Written Testimony of

James J. Hack, Ph. D.

Director, National Center for Computational Sciences
Oak Ridge National Laboratory

Before the
Committee on Commerce, Science, and Transportation
United States Senate

8 May 2008
Hearing on *Improving the Capacity of U.S. Climate Modeling for Decision-Makers and other End-Users*

I thank Chairman Inouye, Vice Chairman Stevens, and the other Members of the Committee for the opportunity to speak with you today on ways to improve the capacity of U.S. climate modeling for decision-makers and other end-users. My testimony draws on over two decades of developing global models of the climate system at the National Center for Atmospheric Research. My name is James J. Hack and I currently serve as director of the National Center for Computational Sciences (NCCS) at the Oak Ridge National Laboratory (ORNL). The ORNL NCCS provides the most powerful computing resources in the world for open scientific research. It is one of the world's premier computational science research environments supporting advances in our understanding of the physical world and using that knowledge to address our most pressing national and international concerns. My role as NCCS Director provides a unique perspective on how the application of leadership-class computing technology in a computational science partnership with scientific investigators can radically accelerate basic progress for a variety of extremely demanding scientific domains. Examples of NCCS research focus areas are the simulation of complex biomolecular systems with applications to pharmaceuticals as well as more efficient biofuel generation, simulations that investigate the fundamental properties of materials, such as high temperature superconductors, and simulations exploring the processes that maintain and regulate Earth's global climate system.

There are many scientific and technical challenges related to monitoring, understanding, predicting and adapting to climate change, especially on local and regional scales. Observations of the entire Earth, for instance, are the foundation for improved understanding of climate change and for computer models that accurately predict weather and climate. A newly emerging issue is the development of optimal methods for assimilating this broad range of physical, chemical, and biogeochemical observations into models of the Earth system in order to more completely describe the current state of the system. This is but one example of how the synthesis of models and observations is critical both for understanding the present climate and for simulating its evolution over the next several decades. Computational research associated with the modeling and prediction of Earth's climate system includes developing methods for simulating complex multiphase fluid motions over a wide range of scales with high fidelity and with high computational efficiency, as well as by the need to continually incorporate new theoretical and observational knowledge into global models. The rapid evolution of computer architectures creates its own challenge to fielding stable computational environments that support Earth system science.

State-of-the-art climate models, such as those developed by NSF, NOAA, DOE Office of Science, and NASA programs embody our best understanding of the physical and biogeochemical processes that are central to the climate system. The goal of such modeling efforts is to accurately represent the collective behavior of these climate processes as an interactive system. These models are continually developed, tested, and evaluated against observations. Although they are the best available tools for exploring how the climate system works, they are not perfect. Uncertainties arise from shortcomings in our scientific understanding of the climate system, and in identifying the best mathematical approaches for representing those processes we do understand in numerical models.

Despite these imperfections, climate models are still able to reproduce the climate of the past, which gives considerable confidence in their ability to simulate changes in future climate. For instance, climate modelers are able to test the role of various forcings in producing observed changes in climate over the past century. Such simulations have now reliably shown that global surface warming of recent decades is a response to the increased concentrations of greenhouse gases in the atmosphere. They are also remarkably consistent in their projections of continued warming of the climate system for the remainder of this century, as discussed in the Fourth Assessment Report (AR4) of the United Nations Intergovernmental Panel on Climate Change (IPCC). The release of the IPCC AR4 report, along with release of a series of Climate Change Science Program (CCSP) reports, signal that the detection and attribution of climate change at global scales has essentially been resolved. The global community is now faced with a new set of urgent problems relating climate change to human health, water resources, food supplies, changing risks of forests to fires and insect disease, and threats to managed and natural ecosystems. Central to these problems is the demand for much more regional detail on climate change on the time scales of resource and infrastructure planning. In order to address these issues, the community needs to develop and undertake a coordinated research program balanced and integrated among observation, theory and computation. Meeting future challenges in climate change science will require qualitatively different levels of scientific understanding, modeling capabilities, and computational infrastructure than are currently available to the climate science community. Many of society's questions will require the development of a new generation of more comprehensive climate models, frequently referred to as Earth

System Models (ESMs) that predict the coupled chemical, biogeochemical, and physical evolution of the climate system. These models will also need to be exercised at unprecedented high resolution. The needed increases in complexity and resolution will require transformational changes in computational capability.

Over the last 30 years, modeling capabilities have advanced considerably in their treatment of complexity, and the ability to treat ever finer scales of motion. Modern atmospheric models represent the observed equator to pole energy transport much more realistically than did earlier model generations. They also do a much better job of representing many detailed features of the observed mean climate state. These improvements have meant that global climate models are now routinely run with fully-interacting atmosphere, ocean, land surface, and sea ice components. These more realistic and complex models can now not only simulate observed changes over the past century in global mean climate, but also climate variability and change on continental scales. This includes the attribution of many of the observed large-scale changes in indicators of climate extremes consistent with a warming climate, such as the annual number of frost days, warm and cold days, and warm and cold nights. Models that contributed simulation results to the IPCC AR4 also generally agree that regions like the subtropics will dry, including the US Southwest, while polar latitudes will receive more precipitation related to large changes and shifts in the extratropical storm tracks.

On finer spatial scales, however, state-of-the-art climate models don't always agree on projected climate change impacts, either on decadal or longer time scales. It is also not clear that they can accurately project changes in extreme events, or can reliably simulate changes in low-frequency climate variability or the likelihood of abrupt change. Near-term investments in the climate science enterprise could lead to a significant quantitative improvement in the scientific community's ability to address these difficult but societally relevant questions, leading to improved guidance to policymakers and stakeholders charged with developing strategies for adapting to climate change.

One immediate scientific challenge and opportunity is the incorporation of chemical and biogeochemical processes in climate models. The science surrounding the chemical and biogeochemical coupling of climate has become central to answering climate change questions, particularly those associated with the global carbon cycle.

Addressing the science issues will require new observations and methods of analysis, new theoretical understanding, and new models of the Earth system that include the interactions between human and natural systems. These models will play pivotal roles in interpreting the paleoclimate records, in synthesizing and integrating observational measurements to study the current carbon cycle, and in projecting the future responses of human society and the natural world to evolving climate regimes.

Another example of a pressing scientific challenge is the rate of sea level rise and the impact of that rise on coastal communities. Recent observations indicate ice sheets can dissipate on much more rapid timescales than from melting alone due to dynamical processes in large outlet glaciers and ice streams within the ice sheet. Faster disintegration of the ice sheets will contribute to faster sea level rise and will pose a greater risk of abrupt changes in the climate system. Abrupt climate change can also result from thresholds and nonlinearities in the response of climate to slower time scale forcing of the climate system. Examples include rapid changes in ocean circulation, large scale vegetation mortality and succession, release of methane frozen in ocean and permafrost, and megadroughts. The climate community will need to use models to identify thresholds of forcing in the climate system and explore the likelihood and impacts of such scenarios. The community's efforts to advance climate modeling and its application to science and technology options for mitigation and adaptation will require advances in essentially every aspect of the models' theoretical, observational, and computational foundation. Quantifying uncertainties in predictions will require a new level of integration between modeling and observational science. New mathematical methods and algorithmic techniques will also be required to address the fundamental challenges of multi-scale coupling of physical, dynamical, chemical and biogeochemical processes. A flexible leadership-class computing infrastructure has been and will continue to be a key factor in making these advances possible.

As mentioned earlier, today's climate models are in strong agreement that global and continental-scale temperatures will continue to rise as a result of human activities. However, it is also important to improve our understanding of the likely changes in regional climate over the next few decades. Climate forecasts on decadal time scales are governed primarily by the history of the ocean circulation and the current atmospheric forcing. Therefore, climate forecasts on these time scales will require retrospective analyses of the global oceans to be able to accurately initialize the forecasts. The ocean is responsible for much of the inertia or near-term "memory" in the climate system. The development of ocean data assimilation techniques, largely an applied mathematics and algorithmic challenge, will be necessary to provide an initial ocean state for decadal prediction and represents a pacing item for seasonal, inter-annual, and decadal prediction. While assimilation has been extensively developed and used in the weather community, the climate community will need to evaluate which assimilation methodology is best suited for climate simulation and the creation of realistic initial states for climate change scenarios. Optimal interpolation and simple methods have so far been adequate for the ocean due to sparseness of data, particularly for salinity and for ocean properties at depths below 1000m. With the influx of new ocean data sets, advanced techniques will need to be examined. Recent progress in deploying large numbers of floats and the launch of new satellites that together will measure salinity profiles will greatly improve our ability to effectively constrain ocean models with assimilation. For example, assimilation of data from ARGO floats with a fully coupled climate model has shown great promise in determining the state of the climate system, although the assimilation process is extremely computationally demanding.

Accurate projections of changes in the frequency of climate extremes at relatively high geographic and temporal resolution will be essential for the development of robust adaptation strategies. However, current climate models have been designed primarily to predict patterns of change at a coarser level. Much more research is required to understand how increasing model resolution and employing increasingly sophisticated parameterized treatments of non-resolvable processes may affect the ability of models to more accurately simulate changes in local extremes. In particular, the relationships between extreme statistics and synoptic-scale low-frequency variability are not understood.

A better understanding of low-frequency variability is critical for the detection of climate-change signals. For Earth system modeling, it is important to characterize the natural modes of coupled variability in the carbon cycle, terrestrial ecosystems, and dynamic vegetation. It is also important to develop a better understanding of external forcing mechanisms, such as the role of solar variability in the broader context of the

Sun-Earth system. Current understanding of these complex systems is limited by the length of the observational record. The wide dynamic range in the relevant space and time scales further complicates the coupling issues. New mathematical methods designed for multiscale systems hold promise for this class of problems, and these methods should be explored for efficient implementation in next generation models.

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

Multiscale interactions also complicate treatment of the climate system. As with the broader issues of climate variability, process-level understanding of things like the water cycle is limited by the lack of basic observations. While the absence of these data still represents a barrier to progress, near-term enhancements in computational capacity would permit the resolution of fundamental phenomena at the process level. Targeted investments in observational programs can provide much of the necessary data to validate high-resolution process modeling studies of critical topics like aerosol-cloud interactions, central to the climate model sensitivities that lead to discrepancies in projections of future climate on century-long time scales.

Finally, there are significant software and computational hardware infrastructure challenges pacing progress in climate science. Many scientists have found the growing requirements to support the software on high performance computers as a distraction from the central scientific goals of improving climate models and answering fundamental questions about climate feedbacks and variability. This drawback is offset by the new scientific opportunities provided by dramatic increases in computational power. This becomes an issue of scientific productivity. What is needed is a software framework that not only scales from desktop to petascale, but also that supports multi-scale model development and process integration. As a closer connection with observational data and process studies is required to advance the science of regional climate prediction, the software must also become more closely integrated and supported across scales. A flexible and powerful software development environment will increasingly be required to support data assimilation and other data intensive frameworks. The limitations of existing software environments have emerged as key bottlenecks to progress where near-term investment would have important scientific payoffs.

There are real opportunities to invest in climate change science to improve the utility of global models for decision makers and the broader end user community. High impact opportunities for investment include computational facilities, theoretical efforts associated with model development, targeted observational programs and the development of novel computational algorithms. Investments in modeling will accelerate progress on improving the predictive skill of global climate models. The climate community needs to develop a new generation of Earth System Models based upon new and expansive requirements including the ability to more accurately reproduce major modes of natural variability, incorporating functionality for decadal-scale ensemble forecasts at very high spatial resolution, the flexibility to incorporate new data on the physical, chemical, and ecological climate system in the form of process representation (thereby increasing the fidelity of climate simulations), stronger connectivity with user communities for exploring adaptation and mitigation strategies, and the capability for two-way interactions among emissions, impacts, adaptation, and mitigation.

Modeling over a large range of time scales to fully evaluate the couplings between biogeochemical cycles, chemistry, and ecology will present a significant computational challenge. The growth requirement of characteristic applications of climate change prediction models already more than doubles every year. High-resolution ocean circulation studies and cloud system resolving atmospheric simulations are already pushing the limits of petaFLOP systems that utilize many tens of thousands of processors. As regional climate prediction on decadal to century time scales becomes more important, the required computational power will approach the exaFLOP scale (one quintillion floating point operations per second) that will utilize 100K – 1M processors. This will require a continued focus on fielding state-of-the-art leadership class computing facilities so that computational capability does not become a more critical pacing factor. Ancillary investments in software, networking, data storage, collaborative tool, and visualization technologies are necessary for balance. For example, climate science is both distributed and collaborative. As interest in climate science continues to grow and its scope broadens to encompass issues of ecosystem and economic impacts, and the evaluation of mitigation and adaptation strategies, the number of participants will also increase. The overall productivity of researchers and the quality of the research output can likely be improved significantly by the use of advanced collaboration technologies that distribute applications and data across the network. It is easy to project that climate research demands on networks will grow yet further as data volumes increase. With a growing number of participants in the climate science enterprise, and a growing diversity and volume of climate data, the need for new data and network resource management strategies and technologies will emerge. Modern visualization capabilities can also play an important role in the discovery of new scientific results and in the communication of the science to a broader community of stakeholders. For an area like climate modeling this is particularly important because of the societal relevance of our results to policy makers and those concerned about the consequences of climate change.

New observational programs and data assimilation systems represent opportunities to improve our understanding of a variety of physical, chemical, biogeochemical, and ecological processes, reducing key uncertainties in modeling assumptions. Meteorological and oceanic analyses have become important tools for studying the mean state and variability of the current physical climate. These analyses are constructed using a model that is adjusted by incorporating observations during its numerical integration. These analyses have proved particularly useful for understanding the relationship between observations and the underlying dynamics of the climate system. It would be especially valuable to have a comparable analysis of biogeochemical and chemical cycles that could relate local and global biogeochemical processes to more completely describe the state of the global system. However, there are no existing analyses that encompass the physical, chemical, and biogeochemical processes in the climate system. Development of these analyses will require significant investment in assimilation systems for chemical and biogeochemical observations from in situ and satellite platforms. Much more advanced models will be required to understand the fidelity of the analysis system, which will further push the sophistication of global modeling activities.

Investments in computational algorithms will increase scientific productivity using leadership-class computers for climate change simulation studies and improved simulation accuracy. There is a broad class of mathematical and numerical algorithms that are ready to be explored for application to the climate problem. For example, there are strong arguments for exploiting higher-resolution variable gridding configurations for the atmospheric component of a climate model. The computational demands of uniform ultra-high resolution configuration of a global atmospheric model would outstrip existing computational capability. An intermediate practical approach to dealing with resolution issues is to use a multi-resolution approach, such as nested refinement. These approaches will allow scientists to improve understanding of the multi-scale interactions in the climate system, to identify those of greatest importance, and to document their effects on climate. Ultimately, such research will help determine the best methods of including these multi-scale interactions in climate models, and it will help differentiate between those processes that can be better or newly parameterized versus those that cannot. Such techniques are already being explored by several research groups including the National Center for Atmospheric Research. With a nested or adaptive resolution approach the computational capability required could be reduced by an order of magnitude or more, and could make the goal of computing with such ultra-high resolution models more feasible. A final example of algorithm opportunities is the need to better characterize the uncertainty in simulation results. Ensembles and basic statistics are currently used to assess uncertainty due to natural internal variability intrinsic to the climate system. More formal methods for verification, validation and uncertainty quantification are needed from the computer science, mathematics and statistical science communities. A particular challenge is the sparse nature of observational data necessary to validate models.

The nation's climate modeling enterprise is likely to be increasingly driven by the need to obtain scientific results for a large and diverse group of users, including government officials, in a timely fashion. In such an environment, the development of innovative models, algorithms, and software must be managed as a project, as opposed to an open-ended research program. Some aspects of such an approach are well-understood, such as the need for planning, schedule visibility, and milestones. A more difficult problem is the potential dependence of success on delivering high-risk products in models, algorithms, and software on a particular schedule. Many of these products, such as new approximation methods, or new programming models, represent non-incremental departures from the current methods used in production climate models, but may be necessary to achieve National goals. Risk management in such a setting requires careful planning and a close and continuing collaboration between the climate, facilities, applied mathematics, and computer science communities. In addition to the research management needs, there will also be a need to ensure that end-users are sufficiently involved in the prioritization of research efforts, and that the resources and institutions exist to transfer the large volumes of information into the decision making processes of various private and governmental users.

In conclusion, there is no single pacing item to the advancement of climate change science, but a collection of interrelated science and technology challenges. Many of the issues discussed in this testimony speak to the need for a balanced investment portfolio in computational infrastructure, climate science, computer science, and applied mathematics. In the short and long term, computational capability remains a significant bottleneck and should remain a high priority investment. But as the science and complexity of climate simulation grows, so will new technical and scientific challenges. Immediate proactive investments in climate science, software, algorithms, data management, and other pacing items are needed for accelerated progress that can keep pace with the rapidly evolving computational environment. The management of these investments is also critical to success. Strategic management of such a broad multidisciplinary activity will likely prove to be the most effective way to ensure that new investments have the desired impact on accelerating progress.