The computational future for climate and Earth system models: on the path to petaflop and beyond

BY WARREN M. WASHINGTON*, LAWRENCE BUJA AND ANTHONY CRAIG

National Center for Atmospheric Research (NCAR), 1850 Table Mesa Drive, Boulder, CO 80305, USA

The development of the climate and Earth system models has had a long history, starting with the building of individual atmospheric, ocean, sea ice, land vegetation, biogeochemical, glacial and ecological model components. The early researchers were much aware of the long-term goal of building the Earth system models that would go beyond what is usually included in the climate models by adding interactive biogeochemical interactions. In the early days, the progress was limited by computer capability, as well as by our knowledge of the physical and chemical processes. Over the last few decades, there has been much improved knowledge, better observations for validation and more powerful supercomputer systems that are increasingly meeting the new challenges of comprehensive models. Some of the climate model history will be presented, along with some of the successes and difficulties encountered with present-day supercomputer systems.

Keywords: climate modelling; Earth system modelling; climate change

1. Introduction

The history of weather and climate modelling shows that progress is often limited by access to top-of-the-line supercomputers. Even with these powerful machines, a 100-year climate simulation at even modest resolutions can require hundreds of thousands of processor hours, or more. As advances in supercomputer technology increase, the speed and memory of the available systems also improve. Recent history shows that the climate model complexity has also grown correspondingly, with both improved and more realistic treatment of physical processes, such as clouds, precipitation, convection, surface hydrology, vegetation and boundary-layer interactions, as well as ocean and sea-ice interactions. Of course, the modelling community cannot wait forever for the ultimate supercomputer before carrying out useful research. Typically, climate modellers carefully balance the resolution, treatment of dynamics, level of physical process detail and overall

*Author for correspondence (wmw@ucar.edu).

One contribution of 24 to a Discussion Meeting Issue ‘The environmental eScience revolution’.
experiment design for a particular climate model with the available supercomputer capability. If one is too ambitious, then it is possible to have a model that executes too slowly to be a useful research tool, or if one is too conservative, it is possible to end up with a climate model that is not state-of-the-art.

A description of the pros and cons of the various computer architectures used in climate and weather modelling can be found in the United States National Research Council Report (2001) entitled *Improving the effectiveness of U.S. climate modeling* (2001). The bottom line is that all current computer architectures have serious limitations when applied to climate modelling. These limitations offer a challenge for the modellers and their computational colleagues to find a ‘sweet spot’ for a particular computer system. The sweet spot is defined as the optimal intersection of real-time integration rate and computational
efficiency. For example, if the execution rate of the computer program does not increase past a certain number of processors, then it does not make sense to use more than that number of processors.

2. Physical processes in climate models

To understand the role of computer architecture in the context of climate, it is important to first describe the composition of the state-of-the-art climate models. The present-day climate models are composites of dynamic models representing each of the major components of the Earth’s climate system. In a sense, they are not yet really complete Earth system models that are designed to deal with all the issues of global change and all the details involved in the understanding of past climates, such as, for example, including all the complexity of biogeochemical cycles or the interactive impacts of mankind and land cover. The standard climate model components are an atmospheric model, an ocean model, a combined land–vegetation–river transport model, which is sometimes a part of the atmospheric model, and a sea-ice model. Some of the climate model versions have embedded chemical cycles, such as carbon, sulphate, methane and...
nitrogen cycles, which are treated as additional aspects of the major components. Figure 1 shows a schematic of the various components used in the present-day climate models. The solar and infrared radiation, different cloud types, mountains, river hydrology, snow and soil moisture, vegetation, land cover, ocean and sea ice are interactive components of the present-day climate system models. One of the most important additional features of the present-day climate models is the addition of various atmospheric aerosols, such as dust, sea salt, sulphate and carbon. Each of these has different sources, transport and radiative properties, which are explicitly included. The recent introductory book by Washington & Parkinson (2005) describes the basic elements of the climate models, the numerical methods and examples of their use. The book has internet links where additional information can be obtained.

3. Resolution requirements

Another important attribute of the climate models is their vertical and horizontal grid resolutions. The computation time of a model with high spatial resolution can take too much real ‘wall clock’ time to be useful for simulations of the order of 100–200 years. Thus, the climate modeller must make compromises in resolution in order to perform a realistic set of simulations, while still completing the integration in a reasonable amount of time. Furthermore, the amount of detail in the physical and chemical processes is a crucial factor in the computing cost of a climate model. Most modelling groups work intensely to increase the realism of the physical processes simulated by their model. Early in the development of the climate models, the general philosophy was to keep the physical processes quite simple because of computer limitations. However, as we learned more from observations and modelling about ‘how the real climate system works’, combined with advances in supercomputers, we are now able to simulate the complex interactions and feedbacks in the climate system at a level of detail and realism never before possible. We have also learned from observational studies that we need a particular resolution to resolve a certain phenomena. Including hurricanes or tropical cyclones in a global climate model requires horizontal resolution of the order of 10–20 km. The oceans have even smaller eddies and narrow current systems, such as the Gulf Stream, which should be resolved or accounted for by a small set of parameters, while in the atmosphere, the most energetic waves or eddies are mostly of a larger scale.

One of the most serious shortcomings in the climate models are the biases. Some of these biases are caused by our limited understanding or the ability to model how various components of the climate system work, such as clouds or precipitation. Through the use of observations, especially observational field studies, we are gaining new insights into how to represent these aspects of the climate models. Not only does this provide a pathway for improving the model, but it helps the researchers reduce the biases. Although progress has been made, the models still have sizeable biases.

Figure 2 shows the various horizontal resolutions of a typical atmospheric model component that uses a spectral transform technique on the sphere. Starting from figure 2a, we show a rhomboidal truncation 15, which has an approximate 400–500 km grid size. This resolution was used mostly in the 1970s
and the 1980s. In the 1990s, many modelling groups used a triangular truncation of T42 (figure 2b) which is of approximately 300 km grid size. T85 (figure 2c), with a resolution of approximately 160 km, and the T170 (figure 2d), or its equivalent, will probably become the norm over the next few years. Higher resolution studies are presently underway, but they are not currently extensively used for century-scale climate simulations because of computer limitations. With increasing resolution, important high-gradient features, such as mountains, coastlines and ocean bottoms, are more realistically resolved. Note, however, that the ocean and sea-ice components often used in fully coupled models run at resolutions near 60 km or less in some regions. There have been some shorter term global spectral atmospheric simulations with an approximate 10–20 km grid size on Japan’s Earth Simulator (see http://www.es.jamstec.go.jp/esc/eng/) that show impressive smaller scale features, such as cyclones in the western Pacific region and more realistic weather frontal structures. However, such high resolutions are still beyond the reach of most modelling groups interested in performing century and longer time-scale simulations. Comparable high-resolution atmospheric studies are being used with novel finite-difference or finite-element dynamical core atmospheric models. Ocean and sea-ice models typically mostly use finite-difference methods of solution. Note that the ocean bottom is better resolved in the figure below with increased resolution.

Figure 3 shows the history of development of the different climate system model components. It is clear from the figure that the climate models have become more comprehensive over the last few decades. Each component is constantly being evolved on an almost yearly basis. Many of the modelling groups use a flux coupler to link the fluxes of energy, momentum, moisture and heat transfers between the various components. The coupled models were first used in the 1970s and the 1980s. The addition of ice sheet modelling has special challenges because the ice sheet streams that actually go into the ocean are of much smaller space scale than the climate model spatial scales. Clearly, innovative techniques will have to be invoked to deal with this subgrid scale ice sheet problem. The size of the community involved with developing the models has also increased, starting with individual ‘hero’ and small group researchers in the 1960s, growing into medium-sized development teams in the 1970s and the 1980s and expanding now to large, internationally distributed, interagency and interdisciplinary communities numbering in the hundreds.

4. Brief history of the treatment of the poles and the search for a more uniform global grid system

Williamson (2007) has written an excellent review article on how to deal with the problem of numerical difficulties near the poles and the various numerical techniques that have been proposed to deal with it. If the model equations are written in spherical coordinates (latitude and longitude), then as the grid approaches the pole, the meridians converge and the longitudinal grid interval distances become very small, which limits the time increment that can be used to solve the equations. Figure 4 shows a typical equally spaced latitude and longitude grid system. The earliest weather prediction models used in the 1950s applied
conformal map projections to give a more uniform grid system, but the climate modelling community has mostly used a global spherical coordinate system. As Williamson points out, one of the earliest suggested methods for avoiding the pole problem was proposed by Sadourny (1972). It is based on a regular polyhedron circumscribed to the sphere, where each face of a cubed sphere has

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>atmosphere/land surface</td>
<td>atmosphere/land surface/vegetation</td>
<td>atmosphere/land surface/vegetation</td>
<td>atmosphere</td>
<td>atmosphere</td>
</tr>
<tr>
<td>ocean</td>
<td>ocean</td>
<td>ocean</td>
<td>ocean</td>
<td>ocean</td>
</tr>
<tr>
<td>sea ice</td>
<td>sea ice</td>
<td>sea ice</td>
<td>sea ice</td>
<td>sea ice</td>
</tr>
<tr>
<td>sulphate aerosols</td>
<td>sulphate aerosols</td>
<td>sulphate aerosols</td>
<td>sulphate aerosols</td>
<td>sulphate aerosols</td>
</tr>
<tr>
<td>solar forcing volcanic aerosols</td>
<td>solar forcing volcanic aerosols</td>
<td>solar forcing volcanic aerosols</td>
<td>solar forcing volcanic aerosols</td>
<td>solar forcing volcanic aerosols</td>
</tr>
<tr>
<td>carbon cycle</td>
<td>carbon/nitrogen cycle</td>
<td>carbon/nitrogen cycle</td>
<td>carbon/nitrogen cycle</td>
<td>carbon/nitrogen cycle</td>
</tr>
<tr>
<td>vegetation</td>
<td>interactive vegetation</td>
<td>interactive vegetation</td>
<td>interactive vegetation</td>
<td>interactive vegetation</td>
</tr>
<tr>
<td>biogeochemical cycles</td>
<td>biogeochemical cycles</td>
<td>biogeochemical cycles</td>
<td>biogeochemical cycles</td>
<td>biogeochemical cycles</td>
</tr>
<tr>
<td>ice sheet</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 3. The time history of the climate model components and coupled climate model development (past, present and future). In order to tie together all of these components, a flux coupler is used. The flux coupler allows passing variables and fluxes of energy, heat, momentum and moisture between components. It should be pointed out that most modelling centres only have computers in the teraflop range ($10^{12}$ floating point operations per second) or roughly equivalent to many thousands times faster than a standard personal computer.

As Williamson points out, one of the earliest suggested methods for avoiding the pole problem was proposed by Sadourny (1972). It is based on a regular polyhedron circumscribed to the sphere, where each face of a cubed sphere has

Phil. Trans. R. Soc. A (2009)
coordinates that do not overlap (figure 5). Note that each face has a common boundary with adjoining faces. The finite differences used in the equations at the boundaries of the faces are solved by a method based on conservative principles rather than simple interpolation.

Another type of grid used in the 1960s and the 1970s was the reduced grid or Kurihara grid that was actually devised independently by a number of researchers in several different forms. Kurihara (1966) wrote in detail about this grid (figure 6). Basically, the method is to increase the longitudinal distance as the pole is approached such that the time step does not have to decrease. Often researchers use either spatial or Fourier filters to eliminate the smaller scale features near the pole in either the regular spherical grid system or the reduced grid system. In any case, this method has some undesirable computational features that have plagued climate modelling for decades. In fact, these problems have revived interest in the use of Buckminster Fuller’s geodesic dome idea, in which the globe is covered with ‘nearly’ uniform triangles as shown in figure 7. Some of the earliest work carried out in this area was by Sadourny et al. (1968) and Williamson (1968).

The spectral transform method became widely used in the 1970s for solving the dynamical equations of atmospheric climate models. The basis for this method is akin to the harmonics in musical instruments or harmonic analysis, where a variable used in a model can be represented in terms of a series of sine and cosine functions if the area is a rectangle. Most introductory physics
textbooks have a discussion of the mathematical representation of a vibrating string. On the sphere, spherical functions are used, which can be expressed in terms of latitude and longitude.

One of the benefits of the spectral transform method is that it provides a natural spatial filter with a given spectral truncation, thus the short longitudinal distances near the poles do not exist. This technique has other very attractive properties, e.g. advection is quite accurate compared to corresponding equivalent finite-difference methods. One reason the method was so successful was the capability of efficiently solving the nonlinear horizontal advection equations by fast Fourier and Legendre transform techniques. However, over time there was growing dissatisfaction with the transport of water vapour and chemical constituents that are always positive definite quantities and they should be both locally and globally conservative. In addition, near mountains and coast lines, there is some spectral ‘ringing’, which results in undesirable computational features that are non-physical. Some of these problems are not only limited to spectral but can also occur with standard finite-difference numerical schemes that can allow non-physical values to occur.

One of the innovative approaches to getting a more uniform grid structure is a composite approach in which two or more overlapping grids are used. Fluxes and variables are transferred between grids by interpolations. Kameyama et al. (2004) proposed a quasi-uniform composite grid system with spherical geometry. The advantage of this system beyond being almost uniform is that it has no

Figure 5. Cubed sphere grid designed by Sadourny (1972).
singular points similar to the standard pole point. They named the grid ‘Yin–Yang’. Figure 8 shows the two grids separately and the composite grid. This grid configuration has some advantages as both grids are based on spherical coordinates and the grids are orthogonal to each other; however, extensive interpolation is needed where the two grids overlap, as shown in figure 8c.

One of the solutions to many of the computational problems was to use semi-Lagrangian transport methods for water vapour and the chemical constituents, along with spectral transport for the other variables, but these models were still not completely conservative. Another factor that became evident for the climate modelling community is the ascendency of the massively parallel computer systems. The spectral transform method is difficult to efficiently implement on modern supercomputer systems that now use thousands to hundreds of thousands of processors because of the need to perform global sums. Some in the climate modelling community started moving towards local conservative grid point flux methods. Lin & Rood (1996) and Lin (2004) developed a new paradigm for the transport, often referred to as flux-form semi-Lagrangian. The method is usually referred to as the finite-volume method. This method is very stable with long time steps, which is important for transport near the poles and has built into it a monotonicity constraint that prevents negative values, which is very important for the transport of water vapour and chemical constituents. Note that this method requires the gravity wave part of the solution to be treated
explicitly. In order to deal with this problem, a longitudinal filter for the gravity wave part of the solution is used to prevent the need for a short time step. Another important feature is to capture the kinetic energy spectra near the grid scale. Lin has added a small diffusion of divergence to produce a kinetic energy power law that is close to the observed. In order to make the Lin–Rood finite-volume method

\[ \text{Figure 7. The grid structure for a spherical geodesic or icosahedral grid. Note that this is not a regular grid in which each of the basic elements are the same size.} \]

\[ \text{Figure 8. The Yin–Yang grids. (a) A Yin grid, a low-latitude, latitude–longitude grid with a gap in the longitude oriented as the traditional latitude–longitude grid. (b) A Yang grid, the Yin grid rotated 90° to fill the gap in the Yin grid and to cover the Polar regions left open in the Yin grid. The gap is on the back. (c) A Yin–Yang grid, the combination of the Yin and Yang grids showing the overlap of the two grids. (Adapted from Williamson 2007).} \]
The method is based on upstream cells, rather than the more usual upstream grid points used in the more conventional semi-Lagrangian numerical methods. Thus, the mass within the upstream cell gives a forecast to the arrival grid cell. As Williamson explains, in order to make the forecast of mass, the upstream cell has to be determined from the mass in the surrounding grid cells by interpolation. This requires a conservative remapping of the mass.

Before leaving this section, it should be pointed out that ocean models have also developed innovative grid systems to avoid the pole problem. The Parallel Ocean Program is an ocean model that uses a tripole grid that avoids the North Pole problem in the Arctic Ocean. Note the computational poles are over land grid points near Alaska and northwest Russia. The third computational pole is at the geographical South Pole. The model equations were modified and discretized to allow the use of any of the three locally orthogonal horizontal grids. Such a displaced pole leaves a smooth, singularity-free grid in the Arctic Ocean. The Northern Hemisphere grids join smoothly at the equator with a standard Mercator grid in the Southern Hemisphere (see Murray (1996) for more details).

Approaches for solving the climate equations on the globe are constantly undergoing evolution with many computational scientists involved in experimenting with a wide spectrum of different approaches. This area of computational science research is unsettled. Whatever choices are made, it will be important that they can be computationally efficient on modern massively parallel supercomputer systems. Depending on the approach taken and the type of computer system, there is great opportunity for clever and innovative solutions to be found.

5. Methods of execution of climate models on supercomputers

One of the challenging issues for climate modellers is to efficiently and accurately couple components on present-day computer architectures. As mentioned earlier, a coupler is responsible for merging fields between model components; mapping fields onto different grids, coordinating communication between models and sequencing the component models in time. The coupler transfers state variables as well as energy, heat, water and momentum fluxes between model components. Jones (1999) of Los Alamos National Laboratory developed a conservative
mapping scheme that allows the use of different horizontal grid systems for model components, and several modelling groups are using that mapping tool.

Figure 9 shows an example of executing a coupled climate models on multiprocessor computer systems. This method is used by many modelling groups including the NCAR Community Climate System Model (CCSM; see Collins et al. 2006). In this approach, each model component is assigned a certain number of unique processors that are used to integrate that component forward in time. At regular intervals (typically 1 hour to 1 day), the components pass variables and fluxes to each other via a coupler. To resolve the diurnal (day–night) cycle properly, the atmosphere, land and sea ice typically communicate hourly. Because the ocean has a large heat capacity, it typically communicates with the rest of the system once per simulated day. The best performance is achieved when each of the model components completes its task at the same time. Clearly, if a component finishes its work before the others, the processors assigned to that component must sit idle until the slowest component is finished. A series of load balancing tests are performed to optimize the models’ use of the available processors. However, there will always be some imbalance because, at certain times of the year and over parts of the globe, a specific component may have more or less work. Another technique that has been used to couple the climate models is the sequential integration method used by the DOE supported parallel climate model (see Washington et al. 2000), where all processors are used for all components and the components integrate sequentially in a single executable. This method is reasonably straightforward.

The climate community has been able to adapt their models to both vector and scalar computer architectures, with varying degrees of success. The optimum coding style can be quite different on different platforms and can be a function of processor scalar performance, vectorization capabilities, cache and memory hierarchy, interconnect and input/output performance. In addition, the tools available, such as compilers, debuggers, performance monitors and libraries can vary greatly in quality between platforms and/or vendors. The programming language of choice for most of the climate modelling community continues to be Fortran; however, there are some uses of other languages, including the use of object oriented programming.

It is difficult to predict future supercomputer designs, but most of the supercomputers being used by the community today are massively parallel computer systems, with a complex hierarchy of cached memory and only a modest amount of memory on the processor. The latest computer systems are pushing development in the climate community towards clever domain decompositions, new algorithms that reduce communication cost and utilities, such as parallel input/output to reduce the memory and computational cost. As a result of this work, it is becoming feasible to run multi-century simulations at much higher resolutions sooner than expected. However, a significant amount of effort is being made to redesign the models for the latest architectures. Scaling improvement is a direct result of today’s hardware being better balanced. In the past, we had relatively fast processors and less capable memory and communication systems. More recently, as manufacturers strive to reduce power consumption and year to year processor speed increases have been less and the memory and communication systems have improved relatively more, so that we now have better balanced massively parallel computer systems.
6. Final comments

Climate modelling has had a successful history, with continually improving climate and Earth system models. The problems, limitations and results of the model simulations have been fully discussed in the International Panel on Climate Change reports (see http://www.ipcc.ch/), which is an assessment of published articles (see Randall et al. (2007) for further details). Most of the major modelling centres are addressing the very important issues of future climate change, especially global warming and its impacts. Such studies are of high importance to the public and policy makers. Other climate model studies concern understanding past climate change and the potential for abrupt climate change. In the early days, small teams of scientists and computation experts developed the climate models. Now, they are being developed by large ‘virtual’ centres over the internet involving, in some cases, large groups of scientists and computational experts. With the CCSM discussed earlier, formal management mechanisms exist to coordinate the distributed development effort and to decide what goes in to the model and what should be the desired resolution. This new way of conducting climate-modelling research must still be sensitive to innovation and the testing of alternate methods. Another important, and often neglected problem, for high-performance computing is how to handle the huge amount of data that flows from the climate model studies. The concepts in the DOE-supported Earth system Grid (http://www.earthsystemgrid.org) are addressing the very important problem of making data available to users in the broader community, even if the computations are performed at multiple supercomputer sites.

Finally, there is a perception by some in the computing community that climate modellers are not in touch with the computing community and that they are not using the most current methods for solving model equations on the present generation of supercomputers. We believe that perception is mostly in error. The majority of scientists and computational experts engaged in climate modelling have had, and continue to have, many close collaborations with their colleagues in the computational field. They continue to work together to seek the best possible computational methods, languages and programming techniques for modern supercomputer systems. Many of the ‘new’ ideas have been investigated by researchers already in the community. The need for increased high-performance computing capability and access remains a very high priority, especially given the increased national and international concerns about global climate change.

Although predicting the future is risky, we will attempt it in our final comments. We envisage the vastly improved Earth system models, which incorporate virtually all of the process that interplay in the climate and Earth system and that we can execute such models on computer systems more than a thousand times the present capability. There will be two types of research and computational scientists. One type will be the ‘generalist’ and the other type is the ‘specialist’. Both are needed to keep this team enterprise working and improving with time. We do not envisage one big research centre that conducts research and operational climate forecasting, but a relatively set of modest size operations (a few hundred staff) enabling some friendly competition between research centres. Redundancy in science, in which different approaches are pursued simultaneously, is a virtue and a necessity in science. The path forward will be very rewarding for mankind and the environment.

*Phil. Trans. R. Soc. A* (2009)
We appreciate David Williamson for providing figures on the various global grid structures being tested and used for the climate and Earth system models. Figure 1 was published as part of the NCAR’s CCSM. We thank Gary Strand (NCAR) for preparing figure 3 and Mariana Vertenstein (NCAR) for providing a version of figure 9. Stephanie Shearer helped with preparing the manuscript and editing. The authors acknowledge the support for the CCSM project from the National Science Foundation (NSF), the US Department of Energy (DOE) Office of Biological and Environmental Research, the National Oceanic and Atmospheric Administration and the National Aeronautics and Space Administration (NASA).

References


