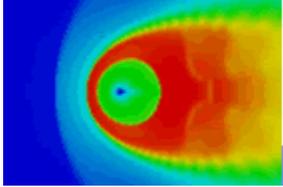




GODDARD SPACE FLIGHT CENTER

SEARCH NASA



ESDCD *news*

Earth and Space Data Computing Division
Earth Sciences Directorate, Goddard Space Flight Center

Code 930 • <http://esdcd.gsfc.nasa.gov>

Summer 2004 ESDCD News Home > Data Drives Land Surface Modeling at GSFC

Home

Archives

Earth and Space Data
Computing Division

Earth Sciences
Directorate

Computational Technologies Project

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*.

"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.

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)

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."

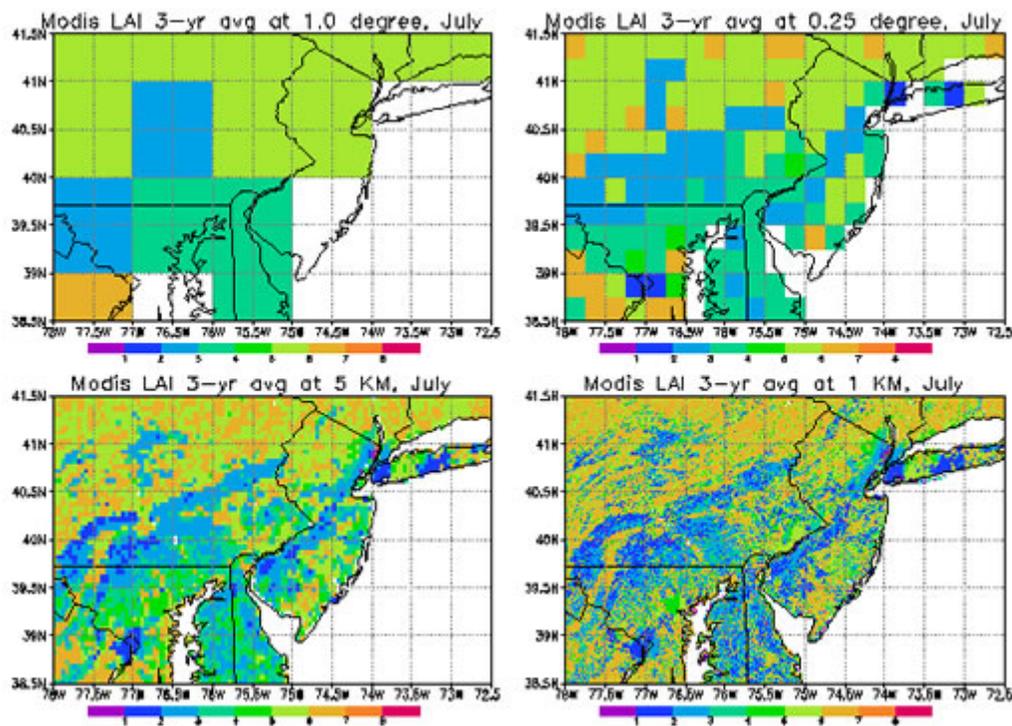
LIS moves the assimilation data onto the land pieces through a peer-to-peer approach, "like the KaZaA file sharing service on the Internet," Tian said. Using software, LIS first makes eight copies of the assimilation data, one for each head node. Next, the software splits the copies into small chunks 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 NOAH¹ Land Surface Model.



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

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.

| [Summer 2004 ESDCD News Home](#) | [Next Article](#) |



+ [NASA Privacy, Security, Notices](#)



Authorizing NASA Official: Dr. Richard Rood, GSFC Code 606
Curator: [Darcy L. Herman](#)
Masthead image credit: NASA/GSFC, SVS