Scalable and Extensible Earth System Models for Climate Change Science

SciDAC progress report April 2007

John Drake and Phil Jones

I. Introduction: Plans and Project Overview

In this, the first progress report of our SciDAC2 project, we can record that the project team was able to hit the ground running. During our kick off meeting in Boulder, October 2006, several key decisions were made. Rather than organize ourselves in strict alignment with the CCSM Working Groups, we have chosen a more integrative and crosscutting strategy that still engages collaboratively with the working groups. Topics of immediate priority were chosen so that schedules for CCSM4 development could be met. The issue of parallel scalability and the ability to run effectively on thousands of processors is an important early emphasis of our Scalability and Performance task. The task is undertaken with a Scientific Application Partnership (SAP) focusing on the IBM BlueGene and Cray XT4 computing platforms. We report on these efforts in this progress report along with the Biogeochemistry task and the Carbon Land Atmosphere Modeling Project (C-LAMP). These are emphasized early in the project because they influence key decisions on the configuration of the emerging CCSM4 and its configuration as an Earth System model. As of this date, a CCSM3.5 model is being assembled as an intermediate development. We believe that both these efforts have contributed to the likelihood of a higher resolution and more comprehensive model for climate change science that will utilize petascale computers.

Other areas of focus in this project are somewhat longer term requiring significant modeling work and scientific research. These include more comprehensive treatment of aerosol radiative effects, the integration of new components such as an ice sheet model, and evaluation of new methods within the modeling framework. Though considerable progress has been made in these topics over the reporting period, they will be highlighted in future progress reports.

II. Scalability and Performance

Responsibility for improving performance scalability is shared between the core SciDAC project and the SAP in performance scalability, with the core project focusing on software engineering and near term performance goals and the SAP focusing on longer term algorithmic changes. Activities to date have been in the following three areas: improving our ability to measure of CCSM performance by introducing performance analysis infrastructure; performance scalability analysis and optimization on the Cray XT3 and XT4, and IBM BG/L; and improvements in performance scalability.

Much of the work on improving scalability was less on performance than on simply enabling running larger problem instances or running standard resolutions on systems with small memory per processor. For example, the number of undecomposed global arrays in the land model was decreased from the hundreds to less than 20, with the goal of reducing this to zero or one. This is a requirement for running at scale, on any system, but in particular on the BG/L system. Reducing memory footprint further (in all of the CCSM components) depends on the introduction of parallel I/O, and work in this area is ongoing in collaboration with CGD and CISL at NCAR (see section below). One recent accomplishment in performance scalability was the introduction of an ability to run the CAM physics with more MPI processes than the spectral dynamics, providing performance on MPI-only systems similar to that available using OpenMP on cluster of shared memory node architectures. (See Fig. 1.) Even more flexibility is being introduced into the finite-volume (FV) dycore, allowing different numbers of MPI processes to be used within different phases of the dycore and for tracer advection (in preparation for the introduction of many more tracers needed in a chemical atmosphere). In another approach to improving the performance scalability of the atmosphere model, the physics/dynamics interface has been generalized to support dycores that do not use longitude/latitude grids.

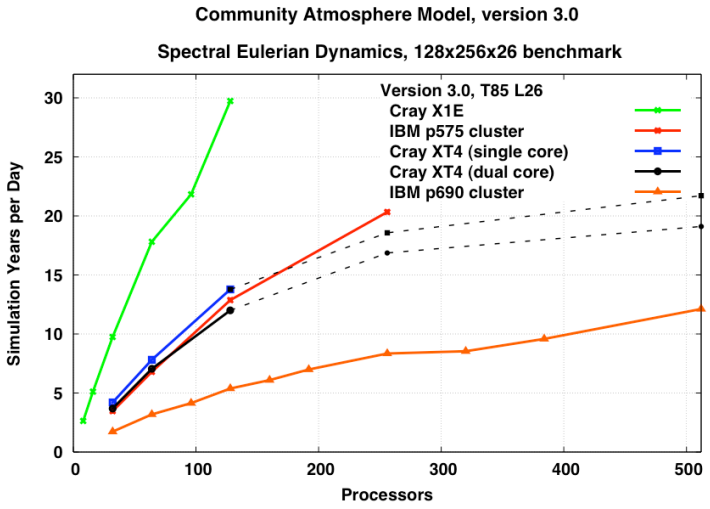

Figure 1. *CAM performance as a function of processor count on several architectures.*

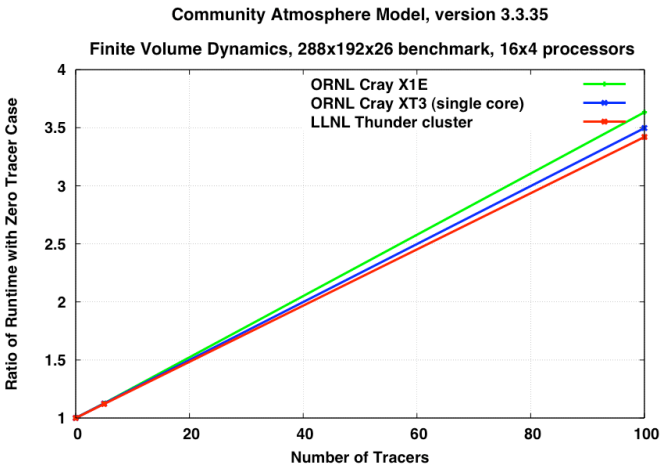

Figure 2. *CAM FV Scaling with tracer count.*

We ported and maintained CCSM and CCSM component models, studied and analyzed performance, and optimized performance on the Cray XT4 and the IBM BG/L. On the XT4 we focused both on production resolutions (T85 atmosphere and 1 degree ocean) and higher 'reach' resolutions (0.5x0.625 degree atmosphere and 0.1 degree ocean). A number of XT4 performance peculiarities were identified (primarily with regard to tuning I/O and tuning

MPI when using both cores) that are motivating the introduction of new performance tuning options into the CCSM. We also studied the performance impact of increasing the number of tracers, as would be required in a chemical atmosphere, comparing with similar data collected on the Cray X1E and the Thunder cluster at LLNL. (See Figure 2.) On BG/L, production resolutions of the full CCSM were run using up to 512 processors. Additional reductions in memory requirements are the primary focus for further improving both problem size and processor count scalability on BG/L. In collaboration with both NASA and NOAA, we also began to investigate the scalability of S-J Lin's latest FV dycore, which utilizes a cubed sphere grid similar to that used by HOMME.

Scalability of the ocean component (POP) has been improved through a collaboration with J. Dennis (NCAR). A new load-balanced distribution scheme was introduced as well as a new data structure for the barotropic solver. These changes improved scalability in displaced-pole grid configurations, but were not tested with tripole grid configurations. Testing of these changes with tripole grids is currently in progress.

IIa. ***Parallel I/O***

SciDAC researchers, in collaboration with NCAR software engineers and computer scientists, have developed a parallel I/O library which provides a generic interface to Parallel Netcdf, direct MPI-IO and serial netcdf. Parallel I/O is necessary to run on low-memory platforms such as BlueGene, a major focus of the performance SAP. In the first few months of this project, the prototype PIO library has been incorporated in to HOMME where it performed 2-3 times faster than the previous implementation of Pnetcdf. SciDAC researchers also developed the serial netcdf "fallback" feature of PIO which avoids the need to hold all data in memory in one node before writing to disk. PIO will be used in upcoming $1/10^{th}$ degree simulations with POP as part of BlueGene "Watson Days".

IIb. ***Cubed Sphere Dynamical Methods and Spectral Element Model Development***

Scalability has motivated us to evaluate the use of cubed-sphere dycores in CAM, while the emphasis on carbon cycle modeling and the associated chemistry requires a dycore with a locally conservative, positive-definite advection algorithm. Such dycores on the cubed-sphere will soon be available, (such as HOMME-DG and cubed-sphere FV) but in the mean time we have been evaluating CAM using the HOMME-SE (spectral element) dycore. Initial runs of CAM-HOMME with this dycore showed that the dissipation mechanisms in HOMME-SE, were far too viscous when there was a strong enstrophy cascade typical of realistic climate simulations. To remedy this problem, we implemented a hyper-viscosity term which replaces the conventional viscosity and element based filter used in HOMME (motivated by the fact that hyper-viscosity has been used in CAM since its inception as CCM0). This work required the development of a weak-form-divergence operator, which we also used in the continuity equation to achieve *exact, local* conservation of mass and semi-discrete conservation of total energy (exact with exact time integration). Both of these improvements have been tested in the shallow water equations and we are currently evaluating their performance in CAM-HOMME using the aqua planet model problem. The shallow water results are shown in the following figures (vortex breaking problem from Polvani et. al., JAS 1996). We

show the reference solution (shallow water version of CAM Eulerian) and two spectral element solutions. All the simulations have the same number of points around the equator. The spectral element solution shows that with hyper-viscosity we achieve results equivalent to the CAM solution, while with conventional viscosity, at the level necessary for a stable solution, is far too viscous.

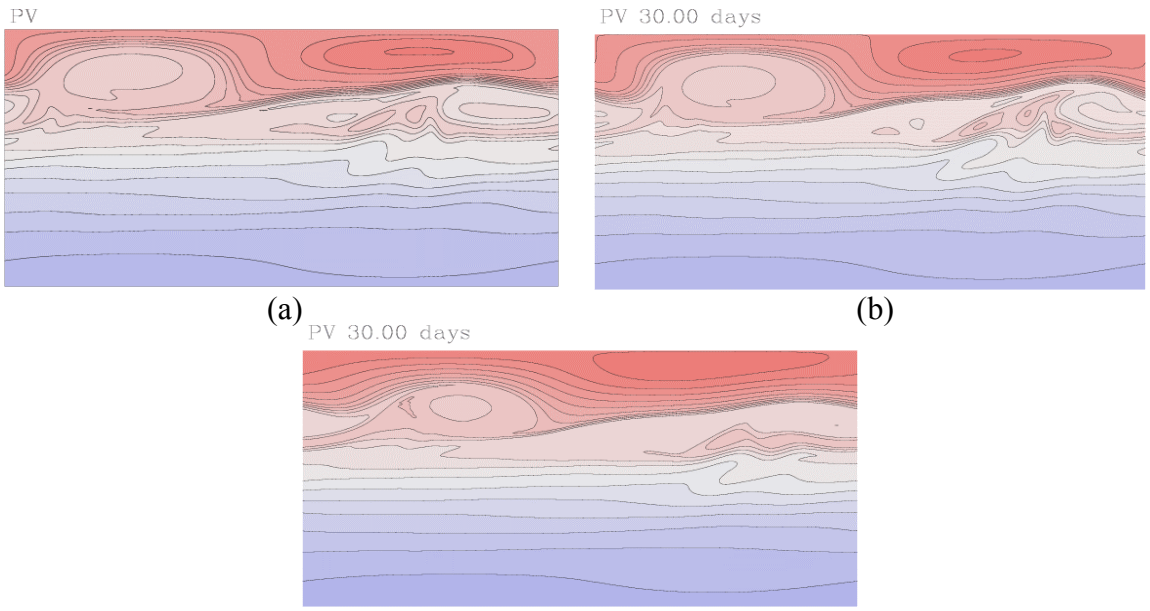

Figure 3. *Potential vorticity after 30 days in lat/lon projection for (a) T85 spectral dycore, (b) spectral elements with hyperviscosity, and (c) spectral elements with conventional viscosity.*

IIc. *Sequential CCSM*

During this reporting period, SciDAC researchers in both the Performance SAP and the main project have contributed to the development of a sequential version of CCSM. Work on sequential CCSM covers two areas of this project: Performance and Scalability and Integration and Evaluation. For the former, a sequential CCSM eliminates many load imbalances seen in the concurrent version. For the later, a sequential CCSM has a simplified data flow which makes it easier to add new components. The new land ice model developed under this program will first be coupled to the sequential CCSM.

Development of the sequential CCSM began prior to this SciDAC. In this reporting period, SciDAC researches have extended the sequential version of CCSM so that it can now accommodate different resolutions. The driver can determine at runtime if the resolutions between the atmosphere/land and ocean/sea ice are different and initialize appropriate data structures. Sequential CCSM uses the same SCRIP mapping weights as the concurrent version. We have also added the ability to modify the mapping weights according to dynamically changing sea-ice fractions. A new low-memory routine which reads in SCRIP mapping weights from files was created and is now being used by both the concurrent and sequential versions of CCSM. We have configured a version of

sequential CCSM that uses dead models for the components and have begun testing memory scaling of the sequential coupling infrastructure on BlueGene/L.

Another outcome of the sequential development is that the same top level driver that is used to run CAM is also being used to run the "standalone" version of CLM which uses a data atmosphere. This used to be a separate driver. Sequential development has already had a positive impact on code reuse.

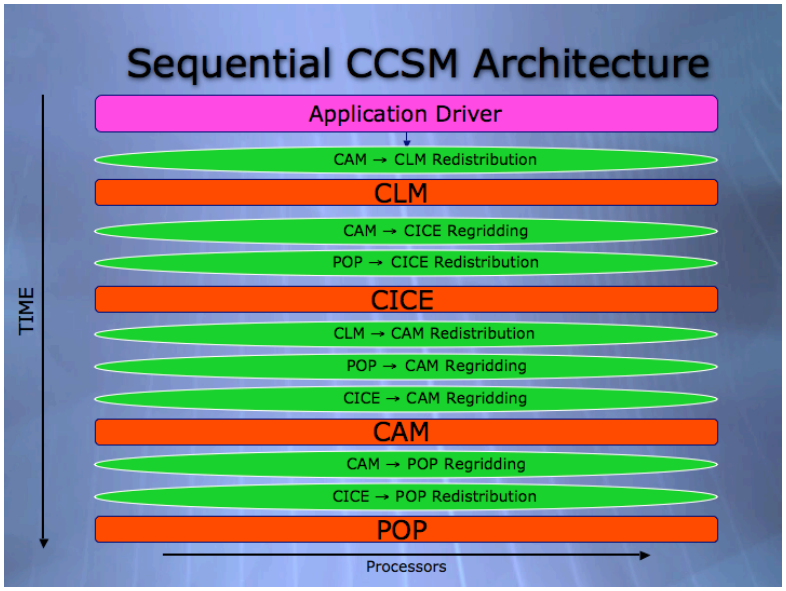

Figure 4. The sequential CCSM execution sequence. It is much easier to modify this execution sequence then the concurrent.

A new release of the Model Coupling Toolkit (MCT 2.3.0) in January of 2007 included support for a more flexible version of the sparse matrix vector multiply routine which allows the attributes to be specified. This reduces the need for duplicate copies of data used for coupling in the sequential case. MCT 2.3.0 also features reduced memory requirements for the initialization of MCT Router's and Rearrangers. Other MCT releases in Fall of 2006 included support for the XT3 cross compiler and the first public release of vector modifications completed at the end of the previous project.

III. Earth System Model

Progress towards a coupled Earth System Model continued with substantial work on the Carbon-Land Model Intercomparison Project (C-LAMP), in preparation for important decisions on processes to be included in future CCSM simulations. In addition, further work on chemistry and biogeochemistry continues in the atmosphere and ocean components in preparation for the next production simulation of a fully-coupled simulation with carbon and sulfur cycles.

IIIa. *Land biogeochemistry and C-LAMP*

In the area of Biogeochemistry, significant effort has been directed toward the Carbon-Land Model Intercomparison Project (C-LAMP). Simulations for Experiment 1 (Forcing with NCEP/NCAR Reanalysis) of C-LAMP were performed on the Cray X1E through the Computational Climate Science End-Station at ORNL. Experiments 1.1 (Spin-Up), 1.2 (Control), 1.3 (Varying climate), and 1.4 (Varying climate, CO2, and N deposition) were completed using the CLM3-CASA' and CLM3-CN models. Because of the unexpectedly long spin-up runs (about 4,000 years for each model), over 10,700 model years were completed resulting in over 9.0 TB of model output being stored in the High Performance Storage System (HPSS) at ORNL. Experiment 1.5 will be run once model code has been written and tested for performing prescribed land use change in the both models.

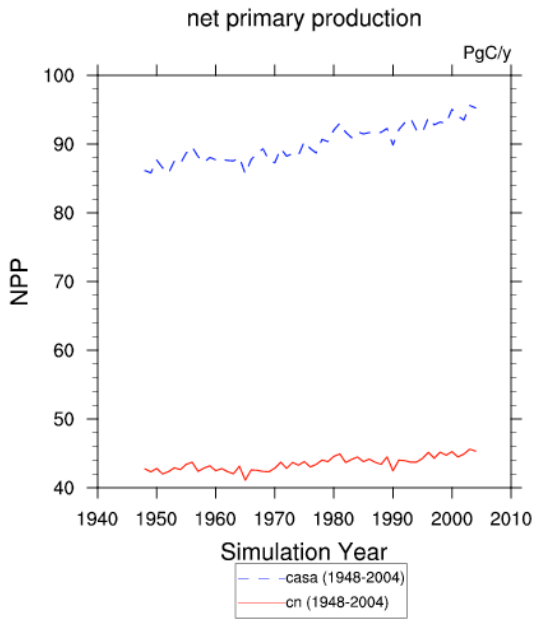

Figure 5. Net primary production for the CLM3-CASA' and CLM3-CN models.

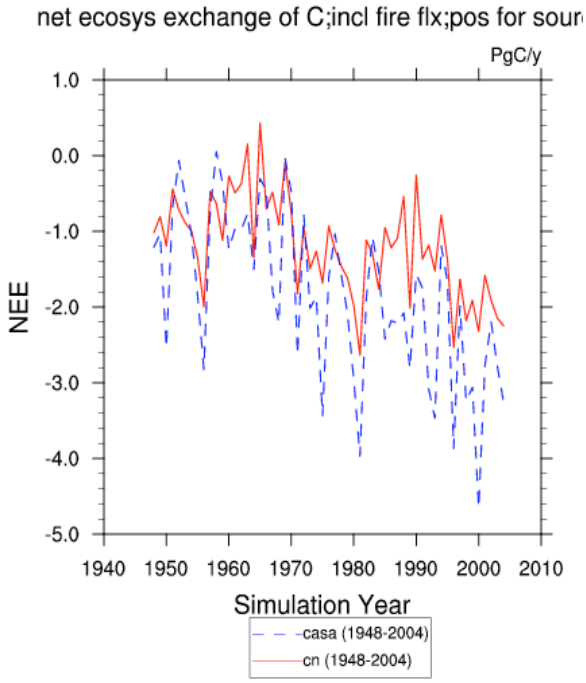

Figure 6. *Net exchange of carbon from CLM3-CASA' and CLM-CN models.*

A preliminary evaluation of the results of Experiment 1.4 was presented at the CCSM Biogeochemistry Working Group meeting on February 28. CLM3-CASA' has a higher net primary productivity than CLM3-CN (Fig. 5); the accepted global value being near the mean of the two models. CLM3-CASA' has a stronger fertilization response to increasing CO2 than CLM3-CN because of increasing nitrogen limitation in CLM3-CN (Fig. 6). Metric scores--resulting from comparisons to best-available datasets for net primary productivity, leaf area index, carbon stocks, and energy--show that both models exhibit similar scientific performance. Comparisons of global leaf area with MODIS observations are shown in Figures 7 (for CLM3-CASA') and 8 (for CLM3-CN). The simulations demonstrated productivity problems at high latitudes affecting both CLM3 models. Thornton at NCAR implemented fixes for these problems as well as fixes for CLM3-CN to solve leaf area distributions problems. All these changes have resulted in a "new" version CLM3-CN, which has been selected for inclusion in CCSM3.5. This unreleased version of CCSM is now being used at NCAR to

evaluate the fully coupled system leading toward CCSM4, the next anticipated release of the model that will be used for IPCC runs.

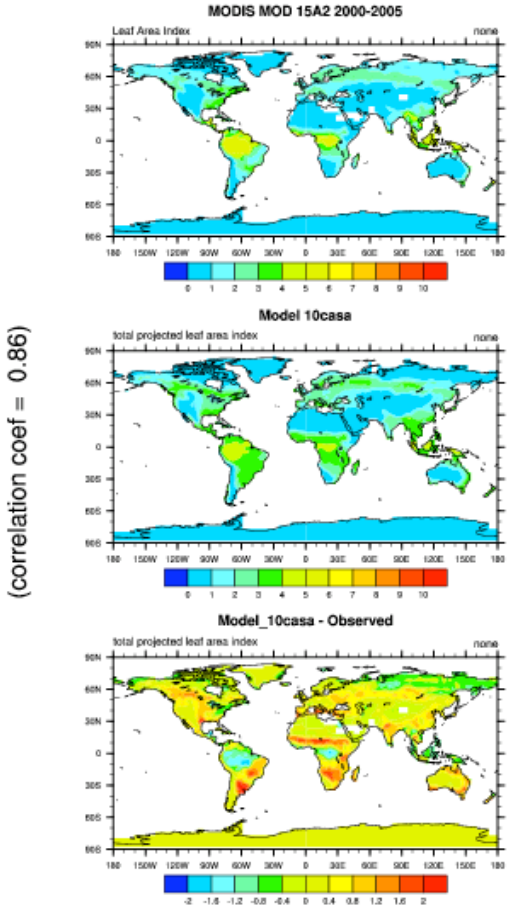

Figure 7. *Comparison of global leaf area with MODIS data for CLM3-CASA'.*

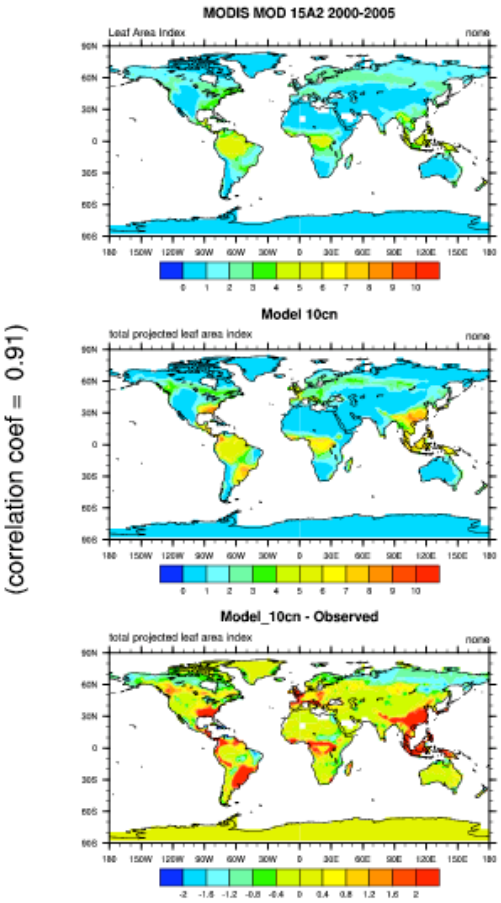

Figure 8. *Comparison of global leaf area with MODIS data for CLM3-CN.*

An initial evaluation of a long spin-up and control run of LSX-IBIS in the CCSM3 framework has been performed. This run exhibited as much as an 8 degree C warm bias in parts of the Northern Hemisphere, overlapping areas where a CLM3 control run had a 4 degree C cold bias. Dynamic vegetation was enabled in the run, and deciduous forests spread across most grasslands. Additional analysis and shorter runs with dynamic vegetation disabled will be performed to further test and tune the model. After tuning, LSX-IBIS will be run through the C-LAMP protocol; however, the bulk of the effort to date has focused on CLM3-CASA' and CLM3-CN to support model evaluation for CCSM Biogeochemistry Working Group recommendations for CCSM4.

Code was developed to post-process hourly output from the models to produce files of monthly means of desired fields by hour-of-day. These model results will be used at a

later time to evaluate the models' performance on the diurnal cycle. Both sets of output (the monthly means and diurnal monthly means) from the completed portions of Experiment 1 will soon be processed using CMOR to produce files for the Earth System Grid (ESG) to be distributed to the wider community. In addition, spin-up simulations for Experiment 2 (Coupled Land-Atmosphere) using CLM3-CASA' and CLM3-CN have begun on the Cray X1E.

The C-LAMP experimental protocol, draft metrics document, model diagnostics, and supporting information are available at http://www.climatemodeling.org/bgcmip/.

IIIb. ***Methane and ozone chemistry***

Over the reporting period, we have continued our analysis of the simulated tropospheric and stratospheric chemistry in our version of CAM3 with a representation of reduced hydrocarbon chemistry and stratospheric chemistry and processes (namely polar stratospheric clouds and stratospheric aerosols), without changing the vertical extent of the model. In collaboration with the WACCM group, we have performed a set of simulations 1970-2005 with observed sea-surface temperatures and time-varying specified greenhouse gas and ozone-destroying agent surface concentrations. In particular, we have performed a simulation in which the surface methane concentration was kept at its 1970 level. The comparison of the base simulation (red curves in Fig. 9, as two realizations have been simulated) with this sensitivity experiment (green curve in Fig1) indicates that methane has a very significant role (at the global scale) on the stratospheric water vapor (lower methane means lower water) but also on the ozone distribution, with the increased methane limiting the impact of ozone loss; the reason for this behavior is being investigated. Similarly, a no-climate change simulation (blue curve) also indicates a much deeper ozone hole.

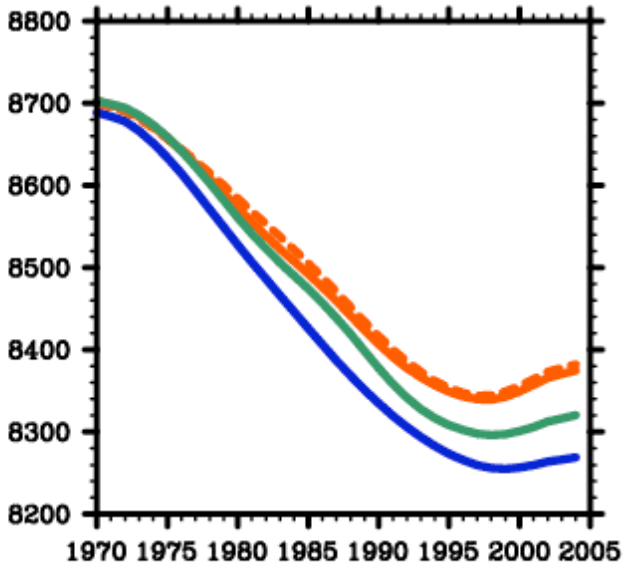

Figure 9. *Time evolution of the model simulated ozone mixing ratio (in ppbv), global and annual average, at 10 hPa. The base simulation is in red (solid and dash), the simulation with 1970 methane is in green and the simulation with 1970 CO_2 levels and climatological SSTs is in blue.*

IIIc. ***Fast atmospheric chemistry***

For the first 6 months of this project we have: (1) successfully extended our fast chemical mechanism to simulate the stratosphere as well as the troposphere for no extra computational cost, (2) received agreement from the CCSM chemistry-climate working group that inclusion of our fast mechanism into the CCSM trunk code is a working group

priority, (3) helped to validate DMS emissions from the 3D ocean interactive sulfur cycle models submitted to an international DMS intercomparison workshop by simulating the atmospheric distribution of sulfate generated by those emissions, which we compared with observations (see figure), and (4) developed a strategy to test the CLAW/GEIA hypothesis for DMS-sulfate negative climate feedback.

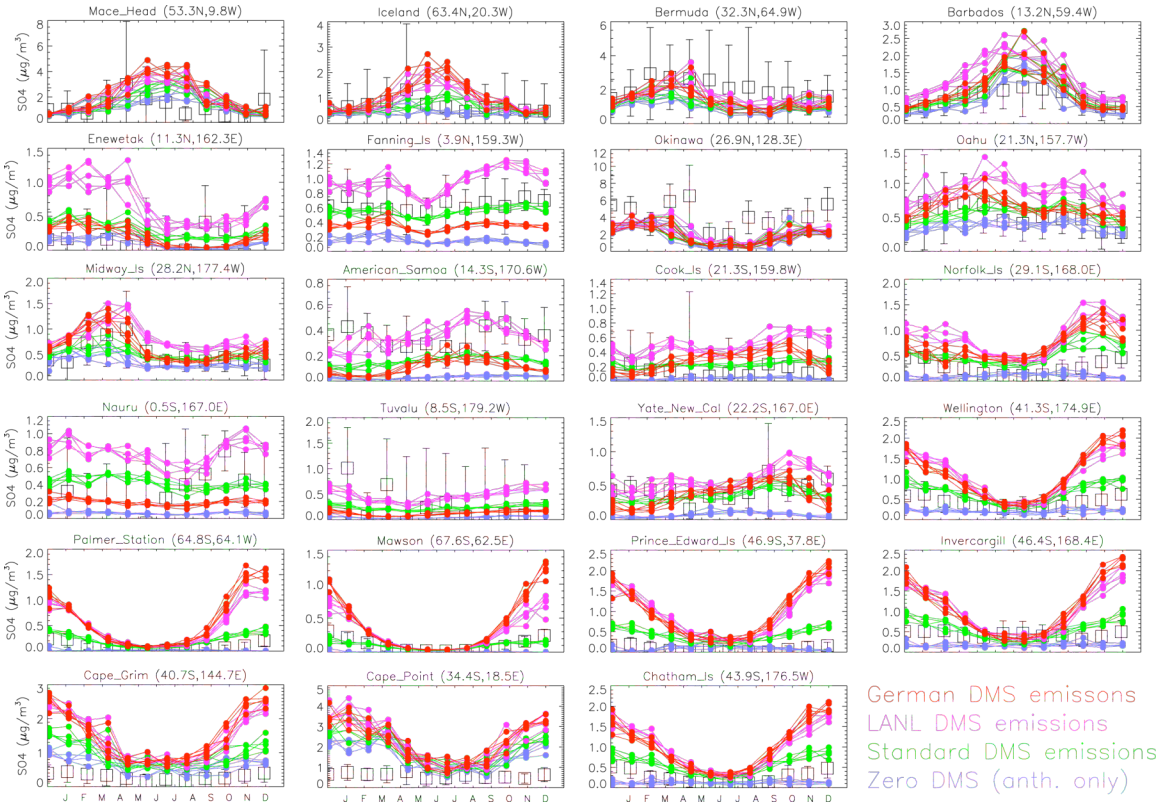

Figure 10. *Validation of DMS emissions for the two 3D interactive ocean sulfur models participating in the international interactive DMS intercomparison workshop, using the CAM atmospheric fast-chemistry model and observed sulfate. Five-year simulations were run using the monthly mean emissions from each of (1) the Hamburg interactive ocean sulfur cycle model, (2) the LANL interactive ocean sulfur cycle model developed as part of this SciDAC project, (3) the standard monthly mean emissions currently used in CAM, and (4) zero DMS emissions that shows the effect of anthropogenic sulfur emissions to the simulations. The DMS emissions from the interactive ocean cycle models generally produce sulfate that compares well to observations in the northern hemisphere and tropics, but over estimates sulfate over the southern ocean. This may be the result of the standard ocean-atmosphere piston-velocity parameterization over-estimating DMS emissions for high wind velocities.*

IIId. *Aerosol Indirect Effect*

The influence of aerosols on clouds through their role as cloud condensation nuclei (CCN) is an important radiative forcing mechanism that is neglected in the CCSM. To correct this weakness we have applied a parameterization of cloud droplet nucleation to a

prognostic treatment of droplet number in the CCSM. This treatment is expressed in terms of the aerosol in the model, so that as aerosols change with emissions, the sensitivity of clouds and cloud radiative forcing to the aerosols can be simulated. We have performed simulations for present day and pre-industrial conditions to estimate the impact of the aerosols on the planetary energy budget both directly (through scattering and absorption of solar radiation) and indirectly (through their role as CCN). The estimated indirect forcing due to anthropogenic sulfate is -2.8 W m^{-2}, even if the influence of droplet number on precipitation (the second indirect effect) is neglected. Such an estimate is much larger than the estimate (-0.36 W m^{-2}) by a different model called MIRAGE that uses the same treatment of droplet nucleation but a completely different (and probably more realistic) treatment of aerosols and a slight different treatment of clouds. To reconcile the different estimates, we have first noted that the CCSM treats the dependence of droplet sedimentation on droplet size, while MIRAGE does not. If that dependence in the CCSM is turned off, the indirect forcing estimated by CCSM is smaller: _2.0 Wm^{-2}. Recent work by others suggests that the sensitivity of the cloud radiative forcing to droplet sedimentation in the CCSM has the wrong sign, so that the smaller estimate is more likely to be correct than the larger estimate. Much of the remaining difference in the estimates of indirect effect by CCSM and MIRAGE is due to the difference in the aerosols. When the monthly mean aerosols from MIRAGE are used in the CCSM, the indirect effect estimated by the CCSM -0.9 W m^{-2}, which is comparable to the indirect effect estimated by MIRAGE with the same monthly mean aerosols: -0.6 W m^{-2}. This suggests that much smaller estimates of aerosol indirect effect by CCSM are likely if the treatment and representation of aerosols in MIRAGE are applied to the CCSM. We are presently in the process of applying the MIRAGE treatment to the CCSM, so that the MIRAGE representation can be used online rather than just offline.

IIIe. ***Marine biogeochemistry***

A complete, global ocean sulfur simulator has been coded and tested which includes 1) production of the precursor species dimethyl sulfoniopropionate (DMSP) within multiple phytoplankton classes, 2) regulation of the intracellular S content by ultraviolet and other stress types, 3) release via grazing and lysis, 4) microbial processing to dimethyl sulfide in the water column, plus 5,6,7) bacterial, photolytic and sea-air transfer losses. All the required parameterizations have been based on details distilled from existing low dimensionality models. Surrogates were employed to represent the yield of DMS from propionate consumption, and for the effects of stratification and ultraviolet dose upon succession and composition. The global code was able to simulate key features of the accepted Kettle climatology including small peaks along the equatorial divergence, evenness moving from the central gyres to middle latitude frontal systems, the very counterintuitive rise in DMS which occurs as chlorophyll levels drop in summer, and polar maxima associated with specialist prymnesiophytes.

Progress on POP sulfur was reported to a joint meeting of land, biogeochemistry and chemistry working groups held at NCAR in late winter 2007. Results of this development process were also submitted to an international sulfur model intercomparison exercise sponsored by the Surface Ocean Lower Atmosphere Studies group, or SOLAS. The

project was given the acronym CODiM, for Comparison of Ocean Dimethyl Sulfide models. The POP entrant fared quite well during ensuing meetings. Only three global dynamic DMS models are currently available around the earth system science community. From among them, SciDAC results were the only ones to fully capture decoupling from chlorophyll. Seasonal correlations with the major data set are offered in Figure 11. The Department of Energy model is labelled POP, and its European competitors are PIS for a version of Pisces currently running at the University of East Anglia, and HAM for the Hamburg global carbon cycle code. Red tones indicate correlation values near one, or equivalently, local temporal agreement with the climatology. All models reproduced the total annual flux of 30 Tg S.

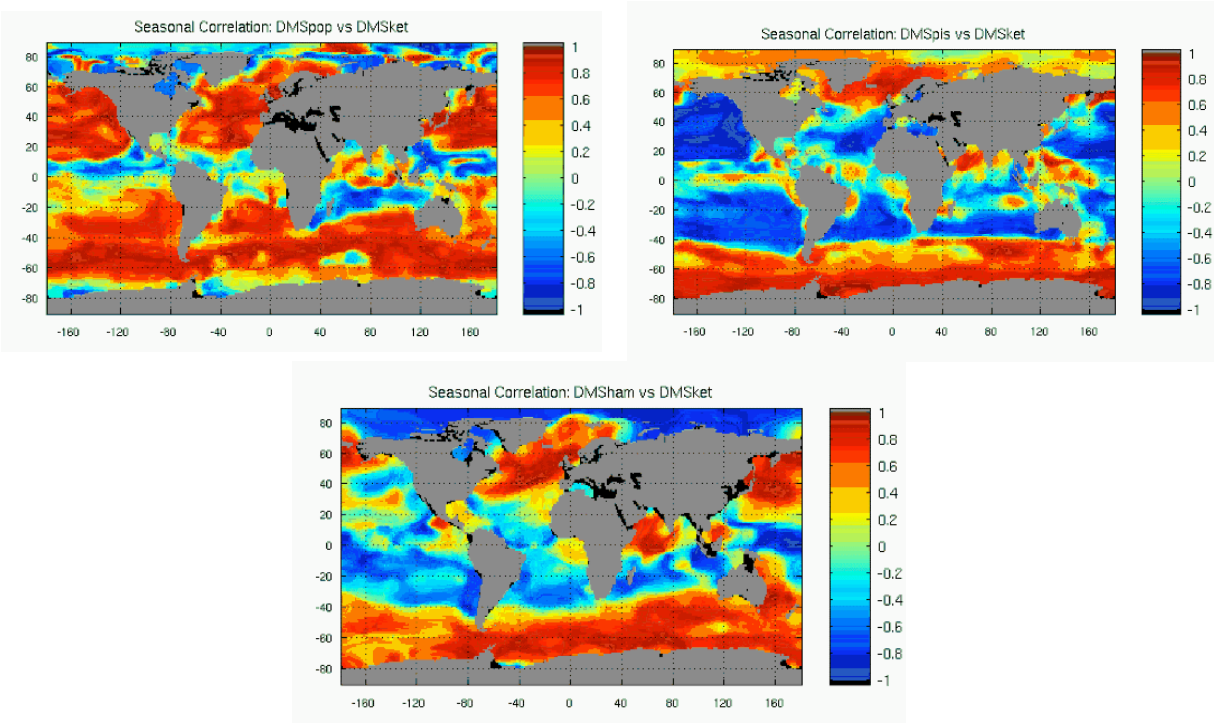

Figure 11. *Seasonal correlation of DMS with the Kettle database for POP (upper left), Pisces-East Anglia (upper right) and Hamburg (lower).*

At the recent combined working group meetings (land, chemistry, biogeochemistry), much interest was expressed in the role ammonia may play in determining acidity and hygroscopicity of the marine tropospheric aerosol. Both the standard ecology and trace gas simulators carry ammonium (NH_4^+) a recycling agent. SciDAC has recently augmented POP with a fast pH simulator and volatile ammonia concentrations/fluxes (NH_3). Preliminary results are shown in Figure 12. Pacific-centric distributions are shown for dissolved ammonium at just below the sea surface, and for gas phase NH_3 which fluxes into the boundary layer to be removed by aqueous droplets. It is likely that during the period of global warming, increases in ocean acidity and stratification will conspire to inhibit the reduced nitrogen transfer.

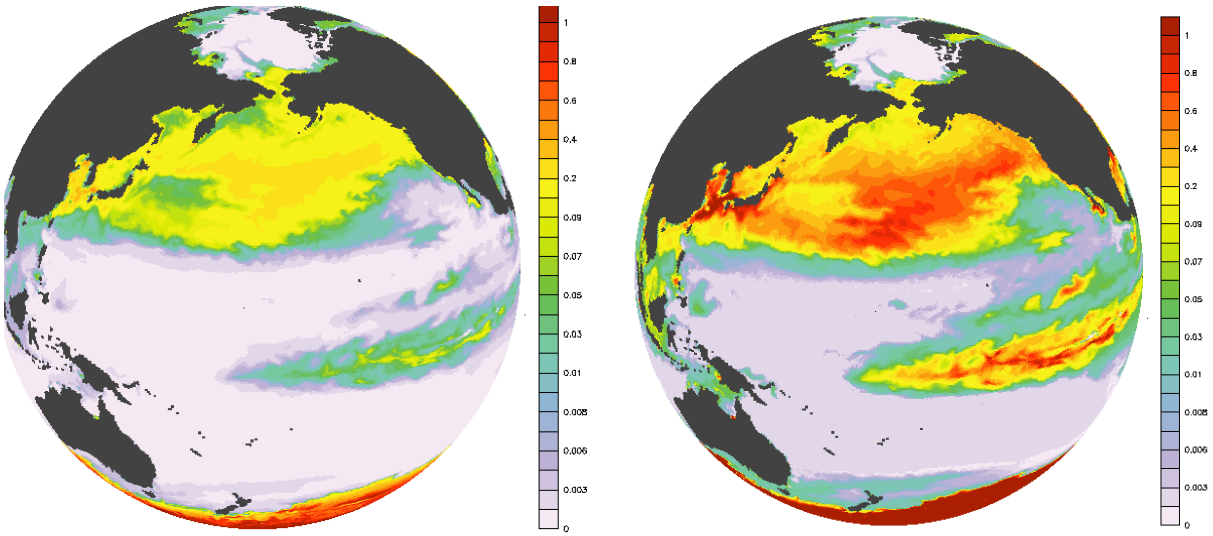

Figure 12. *Mixed layer ammonium and troposphere ammonia distributions simulated in the stand alone Parallel Ocean Program. Left panel gives dissolved NH_4^+ in units of micromolar, the right panel gas phase NH_3 in picomolar.*

IV. Summary of Personnel and Effort

The following table summarizes the personnel and level of effort at each laboratory participating in this SciDAC project. Contact drakejb@ornl.gov or pwjones@lanl.gov for further questions.

Name	Institution	FTE
Rob Jacobs	ANL	0.8
Ray Loy	ANL	0.8 (SAP)
Bob McGraw	BNL	0.2 (SAP)
Yangang Liu	BNL	0.2 (SAP)
Ernie Lewis	BNL	0.2 (SAP)
Wei Zhu	SUNY-Stony Brook	3 months
Tianyi Zang	SUNY-Stony Brook	Doctoral student
Ling Leng	SUNY-Stony Brook	Doctoral student
Scott Elliott	LANL	1.0
Phil Jones	LANL	0.5
Mat Maltrud	LANL	0.5
Bill Lipscomb	LANL	1.0
Philip Cameron-Smith	LLNL	0.6
Art Mirin	LLNL	0.7 (SAP)
Cathy Chuang	LLNL	0.1
Peter Connell	LLNL	0.1
Cyndi Atherton	LLNL	0.1
Michael Wehner	LBNL	0.3
J.-F. Lamarque	NCAR	1.0
Mariana Vertenstein	NCAR	1.0
Bill Collins	NCAR	0.5
Jeff Lee	NCAR	0.5
Tony Craig	NCAR	0.5
Warren Washington	NCAR	0.1
Peter Gent	NCAR	1.0 (NSF)
Pat Worley	ORNL	0.5 (SAP)
Forrest Hoffman	ORNL	0.5
John Drake	ORNL	0.8
David Erickson	ORNL	0.5
Marcia Branstetter	ORNL	0.2
Trey White	ORNL	0.2
Mac Post	ORNL	0.1 (BER)
Steve Ghan	PNNL	0.3
Xiaohong Liu	PNNL	0.6
Richard Easter	PNNL	0.3
Mark Taylor	SNL	0.7
Bill Spotz	SNL	0.2

V. Publications (FY07)

1. Wang, Tribbia, Baer, Fournier and Taylor, *A spectral element version of CAM2*, to appear, MWR 2007
2. M. A. Taylor, *Asymmetric cubature formulas for polynomial integration in the triangle and square,* to appear, J. Comput. Appl. Math. 2007
3. Giraldo and Taylor, *Triangular Diagonal Mass Matrix Spectral Elements based on Cubature Points*, to appear, J. Eng. Math. 2007
4. Kuhn, Gabriel, Shiraj Khan, Auroop R. Ganguly, and Marcia L. Branstetter, "Geospatial- temporal dependence among weekly precipitation extremes with applications to observations and climate model simulations in South America", accepted Advances in Water Resources, 2007.
5. J.B. Drake, P.W. Jones Developing Models for Predictive Climate Science SciDAC Review, Vol 3, pp. 44-56, (2007)
6. Schimel, David, William Hargrove, Forrest Hoffman, and James McMahon. March 2007. "NEON: A Hierarchically Designed National Ecological Network." Guest Editorial, Frontiers in Ecology and the Environment, 59.
7. Hoffman, Forrest. October 2006. "The Future of HPC." Linux Magazine, 8(10): 46-48.
8. Hoffman, Forrest, Inez Fung, Jim Randerson, Peter Thornton, Jon Foley, Curtis Covey, Jasmin John, Samuel Levis, W. Mac Post, Mariana Vertenstein, Reto Stöckli, Steve Running, Faith Ann Heinsch, David Erickson, and John Drake. September 2006. "Terrestrial Biogeochemistry in the Community Climate System Model (CCSM)." Journal of Physics: Conference Series, 46: 363-369. doi:10.1088/1742-6596/46/1/051.
9. Ghan, S. G., and R. A. Zaveri, 2007: Parameterization of optical properties for hydrated internally-mixed aerosol. *J. Geophys. Res*., doi:10.1029/2006JD007927.