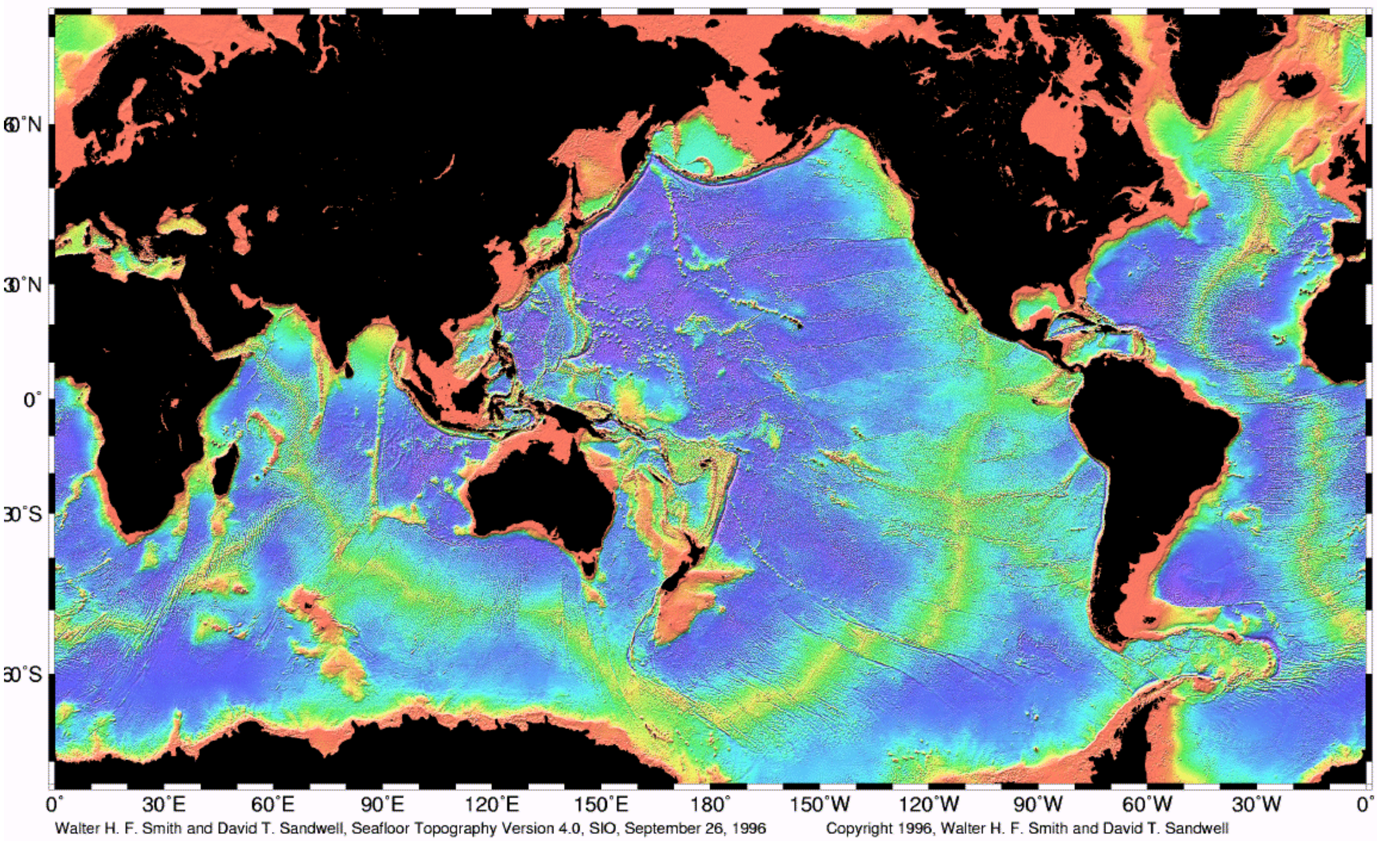
$$+\int_{1120}^{1170} (1-\mathcal{T}^4_{CFC12}\mathcal{T}_{H_2O}\mathcal{T}^4_{H_2SO_4}\mathcal{T}^2_{N_2O})F(B_\nu)d\nu$$

$$+\int_{1170}^{1500} (1-\mathcal{T}_{CH_4}\mathcal{T}^3_{N_2O}\mathcal{T}_{H_2O}\mathcal{T}^5_{H_2SO_4})F(B_\nu)d\nu$$

Atmospheric General Circulation Model

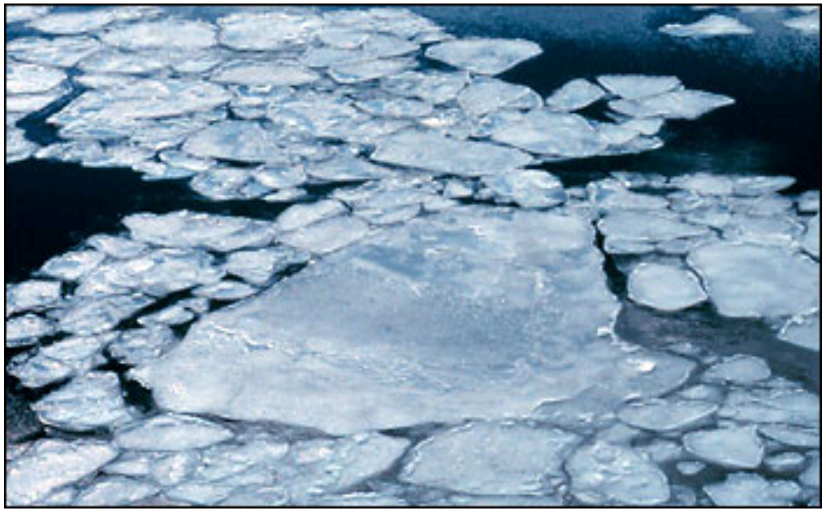
- Algorithms to solve the primitive equations called "the dynamics"; "dynamical core" "dycore"
- Forcing terms: **F(t,u,v,phi)**
    - *Change in temperature due to radiative transfer*
    - *Effect of clouds on radiative transfer*
    - *Change in moisture due to cloud, rain formation*
    - *Change in temperature due to sensible heat transport through the boundary layer*
    - *Change in temperature due to release of latent heat*
    - *Change in momentum due to friction with surface.*
  - Algorithms for the above called "the physics" or "column physics".
  - Major groupings: longwave radiation, shortwave radiation, boundary layer, deep convection, cloud fraction, gravity wave drag.
  - ***Can take as much or more computer time as the dynamics and also dominate the source code.***

Ocean General Circulation Model



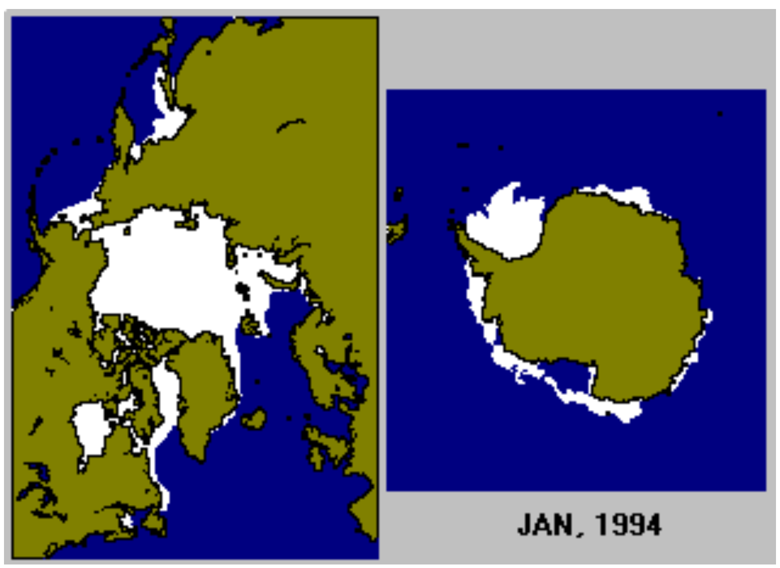
- Very Similar to AGCM except:
  - Presence of side boundaries. Nearly all OGCM's are FD with z-coordinates.
  - Not as much "physics"
  - Motions are slower. Length scales are shorter.
  - Much higher heat capacity. The memory of the climate system is in the ocean.

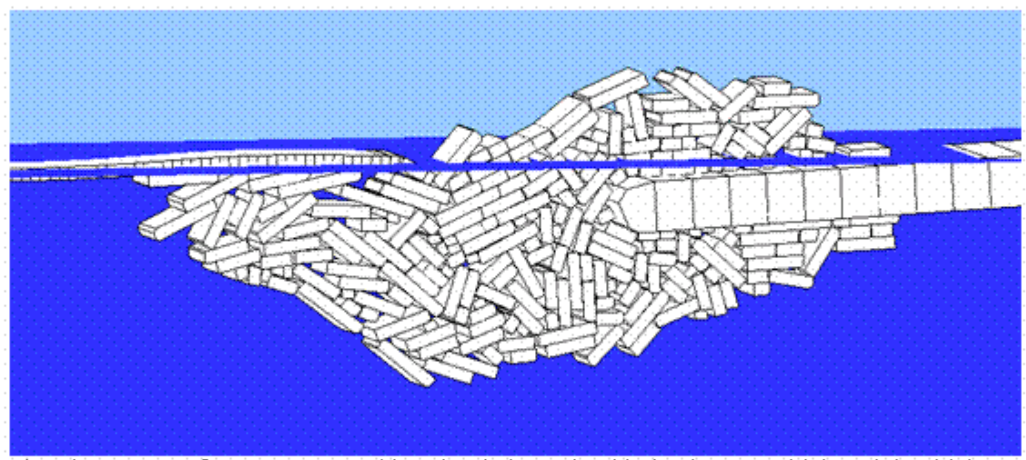
Sea Ice Models



- Thermodynamics: formation, growth, melting, **albedo**, melt ponds.

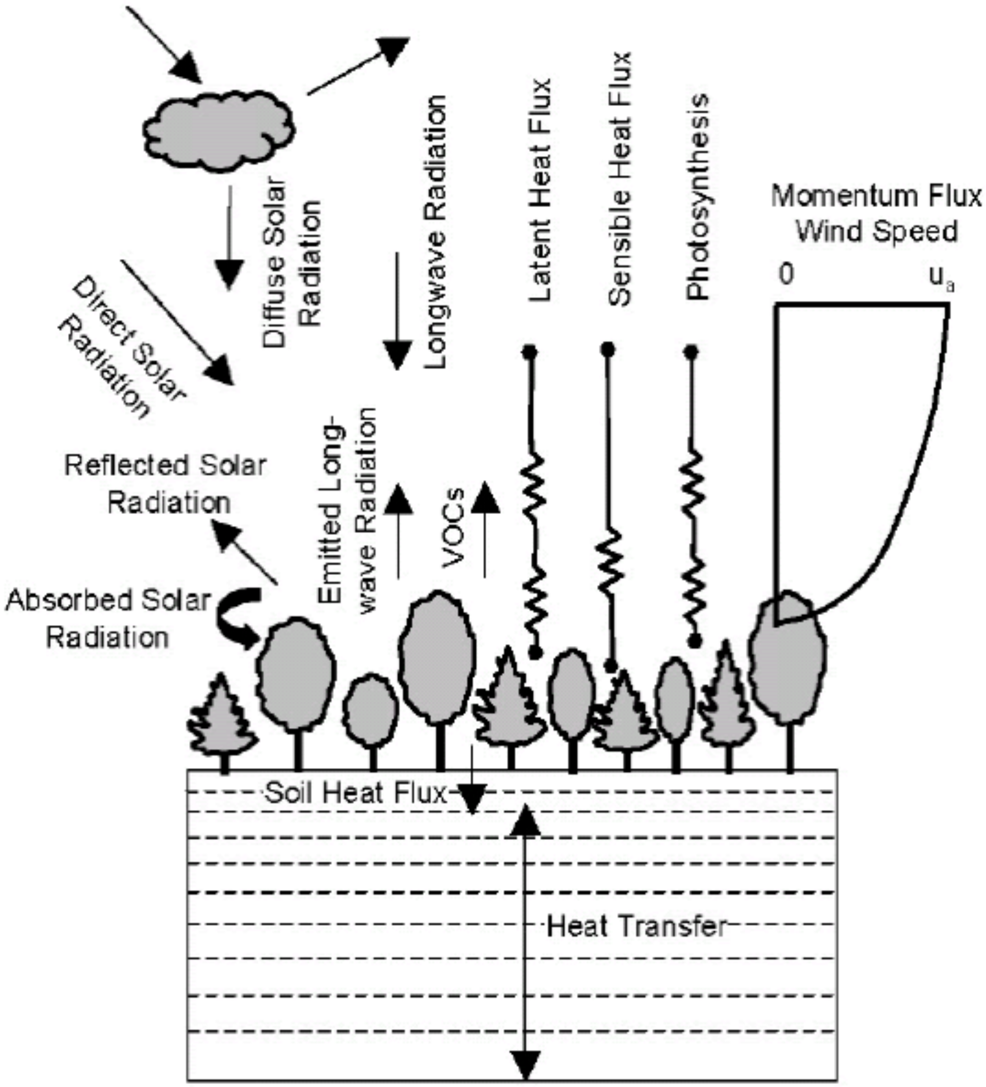
- Dynamics: transport, internal stress, ridging





Showing a scene from a pressure ridge simulation. The thin ice is 0.5 m thick and the thick floe is 2 m thick.

Land Surface Models



- Nearly all "physics":
  - Vegetation composition, structure
  - Vertical heat transfer in soil.
  - Heat, radiation transfer between ground, canopy and free atmosphere
  - Hydrology of canopy, snow, soil moisture
  - River runoff
- Historically, was part of column physics in the atmosphere model.

Figure: Bonan, NCAR

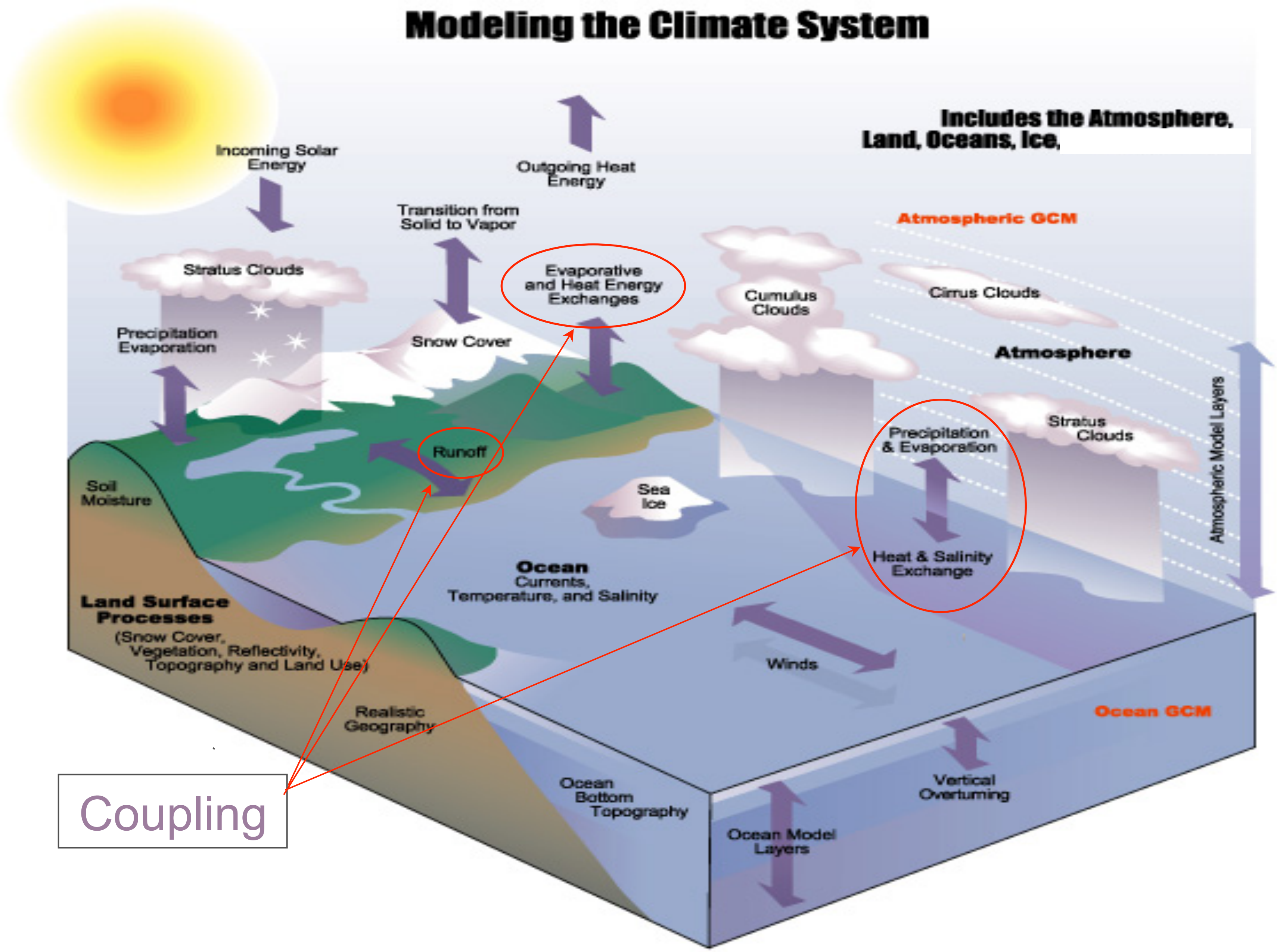
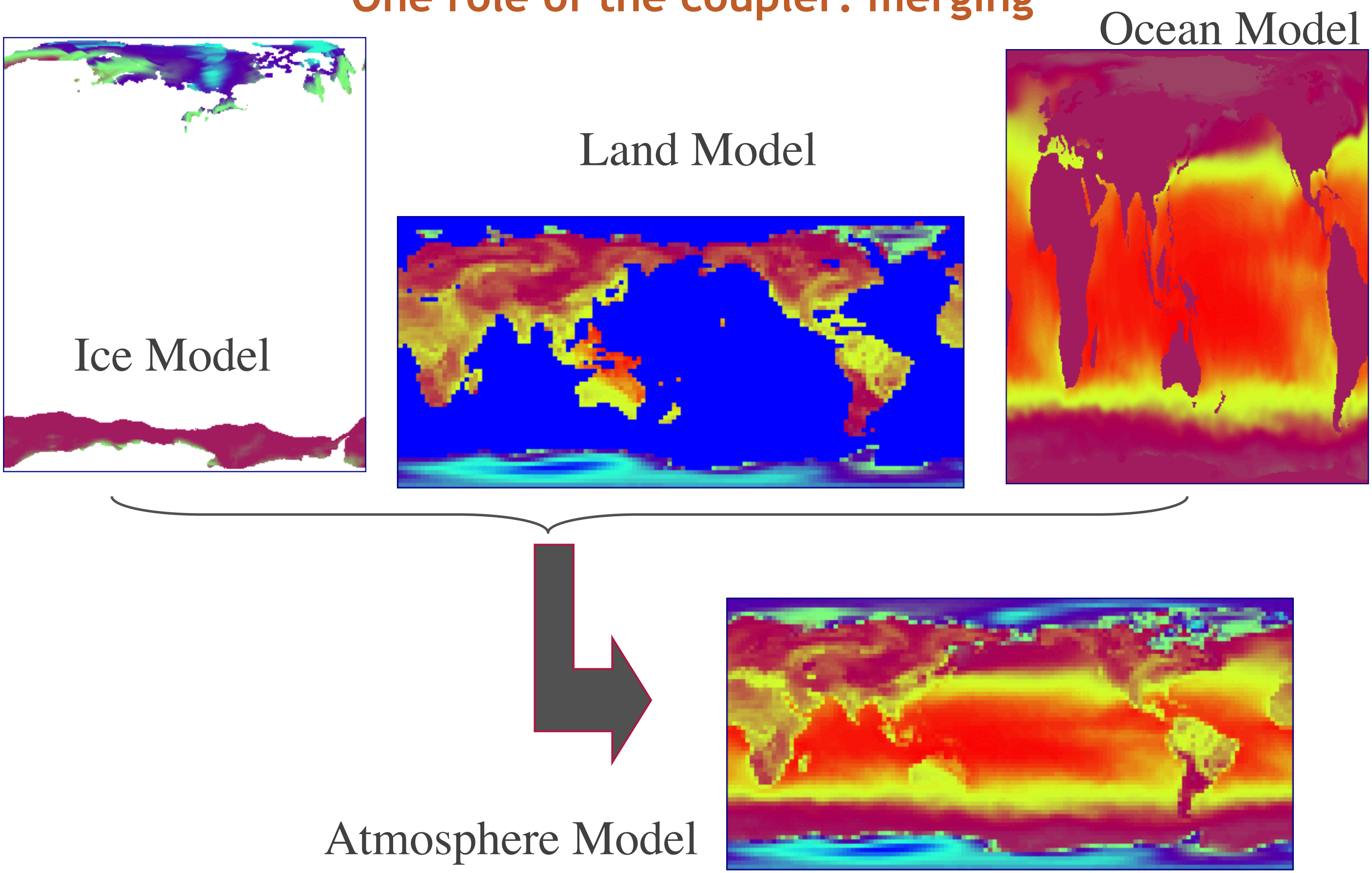


Figure: U.S. Global Change Research Program

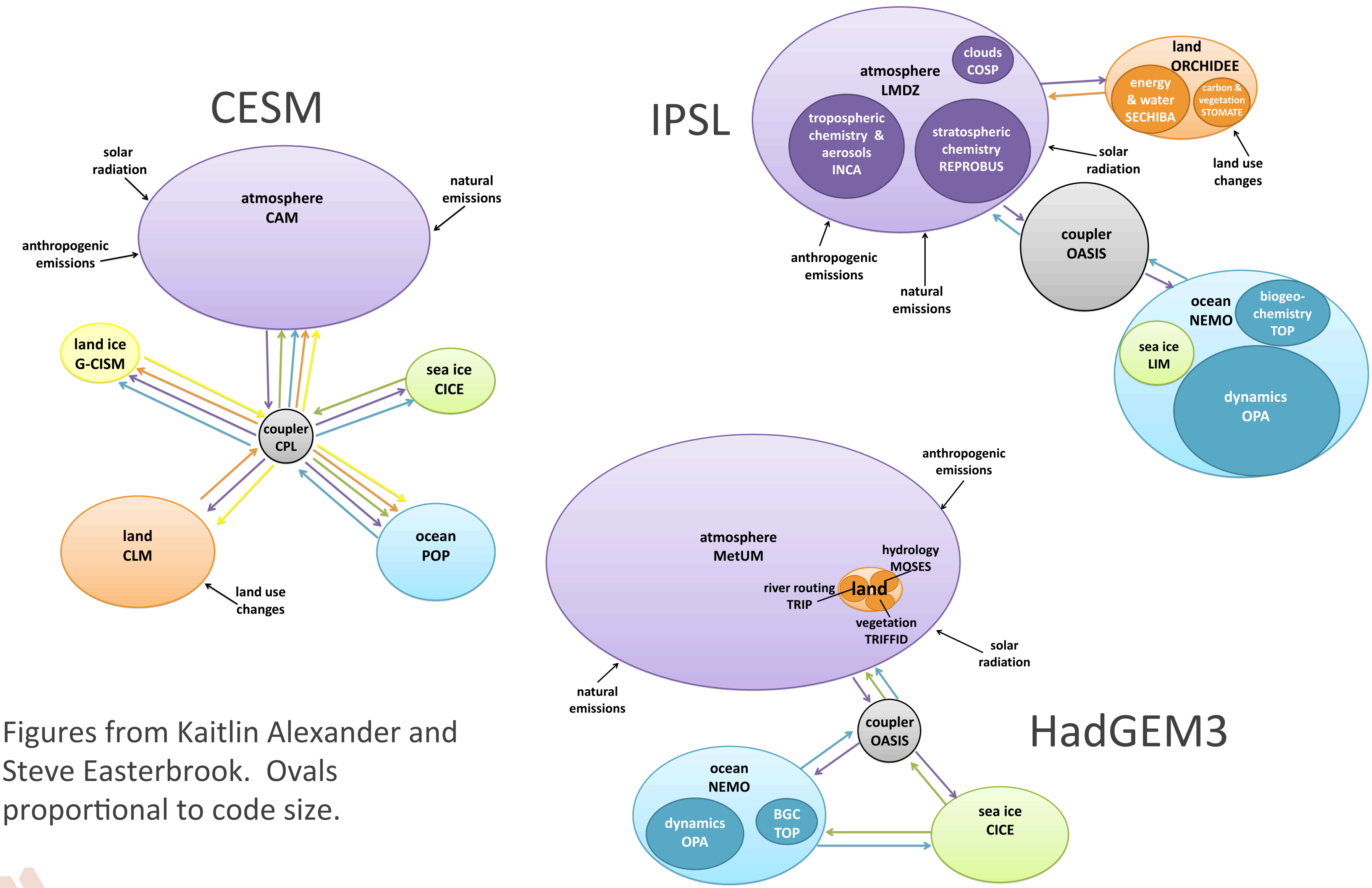
One role of the coupler: merging



The Model Coupling Toolkit: Software for building gridded multi-physics models.

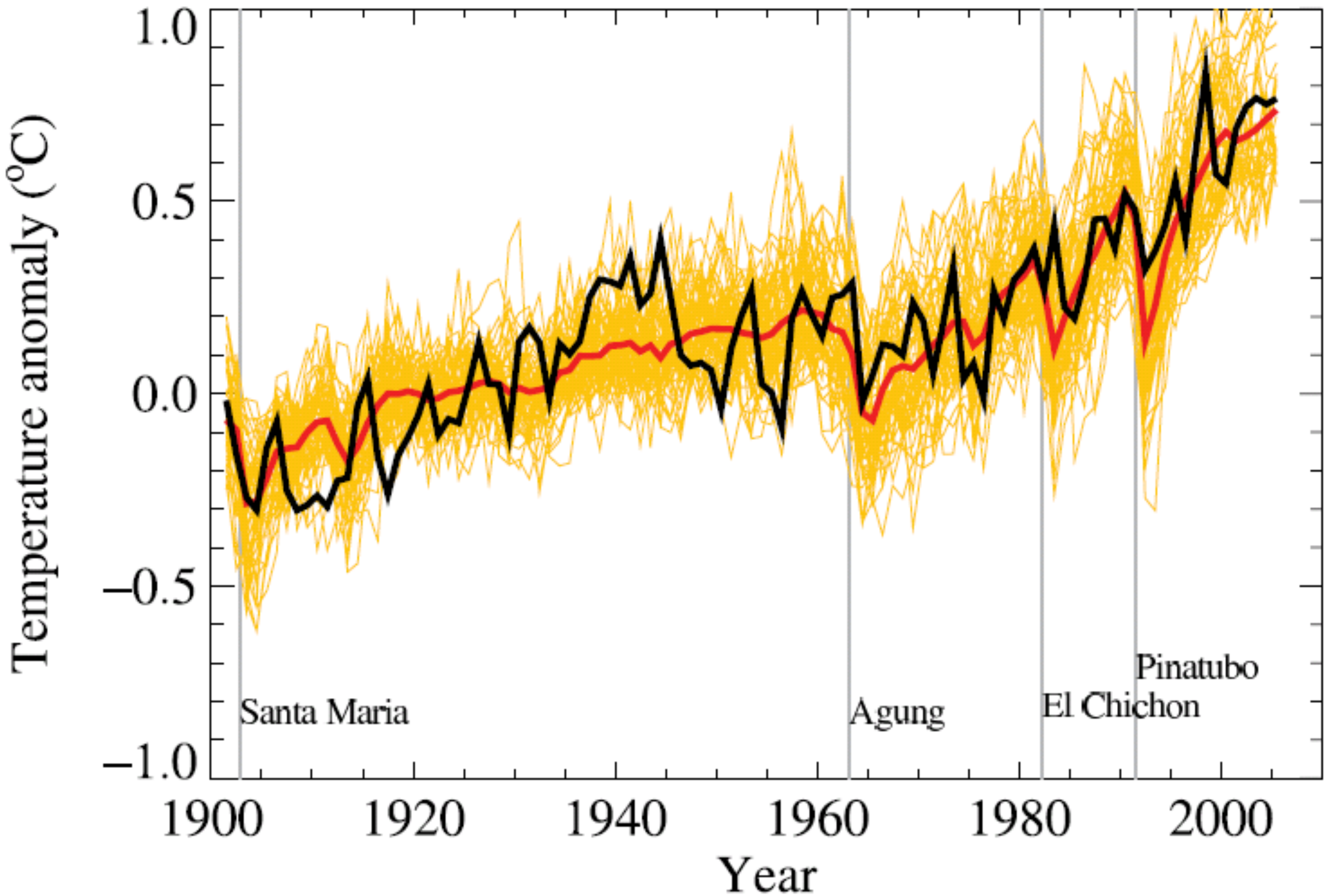
- Main coupling framework in the NSF/DOE Community Earth System Model
- Developed at Argonne
- OASIS-MCT - released in August! Informal collaboration with CERFACS.

Climate model construction



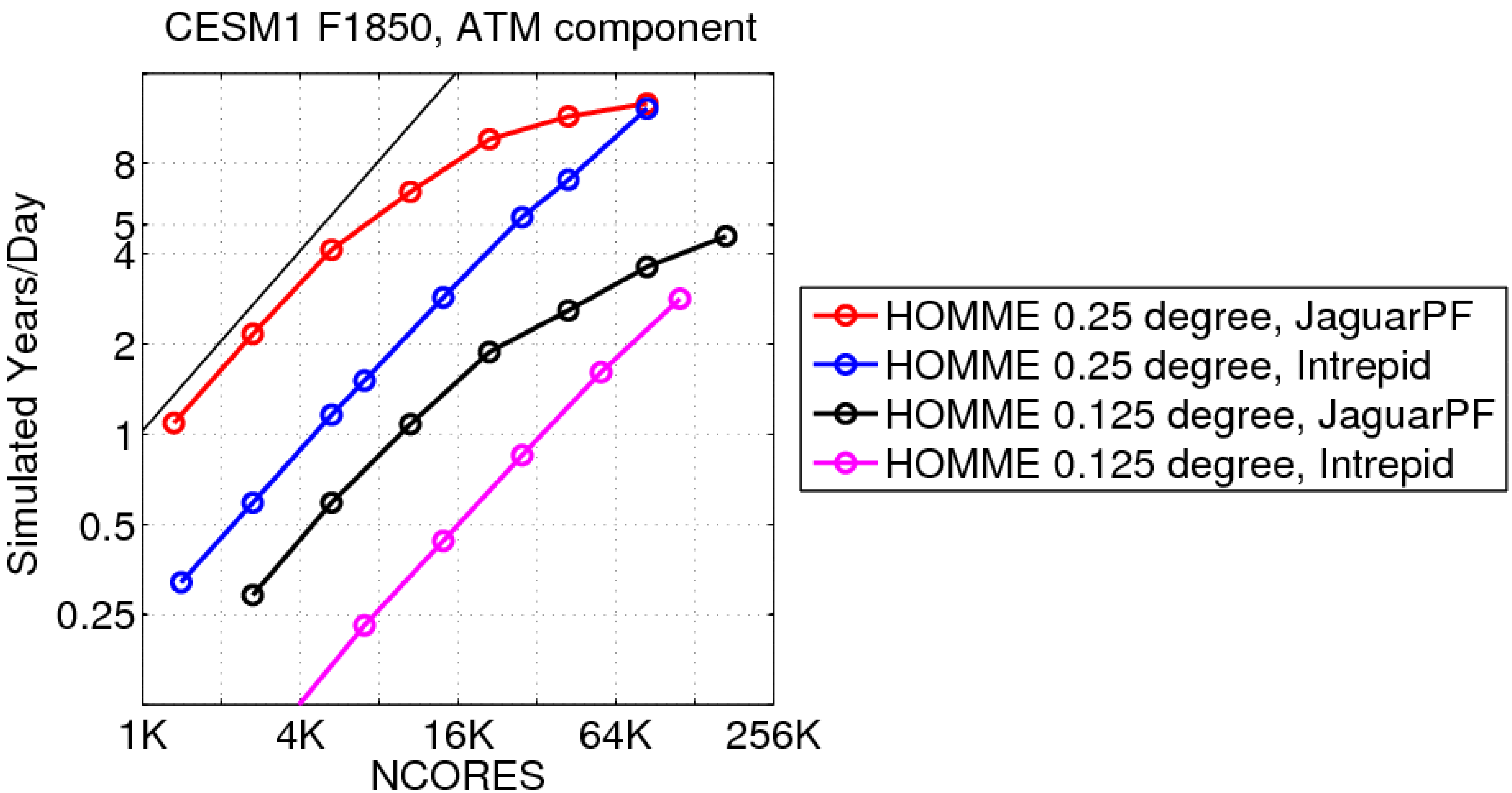
Figures from Kaitlin Alexander and Steve Easterbrook. Ovals proportional to code size.

Multi-model simulations of the 20th Century (IPCC AR4)



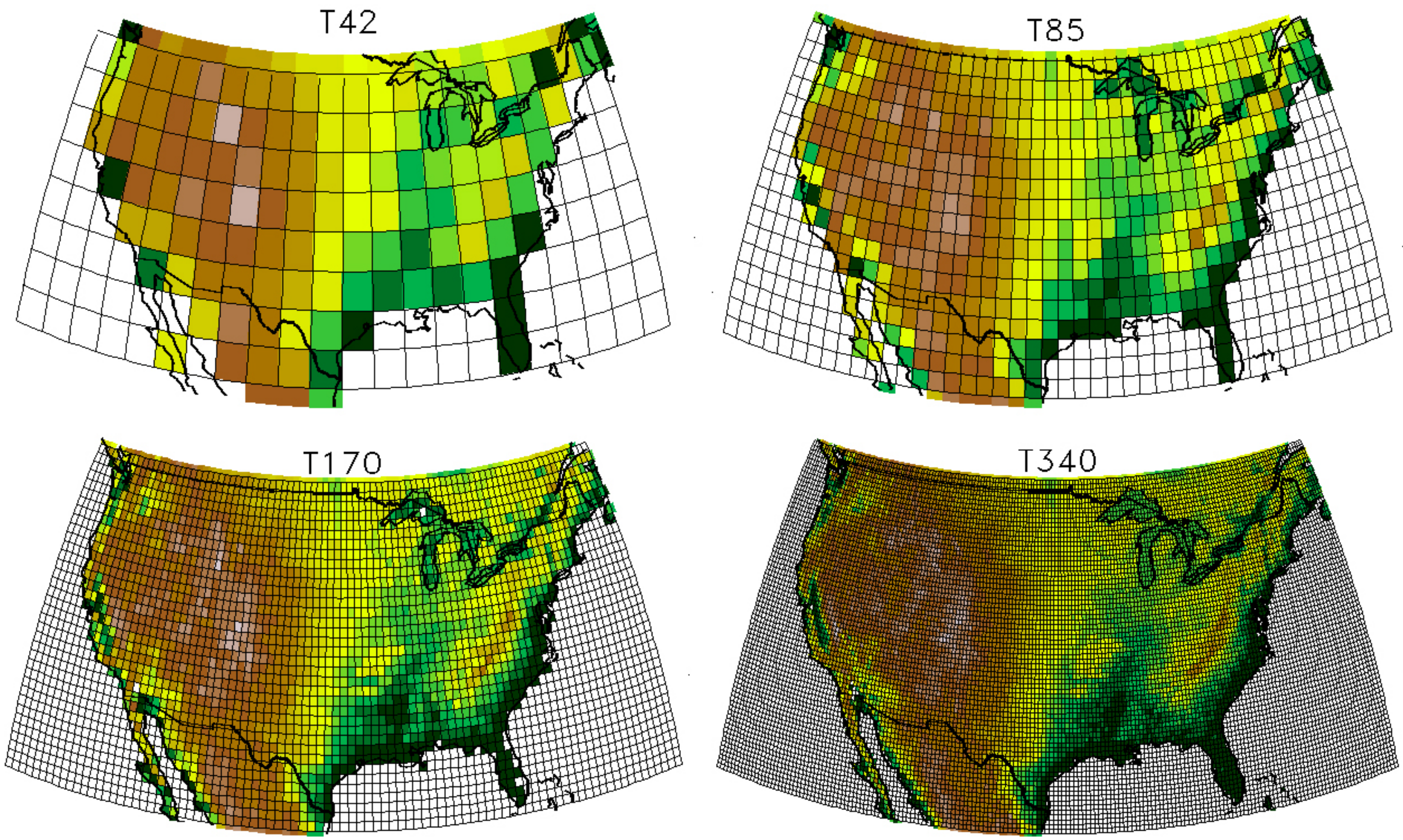
Challenges for Climate Modeling at Exascale

We can run climate model components on 100K cores. How to we get to 1 Billion?



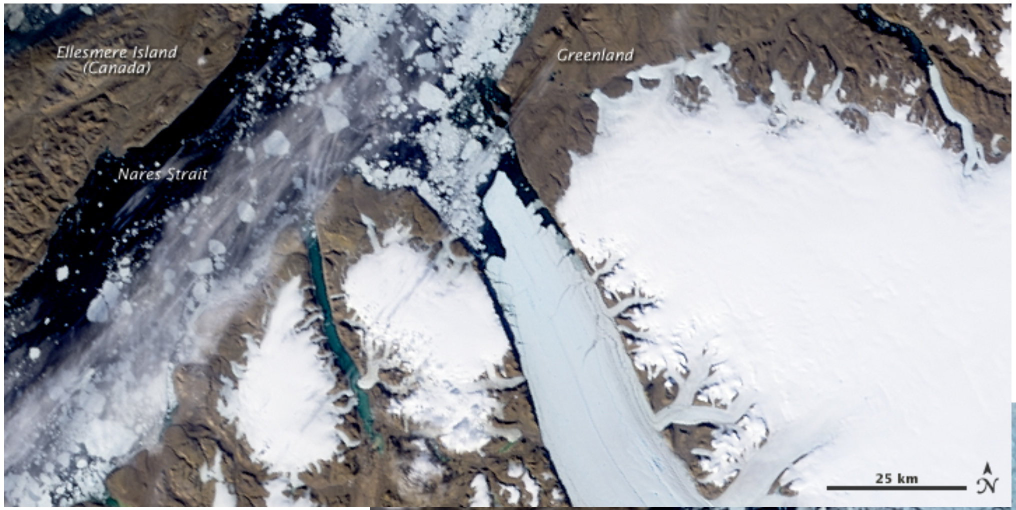
From Mark Taylor, SNL

Increase the resolution…



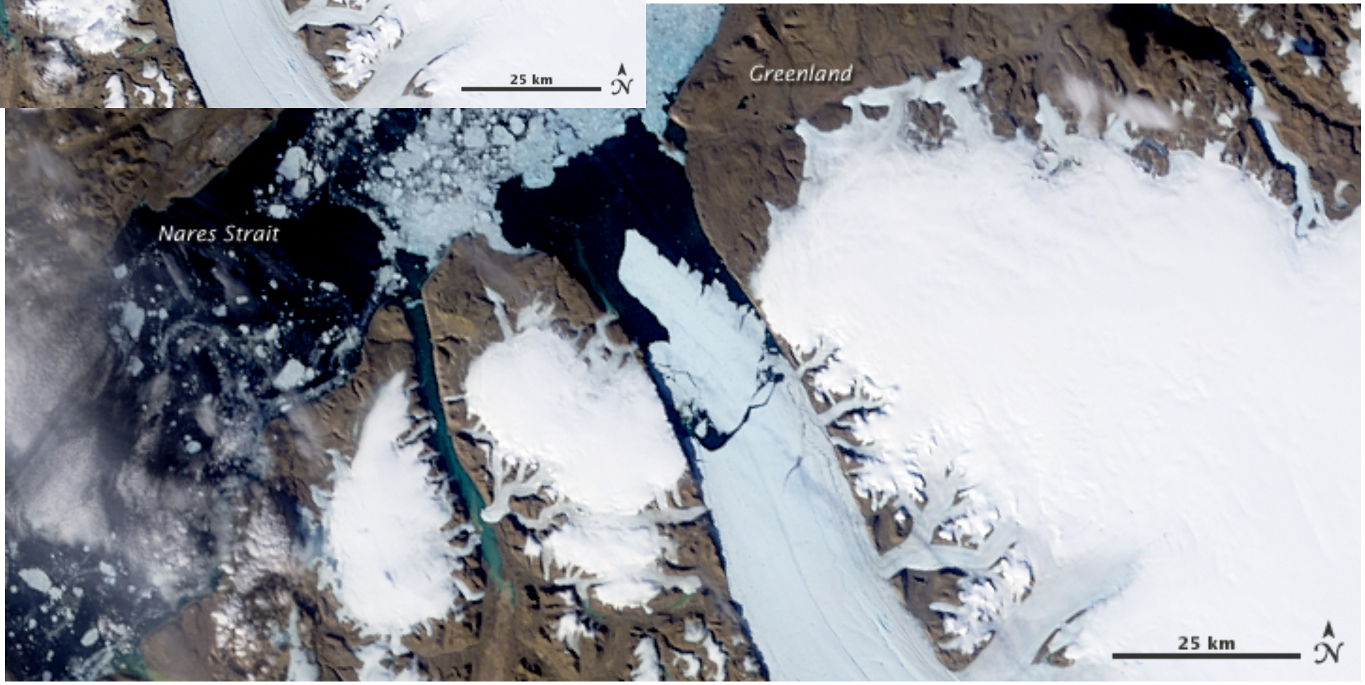
Very small scale features can have global consequences.

July 28, 2010

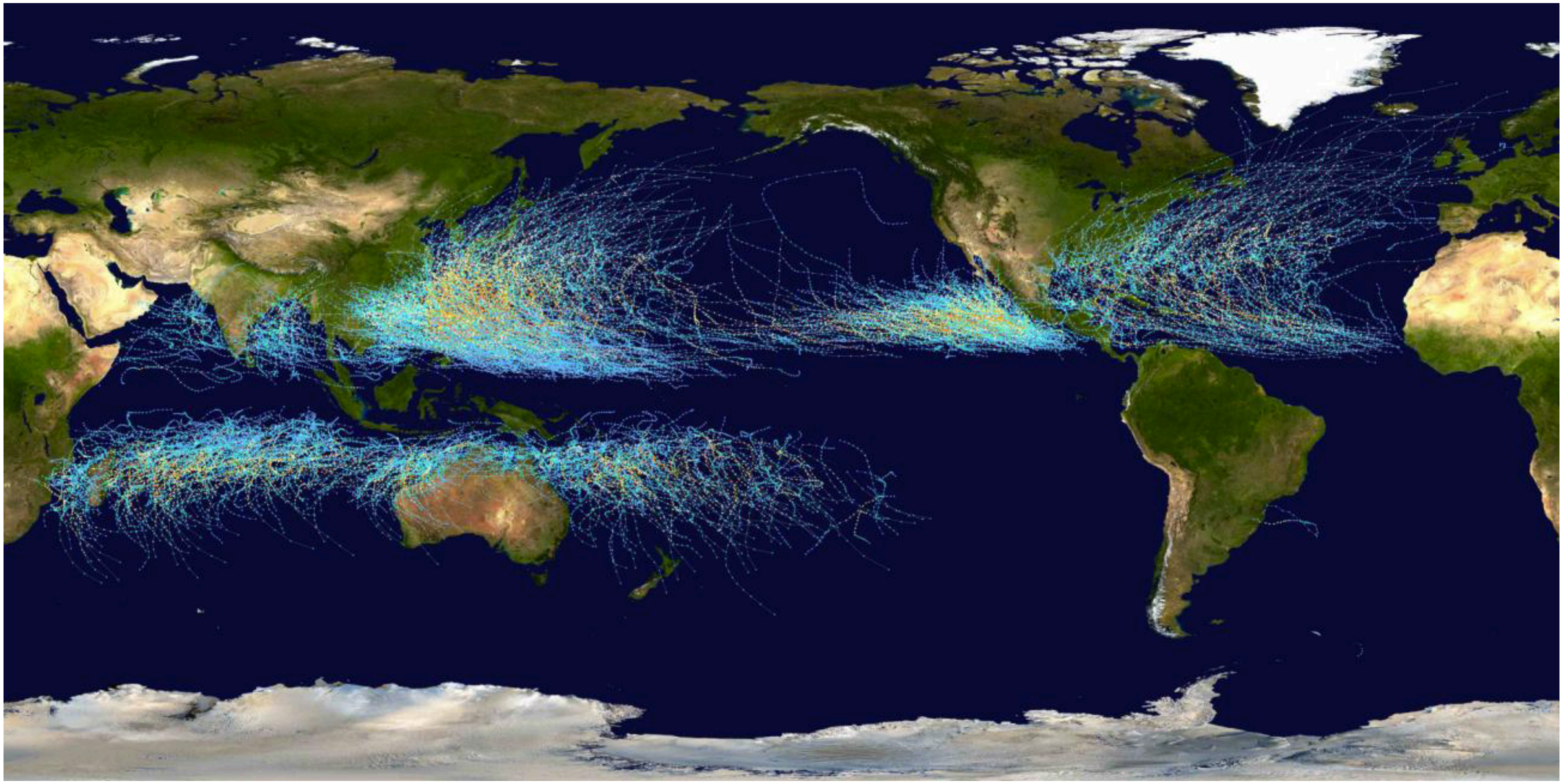


Peterman Glacier, Greenland

August 5, 2010

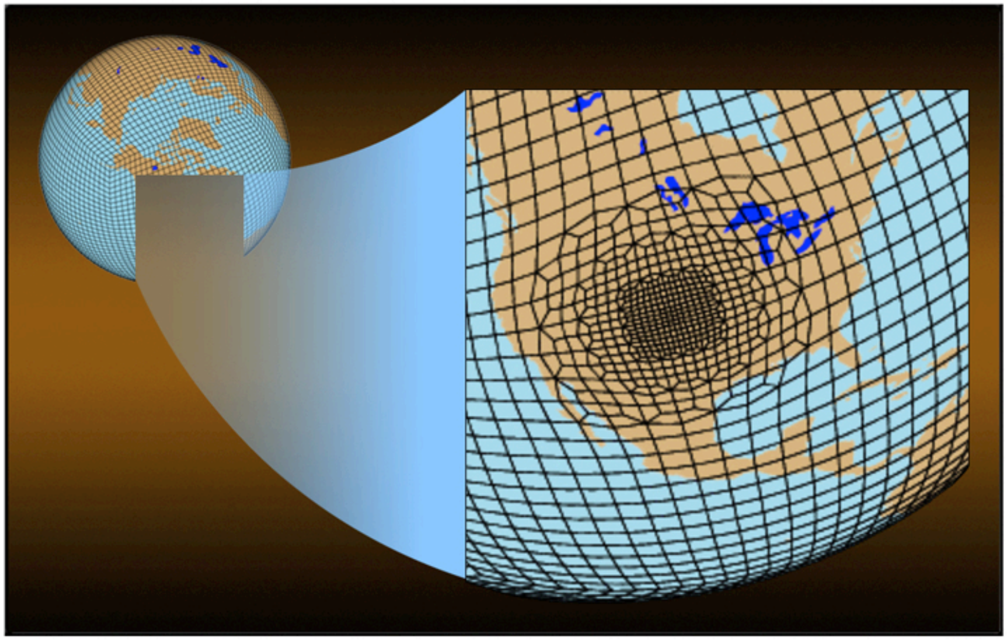


Need high resolution for hurricanes - a vital part of the climate system.

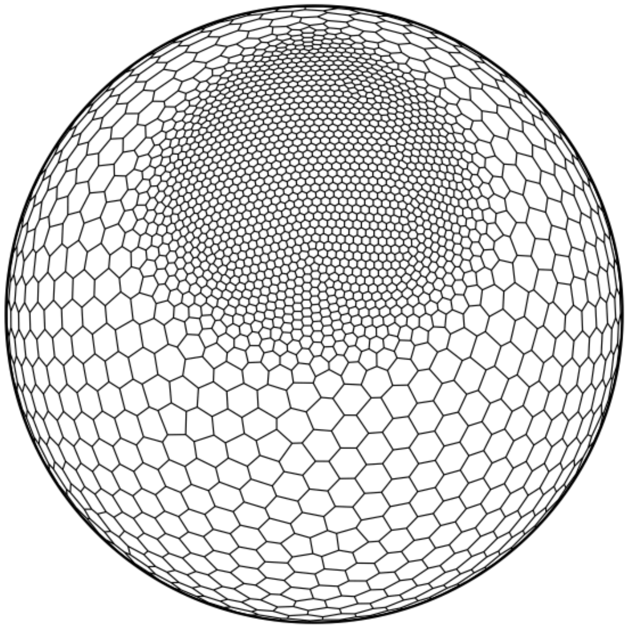


All tropical cyclone tracks 1985-2005. Tracks colored by max wind speed from weak (blue) to red (strong)

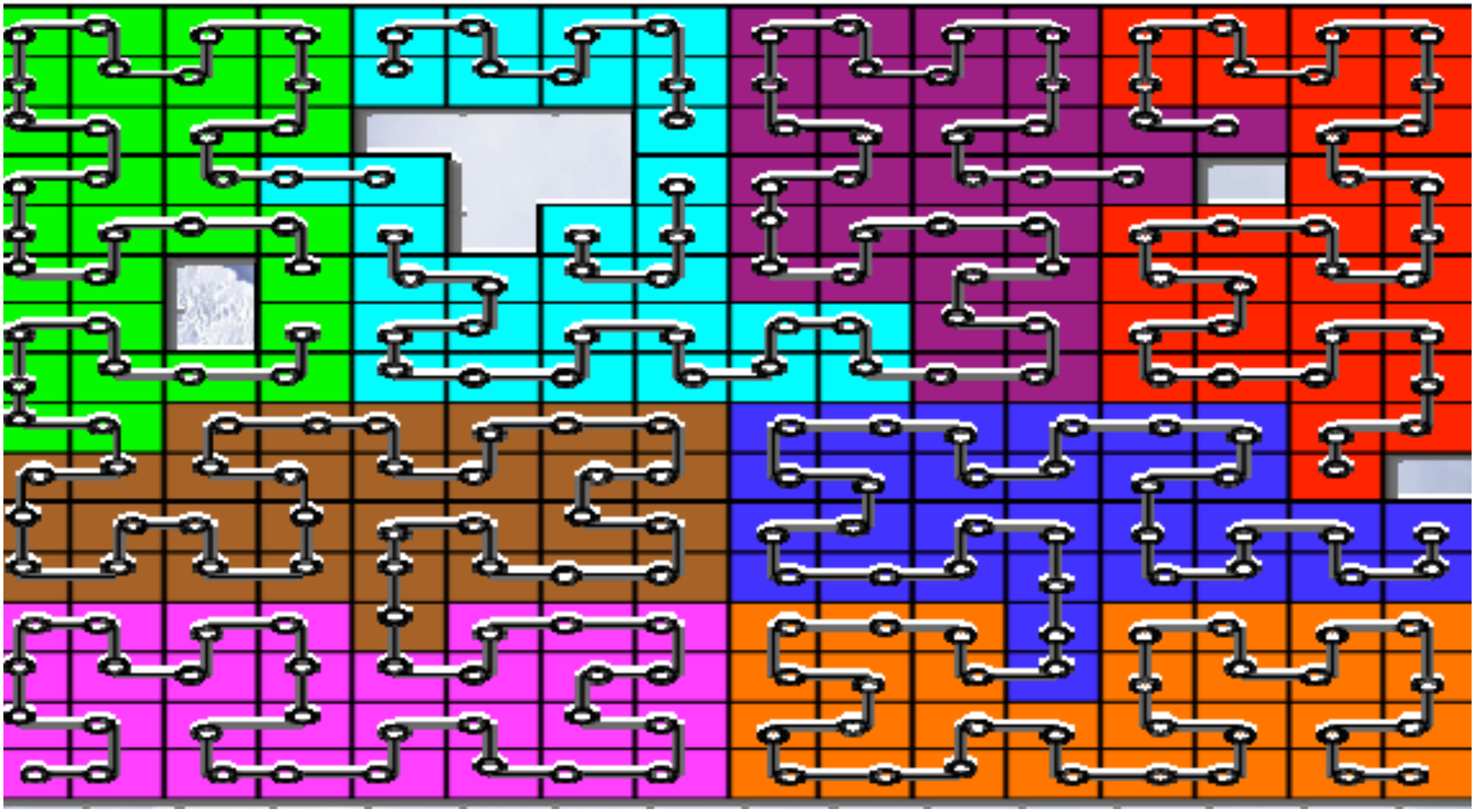
Higher Resolution: Regionally refined grids



Same dynamics throughout



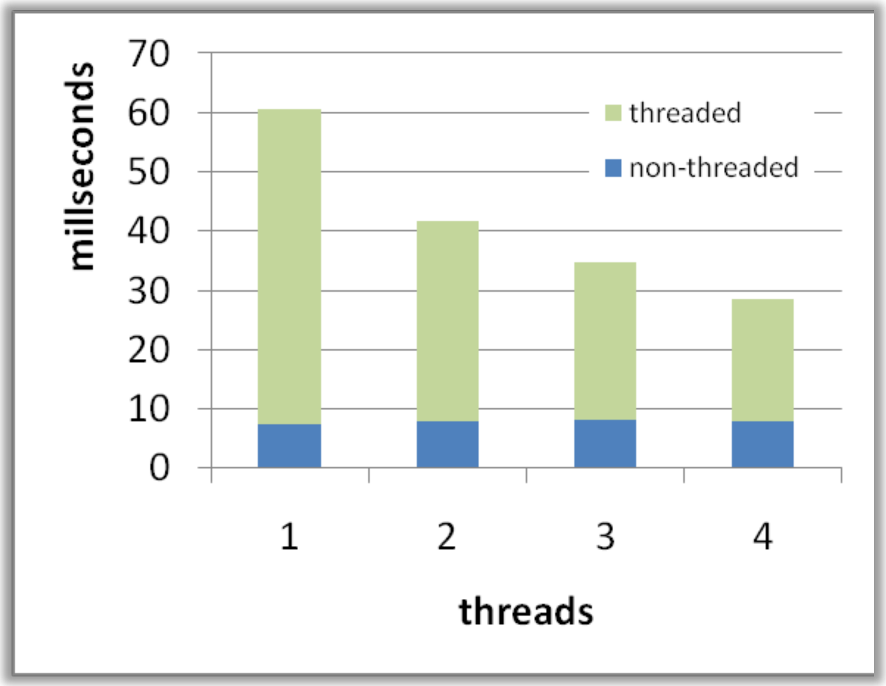
But still limited parallelism if only horizontal (which is what all climate models do.)



Ocean model space-filling curve decomposition

Exploiting parallelism in the vertical dimension

- Problems with vertical parallelism
  - Dynamics generally dependence-free except for implicit solver for W-winds
  - Physics has many vertical dependencies
    - Reductions
    - Recurrences
    - Searches
    - All-to-all dependencies
  - Hopeless?
- Cloud microphysics
  - Computationally expensive
  - Principal source of load imbalance
  - Dependencies
    - Calculating internal time step is a reduction
    - Computing fall speeds is recurrent
  - *The rest is dependency-free in vertical*
  - Initial attempt shows 2.3x on 4 threads



Threading vertical dimension of WSM5 cloud microphysics
(S. Y. Hong et al. Yongsei University)
Dual quad-core Intel 5570 2.93 GHz
Various KMP_AFFINITY granularity settings

Slide from J. Michalakes, NREL

The Data Problem:

- Take your coupled global climate model and calculate evolution of global weather for 100 years, 20 minutes at at a time.
  - CCSM3 (150km): 1 quadrillion operations/simulated year.
- After 100 quadrillion operations, what do you know about the climate?

NOTHING!

Climate is revealed by calculating statistics on "climate" model output

- Averages over time and space.
- Other moments
- More sophisticated analysis: CCA, PCA, etc.

Climate model output practices

- Since running a model is very expensive AND
- Since the science comes from analyzing the output.

- Output everything!
  - Prognostic state variables
  - Derived quantities
  - Approximately 100 different variables. 25% 3D, rest 2D or 1D.

- ….But don't save everything for all times
  - Monthly output of all variables.
  - Daily or 4-hourly output of some of the same variables.

With typical climate model data sizes, not a problem.

- Atmosphere Model (single output file of all variables, one time step)
  - 1 degree: 233MB
  - 0.5 degree: 821MB
- Ocean Model
  - 3 degree: 20 MB
  - 1 degree: 1.1 GB
- Sea Ice Model
  - 1 degree: 69 MB
- Land Model
  - 1 degree: 86 MB

Near future climate model data sizes

CAM-SE 0.125 degrees

```
single 3D variable    616MB (real*8)
single 2D variable     25MB (real*8)
total grid points per 3D variable:  3110402 x 26 (80M points)
single history file    24GB

1 year:                392 GB

100 years:            39.2 TB
```

POP 0.1 degrees

```
single 3D variable     1.45  GB (4 byte reals)
single history file    18.94 GB
single restart file    24.19 GB

1 year:      227 GB

100 years:  22.7 TB
```

The Global Cloud Resolving Model (GCRM) Output Tsunami

4 km, 100 levels, hourly data

~1 TB / simulated hour

~24 TB / simulated day

~9 PB / simulated year

2 km, 100 levels, hourly data

~4 TB / simulated hour

~100 TB / simulated day

~35 PB / simulated year

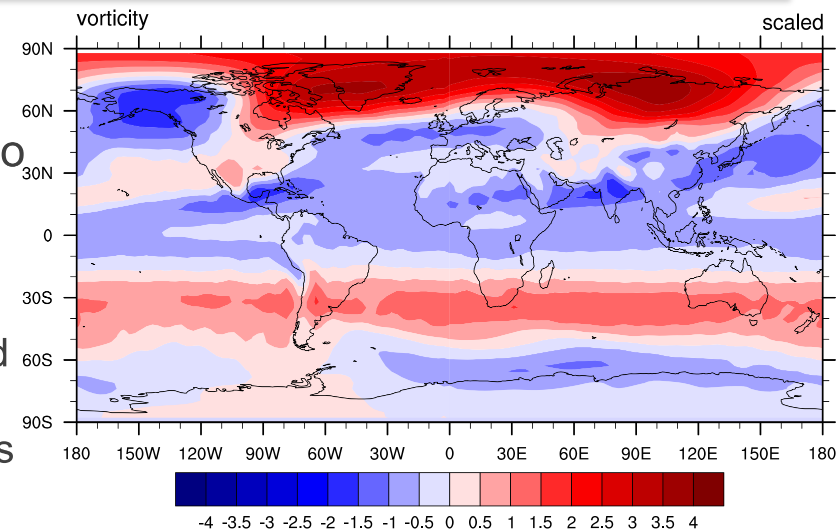
Karen Schuchardt, PNNL

parVis

Parallel Analysis Tools and New Visualization Techniques for Ultra-Large Climate Data Sets

Objective: *Speed up* the production of standard 2D plots of climate model output and allow application to *ultra-large data sets* and *native grids*

- Speed up current diagnostics (e.g the CESM-CAM atmospheric model diagnostics) with **task parallelism**
- Create a **data-parallel version** of the NCAR Command Language (NCL) analysis and visualization package.
- Build a new library: ParGAL – Parallel Gridded Analysis Library.
- Use existing software technology (MOAB, PnetCDF, Intrepid).
- ParNCL (built with ParGAL) will allow users to run their NCL scripts unaltered.
- Explore news ways of doing 3D visualization of climate data



Also anticipate **future hardware landscape** for climate analysis

- Introduce compression within NetCDF to cope with relatively small disk sizes.
- Building MapReduce-based climate analysis tools for cloud-based platforms.

Recent Accomplishments:

- 3x to 4x speedup of CAM and POP diagnostics. Released to community.
- Data parallel time averaging, vorticity (see above) and divergence calculations implemented.
- ParNCL interface to ParGAL working with simple scripts.
- Developed 2x – 3x lossless compression for smooth climate data

See http://trac.mcs.anl.gov/projects/parvis

U.S. DEPARTMENT OF ENERGY | Office of Science | Argonne NATIONAL LABORATORY | Computation Institute | NCAR | Pacific Northwest NATIONAL LABORATORY | Sandia National Laboratories | UCDAVIS UNIVERSITY OF CALIFORNIA

Possible strategies for output at Exascale.

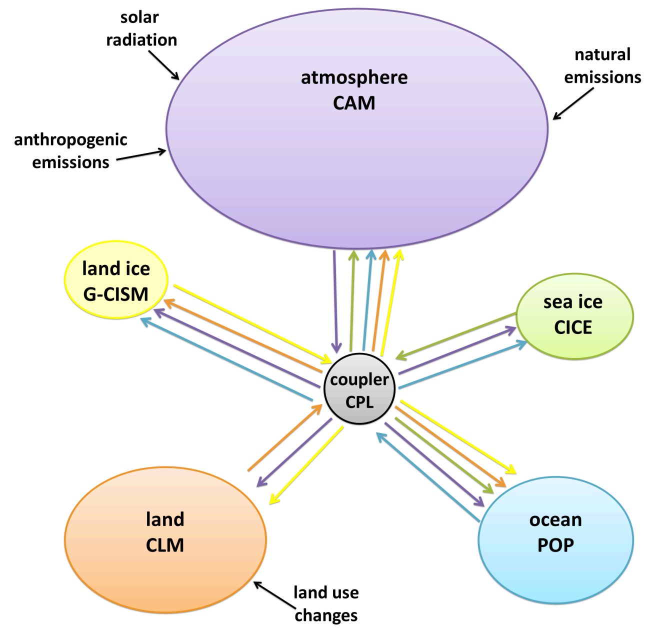
- Still save everything but:
  - Save it compressed
  - One file per variable, append in time up to X years. Better for storage hierarchy.
- Vastly reduce the number of variables saved.
  - Everyone has a favorite.
- In-situ analysis is not necessarily a solution:
  - Calculating averages costs more memory
  - Always need to compare with other climate simulations/observations.

Coupler at the Exascale: Problems

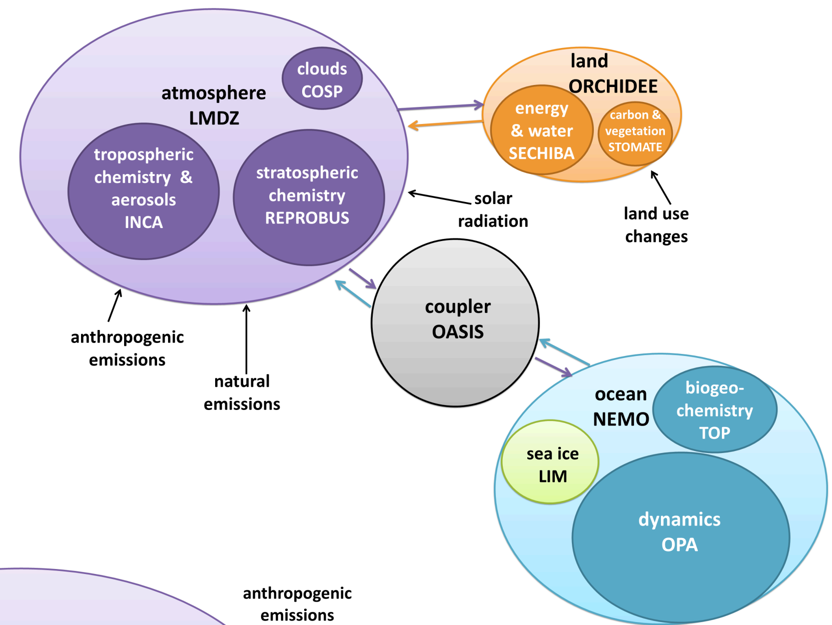
- Coupler is almost entirely 2D
  - Limited amount of parallelism
  - But not a huge number of flops compared to full model
- Coupler does lots of memory movement (which is expensive)
  - Moving data between model's native data type and coupler data type.
  - Moving data from one model's processors to another's.

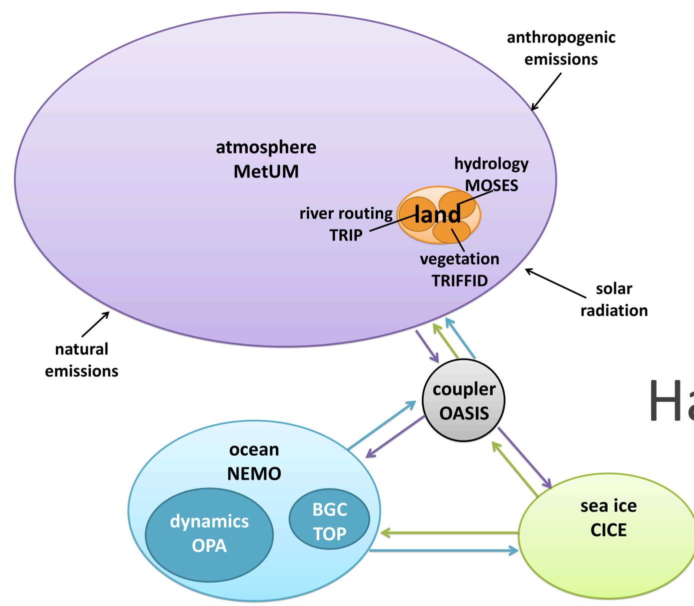
Climate model construction

CESM



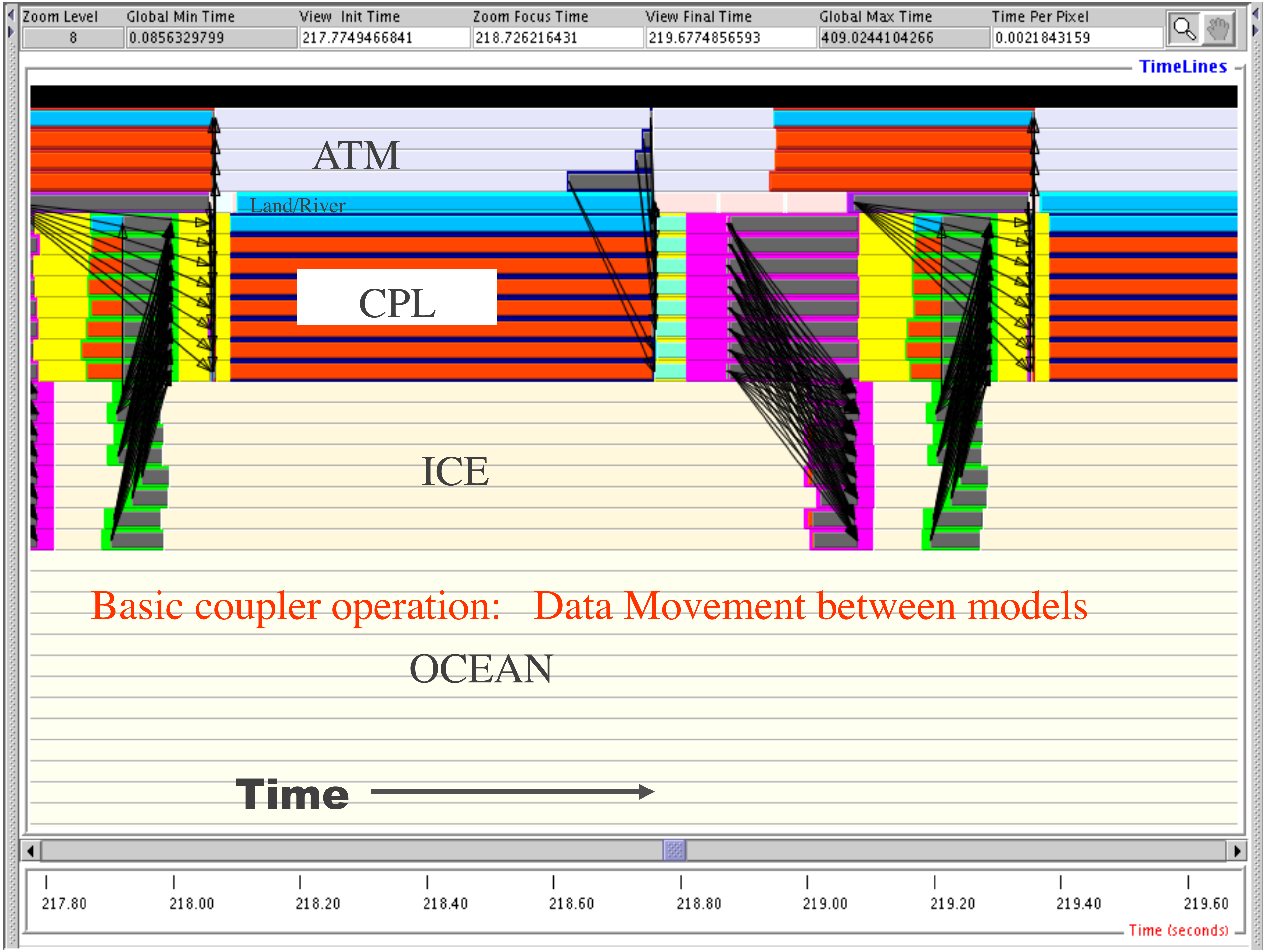
IPSL





HadGEM3

Figures from Kaitlin Alexander and Steve Easterbrook. Ovals proportional to code size.



Zoom Level
8
Global Min Time
0.0856329799
View Init Time
217.7749466841
Zoom Focus Time
218.726216431
View Final Time
219.6774856593
Global Max Time
409.0244104266
Time Per Pixel
0.0021843159
TimeLines
ATM
Land/River
CPL
ICE
Basic coupler operation: Data Movement between models
OCEAN
Time
217.80
218.00
218.20
218.40
218.60
218.80
219.00
219.20
219.40
219.60
Time (seconds)

Coupling at the Exascale: solutions

- More parallelism through more components executing concurrently
  - Ensembles
  - Different models

- Reduce memory movement
  - One data type across all model components?
  - Co-located decompositions.