ESDCD news

Summer 2004

Code 930 • http://esdcd.gsfc.nasa.gov

Earth and Space Data Computing Division

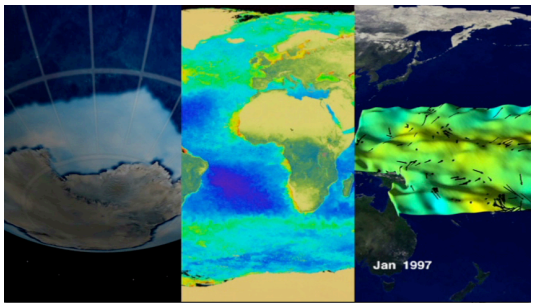
Earth Sciences Directorate, Goddard Space Flight Center

In This Issue

Scientific Visualization Studio

GSFC Video Featured in Computer Animation Festival

"The Edge of History," a 2-minute GSFC video production, was featured in the Computer Animation Festival for SIGGRAPH 2004, the 31st International Conference on Computer Graphics and Interactive Techniques. Selected for the festival's premier event, the Electronic Theater, the video includes numerous Earth science visualizations produced by groups at GSFC such as Public Affairs, the Visualization Analysis Lab, Earth Observatory, and ESDCD's Scientific Visualization Studio (SVS). This marks the first time an SVS effort has appeared in the Electronic Theater.



Images from "The Edge of History" (Image credit: SVS).

"The Electronic Theater is the most prestigious juried venue for computer graphics. It is the Academy Awards for our field," says SVS Manager Horace Mitchell. "We're in there with excellent talent—individual filmmakers, university computer graphics and art graduate programs, and the best film and television computer graphics companies."

Sponsored by ACM SIGGRAPH (Association for Computing Machinery/Special Interest Group in Computer Graphics), SIGGRAPH 2004 was held in Los Angeles from August 8-12 and brought together 27,825 computer graphics and interactive technology professionals from 6 continents. The Computer Animation Festival, an internationally recognized event, is considered by many to be the highlight of the conference. Pieces are selected for their creative merit in both computer-generated imagery and storytelling. This year, 83 projects were selected from a record 643 entities. Of the 83 selections, 30 were selected for the Electronic Theater.

Appearing alongside the likes of "Shrek 2," "The Lord of the Rings: The Return of the King," and "Spider-Man 2," "The Edge of History" illustrates how the Earth science disciplines are crossing the threshold of a new era. By using scientific visualizations,

the video seeks to increase the public's understanding of and enthusiasm about Earth science by showing how space-based data collection plays a principal role in cutting-edge research.

http://svs.gsfc.nasa.gov
http://www.siggraph.org/s2004/

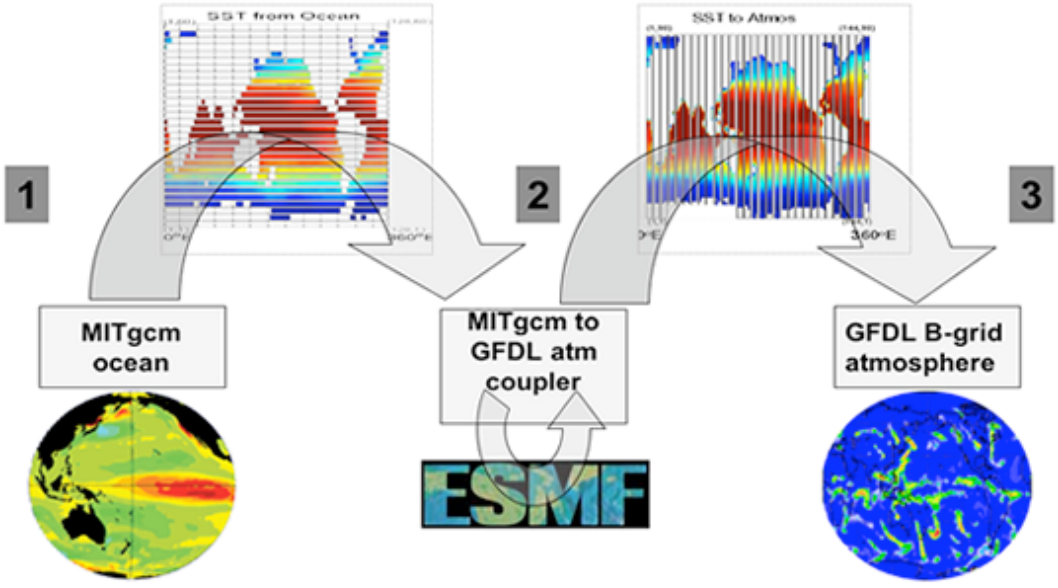
Computational Technologies Project

ESMF Version 2.0 Introduced at 3rd ESMF Community Meeting

More than 100 scientists and model developers from around the world met at the National Center for Atmospheric Research (NCAR) on July 15 to take the next steps towards unifying the weather and climate prediction communities through a common underlying software system. The event was the 3rd Earth System Modeling Framework (ESMF) Community Meeting, where the main topics of discussion were the newly released ESMF Version 2.0 software and early field tests using this software in real models. ESMF is being developed by a national partnership led by NASA and including other government agencies and research universities.

ESMF allows researchers to easily assemble complex weather and climate models by representing the models (atmosphere, ocean, land, etc.) as collections of smaller components that are coupled together. With ESMF, researchers have a standard way to add new capabilities and swap in different options—making it much simpler for them to exchange codes with other groups and institutions. Ideas can move quickly from scientist to code, from group to group, and from research to operations.

The ESMF Version 2.0 release supports modeling applications composed of hierarchies and ensembles of components. It includes software for representing and manipulating components, states, fields, grids, and arrays, as well as utilities for time management, configuration, and logging. The ESMF software comes with an exhaustive



A preliminary version of an ESMF field test couples ocean and atmosphere models that had never interacted before. Sea surface temperature data moves from the MITgcm ocean model run at 2.8-degree resolution to the GFDL B-grid atmosphere model run at 2-degree resolution. The coupler software manages changes in resolution and re-gridding from latitude to longitude orientation (Image credit: Chris Hill, MIT).

reference manual, a user's guide, more than 800 diagnostic tests, and more than 30 examples. It runs in Hewlett-Packard/Compaq, IBM, SGI, Apple Macintosh OS X, and a variety of Linux computing environments. As of this writing, Cray Inc. programmers are also running ESMF on the company's X1 supercomputer.

The Community Meeting included descriptions of ESMF use at GSFC's ESDCD, Global Modeling and Assimilation Office (GMAO), and Goddard Institute for Space Studies; the Massachusetts Institute of Technology; NCAR; NOAA's Geophysical Fluid Dynamics Laboratory and National Centers for Environmental Prediction; and the Naval Research Laboratory-Monterey, which is among several Department of Defense entities that have joined the ESMF effort during the past year. Notably, the GMAO is building its next-generation GEOS-5 atmospheric general circulation model from scratch using ESMF and recently ran the first multi-year test simulation with it. ESMF collaborators also showed preliminary versions of three field tests that couple atmosphere and ocean models from different institutions in entirely new configurations, demonstrating the power of ESMF to forge new collaborations. These new models are still being checked and validated.

http://www.esmf.ucar.edu/
(choose "Downloads & Documentation")

Data Drives Land Surface Modeling at GSFC

When the Land Information System (LIS) investigation set out to run the world's first 1-kilometer global land surface model, they quickly realized that a shared supercomputer center could not meet their data needs. A global run at that resolution, as completed this past July, produces 600 gigabytes of output *each simulated day*.

A 200-processor Beowulf cluster enables the Land Information System (LIS) to model the globe at the same resolution as NASA satellite observations (Photo credit: Yudong Tian, GSFC).

"At 1 kilometer, we can't get the data over there and back fast enough," said GSFC hydrologist Christa Peters-Lidard, who is LIS project manager and co-principal investigator with GSFC's Paul Houser. "You can buy the biggest computer in the world, but if you can't connect the computer to the data, it is not useful."

To serve such a data-intensive application, the LIS team built their modeling program around a $100,000 customized Beowulf cluster with 200 processors. Several hardware and software innovations have empowered LIS "to model the land surface at the scale of NASA observations—from current platforms like MODIS and TRMM to future platforms like HYDROS and GPM," Peters-Lidard said. With this capability, LIS realistically predicts the water and energy cycles, including runoff, evapotranspiration from plants and soil, and heat storage in the ground.

The LIS computing cluster is fine-tuned for high-resolution land surface modeling. Most Beowulf clusters have one or two head nodes to manage jobs. LIS increases that to eight head nodes of two processors each to better handle data flow, especially input and output, or I/O. "For parallel I/O, it gives us

huge data throughput," said Yudong Tian, assistant researcher at GSFC.

LIS is flexible, mixing data from satellites, ground stations, and atmosphere models in a variety of formats. Input starts with data assimilation. Parameter data include static observations of soil type and depth as well as topography. Forcing data encompass land cover, vegetation, and meteorological information. A specialized server converts the data into the same format and supplies them to the cluster.

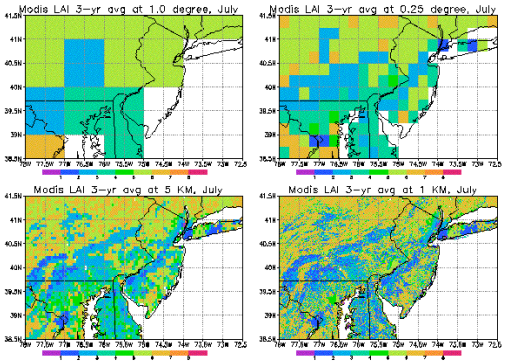
Before the cluster receives the data, programmers must set up a model run. Borrowing nomenclature from a children's song, the LIS job management system is known as the "farmer." The farmer divides the Earth's land surface into 1,200 pieces, called "bones," and throws them to "dogs," the cluster's 192 compute nodes. "The requests come in scattered," Tian said. "Dogs don't have a chance to fight over a single job. They come to the farmer, and he knows what bones to give out."

LIS moves the assimilation data onto the land pieces through a peer-to-peer approach, "like the KaZaA file sharing service on the Internet," Tian said. Using software, LIS first makes eight copies of the assimilation data, one for each head node. Next, the software splits the copies into small chunks and randomly sends them to the compute nodes. These nodes immediately start swapping among themselves until each node gets a whole copy of the data.

"The computer couldn't handle the request if we had to download from the head nodes," Tian explained. As parameter data do not change often, they remain on the compute nodes. By contrast, the forcing data get updated every 3 hours. With the need to input 20 to 30 files the size of a DVD movie, peer-to-peer maximizes computing power.

To pull the model output from the cluster, the team tried writing to the head nodes, but they soon got swamped. After some tinkering, they found success in letting the compute nodes output the data onto their own hard drives. This scheme works because the nodes process their pieces of the globe separately, with no need for data traffic between them. "We developed a system that could treat all the hard drives as one big disk," Tian said, with a total of 48 terabytes. "We can take data directly off the disks on demand for better performance," he added. "It can serve the user 10 times faster."

These speeds allowed the LIS team to exceed the requirements of their Computational Technologies Project milestone. They had to model the entire globe at 1-kilometer resolution fast enough to compute 1 day within 24 hours of wall clock time. LIS can now simulate 3 or 4 days in 24 hours. High performance also makes it possible to run three land surface models at once with the same input. The LIS models solve the equations in slightly different ways for "a statistical sampling of alternate realities," Peters-Lidard said. LIS currently incorporates the Community Land Model, the Variable Infiltration Capacity Model, and the Community



LIS can simulate the global land surface at various spatial resolutions, up to 1 kilometer. These visualizations of Leaf Area Index (LAI) data show more and more details being revealed as the resolution increases from 1 degree (~100 kilometers) to 1 kilometer (Image credit: Yudong Tian, GSFC).

NOAH[1] Land Surface Model.

Having passed the performance milestone, the LIS team is running 15-year retrospective studies on 100-square-kilometer sites around the world, validating model output with observations. On broader scales, they are exploring a diverse group of applications.

Since land surface feedback to the atmosphere affects weather and climate patterns, LIS is being coupled to several important atmosphere models. NASA's Earth Science Technology Office is funding couplings to the Weather Research and Forecasting Model and the Goddard Cumulus Ensemble Model using ESMF Version 2.0 (see "ESMF Version 2.0 Introduced at 3rd ESMF Community Meeting" in this issue). Researchers at the National Centers for Environmental Prediction are testing LIS with their operational models, including those used to generate the daily weather forecasts.

"Coupling to Earth system models is an important goal, but the land surface energy and water cycles are of interest to a number of groups because that is where we live," Peters-Lidard stressed. A project with the Bureau of Reclamation is focusing on the Rio Grande area of the American Southwest. "We want greater understanding of the things LIS predicts, such as snowpack and evaporation, which can help them better manage water resources," she said.

For air quality studies, collaboration with the Environmental Protection Agency will add a model of atmospheric emission and deposition of ammonia to LIS. Because LIS predicts soil moisture, temperature, and various aspects of vegetation, it could prove useful for planning crops. LIS also has potential for military applications. The U.S. Army is funding a special 30-meter version of LIS to model the impact of soil moisture on troop mobility.

http://ct.gsfc.nasa.gov/
http://lis.gsfc.nasa.gov/

[1] **NOAH** is an acronym for **N**ational Centers for Environmental Prediction, **O**regon State University, **A**ir Force, and **H**ydrologic Research Lab.

NCCS Highlights

System Metrics at the NCCS

A lot goes on behind the scenes at the NASA Center for Computational Sciences (NCCS) that most users never see but that is important to the effective management of the Center. Information is gathered and analyzed each month on the usage of the computing and storage facilities. This information allows NASA management to track trends in usage, predict the impacts of system changes, and plan for High End Computing (HEC) and storage needs for the Earth and space science community.

Figure 1 shows the increase in usage of the largest NCCS HP/Compaq system over the past year and a half.

The maximum number of processors available for batch processing is 1,292, so the system utilization consistently exceeded 90% from November 2003 through February 2004. Large parallel systems like this one are commonly considered to be saturated when they reach 85-90% utilization. Be-

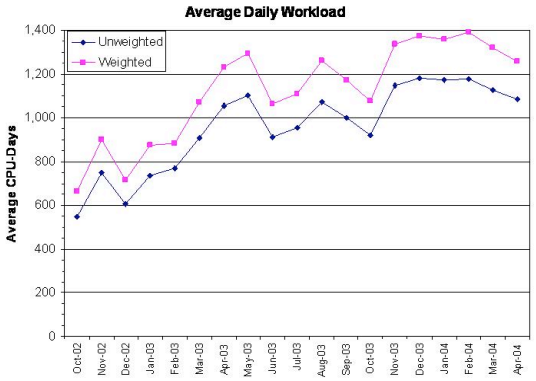


Figure 1. NCCS Average Daily Workload vs. Average CPU-Days, October 2002-April 2004

cause the system is composed of processors with two different speeds, the weighted average can exceed 1,292 CPU-Days. The information gained from tracking the metrics at the level shown above is important for tracking the overall utilization of the various systems at the NCCS and for long-term capacity planning, but system metrics are also tracked at a much finer level of granularity to determine the usage of major organizations over time. The usage of the same system by organization for March 2004 is shown in Figure 2.

Many other breakdowns of the data are examined for a more complete perspective of how the systems are being used by hour of the day, day of the week, job size, workload, queues, and groups. This allows NASA management to identify trends or deficiencies in system utilization and take action based upon quantitative data. The NCCS has changed batch queue structures in response to trends identified by this analysis and provided feedback to major user groups to improve job scheduling and make better use of the resources. The information is also presented to NASA management and the NCCS Customer Board as part of the decision making process on requests for resource allocation. For example, as the full HP/Compaq system was being brought into production, the NASA Seasonal-to-Interannual Prediction Project had almost exclusive use of a large part of the system. As usage grew on the system, other groups' workloads were shifted through queues to better distribute the load. More recently the Goddard Institute for Space Studies (GISS) needed additional capacity to meet dead-

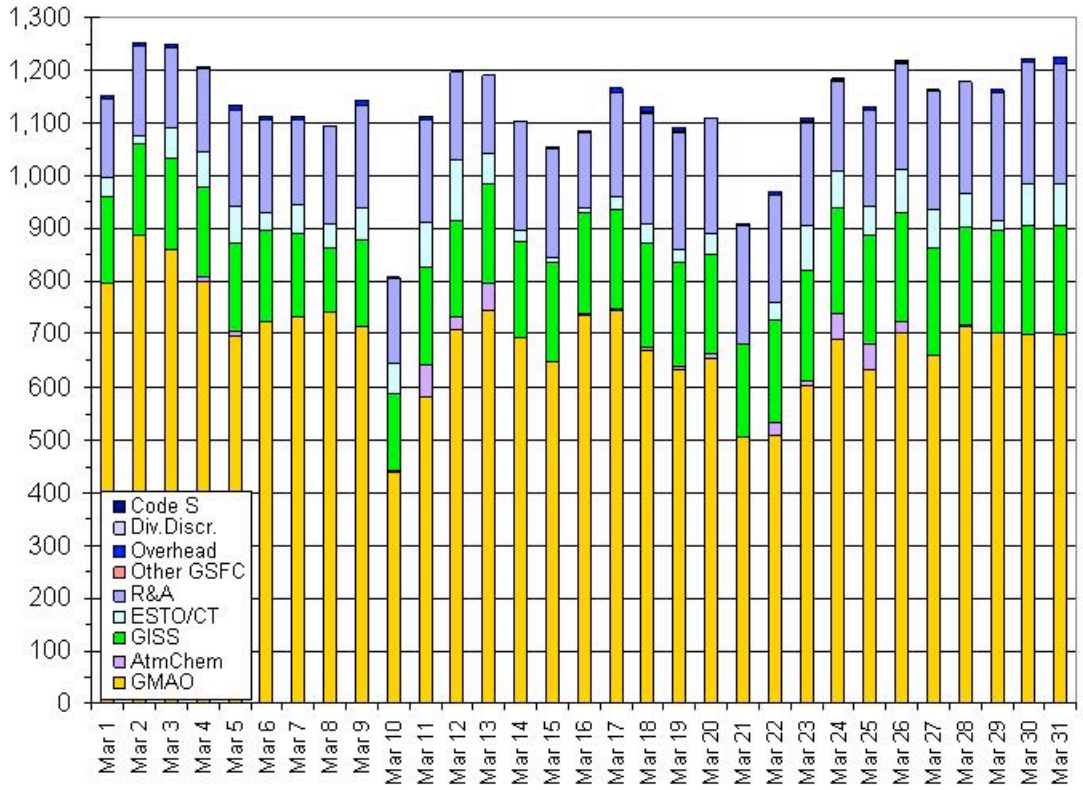


Figure 2. March 2004 Unweighted CPU-Day Utilization at the NCCS

lines for the Intergovernmental Panel on Climate Change. To help GISS meet its requirements, queue limits were adjusted to allow the Institute to use more resources on the system; its usage has more than doubled over the past three months.

The NCCS maintains a database of detailed system accounting information that allows reports to be generated down to the level of individual user batch jobs. Although this level of detail is rarely needed, when the NCCS rebuilt and reconfigured the systems this summer, the usage data by individual user was one of the criteria that determined the scheduled return to service of the user community. Similarly, detailed usage information also impacts management decisions during the Fiscal Year Initiation process. Questions about system accounting data should be directed to NCCS User Services at support@nccs.nasa.gov.

Project Columbia to Revolutionize NASA Supercomputing

On July 27, NASA announced a partnership with two major Silicon Valley corporations, SGI and Intel, to dramatically increase the Agency's supercomputing capacity to meet critical national goals. The three organizations have formed a unique, innovative partnership in which each is contributing resources and capabilities to the collaboration.

As part of 'Project Columbia,' NASA will integrate a cluster of 20 interconnected SGI Altix 512-processor systems, for a total of 10,240 Itanium 2 processors, to significantly increase the Agency's capability and capacity through creation of the 'Space Exploration Simulator.' The new machine will be based at NASA Ames Research Center in the heart of California's Silicon Valley. It will provide an estimated 10-fold increase in NASA's current supercomputing capacity. The primary purpose of Project Columbia is to revitalize NASA's supercomputing capability by deploying an integrated computing, visualization, and data storage environment tailored to the NASA mission.

Work using the initial 512-processor Altix at Ames led to major advances in modeling the Space Shuttle as well as in the Earth sciences in high-resolution modeling of the world's oceans. These advances demonstrated the power of simulation to significantly advance NASA's goals, and NASA's powerful new supercomputer will allow the Agency to continue this important work. The NCCS will be working with Earth and space science investigators to quickly port codes that can take advantage of this substantial shared resource and ensure that NCCS users will have sufficient network capacity to effectively access the system and their simulation results.

In line with Office of Science and Technology Policy recommendations, a portion of the new system will be made available on a broad basis to ensure that the nation's entire science and engineering community has access to this highly advanced supercomputer architecture.

Education and Public Outreach

MU-SPIN Awards Ceremony Held at GSFC

On Tuesday, May 25, the Minority University-Space Interdisciplinary Network (MU-SPIN) hosted a NASA Student Involvement Program (NSIP) Awards ceremony at GSFC for students that participated in the 2004 science competition projects. These students were from Flowers High School, Drew Freeman Middle School, and Duval High School in Prince George's County and from Gwynn Falls Elementary School in Baltimore City.

MU-SPIN NSIP Awards ceremony (Photo credits: Debora McCallum)

Each student from the participating schools received a Certificate of Recognition from MU-SPIN for his or her hard work and effort in submitting a science project to the NSIP review board. All teachers who facilitated the involvement of their students were also recognized at the ceremony and received a Certificate of Appreciation from MU-SPIN.

Speakers for the evening included representatives from MU-SPIN, the Elementary and Secondary Education Division Office of Education at NASA, the Prince George's County Workforce Services Corporation, and the Department of Computer Science at Morgan State University.

MESSENGER Launch Culminates MU-SPIN Program for Undergraduate Students

On August 3 at Cape Canaveral, Florida, NASA successfully launched the MESSENGER spacecraft on its 7-year journey to orbit Mercury. MESSENGER (short for MErcury Surface, Space ENvironment, Geochemistry) is the first spacecraft to return to Mercury since Mariner 10 made three brief flybys 30 years ago.

Joining the scientists and engineers celebrating MESSENGER's successful liftoff were student interns who, through the MESSENGER Education and Public Outreach (EPO) program, had the opportunity to be a part of this exciting mission.

ESDCD's Minority University-Space Interdisciplinary Network (MU-SPIN) program, a MESSENGER EPO partner, sponsored 14 minority students to work at Johns Hopkins University Applied Physics Laboratory, where engineers designed and built the spacecraft. The students, from colleges and universities in New York, North Carolina, South Carolina, and Texas, honed their technical skills and assisted mission project staff in areas such as spacecraft integration and testing, mission design, mission operations, mechanical support, and web site management.

August 3, 2004 MESSENGER spacecraft launch at Cape Canaveral, Florida (Photo credit: NASA).

The MESSENGER spacecraft serves as a backdrop for the summer interns (Photo Credit: JHU/APL).

"It was a really exciting moment for all who attended the launch, but especially so for MU-SPIN, considering that we actually contributed to the development of the spacecraft," remarked MU-SPIN Project Manager James Harrington. "[The students] knew that very few people had seen the spacecraft in person, and for that reason, it was extra special. It was particularly special to me, because it was a rare experience for underrepresented minorities."

http://messenger.jhuapl.edu/
http://muspin.gsfc.nasa.gov/

ESDCD Updates

ESDCD Scientist's Research Noted in *Science News*

ESDCD scientist Jan M. Hollis and colleagues' research on organic molecules in space was mentioned in the "Of Note" section of the July 24, 2004 issue of the prestigious newsmagazine *Science News*. Hollis noted in the article that evidence of the aldehydes propenal and propanal, which were identified by detecting specific radio wave frequencies emitted from the stellar cloud Sagittarius B2, will help researchers understand how molecular building blocks assemble into more-complex organic molecules in space. The note in *Science News* summarizes a more extensive research publication in the *Astrophysical Journal* (610:L21-L24, July 20, 2004).