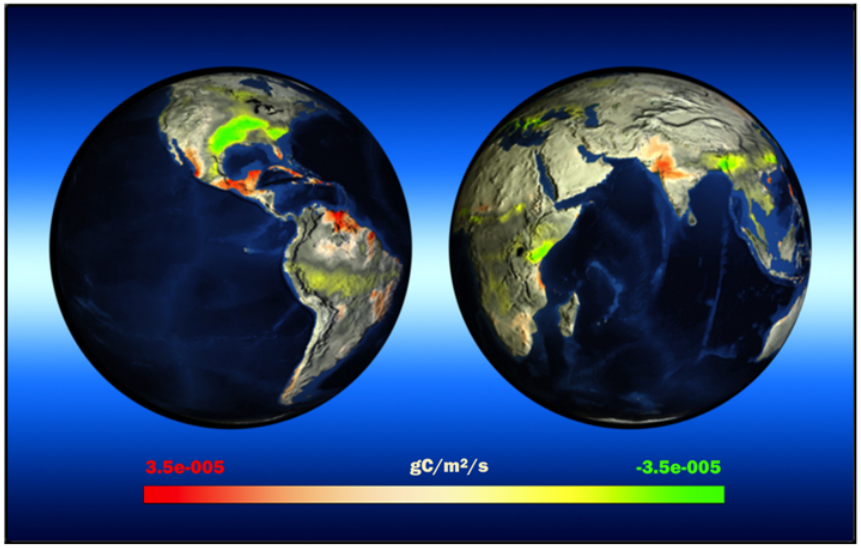
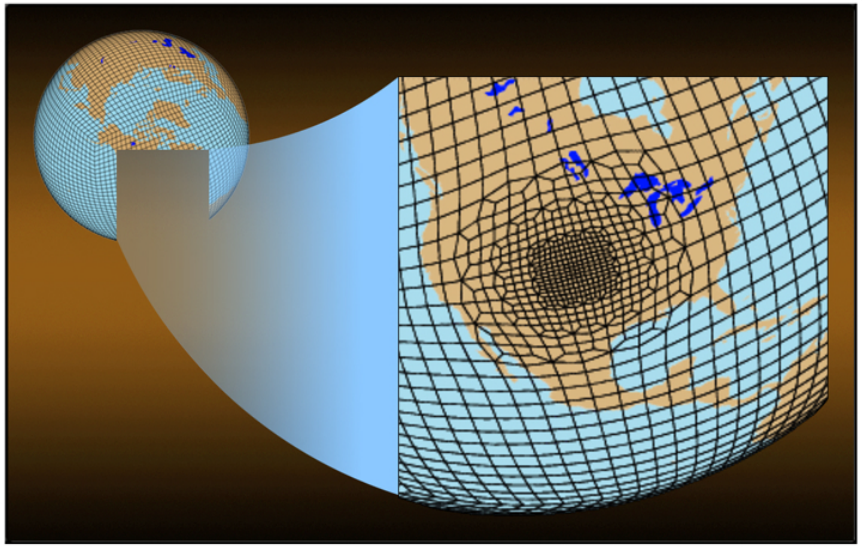
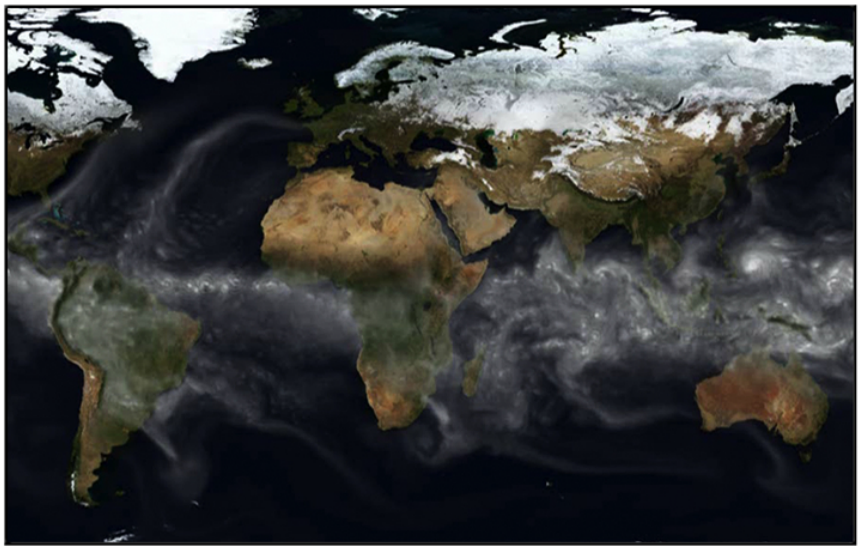
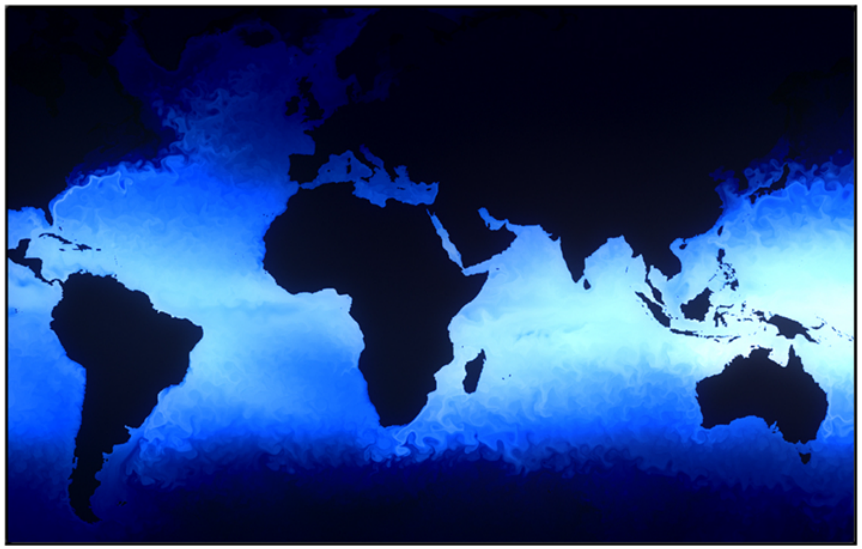








Large Scale Computing and Storage Requirements for Biological and Environmental Research Workshop, September 11-12, 2012

D. Bader, D. Koch, S. Klein, M. Taylor, D. Williams, W. Collins, J. Hack and R. Neale

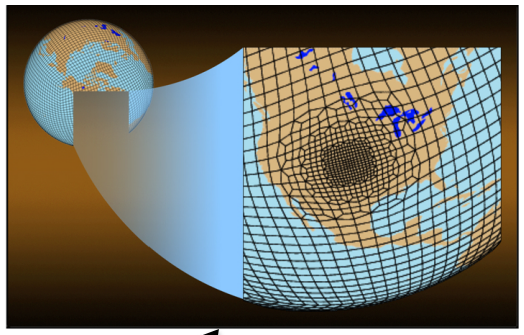
Cross-cutting Themes and Labs to advance CESM to address high priority DOE climate research

3 Science Themes:
- Numerics
- Testbeds
- Uncertainty Quantification

3 Components:
- Atmosphere
- Land
- Ocean and Sea-Ice

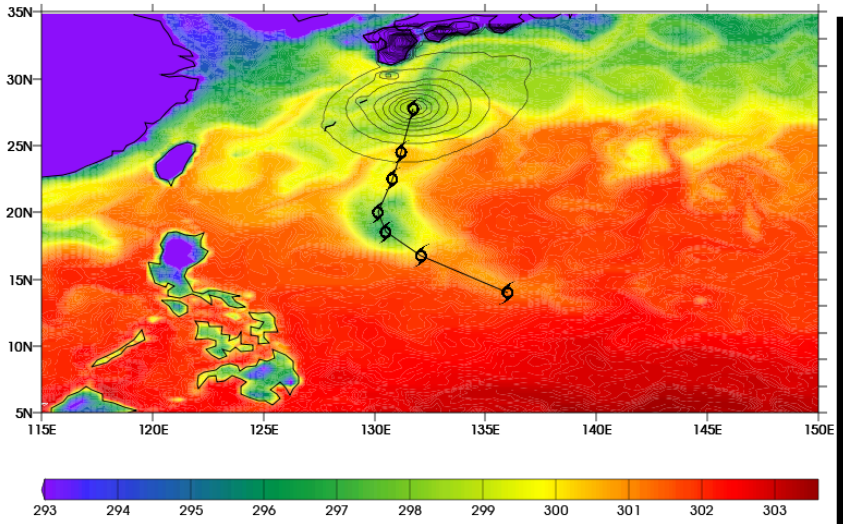
3 Research Directions:
- Hydrologic simulation improvement
- Variable-resolution numerical methods
- Carbon cycle uncertainty reduction

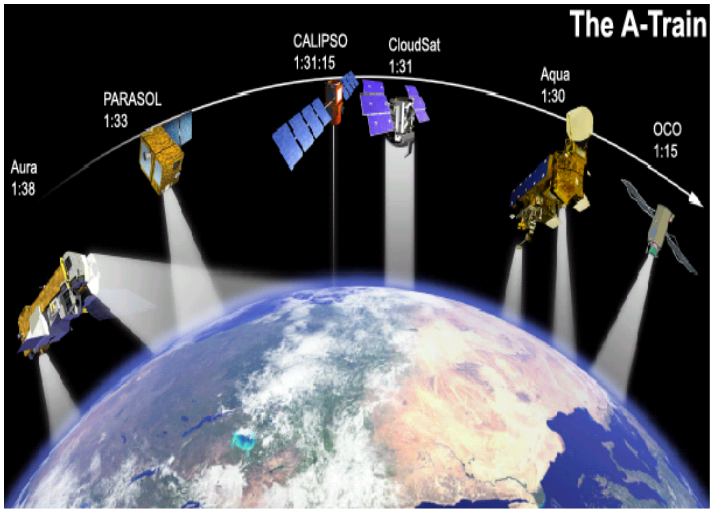


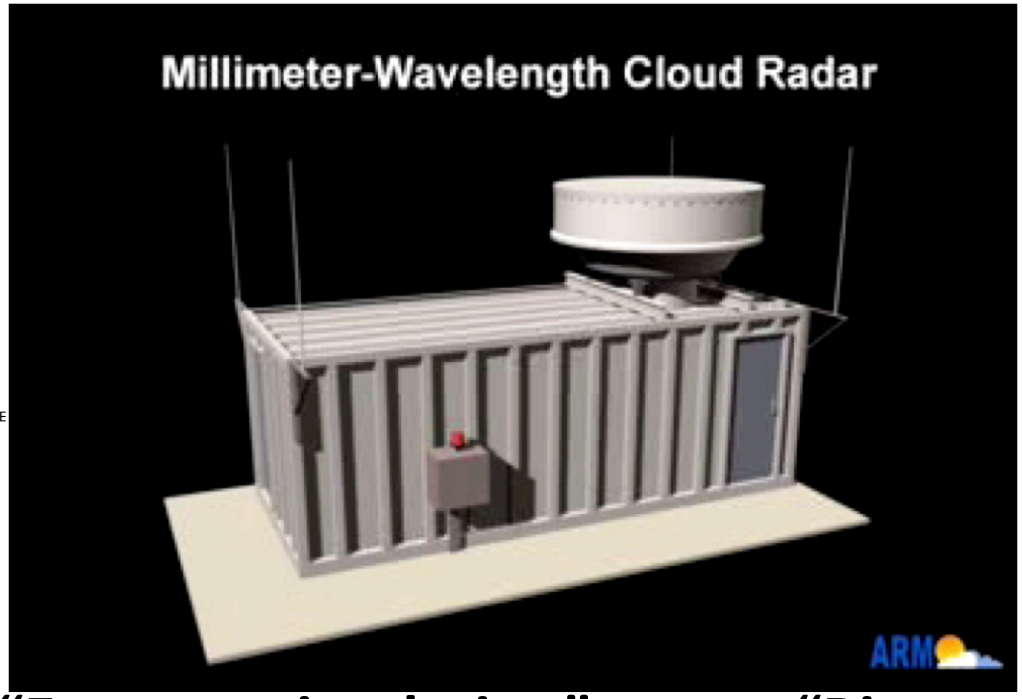




8 Labs:
ANL
BNL
LANL
LBNL
LLNL
ORNL
PNNL
SNL

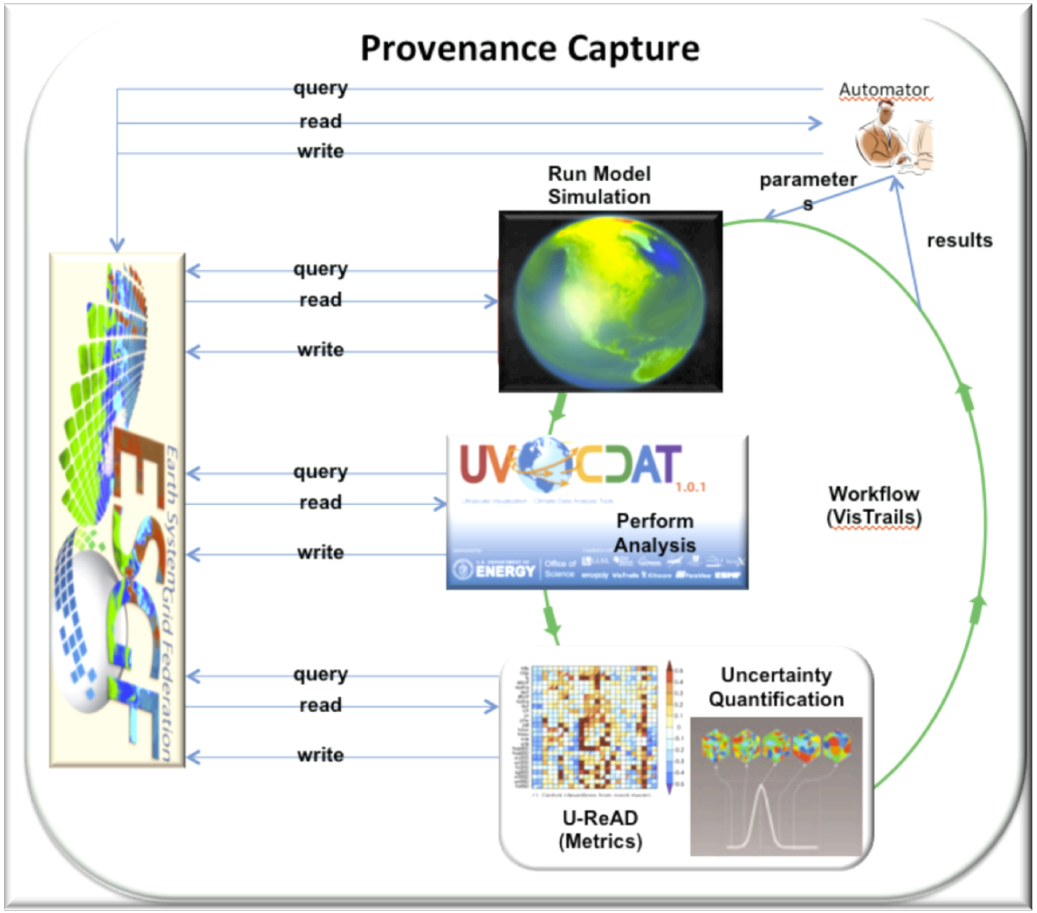






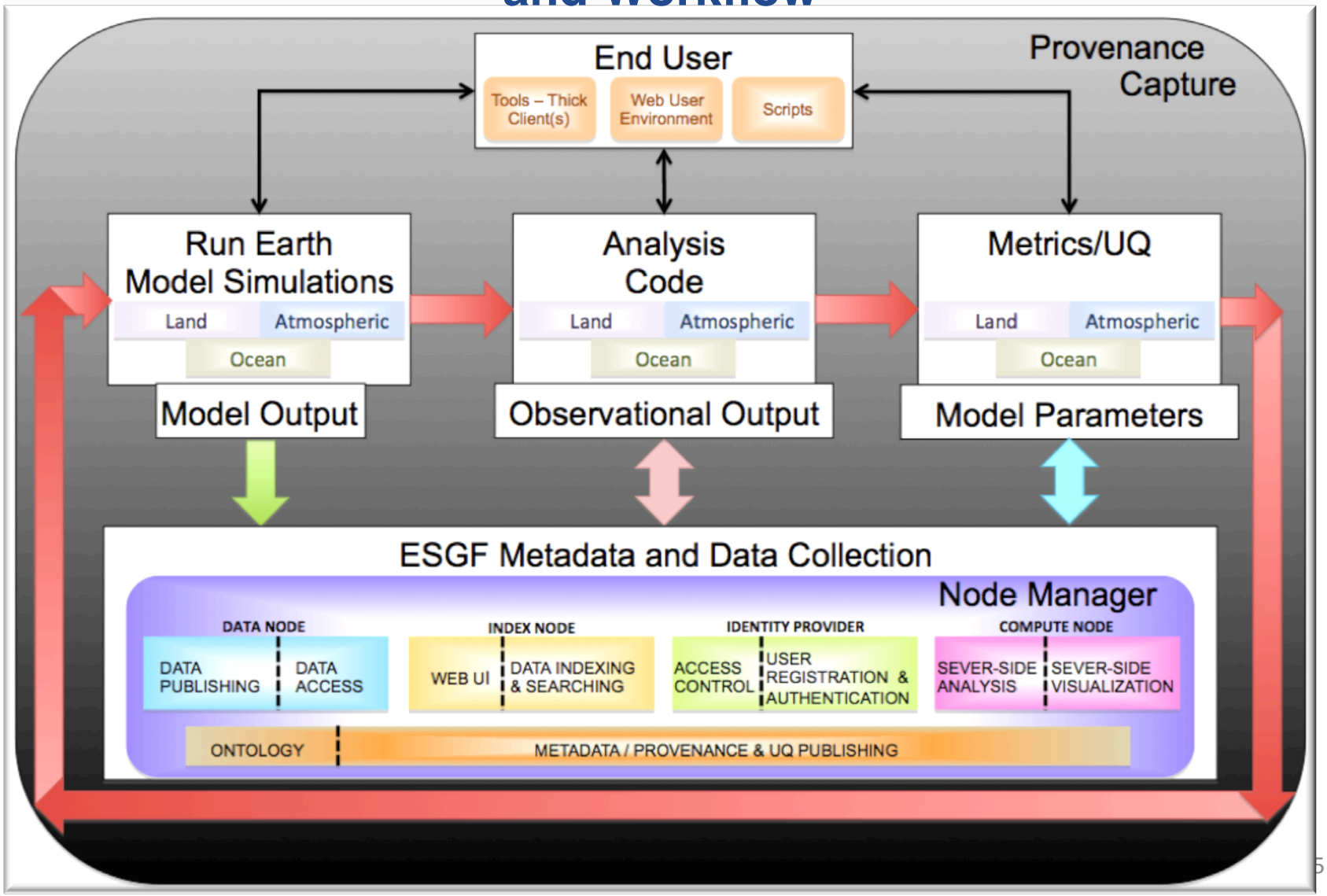
"Extreme simulation" meets "Big Data"!!!

Workflows and Provenance

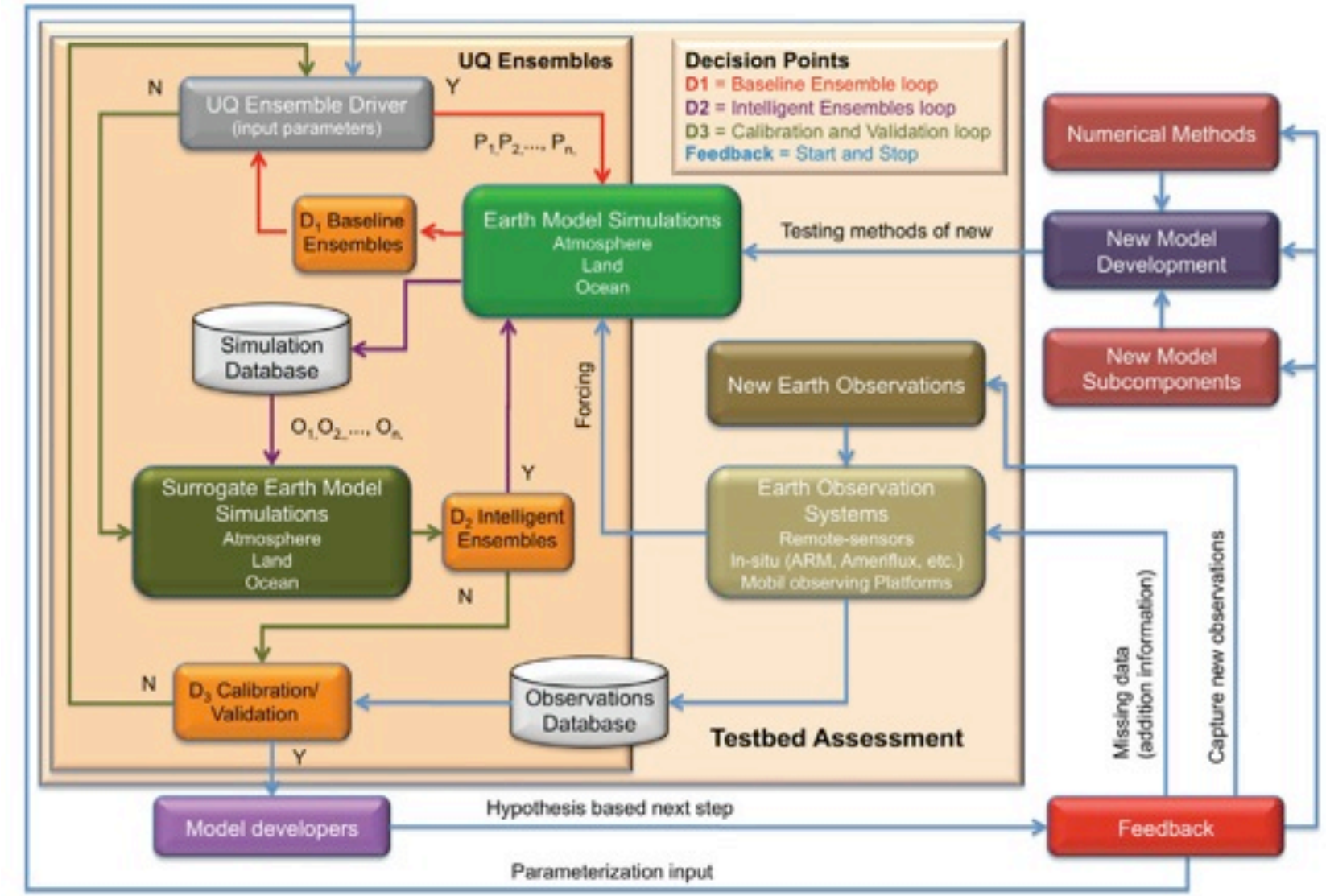


- **Why Use Workflows in CSSEF?**
  - Standardize scientific experiments, data processing, and analysis under one unified technical approach
  - Workflows will work as building blocks to support more complex problems.
  - Can automatically record what processes executed, what data was used, and what data resulted (Provenance)
- **Why is provenance important?**
  - Without knowing the origin of published data, scientists would have no way to trust the results:
    - What version of the model was used to generate it?
    - What model forcings and parameters were used to produce this result?
    - Real world atmospheric example: LLNL and PNNL both published ARM best estimate data sets. Are they redundant, interdependent or completely independent? The historical record of how they were produced would be the only way to truly differentiate the two efforts.

High-level Conceptual View of CSSEF Test Bed Architecture and Workflow

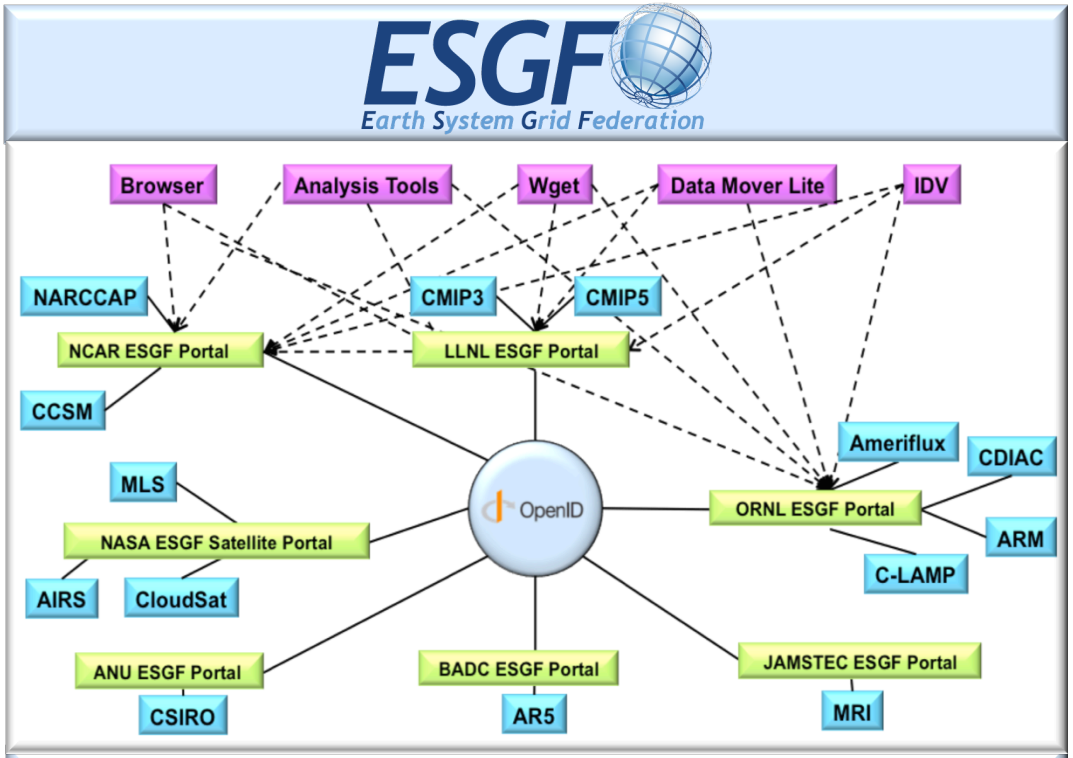


Test-bed Workflow with UQ



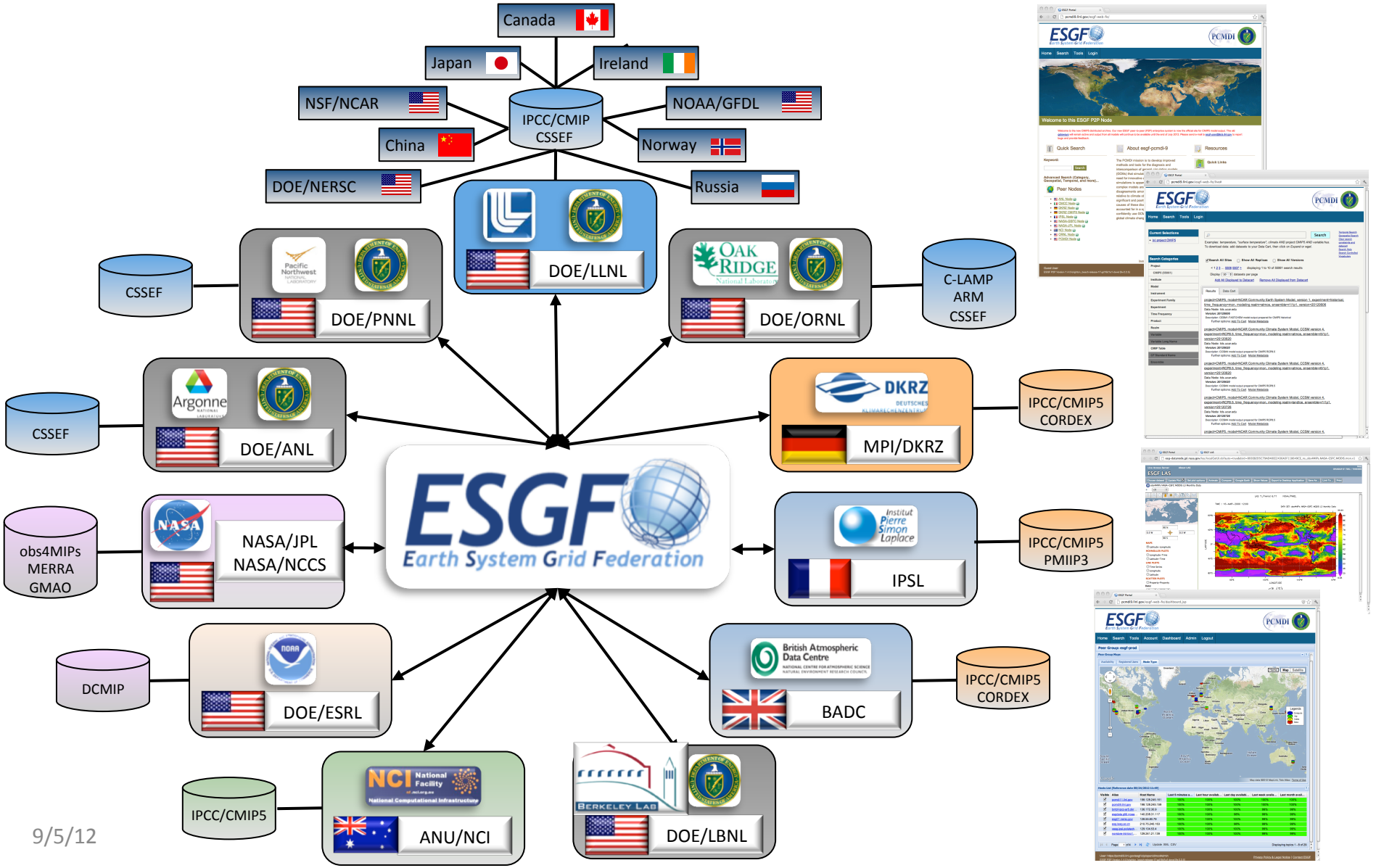
The ESGF distributed data archival and retrieval system

- Distributed and federated architecture
- Support discipline specific portals
- Support browser-based and direct client access
- Single Sign-on
- Automated script and GUI-based publication tools
- Full support for data aggregations
  - A collection of files, usually ordered by simulation time, that can be treated as a single file for purposes of data access, computation, and visualization
- User notification service
  - Users can choose to be notified when a data set has been modified



Large amounts of data movement and management are needed for the complex and diverse climate data community.

ESGF : Federated and integrated data from multiple sources



The UQ Process in CSSEF

1. Sensitivity analysis

   Response of climate to uncertain input parameters

2. Surrogate models

   Climate model emulators for UQ and optimization

3. Model calibration

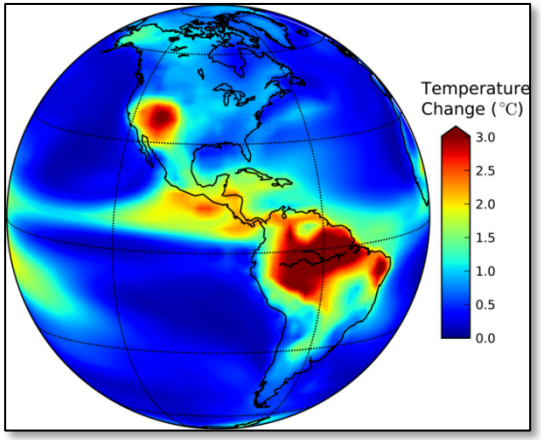
   Tuning of uncertain parameters using observations

4. Forward UQ

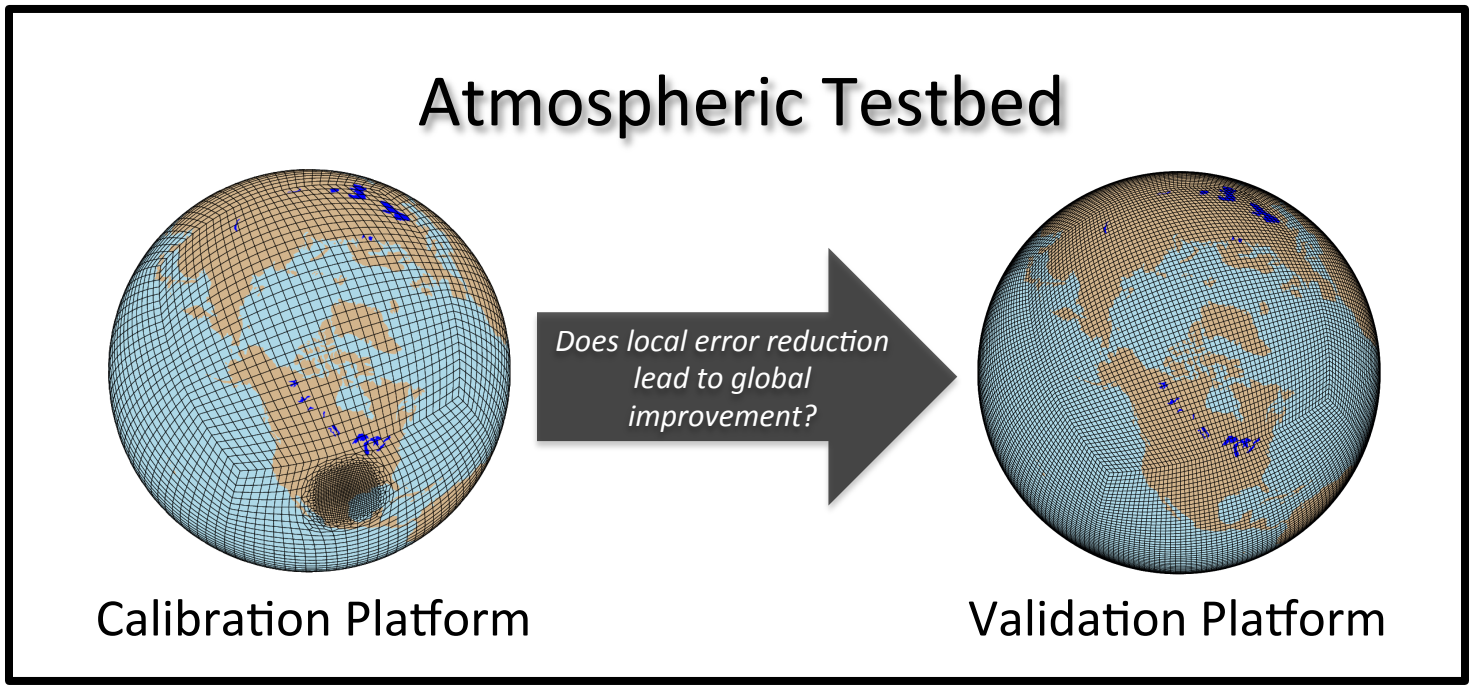
   Characterize predictive accuracy from UQ on CESM

	Sensitivities & Uncertainties	Observations & Calibration	Model Development
Land	• About 40 parameters (6-12 main drivers) • Forcing from CAM ensembles	• FluxNet data at sites • Global gridded data • Multi-site calibration (real and 'proxy' sites) • Global calibration	• Ecosystem demography • Soil carbon dynamics • Subgrid scale hydrology
Ocean	• Eddy mixing in Gent-McWilliams scheme • Resolution of static refined meshes	• T, S, ACC transport, CFC ventilation • Calibrate kappa-i as a function of (x,y,z)	Optimal placement of high resolution meshes using calibrated eddy mixing
Atmosphere	• About 20-30 parameters (cloud formation, convection, etc)	• ARM data at individual sites • Calibration at sites using statically refined mesh	New parameterizations of convection and cloud processes

Baseline assessment → Optimized models → Model improvements



Testbed to Better Model?



All atmospheric models need to be calibrated with observations because they contain parameterizations

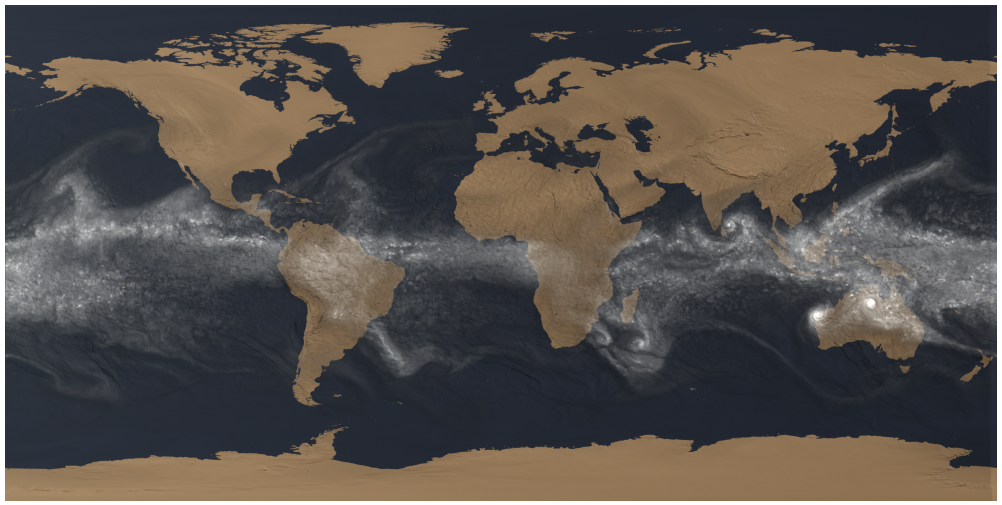
Preparing CESM for DOE leadership class computing facilities: Global high-resolution and variable resolution capabilities

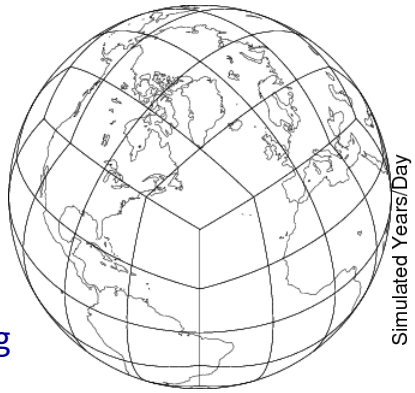
CESM next generation spectral element based atmospheric dynamical core for improved parallel scalability and variable resolution capability.

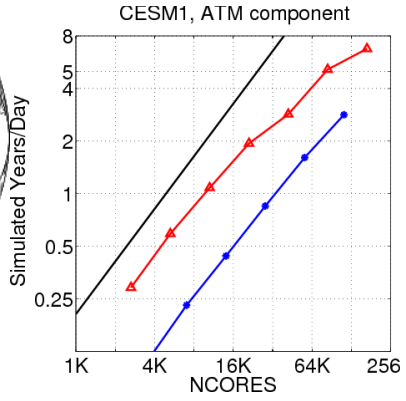
Enables ultra high resolution simulations on petascale platforms: 1/8° CESM time-slice simulations running well on 170K cores (ORNL Jaguar), achieving 4.6 SYPD. Scaling to O(500K) cores expected.

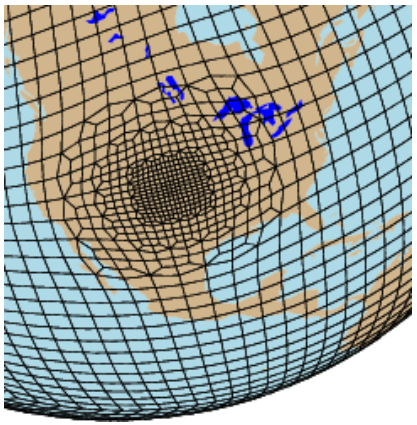
CESM's spectral element method is 4'th order accurate with local conservation of mass and energy and quasi-monotone transport.

Examples: cubed-sphere grid for uniform high resolution (left), scaling to 170K cores at global 1/8° (middle), a global 1° grid smoothly transitioning to 1/8° over the SGP ARM site (right).









Initial application of production UQ tools for atmospheric test-bed at LLNL

Status of Global Sensitivity Analysis of CAM5

Goal: Sensitivity analysis of low-resolution CAM5 with respect to important and uncertain hydrological parameters.

Approach:

- **Adopt CAM5 from CESM at 1.9°×2.5° resolution using 5 years simulations.**
- **Select a targeted set of uncertain parameters from a larger superset elicited by the atmospheric testbed group.**
- **Sample uncertain parameters using various sampling schemes.**
- **Carry out sensitivity analysis of hydrological metrics and statistics using multiple sensitivity measures**

Analysis:

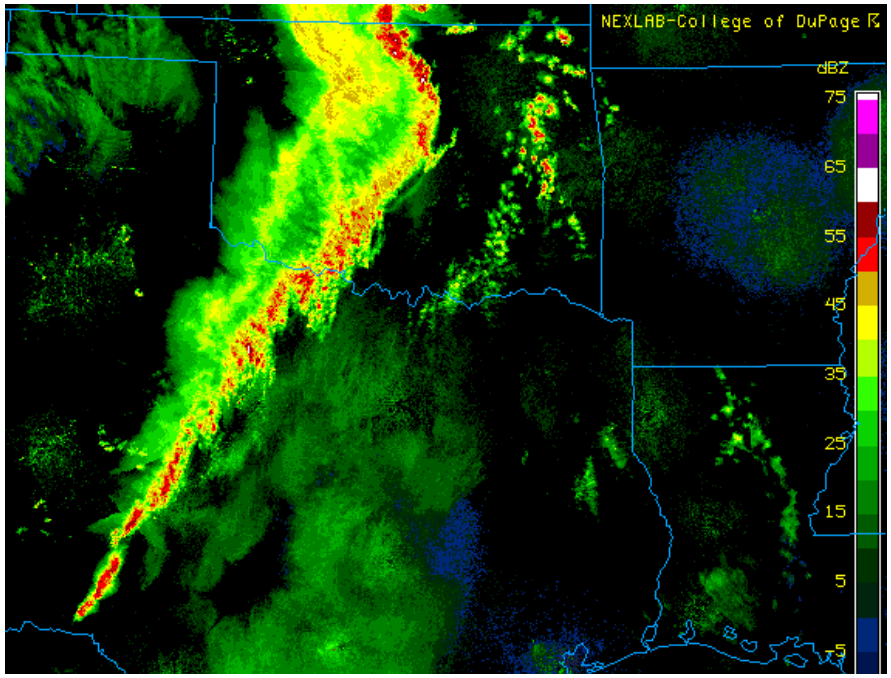
- Focus on 22 uncertain parameters mostly related to the macro physics of CAM5.
- First batch of runs consisted of three Latin hypercube (LH) ensembles, each of size 220.
- Follow up ensembles with revisited uncertainty ranges consisted of two LH ensembles of size 220 each, plus a one-at-a-time ensemble of size 45. (Finished the first week of Nov).
- ***Total of 1,145 CAM5 simulations = 5,725 simulated years.***
- **Two main analyses:**
  - *Comparison of CAM4 and CAM5 sensitivity analysis*
  - *Comparison of multiple response-models and sensitivity analysis methods using the CAM5 ensembles.*

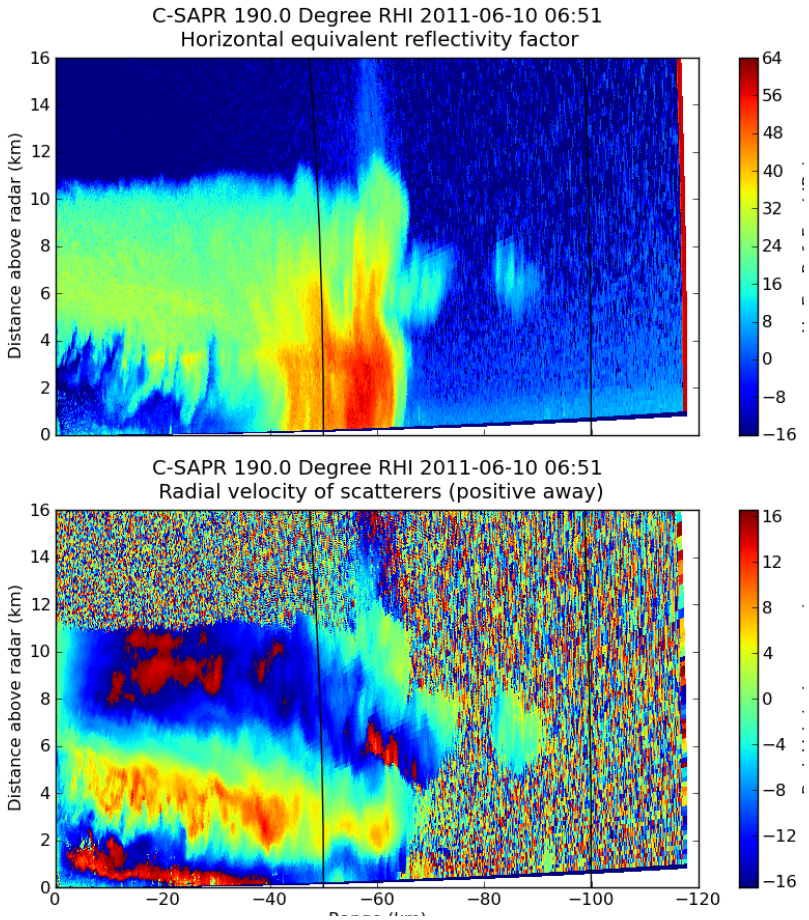
Table IV.1. Characteristics of Calibration and Validation Platforms

	Calibration Platform	Validation Platform
Model Construction	Global model with static mesh refinement above sites of interest	Global model with uniform grid
Forcing	Weather-forecast mode with nudging to analysis data on coarse outer grid	Free-running climate mode with initial condition from analysis data
Initial Resolution	1/8° fine mesh transitioning to 1°	1/8° everywhere
Uncertainty Quantification Methods	Parameter tuning: calibration of uncertain parameters with local data sets	Ensemble of simulations with parameter sets generated by parameter tuning
Data sets	Local water cycle data sets such as ACRF site data	Global data sets such as satellite or re-analysis data

Utilizing ARM Instruments

- With ARRA funds, ARM installed over 60 new advanced radars (many of which see precipitation)
- We'll use these observations in our model calibration
  - Mid-latitude Continental Convection
  - http://campaign.arm.gov/mc3e/
  - Organized Tropical Convection and MJO
  - (http://campaign.arm.gov/amie/)





UV-CDAT - complementary BER funded project

http://uvcdat.llnl.gov

Ultra-scale Visualization Climate Data Analysis Tools (UV-CDAT):

- **Integrate DOE's climate modeling and measurements archives**
- **Develop infrastructure for national and international model/data comparisons**
- **Deploy a wide-range of climate data visualization, diagnostic, and analysis tools with familiar interfaces for very large, high resolution climate data sets (CDAT, VTK, R, VisIt, ParaView, DV3D, …)**
- **Workflow – data flows are directed graphs describing computational tasks**
- **Takes advantage of ESGF data management**

