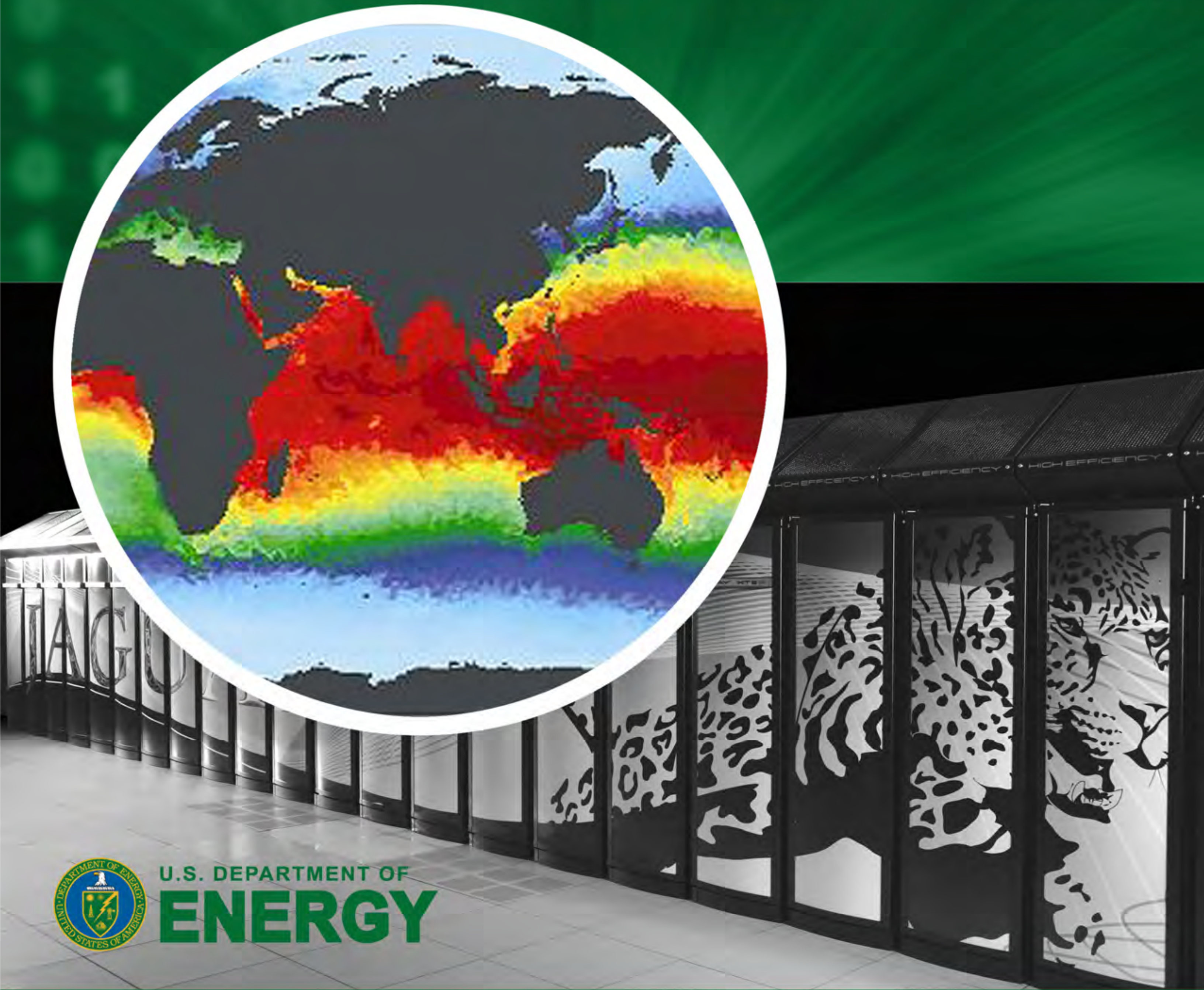
Scientific Grand Challenges

CHALLENGES IN CLIMATE CHANGE SCIENCE AND THE ROLE OF COMPUTING AT THE EXTREME SCALE

November 6-7, 2008 • Washington D.C.



U.S. DEPARTMENT OF
ENERGY

Sponsored by the Office of Biological and Environmental Research and the Office of Advanced Scientific Computing Research

DISCLAIMER

On the cover: Oak Ridge National Laboratory's Cray XT5^{TM} supercomputer. The computer, dubbed Jaguar, is the largest in the U.S. Department of Energy's Office of Science. The Jaguar's computing power makes it among the most powerful open scientific systems in the world. Future reports in the Scientific Grand Challenges workshop series will feature different Office of Science computers on their covers.

SCIENTIFIC GRAND CHALLENGES: CHALLENGES IN CLIMATE CHANGE SCIENCE AND THE ROLE OF COMPUTING AT THE EXTREME SCALE

Report from the Workshop Held November 6-7, 2008

Sponsored by the U.S. Department of Energy, Office of Biological and Environmental Research and the Office of Advanced Scientific Computing Research

Chair, **Warren Washington**
National Center for Atmospheric Research

Co-Lead, Model Development and Integrated Assessment **David Bader**
Lawrence Livermore National Laboratory

Co-Lead, Model Development and Integrated Assessment **Bill Collins**
Lawrence Berkeley National Laboratory

Co-Lead, Algorithms and Computational Environment **John Drake**
Oak Ridge National Laboratory

Co-Lead, Algorithms and Computational Environment **Mark Taylor**
Sandia National Laboratories

Lead, Decadal Predictability and Prediction **Ben Kirtman**
University of Miami, Rosentiel School of Marine and Atmospheric Science

Co-Lead, Data, Visualization, and Computing Productivity **Dean Williams**
Lawrence Livermore National Laboratory

Co-Lead, Data Visualization, and Computing Productivity **Don Middleton**
National Center for Atmospheric Research

Office of Biological and Environmental Research, **Anjuli Bamzai**

Office of Advanced Scientific Computing Research, **Lali Chatterjee**

EXECUTIVE SUMMARY

High-resolution climate system models and Earth system models that better simulate the interactions and feedbacks among the physical and biological component processes will soon require greater computing capabilities. These complex component processes include the following:

- atmosphere
- ocean
- land, soils, permafrost, and vegetation (specified and interactive) cover
- land ice
- sea ice
- carbon and other biogeochemical cycle
- clouds and related microphysics
- hydrology
- atmospheric chemistry
- aerosols
- ice sheets
- human systems.

Current models already provide improved simulation and prediction of changes in temperature and precipitation, and extreme weather events. In addition, Earth system models are becoming more effective at finer scales and for shorter periods. Regional-scale projections of climatic change will also require an increase in the spatial resolution of climate models and Earth system models and an acceleration of computational throughput with a combination of software and hardware advances.

The next generation of Earth system models will challenge current frameworks for computation, communication, the free exchange of observations and simulations, and analysis. For policymaker decisions at the regional level, high resolution in space and time will be required for assessment of changes to regional climate variability. An increase in spatial and temporal resolution and the need for accelerated throughput will require expanded high-end computing resources at the extreme scale.

As the science and complexity of climate simulation grows, so will new technical and scientific challenges. Immediate proactive investments in software, algorithms, and data management are strongly recommended so that the needed advances can keep pace with the evolving science and computational infrastructure.

This report is an account of the deliberations and conclusions of the workshop on Scientific Grand Challenges: Challenges in Climate Change Science and the Role of Computing at the Extreme Scale, held November 6-7, 2008. Representatives from the national and international climate change research community as well as representatives from the high-performance computing community participated in the workshop. This report reflects their opinions as representatives of the scientific community.

This report highlights the research needs and opportunities for climate change science over the next two decades, assesses the current state of science and technology, and recommends directions for research that should be pursued to meet the goals described. Finally, the report summarizes the results and conclusions of the workshop.

Technical panel discussions focused on four major areas where extreme scale computing is clearly relevant to climate change research: 1) Model Development and Integrated Assessment, 2) Algorithms and Computational Environment, 3) Decadal Predictability and Prediction, and 4) Data, Visualization, and Computing Productivity. A common goal of these four panel discussions was to define those research needs in climate change science that require extreme scale computing for their satisfaction. These four panels met, resulting in four panel reports and 12 priority research directions, which are summarized below.

PRIORITY RESEARCH DIRECTIONS

Model Development and Integrated Assessment

- How will the sea level, sea-ice coverage and ocean circulation change as the climate changes?
- How will the distribution and cycling of water, ice, and clouds change with global warming?
- How will extreme weather and climate change on the local and regional scales?
- How do the carbon, methane, and nitrogen cycles interact with climate change?

Algorithms and Computational Environment

- Develop scalable algorithms for non-hydrostatic atmospheric dynamics with quasi-uniform grids, implicit formulations, and adaptive multiscale and multiphysics coupling.
- Foster international consortia for parallel input/output, metadata, analysis and modeling tools for regional and decadal multimodel ensembles.
- Develop multicore and deep memory languages to support parallel software development.

Decadal Predictability and Prediction

- Identify sources and mechanisms for decadal predictability.
- Develop strategies for tapping into this predictability and ultimately benefiting society.

Data, Visualization, and Computing Productivity

- Develop new, robust techniques for dealing with the input/output, storage, processing, and wide-area transport demands of extreme scale data.
- Integrate diverse and complex data.
- Dedicate resources to the development of standards, conventions, and policies.

TABLE OF CONTENTS

INTRODUCTION

The purpose of this workshop was to examine the forefront scientific challenges in climate change science and how computing at extreme scales (i.e., peta, exa, and beyond) can contribute to meeting these challenges by the end of the next two decades.

GOALS

Key goals of this workshop were to engage the national and international scientific leaders, including participants of the 2008 World Climate Summit, in identifying the crucial global scientific challenges and to provide them with an opportunity to shape scientific computing at extreme scales. Other goals included:

- reviewing and identifying the critical scientific challenges
- prioritizing the challenges in terms of decadal or annual timelines
- identifying those challenges where computing at the extreme scales is critical for climate change science success within the next two decades
- engaging international scientific leaders in discussing opportunities to shape the nature of extreme scale scientific computing
- providing the high performance computing community with an opportunity to understand the potential future needs of the climate change research community.

WORKSHOP STRUCTURE AND REPORT PREPARATION

The U.S. Department of Energy (DOE) Office of Biological and Environmental Research (BER) in partnership with the Office of Advanced Scientific Computing Research (ASCR) held a workshop on the challenges in climate change science and the role of computing at the extreme scale, November 6-7, 2008, in Bethesda, Maryland. At the workshop, participants identified the scientific challenges facing the field of climate science and outlined the research directions of highest priority that should be pursued to meet these challenges. Representatives from the national and international climate change research community as well as representatives from the high-performance computing community attended the workshop. This group represented a broad mix of expertise. Of the 99 participants, 6 were from international institutions.

All attendees were provided with previous DOE and other documents as well the charge from Dr. Raymond Orbach, the DOE Undersecretary and head of the Office of Science. The workshop agenda and the names of the plenary speakers are presented in Appendix 1. The workshop participants are listed in Appendix 2.

Before the workshop, each of the four panels prepared a white paper, which provided the starting place for the workshop discussions. These four panels of workshop attendees devoted to their efforts the following themes:

- Model Development and Integrated Assessment
- Algorithms and Computational Environment
- Decadal Predictability and Prediction
- Data, Visualization, and Computing Productivity.

This workshop report is one of a series resulting from the Scientific Grand Challenges Workshops hosted by ASCR in partnership with other Office of Science programs. The workshop series focuses on the grand challenges of specific scientific domains and the role of extreme scale computing in addressing those challenges. Dr. Paul Messina, interim director of science at the Argonne Leadership Computing Facility, is overseeing the workshop series.

CROSSCUTTING FINDINGS

During the course of the workshop, nearly all of the breakout groups discussed the following crosscutting findings:

- Educate the next generation of climate scientists in extreme computing and train current scientists in the use of high-performance computing.
- Because computer architectures have become increasingly complex, it is important to have better interfacing tools to make them easier to use.
- Improve ability to predict changes in land cover, vegetation types, oceanic biology, and atmospheric and oceanic chemistry. We need to know how carbon, methane, and nitrogen cycles interact with climate change and how local and regional water, ice, and clouds change with global warming.
- Develop scalable algorithms that can use upcoming petascale and extreme scale architectures efficiently. New, robust techniques must be developed to enhance the input/output, storage, processing, visualization, and wide-area transport demands of extreme scale data sets.

The panels were charged with defining the gaps in current knowledge of climate science and identifying the major scientific and technological challenges to be overcome to facilitate advancing climate and Earth system modeling science. One problem that was discussed extensively was how to improve the computer performance of present-day and future climate models and Earth system models. Also discussed was how to process, archive, and visualize the data. In addition, a key question for the science community was how to predict the climate over the next few decades. At the end of the workshop, each panel presented priority research directions that should be pursued to address the identified opportunities and challenges.

The recommendations of the panels are summarized in the body of this report.

PANEL REPORTS

MODEL DEVELOPMENT AND INTEGRATED ASSESSMENT

ALGORITHMS AND COMPUTATIONAL ENVIRONMENT

DECADAL PREDICTABILITY AND PREDICTION

DATA, VISUALIZATION, AND COMPUTING PRODUCTIVITY

MODEL DEVELOPMENT AND INTEGRATED ASSESSMENT

Co-Leads: David Bader, Lawrence Livermore National Laboratory
Bill Collins, Lawrence Berkeley National Laboratory

INTRODUCTION

One of the most significant findings from the Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report is the near certainty that the Earth's climate will change dramatically over the next several decades (IPCC 2007). With this knowledge, several critical questions confront both scientists and policymakers.

1. How will the climate change over the next few decades to centuries in response to natural and human-induced climate forcing?
2. What are the potential impacts and consequences to ecosystems and human systems as climate change progresses in response to such forcing?
3. What are the viable options for adapting to and mitigating climate change in various regions? What is the efficacy of these methods? When must specific alternatives be deployed?

Although substantial uncertainty exists as to the degree and impacts of future climate change, especially at local and regional scales, it is generally agreed that significant adaptation will be required. Furthermore, the magnitude of climate change later in the century depends upon the near- and intermediate-term mitigation strategies used to reduce the emission of greenhouse gases. These strategies also must satisfy an increasing energy demand of a growing global population experiencing an improvement in its standard of living. Predicting these future climate changes and evaluating the effects of mitigation strategies require Earth system models (ESMs) that are far more accurate and comprehensive than those in use today. Integrated assessment models provide the framework for climate predictions by defining the emissions scenarios and elucidating the relationships among the natural and human systems that are at the core of climate change studies. In the next decade, integrated assessment and comprehensive ESMs will probably be combined into a single system that could be used to investigate scientific issues and to formulate policy options for adaptation and mitigation.

BASIC SCIENCE CHALLENGES, OPPORTUNITIES, AND RESEARCH NEEDS

The Computational Challenge

The predictions from integrated ESMs will be most credible if the important processes in the climate system, for example mixing by ocean eddies, are simulated at their native spatial and temporal scales. Critical organized features in the atmosphere and ocean including clouds and eddies have characteristic sizes of 1 to 10 km. Some of the major sources of uncertainty in climate predictions from existing models are associated with the aggregate effects of these phenomena. Experience with current climate models suggests that simulation of climate change with a model with 10-km grid resolution is inherently a petascale problem. In fact, even higher resolution is required to resolve these features with sufficient fidelity to the physical principles underlying their formation and evolution. Since the computational cost increases nonlinearly with higher resolution, it is likely that predictions of societal and environmental change at 1-km resolution would require truly extreme scale computers.

Goals and Metrics

Given these drivers for the continued development of climate, Earth system, and integrated assessment models, it is clear that extreme scale computers and ultrafast networks, data systems, and computational infrastructure will be required by 2020. A primary goal for deploying these resources is to enhance climate models as scientific, prediction, and decision-support tools. The metric for the impact of these resources on the field is to improve accuracy of adaptation and mitigation simulations through the quantification, attribution, and reduction of significant sources of error in ESMs.

Scientific Challenges and Potential Impact on Climate Assessment

Several major scientific questions remain unanswered, but they could be addressed with much greater fidelity to the real climate system using petascale to extreme scale computation.

1. **How will the sea level, sea-ice coverage, and ocean circulation change as the climate changes?**

   Coastal communities around the world could face an uncertain future as sea levels increase because of human-induced climate change. While the IPCC concluded that sea levels would rise by 30 cm by 2100, empirical estimates suggest that the rise could be 1 m or higher. The uncertainties stem primarily from the dynamic and thermodynamic evolution of the ice sheets on Greenland and Antarctica. The fresh water released by these ice sheets and other influences on the ocean mixed layer could affect the speed and even the stability of the thermohaline circulation. The principal scientific and computational challenges include determining the important processes that govern the melting of ice sheets, developing accurate treatments of mixing between the surface and the deep ocean, and understanding how mixing by ocean eddies and surface forcing combine to affect the stability and variability of the Meridional Overturning Circulation. Over the next 5 to 10 years, major advances in these areas would lead to more robust predictions of sea-level rise and its impacts on coastal communities. Resolution of these questions would also help quantify the risk of abrupt changes in ocean circulation.

2. **How will the distribution and cycling of water, ice, and clouds change with global warming?**

   The evolution of the hydrological cycle is a central issue for the sustainability of the natural environment and human society. At present, climate models consistently predict that the storm tracks in both hemispheres will shift northwards, leading to greater precipitation over the southern- and northernmost continents. However, predictions from multimodel ensembles diverge on the sign and magnitude of rainfall changes in the tropics and sub-tropics. This uncertainty applies to roughly half of the Earth's surface. The ensembles also diverge on the sign and magnitude of tropical cloud radiative effects from increasing temperatures. The main scientific and computational challenges include quantifying the critical cloud controls on the climate system and understanding the significance of small-scale cloud dynamics and cloud processes for climate change. Over the next 2 to 5 years, cloud-resolving models should be used to advance traditional climate models used for climate projection. On time scales of 5 to 10 years, ESMs configured for ultra-high resolutions could improve projections of regional water cycles, a critical element of integrated assessment. These models could quantify impacts of water resources on energy production and use.

3. **How will extreme weather and climate change on local and regional scales?**

   The IPCC has concluded that changes in extreme weather and climate events pose the greatest risk to human health and societal stability. These phenomena include winter windstorms, tropical cyclones, blizzards, mesoscale storms, heat waves, droughts, floods, and frost. The primary reason is that these phenomena can exceed thresholds for adaptation because of the rate and magnitude of climate change. Changes in the climate extremes could be particularly important for societies with limited access to the financial and institutional resources required to cope with the consequences. By their nature, climate extremes represent the “statistical tails” of climatological distribution of meteorological events. Climate models have been designed to predict the average climate over large areas and not the probability of locally extreme changes in climate. Robust predictions of the changing magnitudes and frequencies of extreme phenomena could require the development of truly multiscale ESMs. The principal scientific and computational challenges are in understanding what processes govern the statistics of weather events, in particular extreme weather; creating the computational capacity to generate the massive ensembles necessary to quantify rare extreme events; and developing the mathematical tools to analyze high frequency and high resolution data on climate time scales. The potential benefits from advances in these areas are better assessments of how changes in weather extremes will affect environmental security and societal stability. The research also could yield more reliable predictions of how the frequency and intensity of disruptive extreme events will evolve with global warming.

4. **How do the carbon, methane, and nitrogen cycles interact with climate change?**

   The emissions of well-mixed greenhouse gases from human activity are superimposed on massive natural exchanges of carbon dioxide and other carbonaceous compounds among the ocean, atmosphere, and land. The response and feedbacks from this global climate carbon cycle represent two of the most significant sources of uncertainty in projections of climate change for the 21st century. Current models give divergent projections for the evolution of terrestrial ecosystems and their interaction with the rest of the carbon cycle. The divergence stems from significant uncertainties in how best to model terrestrial vegetation and from the major discrepancies in the projections of future rainfall patterns. ESMs that more accurately and completely simulate physical, chemical, and biological processes and their interactions are required to explore changes on longer time scales and explore the potential for mitigation strategies. The main scientific and computational challenges involve thoroughly characterizing the natural carbon, methane, and nitrogen cycles; measuring how these cycles are being altered by climate change; and quantifying the current and future strength of the global carbon sink. On time scales of 5 to 10 years, advances in these fields could help quantify the impacts of land use change on weather and climate driven by biofuel and other energy production and use. This information would be critically important for the identification and development of sustainable sources alternative energy. On time scales of 10 to 15 years, transformations in carbon-cycle science could lead to rigorous assessments of geo-engineering measures including carbon sequestration in soils and other strategies.

Research Directions and Potential Scientific Impacts

The major challenges in climate change science are directly connected with the fidelity of integrated assessments of mitigation strategies. The connection stems from the need to forecast as accurately as possible the response of the Earth system to reductions in greenhouse-gas emissions. Research directions enabled by computing at extreme scale could yield potentially significant advances in the state of the science.

1. **How will the sea level, sea-ice coverage, and ocean circulation change as the climate changes?**

   Fully dynamic models of land ice should be integrated with ESMs to explicitly account for the contributions of melting glaciers and ice shelves needed to predict sea-level rise. Current estimates suggest that a quarter of sea-level rise results from the melting of land ice. In addition, land glaciers supply an appreciable amount of the potable water in central Asia and other mountainous regions, and projections for this water supply are a critical component for adaptation planning.

   Ocean general circulation and sea-ice models should be developed for operation at ultra-high spatial resolution to simulate deep-water formation, the analog to deep convection in the atmosphere. Accompanying these increases in spatial resolution, multiscale treatments of physics and mixing should be introduced in ocean and sea-ice models to correctly determine the large-scale meridional overturning circulation. Wave and storm surge models should be integrated with the ocean model to simulate the impacts of extreme events.

   These research initiatives would lead to more reliable predictions of future sea-level rise and would improve quantification of oceanic heat uptake and the exchange of carbon dioxide.

2. **How will the distribution and cycling of water, ice, and clouds change with global warming?**

   Over the next decade, the community will extend current models to operate at grid resolutions of 25 km to investigate and simulate jets, moisture transport, and mesoscale storm systems while it explores the next generation of modeling frameworks (Figure 1). The creation and initial tests of global cloud-resolving models with 1-km grid resolution should be accompanied by ongoing development of parameterizations for the inherently sub-grid processes involved in cloud and aerosol microphysics. These new process representations could be deployed both in experimental cloud-resolving models and in operational ESMs used for assessments and policy formation. The representation of aerosol and cloud microphysics should be treated as a single integrated system. Unification of atmospheric microphysics should produce much more accurate simulations of the principal effects of clouds on the Earth's radiative energy budget and hydrological cycle. These new parameterizations should be tested in deterministic fashion against field and satellite data using regional cloud-resolving models to simulate cloud systems at their native scales. The implications of these parameterizations for both weather and climate could then be examined in a self-consistent framework using global extensions of these cloud-resolving models.

   Global cloud-resolving models would transform research on the connections between weather and climate by explicitly resolving the space and time scales relevant to both meteorology and climatology. These models will enable rapid progress in a wide variety of climate science issues where clouds play an important or central role.

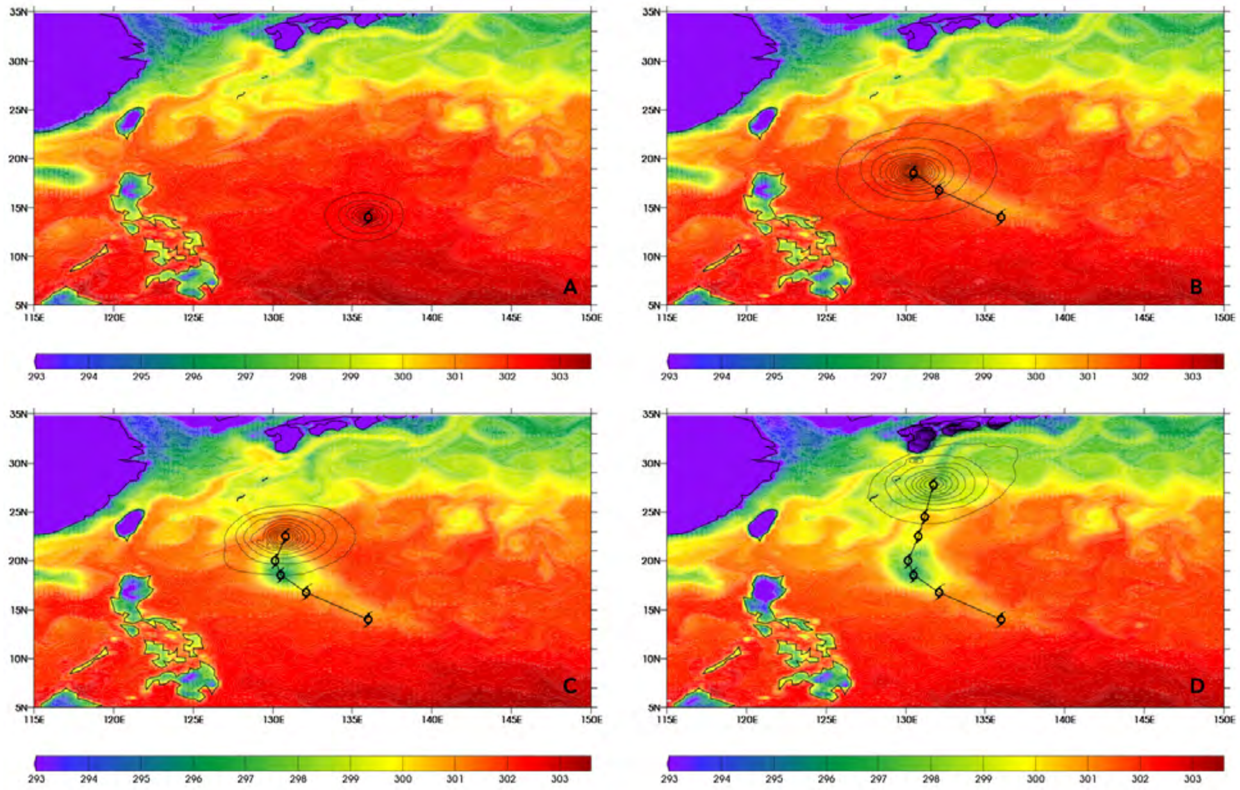


Figure 1. A simulation of a self-generated Category 4 tropical cyclone at Day 0 (A), Day 2 (B), Day 4 (C), and Day 6 (D) from the Ultra-high Resolution Community Climate System Model Simulation. This simulation was run on the Atlas supercomputer at Lawrence Livermore National Laboratory by a team of scientists from DOE laboratories and the National Center for Atmospheric Research. The mode uses 0.25-degree grid spacing for the atmosphere and 0.1-degree grid spacing for the ocean. The colors show sea-surface temperatures and the contour lines display surface pressure. At this resolution, the phenomenon of cold water upwelling produced by the storm's winds can be realistically simulated, and appears as a cold water "wake" behind the storm track.

3. How will extreme weather and climate change on the local and regional scales?

High priority should be given to the development of simulation tools that can identify which specific processes are critical to the prediction of various types of extreme events. Because of the highly localized character of many extreme phenomena, e.g., severe rainstorms and hurricanes, the climate community should explore the utility of hierarchical multiscale models that unify weather and climate prediction in a single simulation framework. Given the centrality of extremes in the hydrological cycle, the models should include detailed representations of modifications to the land and water systems by humans. Detailed projections of the impact of more extreme dry spells, droughts, and rainfall at the scale of individual watersheds will important for hydrological applications, including hydropower and water management.

Advances in the simulation of these phenomena would accelerate research in the processes that govern long-distance interactions between weather and climate. New models that exploit extreme scale computing could determine the future frequency, duration, intensity, and spatial distribution of droughts, deluges, heat waves, and tropical cyclones.

4. **How do the carbon, methane, and nitrogen cycles interact with climate change?**

   The development of ESMs that simulate the coupled interactions among physical, chemical, and biological processes is required to explore changes on longer time scales and explore the potential for mitigation strategies. These models should incorporate more advanced modeling of ocean biology and chemistry at the boundaries of the global oceans, including better representations of the biogeochemistry of coastal zones and the evolution of clathrates in the sea floor. Other important advances include better simulation of terrestrial biogeochemistry, particularly where there are large gradients in vegetation and land cover; better representation of biogeochemical cycling that captures ecosystem demography and its temporal evolution; and inclusion of process models of land surface disturbance including the effects of forest and grass fires.

   Enhanced models could transform understanding of the significant roles of human-driven and natural disturbances on the carbon and nitrogen cycles. These models could make explicit connections between the diversity and biological activity of complex ecosystems and the carbon cycle.

Coevolution of Resolution, Processes, and Integrated Assessment

The implications of these new requirements are three major model development directions that must proceed in parallel and, in fact, are dependent on each other. They are 1) increasing the spatial resolution of both atmosphere and ocean general circulation models (GCMs); 2) incorporating additional processes into the models for elements of the Earth system that are missing; and 3) linking the ESMs to integrated assessment frameworks that link natural and human systems.

When speaking of model resolution, it is necessary to distinguish true resolution from grid spacing. For spectral models based on spherical harmonics, the resolution is defined by the wave number truncation. For finite difference and volume approaches, the resolution is typically a factor of four lower than the grid resolution, as this is the smallest feature that can be simulated with any fidelity by these methods. Future generations of GCMs will have higher resolution, but their development will require significant advances in the grid structure and in the mathematical formulation and numerical implementation of the dynamics.

Climate models benefit greatly from atmospheric GCM development by the global Numerical Weather Prediction centers. In general, a state-of-the-science climate model is one to two generations behind current Numerical Weather Prediction models in terms of resolution because of the increased simulation times and the inclusion of processes that are important to climate, but they are inconsequential to influence 10-day predictions. The global weather prediction model used today by the European Centre for Medium Range Weather Forecasts produces 10-day forecasts twice per day at approximately 50-km horizontal resolution using spectral dynamics transformed from a spatial grid with an average spacing of 18 km. The European Centre for Medium Range Weather Forecasts employs their very high-resolution system to accurately simulate severe weather phenomena, such as localized intense rain or wind areas, that are often embedded in larger weather systems. The prediction skill of its model is improved because more of the energy transfer among the smaller scales is explicitly resolved, increasing the forecast accuracy by reducing errors.

In the longer term, research into cloud-resolving global models, coupled with "extreme scale" computing power is expected to lead to climate models that can explicitly resolve clouds and convection. The potential for cloud-resolving atmospheric GCMs is huge, but the scientific and technical challenges for their successful development and adoption are similarly large. While small-scale atmospheric phenomena can be simulated with much greater fidelity, simultaneously satisfying the global energy and mass constraints is an unsolved challenge to the production of accurate climate simulations.

Development of future higher-resolution versions must anticipate and integrate progress in numerical algorithms, programming models and computer architectures. Open and thorough testing of competing ideas for the next generations of high-resolution atmospheric GCMs is critical to successful development.

Improvement of existing process parameterizations and the incorporation of new ones are necessary for future models. Nevertheless, extensive observations are required to test the parameterizations, both as individual submodels and as part of the complex model system with many feedbacks. These phenomena are challenging to observe. To understand the changes globally, observations are necessary but not sufficient; model-data syntheses will be required.

Observations in Support of Climate Change Science

Extreme scale models of weather and climate will require a prodigious amount of data to evaluate the fidelity of their predictions across the full range of scales treated in the simulations. The climate modeling and integrated assessment communities should explore how best to collaborate with the Earth-observing communities, especially scientists involved in remote sensing and its fusion with in situ observations. The high spatial resolution and worldwide extent of the satellite data are essential qualities of remote sensing required to bridge the local, regional, and global scales of meteorological and climatological interactions. These data sets will be particularly important in analyses and simulations of land cover and land use change since human alteration of the landscape is frequently highly localized and specific to a given geographic region. The volume of the data sets, which currently exceeds 1 terabyte per day of multispectral imagery, poses a significant challenge to its application by the modeling community. Assimilation of land-surface data into Earth system and full-resolution integrated assessment model would be desirable for short-range predictions, but to date it has not been feasible with the modern voluminous satellite data sets. Concurrent investments in the computational and mathematical aspects of petascale data set analysis and assimilation are critical for the success of climate change science conducted using extreme computing.

CONCLUSIONS

In past national and international assessments, detailed ESMs for predicting the climate have been kept separate from integrated assessment models for simulating measures to slow or mitigate climate change. However, it is likely that the intersection and eventual integration of these two modeling approaches will accelerate significantly over the next decade. One of the primary drivers will be the simulation of the socioeconomic implications of second-generation biofuels and their interactions with the natural land surface and the carbon cycle. The differences between physical climate and assessment modeling, for example in local and regional resolution and in process representation, must be reduced or eliminated in the near future if we are to make progress on this and other critical issues in climate-change mitigation.

The panel group on model development and integrated assessment recommends detailed, near-term studies to identify joint projects to foster collaborations between the Earth system modeling and integrated assessment communities. From the perspective of policy formation, the primary objective of these collaborations should be the quantification of risks associated with alternate mitigation strategies, including strategies relying primarily on adaptation. Adopting a metric of reliable enumeration of risk rather than reduction in predictive uncertainty provides a better pathway for interfacing the climate and integrated assessment communities. It also provides a more direct link to decision makers who routinely must formulate policy in the face of appreciable uncertainty. With the advent of extreme computing, it may soon be possible to create integrated assessment models built on the foundation of the most comprehensive and computationally intensive ESMs. The development, analysis, and application of these models will pose an unprecedented set of challenges in both the physical climate and macroeconomic sciences. Overcoming these challenges will be an important step toward full engagement of the climate sciences in the exploration of policy for climate change mitigation.

ALGORITHMS AND COMPUTATIONAL ENVIRONMENT

Co-Leads: John Drake, Oak Ridge National Laboratory
Mark Taylor, Sandia National Laboratories

INTRODUCTION

Mathematical and numerical methods and their associated algorithms and computer software are at the intersection of science and extreme scale computing. These are the instruments that will make it possible to use the extreme scale resources required to further the science discovery process and support the proposed assessment activities. Resolving the dynamic scales of motion needed for decadal prediction and multiscale, cloud-resolving models will dictate the algorithmic and computational challenges. Programming models that effectively identify parallelism and tools that schedule memory flow and computation are needed to scale the models for the new architectures.

BASIC SCIENCE CHALLENGES, OPPORTUNITIES, AND RESEARCH NEEDS

To predict regional changes in water, ice, and clouds, a global cloud-resolving modeling capability is needed within 2 to 5 years. Models with this capability require unprecedented levels of parallel scalability in all components, including the input/output subsystem. One study of throughput rates (with a simulation time one thousand times wall clock) and computer capability requirements matches a 30-km global resolution with a 1 petaflop sustained computational speed. A 10-km grid resolution requires 5 petaflops, while a 1-km grid resolution requires 20 petaflops sustained and 100 terabytes of main memory. The science goals over the next 2 to 5 years will focus on characterizing the impacts of clouds on the Earth system, but within 5 to 10 years ensembles of century-long simulations using cloud-resolving models will be needed. This requires new programming models to allow the efficient use of extreme scale architectures and new approaches to overcome the algorithmic barrier associated with time integration techniques.

A second science goal involves modeling future sea-level rise and ocean circulation with the fidelity of eddy-resolving ocean models and new dynamic ice-sheet models for the massive land ice on Greenland and the Antarctic continent. The ever-increasing complexity of these models would introduce fundamental challenges in multiscale and multiphysics modeling. A third science goal calls for improving our ability to model extreme weather and abrupt climate events. This requires employing hierarchical multiscale models and large ensembles of long time simulations to fully characterize rare events. This goal reinforces the need for multiscale techniques and brings in new challenges in creating and analyzing the petabytes of simulation data. The same modeling, analysis and data challenges must be overcome to enable a true decadal climate prediction capability for adaptation to climate impacts and mitigation of future climate change.

New Approaches and Paradigms

The present petascale computers comprise tens of thousands of processing elements (e.g., the Oak Ridge National Laboratory's Jaguar system achieved 1 petaflop sustained performance with 149,504 cores) with a complex memory hierarchy. The machines being planned have factors of hundreds more processing elements requiring millions of parallel threads of execution. Computer chip manufacturers are challenging application programmers to identify finer grain parallelism and manage deeper memory hierarchies for efficient computation. The extreme scale platforms with millions of processors may arrive as early as 2015; therefore, new approaches and paradigms need rapid development to take advantage of the science opportunity this computational power represents.

New Algorithms and Mathematical Approaches

Models run at low resolutions for paleoclimate and multicentury climate simulations are approaching their scalability limit. This means that increased computing power, which now comes mostly through increased parallelism as opposed to single processor performance, will not allow simulations to run any faster than they do today. The only way to improve throughput is with new algorithms and mathematical approaches. Breakthroughs in this area would have an immediate impact on climate science because they directly benefit models already in extensive use by climate scientists.

Uniformly Distributed Grids

Next-generation atmospheric components that use more uniformly distributed grids (e.g., icosahedral and cubed-sphere based grids) or adaptive grid technologies would allow weak scalability on the new architectures. However, adapting to these less-structured grids introduces new challenges. These challenges include developing accurate, conservative numerical methods and associated dissipation mechanisms for unstructured grids and coupling them with the many subgrid scale parameterizations used in climate models. Because the community would require time to develop and gain experience with climate models in the million-processor regime, we recommend that the development of these components should start now.

Non-hydrostatic Model Requirements

Petascale computers would allow the first generation of cloud-resolving, global simulations. Initially, these simulations would be limited to forecast regimes and would affect climate models by providing a better understanding of convective processes and how these processes should be parameterized. Extreme scale resources would be required to perform century-long cloud-resolving simulations. It may be possible to provide a single multiscale model or code framework that spans all these scales and addresses the algorithmic and computational challenges. Non-hydrostatic dynamical cores are required to span these scales, bringing in the challenge of determining optimal equation formulations, numerical methods for unstructured grids with properties needed in hydrostatic models but also suitable for resolved convection. Significant synergy and strength may be gained with such an approach.

More Flexible Utility Structure

Practical simulations for decadal prediction would require true and effective parallel input/output that achieves rates of hundreds of gigabytes per second. Because of the many requirements on the input/output subsystems of these models, a more flexible utility structure is needed that is supported and available across centers and hardware/software vendors. This issue would only be aggravated by using unstructured grids or the selective output of extreme events. The large volume of output and the deep memory architecture that includes archive and analysis structures of the new computational environments might suggest non-hierarchical data distribution mechanisms. To democratize the input/output process would increase the use of existing network capacity and increase fault resilience. Workflow and analysis tools that are able to tap into the datastreams and process in parallel should be considered a part of an end-to-end science capability provided in the computational environment.

Prototyping the Programming Model

The programming model for extreme scale computing architectures has not been determined. Models are required to run efficiently on processors with many more cores than are available and with little increase in the processor's memory bandwidth. This places doubt in the current programming practice. The hybrid MPI and OpenMP paradigm would likely be adequate for petascale computing because the number of cores per socket is relatively small (Drake et al. 2008). Extreme scale computers must be programmed with techniques that expose finer grain parallelism and coordinate computation with the memory hierarchy. Because many options exist for programming models over the next 5 to 10 years and because code conversion would take nearly as long, we recommend that an experimental approach to prototype parts of the model be pursued. It is not necessary to explore radically different model (coupling) organization and abstraction because the current abstractions are effective and based in physical reasoning.

Increased International Collaborations

The European community is developing frameworks and modeling infrastructures to support the development of ESMs with the European Network for Earth System Modeling (http://www.enes.org). The European community shared the following goals with U.S. efforts:

- foster integration of the ESM research community
- foster development of ESMs
- foster high-end simulations
- foster application of ESM simulations for climate change impacts.

Currently, DOE is active in efforts such as the Global Organization for Earth System Science Portal and the Earth System Grid (ESG), which target a working system for distribution of climate science data and Fifth Coupled Model Intercomparison Project (CMIP5) results by early 2009 (see http://go-essp.gfdl.noaa.gov and http://www.earthsystemgrid.org). Joining the international collaborations with a more expansive agenda of parallel tools and modeling frameworks in support of climate change science would be fruitful (Simon et al. 2008).

Adapting to Embedded Processor Technology

Computer vendors are considering embedded processor technology. These chips enable the proliferation of portable consumer electronic devices such as cell phones, music players, and digital cameras. Power-efficient extreme scale computing systems using embedded processor technologies would be less general purpose than current systems. However, the cost efficiencies of such an approach could permit multiple systems to be built, with each system specialized to a class of problems of interest to the DOE Office of Science. If this can be applied successfully to exaflop scientific computing problems while eliminating the cost and electrical power barriers faced by conventional approaches, then the climate science community would face even greater programming challenges with a potential to disrupt the progress of climate science. Because there also may be potential benefits to our community, we must develop techniques that minimize the possible impact on science if this proves to be the technological path that vendors follow.

CONCLUSIONS

Future computing environments would offer new scientific insights for climate research only with increased effort on the part of software and algorithm research teams. This is a common problem across many application areas and within the international climate community that manifests itself in the problems of scaling climate models to millions of processors. A final concern of the panel group is the need to train scientists in the use of high-performance computers, a task that would be easier if the computing environment offered a richer set of tools and algorithmic options.

DECADAL PREDICTABILITY AND PREDICTION

Lead: Ben Kirtman, University of Miami, Rosenstiel School of Marine and Atmospheric Science

INTRODUCTION

Designing mitigation and adaptation strategies to respond to the consequences of global climate change are some of the nation’s greatest challenges. The decision support that is essential to meet this challenge requires reliable high-resolution regional climate predictions over the next 20 to 30 years. The use of the word “prediction” here has significant implications because the requirements are much more substantial than simply isolating the climate response to anthropogenic forcing 50 to 100 years from now.

For the next 10, 20, or 30 years, the natural variability of the climate system would be on the same order of magnitude as the response to external forcing (e.g., greenhouse gases, aerosols, and land cover). Therefore, practical planning of adaptation and mitigation options requires decadal forecasts of the natural variability of climate, the climate change commitment we have already made, and the response to additional greenhouse gases and aerosol forcing. Unfortunately, there are no robust estimates of what part of natural variability can or cannot be predicted during the next 30 years. Additionally, no consensus exists on how to delineate clearly between natural climate variability and externally forced climate response on these time scales.

Some of the collective and daunting climate-science challenges that must be addressed are as follows:

- What are the best modeling methodologies and computational requirements, based on a clear scientific consensus, for producing these forecasts?
- How do we best initialize the entire climate-system model?
- How do we do coupled ocean-land-atmosphere-cryosphere-biogeochemistry initialization so that little “shock” or climate drift occur in the forecasts?
- How would we generate quantitative estimates of forecast uncertainty?
- How would we include the impact of unpredictable components (e.g., volcanoes) of the climate system in the forecasts?
- Given the limited observational records, how would we provide quantitative assessments of forecast skill?

BASIC SCIENCE CHALLENGES, OPPORTUNITIES, AND RESEARCH NEEDS

Predictability, initialization, data assimilation, and modeling of the climate system present the underlying scientific and computational challenges. To understand 1) the global and regional decadal predictability and 2) prediction skill and where it originates, the following issues must be addressed:

The Limit of Predictability

We need to be able to make quantitative statements about the predictability of regional climatic variables that are of use to society. To do this, we need to define the climatic variables of societal use, identifying the potential sources of predictability for these variables and the modeling capabilities required to potentially predict these variables.

Translating Predictability into Actual Prediction Skill on both Regional and Global/Continental Scales

Quantifying the limit of predictability is an important part of developing a predictive capability. However, as a practical problem, potential predictability does not always translate into actual forecast skill. We need to investigate the bio-geophysical modeling and data requirements to translate some of the potential predictability into actual useful forecast skill. We need to understand how model errors limit our ability to tap into potential predictability.

Interactions between the Physical and Biogeochemical Components

Traditionally, the climate prediction community has focused on interactions among physical components; however, it is possible that some predictability would result from interactions between the physical and biogeochemical components of the climate system.

What is the predictability of changes in extreme weather events, regional ecosystems and regional hydrology, and can these changes be predicted with sufficient accuracy? It is important to emphasize that a forecast capability that is of societal use needs to go beyond traditional large-scale time averaged climatic variables. The need is to assess the potential predictability of the changes of the weather statistics within climate. Additional questions to be considered include the following:

- How do land-use changes (i.e., biofuels, crop production, energy production, forest management, etc.) affect decadal regional climate, and can these changes be predicted?
- Are there regional changes in forcing scenarios such as changes in short-lived aerosols that impact regional decadal predictability?

Much of this discussion emphasizes the ultimate goal of making climate predictions of societal use and how this goal demands that the climate research community extend beyond traditional boundaries. Despite this need, it is also clear that much of the societal relevant regional predictability would depend on the predictability of known phenomena such as the Pacific Interdecadal Oscillation, Interdecadal Pacific Oscillation, North Pacific Oscillation, Atlantic Multidecadal Oscillation, Tropical Atlantic Variability, and North Atlantic Oscillation. The climate prediction community recognizes that we do not have quantitative estimates of the limit of predictability, a full understanding of the sources of predictability or estimates of the prediction skill of these physical phenomena. In fact, the climate research community is still beginning to understand the potential sources of predictability associated with the interactions among the components of the climate system.

Challenges at the Science-Computation Boundary

The core science-computational decadal predictability and prediction challenge is twofold. On one hand, it is necessary to identify which components (i.e., physical, chemical, and biological) are needed, which component interactions and feedbacks need to be modeled, what resolution is required and how large of an ensemble of predictions is needed. In other words, how do we best balance resolution versus complexity versus ensemble size to give current and future computational resources (Figure 2)? On the other hand, what are the appropriate strategies and methodologies for initializing decadal forecasts? What datastreams are needed to best access potential predictability? It is clear that the assimilation of disparate

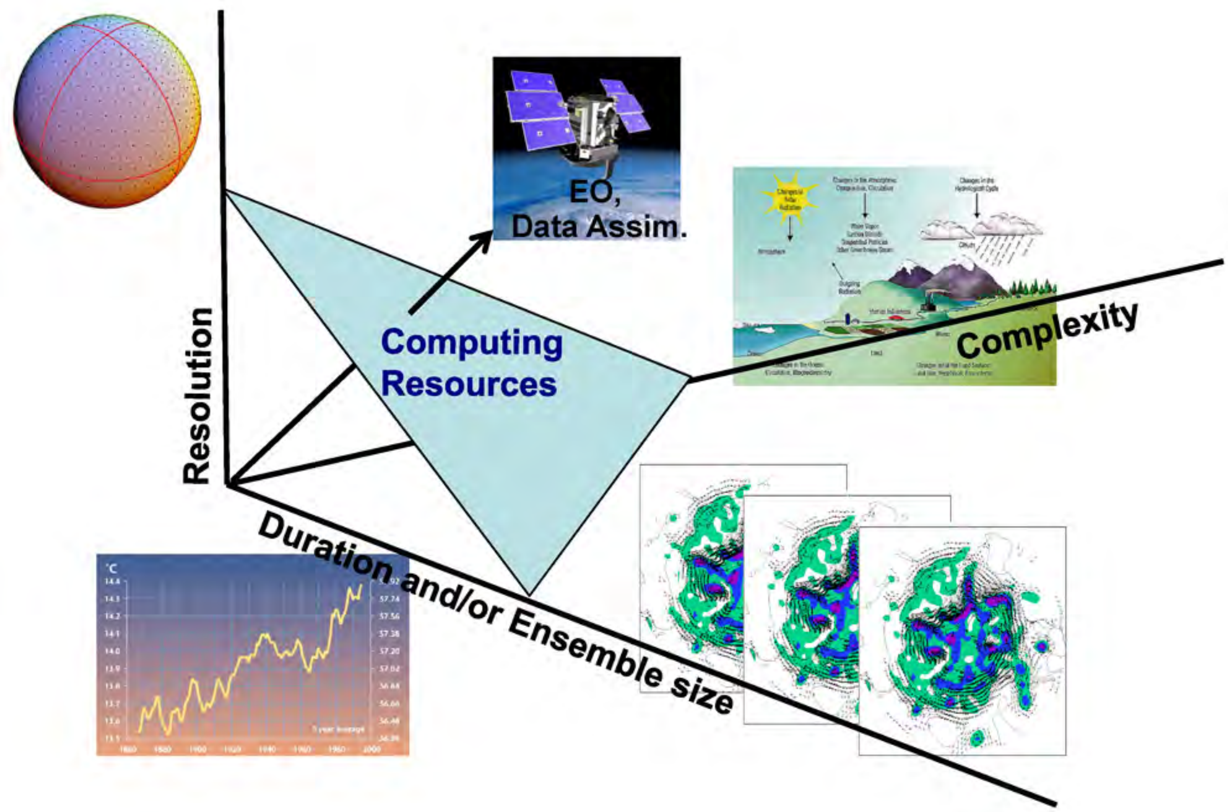


Figure 2. Resources tradeoffs. Image courtesy of Jim Kinter (Center for Ocean-Land-Atmosphere Studies).

observational datastreams into climate models would be required. It is, however, unclear how to readily translate this data assimilation into the best possible initial condition for the prediction problem. Indeed, multiple approaches are likely for initializing decadal predictions, and the research community needs to develop strategies and processes for documenting best practices in forecast initialization. Finally, it is critical to acknowledge the importance of a continued and vigorous effort in model improvement. This problem has implications in both defining the modeling requirements and in the data assimilation/forecast initialization issues.

Addressing These Challenges

The recommendations outlined here are separated into two time scales: near-term (2009 to 2015) and long-term (beyond 2015). We also discuss an important challenge that crosscuts both time horizons; namely, the training of the next generation of climate scientists that have a high level of competency in high-performance computing.

A Partnership for Decadal Predictability and Prediction: Using extreme computing to meet the scientific challenge of the next decade. In the near-term, we recommend that a multi-agency, multi-institutional partnership be established for sharing models, data, and predictions. Although this partnership should extend beyond the near-term, on this shorter time scale, it should focus on supporting the international CMIP5 Fifth Intergovernmental Panel on Climate Change Assessment Report (AR5) decadal prediction experiment. This support should not only include meeting the computational needs for completing the numerical experiment but also facilitating the collection and sharing (i.e., dissemination) of the model-based, retrospective forecast data and the forecasts from 2010 to 2030.[a] The partnership should contribute to the understanding the CMIP5/AR5 results by providing the necessary research tools (i.e., initial condition data, verification data, observational estimates of both raw and analyzed, data assimilation systems). This partnership also should share model codes, documentation, user guides, and computational facilities for numerical hypothesis testing.

It is critical that this partnership provide the computational infrastructure for decadal prediction and predictability hypothesis testing. For example, a thorough investigation of the fundamental uncertainties associated with initialized decadal simulations must be carried out before the results of such predictions may be interpreted. "Perfect model" experiments examining the spread in individual models when initialized with small differences in initial state are an essential first step in this process, placing a lower limit on the inherent uncertainty associated with predictions. These experiments also would allow for an investigation of the system sensitivity to errors in different components of the initial state, thus prioritizing future investment in observational systems. Experiments such as these require extensive computing resources, which are well beyond the reach of individual investigators.

A multimodel collaboration is recommended for validating and constructing decadal predictions using the current generation of coupled general circulation models. Because the instrumental observational record is comparatively short when measured on the decadal time scale, it would be necessary to perform a series of cross-model validation experiments to estimate the predictive skill, which remains when inter-model biases are included. Such experiments would initialize multiple models using approximately the same initial conditions and compare their evolution. Repeating future simulations using different models would improve confidence in those predictions by identifying regions in which the models provide consistent forecasts. In addition, combining forecasts for multiple model environments also is likely to reduce biases in future predictions, given the already demonstrated reduction in error observed in past simulations when a multimodel mean is used rather than any single model. Such experiments fit in well with the proposed CMIP5/AR5 decadal experimental design, but cross-institution cooperation must be encouraged in the design stages of inter-model validation experiments. Again, this represents a significant challenge regarding computational resources.

The panel recommends a study of the most effective use of future computing resources. This study should fully examine the most effective combination of model resolution and ensemble size needed to maximize prediction accuracy while properly evaluating the associated uncertainties. Test simulations must be conducted to determine the ensemble size required to properly sample the distribution of model outcomes possible from a given initialized state. In addition, simulations using sub-50-km resolution must be initiated to test the potential benefits of increasing resolution beyond that of those models proposed for CMIP5. This enhanced resolution would be critical to assessing regional decadal predictability and prediction skill, and it would be a necessary first step in developing global-coupled, process-resolving models (e.g., see beyond 2015 time horizon below). Much of this research would extremely tax existing computational resources.

---

[a] The data dissemination strategy requires careful consideration in how much of the data are provided through a centralized facility versus a distributed approach. Regardless of the final design, the end goal is to provide a single point-of-entry for data access.

Using perturbed physics ensembles in decadal predictions also could be considered by taking different, physically plausible values for model parameterizations to test for potential improvements in prediction skill. A problem with simply increasing the size of initial condition ensembles is that they may yield an unrealistically small spread in predictions from a single model, while other GCMs may show an equally small spread about a different mean. Perturbed physics ensembles could resolve some of this discrepancy by widening the artificially small error estimates by including the uncertainties in model parameterization values.

Verification and skill assessments of decadal predictions also present a daunting challenge. The observational record that can be used for retrospective verification of the decadal predictions is limited. The first set of forecasts would have 10 initial states (two per decade), and it would sample only about one cycle of major modes of what is thought to be natural climate variability. Additionally, the observational record includes trends resulting from climate change, which also must be predicted. To have a sufficiently large sample of independent forecasts to determine the predictability of the models with a reasonable level of confidence, it would be necessary to apply the decadal forecast systems to synthetic data generated by long model "control" simulations. These control simulations can include external forcing, such as greenhouse gas scenarios and injection of volcanic aerosols. These model world predictability studies would be limited computationally only, because as much verification data as desired can be generated. Identical twin-perturbed, initial-condition experiments would yield estimates of the intrinsic predictability of each model. The impact of model error on the decadal predictability could be addressed by attempting to predict the variability obtained from a simulation of one model using another model, by changing the physical parameterizations of the predicting model, or by other experiments. Similarly, the data assimilation systems and initialization strategies could be evaluated by applying the systems using one model to the data generated by another model, and multimodel ensemble strategies also could be tested (e.g., how does a multimodel ensemble predictability compare to "perfect model" predictability). It would be of particular interest to examine the impact of the differing climate sensitivities of the models on the predictability and the impact of differing sensitivities on the attribution of decadal predictability to external forcing or natural variability.

Regional decadal prediction: Pushing the envelope of extreme computing beyond 2015. After the initial period of evaluating the viability and impact of decadal prediction systems (around 2015), the following improvements are envisioned. One improvement would be adding components that can predict changes in land plants, oceanic biology, and atmospheric and oceanic chemistry. These components are being tested in atmosphere-only models and should be ready to incorporate into the atmosphere-ocean coupled general circulation model. Incorporating these components would allow evaluation of impacts of predicted climate changes on land plants, crop production, and oceanic biology. Another improvement foreseen as decadal prediction systems mature would be better estimates of past and future changes in short-lived atmospheric species (e.g., aerosols, tropospheric ozone, etc.) or in changes in land use (e.g., biofuels, crop production, forest management, etc.). Predictions that incorporate these complex regional forcing scenarios are critical for socioeconomic decision-making associated with, for example, energy production through hydropower generation or even wind-energy generation. A third potential improvement in decadal prediction systems by 2015 would be the incorporation of higher resolution, which would enable clouds to be simulated directly without the use of parameterization. This improvement would reduce the climate response uncertainty associated with cloud changes, would be critical for regional decadal prediction, and would be a key component in designing regional mitigation or energy production strategies that rely on complex regional forcing scenarios.

In the longer term, the decadal predictability and prediction initiative could contribute to the following outcomes:

- identifying observations that enable more accurate decadal predictions
- developing new observing-system components
- designing a broader climate-observing system that would be developed in the future.

For example, the sensitivity of decadal predictions to the climate system initial state provides information about the type, accuracy, frequency, and location of new observations that could improve decadal prediction accuracy and reduce prediction uncertainty. The models that would be used for decadal predictions would be confronted with observations, both through more advanced data assimilation methodologies and detailed verification of the predictions. This would lead to a better understanding of how to improve the models and would suggest changes in the observing system that could improve predictions in a manner analogous to numerical weather prediction practices today. The models also could be "observed" by applying a synthetic observing system to model simulations—so-called observing system experiments—to determine where and how to observe the actual climate system. This system would generate better initial conditions that would lead to improvements in the quality of the predictions. By carefully evaluating these results, new observational requirements for decadal prediction could be generated. These new observational requirements would help in the design of future climate-observing systems and the development of new instruments. It is envisioned that in the 2015 to 2025 period, these results could begin to translate into the development and deployment of new observing capabilities that would be added to either research or operational observing programs and missions.

Educating the next generation of climate sciences in extreme computing. One of the more difficult challenges facing the climate science community is how to ensure the next generation of scientists has the capacity to address the problems discussed in this report. Climate system models and computer architectures are becoming increasingly sophisticated and complex. Moreover, the mechanistic understanding, hypothesis testing, and model development that is vital to advancing decadal prediction approaches require scientists who have the capacity to exploit emerging computational advances and can delve into the details of the model formulation without trepidation. A great deal of anecdotal evidence shows that as a community we are not adequately meeting this challenge. Therefore, one of the key components of the partnership discussed above—one that crosscuts both time horizons—is a renewed emphasis on workforce development. This workforce development should enhance current government-sponsored postgraduate education and post-doctoral training activities, with an emphasis on the interface between high-performance computing and climate modeling and simulation. The partnership should enable focused educational collaborations among universities, government laboratories, and computational centers through extended summer workshops and internship programs. This effort should focus on "hands-on" training so the next generation of scientists would not only be capable of running a climate model but also designing and implementing sophisticated hypothesis-testing experiments.

CONCLUSIONS

Developing robust climate change adaptation and mitigation strategies requires climate predictions that have quantitative estimates of the prediction skill. This is a daunting challenge, given our current understanding of climate predictability and of the potential sources of predictability, our biogeophysical modeling capabilities, and existing sources of observational data. To meet this challenge, the climate research community would need to extend well beyond traditional scientific and computational boundaries. In quantitatively estimating the limit of predictability and making predictions, there needs to be multi-institutional and multi-agency partnerships that would facilitate, for example, determining the proper balance among model resolution, model complexity, and ensemble size given existing computational resources and the development of best practices in forecast initialization. Finally, there needs to be a concerted effort to educate the next generation of climate scientists that have the capacity to exploit emerging computational advances.

DATA, VISUALIZATION, AND COMPUTING PRODUCTIVITY

Co-Leads: Dean Williams, Lawrence Livermore National Laboratory
Don Middleton, National Center for Atmospheric Research

INTRODUCTION

Projections of potentially catastrophic changes to human habitat associated with anticipated climate change make it imperative to develop a “built-to-share” scientific discovery infrastructure that is funded nationally and internationally for the benefit of climate researchers, policymakers, and society. Currently, climate researchers have a difficult time locating and analyzing data for scientific studies, which leads to untenable delays and gaps in the evolving scientific understanding of our changing world. Existing data-sharing processes associated with data discovery, access, analysis, and visualization process are labor intensive and collaborative communication and governance mechanisms are inadequate. Today, human intervention is needed at every phase of data management, access, analysis, and visualization.

Building on the Earth System Grid

Making available extreme scale data, distance computing, and distributed applications from the network for participating national and international climate research institutions requires an infrastructure that is realized in embryonic form as the ESG. Moreover, to improve research ability and productivity and to solve complex scientific climate problems, this next-generation infrastructure requires advances in all sub-components of the ESG science environment needed to support true distance computing, databases, visualization, and other distributed applications. Projecting a glimpse of the future, the ESG Center for Enabling Technologies is striving to accomplish the following:

- make data more accessible to climate researchers by making access location transparent to heterogeneous data sets
- meet specific needs of national and international climate projects for distributed database, data access, and data movement
- provide secure web-based access portals to multimodel data collections
- provide a wide range of grid-enabled climate data analysis tools and diagnostic methods to international climate centers and U.S. government agencies.

Vision

Future technological advances envisioned over the next 5 to 20 years must address extreme scale data warehousing, scalability, and service-level requirements that far exceed what exists today. During this period of rapid expansion, users worldwide would have access to hundreds of exabytes of data, which would need to be stored at multiple disparate sites and connected to exaflop servers operating with high reliability at unprecedented Internet speeds. Scientists and policymakers could retrieve information and derive sophisticated data products within milliseconds. Transcending geographical and organizational boundaries, scientists would employ virtual collaboration analysis tools to instantaneously manipulate data that are viewable by all users. Unlike the current static analysis tools, these tools would support the co-existence of many users in a productive shared virtual environment. This advanced technological world, driven by extreme scale computing and the data it generates, would increase scientists' productivity, exploit their national and international relationships, and push their research to new levels of understanding (Figure 3). Thus, although the targeted primary users are domain experts, it is essential that non-experts (e.g., politicians, decision makers, health officials, etc.) also are able to access much of the data. Rich but simple interfaces will allow the non-experts to accomplish difficult data manipulation and visualization tasks without having to understand the complexities of application programs or the computing environments on which they depend.

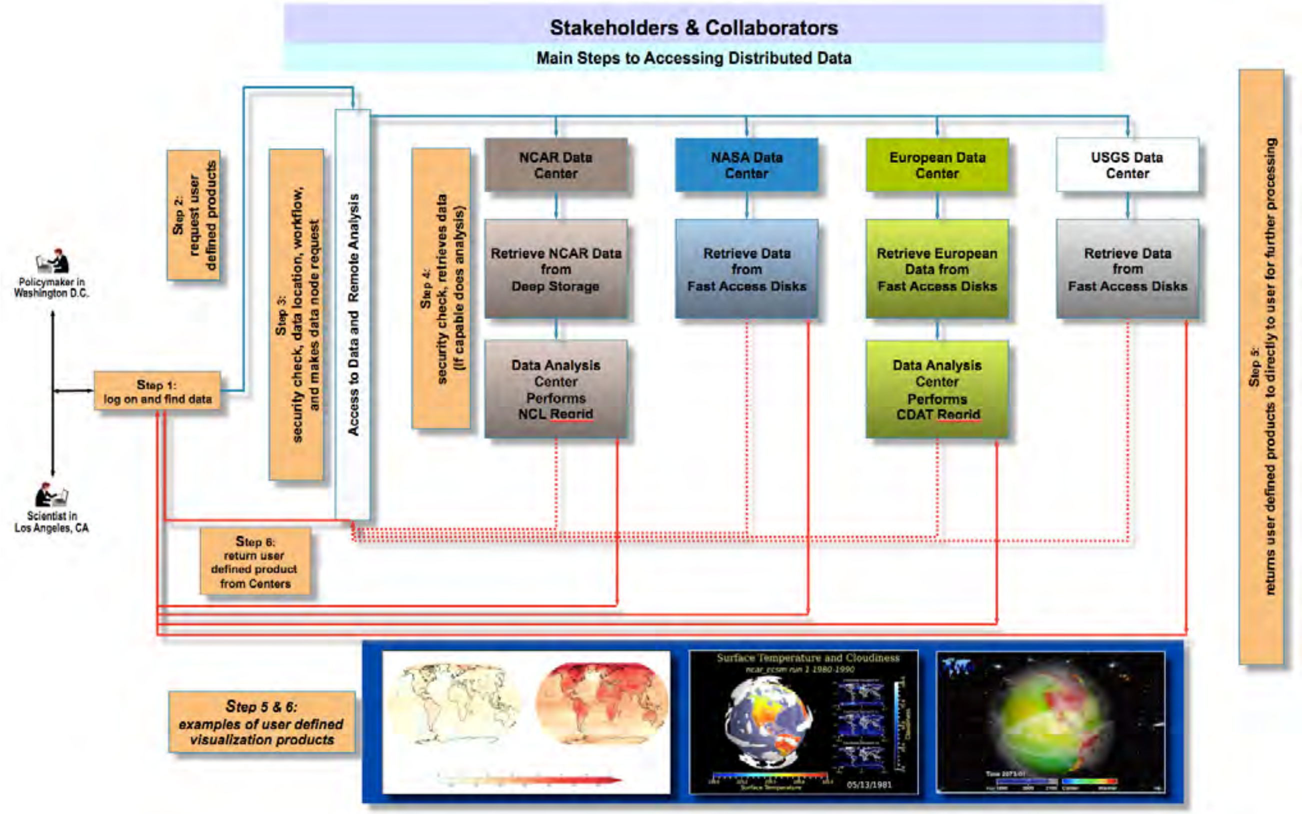


Figure 3. Schematic depiction of a use-case scenario supporting remote login for experts (e.g., model developers and climate researchers) and non-experts needing fault-tolerant end-to-end system integration and large data movement. For the expert user, the system expresses all the capabilities and complexities needed for rich data exploration and manipulation. For the non-expert, however, a simple abstraction layer includes easy-to-use controls for maneuvering within the system. Image courtesy of Dean Williams (Lawrence Livermore National Laboratory).

BASIC SCIENCE CHALLENGES, OPPORTUNITIES, AND RESEARCH NEEDS

Wide Area Network

Climate model data sets are growing faster than the data set size for any other field of science (Figure 4). Based on current growth rates, these data sets would be hundreds of exabytes by 2020. To provide the international climate community with convenient access to these data and to maximize scientific productivity, these data would need to be replicated and cached at multiple locations around the globe. Unfortunately, establishing and managing a distributed data system presents several significant challenges not only to system architectures and application development but also to the existing wide area and

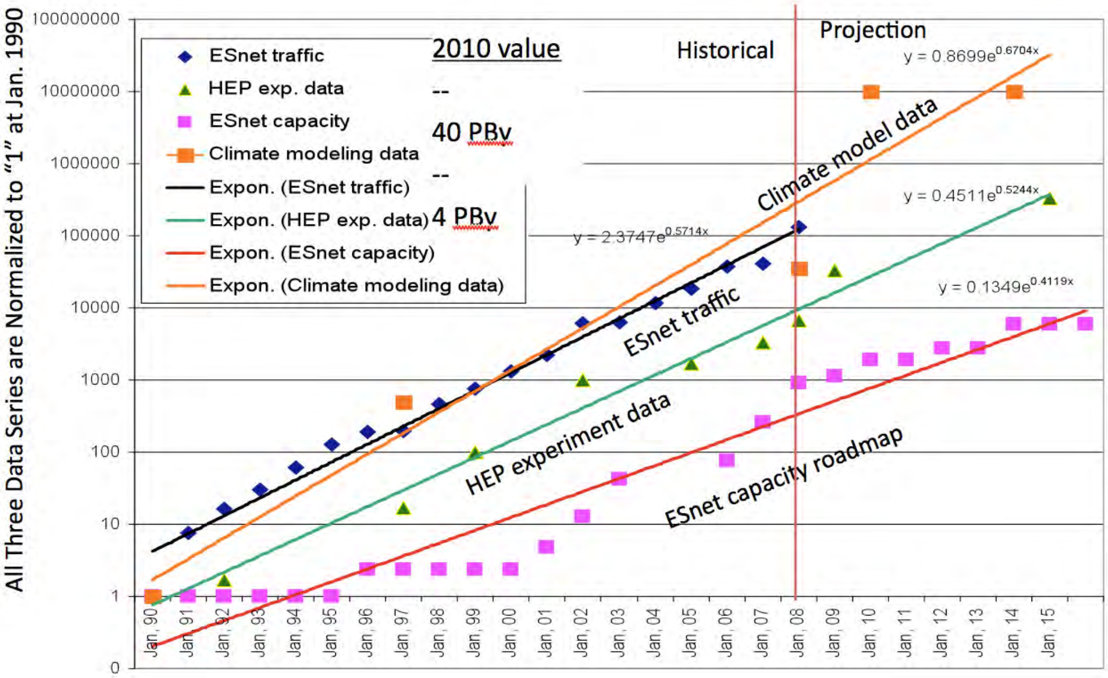


Figure 4. Long-term trends in future network traffic and capacity. Image courtesy of Brian Tierney and Steve Cotter (Lawrence Berkeley National Laboratory)

campus networking infrastructures. For example, transport technologies currently deployed in wide area networks do not cost-effectively scale to meet the scientific community's projected aggregate capacity requirements based on the growth rates for data set size. Even if backbone network technology improvements increase link speeds from the current 10 Gigabits per second to 100 Gigabits per second and are in production service by 2012, as anticipated, more efficient use of networking resources would be essential. Efforts are under way to develop hybrid networks with dynamic virtual circuit capabilities, such as those networks currently being tested and deployed by research and development networks like ESnet. Although dynamic virtual circuits allow high-capacity links between storage and computer facilities to be created as needed and then deactivated quickly to free up network capacity for other users, much work still is required to optimize and harden the software.

Easy and efficient data transport across wide area networks is not the only networking challenge to be faced over the next 5 to 10 years. In the use case mentioned above, the policymaker easily downloaded 20 terabytes of data to their site. To achieve easy downloads and better performance for the ordinary user, training is key. Often too few staff members or not enough outreach are available for this effort, which may leave the user of the system frustrated.

To maximize throughput between distributed systems, sophisticated network monitoring tools are required to enable real-time monitoring of the entire end-to-end network link—including campus networks—to avoid congestion and provide the necessary performance data to enable fine tuning of the protocol stack and assist with troubleshooting. This network monitoring software would need to be tightly integrated, preferably facilitated via the development of applicable standards, with the software systems that control data movement and dynamic circuit control to achieve our networking goals. Data movement infrastructure—both hardware and software—would have to be maintained and upgraded to enable the continued efficient use of the network. Long-term investments in middleware and test/measurement tools are critical. The implication is that tools would become more essential over time, and thus would need to be officially supported with long-term funding.

Data Management

Managing collections of extreme scale scientific databases, consisting of diverse Earth system data sources, present several significant technological challenges. As mentioned previously, data archives would be distributed over a wide geographic area in federated databases. While such data archives would be diverse depending on the country or agencies involved, they should have uniform data access layers that permit them to operate transparently. Furthermore, adding new archives and data to the infrastructure should be a simple dynamic process that does not disrupt ongoing operations. Automating monitoring of storage usage and the effectiveness of data access would be essential to identify and correct bottlenecks. In moving toward increased collaboration among groups of researchers, the technical software infrastructure must support shared access to distributed data archives and applications, and must provide flexible software systems designed for individual or group needs. Fundamental to this infrastructure is the capacity and collaborative mechanisms to specify, standardize, and share metadata, in a manner that balances autonomy and flexibility. For example, the use of distributed ontologies, that support real-time merging to discover needed data sets and their locations at a sufficiently granular level to efficiently take action (e.g., move, compute, analyze, or visualize). Online registries would be in place to enable researchers to access the contents of remote databases, formulate queries, transfer data, and combine disparate data management systems such as health, economics, etc. It then would become feasible to issue data requests that access many different data archives simultaneously, while rendering the underlying systems complexity transparent to the user. This model also can be extended to the sharing of applications, libraries, and software modules as follows:

- support services for collaborative visualization, where users can make scientific understanding of the data via visualization tools in a collaborative environment
- support for distributed data discovery through ontologically based metadata (i.e., Resource Description Framework and Web Ontology Language), including geo-spatial metadata
- support for computation, analysis, and management of distributed data
- support for application/module sharing, including virtualization strategies (e.g., diagnostic libraries).

Storage Management and Data Movement

The data volume expected from future simulations and observational devices is so large that it presents special challenges in the areas of storage management and large-scale data movement. While we foresee that storage technology would be able to store petabytes of data on disks, the mirroring of such data presents the following requirements:

- Moving 1 petabyte of data, even at the rate of 1 gigabyte/s (10 gigabit/s), takes 10^6 seconds, or 277 hours (about 12 days). The entire end-to-end system must support that transfer rate and must be robust and able to recover from transient failures of the storage systems and networks involved.

- All storage systems should interoperate seamlessly. Since storage systems are diverse (based on various standards and storage components), that goal requires a standard interface to allow seamless access of data from online storage (disks or solid-state drives) and offline storage (robotic tape systems). Such standards are starting to emerge in the High Energy Physics communities (called Storage Resource Management) and could help mitigate this problem.
- When data are replicated, mechanisms should be in place to manage the lifetime of the replicated data, so over time the storage systems can reclaim the space the replicas occupy, while still preserving the "master" copies.
- It is necessary to keep rich metadata of the data holdings content based on agreed upon terminology to search the data (referred to as "ontologies" or "controlled vocabularies"). It is also necessary to keep track of the locations of replicated data, and propagate version changes to the replicas. Furthermore, algorithms should be developed to allow users to get data from the sites that maximize transfer efficiency based on network traffic, storage system transfer capacity, and storage system load.

Data Analysis Center

Data replication allows data to be closer to users to reduce the latency of data movement. However, it is unreasonable to require that each user would download large volumes of data to perform the analysis, as is the current practice. Given the large volume of data, it may be more practical to perform the analysis tasks close to where the data are stored. This can be achieved by having "analysis centers" that are placed close to the data (Figure 3). The analysis center facilities (usually small computer clusters) should be designed to permit users to invoke their preferred analysis components and introduce their own analysis codes into such facilities. We believe that the model of "analysis centers" would simplify the tasks of climate analysts and avoid moving and replicating data unnecessarily. The analysis centers also should provide software components that compare model output in models have different grids, different resolution, and compare/combine data coming from different sites. Finally, the analysis centers should be designed to take advantage of multiple processors and multiple cores to run parallel analysis whenever possible. This requires an effort to parallelize the analysis tools used most by the climate community.

Data Environment and Workflow Management

Data analysis usually involves multiple steps. An analysis center would benefit from using workflow management systems to orchestrate these steps. Furthermore, the workflow process, and the details on the parameters used, what input data were used, and what was produced needs to be captured. Such information is referred to as "provenance." A similar need exists when running simulations, where a workflow and all the conditions for running the simulations need to be captured, so that simulations can be reproducible. Although large-scale simulations do not need to be rerun, the provenance data ensure that all information about the runs are captured, preserved, and easily retrieved.

The workflow management system should allow some tasks to be executed on-demand. For example, in the above use case the user ran several climate models in real-time that generated multimodel ensembles for a desired region and period. It should be possible to have facilities that allow such on-demand runs. Here again, establishing new workflows or re-using existing ones should be possible.

Another capability that we believe would be most valuable to the climate community is the ability to support tagging data by community users. Such annotations have proved extremely useful as a mechanism to inform all users of observation of interesting patterns, anomalies, etc. Related to such services, is to allow users to see changes made to the data holdings and the history of such versions.

Analysis

Intelligence analysis, emergency response, disaster prevention, and border security concerns would play an important role in the shaping of advanced analysis and visualization tools. A massive increase in climate simulation capacity would demand major advances in our analysis and visualization tools to elevate scientists' productivity. Many of the existing tool-development efforts (e.g., Climate Data Analysis Tools, National Center for Atmospheric Research Command Language, Grid Analysis and Display System, Ferret, etc.) are grossly understaffed relative to future challenges; however, the potential exists for an effective interagency collaboration to address this problem. The current tools are inadequate for our high-resolution runs and will become even more problematic in the future. These tools need to advance from single-threaded desktop applications to more powerful ensemble analysis capabilities that can leverage the same systems that produced the data. Analysis capabilities must scale along with the simulation capabilities.

The scale of data that needs to be analyzed would require that many analysis functions be performed near the data. For example, extracting desired data subsets would require applying indexing technology at the data source. Similarly, summarization operations (such as monthly means) would have to be performed at the data source. In general, multistep coordination of the analysis pipeline would have to be managed using automated workflow capabilities that can control distributed processing.

Climate analysis is based on climate models and their verification with observed data sets. One of the most difficult challenges is comparing large-scale simulations with observed data because of the differences in scale and layout of the data. Observed data can originate from various devices, such as monitoring stations and satellite images. The analysis tools that integrate and compare such data with simulated data would present unprecedented challenges as the scale and diversity of the data grows over time.

In the next few years, non-experts will increasingly be analyzing climate data. This would make the strong case for using other data formats such as Geographic Information System, Google Earth Keyhole Markup Language, and conversion tools. This effort would require richer metadata standards to provide information that is more comprehensive.

Visualization

Scientific visualization, which is the science and art of transforming abstract data into readily comprehensible images, is an integral and indispensible part of the modern scientific process (Figure 5). It leverages the high-bandwidth human visual system along with humans' ability to assimilate visual information into comprehension and understanding.

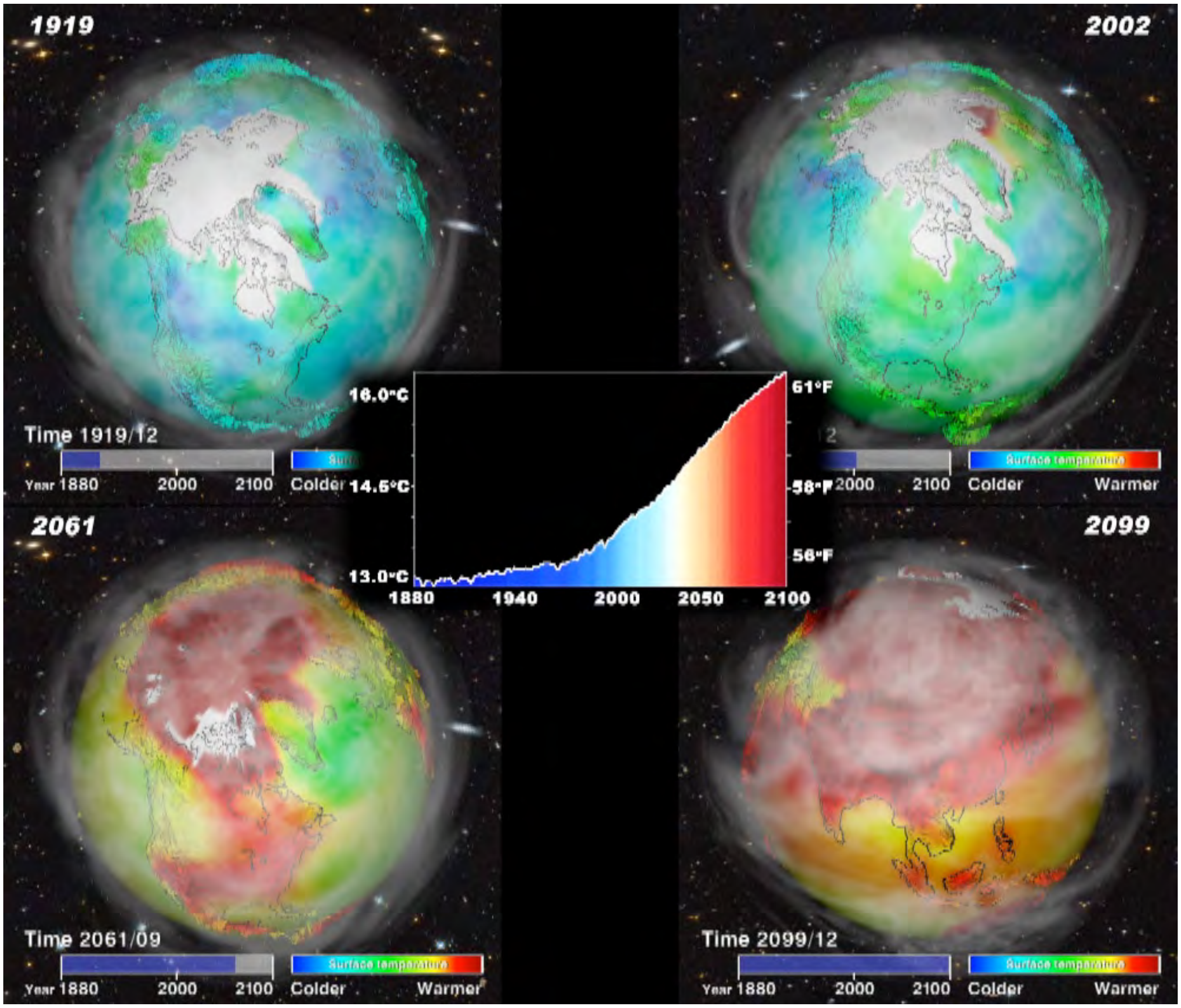


Figure 5. The image depicts four three-dimensional timestamps, showing surface temperature, clouds, and sea ice of the Coupled Model Intercomparison Project Phase 3 (CMIP3) Multimodel ensemble of 20th century run (1880-1999) and the 21st century (2000-2100) experiment. In aggregation, the 23-model ensemble combines for 220 years of science observation and computer simulated global warming. The center image shows progressive warming of the planetary surface as greenhouse gases concentrations increase over time. Image courtesy of Dean Williams (Lawrence Livermore National Laboratory).

Visualization research and application efforts tend to fall into one of three primary use modalities. The first is “exploration visualization” where the user has no idea what they are looking for. This type of use model is typically the most challenging since its success relies on interactive, “random-access” exploration of large and complex data sets. Another use model is “analytical visualization” where the user knows what they are looking for in data. Analytical visualization techniques often are those reduced into optimal practice after being established during exploratory visualization. Finally, “presentation visualization” is where the user wishes to convey a specific concept to others.

It is clear that the coming evolution in climate science would result in an increased ability to generate, collect, and store larger and more complex data sets. The size and complexity of these data sets would result in several challenges. First, these data would be of unprecedented spatial and temporal resolution with spatial resolution far exceeding screen resolutions. Second, many of the new simulation data sets would likely exhibit advanced spatial mappings like adaptively refined meshes (space and time) and icosahedral/geodesic grids. Third, future visual data analysis software infrastructure would be deployed in diverse ways to support several use modalities ranging from standalone-desktop applications to highly parallel tools that are integrated as part of easy-to-use, web-interface brokered visual data analysis infrastructure.

The climate science and visualization communities have done an admirable job over the past few decades in producing visual data analysis software infrastructure. However, we are at a crucial juncture marked by the growth of data size and complexity, the proliferation of parallel computational platforms, and the need for a large, global community of climate scientists to share data and software to solve some of our generation's most challenging and urgent problems. Some of the current challenges facing visual data analysis are as follows:

- Existing tools and algorithms are not suitable for use on larger data sets. Tools from the 1970s and 1980s that have been the “workhorses” of the past few decades simply will not run on many current and virtually all future large, complex data sets.
- Most existing tools/algorithms, in serial form, lack the capacity or responsiveness to meet the future visual data understanding needs of the climate community. For example, most existing tools and algorithms are serial. In contrast, it is generally accepted that the path towards the target capability would require effective use of parallel computational platforms and parallel data input/output infrastructure.
- The explosion of data size and complexity would give rise to a new set of needs in visual data understanding. Existing techniques, which have proven effective for single-variable, coarse data sets, would not be adequate for future science needs. For example, existing visual data analysis algorithms are not effective for studying the relationships between the dozens or hundreds of runs produced in a single ensemble or for comparative analysis of dozens or hundreds of ensemble runs. Other examples include the need to discover and understand the relationships between variables in time-evolving data, comparing simulation and experiment/observed data, and comparing/understanding data that exist on different grid types and have different resolutions.
- Software engineering to enable deployment of new capabilities in support of diverse use modes. The target here is to have visual data analysis software infrastructure that can be deployed as a standalone application on a desktop or as part of workflow that runs on large parallel computational platforms.
- Computational platforms and infrastructure for visual data exploration/analysis post-processing need to be designed to accommodate data intensive computing needs, which include vast amounts of input/output bandwidth and memory. These systems need to be available for on-demand, real-time visual data exploration to support the “exploratory” form of visual data analysis, as well as to support on-line execution of “analytical” and “presentation” visualization.

CONCLUSIONS

Our objective is to envision the critical and unfolding challenges in data management, analysis, and visualization, and to advance the capacity for collaborative research communities to become high-performing stewards of all distributed, scientific workflow processes associated with extreme scale data sets. This is a complex and exploratory process. Climate modeling as a community activity is becoming integral to an interdependent chain of operations in which global models force other models that predict societal impacts of climate, for example, in agriculture, water resources, energy usage, etc. The ability to quickly implement downscaled dependencies between such operational modeling chains must become routine. Extreme data volumes would tax any foreseeable network bandwidth. Therefore, we recommend that robust multiple data reduction capabilities be developed and deployed at provider sites and readily provisioned to users via common standards.

In the short-term, much of the work we discuss continues to be neglected and needs immediate action if the community is to realize extreme scale data management, analysis, and visualization for the future.

PRIORITY RESEARCH DIRECTIONS

MODEL DEVELOPMENT AND INTEGRATED ASSESSMENT

Co-Leads: David Bader, Lawrence Livermore National Laboratory
Bill Collins, Lawrence Berkeley National Laboratory

INTRODUCTION

The panel participants identified the following priority research directions:

- How will the sea level, sea-ice coverage, and ocean circulation change as the climate changes?
- How will the distribution and cycling of water, ice, and clouds change with global warming?
- How will extreme weather and climate change on the local and regional scales?
- How do the carbon, methane, and nitrogen cycles interact with climate change?

HOW WILL THE SEA LEVEL, SEA-ICE COVERAGE, AND OCEAN CIRCULATION CHANGE AS THE CLIMATE CHANGES?

This priority research direction focuses on multiscale models of ice sheets and ocean/ice interaction, models for ocean circulation at truly eddy-resolving resolution, and fully dynamic models of land ice for sea-level rise.

Scientific and Computational Challenges

- What are the important processes governing ice sheet melt?
- How do we more accurately represent important vertical mixing in the ocean?
- How do mixing eddies and surface forcing combine to affect the stability and variability of the Meridional Overturning Circulation?

Potential Scientific Impact

- These models will quantify the rate of future sea-level rise.
- These models will improve quantification of ocean heat uptake and exchange of carbon dioxide.
- These models will provide more robust estimates of sea-level rise and its impact on coastal communities and energy infrastructure (time scale: 5 to 10 years).
- These models will quantify the risk of abrupt changes in ocean circulation (time scale: 5 to 10 years).

HOW WILL THE DISTRIBUTION AND CYCLING OF WATER, ICE, AND CLOUDS CHANGE WITH GLOBAL WARMING?

This priority research direction focuses on the development of global cloud-resolving models and their applications.

Scientific and Computational Challenges

- Determining the critical cloud controls on climate
- Determining the importance of motions and particle-scale processes that are still unresolved

Potential Scientific Impact

- These models will enable rapid progress in a wide variety of climate science issues (where clouds play an important role).
- These models will bridge scales from weather to climate for the first time.
- These models ultimately will improve our ability to project changes in regional water cycles, a critical element of integrated assessment (time scale: 5 to 10 years).
- Cloud-resolving models will be used to improve traditional climate models used for climate projection (time scale: 2 to 5 years).
- Impacts of water resources will be quantified on energy production and use (time scale: 5 to 10 years).

HOW WILL EXTREME WEATHER AND CLIMATE CHANGE ON THE LOCAL AND REGIONAL SCALES?

This priority research direction focuses on hierarchical multiscale models, a unified weather and climate prediction framework, and ways in which specific physical processes are connected with the prediction of extreme events.

Scientific and Computational Challenges

- Requiring massive ensembles and long-term integrations to quantify extreme events
- Using tools for analyzing high-frequency, high-resolution data on climate time scales
- Determining what governs the statistics of weather, especially extreme weather

Potential Scientific Impact

- Understand the role of tropical cyclones in climate
- Understand the frequency, duration, intensity, and spatial distribution of droughts, deluges, and heat waves
- Understand the processes behind long-distance interactions of weather and climate
- Understand how changes in weather extremes affect environmental security and societal stability
- Understand how the frequency and intensity of disruptive extreme events evolve with global warming

HOW DO THE CARBON, METHANE, AND NITROGEN CYCLES INTERACT WITH CLIMATE CHANGE?

This priority research direction focuses on modeling the coastal zone and biogeochemical cycles that capture ecosystem demography and its temporal evolution. It also covers process models of land-surface disturbance, anthropogenic forcing of the natural cycles, and process models of natural biogeochemical nutrients.

Scientific and Computational Challenges

- Determining the natural methane and nitrogen cycles and how are they being altered by climate change
- Determining the strength of the global carbon sink and how it will change

Potential Scientific Impact

- Understand the carbon cycle over the Holocene
- Understand the significant roles of life forms and cycles in the carbon cycle
- Understand the significant roles of human-driven and natural disturbance on the carbon and nitrogen cycles
- Quantify impacts of land use change on weather and climate driven by biofuel and other energy production and use over the next 5 to 10 years
- Assess related geo-engineering measures over the next 10 to 15 years

ALGORITHMS AND COMPUTATIONAL ENVIRONMENT

Co-Leads: John Drake, Oak Ridge National Laboratory
Mark Taylor, Sandia National Laboratories

INTRODUCTION

The ability of our modeling and analysis tools to take advantage of extreme scale computing would have a significant impact on climate science. Many of the science questions require high-resolution models with large ensembles of forecasts, a setting well suited for extreme scale computing. However, algorithmic development and exploration of new scalable programming paradigms are necessary to enable and take full advantage of the new computers. The workshop recommended research and development take place with international and national collaborations. The following priority research directions were identified where the climate modeling community would face significant technical challenges:

- develop scalable algorithms for non-hydrostatic atmospheric dynamics with quasi-uniform grids, implicit formulations, and adaptive and multiscale and multiphysics coupling
- foster international consortia for parallel input/output, metadata, analysis and modeling tools for regional and decadal multimodel ensembles
- develop multicore and deep memory languages to support parallel software development.

These research directions must overcome computational barriers to achieve the scientific goals outlined in the climate modeling and decadal predictability panel reports. Research efforts in these directions would yield productivity gains and enable the next generation climate models to use scalable computing environments.

DEVELOP SCALABLE ALGORITHMS

This priority research direction focuses on the development of numerical algorithms that efficiently use upcoming petascale and extreme scale architectures.

Scientific and Computational Challenges

- Increasing throughput of ESMs, especially in the 10- to 50-km regime
- Building global cloud-resolving models
- Improving both the fidelity and computational performance of multiphysics and mutliscale model coupling approaches
- Improving the speed at which new components can be integrated and evaluated in ESMs
- Developing efficient, scalable-data-assimilation approaches for decadal climate predictability work
- Building and running global cloud-resolving climate models at the necessary throughput rates would introduce the largest algorithmic challenges

Scalability would be needed in every component of the ESM. The biggest scalability bottleneck in today's ESMs is the atmospheric dynamical core. This bottleneck is created partly by the pole problem introduced by the use of latitude/longitude grids. New dynamical cores with cubed sphere and icosahedral girds along with yin-yang (baseball) and other overlapping systems show promise for scalability in highly resolved models (see Figure 6). Cloud-resolving models (and maybe even next-generation ESMs) would require a non-hydrostatic dynamical core; thus, determining the best formulation of the non-hydrostatic equations and solvers with appropriate levels of accuracy and conservation for these more uniform grids would remain a priority research direction. The dynamical core is just one example. New, more scalable algorithms would be needed in several other ESM components as well.

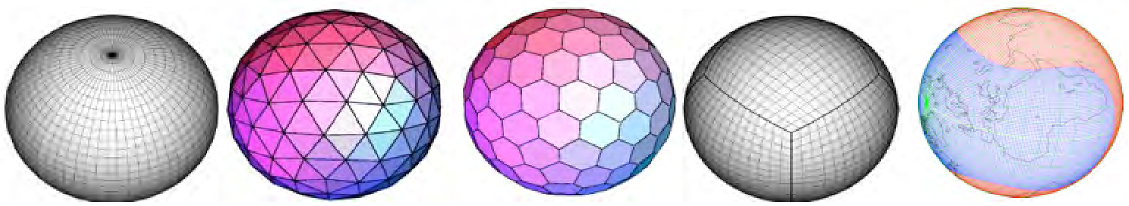


Figure 6. Each type of grid depicted gives a different ordering and type of computation. For example, the first grid allows an ordering of the computation along lines in the east-west direction followed by lines in the north-south direction. The model physical, chemical, and biological calculations relate the variables at a grid point or enclosed area with the neighboring areas. In this way, changes in the variables at one grid point can affect the values at all the other grid points. This communication of variables requires information exchange among processors of the parallel computer. Coordinating the information exchange with the calculation defines current programming of high performance computers. Image courtesy of David Randall (Colorado State University).

In addition to scalability, new coupling mathematics algorithms are needed. To develop the ability to predict future sea-level rise and ocean circulation and to understand the impacts of climate change on extreme events, new components would be added to ESMs. These models would introduce new physics and many new scales, and thus, require research into the best multiphysics and multiscale coupling strategies. All new techniques must be developed from inception with scalability as a key requirement. In addition, the time it takes for model developers to incorporate and evaluate new components must be dramatically reduced. This would require research into improved verification and validation strategies and an investment in associated software engineering support.

Improvements in scalability alone will not be sufficient to obtain the needed throughput (the time it takes to complete a climate simulation). Obtaining the needed level of throughput will also require incorporating as much implicitness as possible and parallelizing methods in time and local time stepping and local mesh refinement approaches, where small time steps and/or the highest resolution is only used in key regions on the globe.

Potential Scientific Impact

- Improved scalability would allow ESMs to effectively use the millions of processing cores envisioned on petascale and extreme scale systems.
- Higher fidelity models with reduced uncertainty for policy and planning.
- New mathematical techniques for multiscale problems.

- High resolution for extremes and cloud characterization for parameterization in global climate models.
- Quickly enable new algorithms and components to improve decadal to century within 5 years
- Enhanced throughput for regional modeling.

FOSTER INTERNATIONAL CONSORTIA

Many of the other science goals require running ensembles of ESMs where the input/output requirements for each individual simulation may not be a challenge, but archiving and analyzing the terabytes of data that would be produced by a large ensemble would require significant enhancements in the parallel performance of all components in the analysis toolset chain. All of this work is common to modeling centers worldwide; therefore, we recommend that petascale-ready tools be provided for the user community, join existing collaborations, and coordinate application for regional and decadal climate prediction.

Scientific and Computational Challenges

- Enhance climate models to exploit parallel input/output and analysis tools. Without an immediate focus on enhancing these tools, it would be difficult to perform global cloud-resolving simulations in the next 2 to 5 years.
- Make new tools accessible to international collaborators.
- Encourage international cooperation between computer hardware and data storage vendors, modeling centers, and research institutes. We recommend that these efforts be catalyzed in the form of an international consortium for parallel input/output, metadata, analysis, and modeling tools.
- Improve input/output performance to support the dramatic increase in data generated by ensembles of long time simulations required for studying extreme events and the large multimodel ensembles needed for decadal predictability efforts.
- Focus on improving workflow issues. These issues are especially important in the realm of decadal predictability, where many collaborators, including international collaborations, need the ability to analyze petabytes of data distributed at modeling centers worldwide.

Potential Scientific Impact

- International collaborations would ensure common metadata and many other requirements to allow analysis of ensembles of multiple models from multiple international efforts.
- Scalable input/output and analysis tools for large distributed data sets are a near-term limiter. Research in this area would enable regional climate change information from ensembles of high-resolution simulations and helping answer outstanding questions of deep-water formation, the role of land-use change on the hydrologic cycle, and decadal prediction.
- Among the international research community, modeling and analysis challenges are effective tools and libraries for input and output, international collaboratory access, workflow for ensemble decadal prediction, analysis, and code parallelization support.

- Global cloud-resolving resolutions would require scalable input/output performance unimaginable in today's ESMs.
- An international collaboration would encourage vendor participation. As climate models provide regional information that is more detailed, access to climate model data would be increasingly in demand.
- New tools making these data accessible to a wider range of scientists and non-scientists make the continued support of tools research and development a priority with impacts in the 10- to-20-year period.
- This research also would be required to realize the scientific impacts of multicenter, multimodel ensembles for decadal prediction, as called for by the Decadal Prediction and Predictability panel of this workshop.
- An international consortium to provide tools for high-end modeling and for data manipulation and analysis would be important for the international community in analyzing climate model output for analysis of climate change impacts.

DEVELOP MULTICORE AND DEEP MEMORY LANGUAGES

No other technology has increased in power by a factor of a billion in 30 years. The high-performance computing industry tracks the power of a system in the number of arithmetic calculations that can be performed in one second. By this measure, the computational power of the top computers over the last 30 years has increased exponentially through the thousands, millions, billions, and trillions to the present petaflop (10^{15} floating-point operations) systems. Following the industry roadmap and the historical trend, the systems of 2020 would boast systems 1000 more powerful with exaflop (10^{18} floating-point operations per second) performance. To take advantage of the computational power and advance climate science, the numerical methods used in global general circulation models must map onto the processors of the computer. If the Earth is represented by a number of grid points, as shown in Figure 6, then one way to map the computation is to divide these points among the processors of a parallel computer. The roadmap for the development of high performance computing has broad implications for the economic development as well as scientific progress (HECRTF 2004).

Scientific and Computational Challenges

- As we move towards extreme scale computing systems, we expect applications would need to run on systems with millions of cores and diverse architectures. Current programming models (MPI+FORTRAN) and fault tolerance approaches may not scale to the numbers of processors and concurrent tasks anticipated in extreme scale systems. We recommend improving fault detection and resilience strategies. If this does not happen, the likelihood of undetectable errors would increase and compromise the value of large data sets.
- High-performance computing architectures are changing quickly, and the modeling community needs to develop programming models and auto-tuning technologies (as a risk mitigation strategy) that offer performance portability and fault resilience in the face of uncertainty in the future computing architectures.

- Obtaining reasonable performance on multicore platforms. Therefore, we recommend better performance introspection to do auto-tuning and a greater understanding of causality with sufficient coverage by instrumentation to understand performance.
- Scaling to millions of processing cores requires parallel scalability in every aspect of the climate model; therefore, domain scientists would have to be parallel-computing experts unless the community can define software engineering guidelines encoded in community frameworks, which are built around any new programming model. We cannot run on these systems in 10 years if substantial progress is not made during the next 5 years.
- A new programming infrastructure would enable sustained extreme scale performance on extreme scale systems. Without this, extreme scale systems would fail to meet the performance requirements necessary to achieve the desired science impacts.
- ESMs are becoming extremely complex requiring extensive software engineering to change programming models. We recommend a concerted effort be made to prototype key climate algorithms to understand the relative merits of alternative programming technologies.

Potential Scientific Impact

- Higher-level programming models would reduce the time-to-production status for these new/novel climate models in the 10- to-20-year period.
- Enable the highest resolution global cloud-resolving simulations and other simulations with the most complex ESMs that require full use of million-core extreme scale class systems
- The impact of performance portability on the existing diversity of multicore architectures and accelerators (Cell/GPGPU, etc.) would be to reduce effort on algorithm tuning and redevelopment and thus accelerate model development and deployment.
- Fault resilience and auto-tuning enables better throughput and efficient use of expensive hardware resources.
- Enable high-resolution runs with higher-fidelity climate science and more climate model complexity.
- Improve productivity of the software engineers to reduce time-to-production status for new-novel climate models.

DECADAL PREDICTION AND PREDICABILITY

Lead: Ben Kirtman, University of Miami, Rosenstiel School of Marine and Atmospheric Science

INTRODUCTION

The panel participants identified the following priority research directions:

- identify potential sources and mechanisms for decadal predictability
- develop strategies for tapping into this predictability and ultimately realizing predictions that have societal benefit.

IDENTIFY SOURCES AND MECHANISMS FOR DECADAL PREDICTABILITY

This priority research direction focuses on identifying potential sources for decadal predictability.

Scientific and Computational Challenges

- Determine how best to balance resolution versus complexity versus ensemble size given current and future computational resources.
- Establish what time and space scales should be included in the models; we need to understand how the space and time scale interaction potential contribute and limit the predictability.
- Determine how complexity within the component systems contributes to predictability. For example, there may be predictability to be mined in the growth and life cycle of vegetation, and these processes may need to be accurately modeled.
- Determine precisely which interactions are needed and how to accurately simulate the feedbacks requires focused research efforts.

Potential Scientific Impact

Focused efforts are needed specifically designed to identify the potential sources of decadal predictability. Presumably, the sources and mechanisms of predictability come from interactions among all the components of the climate system (physical, chemical, and biological). There is predictability because of interactions among the physical components alone (e.g., air-sea, air-land, ocean-ice, etc.) and potential predictability because of interactions across disciplinary boundaries (e.g., how ocean biology affects sunlight penetration). Concerted efforts are needed to determine how large of a (multimodel) ensemble is required to adequately quantify the uncertainty at all space and time scales.

DEVELOP STRATEGIES FOR TAPPING INTO THIS PREDICTABILITY AND ULTIMATELY REALIZING PREDICTIONS THAT HAVE SOCIETAL BENEFIT

This priority research direction focuses on producing predictions that are of societal benefit. This effort requires careful research into the predictability and the prediction of weather within climate. Multiple approaches are likely for initializing decadal predictions, and the research community needs to develop strategies and processes for documenting best practices in forecast initialization. It is critical to acknowledge the importance of a continued and vigorous effort in model improvement across components and time scales. This problem has implications in identifying sources and mechanisms of predictability and in realizing predictions that are of societal benefit. It is important to emphasize that a forecast capability that is of societal use needs to go beyond traditional large-scale time-averaged climatic variables. Target research efforts should document the limit of predictability of changes in extreme weather events and ways to make predictions of potential changes in extreme weather events.

Scientific and Computational Challenges

- Determine how to initialize the meridional overturning circulation in climate system models. The meridional overturning circulation in the Atlantic Ocean is a potential source of predictability.
- Determine the appropriate strategies and methodologies for initializing decadal forecasts.
- Find out how to readily translate data assimilation into the best possible initial condition for the prediction problem.
- We recommend the assimilation of disparate observational data streams into climate models.
- As sources of predictability are identified among interactions of climate system components, new input data streams and new strategies will need to be identified for how to ingest these data streams into the climate model.
- In the near-term, we recommend the establishment of a multi-agency and multi-institutional partnership for sharing models, data, and predictions. Although this partnership should extend beyond the near-term, on this shorter time scale it should focus on supporting the international CMIP5/AR5 decadal prediction experiment. This support should not only include meeting the computational needs for completing the numerical experiment but also facilitate the collection and sharing (i.e., dissemination) of the model-based, retrospective forecast data and the forecasts for the period 2010 to 2030. The partnership should contribute to the understanding the CMIP5/AR5 results by providing the necessary research tools (i.e., initial condition data, verification data, observational estimates of both raw and analyzed, and data assimilation systems). This partnership also should share model codes, documentation, user guides, and computational facilities for numerical hypothesis testing.

Potential Scientific Impact

- Research should continue in translating or downscaling climate forecast to regional ecosystems and regional hydrology.
- This multi-agency and multi-institutional partnership would foster a renewed emphasis on workforce development, which would enhance educational collaborations as well as current government-sponsored postgraduate education and post-doctoral training activities. This effort would focus on “hands-on” training so the next generation of scientists would not only be capable of running a climate model but also designing and implementing sophisticated hypothesis-testing experiments.

DATA, VISUALIZATION, AND COMPUTING PRODUCTIVITY

Co-Leads: Dean Williams, Lawrence Livermore National Laboratory
Don Middleton, National Center for Atmospheric Research

INTRODUCTION

The panel participants identified the following priority research directions:

- develop new, robust techniques for the input/output, storage, processing, and wide-area transport demands of extreme scale data
- integrate diverse and complex data
- dedicate resources to the development of standards, conventions, and policies.

DEVELOP NEW, ROBUST TECHNIQUES FOR THE INPUT/OUTPUT, STORAGE, PROCESSING, AND WIDE-AREA TRANSPORT DEMANDS OF EXTREME SCALE DATA

This priority research direction would demonstrate how to build and operate internationally federated data and knowledge systems that keep pace with societal needs. This effort is essential for the advancement of climate science and is at the very foundation of community-wide sharing.

Scientific and Computational Challenges

- Community sharing of extreme scale climate data resources is fundamental to the advancement of climate science. The ESG represents an embryonic foundation for this systems-of-systems approach.

Potential Scientific Impact

- Such systems would allow the community to expand available data resources, share broadly, and readily undertake new research assessment activities.
- The ESG represents an embryonic base for this systems-of-systems approach, which enables activities for future climate assessment reports, such as IPCC (2007).
- This effort would demonstrate how to build and operate internationally federated data and knowledge systems that keep pace with societal needs.
- As a national and international challenge, such systems would allow the community to expand available data resources, share broadly, and readily undertake new research assessment activities.

INTEGRATE DIVERSE AND COMPLEX DATA

This priority research direction would involve integrating data of different formats, grid structures, metadata, standards, and conventions.

Scientific and Computational Challenges

- Climate research is impeded by the difficulties posed by diverse data products with different formats, different grid structures, ill-defined metadata, and inadequate standards and conventions.
- The community lacks powerful facilities for analyzing large climate research data sets, and the current analysis and visualization tools are not scalable.

Potential Scientific Impact

- Advances here would enable multiple, diverse communities to interact and collaborate. The Exascale Climate Analysis Facilities (ECAF) can serve in a strategic role by deploying new integration tools.
- ECAF can help to improve vocabulary services, semantic services, and powerful regridding services.
- The combination of facilities and tools would revolutionize the productivity of the climate science research community and use computer/storage resources more efficiently.
- Dedicate resources to the development of standards, conventions, and policies, and contribute to related committees.

DEDICATE RESOURCES TO DEVELOPING STANDARDS, CONVENTIONS, AND POLICIES

This priority research direction would serve to regulate the national and international communities and help keep pace with societal demands.

Scientific and Computational Challenges

- International climate science relies upon standards, policies, and even international security agreements among unfunded heroic volunteers. For example, at the extreme scale level, converting and reformatting data sets are untenable.
- Developing new robust techniques for dealing with the input/output, storage, processing, and wide-area transport demands for extreme scale data.
- To help with the integration of diverse complex data, advancements in the development of format-neutral data interfaces, along with effective methods for dealing with diverse grid structures and advanced metadata are required.
- Continuous efforts must be dedicated to developing standards, conventions, and policies, and contributing to related committees. These collective efforts would serve as an exemplar for other disciplinary communities that need to mobilize nationally and internationally.

Potential Scientific Impact

- This priority research direction can serve as a demonstrative model for delivering facilities and tools for a select disciplinary community, and it can transform statistical processing for large ensembles of models.

- The combination of facilities, tools, and governance would revolutionize the productivity of the climate science research community and use computing and storage resources more efficiently.
- Scalable analysis and robust visualization infrastructures would be deployed on open, dedicated ECAF.
- These infrastructures will foster an interagency effort to advance the development of tools and related algorithms and workflow capabilities.
- This priority research direction is fundamental to accomplishing our other research and development priorities and to sustaining the tools and systems. It will enable the realization of high-performing climate science that keeps pace with society’s demands.

- The combination of facilities, tools, and governance would revolutionize the productivity of the climate science research community and use computing and storage resources more efficiently.

REFERENCES

Drake JB, PW Jones, M Vertenstein, JB White III, and PH Worley. 2008. "Software Design for Petascale Climate Science." Chapter 7 in Petascale Computing: Algorithms and Applications," ed. D.A. Bader, Chapman & Hall/CRC; Boca Raton, FL.

European Network for Earth System Modeling (http://www.enes.org), Max-Planck Institute for Meteorology, Hamburg, Germany.

HECRTF – High End Computing Revitalization Task Force. 2004. *Federal Plan for High End Computing: Report of the High End Computing Revitalization Task Force.* Executive Office of the President, Washington D.C. Available online at: http://www.nitrd.gov/Pubs/2004_hecrtf/20040702_hecrtf.pdf

IPCC - Intergovernmental Panel on Climate Change. 2007. *Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Intergovernmental Panel on Climate Change, Geneva, Switzerland.

Simon H, T Zacharia, and R Stevens. 2008. *Modeling and Simulation at the Exascale for Energy and the Environment.* U.S. Department of Energy, Washington D.C. Available at http://www.sc.doe.gov/ascr/ProgramDocuments/ProgDocs.html

APPENDICES

APPENDIX 1: WORKSHOP AGENDA

APPENDIX 2: WORKSHOP PARTICIPANTS

APPENDIX 3: ACRONYMS

APPENDIX 4: PREVIOUS DOCUMENTS

APPENDIX 1: WORKSHOP AGENDA

Thursday, November 6, 2008

7:45 a.m. – 8:00 a.m.	Warren Washington: Introduction
8:00 a.m. – 8:30 a.m.	Dr. Raymond Orbach (video): Welcome Address
8:30 a.m. – 8:45 a.m.	DOE/ASCR Representative: ASCR Workshop Goals and Welcome
8:45 a.m. – 9:00 a.m.	DOE/BER Representative: BER Workshop Goals and Welcome
9:00 a.m. – 9:20 a.m.	John Drake and Mark Taylor: Overview of Climate Algorithms and Computational Environment
9:20 a.m. – 9:40 a.m.	Don Middleton and Dean Williams: Overview of Data, Visualization, and Computing Productivity
9:40 a.m. – 10:20 a.m.	Warren Washington: Informal Group Discussion and Networking
10:20 a.m. – 10:40 a.m.	David Bader and Bill Collins: Overview of Model Development
10:40 a.m. – 11:00 a.m.	Ben Kirtman: Overview of Decadal Predictability and Prediction
11:00 a.m. – 11:30 a.m.	Horst Simon: Future Trends in Computing
11:30 a.m. – 12:30 p.m.	Warren Washington: Working Lunch- Breakout session agenda and expectations
12:30 p.m. – 4:00 p.m.	**Breakout Sessions**
	Model Development and Integrated Assessment Algorithms and Computational Environment
	Data, Visualization, and Computing Productivity
	Decadal Predictability and Prediction
4:00 p.m. – 5:00 p.m.	Slide Preparation and Summary
5:00 p.m. – 7:00 p.m.	Working Dinner: Breakout Session Reporting Breakout Session Co-Leads
7:00 p.m. – 7:30 p.m.	Wrap up and next day’s agenda

Friday, November 7, 2008

7:00 a.m. – 8:00 a.m.	Warren Washington: Working Breakfast: Agenda and Expectations for Day 2 Breakout Sessions
8:00 a.m. – 10:30 a.m.	**Breakout Sessions** Model Development and Integrated Assessment Algorithms and Computational Environment Data, Visualization, and Computing Productivity Decadal Predictability and Prediction
10:30 a.m. – 11:00 a.m.	Warren Washington: Plenary Session-Closing Remarks
11:00 a.m. – 2:00 p.m.	Report Writing Team: Working Lunch-Letter Report and Initial Draft/Outline Workshop Report Writing
2:00 p.m. – 3:00 p.m.	Warren Washington: Workshop Coordinating Committee-Discussion of Next Steps

APPENDIX 2: WORKSHOP PARTICIPANTS

Name	Affiliation	E-mail
Panel Co-Leads		
Bader, David	Lawrence Livermore National Laboratory	Bader2@llnl.gov
Collins, Bill	Lawrence Berkeley National Laboratory	wdcollins@lbl.gov
Drake, John	Oak Ridge National Laboratory	drakejb@ornl.gov
Taylor, Mark	Sandia National Laboratories	mataylo@sandia.gov
Williams, Dean	Lawrence Livermore National Laboratory	Williams13@llnl.gov
Middleton, Don	National Center for Atmospheric Research	don@ucar.edu
Kirtman, Ben	University of Miami	bkirtmand@rsmas.miami.edu
International Participants		
Duemenil-Gates, Lydiea	Institut für Meteorologie	lydia.dumenilgates@met.fu-berlin.de
Flato, Greg	Canadian Centre for Climate Modeling and Analysis	Greg.flato@ec.gc.ca
Foujols, Marie-Alice	IPSL	Marie-alice.foujols@ipsl.jussieu.fr
Hamrud, Mats	European Centre for Medium-Range Weather Forecasts	Mats.Hamrud@ecmwf.int
Nakashiki, Norikazu,	Central Research Institute of Electric Power Industry, Japan	nakasiki@criepi.denken.or.jp
Vidale, Pier Liugi	University of Reading, United Kingdom	P.L.Vidale@reading.ac.uk
Participants		
Abeles, Jim	IBM	jabeles@us.ibm.com
Anderson, Don	National Aeronautics and Space Administration	Donald.anderson-1@nasa.gov
Archibald, Rick	Oak Ridge National Laboratory	archibaldrk@ornl.gov
Balaji, Venkatramani	Princeton University	V.Balaji@noaa.gov
Bamzai, Anjuli	DOE Office of Biological and Environmental Research	Anjuli.bamzai@science.doe.gov
Bansil, Arun	DOE Office of Basic Energy Sciences	bansil@neu.edu
Beckman, Peter	Argonne National Laboratory	Beckman@mcs.anl.gov
Biven, Laura	DOE	Laura.biven@science.doe.gov
Buja, Lawrence	National Center for Atmospheric Research	southern@ucar.edu
Carr, George	SGI	gcarr@sgi.com
Carruthers, Julie	DOE Office of Science	Julie.carruthers@science.doe.gov
Chatterjee, Lali	DOE	Lali.chatterjee@science.doe.gov
Collela, Phillip	Lawrence Berkeley National Laboratory	pcolella@lbl.gov
Cotter, Steve	DOE	steve@es.net
Craig, Tony	National Center for Atmospheric Research	tcraig@ucar.edu
Dahlman, Roger	DOE Office of Science	Roger.dahlman@science.doe.gov
Dattoria, Vince	DOE	Vince.dattoria@science.doe.gov
Dennis, John	National Center for Atmospheric Research	dennis@ucar.edu
Dickinson, Bob	University of Texas	robted@jsg.utexas.edu
Dimotakis, Paul	Jet Propulsion Laboratory	Paul.e.dimotakis@jpl.nasa.gov
Donner, Leo	Geophysical Fluid Dynamics Laboratory	Leo.j.donner@noaa.gove
Erickson, David	Oak Ridge National Laboratory	ericksondj@ornl.gov
Fein, Jay	National Science Foundation	jfein@nsf.gov
Ferrell, Wanda	DOE Office of Biological and Environmental Research	wanda.ferrell@science.doe.gov
Ghan, Steven	Pacific Northwest National Laboratory	Steve.ghan@pnl.gov
Gross, Brian	Geophysical Fluid Dynamics Laboratory	Brian.gross@noa.gov
Hack, James	Oak Ridge National Laboratory	jhack@ornl.gov

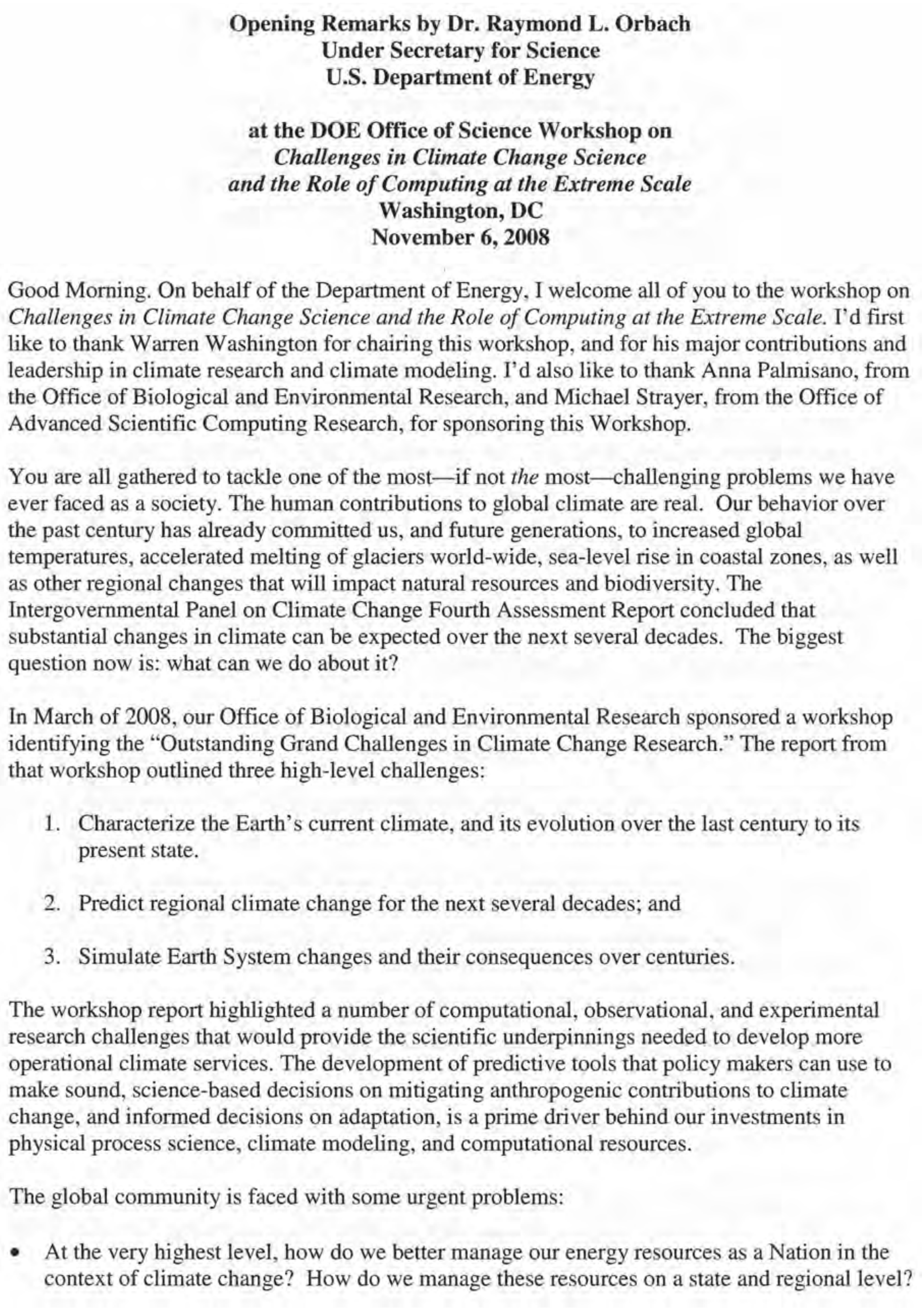
Hitchcock, Dan	DOE Office of Advanced Scientific Computing Research	Daniel.hitchcock@science.doe.gov
Hoffman, Forrest	Oak Ridge National Laboratory	hoffmanfm@ornl.gov
Huber, Melissa	Pacific Northwest National Laboratory	Melissa.huber@pnl.gov
Jacob, Rob	Argonne National Laboratory	Jacob@mcs.anl.gov
Janetos, Tony	Pacific Northwest National Laboratory	Anthony.janetos@pnl.gov
Johnson, Fred	DOE	
Johnson, Gary	CSS	garymichaeljohnson@gmail.com
Jones, Philip	Los Alamos National Laboratory	pwjones@lanl.gov
Kellie, Allen	National Center for Atmospheric Research	kellie@ucar.edu
Khaleel, Moe	Pacific Northwest National Laboratory	Moe.khaleel@pnl.gov
Kinter, Jim	COLA	kinter@cola.uges.org
Kuperberg, Mike	DOE Office of Biological and Environmental Research	michael.kuperberg@science.doe.gov
Landsberg, Alexandra	DOE	
Latham, Rob	Argonne National Laboratory	robl@mcs.anl.gov
Lee, Steven	DOE	
Lee, Tsengdar	National Aeronautics and Space Administration	Tsegdar.j.lee@nasa.gov
Lesmes, David	DOE Office of Biological and Environmental Research	david.lesmes@science.doe.gov
Leung, L. Ruby	Pacific Northwest National Laboratory	Ruby.leung@pnl.gov
Marques, Osni	DOE	
McClean, Julie	University of California, San Diego	jmcclean@ucsd.edu
Meehl, Gerald	National Center of Atmospheric Research	meehl@ucar.edu
Messina, Paul	Argonne National Laboratory	messina@mcs.anl.gov
Morss, Sue	DOE	Helaine.morss@science.doe.gov
Ndousse-Fetter, Thomas	DOE	
Nichols, Jeff	Oak Ridge National Laboratory	nicholsja@ornl.gov
Nyberg, Per	Cray, Inc.	Nyberg@cray.com
Palmer, Bruce	Pacific Northwest National Laboratory	Bruce.palmer@pnl.gov
Palmisano, Anna	DOE Office of Biological and Environmental Research	Anna.palmisano@science.doe.gov
Pederson, Mark	DOE Office of Basic Energy Sciences	Mark.Pederson@science.doe.gov
Polansky, Walter	DOE	Walt.polansky@science.doe.gov
Randall, Dave	Colorado State University	randall@atmos.colostate.edu
Rasch, Philip	National Center for Atmospheric Research	pjr@ucar.edu
Ray, Doug	Pacific Northwest National Laboratory	Doug.ray@pnl.gov
Rienecker, Michele	National Aeronautics and Space Administration	Michele.m.rienecker@nasa.gov
Robinson, Walt	National Science Foundation	warobins@nsf.gov
Rood, Richard	University of Michigan	rbrood@umich.edu
Sanderson, Ben	National Center for Atmospheric Research	bsander@ucar.edu
Schneider, Edwin	COLA	Schneider@cola.iges.org
Schuchardt, Karen	Pacific Northwest National Laboratory	Karen.schuchardt@pnl.gov
Sekine, Yukiko	DOE Office of Advanced Scientific Computing Research	yukiko.sekine@science.doe.gov
Semtner, Bert	Naval Postgraduate School	bsemtner@sbcglobal.net
Shalf, John	Lawrence Berkeley National Laboratory	jshalf@lbl.gov
Shoshani, Arie	Lawrence Berkeley National Laboratory	shoshani@lbl.gov
Simon, Horst	Lawrence Berkeley National Laboratory	hdsimon@lbl.gov
Spotz, Bill	DOE Office of Advanced Scientific Computing Research	wfspotz@ascr.doe.gov
Stevens, Rick	Argonne National Laboratory	stevens@anl.gov
Stouffer, Ron	Geophysical Fluid Dynamics Laboratory	Ronald.stouffer@noaa.gov

Straatsma, TP	Pacific Northwest National Laboratory	tps@pnl.gov
Strand, Gary	National Center for Atmospheric Research	strandwg@ucar.edu
Strayer, Michael	DOE Office of Advanced Scientific Computing Research	Michael.strayer@science.doe.gov
Tufo, Henry	National Center for Atmospheric Research	tufo@ucar.edu
Turnbull, Susan	DOE	Susan.turnbull@ascr.doe.gov
Vallario, Bob	DOE Office of Biological and Environmental Research	bob.vallario@science.doe.gov
Vertenstei, Mariana	National Center for Atmospheric Research	mvertens@ucar.edu
Washington, Warren	National Center for Atmospheric Research	wmw@ucar.edu
White, James	Oak Ridge National Laboratory	Trey@ornl.gov
Williamson, Ashley	DOE	ashley.williamson@science.doe.gov
Winchell, Kent	IBM	kentwin@us.ibm.com
Worley, Pat	Oak Ridge National Laboratory	worleyph@ornl.gov

APPENDIX 3: ACRONYMS AND ABBREVIATIONS

AR5	Fifth IPCC Assessment Report
CMIP	Coupled Model Intercomparison Project
DOE	U.S. Department of Energy
ECAF	Exascale Climate Analysis Facilities
ESG	Earth System Grid
ESM	Earth system model
GCM	general circulation model
IPCC	Intergovernmental Panel on Climate Change

APPENDIX 4: PREVIOUS DOCUMENTS



Opening Remarks by Dr. Raymond L. Orbach
Under Secretary for Science
U.S. Department of Energy

at the DOE Office of Science Workshop on
Challenges in Climate Change Science
and the Role of Computing at the Extreme Scale
Washington, DC
November 6, 2008

Good Morning. On behalf of the Department of Energy, I welcome all of you to the workshop on *Challenges in Climate Change Science and the Role of Computing at the Extreme Scale.* I'd first like to thank Warren Washington for chairing this workshop, and for his major contributions and leadership in climate research and climate modeling. I'd also like to thank Anna Palmisano, from the Office of Biological and Environmental Research, and Michael Strayer, from the Office of Advanced Scientific Computing Research, for sponsoring this Workshop.

You are all gathered to tackle one of the most—if not *the* most—challenging problems we have ever faced as a society. The human contributions to global climate are real. Our behavior over the past century has already committed us, and future generations, to increased global temperatures, accelerated melting of glaciers world-wide, sea-level rise in coastal zones, as well as other regional changes that will impact natural resources and biodiversity. The Intergovernmental Panel on Climate Change Fourth Assessment Report concluded that substantial changes in climate can be expected over the next several decades. The biggest question now is: what can we do about it?

In March of 2008, our Office of Biological and Environmental Research sponsored a workshop identifying the "Outstanding Grand Challenges in Climate Change Research." The report from that workshop outlined three high-level challenges:

1. Characterize the Earth's current climate, and its evolution over the last century to its present state.
2. Predict regional climate change for the next several decades; and
3. Simulate Earth System changes and their consequences over centuries.

The workshop report highlighted a number of computational, observational, and experimental research challenges that would provide the scientific underpinnings needed to develop more operational climate services. The development of predictive tools that policy makers can use to make sound, science-based decisions on mitigating anthropogenic contributions to climate change, and informed decisions on adaptation, is a prime driver behind our investments in physical process science, climate modeling, and computational resources.

The global community is faced with some urgent problems:

- At the very highest level, how do we better manage our energy resources as a Nation in the context of climate change? How do we manage these resources on a state and regional level?

What are the most promising near-term strategies for reducing carbon emissions through economic incentives and technology development?

How we optimize our investments in energy technology research and development is coupled to managing our energy resources. The Department of Energy's applied technology programs support the development of a broad range of energy technologies. But we need better tools informed by basic research to help inform technology down-selects in the context of mitigation of carbon emissions, along with other considerations such as technological feasibility, cost, and market potential.

- Another area is the availability of water, central to our quality of life and national security. Our ability to produce food and energy, ability to maintain healthy ecosystems—and the services we depend on from these ecosystems, are inextricably linked to water.

  How will changes in climate affect trends in precipitation on a regional level, and how will those trends impact hydrological systems and water resources? How can we incorporate this knowledge into water management? How can we better prepare for events such a floods and mega-droughts?

  Likewise, land-use changes, crop production for food and biofuels, and water resources and how they affect and are affected by climate are deeply intertwined.

There are other challenging issues that society and policy makers must face:

- How, and to what degree, will climate change affect national economies?
- What are the most urgent threats from sea level rise and changes in patterns of extreme weather events, and how can we better prepare for them?
- Who will be the most vulnerable, and what will the consequences for human health and other aspects of our society be?

The limitation of current climate models to address regional and local-scale impacts on time-scales of interest to society and decision makers presents a challenge, and an opportunity. To address the types of questions I just mentioned with the scientific understanding and reliable tools that are required, we must have the following:

1. Robust long-term climate models of the coupled climate system—based on reliable physical and biological observations—that can project to a century-and-beyond time scale.
2. Human behavior fully integrated into long-term climate models and integrated assessment models; and
3. Sufficient computation resources to enable policy makers to assess—in a reasonable time frame—the consequences of different policies on long-term global and regional climate change.

These points bring me to the important work you have ahead of you over the coming days:

All four breakout sessions of today's workshop will address issues related to the achievement of robust long-term climate models that can project to a century-and-beyond timescale. The first session—on Model Development and Integrated Assessments—will identify critical model improvements. The second session—on Algorithms and Computational Environment—will deliberate how dynamics can be improved through algorithms that have enhanced numerical accuracy. The third session—on Data, Visualization and Productivity—will tackle challenges of data mining and visualization challenges associated with 'extreme-scale' data sets. Collectively, the second and third sessions will be looking into making the climate models more robust in terms of computational reliability. The fourth session—on Decadal Predictability and Prediction—will address the new emerging topic of observation-based prediction. This adds the experiment of numerical weather prediction techniques to the climate change arena.

The rigorous incorporation of the human dimension should play a role in each of the break-out sessions of this workshop. How will humans adapt to a changing climate? What is the role of human ingenuity through science and technology and innovation? What role will climate change and its impact play on any or all of the behaviors that drive climate change? These are the compelling questions that will determine how we respond to the climate change challenge—whether it is mitigation, adaptation, or a balance in both.

The 2006 Stern Review, *The Economics of Climate Change*, was a serial approach to the integration of human behavior. Nicolas Stern took the consequences—as they were known then—of human behavior on climate, and then he calculated the economic consequences of climate change. He then explored the economic costs of stabilizing green house gases.

Current Integrated Assessment Models do indeed model human behavior, but primarily through an economic lens. The development of fully integrated assessment would incorporate improved understanding of human behavior. It is essential that social scientists become a part of the climate change science community. DOE is looking to partner with the social science community to build this capability into future fully integrated models.

As society explores specific mitigation strategies and prepares to cope with and adapt to a changing climate, significant improvements to both Integrated Assessment Models and to Earth System Models are required: in particular, closer coupling and interoperability between the two are necessary. Integrated Assessment Models are prime beneficiaries of the improved predictive power and higher resolution made possible by Earth System Models, and often bridge the divide from deep science-based process models to broad science-based decision support tools. In turn, realistic scenarios by Earth System Models are dependent upon human behavior examined from decades to centuries in Integrated Assessment Models. The next round of IPCC will build on the strong interactions between members of the Earth System Modeling and Integrated Assessment Modeling communities.

Incorporation of human behavior is central to understanding where climate is heading, and our options for mitigation and adaptation. The interdependencies between the Earth System and Integrated Assessment Models…and the modeling communities…signal a major shift—the emergence of a "system of systems." It is my hope that DOE will provide the foundations and resources to accelerate this shift, having expertise and leadership in both classes of models…and the computational infrastructure to explore and strengthen the connections…to improving understanding of the human dimensions of climate change.

The next generation of high resolution Earth System models that better simulate the interactions and feedbacks among the physical and biological component processes will require greater computing capabilities. These complex component processes include the carbon cycle, clouds, atmospheric chemistry, aerosols, large ice sheets, vegetation, as well as human systems. Ultimately these models will provide improved simulation and prediction of changes in temperature, precipitation, and extreme weather events. In addition, these models will be effective at finer scales and shorter time periods. Regional-scale projections of climatic change will also require an increase in the spatial resolution of climate models, and an acceleration of computational throughput with a combination of software and hardware advances. The break-out sessions on Model Development and Integrated Assessment will cover these points.

The next generation of Earth System models will challenge current frameworks for computation, communication, the free exchange of observations and simulations, and analysis. For decisions at the regional level, high resolution—both in terms of space and time—will be required for assessment of regional climate variability and change. The concomitant increase in spatial and temporal resolution, and need for accelerated throughput, will require expanded high-end computing resources. The breakout session on Data, Visualization and Productivity, as well as session on Algorithms and Computational Environment, will deal with some of these issues.

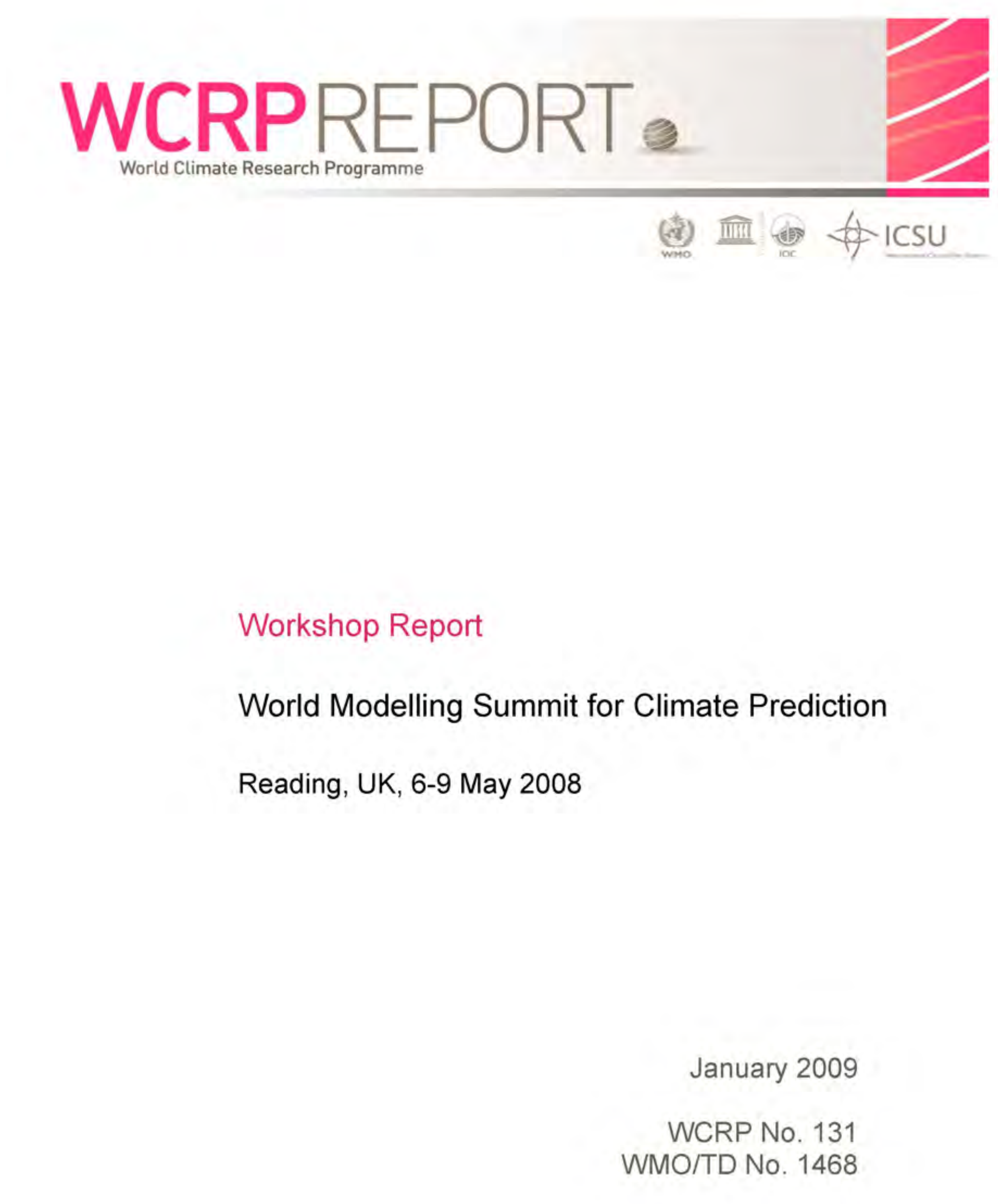
To tackle the current and future challenges of climate change, a qualitatively different level of scientific understanding, modeling capabilities, and computational resources and infrastructure will be required from what is currently available now. A new level of integration between modeling and observational science, new mathematical methods and algorithm techniques, and a flexible high-performance computing infrastructure is required to better quantify prediction uncertainties.

In this vein, the Lawrence Berkeley, Oak Ridge, and Argonne national laboratories held three town hall meetings in 2007 to engage the computational science community in a series of open discussions about the potential benefits of advanced computing at the exascale, addressing such problems as energy, the environment, and basic science. There are significant challenges to reach the exascale, particularly in the areas of architecture, scale, power, reliability and cost. The climate modeling community will have much to gain from these advancements. The exascale frontier will enable dramatic improvements in climate model representations of physical, chemical, and biological processes; it will enable much higher model resolution to understand regional impacts of climate change; and it will enable a greater understanding of the sources of uncertainty in climate models.

DOE has been partnering with industry to push computing and network capabilities for open science at a remarkable pace in recent years. In 2003 we had 7 teraflops on the floor at NERSC. In a matter of weeks from now, DOE's Oak Ridge National Leadership Computing Facility will complete the final upgrade to the Cray XT Jaguar supercomputer. Jaguar will be the world's first petaflop system dedicated to open research. While scientific opportunity increases dramatically with increases in computational power, maintaining support of the software and codes on frequently changing architectures is essential for scientific productivity. A software framework that scales from the desktop to the petascale, and supports multi-scale model development and process integration, is a formidable challenge.

Developing the knowledge and the tools to understand and address the challenges associated with climate change requires a coordinated national effort and an international effort. I want to thank those of you that have traveled from abroad to be here this week to help foster an international dialogue at this workshop. It is difficult to overestimate how important the work of your community is to the future of our planet. I wish all of you the most productive discussions in the coming days. Our future as a world depends on it.

Thank you.



WCRP REPORT

World Climate Research Programme

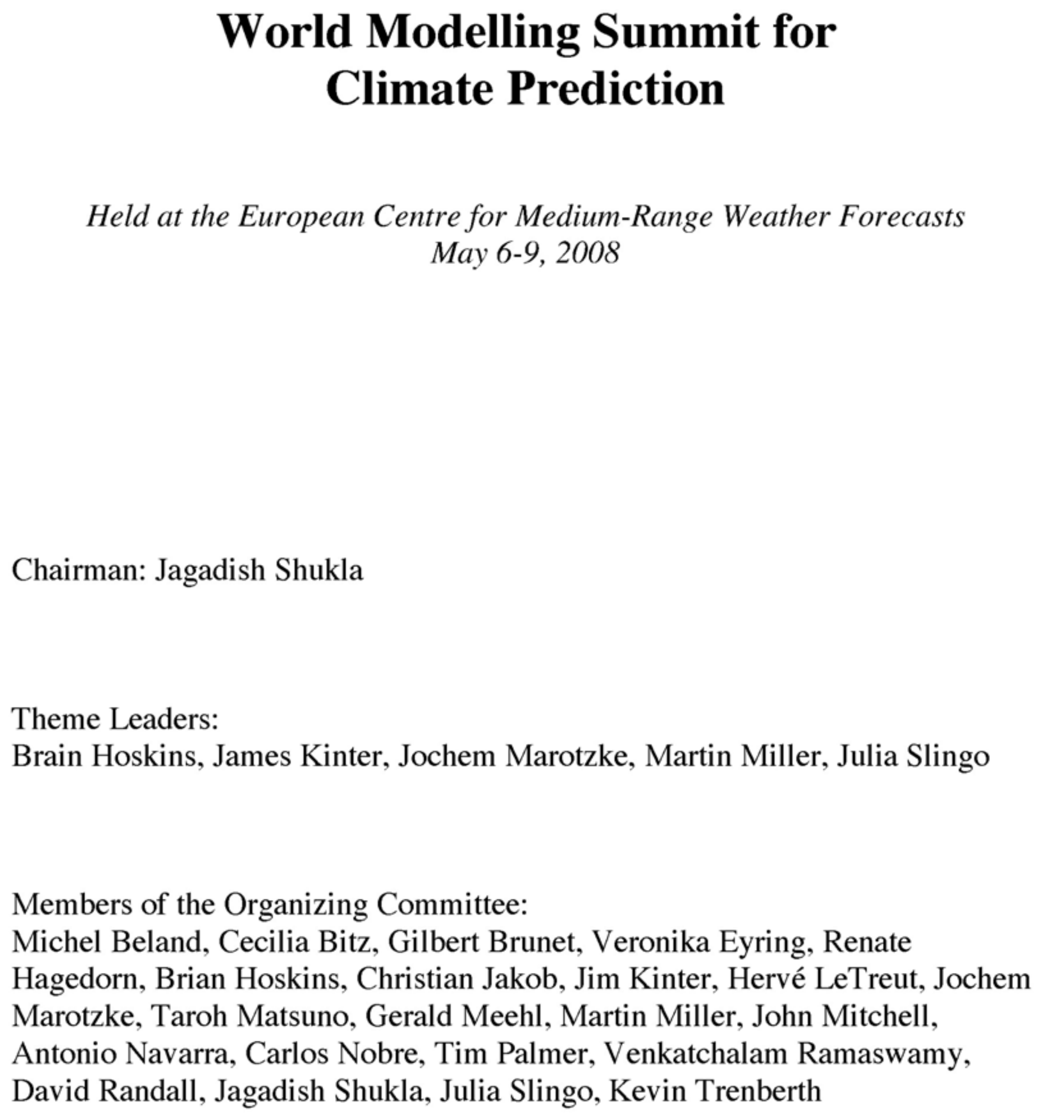
Workshop Report

World Modelling Summit for Climate Prediction

Reading, UK, 6-9 May 2008

January 2009

WCRP No. 131
WMO/TD No. 1468



World Modelling Summit for Climate Prediction

Held at the European Centre for Medium-Range Weather Forecasts
May 6-9, 2008

Chairman: Jagadish Shukla

Theme Leaders:
Brain Hoskins, James Kinter, Jochem Marotzke, Martin Miller, Julia Slingo

Members of the Organizing Committee:
Michel Beland, Cecilia Bitz, Gilbert Brunet, Veronika Eyring, Renate Hagedorn, Brian Hoskins, Christian Jakob, Jim Kinter, Hervé LeTreut, Jochem Marotzke, Taroh Matsuno, Gerald Meehl, Martin Miller, John Mitchell, Antonio Navarra, Carlos Nobre, Tim Palmer, Venkatchalam Ramaswamy, David Randall, Jagadish Shukla, Julia Slingo, Kevin Trenberth

I. PREFACE:

The World Modelling Summit for Climate Prediction was co-sponsored by the World Climate Research Programme (WCRP), World Weather Research Programme (WWRP), and the International Geosphere Biosphere Programme (IGBP), "to develop a strategy to revolutionize the prediction of the climate to address global climate change, especially at regional scale." The primary emphasis of the Summit was on the simulation and prediction of the climate system, but the participants recognized similar challenges/opportunities in weather and other environmental simulation and predictions and that these fields can also benefit from the discussions and recommendations of the Summit. They acknowledged that challenges/opportunities in research, development and verification of climate models span across a wide range of time (intra- and inter-seasonal, decadal, centennial, and longer), and space (global, continental, regional, and other) scales that require immediate attention of climate, weather and environmental scientists, funding agencies and political leaders, globally. The Summit was very well organized and effectively hosted by the European Centre for Medium-Range Weather Forecast (ECMWF). The success of the Summit was due in large part to the major efforts of the organizers, host, sponsors and the participants. This Summit was indeed a major accomplishment in bringing the world leaders together and in defining a set of common objectives/priorities.

The participants identified four major objectives/priorities of: 1) developing models that represent realistically all aspects of the climate system; 2) confronting these models with observations to evaluate their adequacy, accuracy and shortcomings towards building confidence in their future projections; 3) obtaining computational capabilities that are three to four orders of magnitude greater than the best available capabilities today; and 4) establishing a world climate modelling project/programme that benefits from the expertise and investments of the nations around the world to achieve these priorities. The participants recognized all four objectives to be challenging beyond the resources and capabilities of any single nation, thus identified the opportunity for global coordination and cooperation as we move on towards accomplishing them. The participants called on the global environmental research programmes (IGBP, WCRP, WWRP, etc.) and their sponsoring organizations the International Council for Science (ICSU), United Nations Education, Science and Cultural Organization (UNESCO), Intergovernmental Oceanographic Commission (IOC), United Nations Environment Programme (UNEP), and the World Meteorological Organization (WMO), and their member countries to adopt and support these recommendations for implementation in the near-, intermediate- and long-term.

We believe the Summit was successful in achieving its primary goal of identifying the common priorities that are shared and endorsed by the participants, but the challenge of realizing them will be with us during the next decades. The Summit recommendations are quite timely as the world leaders are preparing to convene the World Climate Conference Three (WCC-3), in 2009, three decades after they established the climate research programme, WCRP, and the IPCC process, and two decades after the establishment of coordinated observations (GCOS) and the policy framework (UNFCC). The theme of WCC-3 is "climate information and prediction for decision making" which will depend to a large extent on our ability to predict and project reliably the state of Earth's climate system on seasonal, decadal and longer time scales, but more importantly to synthesize our best scientific knowledge about the climate variability and change and make it available to decision makers in a timely and effective manner. This implies establishing a climate information development and dissemination system that captures the outcome of climate observations, research, analyses and assessments in an end-to-end and seamless manner to serve effectively the providers and users of such information. This is indeed a multi-generations' challenge and opportunity that WCRP is pleased and privileged to embrace and support through its network of global partnerships with the national and international climate research programmes, and the network of its scientific experts from more than 190 countries around the world.

Ghassem R. Asrar
Director, World Climate Research Programme

Report on the DOE/BERAC Workshop

Identifying Outstanding Grand Challenges in Climate Change Research: Guiding DOE's Strategic Planning

September 5, 2008

Preface

Efforts to reduce greenhouse gas emissions, the bulk of which are from energy-related activities, will add to the challenges facing the nation and the world in meeting the energy needs of the 21st Century. As Secretary of Energy Samuel Bodman recently stated, "By 2030, global energy consumption is expected to grow by over 50 percent. . . . . U.S. electricity demand is projected to increase by about 50 percent by 2030, with global demand nearly doubling. . . . We must recognize the realities of global climate change and work to develop cleaner sources of energy that at the very least do not worsen—and hopefully can improve—the health of the environment."

Much has been learned about climate from research conducted by the Department of Energy (DOE), other agencies under the Climate Change Science Program (CCSP) and its predecessors, and by the international scientific community. Current understanding about climate forcing, climate response, and consequences of climate change is outlined in recent reports by the Intergovernmental Panel on Climate Change (IPCC).

In recognition of the need to develop further research directions, Dr. Raymond L. Orbach, DOE Under Secretary for Science, charged DOE's Biological and Environmental Research Advisory Committee with conducting a workshop identifying the "Outstanding Grand Challenges in Climate Change Research."

In response to this charge, a workshop was held in Crystal City, Arlington, Virginia, March 25-27, 2008, co-chaired by Drs. Robert E. Dickinson and Gerald A. Meehl. Workshop participants included many of the leading U.S. experts in climate change science from academia, other federal agencies, and DOE national laboratories. The workshop consisted of plenary lectures and breakout sessions on several Grand Challenge areas delineated by Dr. Orbach. Participants in the breakout sessions discussed the key challenges in three areas and the research, observational, and computational capabilities needed to meet the identified Grand Challenges.

The outstanding Grand Challenges in climate research proposed by this workshop should guide DOE in its strategic planning activities to meet the energy needs of the nation, while minimizing environmental damage caused by use of this energy.

Executive Summary

The workshop identified three Outstanding Grand Challenges in climate change science. These challenges are consistent with research priorities identified by other national and international bodies after the IPCC Fourth Assessment Report (AR4). Meeting them would provide a major advance in the Nation's capacity to address the climate issue.

1. *Characterize the Earth's current climate, and its evolution over the last century to its present state*.

The ability to predict future climate accurately requires a clear understanding of what has caused climate change over the recent past. The simulation of this observable time period contributes an important validation of the models used to project into the future and provides a better description of the current state that also benefits future projections. Advancing our understanding will require novel approaches to atmospheric reanalysis, combining models and observations to characterize changes to the Earth System to account for cloud and aerosol interactions and cloud feedbacks. These new approaches need high resolution models, depending on computational and computing advances, to greatly extend the past methodology of reanalyses to assimilate many further sets of measurements about the physical state of the atmosphere, ocean, ice-sheet, and land systems into comprehensive data sets. Such datasets provide a fundamental basis for evaluating climate simulations and learning more about the mechanisms and processes that characterize the global coupled climate system.

***2.* Predict regional climate change for the next several decades.**

Predictions over decades at fine enough spatial resolutions are needed to support the decisions being made about adapting to future climate change. Such predictions require the use of high-resolution climate models and improved observations of the current climate to begin model integrations (*i.e.*, Challenge 1). The decadal evolution of the climate state includes climate change to which we already are committed, the change from natural modes of variability, and additional changes resulting from future greenhouse gases, aerosols, and land cover. Providing such predictions on the decadal time scale (*e.g.*, for the next two to three decades) and with regional fidelity is a new challenge for the climate science community.

3. Simulate Earth System changes and their consequences over centuries.

Current climate model simulations summarized in the IPCC report do not adequately include feedbacks that contribute to overall Earth

System changes on century-long time scales (*e.g.*, carbon-cycle feedback). Grand Challenge 3 is to develop and integrate important long-term components of the Earth System, such as large ice sheets and the natural and managed components of the carbon cycle. Earth System models would be developed and used to address the character and magnitude of climate changes for the next century and beyond, when large climate change and subsequent forcing feedbacks are expected. This information should allow decision makers to formulate "midcourse corrections" as to the optimum mitigation and adaptation measures needed. Grand Challenge 3 relates directly to the economic, technological, and environmental choices that future generations may need to make in response to climate consequences.

The workshop identified the following research initiatives that could extend areas in which DOE has particular program strengths to meet the Grand Challenges:

a. Characterization of impacts of radiatively active atmospheric constituents, especially aerosols and clouds, on climate and air quality through their interactions with precipitation, moist convection, and atmospheric chemical processes, as these change with climate change.

b. Interactions between ecosystem processes and changes and the climate system.

c. Determination of interactions between changing climate, hydrological systems, and their management.

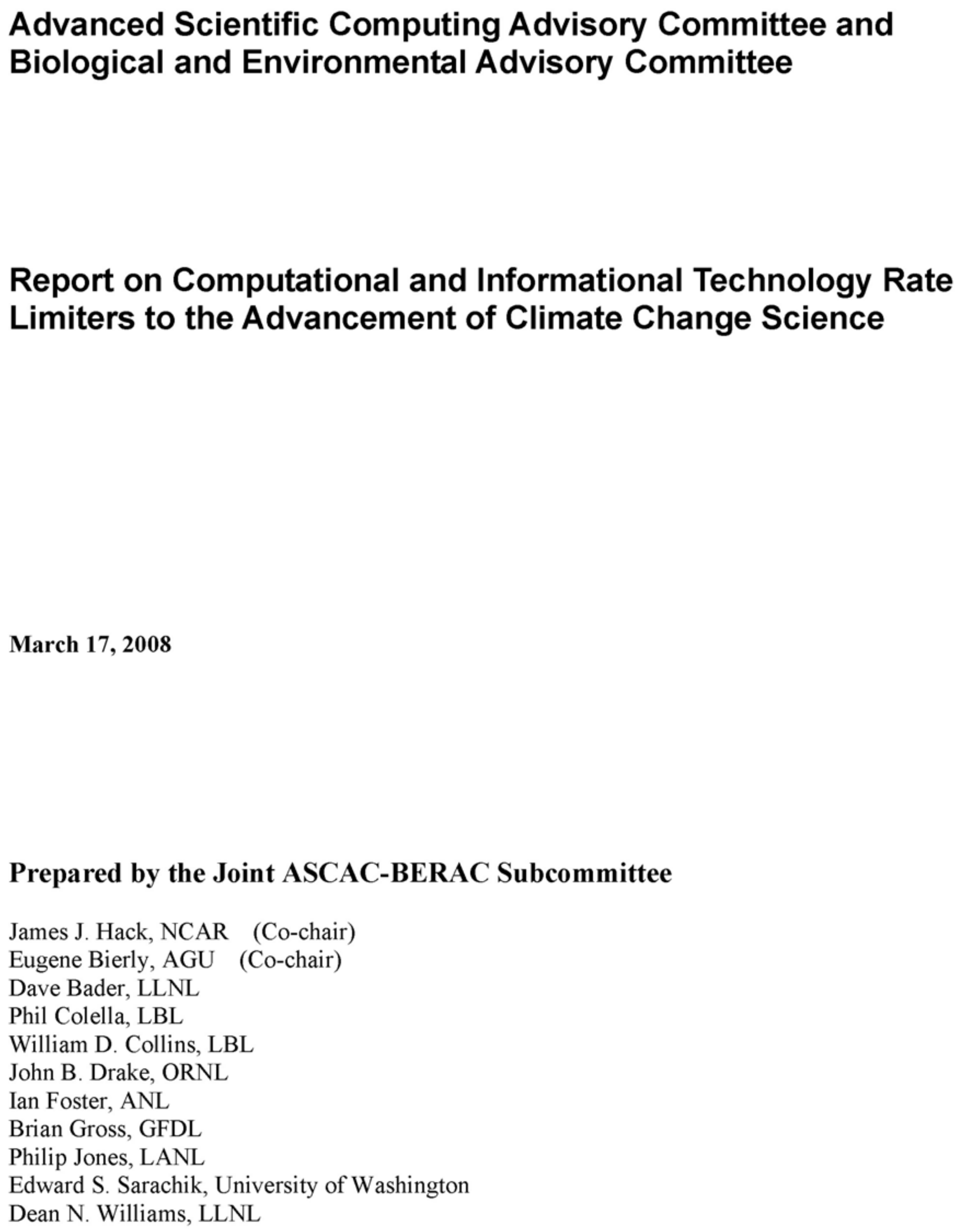
d. Incorporation of knowledge gained from observational and modeling process studies into multiple generations of Earth System models; such models will contain more complexity, improved parameterizations, and better initialization procedures that are validated through observational experimental and focused modeling activities.

e. Determination of the implications of climate change for energy systems, including a focus on supporting strategies for mitigation actions that affect climate change and influencing selection of adaptation choices.

The workshop demonstrated that although climate change science intersects a broad range of physical, biological, and social sciences, the science community concurs on the major cross-disciplinary issues that need to be addressed through enhanced research. Individual research conducted along disciplinary lines is a requisite component of the challenges, but the workshop stressed that such research must now connect across disciplinary boundaries in a more integrated approach and more explicitly include human dimension components to address the

Grand Challenges involving the entire global Earth System.

Incorporation of the advances outlined in this report will require accelerated investments in new observational and experimental research for developing complex models, but also investments in highly sophisticated software and in the nation's most advanced computational capabilities, and in the training of new scientists to develop and use these capabilities. Meeting these challenges will require an energized research community with strong contributions from new generations of young scientists. This effort will require tackling many science issues not in isolation but in the context of how each contributes to modeling and understanding of the overall system.



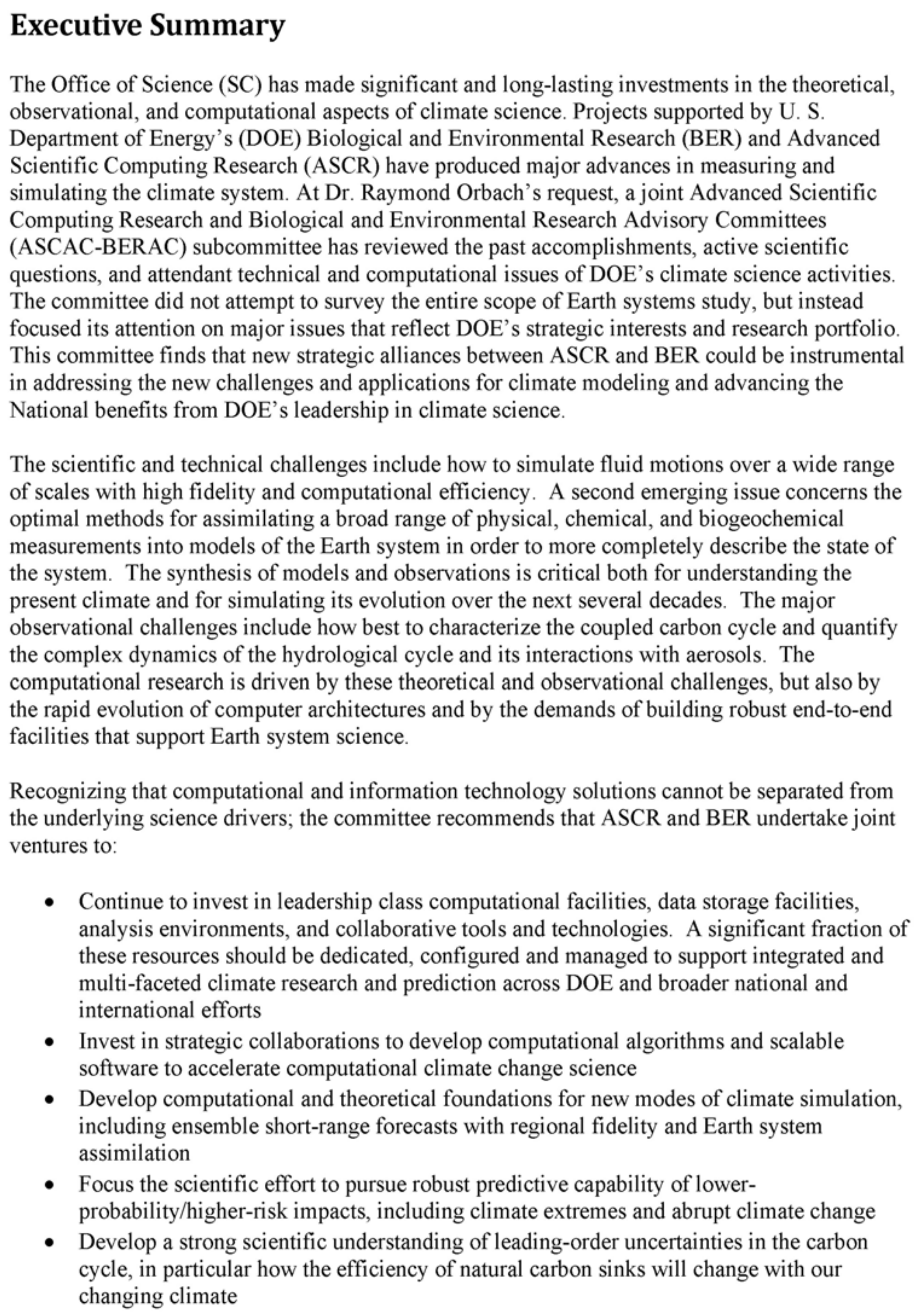
Advanced Scientific Computing Advisory Committee and Biological and Environmental Advisory Committee

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



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

Prepared for the U.S. Department of Energy under Contract DE-AC05-76RLO1830

Production support provided by Pacific Northwest National Laboratory,
Fundamental & Computational Sciences Directorate

PNNL-18362