The Earth System Modeling Framework

V. Balaji
GFDL Princeton University

APAN eScience Workshop
Honolulu HI
28 January 2004

Talk overview

- Why frameworks have become hot: science and technology trends

- The ESMF project

- Beyond ESMF: convergence of models and data

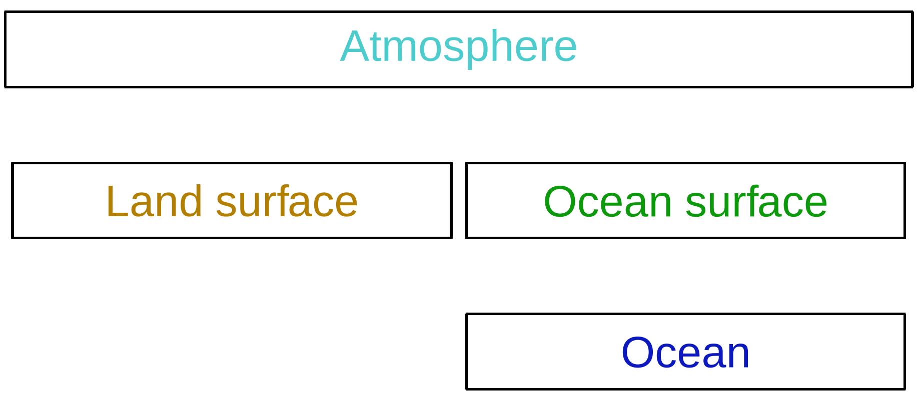
Components of the Earth system

Atmosphere atmospheric fluid dynamics and thermodynamics, moist processes, radiative transfer, transport and chemistry of trace constituents.

Ocean World ocean circulation, ocean biogeochemistry.

Land surface Surface processes, ecosystems, hydrology.

Ocean surface Sea ice, wave processes.



Complexity of climate simulations

Models have grown increasingly complex with time.

70s	80s	early 90s	late 90s	today	early 00s	late 00s
Atm	Atm	Atm	Atm	Atm	Atm	Atm
	Land	Land	Land	Land	Land	Land
		Ocn, SeaIce	Ocn, SeaIce	Ocn, SeaIce	Ocn, SeaIce	Ocn, SeaIce
			Aerosols	Aerosols	Aerosols	Aerosols
					C Cycle	C Cycle
		Aerosols			Ecosystems	Ecosystems
		Land C			Chemistry	Chemistry
	Ocn, SeaIce	Ocn Carbon			Ocn Eddies	Ocn Eddies
	Clouds	Chemistry	C Cycle	Ocn Eddies		Clouds

Components are developed “offline” (bottom left) and then are integrated into comprehensive coupled models.

Component-based design

Each process has its own intrinsic time and space scales.

Older models did not allow subcomponents to be on independent grids and timesteps.

- Old way: sharing of data through arrays in common blocks.

- New way: independent model grids connected by a coupler.

Each physical *process component* becomes an independent *code component* that can be separately instantiated, initialized, stepped forward, and terminated.

The underlying architecture of high-end computing

Shared memory signal parallel and critical regions, private and shared variables. Canonical architecture: UMA, limited scalability.

Distributed memory domain decomposition, local caches of remote data ("halos"), copy data to/from remote memory ("message passing"). Canonical architecture: NUMA, scalable at cost of code complexity.

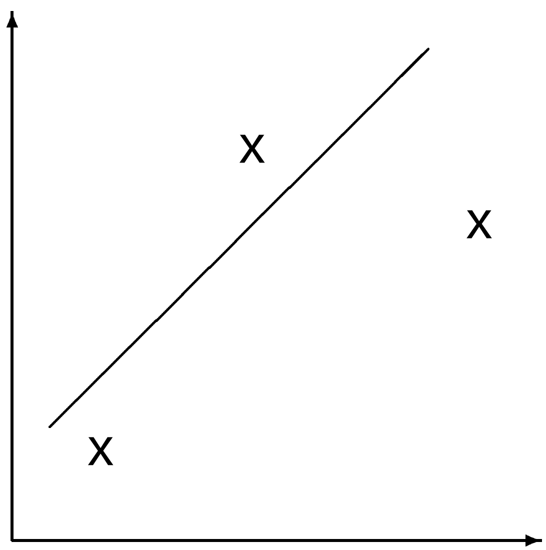
Distributed shared memory or ccNUMA message-passing, shared memory or remote memory access (RMA) semantics. Processor-to-memory distance varies across address space, must be taken into account in coding for performance. Canonical architecture: cluster of SMPs. Scalable at large cost in code complexity.

Data assimilation

Data can appear at unpredictable locations in time and space (radiosondes, buoys, satellites) and have no clear radius of influence ("location streams").

Models will "slosh" if incompatible with data.

Data assimilation involves bringing models and data into acquiescence. Assimilation algorithms can be treated as gridded components.



Technological trends

In climate research... increased emphasis on detailed representation of individual physical processes governing the climate; requires many teams of specialists to be able to contribute components to an overall coupled system;

In computing technology... increase in hardware and software complexity in high-performance computing, as we shift toward the use of scalable computing architectures.

Technological trends

In software design for broad communities... The open source community provided a viable approach to the construction of software to meet diverse requirements through "open standards". The standards evolve through consultation and prototyping across the user community. Software is designed as coupled independent components.

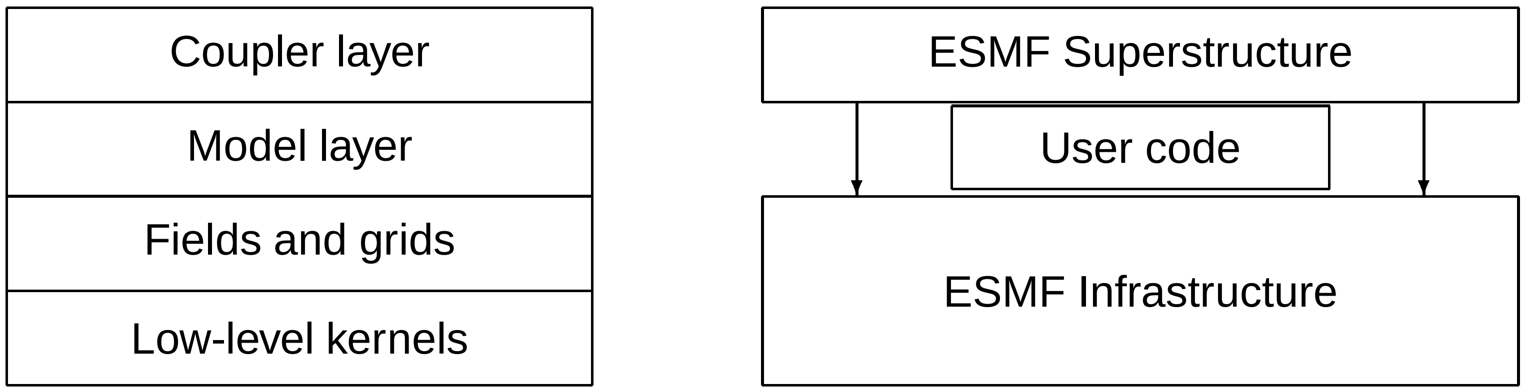
Prototype frameworks based on standards and component-based design began to appear in the climate modeling community starting in 1997 (FMS: GFDL Flexible Modeling System, ...)

ESMF

The Earth System Modeling Framework is an end-to-end solution for the problem outlined here: supporting distributed development of models with many interacting components, with independent space and time discretization, running on complex modern scalable architectures. A capsule history of ESMF:

- The need to unify and extend current frameworks achieves wide currency (c. 1998).
- NASA offers to underwrite the development of an open community framework (1999).
- A broad cross-section of the community meets and agrees to develop a concerted response to NASA uniting coupled models, weather, and data assimilation in a common framework (August 1999). Participants include NASA/GMAO, NOAA/GFDL, NOAA/NCEP, NCAR, DOE, and universities with major models.
- Funding began February 2002: $10 M over 3 years.
- First Community Meeting, Washington, May 2002: requirements review.
- Second Community Meeting, Princeton, May 2003: design review.
- Third Community Meeting, Boulder, July 2004: prototype release.

Architecture of an Earth System Modeling Framework: the sandwich



ESMF features

- ESMF is usable by models written in f90/C++.
- ESMF is usable by models requiring differentiability.
- ESMF is usable by models using shared or distributed memory parallelism semantics.
- ESMF supports serial and concurrent coupling.
- ESMF supports multiple I/O formats (including GRIB/BUFR, netCDF, HDF, native binary).
- ESMF has uniform syntax across platforms.
- ESMF runs on many platforms spanning desktops (laptops, even!) to supercomputers.

Summary

- The Earth System Modeling Framework supports distributed development of models with many interacting components, with independent space and time discretization, running on complex modern scalable architectures.
- Prototype release scheduled for July 2004.
- Future maintenance guarantee from NCAR: seeking funding for further development.
- Future directions: link to community data portals, runtime environment.

Selected web references

`http://www.esmf.ucar.edu` General website for ESMF: documentation, code, examples, contacts.

`http://prism.enes.org` General website for PRISM: PRogram for Integrated Earth System Modeling.

`http://www.gfdl.gov/~fms` The GFDL Flexible Modeling System. Also links production models and climate model simulation data.

`http://nsipp.gsfc.nasa.gov` Focus on short-term climate variability.

`http://mitgcm.org` The MIT GCM. Considerable emphasis on data assimilation.

`http://www.wrf-model.org` The NCAR/NOAA Weather Research and Forecasting model.

`http://www.ccsm.ucar.edu` The NCAR Community Climate System Model.

`http://www.ipcc.ch` Intergovernmental Panel on Climate Change. Comprehensive scientific synthesis of current thinking on climate.

Beyond ESMF and ESG: linking model and data frameworks

Community data frameworks are under development. For model output data to be scientifically useful, the researcher must have some knowledge of how the data was produced. Model data requires a *model's eye view* description of the data, another layer of metadata, which includes:

- Description of model components: e.g FMS BGRID atmosphere, land and sea ice coupled to MITgcm ocean.
- Description of grid configurations and resolutions.
- Choice of physics packages and input parameters.
- Model state and its fields.

ESMF and PRISM are emerging standards that allow the development of the model metadata layer, based on the state data structures and its base classes. Modeling framework data structures map directly on to community hierarchical metadata.

Convergence of models and datasets

Given the existence of a model metadata layer, *the same descriptor can be used as model input and model output.* This means:

- the files that are used to configure, build and launch a model (written in, say, XML) contain the same physical information that must be written to the output dataset for a comprehensive description of how the data was generated.

- This information can also be stored in a relational database of model configurations and datasets: the Earth System Model Curator. Such a DB would allow experiment comparisons, high-level queries, experiment redesign, next-generation publication of scientific results.

Potential use scenarios

Climate scientists setup (assemble components, configure input parameters); comparisons (run configurations, results, with data); branch runs, ...

Impacts studies query models by pattern, couple biogeochemistry model either offline with dataset or online with model.

IPCC, MIPs descriptions of intercomparisonns, setup new MIPs, archive MIP results.

Policymakers, industry and educators High-level access to swathes of model data.

Publication link datasets to publications; introduce interactive aspect to publication; annotation of data, certification and quality control.

Portability automatic best-practice configuration appropriate for platform.

Operations higher rate of technology transfer from research to operations.

Proposal for an Earth System Modeling Environment (ESME)

We seek to unite the data (ESG) and model (ESMF) communities with climate scientists (IPCC, CMIP) to develop the model metadata layer, and the relational database of models and data that would be based on it.

This effort would be closely allied with the PRISM/CAPRI efforts in the same domain.

Elements of the ESME

Physical interfaces development of comprehensive physical interfaces for model components.

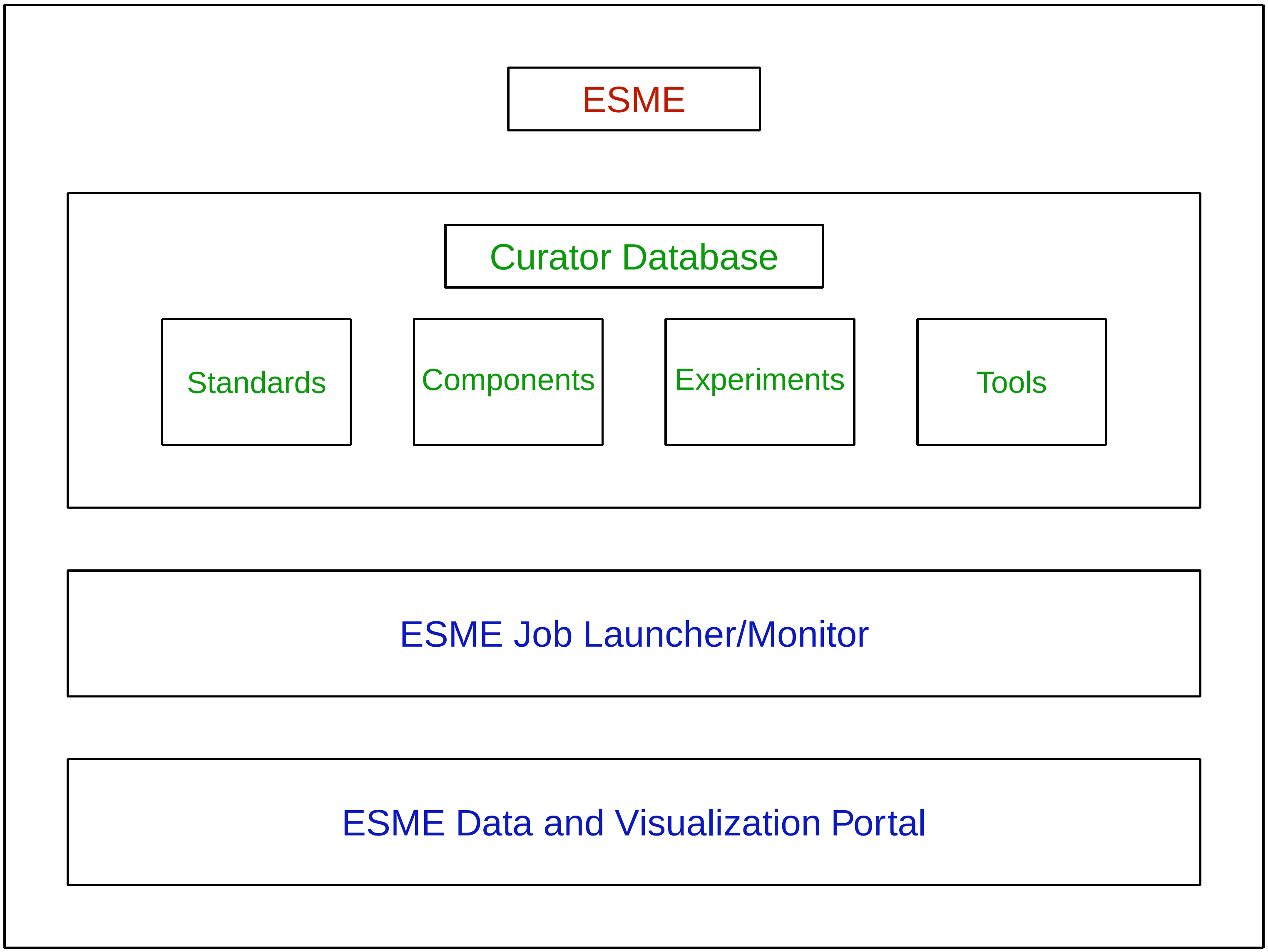
Hierarchical metadata development of a semantic web of model and data descriptors.

Relational database of model experiments and observational and model datasets.

Data annotation certification by assigned authority, or *à la* Google. Links with scientific results and peer-reviewed literature.

Web portal interfaces to query operations, comparisons, client- and server-side data analysis.

Structure of the ESME



ESME workflow