
# Software Engineering for Climate Modeling


Spencer Rugaber

College of Computing

Georgia Institute of Technology

The Earth System Curator Project
Collaborators:
National Center for Atmospheric Research
(Cecelia DeLuca and Don Middleton PIs)
Geophysical Fluid Dynamics Laboratory – Princeton University
(V. Balaji – PI)
Georgia Tech: Leo Mark (Co-PI)





# Context


- **Problem:** predict climate change
- **Approach:** simulation
- **Validation:** comparison with climate record, model intercomparison, comparison vs. theory, peer review
- **Impact:** Intergovernmental Panel on Climate Change (IPCC) Assessment Reports

GFDL Geophysical Fluid Dynamics Laboratory

The National Center for Atmospheric Research
NCAR
managed by UCAR to serve the community
sponsored by NSF


# Background Concepts


- *Climate* vs. *weather*
- *Model*: simulation software
- *Grid*: numerical mesh
- *Coupling*: multiple models combined for purpose of increased accuracy; different update frequencies; different grids
- *Framework*: API and support library for coupling models
- *Run*: one execution of a model, combined into *realizations*
- *Ensemble* – multiple realizations with varied initial conditions or physics assumptions for the purposes of sensitivity analysis
- *Experiment* – (scientific hypothesis) Climate of the 20th Century




# Example Model – CCSM3


- Community Climate System Model
- Source: NCAR
- Models: atmosphere, ocean, sea ice, land
- Source code
  - 4 models; 1475 files; seven languages (plus variants)
  - 285,962 lines of code
- Output data (one run; 100 year period)
  - ~685 gigabytes
  - NetCDF format




# CCSM3 Details


- Fully-coupled global climate model
- Atmosphere: “T85” resolution (~155 km), 26 levels
- Ocean: 320x384, 40 levels, curvilinear grid
- Sea ice: 320x384, curvilinear grid
- 70+ experiments submitted to the IPCC

GFDL Geophysical Fluid Dynamics Laboratory

The National Center for Atmospheric Research
NCAR
managed by UCAR to serve the community
sponsored by NSF


# IPCC AR5 Timetable


| Date | Event |
|---|---|
| 9/07 | Benchmark experiments defined |
| 3/08 | Standard output finalized (CMIP panel) |
| 9/08 | All required experiments defined (WGCM meeting) |
| Late 08 | Modeling groups begin running benchmark experiments |
| 2009 | Modeling groups run models and produce output |
| 1/10 | Model output starts to be made available to community |
| 9/10 | First meeting of WG1 AR5 lead authors |
| 3/11 | Research paper drafts completed and available to WG1 AR5 authors |
| 9/11 | First order draft of WG1 AR5 completed |
| 8/12 | Deadline for acceptance of papers cited by IPCC |
| 6/13 | Release of WG1 AR5 |




# Example Infrastructure Project: ESMF


- Earth System Modeling Framework
- Source: NCAR
- Purpose
  - Software framework for building and coupling models
- Classical software engineering tradeoff
  - Severe performance requirements
  - Perceived value related to abstractions provided




# ESMF Details


- Source code
  - 1064 files
  - 4 languages (mostly Fortran)
  - 356K lines of executable code (+100K tests)
  - 340 subroutine (including overloads)
- Approach: infrastructure/superstructure
- Normative: clients need to subscribe to usage protocol




# ESMF Process Element


- Software bundled with exhaustive regression test suite, run nightly on many platforms
- Use of an external change review board for prioritization of development tasks
- Close coordination of a distributed development team
- User training
- Community meetings


# ESMF Challenges


- Fortran
- Super-computers as target platforms
- Standardization of behavior, interfaces, and documentation across a large code base
- Balancing generality with ease of use
  - General subroutines with lots of arguments vs. lots of subroutines





# FMS


- Flexible Modeling System
- Source: GFDL
- Purpose: software framework for development, construction, execution, and interpretation of climate models
- Source: 100KLOC (infrastructure); 900KLOC (total)
- FRE: FMS Runtime Environment; custom workflow engine; 50KLOC PERL generating CShell scripts to control supercomputer




# FMS Process Elements


- "Unprincipled use of software engineering principles"
- Data modeling (XML; SQL)
- Oversight committee generates requirements
  - No formal issue tracking
- Serious (black box) regression testing


# FMS Challenges


- Super computer; site specific
  - Means general solutions like Kepler can't be applied
- Massive parallelism
- Precision
- Quicker feedback on tests
  - One-hundred year test run may take two weeks to complete




# Earth System Curator


- NSF funded
  - *Information Integration and Informatics Program*
- Collaborative: NCAR, GFDL, Georgia Tech
- Vision
  - Persistence transparency with respect to models and data sets

GFDL Geophysical Fluid Dynamics Laboratory

The National Center for Atmospheric Research
NCAR
managed by UCAR to serve the community

sponsored by NSF


# Curator Goals


- Community model metadata
- Gateway implementation
  - Navigation; search; retrieval; comparison
  - Connect model components with datasets
- Automatic generation of couplers and drivers




# Non-Challenges


- Component-based design
- Data services
- Scientific validation
- White box testing; reverse engineering
- Semantics (use of constraints)
- Turf




# Software Engineering Challenges


- Performance, performance, performance
- Scalability
- Community building, "standards", and agile governance
- Technology transition (unknown impacts researchers)
- Reproducibility (in the face of multiprocessing)
- Configuration and deployment
- Search and visualization of large datasets
- Representation
- Workflows: provenance and explanation




# More


- http://www.earthsystemcurator.org/
- Dunlap, R., et. al, (2008), Earth system curator: metadata infrastructure for climate modeling, *Earth Science Informatics*, 1, 131-149
- Acknowledgements: NSF, Rocky Dunlap, Cecelia DeLuca, V. Balaji