Advanced Scientific Computing Advisory Committee and Biological and Environmental Advisory Committee

Report on Computational and Informational Technology Rate Limiters to the Advancement of Climate Change Science

March 17, 2008

Prepared by the Joint ASCAC-BERAC Subcommittee

James J. Hack, NCAR (Co-chair)
Eugene Bierly, AGU (Co-chair)
Dave Bader, LLNL
Phil Colella, LBL
William D. Collins, LBL
John B. Drake, ORNL
Ian Foster, ANL
Brian Gross, GFDL
Philip Jones, LANL
Edward S. Sarachik, University of Washington
Dean N. Williams, LLNL

Table of Contents

Acronyms ... 3
Executive Summary ... 4
Introduction ... 5
Climate Change Science ... 5
Scientific Opportunities ... 6
Rate Limiting Issues ... 7
Implications for Investments ... 10
- Models ... 10
- Observations ... 11
- Computational Algorithms ... 12
- Facilities and Infrastructure ... 14
- Computational Facilities ... 14
- Data Storage Facilities ... 15
- Networking and Data Management ... 15
- Analysis Tools and Collaborative Technologies ... 16
- Visualization Technologies ... 17
- Management ... 17

Summary ... 20
Bibliography ... 21

Acronyms

ASCR	Advanced Scientific Computing Research
ASCAC	Advanced Scientific Computing Advisory Committee
AR4	Fourth Assessment Report
ARM	Atmospheric Radiation Measurement
ASP	Atmospheric Science Program
BER	Biological and Environmental Research
BERAC	Biological and Environmental Research Advisory Committee
CAPT	CCPP ARM Parameterization Testbed
CCPP	Climate Change Prediction Program
CCSM	Community Climate System Model
CCSP	Climate Change Science Program
DOE	Department of Energy
ESG	Earth System Grid
ESM	Earth System Model
ESNet	Energy Science Network
IPCC	Intergovernmental Panel on Climate Change
LANL	Los Alamos National Laboratory
NCAR	National Center for Atmospheric Research
NRC	National Research Council
ORNL	Oak Ridge National Laboratory
PCMDI	Program for Climate Model Diagnosis and Intercomparison
PI	Principal Investigator
SC	Office of Science
SciDAC	Scientific Discovery through Advanced Computing

Executive Summary

The Office of Science (SC) has made significant and long-lasting investments in the theoretical, observational, and computational aspects of climate science. Projects supported by U. S. Department of Energy's (DOE) Biological and Environmental Research (BER) and Advanced Scientific Computing Research (ASCR) have produced major advances in measuring and simulating the climate system. At Dr. Raymond Orbach's request, a joint Advanced Scientific Computing Research and Biological and Environmental Research Advisory Committees (ASCAC-BERAC) subcommittee has reviewed the past accomplishments, active scientific questions, and attendant technical and computational issues of DOE's climate science activities. The committee did not attempt to survey the entire scope of Earth systems study, but instead focused its attention on major issues that reflect DOE's strategic interests and research portfolio. This committee finds that new strategic alliances between ASCR and BER could be instrumental in addressing the new challenges and applications for climate modeling and advancing the National benefits from DOE's leadership in climate science.

The scientific and technical challenges include how to simulate fluid motions over a wide range of scales with high fidelity and computational efficiency. A second emerging issue concerns the optimal methods for assimilating a broad range of physical, chemical, and biogeochemical measurements into models of the Earth system in order to more completely describe the state of the system. The synthesis of models and observations is critical both for understanding the present climate and for simulating its evolution over the next several decades. The major observational challenges include how best to characterize the coupled carbon cycle and quantify the complex dynamics of the hydrological cycle and its interactions with aerosols. The computational research is driven by these theoretical and observational challenges, but also by the rapid evolution of computer architectures and by the demands of building robust end-to-end facilities that support Earth system science.

Recognizing that computational and information technology solutions cannot be separated from the underlying science drivers; the committee recommends that ASCR and BER undertake joint ventures to:

- Continue to invest in leadership class computational facilities, data storage facilities, analysis environments, and collaborative tools and technologies. A significant fraction of these resources should be dedicated, configured and managed to support integrated and multi-faceted climate research and prediction across DOE and broader national and international efforts
- Invest in strategic collaborations to develop computational algorithms and scalable software to accelerate computational climate change science
- Develop computational and theoretical foundations for new modes of climate simulation, including ensemble short-range forecasts with regional fidelity and Earth system assimilation
- Focus the scientific effort to pursue robust predictive capability of lower-probability/higher-risk impacts, including climate extremes and abrupt climate change
- Develop a strong scientific understanding of leading-order uncertainties in the carbon cycle, in particular how the efficiency of natural carbon sinks will change with our changing climate

Introduction

At the August 2007 meeting of the ASCAC for the Office of Science, United States Department of Energy, Dr. Raymond Orbach, Director of the SC, charged the ASCAC committee and the BERAC to work together to identify the key computational and information technology obstacles to advancing climate change science and improving climate change projections using state-of-the-science coupled climate models. In response, the ASCAC and BERAC co-chairs assembled a subcommittee of experts to address this charge in the form of a report due by the November 2007 meeting of the ASCAC. The committee held one teleconference call and assembled in Washington on 16-17 October to complete drafting the major elements of this report.

The issues pacing progress in climate change science are far broader than historically have been, or can be, unilaterally addressed by the DOE SC. For example, the establishment of a comprehensive global climate observing system is something that is beyond the reach of a single agency and is among a number of highly important climate activities that need to be discussed, prioritized and coordinated at the agency level. Recognizing that there are obstacles to advancing climate science that require broader Federal coordination, the committee's discussions focused on DOE's strengths and opportunities to leverage broader efforts in the climate change and climate modeling community. Given the fact that other efforts are currently underway to address similar issues (e.g., an ongoing National Research Council [NRC] study of *The Potential Impact of High-End Computing on Four Fields of Science and Engineering*) and given the production schedule for the report, the committee has elected to produce a relatively short, balanced response to the charge.

In parsing the charge the committee looked at the respective investments and roles of ASCR and BER in climate change science. These investments are complementary in many respects, with a number of clear opportunities for partnerships in specific topical areas. The opportunities for partnership include increasing demands for new computational capabilities, advanced software methods, advanced algorithms and applied mathematics suitable for evolving computational architectures, data management techniques, networking solutions, and technologies to support scientific collaboration. There are also comparably important and unique investments made by BER in contributing to basic scientific knowledge about the climate system, in specialized observational programs, and in the exploration and enhancement of global modeling approaches and techniques. With this in mind, the committee decided to approach the charge by examining rate limiting issues in specific science fields, and by exploring options for optimizing investments by ASCR and BER to address these pacing factors.

Climate Change Science

The climate community is facing significant challenges and opportunities in its efforts to advance basic science and its application to policy formation. With the release of the 2007 Intergovernmental Panel on Climate Change (IPCC) assessment and the Climate Change Science Program (CCSP) reports, climate science is entering a new phase. Several of the classical problems -- in particular the detection, attribution, and "finger-printing" of climate change at global scales -- have essentially been resolved in these latest assessments. The global community is now faced with a new set of urgent problems, including robust projections of regional impacts; forecasts of abrupt and extreme climate change; simulation of shifts in the water cycle; and prognosis of carbon-cycle feedbacks. In order to address these issues, the community needs to develop and undertake a coordinated research program balanced and integrated among observation, computation, and theory. Collaborative efforts between DOE's ASCR and BER have produced world-class climate models adapted to exploit the most advanced computing platforms in existence. Meeting future challenges in climate change science will require qualitatively different levels of scientific understanding, modeling capabilities, and computational infrastructure than are currently available to the scientific community. Many of the questions now facing climate change science will require the development of a new generation of more comprehensive climate models, most frequently referred to as Earth System Models (ESMs) that predict the coupled chemical, biogeochemical, and physical evolution of the climate system. A strategic partnership between ASCR and BER could accelerate the nation's progress on many of the major new challenges in climate science.

Climate science has become increasingly international in scope. The success of national and international assessments depends to a large degree on the free exchange of observations and simulations. One of the premier examples is the multi-model archive of simulations for the IPCC Fourth Assessment Report (AR4) assembled and maintained by the Program for Climate Model Diagnosis and Intercomparison (PCMDI), which resulted from current BER and ASCR investments. The global climate community has authored 324 peer-reviewed publications based upon this archive as of October 2007. These articles have been instrumental in assessing the fidelity of global climate models and in developing new initiatives to advance the science of climate simulation. The scientific productivity of the PCMDI archive suggests that the development of comparable (perhaps virtual) collections would be tremendously beneficial for future assessment activities. However, the next generation of ESMs will challenge existing frameworks for computation, communication, and analysis. A strategic partnership between ASCR and BER, in collaboration with other US climate change efforts, will be essential to ensure that the U.S. can enhance its leadership in climate science through advances in these key fields shared with the international community.

Scientific Opportunities

One of the most promising pathways to improving our understanding of climate change has been the development of models that represent the full complexity of interactions in the Earth system as accurately as possible. Over the last 30 years, these models have advanced considerably in spatial and temporal resolution and in the representation of key climate processes. The opportunity to credibly contribute to the current global change debate is hindered by the current limits of climate models to address regional and local-scale impacts on time scales of greatest interest to society. Examples include the ability to accurately project climate change impacts on regional space scales and decadal time scales, to project changes in extreme events, and to accurately anticipate changes in low-frequency climate variability. Additional challenges include characterization of changes to the water cycle, quantification of sea level rise, and exploration of processes that might contribute to abrupt climate change. Significant near-term investment could lead to a quantitative improvement in the scientific community's ability to address these difficult but societal relevant questions.

One example of an immediate scientific challenge and opportunity is the incorporation of chemical and biogeochemical processes in climate models. The science surrounding the chemical and biogeochemical coupling of climate has become central to answering climate change questions. The resolution of the open scientific issues has become increasingly important as we learn more about how the coupled carbon cycle has changed in the fossil record, how it is changing in the present day, and how it might change in response to global climate change. Addressing the science issues will require new observations and methods of analysis, new theoretical understanding of the carbon cycle, and new models of the Earth system that include the interactions between human and natural systems. These models play pivotal roles in interpreting the paleoclimate records, in synthesizing and integrating measurements to study the current carbon cycle, and in projecting the future responses of human society and the natural world to evolving climate regimes.

Other scientific challenges are related to developing an understanding of the significant impacts that could follow from abrupt changes in the climate system. One of the largest uncertainties in current climate assessments is the rate of sea level rise. Recent observations indicate ice sheets can dissipate on much more rapid timescales than melting due to dynamical processes in large outlet glaciers and ice streams within the ice sheet. A high priority for climate models is the inclusion of fully dynamic ice sheet models and the ocean/ice shelf interactions needed to assess the rate and magnitude of sea level rise due to rapid ice sheet melting. Abrupt climate change can also result from thresholds and nonlinearities in the response of climate to slower time scale forcing of the climate system. Examples include rapid changes in ocean circulation, large scale vegetation mortality and succession, release of methane frozen in ocean and permafrost clathrates, and megadroughts and dust storms. The climate community will need to use models to identify thresholds of forcing in the climate system and explore the likelihood and impacts of such abrupt change scenarios. These are but a few of the many immediate science opportunities that can be exploited through targeted partnership investments in ASCR and BER.

Rate Limiting Issues

The community's efforts to advance climate modeling and its application to science and technology options will require advances in essentially every aspect of the models' theoretical, observational, and computational foundation. These advances represent near-term opportunities for targeted investment by ASCR and BER.

To a large extent climate science is data limited, and the success of the research is contingent on basic measurements and observations necessary to validate, verify, and constrain the formulation of ESMs. Therefore, quantifying the uncertainties in predictions is expected to require a new level of integration between modeling and observational science. New mathematical methods and algorithmic techniques will also be required to address the fundamental challenges of multi-scale coupling of physical, dynamical, and chemical and biogeochemical processes. A flexible high-performance computing infrastructure has been and will continue to be a key factor in making these advances possible.

Traditional projections on centennial time scales are strongly influenced by the future trajectory of anthropogenic emissions, while forecasts for decadal time scales are governed primarily by the past history of the ocean. Therefore near-term climate forecasts will require two *new* developments: multi-scale models that can explicitly resolve important meteorological systems at regional scales, and retrospective analyses of the global oceans to initialize the forecasts. For multi-scale atmospheric models, one of the primary challenges is the reproduction of weather and climate-related phenomena with sufficient fidelity for both meteorologists and climate scientists. The field of ocean data assimilation is still in early stages of development and exploration, and it would benefit from the transfer of adaptive assimilation methods under development for atmospheric applications. Both of these issues are explored further below.

The ocean is responsible for much of the inertia or “memory” in the climate. Ocean data assimilation will be necessary to provide an initial ocean state for decadal prediction and represents a pacing item for seasonal to inter-annual to decadal prediction. Ocean assimilation has been hampered by a lack of data, particularly for salinity and for ocean properties at depths below 1000m. Recent progress in deploying large numbers of floats and the launch of new satellites that together will measure salinity profiles will greatly improve our ability to effectively constrain ocean models with assimilation. Efforts to improve the data assimilation in current ocean climate models will require the adoption of much more advanced assimilation methodologies. For example, assimilation of data from ARGO floats with a fully coupled climate model using an extended Kalman filter has shown great promise in determining the state of the climate system, although it is computationally demanding.

Accurate projections of changes in the local frequency of climate extremes will be essential for the development of robust adaptation strategies. However, extremes represent the high-order moments of the climate system, and climate models have been designed primarily to predict the low-order moments. Much more research is required to understand how simulated extremes change with increasing model resolution and increasingly sophisticated parameterized treatments of non-resolvable processes. In particular, the relationships between extreme statistics and synoptic-scale low-frequency variability are not understood.

Better understanding of low-frequency variability is critical for the detection of climate-change signals. For Earth system modeling, it is important to characterize the natural modes of coupled variability in the carbon cycle, terrestrial ecosystems, and dynamic vegetation. It is also important to develop a better understanding of external forcing mechanisms, such as the role of solar variability in the broader context of the Sun-Earth system. Current understanding of these complex systems is limited by the length of the observational record and, more fundamentally, by open issues regarding the stationarity of climate statistics. The wide dynamic range in the relevant space and time scales complicates resolution of the coupling issues. New mathematical methods designed for multiscale systems hold promise and should be actively explored for this class of problems, and these methods should be constructed for efficient implementation in ESMs.

As suggested earlier, a large number of significant impacts could follow from abrupt changes in the climate system. These occur when the gradual increases in climate forcing trigger an abrupt transition of the coupled system to a new state. Potential examples of abrupt change include dynamic dissolution of the ice sheets and bifurcations of the ocean circulation system. Characterization of abrupt climate change requires a new paradigm for climate change modeling, one in which the models are integrated over the full range of uncertainties in forcing and parameterized physics. Exploration of this phase space will require implicit formulations of the coupled system designed for fast equilibration combined with parametric continuation techniques and sustained petascale computing.

Multiscale interactions also complicate investigations of the water cycle. As with variability, process-level understanding of the water cycle is limited by the lack of basic observations. While the absence of these data still represents a barrier to progress, near-term enhancements in computational capacity would permit the resolution of fundamental phenomena involved in both weather and climate change. Continued targeted investments in observational programs like the BER Atmospheric Radiation Measurement (ARM) program and the BER Atmospheric Science Program (ASP) could provide much of the necessary data to validate high-resolution process modeling studies of critical issues in topics like aerosol-cloud interactions central to the climate model sensitivities that lead to large ranges in projections of future climate.

Finally, there are significant software and hardware infrastructure challenges pacing progress in climate science. Many scientists have found the growing requirements to support the software on high performance computers as a distraction from the central scientific goals of improving climate models and answering fundamental questions about climate feedbacks and variability. This view is offset by the new scientific opportunities provided by dramatic increases in computational power. The issue is scientific productivity. What is needed is a software framework that not only scales from desktop to petascale, but also that supports multi-scale model development and process integration. The same modules that are used in a global climate simulation should be used for regional and site-specific process studies across bench to field to global spatial scales. This vision for a seamless modeling environment has only been realized in a few areas, e.g. column radiation models. As a closer connection with observational data and process studies is required to advance the science of regional climate prediction, the software must also become more closely integrated and supported across scales. Software will increasingly be required to support data assimilation and other data intensive frameworks like DOE's CAPT activity. These software frameworks will emerge as key bottlenecks to progress. An investment now would have important payoffs in the not very distant future.

Some of the more important challenges include:

Scalability:
Climate models need to be able to exploit petaflop computer systems. This requirement places a severe demand on the scalability of very complex modeling systems. The challenge of many-core chips has implications for programming paradigms and the way model developers identify parallelism and exploit memory hierarchies. Input/output and check pointing are particularly problematic for models running on many thousands of processors.

Operating systems:
Climate models should be highly portable across a wide variety of system architectures and operating software. The solution to this issue includes stable operating systems that facilitate robust and generic abstractions of the interactions between the computer hardware and plug-compatible utilities layer within the models.

Optimizing software:
Single processor performance is highly dependent on local thread and data stream management optimized by the program language compiler. Combining and optimizing the extensibility and functionality of disparate software components at each stage of application development can make them more efficient and/or allow the use of fewer resources.

Maintenance:
Long term maintenance and validation procedures for application software, utilities and numerical libraries are needed to improve the reusability of code and accelerate the development process.

Implications for Investments

The current investment portfolio that established ASCR and BER collaboration under Scientific Discovery through Advanced Computing (SciDAC) should be continued. That investment has produced a successful series of increasingly realistic and sophisticated physical climate models. It has also funded the development of biogeochemical cycles research for Earth system modeling and an active and critical partnership with NSF to support the Community Climate System Model (CCSM) project. Further investments would accelerate progress by eliminating or reducing some of the rate limiters to progress in climate science and bolster the DOE's contribution and leadership in the field. The specific high impact areas we recommend are modeling, observations, computational algorithms, facilities, and management.

Models

The implication for investment in modeling will be an advance in the predictive capability of the extent and implications of climate change. The climate community needs to develop a new generation of ESM based upon new and expansive requirements:

- Ability to more accurately reproduce major modes of natural variability to enable predictive capabilities from intraseasonal to decadal time scales
- Functionality for decadal-scale ensemble forecasts at very high spatial resolution;
- Flexibility to incorporate new data on the physical, chemical, and ecological climate system in the form of process representation, thereby increasing the fidelity of climate simulations;
- Connectivity with user communities for adaptation and mitigation strategies; and
- Capability for two-way interactions among emissions, impacts, adaptation, and mitigation

The community will work to meet these requirements by leveraging ongoing investments in geophysical and computational science supported by the SC and other Federal agencies. However, the anticipated complexity of the models and applications are sufficiently demanding that new frameworks are needed for prototyping, testing, and evaluation. Based upon current methods, it has proved challenging to attribute systematic features in the simulations to the specific aspects of the dynamical, physical, or numerical formulation of the models. For this reason, new ESMs should be modular and hence easy to disaggregate and reassemble. Each functional module should be accompanied by test cases and the observational and/or simulation data required for rigorous and reproducible evaluation. The modules should be assembled in a flexible model superstructure that enables staged increases in process complexity. A strategic partnership between ASCR and BER could provide the new mathematical and computational frameworks required for this kind of robust and extensible model development.

In order to understand climate's effect on ecosystems and the feedbacks between land and ocean ecosystems, global climate models are being extended to ESMs with a full treatment of the carbon cycle. Simulation of biogeochemical cycles also requires detailed understanding of terrestrial and oceanic ecosystems; the exchange of organic and inorganic carbon compounds with other parts of the climate system; and the fluxes of energy, water, and chemical compounds (e.g., nutrients) that affect these ecosystems. The critical nutrient cycles for ocean and land ecosystems span time scales ranging from a few days (such as nitrogen) to over a thousand years. Modeling over this large range of time scales to fully evaluate the couplings between biogeochemical cycles, chemistry, and ecology will be a significant computational challenge. The spatial heterogeneity in the biosphere, below ground and above ground ecology is a fundamental issue overlying much of this science. Some other major open challenges are the sophistication of the ecological representations, the effects of high-frequency spatial and temporal variability on the carbon cycle (e.g., fronts and eddies); and the behavior of the biogeochemical cycles in coastal zones. A partnership between ASCR and BER could be instrumental to develop the measurements, models and computer resources required to understand the carbon cycle across the huge range of relevant space-time scales and process fidelity.

While the uptake of carbon by ocean and terrestrial ecosystems is a key element of the carbon cycle, dynamic changes in vegetation can also appreciably affect the exchange of heat, moisture, and radiation with the land surface. Since vegetation is dependent on regional precipitation patterns of the atmospheric circulation, it has become imperative to improve the hydrologic biases in the physical climate system in order to extend the models to include biogeochemical cycles. Tropical biases and cloud forcing have remained important areas of research for the DOE with the ARM program providing essential data and radiative parameterizations to the global modeling program. The analysis of linkages between vegetation and precipitation in a changing climate represents another opportunity to exploit new mathematical methods and algorithmic techniques to address the multi-scale coupling of these physical, dynamical, and biogeochemical processes. The committee suggests that ASCR and BER explore a hierarchy of simulation capabilities ranging from local process-oriented systems to global climate models to understand how vegetation will interact with a changing climate.

Observations

The expected outcome of investments in observations will be the improvement of our understanding of a variety of physical, chemical, biogeochemical, and ecological processes and how these processes and interactions are affected by climate change. Targeted investment can greatly aid in the improvement of climate modeling capabilities on decadal and century timescales, reducing key uncertainties in modeling assumptions.

At the core of BER's climate change program is the study of Earth’s carbon cycle, a research endeavor that began fifty years ago and continues today with strong measurement programs, processes studies and links between the CCSP and the Climate Change Technology Initiative. A unique opportunity exists for DOE and its partners to integrate observations to improve models. The strengths of the Terrestrial Carbon Program complement the Climate Change Prediction Program and SciDAC Program projects in this area, as well as ASP projects concerned with organic aerosols. The possibilities of assimilating flux tower data and employing process model studies to predict and design ecosystem manipulation experiments are largely untapped. Given the significant algorithmic and mathematical challenges in this field, the committee suggests that ASCR and BER commission the development of new methods and optimization techniques to couple observations and models of biological and physical systems.

Meteorological and oceanic analyses have become an important tool for studying the mean state and variability of the current physical climate. These analyses are constructed using a model that is adjusted by incorporating observations during its integration. These analyses have proved particularly useful for understanding the relationship between observations and the underlying dynamics of the climate system. It would be especially valuable to have a comparable analysis of biogeochemical and chemical cycles that could relate local and global biogeochemical processes to better describe the state of the global system. However, there are no extant analyses that encompass the physical, chemical, and biogeochemical processes in the climate system. Development of these analyses will require significant investment in assimilation systems for chemical and biogeochemical observations from in situ and satellite platforms. Much more advanced models will be required to understand the fidelity of the analysis system. This field appears to be ready for strategic investment in pioneering studies.

It's also important to point out that the ARM Program is the largest global change research program supported by the U.S. DOE. The program was initially created to help resolve scientific uncertainties related to global climate change, with a specific focus on the crucial role of clouds and their influence on radiative feedback processes in the atmosphere. ARM's primary goal remains the improvement of the treatment of cloud and radiation physics in global climate models in order to improve the climate simulation capabilities of these models. This program has developed a great deal of experience in fielding complex observational systems including in-situ and remote sensing systems, coordinating and facilitating scientific investigations of physical processes, and developing techniques for interpreting measurements on global scales. As such, the program can serve as a valuable example for the observational needs discussed earlier in this section.

Computational Algorithms

The implications for investment in computational algorithms will be increased scientific productivity using high end computers for climate change simulation studies and improved accuracy of climate models. There is a broad class of mathematical and numerical algorithms that need to be explored for application to the climate problem. We list several of the more obvious opportunities for enabling higher resolution simulations with shorter time to solution.

Spatial discretization techniques that provide the resolution required to address the underlying scientific questions is one class of algorithms that need to be explored in the context of advancing climate science capabilities. Climate modelers are beginning to introduce new vertical discretizations to better capture both boundary layer processes and isentropic/isopycnal flow outside the boundary layer. In particular, the use of quasi-Lagrangian coordinate schemes will permit better simulation of flow and minimize numerical diffusion. In the ocean, Arbitrary Lagrangian-Eulerian schemes are being introduced to maintain Lagrangian coordinates in the deep ocean while still resolving the surface mixed layer with fixed Eulerian levels. These new vertical methods require new techniques for determining and generating the optimal vertical grid based on physical properties of the simulation. In addition, methods for high-order conservative remapping of variables will be required as grids evolve in time. In a similar context, a fundamental issue in coupled climate models is the communication of energy, water and tracer fluxes between system components in a conservative and accurate manner. Current methods work reasonably well, but will not scale for more comprehensive configurations planned for these models. Robust grid remapping algorithms that work efficiently for high resolution and future dynamic grids will be required for ESMs, such as higher-order regridding with monotone limiters.

For the atmospheric component of the climate model, there are strong arguments for exploiting higher-resolution variable gridding configurations. The computational demands of uniform ultra-high discretization of a global atmospheric model would exceed the capacity of a petascale system. A more practical approach to dealing with resolution issues is to use a multi-resolution discretization, such as nested refinement, provided that the regions that require the finest resolution are a small fraction (10% or less) of the entire domain. In that case, the computational capability required could be reduced by an order of magnitude or more, and make the goal of computing with such ultra-high resolution models more feasible. There are a broad range of design issues that would need to be addressed for such models to be used routinely in atmospheric models, including choice of discretization methods, coupling between grids at different resolution, and dependence of sub-grid models on grid resolution, an important and unresolved issue for nested modeling techniques. The climate community has some experience in this area, but could easily take advantage of the extensive expertise and software already developed under ASCR support.

As the resolution of global climate models is increased, the allowable time step must decrease in current explicit forward time integration methods. Smaller-scale phenomena that are admitted by higher spatial resolution also have shorter time scales that need to be captured in the solution. However, processor performance will not be increasing rapidly enough to make up for the reduction in time step size and the resulting increase in number of explicit time steps required for multi-century integrations. Because the climate system exhibits very long timescales, and thus requires the ability to conduct long simulations, the community will need to begin to explore the use of alternative strategies for increasing time step size, such as fully implicit models, or other advanced solver techniques.

Another increasingly important algorithmic opportunity is data assimilation, which will become more important as climate science enters a more predictive paradigm. While assimilation has been extensively developed and used in the weather community, the climate community will need to evaluate which assimilation methodology is best suited for climate simulation and the creation of realistic initial states for climate change scenarios. Optimal interpolation and simple methods have so far been adequate for the ocean due to sparseness of data, but with the influx of new ocean data sets, advanced techniques like ensemble Kalman filters or 4-D variational assimilation will need to be examined.

A significant barrier to future model development using new algorithmic approaches is the highly non-linear, multi-scale, multi physics nature of climate system components, particularly the atmosphere. While there are many approaches to solving the basic fluid dynamical equations, most attempts to build full atmospheric general circulation models from these “dynamical cores” have come up short because of the difficulty of incorporating physical parameterizations and reproducing observed climate behavior. Currently, there exist a variety of test problems, ranging from the simple shallow-water equation test set to more sophisticated baroclinic test cases recently developed in the climate modeling community. The most rigorous test of a new system is a 20-year integration of the Atmospheric Model Intercomparison Project (AMIP) observational period. Application of these tests is now done on an ad hoc basis, and much could be achieved by standardization of a protocol, and the addition of additional tests to guide development. Algorithm researchers would then have clear expectations of what is needed beyond the solution of the equations of motion alone.

In order to provide useful information to policy makers, the climate modeling community will need to better characterize the uncertainty in simulation results. Ensembles and basic statistics are currently used to assess uncertainty due to natural internal variability intrinsic to the climate system. More formal methods for verification, validation and uncertainty quantification are needed from the computer science, mathematics and statistical science communities. A particular challenge is the sparse nature of observational data necessary to perform model validation.

Facilities and Infrastructure

The need for continued investment in facilities and infrastructure is to allow for the prediciton of climate change with greater levels of regional detail, to enable more comprehensive models of Earth's climate system (e.g., incorporation of the carbon cycle and chemistry), and to form an integrative, collaborative science of climate change consequences. This investment comes in a variety of forms, all of which need to be appropriately balanced with each other.

Computational Facilities

The major climate modeling centers have established a modeling pipeline in which there are present generation workhorse models that are scientifically proven through peer-reviewed publications, next generation workhorse models in the process of being scientifically proven, and models being used to explore parameter space beyond the next generation model. It is important that high-performance computing resources continue to be made available to provide adequate turnaround for all of these types of models, from production runs to debugging large, less-mature models. This is necessarily a mix of capability and capacity computing. The atmospheric sciences and climate research community has been well served by discipline specific computing centers like the National Center for Atmospheric Research (NCAR) that have managed the computing environment in order to maximize scientific productivity for the discipline specific applications and minimize the learning curve of young researchers entering the field. This is accomplished by a rich set of development and analysis tools, a well-balanced high performance computer system with adequate memory, online disk, and mass storage all functioning with high availability and stability. When systems are designed to support a discipline specific job mix, it is possible to get the most appropriate balance.

Existing computational capacity continues to be inadequate, both in real terms and via existing allocation mechanisms. For progress in the BER research areas, there is a demonstrated need for reliable and easy access to long runs with smaller processor counts. Currently, the access to order 1000 processor level machines is inadequate due to machine availability and allocation strategies that favor large processor counts at the expense of everyday development and debugging work. ASCR and BER should work together to provide for a more rational computational support of this discipline specific research area. Current demands continue to require enhancements to data management, migration and analysis mechanisms, which argues for attention to be paid to suitable storage hierarchy, bandwidth, support for workflow and analysis for climate science applications, which also provides for ways of dealing with both model and observationally generated data. Part of this involves making adjustments to optimally manage facilities for production, high-throughput debug, and analysis work. Priority needs to evolve toward providing stable environments which will enhance scientific productivity.

The future growth requirements of characteristic applications of BER climate change prediction, terrestrial carbon and atmospheric science programs more than doubles every year. Eddy resolving ocean circulation studies and cloud resolving atmospheric simulations are already pushing the petaflop requirement that will utilize tens of thousands of processors and as regional climate prediction on decadal to century scales becomes more important, the required computational power will enter the exaflop scale that will utilize 100K – 1M processors. This will require a continued focus on fielding state-of-the-art leadership class computing facilities so that computational capability does not become a more critical pacing factor.

Data Storage Facilities

Due to the special requirements of several areas in the BER research portfolio, special purpose, dedicated machines may be needed. Traditionally, BER has provided data centers serving climate data such as the ARM archive and the multi-ensemble analysis capabilities at PCMDI. With a growing emphasis on the integration of observation, model development and diagnosis and model validation, the specialized computing needs of BER centers may not be best served by the ASCR computing centers which are currently more focused on providing an environment for the computationally-intensive component of a broad class of scientific research needs. This is not meant as a criticism, but rather the acknowledgment that one size does not always fit all and BER will need to continue to provide support where appropriate. A hierarchical approach with the high end computing need provided by the DOE Leadership Computing Facilities federated with development and data centers is a facilities model that should be jointly explored and pursued by BER and ASCR. This provides a balance of discipline specific computing and data stewardship with the economies of scale of large scale, shared resources. The software layer supporting federation of the data grid takes several forms. As an example, the joint ASCR and BER funded Earth System Grid (ESG) provides a common directory and file retrieval service from a web based interface.

Networking and Data Management

To be useful, the data and software that underpin climate change research must be made freely available to global change researchers worldwide, in a manner that allows convenient access, analysis, evaluation, discussion, intercomparison, and application. This requires a plan for an infrastructure and collaborative environment that links centers, users, models, data, and resources on a global scale. The creation of such an infrastructure and environment is vital to the success of climate change research and critical for the impact sought by ASCR and BER. It demands continued investment in data management, software, networking and collaboration technologies.

The beginnings of such an effort have already begun with the ESG, which has as its mission the construction of a universal high-performance seamless access point for petascale data and computing resources. This effort involves distributed resources and data management, high-bandwidth wide-area networks, and remote computing using climate data analysis tools like PCMDI’s CDAT in a highly collaborative problem-solving environment. It is already enabling 1000s of climate researchers worldwide to access >200 terabytes of data products from CCSM and IPCC simulations. The work of ESG can be leveraged to meet some fraction of what will be required to resolve current bottlenecks.

Access to the terabytes of data produced by high-end climate models remains a bottleneck. However, there are a number of steps that can be taken to reduce this bottleneck. Efficient data management requires a standard output or community convention for processing and producing data output. Currently, a large part of the climate community modeling of physical oceanography, atmospheric sciences, and atmospheric chemistry has adopted the netCDF CF metadata conventions. Biogeochemistry and chemistry modelers are now working with the CF committee members to get their requirements into the CF standard. Distributed metadata catalogs must meet requirements for consistency and security of metadata and data information. Individual researchers must be able to search, browse, and discover metadata and data regardless of physical location.

In the not too distant future, large coupled runs will produce much larger data sets. With this increased complexity of data, we must rethink our storage and retrieval paradigm. It will be impractical for most researchers to download more than a small fraction of climate simulation datasets for local analysis (indeed, it is already impractical today for many). Thus, to allow any substantial use of these data, new approaches such as large-scale server-side analysis, replication to multiple national or regional centers, and caching of popular simulation and derived data must be supported. It may be desirable to include several of these processes (e.g., popular analyses, replication) in automated data generation pipelines. Overall, data creation, publication, and analysis processes must become distributed, more automated and closely integrated in terms of running models and directly archiving data for immediate use.

ESG and related programs already make heavy use of the Energy Science Network (ESNet) and other networks, for example to transfer data from supercomputers to archives and from archives to users. Climate research demands on networks will grow yet further as data volumes increase, as systems such as ESG make data more accessible, and as data publication and analysis procedures become more automated. While ESG is a start, it is far from being complete. A significantly enhanced and more integrated system is needed to bring together simulations and experimental data from a variety of sources and a variety of sensors to accelerate global change studies. By so doing, we can enable a growing community of climate and impacts researchers to leverage these studies to gain insight into Earth science process, trends, and interactions, with the goal of answering new scientific questions. Specific quantitative needs have been included in a recent draft ESNet requirements document, generated as a result of a joint ESNet-BER requirements meeting held over the summer of 2007.

Analysis Tools and Collaborative Technologies

As discussed earlier, climate science is necessarily distributed and collaborative. As interest in climate science continues to grow and its scope broadens to encompass issues of ecosystem and economic impacts, and the evaluation of mitigation and adaptation strategies, the number of participants will also increase. The overall productivity of researchers and the quality of the research output can likely be improved significantly by the use of advanced collaboration and workflow technologies.

The analysis of climate model output and comparison with observational data is supported by BER as part of the PCMDI. Analysis tools such as the CDAT software are continuing to evolve and expand to cover new analysis needs. These tools, along with the NCAR Command Language and NCO tools are being integrated with the ESG to allow remote analysis of data where it resides. This is a rich area of collaboration between ASCR and BER researchers with many implications for scientific productivity.

With a growing number of participants in the climate science enterprise, and a growing diversity of climate data, the need emerges to be able to document clearly the provenance (who, what, how) of derived data products. It will also be important to be able to manage who can consume what will sometimes be substantial amounts of computing, storage, and network resources. Trust management mechanisms must scale to far larger user communities than today.

Visualization Technologies

Modern visualization capabilities can play an important role in the discovery of new scientific results and in the communication of the science to a broader community of stakeholders. For an area like climate modeling this is particularly important because of the societal relevance of our results to policy makers and those concerned about the impacts of climate change. Dramatic, scientifically relevant images and animations are now being developed by the ASCR SciDAC visualization centers for BER scientists. As the tools develop to analyze and represent extraordinarily large data sets produced by climate models, visualization technologies will be needed that span the high end power walls and virtual environments down to laptop displays. Ease of use of these tools should be an objective so that working scientists can explore their analysis results without requiring expert visualization help. With powerful graphic accelerators available in most desktops, ways should be found to exploit these technologies to the benefit of the working scientist.

Management

In order to get maximum return on the investments discussed earlier, the activities of a vigorous climate science program must be well coordinated and managed. Here we cite a few opportunities for ensuring efficiencies in such a complex program.

Currently, the allocation of ASCR computer cycles is decided using a peer-reviewed proposal-driven process focused on breakthrough science. Such an allocation process not only requires scientists to undergo two peer reviews for their science, but can also place required program deliverables at the mercy of a second external review process. One of the missions of the BER climate modeling community is to provide input to policy makers on the impacts of energy portfolio choices. This programmatic need drives a large fraction of the computer cycles used by climate modelers and involves assessments with ensembles of relatively coarse resolution models. These assessment products have firm deadlines and a well-defined product and simulation schedule. Future demands may require rapid turnaround in response to queries from policymakers. These everyday production simulations do not often fit into the Innovative and Novel Computational Impact on Theory and Experiment paradigm of large-scale breakthrough science, yet are critical to programmatic deliverables.

The community has been able to obtain resources for programmatic work by bundling these deliverables together with a few large-scale science simulations through a large Computational Climate End Station proposal, under which many of the assessment cycles are managed internally. The End Station provides the development team with computer resources for control runs as well as high priority climate change simulations that directly target the DOE mission.

Since many of the results will be made available on the ESG, this approach has worked with a close collaboration of Oak Ridge National Laboratory (ORNL), PCMDI, Los Alamos National Laboratory (LANL) and NCAR. While this strategy has been successful to date, there is a possibility that peer-reviewers in the future might favor large flashy results and the more routine but important climate program deliverables will be placed at risk in a proposal-driven process. ASCR management should work with other SC program offices to ensure computing capability for required programmatic work is being adequately provided.

The DOE investment in software for climate modeling has been largely on the part of BER. The SciDAC2 program, which started in 2007, is the notable exception. A Scientific Application Partnership (Principal Investigator [PI]: Worley) is funded by ASCR dealing with the scalability of the CCSM as part of the SciDAC2 "Scalable and Extensible Earth System Model" project (PI: Drake and Jones). Since model formulation, building and testing require close coordination between climate scientists, mathematicians and computer scientists, the BER and ASCR partnership is natural and offers many opportunities for gains in scientific productivity. Software is the common currency for translating algorithmic and scientific hypotheses into computational experiments. Climate models, such as the CCSM3, are sophisticated software projects that support research by a large community of scientists as well as major assessment studies such as the 2007 IPCC AR4. The software engineering management must support scientific needs, production schedules as well as reliability and performance requirements. With a distributed team of developers and application scientists, the codes require full time coordination of gatekeepers and a scientific staff available to diagnose problems and provide solutions. The task of integration and coordination should be clearly designated and supported.

The development of new methods, especially new dynamical cores for ocean and atmosphere components, requires a concerted effort over several years by a small team. The steps for bringing new methods into consideration for production use are well delineated but difficult to traverse without becoming a climate domain expert. Mechanisms for mathematicians to be included in these joint ventures are needed. As we move forward the scalability challenge is forcing software architectures to expose higher degrees of parallelism. Utilizing hundreds of thousands of processors requires re-factoring the modeling system code and using different data structures in the middle layer. Tools that support software analysis of dependencies and aid in the identification of parallelism are sorely missing. With computer languages and compilers lagging years behind advanced hardware (for example the cell processor has no scientific programming language supporting its use) the model development community has opted for reliable and portable, easily optimized languages like Fortran90, Fortran95 and gnu C++. The burden of development on scalable systems with these languages is increasingly problematic and requires a larger investment in software engineering support personnel and their coordination.

Finally, the climate modeling enterprise in the DOE is increasingly driven by the need to obtain scientific results for a large and diverse group of users, including government officials, in a timely fashion. In such an environment, the development of innovative models, algorithms and software must be managed as a project, as opposed to an open-ended research program, in order to have the desired impact. Some aspects of such an approach are well-understood, such as the need for planning, schedule visibility, and milestones. A more difficult problem is the potential dependence of success on delivering high-risk products in models, algorithms, and software on the required schedule. Many of these products, such as new discretization methods, or new programming models, represent non-incremental departures from the current methods used in production climate models, but may be necessary to achieve the goals of the project. Risk management in such a setting requires careful planning and a close and continuing collaboration between the climate and math / computer science communities.

Summary

Our joint ASCAC-BERAC subcommittee has reviewed the state of climate change science with the goal of identifying computational and information technology obstacles pacing progress. There are many opportunities for strategic partnerships between ASCR and BER that would accelerate progress in climate change science, but they are not only related to computational and information technology barriers. Indeed, there is no obvious single pacing item but a collection of interrelated science and technology challenges. Many of the issues discussed in this report have been identified in earlier studies (see bibliography) and speak to the need for a balanced investment portfolio in computational infrastructure, climate science, computer science, and applied mathematics. In the short term, computational capability, albeit growing at a relatively healthy rate due to ASCR investments, remains a bottleneck and should remain a high priority investment. But as the science and complexity of climate simulation grows, so will new technical and scientific challenges. Immediate proactive investments in software, algorithms, data management, and other pacing items is strongly recommended so that the needed advances can keep pace with the evolving science and computational infrastructure. The management of these investments is also critical to success. We note that the development of innovative models, algorithms and software must be managed as a project, as opposed to an open-ended research program, in order to have the desired impact on accelerating progress.

Recognizing that computational and information technology solutions cannot be separated from the underlying science drivers, our specific recommendations include that ASCR and BER undertake joint ventures to:

- Continue to invest in leadership class computational facilities, data storage facilities, analysis environments, and collaborative tools and technologies. A significant fraction of these resources should be dedicated, configured and managed to support integrated and multi-faceted climate research and prediction across DOE and broader national and international efforts
- Invest in strategic collaborations to develop computational algorithms and scalable software to accelerate computational climate change science
- Develop computational and theoretical foundations for new modes of climate simulation, including ensemble short-range forecasts with regional fidelity and Earth system assimilation
- Focus the scientific effort to pursue robust predictive capability of lower-probability/higher-risk impacts, including climate extremes and abrupt climate change
- Develop a strong scientific understanding of leading-order uncertainties in the carbon cycle, in particular how the efficiency of natural carbon sinks will change with our changing climate. A more tightly coordinated effort in climate change science via well defined partnerships with ASCR and BER is highly desirable and will provide the best path forward for accelerating progress in this important scientific area

Bibliography

Improving the Effectiveness of US Climate Modeling, NRC Report, 2001, available at http://www.nap.edu

Capacity of Us Climate Modeling to Support Climate Change Assessment Activities, NRC Report, 1998

High-End Climate Science: Development of Modeling and Related Computing Capabilities, USGCRP report to OSTP, December 2000

Federal Plan for High End Computing: Report on the Federal High-End Computing Revitalization Task Force, OSTP, 2004 (especially Appendix D: Discussion on System Specifications and Applications Requirements)

NRC 2007: Evaluating Progress of the U.S. Climate Change Science Program: Methods and Preliminary Results, Committee on Strategic Advice on the U.S. Climate Change Science Program, National Research Council, 2007

Office of Science, USDOE, 2004: “Computing the Climate, Technology Barriers” in A SCIENCE-BASED CASE FOR LARGE-SCALE SIMULATION, Volume 2, pp71-73.(ScaLes report)

Eos summary of Aspen Global Change Institute Summer 2006 Workshop Report http://www.agu.org/pubs/crossref/2007/2007EO200002.shtml