Building Earth System models

V. Balaji
SGI/GFDL Princeton University

APAN Natural Resources Workshop
Honolulu HI
29 January 2004

Weather and climate

“Climate is what you expect, weather is what you get.” – Ed Lorenz.

Climate is the study of long-term time averages.

The climate, however, has variations on all time scales: monsoons, ice ages.

More recently, the chemical and radiative of the atmosphere has been abruptly altered, with consequences that need to be explored.

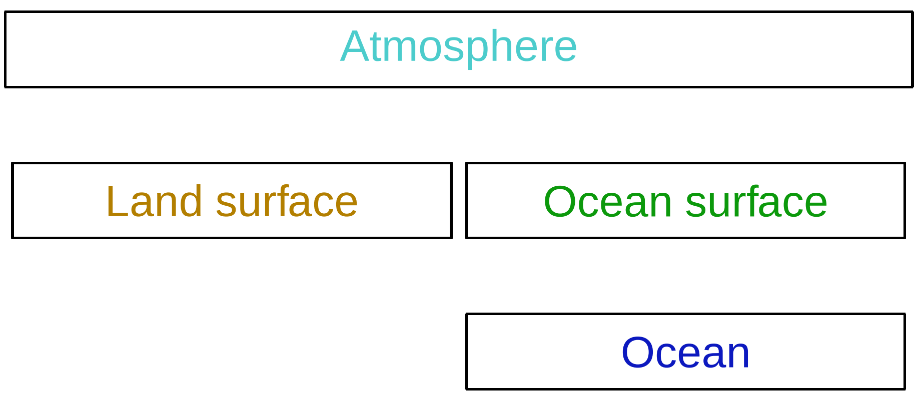
Components of the Earth system

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.

Timescales and space scales

Clouds minutes to hours; 10 m – 100 km.

Ocean eddies hours to days; 10 km.

Weather patterns days; continent scales.

Ocean currents days; ocean basin scales.

Short-term climate seasonal-interannual; El Niño; monsoons.

Climate change Decadal-centennial; ocean overturning and deep water formation.

Paleoclimate Millennia to deep time; ice ages; orbital changes.

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.

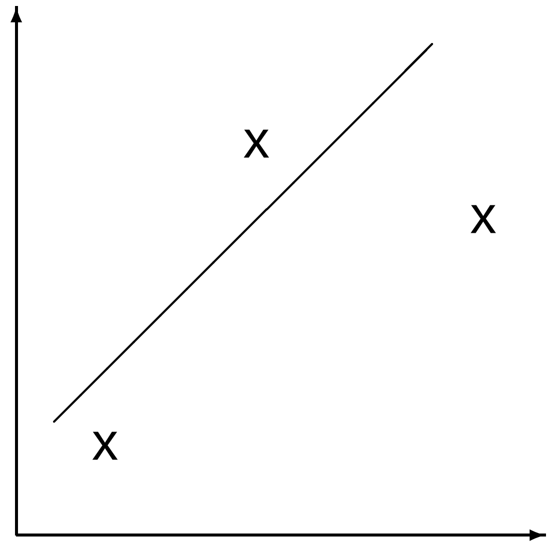
Each physical *process component* becomes an independent *code component* that can be separately instantiated, initialized, stepped forward, and terminated.

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.



Model sizes and times

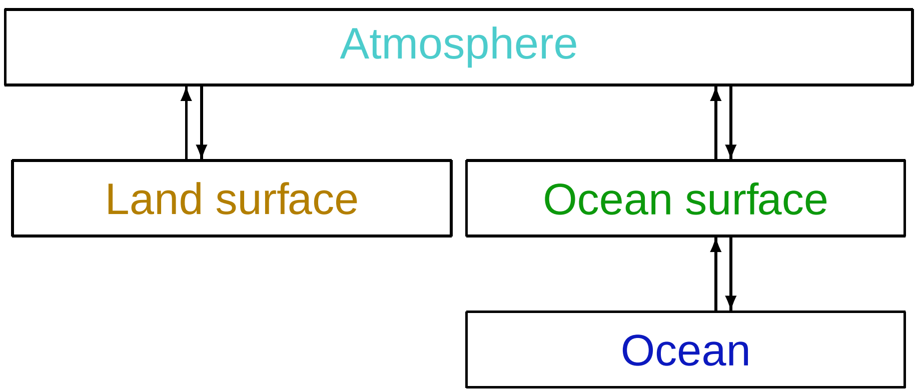
Climate Typically measured in simYears/wallDay. Resolution is adjusted till an experiment runs within a time useful for research. Current high-end research: $\sim 10^2$ variables integrated for $\sim 10^6$ timesteps on $\sim 10^6$ gridpoints. (Earth Simulator considerably raises this bar, but there is currently no relevant experience to compare against).

Weather Operational forecasting is measured in simHours/wallHour. Resolutions about $4{\times}4$ higher; $\sim 10^3$ timesteps.

Since climate simulations involve very long runs of a choatic dynamical system, results are extremely sensitive to small changes at the bit level (platforms, compilers, libraries, algorithms...) *Ensemble* simulations are used to reduce uncertainty.

Data exchange

Data exchanges between components take the form of fluxes of heat, momentum and tracers.



Conservation is very important for climate models; less so for weather.

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.

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.

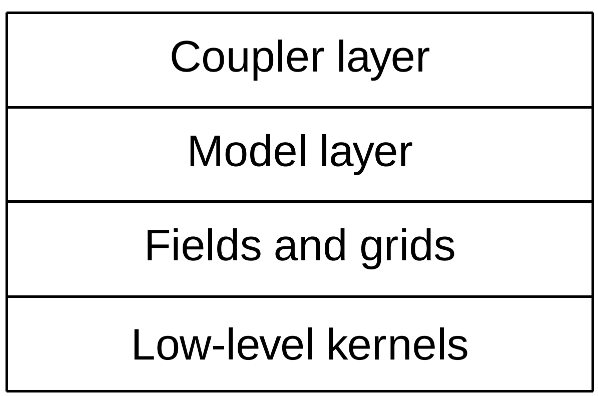
Prototype frameworks based on standards and component-based design began to appear in the climate modeling community starting in 1997 (FMS: GFDL Flexible Modeling System, ...)

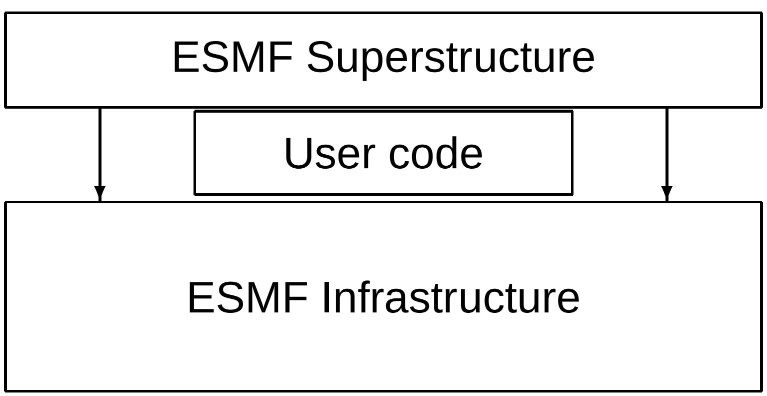
ESMF

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



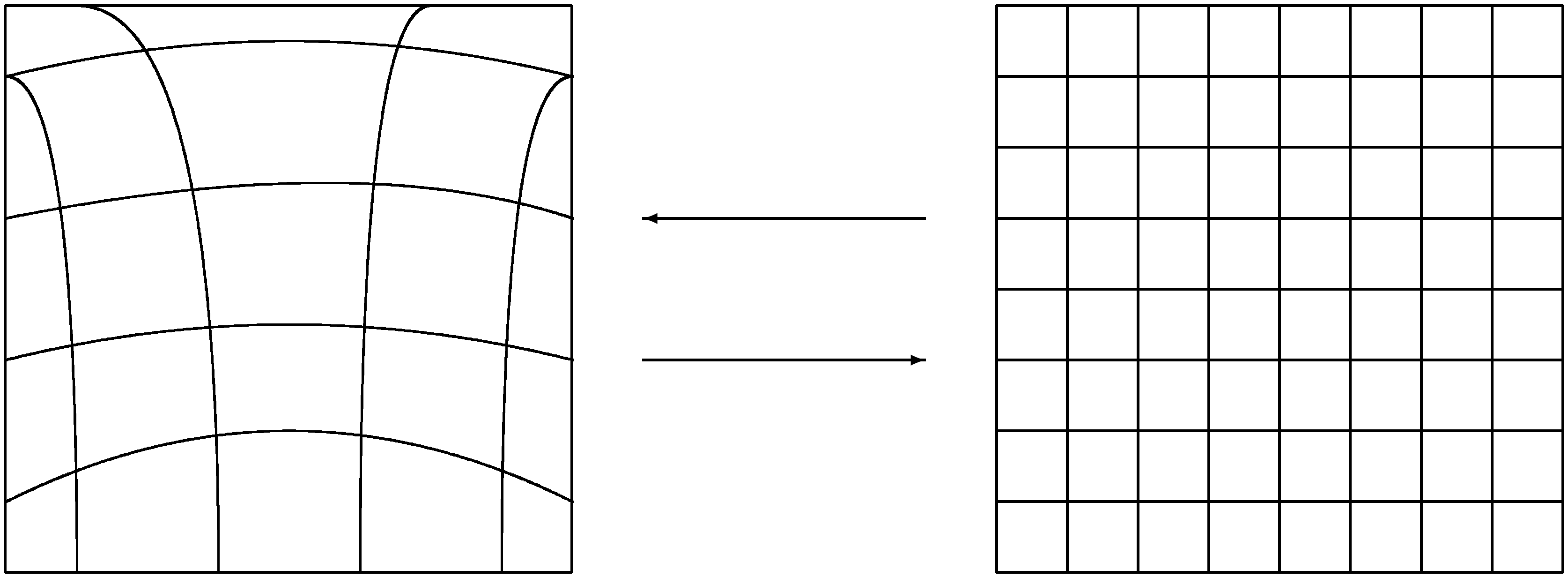


ESMF features

- 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.

Interpolate data between grids: `reGrid`

When components on independent **physGrid**s must exchange data, they require a **reGrid**.

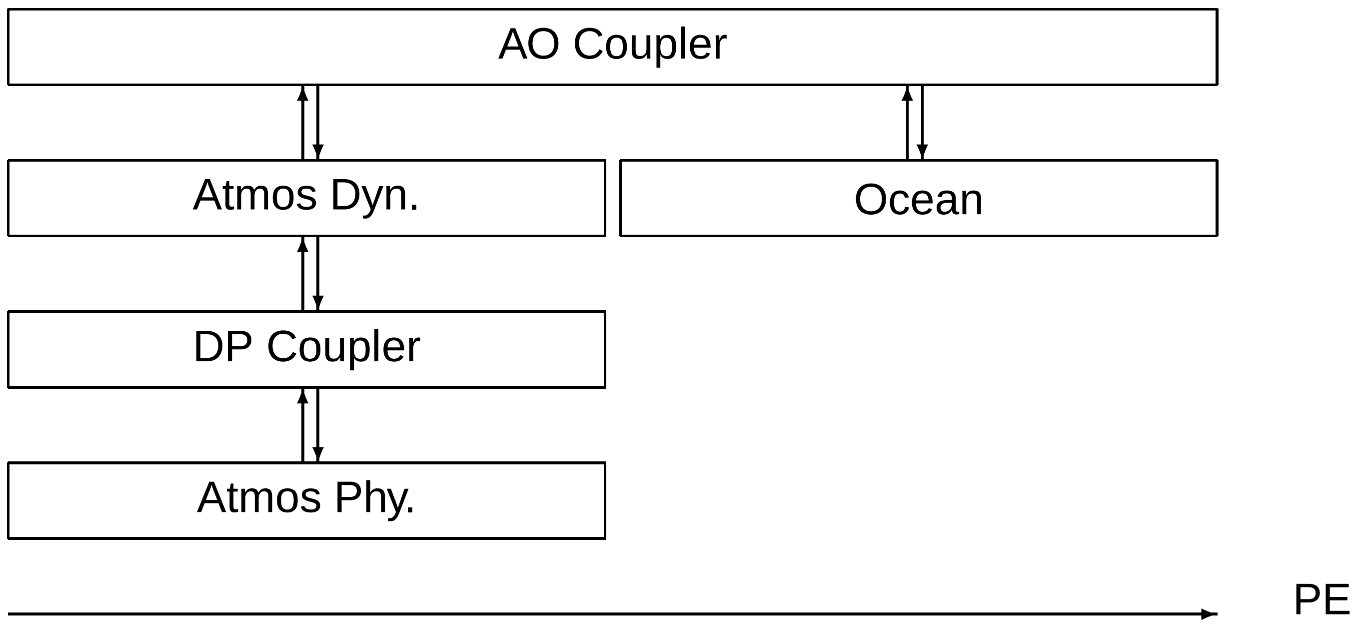


On parallel systems, we require all calls to be local; the **reGrid** object contains the optimized **route**s for communication. Typically invoked $\sim$ 10 msec.

Model composition

We develop a hierarchy of components and assign them to PEs as appropriate.

A **coupler** mediates between pairs of components needing to exchange data, and matches fields in their **import** and **export States**.



The coupler runs on the union of PEs of its components. Typically invoked $\sim$ 100 millisecond.

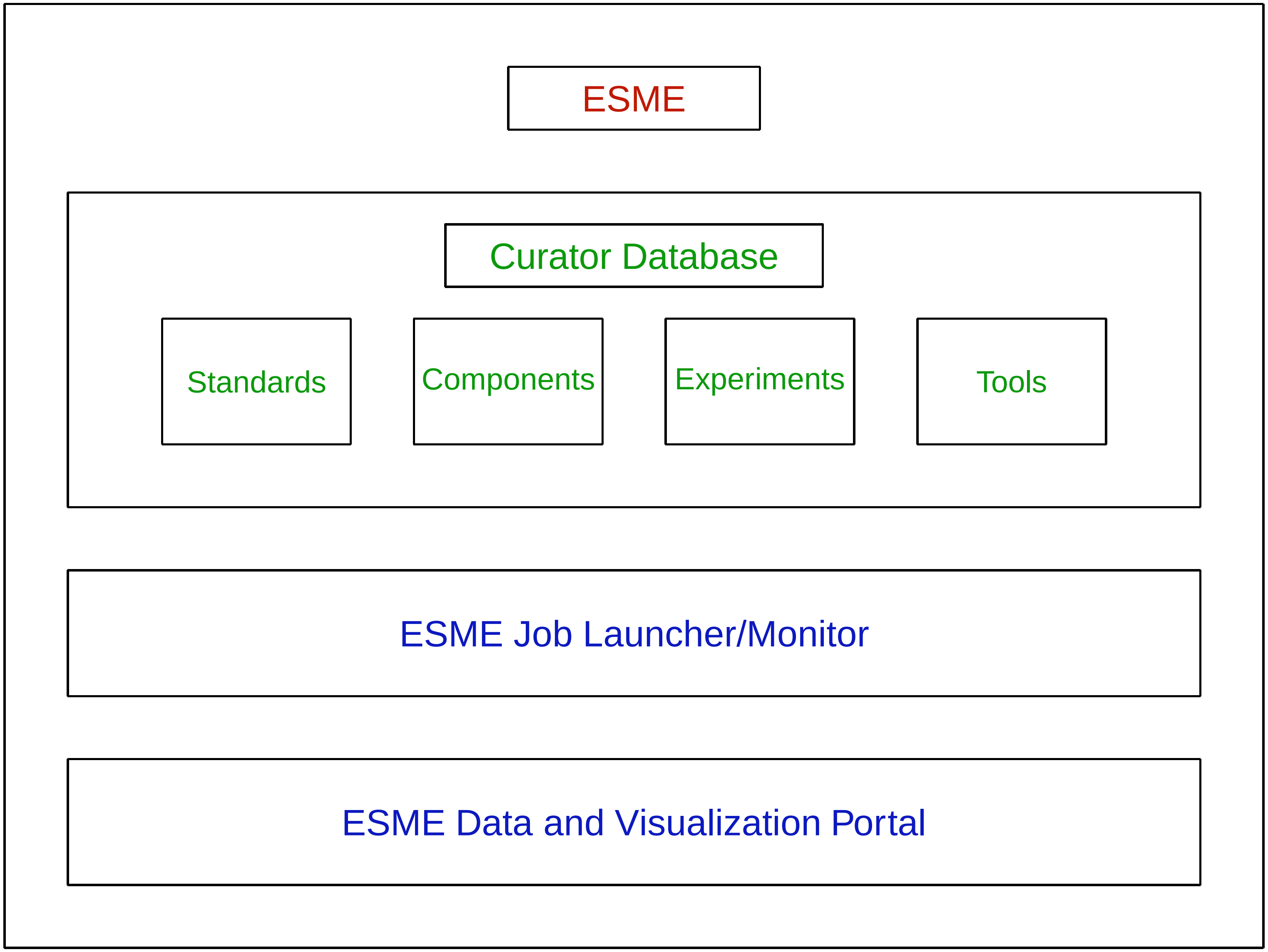
Link to data frameworks

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:

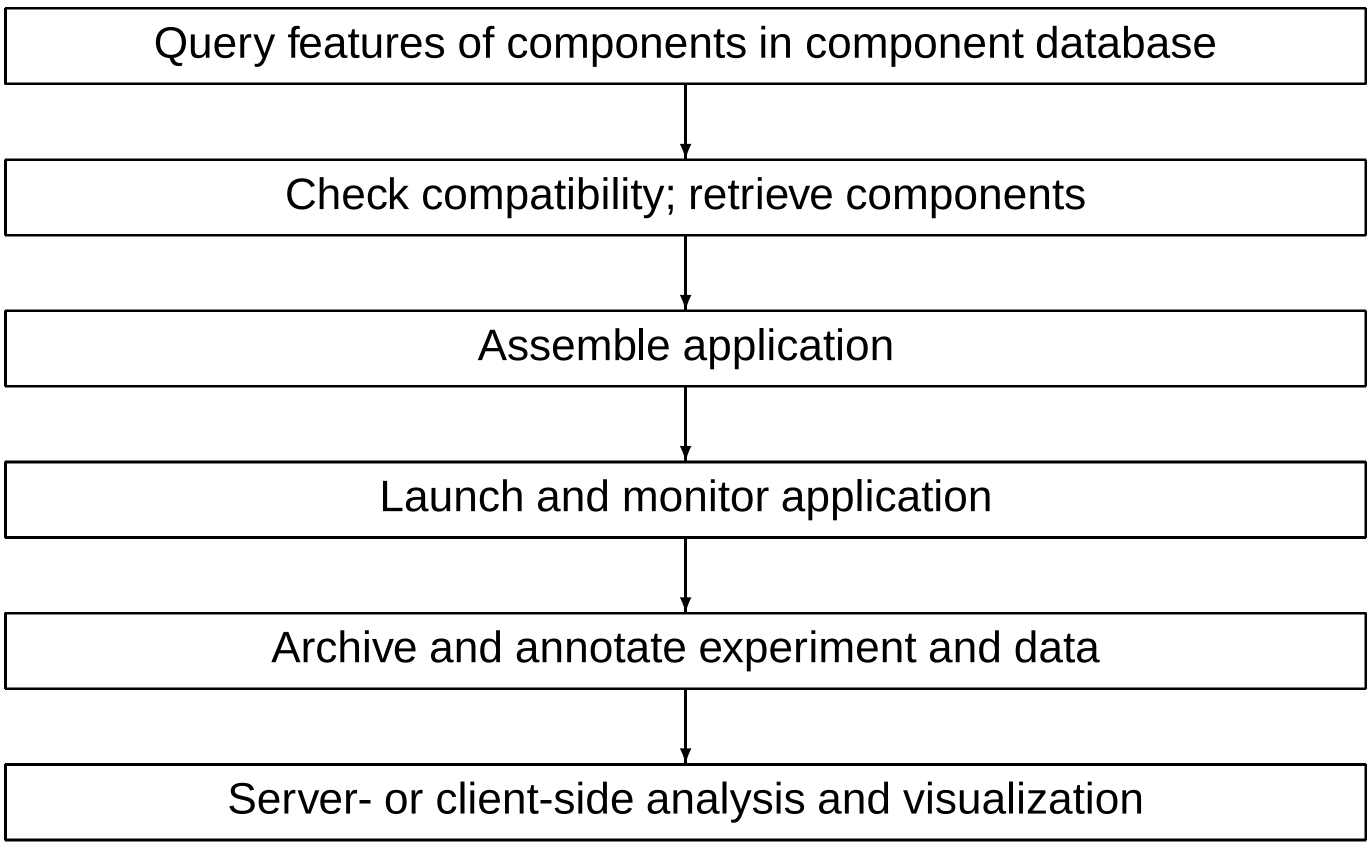
- Description of model components: e.g FMS BGRID atmosphere, land and sea ice coupled to MITgcm ocean.
- Description of grid configurations and resolutions.
- Choice of physics packages and input parameters.
- Model state and its fields.

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.

Structure of the ESME



ESME workflow



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.

Implications for GRID technologies

- Data GRIDs supporting both server- and client-side data analysis, are essential, and are already actively used by the Earth system science community.
- Computational grids are generally not useful for the foreseeable future for comprehensive coupled Earth system models: network *latencies* are prohibitively large.
- Other kinds of climate experiments (large ensembles: see e.g `http://www.climateprediction.net`) are designed for the GRID.
- The job launcher/monitor layer of the ESME could use GRID for job submission across many resources for short-term simulations (weather).

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.