

The Earth System Model

**Mark Schoeberl , Richard B. Rood, and Peter Hildebrand
NASA Goddard Space Flight Center**

**Carol Raymond
NASA Jet Propulsion Laboratory**

Introduction

Over the past 50 years numerical models of the earth's environment have become increasingly complex and realistic. Early numerical modeling efforts had their roots in numerical weather prediction spurred by the post World War II development of the digital computer. Useful weather forecasts are now a part of everyday life, but as was shown by Lorenz, deterministic weather prediction is limited by the non-linear fluid dynamics equations to about two weeks. However, even before such limits are reached, forecasters have noted that forecast improvements appear with the inclusion of increasingly realistic air-sea and land interactions. In other words, as the forecast is pushed further out, there is a lessening influence of the initial state and an increasing influence on the internal dynamics and atmospheric boundary forcing.

Based on the development of ocean models that can predict tropical sea surface temperatures several months in advance, the weather regime that is likely to affect a continent is increasingly predictable. This leads to the concept of forecasts that provide the probability of, for instance, drought or flood, above normal climate states. This provides useful information for a wide range of decisions such as what type of crops to plant or whether to buy road salt or sandbags. As the length of forecasts extend to decades and centuries, we become completely reliant on probabilistic forecasts. The quality of these forecasts will depend on the robust representation of the oceans, the cryosphere, the land, the biosphere, the chemistry, the atmosphere, and the physics and chemistry that connect these elements. This leads to the realization that any kind of successful climate prediction model will need to include so many processes and feedbacks that it becomes in effect an Earth System Model.

Modeling of crustal deformation, with the eventual goal of forecasting earthquakes is also a non-deterministic problem requiring a statistical physics approach. At the present, understanding of crustal rheology at appropriate scales to forecast earthquakes is lacking, but promising results have been obtained from analyzing patterns in the historical catalog of seismicity. Modeling of solid earth deformation is not at the same level of sophistication as weather and climate modeling, although the General Earthquake Model is attempting to close that gap. Dense space geodetic observations are needed to drive future model developments to rapidly advance understanding of crustal rheology necessary to improve the solid earth components of the Earth System Model. Data and modeling techniques that will enable higher fidelity earthquake forecasting will also allow better predictive capabilities for other events, such as volcanic eruptions and landslides.

Development of Earth System Models

Current climate models, that are on their way to becoming Earth System Models, are constructed by coupling atmosphere, ocean, ice, chemistry, biosphere, and land models. For example, the National Center for Atmospheric Research's Community Climate System Model (CCSM) (<http://www.ccsm.ucar.edu/>) consists of four coupled models: atmosphere, ocean, land and sea ice. The models have to be carefully synchronized so that exchange of information between the models is consistent with the individual model's internal dynamics. In CCSM this inter-model information exchange is done using a special model (the Coupler), which has its own architecture.

The Solid Earth Virtual Observatory (SERVO) has the goals of handling and archiving numerous distributed heterogeneous real-time datasets, allowing seamless access to large distributed volumes of data, and development of tools for visualization, datamining, pattern recognition, and data fusion.

Another Earth System Model effort is the Japanese Earth Simulator (ES) (<http://www.es.jamstec.go.jp/esc/eng/index.html>) Project. Unlike the CCSM the ES has divided goals with atmospheric-ocean simulations and solid earth simulations separated. The ES has divided goals with atmospheric-ocean simulations and solid earth simulations separated, and also includes a number of computational goals that contribute to the Japanese government's strategy for high-end computing.

In order to focus the diverse scientific resources needed to build and validate an Earth System Model a commonly used strategy is that the component models, developed across the community, can be meshed into a framework that will allow the interfaces to work seamlessly. NASA has formally invested in the development of a multi-agency framework, the Earth System Modeling Framework (<http://www.esmf.ucar.edu/>). There is proposed European Union Program for Integrated Earth System Modeling (PRISM, <http://www.prism.enes.org/>) which supports a similar activity in Europe.

Another important consideration in the Earth System Model is the practical constraint of available computing power. Including more and more components increases the computing requirements. Typically, since computing resources are fixed, each of the components has to be degraded to make room. For example, without the interactive ocean model, the climate model might be able to run at a 1° (~100km) resolution for 50 years in a week of wall clock time. Including the ocean model, and given the same week of wall clock time, the model might be able to run for only 5 years, or the model could run for 50 years if the resolution was degraded to 2.5°. It is well known that atmospheric simulations become much more realistic as resolution increases. Figure 1 shows an example of an ES simulation run at 10km x 10 km resolution. The realism of the precipitation pattern – including cloud bands is impressive.



Figure 1. Precipitation pattern from the Japanese ES running at 10km x 10km resolution. Note the cloud bands and complex structure normally not seen in 100km by 100km resolution models

Every time the resolution is doubled in atmospheric models; however, the computational load increases by a factor of four or larger. The constraint of fixed computing resources forces the modelers to have to balance the improvements through inclusion of more realistic physical processes versus improvement that arise from increased grid resolution.

An obvious alternative to the dilemma posed above is to increase the amount of “off-line” model testing. Off-line testing is the procedure where the interaction between the models are set so that the information only flows in one direction. For example, the atmosphere is run once. Then the ocean and atmosphere models are run separately at higher resolution to test the sensitivity of the results to resolution or to test the model internally.

3.0 Design Goals of The Earth System Model

The design of the Earth System Model must be oriented to its science objectives and these objectives must be considered in the inevitable design compromises. The model design elements should also include some of the following:

- Modular construction so that physics, dynamics and transport modules can be exchanged. Design standards are already being established for atmospheric models under the Earth System Modeling Framework (ESMF) (<http://www.esmf.ucar.edu/>) initiative, and will be under the Solid Earth Virtual Observatory (SERVO). The goal of ESMF is “to develop a robust, flexible set of software tools to enhance ease of use, performance portability, interoperability, and reuse in climate, numerical weather prediction, and data assimilation applications.” ESMF and SERVO are funded by NASA’s Earth Science Technology Office (ESTO).
- Scalable implementation of component models so that resolution can be adjusted easily. Scalability needs to include the adjustment of physics packages as well. For example, cloud

parameterization schemes used on a 1 km grid will be significantly different from the same scheme operating within a 100 km grid.

- Standardized diagnostics packages and criteria so that rational testing and comparison of modules can be performed. The community needs to agree on certain tests of model performance to prioritize improvements and to understand model shortfalls.

3.1 Assimilation of earth observations to produce climate records

A fundamental aspect of Earth science is the collection of observations. These observations are used to motivate theories and hypotheses that describe the observations. As these theories develop and as interdependencies are discovered in the observations, models are used to express the theories. With the models the observations and theory are synthesized, reconciled, and predictions are made which can be verified by observations. Successful forecasts substantiate theory, and failed forecasts point the direction to new discoveries. The blending of information from observations and models can be performed in many ways. Data assimilation is a formalized process by which observational data is melded with a model. Data assimilation is increasingly important to Earth science. The original successes of data assimilation were to improve the initial conditions of weather forecasts; thus, extending the useful predictive skill from a couple of days to more than a week, beginning to approach the Lorenz limit. Today data assimilation is being used to initialize ocean models to support seasonal prediction. Following these improvements to prediction, data assimilation has been extended to climate and chemistry applications. Long-term reanalysis data sets have been performed in both the United States and Europe, and these data sets have been used in thousands of publications in the last decade. Aside from providing global time series of observed parameters, assimilation also provides estimates of unobserved parameters. For example, a land assimilation model could be updated using space borne soil moisture and biomass data and then queried for the CO₂ flux which would not be observed.

Assimilation systems based on the Earth System Model will clearly produce an improved observational analysis, which can be used in developing climate data sets. However, work with atmospheric assimilation models has shown that while, to first order, data assimilation procedures produce a superior analysis products compared to non-assimilation methods, the process of inserting observations into the model generates inaccuracies that limit their use in climate studies. Thus a consistent data-driven representation of the Earth system remains illusive. In general, the closer the atmospheric model's climatology is to the real climate, the less the error that will be generated by the data insertion process. The important point here is that the success of the assimilation approach is not just dependent on the quality and frequency of observations, but on the capability of the base model as well.

Given models, observations, and a data assimilation system it is possible to develop more quantitative methods for studying the Earth. A high resolution Climate System Model provides an environment in which to develop all aspects of observing systems: sensor type, which parameters of measure, spatial and temporal measurement frequencies, impact of observations on forecasting problems, the ability to determine tracer and energy budgets. Early implementation of Observing System Simulation Experiments (OSSE's) have already contributed to the design and use of observing systems that measure ocean surface winds. In other applications, information from assimilation systems has been used to improve the ability to

extract information from satellite radiances and to determine subtle changes in observation quality that confounds our ability to extract climate signals.

4.0 Computing Requirements

Earth System Model computing requirements are difficult to evaluate in any kind of quantitative way because component models are still under development. Significant leaps forward in computing such as the Japanese Earth Simulator have inspired the community that focused initiatives on computing may produce significant progress. Despite the uncertainty in computing requirements or capability certain definitive statements can be made:

- 1) Regional climate modeling will be the future goal of nascent earth system models. Regional models will need to have a grid resolution of 30km or better although by 2010 coupled global atmosphere ocean climate models making 100 year runs will probably only be able to run with 50 km atmospheric resolution with 50 atmospheric levels.
- 2) Computational costs for fluid dynamical models increase by a factor of 8 each doubling of the resolution. Current models have a grid size of 100 km with 55 levels. Future weather prediction models (2010) are expected to have a grid size of 10 km and 100 vertical levels. It is not known if the parameterization schemes follow the same law.
- 3) Climate predictions are stochastic which means that ensembles of runs will have to be made to determine a future climate state or to quantify state transitions. To be statistically robust, hundreds to thousands of runs will need to be made.
- 4) The data assimilation process currently exceeds the computational cost of running the numerical model.
- 5) Computing capability continues to approximate Moore's Law which means that computing power for a fixed cost will double about every 1.5 years. Using item 2, this means that models can double their resolution every 4.5 years assuming there is sufficient budget to replace the computer system. Table 1 below summarizes the projected computing requirements. The reader may note that the goal of 50 Tflop performance indicated below is just beyond the ES theoretical performance of 40 Tflops. Thus it appears that earth simulation is within the reach of today's researchers.

	2002 System	2010+ System	
Resolution <ul style="list-style-type: none"> • Horizontal • Vertical levels • Time step • Observations <ul style="list-style-type: none"> ○ Ingested ○ Assimilated 	100 km 55 30 minutes 10^7 / day 10^5 / day	10 km 100 6 minutes 10^{11} / day 10^8 / day	
System Components:	Atmosphere Land-surface Data assimilation	Atmosphere, Land-surface, Ocean, Sea-ice, Next-generation data assimilation Chemical constituents (100)	
Computing: <ul style="list-style-type: none"> • Capability (single image system) • Capacity (includes test, validation, reanalyses, development) 	10 GFlops 100 GFlops	Must Have 20 TFlops (2000x) 400 TFlop (4000x)	Important 50 TFlops 1 PFlops

Table 1 above summarizes an estimate of computing requirements needed for Earth System Modeling from NASA's Earth Science Enterprise.

- 6) Storing the observations and model output in easily accessed data base will also be a challenge. Current experience with the EOS Data Information Systems shows that providing quick access to data archived in self describing formats (i.e. HDF) has been a challenge. Even with the 2003 cost of hard drive storage approaching \$1 per gigabyte, the 10 PB/day requirements shown in the table below implies the expenditure of \$10 M/day.

	Observational Data	Access Modes Rates	Output Data	Storage Term/Re-access Mode
Weather Forecast	1 TB/day Multiple Sources Continuous	Streamed input 20 GB/s	10 PB/day – Archival 10 TB/day – external distribution	Medium – Long Catalogued
Climate Modeling	10s of GB from archival sources	Data archive request 2 GB/s (latency tolerant)	100s TB/day	50% Short term - Immediate analysis 50% Medium term - Catalogued
Solid Earth Research	100s of GB/day Distributed sources	Distributed archives – low latency access	1 PB/day – ingested into distributed archives	Medium – Long Catalogued access

Table 2 summarizes an estimate of storage requirements needed for Earth System Modeling from NASA's Earth Science Enterprise.

5. Summary

The Earth System Model is the natural evolution of current climate models and will be the ultimate embodiment of our geophysical understanding of the planet. These models are constructed from components – atmosphere, ocean, ice, land, chemistry, solid earth, etc. models and merged together through a coupling program which is responsible for the exchange of data from the components. Climate models and future earth system models will have standardized modules, and these standards are now being developed by the ESMF project funded by NASA.

The Earth System Model will have a variety of uses beyond climate prediction. The model can be used to build climate data records making it the core of an assimilation system, and it can be used in OSSE experiments to evaluate.

The computing and storage requirements for the ESM appear to be daunting. However, the Japanese ES theoretical computing capability is already within 20% of the minimum requirements needed for some 2010 climate model applications. Thus it seems very possible that a focused effort to build an Earth System Model will achieve success.